The World Trade Web: Using Network Analysis and Machine Learning as Tools for Public Policy Decision-Making

By

Miguel Lozano

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Master of Science in Computer Science

Faculty of Business and Information Technology

University of Ontario Institute of Technology (Ontario Tech University)
Oshawa, Ontario, Canada
November 2020

THESIS EXAMINATION INFORMATION

Submitted by: Miguel Lozano

Master of Science in **Computer Science**

Thesis Title: The World Trade Web: Using Network Analysis and Machine Learning as Tools for Public Policy Decision-Making

An oral defense of this thesis took place on November 13, 2020 in front of the following examining committee:

Examining Comittee

Research Supervisor	Karthik Sankaranarayanan
Examining Committee Member	Amirali Salehi-Abari
Examining Committee Member	Stephen Marsh
Thesis Examiner	Amir Rastpour

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

ABSTRACT

The World Trade Web (WTW) contains a wealth of information that upon rigorous analysis can aid governments in public policy decision-making. In my attempt to provide this valuable input, this dissertation uses two main methods: weighted network analysis and machine learning. First, the topology of the WTW is explored, described, and analyzed. Secondly, the relationship between countries' trade network characteristics and their income is modeled. Lastly, deep learning is used to predict trade interactions between countries using quantitative, dyadic binary, and categorical variables. Insightful remarks are obtained: countries with higher PCGDP tend to associate with more neighbors that are themselves weaker, reciprocate fewer of their trade links, and trade more strongly with countries that are themselves stronger, and have a higher export to GDP Ratio. The improved trade forecasting model obtained can result in better GDP forecasts, which can aid with the optimization of tariffs, quotas, and subsidies.

Keywords: World Trade Web; Weighted Network Analysis; Machine Learning; Public Policy

CO-AUTHORSHIP STATEMENT

I hereby declare that I am the main author of this thesis and the papers included in the "Statement of Contributions" section. Dr. Amirali Salehi-Abari and Dr. Karthik Sankaranarayanan are both co-authors and have contributed with high level ideas on said papers. All of the writing, experimenting, programming, formatting, and testing have been performed by me.

AUTHOR'S DECLARATION

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

Miguel Lozano

STATEMENT OF CONTRIBUTIONS

The work described in Chapter 3 is being considered for publication under the title "The World Trade Web: Countries' Topologies and Their Effect on Income Level", where I am the main author. Dr. Amirali Salehi-Abari and my supervisor Dr. Karthik Sankaranarayanan are both valuable co-authors. Dr. Salehi-Abari has contributed with his expertise in network analysis and the structuring of ideas. Dr. Sankaranarayanan has aided in guiding the direction the project takes.

The work described in Chapter 4 is also being considered for publication under the title "The World Trade Web: A Deep Learning Approach to Link Weight Prediction", where I am the main author. My supervisor, Dr. Karthik Sankaranarayanan is my co-author, contributing with his extensive research and academic expertise.

DEDICATION

This thesis is dedicated to my family, friends, and girlriend, who provided emotional support; my supervisor Dr. Karthik Sankaranarayanan, for seeing talent in me and providing academic counselling, expertise, and funding; Canada, for welcoming me and providing an adequate environment for studying, and personal and professional development.

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. Karthik Sankaranarayanan for his professional, academic, and moral support. Knowing me for a short time, he saw potential in me and decided to support me in my academic journey towards my master's degree completion. Without him, this wouldn't have been possible. Dr. Amirali Salehi-Abari had a remarkable impact on my academic journey. He helped me continually improve my writing skills with his academic expertise and rigorous paper reviews. Ontario Tech University also played a major role by allowing me to cooperate through a teaching assistantship, which helped me sharpen my professional skills and supported me financially. The government of Ontario also played a crucial role by providing support through the Ontario Graduate Fellowship. Finally, I want to thank my parents, friends, family members, and partners for making this journey significantly more enjoyable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

WTW	World Trade Web
GDP	Gross Domestic Product
ML	Machine Learning
DL	Deep Learning
DNN	Deep Neural Networks
WNA	Weighted Network Analysis

Chapter 1

Introduction

The World Trade Web (WTW) has been studied widely in the field of economics, and there is vast literature that attempts to describe its topology. The importance of studying the WTW underlies in the fact that trade plays a crucial role in countries' economies, where to date 20% of the Global World Product comes from trade. Furthermore, for some countries trade is even more predominant within their Gross Domestic Product (GDP), where it could even surpass 100% of their GDP in like it does for Luxembourg, Hong Kong, Singapore, and a couple dozen other countries [1]. This phenomenon occurs given GDP is only the value that was added to products and services domestically, so small countries' exports can exceed the production within the country, as well as their imports can exceed national consumption. It's with a high degree of confidence that one can state that trade plays a crucial role in increasing interactions between countries, which in turn has been previously shown to accelerate globalization and increase interdependence [2].

Studying the WTW using network analysis has proven to be insightful in, but not limited to, the following cases: the analysis of globalization and regionalization in international trade [3]; understanding the potential and risks of economic systems [4]; empirically derive the structure of the world economy [5]; understand global interdependencies [6]; better understand the role of network characteristics in countries' incomes [7, 8].

Due to the aforementioned importance of the WTW, as well as the dependence of countries on others for raw materials, finished products, labor, technology transfers, and numerous other items, studying the WTW has become an attractive way for economist to better understand how trade shapes countries. Additionally, the recent COVID-19 pandemic has shown how supply chains are highly interdependent and countries prioritize the distribution of their supplies based on the diplomatic relationships they have with other countries. Hence, this could add to the importance of understanding the commercial connections between countries to understand this behavior. The bulk of the available literature pertaining to the study of the WTW from the network analysis perspective has been made without taking into account the magnitude of trade between the countries, and instead just take into account trade links in a binary way, where a trade flow either exists or does not, commonly referred to as unweighted network analysis. The metrics to study unweighted networks have been widely studied (see [9, 10]). Using unweighted network analysis to study the WTW can lead to a massive loss of information, hence there has been a recent movement towards using weighted network analysis when analyzing the WTW, i.e. including the volume of trade. However, metrics for the study of weighted networks are more novel, and a vast variety of them with different uses have been proposed (see [11–20]).

The motivation for this dissertation arises from the lack of in-depth weighted network analysis in the literature, where the bulk of it focuses on describing its topology and not on the impact of countries' trade network characteristics in their development. Furthermore, other authors have used a less comprehensive database of just 159 countries, neglecting numerous African countries and small island developing states, hence not finding the true topology of the WTW when analyzing it. I have a found a more comprehensive database, reported by UN Comtrade [21], which includes most of these African countries and small countries that were neglected in previous studies, for a total of 238 countries and territories.

In this dissertation, unweighted and weighted network analysis are used as a tool to address several concerns pointed out by Faioglo, Reyes, and Schiavo [7], such as the lack of in-depth analysis into the topological characteristics of individual countries and regions from the cross-sectional perspective, as well as the lack of studies in the role of geographical proximity in shaping the WTW to determine how fragile the network is. Stemming from this void, an attempt is made to answer the following questions: How do the WTW's trading communities look like? Which continents are more susceptible to instability originating from their trade partners that could spread through trade? Which countries are the most central? Which countries have a high dependency on others? Do countries trade with partners similar to them? Does geographical proximity and trade agreements influence the relative intensity of trade among countries? What actions (pertaining to trade) can countries take to improve their GDP? What continents are the major players in trade? How do continental flows look like?

Finally, machine learning is used in an attempt to predict trade flows in the World Trade Web using a feed forward deep neural network (DNN), and building on the gravity model of trade, as defined by Isard [22]. The reasoning behind the use of a DNN is to improve upon the prediction accuracy and lowering variance. This is another contribution of this thesis when compared to the work by Rose [23] and Head [24]. The same datasets and variables are used, with the difference being the methodology, where they use Ordinary Least Squares Regression (OLS) and in this thesis a a feed forward DNN is used instead.

This thesis is structured as follows: Chapter 2, section 2.1 explains the definition and measurement of the distinct unweighted and weighted network metrics used in section 3 (Chapter 3). Section 2.2 goes through the basics of machine learning and the reasoning behind the model chosen for section 4. Chapter 3 analyzes the WTW using both unweighted and weighted network analysis, and then finds the impact of countries' topologies on their income in order to match network trends with income levels. In chapter 4, deep learning is used to predict trade magnitudes in the WTW more accurately than the current status quo models. Chapter 5 summarizes and concludes the thesis, supported by a discussion in section 5.1 and the limitations and future work in section 5.2.

Chapter 2

Research Methodologies

2.1 Network Analysis

Networks are structures that surround our everyday lives and shape the way we live. Simply put, a network is a collection of nodes that have relationships with other nodes, and these relationships are represented via edges. Some examples of networks that are around us are: the internet, where a computer or router is the node and a cable or wireless data connection is an edge; the world wide web, where a web page is a node and hyperlinks are edges that connect web pages with one another; citation networks, where a node could be a legal case, patent or an article, and the edge is the citation, that refers one resource to another; a friendship network, where people are nodes and friendships are the edges; airports, where a node is an airport and the edges are the flights that connect the airports. Networks are studied in numerous fields: in mathematics, networks are known as graphs; in physics, nodes and edges are called sites and bonds respectively; in sociology, nodes and edges are referred to as actors and ties. In the introductory portion of this section, the mathematics and unweighted metrics of networks are presented, based on Newman [10].

It is relevant to be familiar with the notation used in the mathematics of networks, where n denotes the number of nodes in a network, and m for the number of edges. The

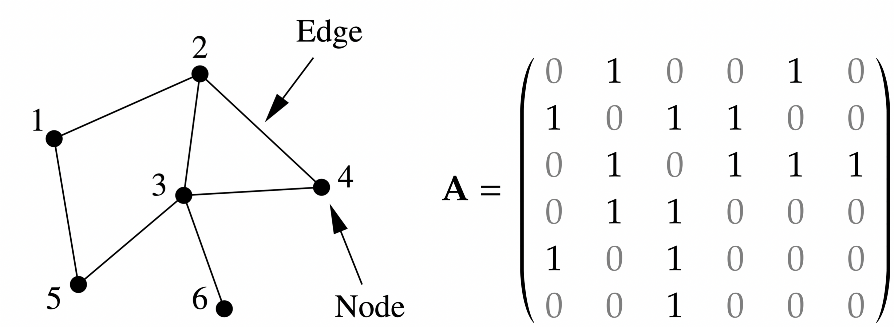


Figure 2.1 Example of Adjacency Matrix.
Retrieved from *Networks* in Oxford University Press, available in [10]

basic mathematical representation of networks is known as an *adjacency matrix*. Adjacency matrices are of dimensions n x n, where each element of the matrix is represented by A_{ij}, where the value is equal to 1 if there is an edge between nodes i and j, and 0 otherwise. An example of a network and its corresponding adjacency matrix can be seen in figure 2.1.

The most common representation of networks is using binary edges between nodes, where the connection either exists or doesn't. This is commonly referred to as unweighted network analysis in the literature. This type of network is useful for cases where it proves challenging to assign a magnitude to a relationship. For example, in a social network of friends, people are the nodes, and edges represents whether two people have met before or not, it's challenging to assign a weight to the edge or connection. The edges either exist or they don't because there are only two possible cases: two people have either met or they haven't.

For the cases where it's relevant to represent a weight, intensity or magnitude in the edges or connection between nodes, weighted network analysis comes into play. In the context of this thesis, in chapter 3.1, the World Trade Web is analyzed, and given that countries have trade relationships of varying magnitudes, it's relevant to include them in the analysis, hence using the weighted network analysis approach. Failing to include the weights in the analysis could be detrimental to the quality of the analysis due to loss of information. In this case, the user would be assuming that the trade flows between the USA and China are equally as important as the ones between Lithuania and Estonia, which could easily be argued inadequate.

Another important aspect of networks to understand is that they can be either directed or undirected. There are some cases where representing the direction of the flow could add important information to the network and allow for a more comprehensive analysis. In a directed network, edges have directions or flows, where the relationship flows from one node to another one. An example of directed network would be a network of people comprising of investors and entrepreneurs, and the edges between them represent investments. In that case, the edges will be flowing from the investors to the entrepreneurs, which in graphs is represented as an arrow that points from the investor that invests the money to the entrepreneur that is receiving it. In the case of the work presented in chapter 3, the network is directed and weighted, given that trade flows go from an exporting country to an importing country, and there's a magnitude associated to the value of the goods that are flowing between the territories.

Different types of networks require distinct approaches in order to be properly analyzed, hence there are two main distinctions between networks that need to be made to determine how the analysis is to be performed, and they are: whether the analysis is more adequate by doing an unweighted network analysis or a weighted network analysis. While the metrics to analyze unweighted networks are well documented (see [9, 10]), weighted network analysis metrics have been developed and discussed by numerous authors (see [11, 15–20]) who have attempted to create homologous metrics to those in unweighted network analysis. The complexity of weighted network analysis metrics allows for discussions to arise on the adequacy of distinct metrics, see for example work around the development of the weighted clustering coefficient [11–14].

In unweighted network analysis, one of the most common centrality metrics is the *degree centrality* of a node. Centrality refers to which are the most important nodes in a network, however, importance can be defined in numerous ways. The degree centrality of a node is simply defined as the number of edges that are connected to the node. In a network of academic citations, the nodes would be researchers, the edges would be the citations, and the degree would be the amount of citations that each researcher has. Using degree, it would be rather simple to determine which authors are more "important" or more "central" to the network by looking at their degrees. It's common to denote the degree of a node i with k_i. The degree of a node that is part of a network that contains n nodes can be denoted as seen in equation 2.1. Keep in mind that for directed networks, each node will have two degrees assigned to it: outdegree and indegree, where the former refers to the edges that flow out of it into other nodes, and the latter to the ones that flow into itself. Degree centrality is a very standard and useful method, however, one of its main drawbacks is that it gives every single connection the same importance, and here is where *eigenvector centrality* comes into play.

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{2.1}$$

Eigenvector centrality is formed from the notion that a node's importance in a network is increased if it has connection to other nodes that are themselves more important, central, or powerful. For example, you might just know one person in the world, but if that person is Jeff Bezos (Amazon's CEO), then you are in a more influential position than any other mortal (arguably). Hence, eigenvector centrality assigns a score to a node proportional to the centrality scores of its neighbors. Eigenvector centrality's computation is shown in equation 2.2, where the metric is denoted by x_i for node i, where the centrality of this node is proportional to the sum of the centralities of its neighbors. The term k^{-1} denotes the constant of proportionality, and y stands for the nodes j that are neighbors of i. With this metric, nodes can have a higher centrality either by having many neighbors with a low centrality, or few neighbors with a high centrality.

$$x_i = k^{-1} \sum_{y} x_j \tag{2.2}$$

PageRank is another widely used centrality metric, where the score that a node receives by having an incoming edge from the neighbors is proportional to the centrality of the neighbor divided by their out-degree. This prevents the type of problem that could happen where an important node points to numerous nodes and they all receive a very high score just by being pointed at by an important node. The centrality can be defined as shown in equation 2.3

$$x_i = \alpha \sum_{j} A_{ij} \frac{x_j}{k_j^{out}} + \beta \tag{2.3}$$

Betweenness centrality is another measure of node importance, which measures to what extent a node lies on the path between other nodes. In other words, it measures the number of shortest paths that go through a node. The reason why betweenness centrality is relevant, is because those nodes can exert some level of power or control over the network given the amount of information that has to pass through them to reach other nodes. The formula to measure betweenness centrality is shown in equation 2.4 where n_i^{st} is 1 if node i is in the shortest path from node s to t and 0 if it does not.

$$x_i = \sum_{st} n_{st}^{i} \tag{2.4}$$

Clustering coefficient is another metric of importance. It allows to quantity network transitivity, where if person "a" is friends with person "b", and person "b" is friends with person "c", then person "a" and "c" have a higher likelihood of becoming friends than other two random nodes. In the cases where person "a" is actually friends with person "c" as well, we say that there is a closed triad, otherwise it's an open triad. In order to

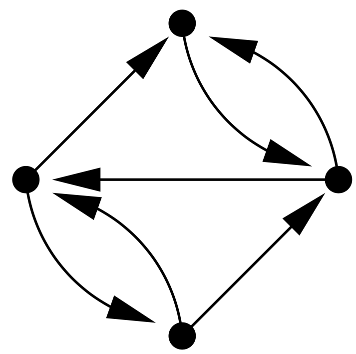


Figure 2.2 Reciprocity, available in [10]

calculate the clustering coefficient, the proportion of closed triads over the number of open triads is computed, as seen in equation 2.5

$$C = \frac{(closed\ triads)}{(open\ triads)} \tag{2.5}$$

Clustering coefficients quantify the loops of length three, however, in directed networks there can be paths of length two, whose frequency is measured by *reciprocity*, which is the chance that a node you are directing to, is directing towards you as well. Figure 2.2 shows a network where 4 out of 7 edges are reciprocated, hence reciprocity r is 0.57.

Homophily and assortative mixing are another notion of networks worth taking into consideration when analyzing a network. The fact that people tend to assimilate with others who are similar to them is known as homophily or assortative mixing in the context of network analysis. Disassortative mixing is also prevalent in social networks, and a good example of this is a marital network, where the majority of the partners are of opposite sex, so they are assimilating with someone who isn't like them. This can also be extrapolated to other networks like academic citation networks, where a paper in network analysis is more likely to cite other papers within the same field. When discussing assortative mixing, an important issue to take into consideration is that mixing can happen based on unordered characteristics (non-numerical characteristics where mathematical operations can't be performed) like nationality, gender, race, etc. or it could happen with ordered characteristics like income or age, which are mathematically treated differently. *Modularity* is the measure to quantify the extent to which, in a network, similar nodes are connected to other similar nodes with unordered characteristics, and is calculated as shown in formula 2.6, where $\delta_{g_i g_j}$ is the Kronecker delta. For the computation of assortative mixing by ordered characteristics, one can refer to Chapter 7, section 7.7.2 from Newman's "Networks" book [10].

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta_{g_i g_j} \tag{2.6}$$

As mentioned earlier in the chapter, weighted network metrics aren't as standard as the unweighted counterparts, hence sometimes there are various accepted metrics for the same notion of measure, proposed by various authors. In the remainder of the present chapter, the weighted network metrics used in chapter 3 are presented.

Weighted out-degree is the sum of the magnitude of all the outgoing edges from a node, and weighted in-degree is homologous but for incoming edges. In the context of the World Trade Web (WTW), the weighted out-degree of a country is its total exports, and the weighted in-degree is its total imports.

Random walk betweenness centrality (RWBC) is a homologous measure to betweenness centrality in unweighted networks. RWBC was developed by Newman [25] and Fisher and Vega-Redondo [26] and its computation can be explained intuitively with the use of signals. Random signals are sent though all the edges of the network, and each one of these has a target node. The signals then perform a random walk, where nodes with a higher weighted degree have a higher likelihood of being chosen as a route to the final destination. The algorithm keeps a track of how many signals go through each node, so the ones with the highest count will be the ones with the highest RWBC.

Random walk closeness centrality (RWCC) is a homologous measure to closeness centrality in unweighted networks. RWCC was developed by Stephenson and Zelen [27].

After its original proposal and development, it was reworked and improved by Brandes and Fleischer [28] by reducing its computational demand. The higher the RWCC, the more important the node is.

The weighted clustering coefficient is the weighted homologous to the unweighted clustering coefficient, and the one we use in chapter 3 is developed by Fagiolo [14]. This coefficient is the geometric average of the subgraph edge weights, which ultimately measures how likely a network is to create neighborhoods that are tightly connected.

In order to get an approximate measure of weighted assortativity, average nearest neighbor strength (sum of weighted out-degree and weighted in-degree of neighbors) is correlated to neighbor strength, in line with the approach of Fagiolo, Squartini, and Garlaschelli [29], and Fagiolo, Reyes, and Schiavo [7, 30].

2.2 Machine Learning

Machine learning (ML) refers to algorithms that enable computers to learn complex patterns from data, usually those that wouldn't be possible to model with other multivariate techniques. The study of ML can be traced to 1959 with the developments made by Arthur Samuel, who was a pioneer in the fields of artificial intelligence (which ML is a subset of) and computer gaming [31]. Recently, ML has developed closer ties to optimization, where the ultimate goal is to minimize a loss function, which can be defined in many ways depending on the use case, such as mean average percentage error, root mean squared error, accuracy, recall, among many others.

The general workflow of an ML algorithm is as follows: there is a training set that contains the true attributes of the variable (the object of prediction), as well as features that will serve as an input to generate that prediction. For example, if one wants to predict the price of a house using characteristics of the house such as its size in squared meters, latitutde, longitude, number of bathrooms, and other similar features, the data could be arranged in a relation similar to an ordinary least squares (OLS) regression. In this sample dataset, we could think as the price of the house as the dependent variable that we are trying to predict, and the independent variables would be the aforementioned features. The goal would be to minimize a loss function, where the optimal choice depends on the type of dependent variable under scrutiny, which in this case would be any loss function that allows the user to determine how far off you are from the prediction, such as root mean squared errors, mean squared logarithmic errors, mean average percentage error, and so forth. The user could experiment with different loss functions to see what's appropriate for the use case.

The above example falls under the subcategory within machine learning, known as supervised learning, which is used in this thesis. Supervised learning refers to those ML applications where there is a true, known value on what the user is trying to predict, in other words, there is a value for a dependent variable, a label for classification, or something of similar nature, that would allow the ML algorithm to determine how good its predictions are and fine tune its parameters to minimize the error. Additional examples on cases where supervised learning has proved its usefulness are: classification problems, such as identity fraud detection, image classification, object detection, medical diagnostics; and regression problems, such as weather forecasting, trade flow prediction, estimation of life expectancy, and a much longer list. An illustration of the process of classification of emails as spam or not spam using supervised learning is shown in figure 2.3

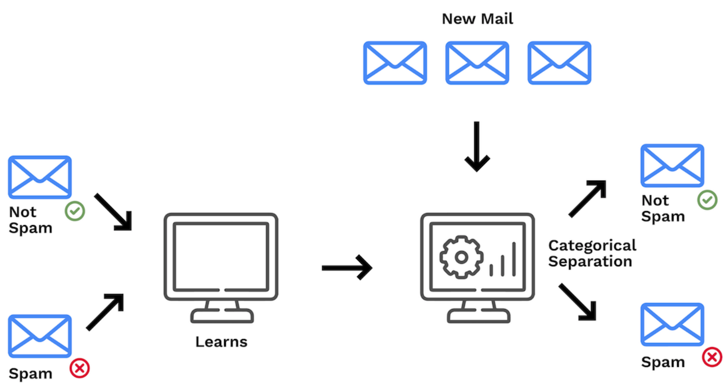


Figure 2.3 Example of Supervised Learning.

Another subcategory of ML algorithms that isn't used in this thesis, but is worth noting for a better context, is unsupervised learning, where the objective isn't prediction, given there is no ground truth, so data is unlabeled and uncategorized. Examples of use cases are: clustering algorithms (K-means clustering, non hierarchical algorithms, and the like), such as recommender systems and customer segmentation; and dimensionality reduction (principal component analysis, factor analysis, and the like) such as creation of indices, structure discovery, big data visualization, and numerous others. An illustration of the process of clustering for circles based on their color scale is shown in figure 2.4

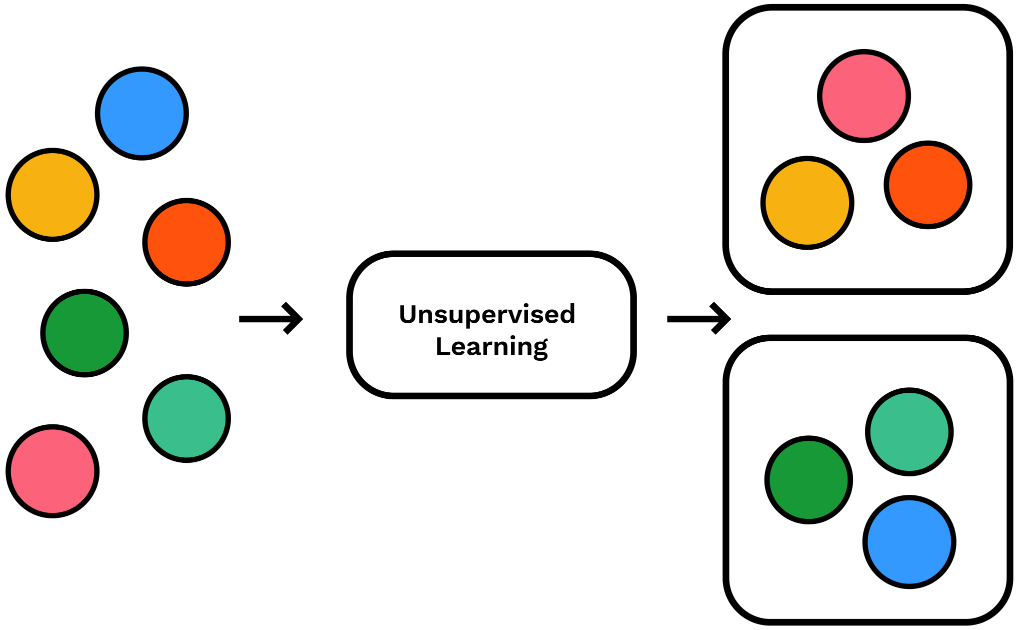


Figure 2.4 Example of Unsupervised Learning.

A last subcategory of machine learning that is worth noting is reinforcement learning. In such a structure, the algorithm learns by interacting with the environment, and there is a reward/penalty system that allows the ML algorithm to know whether it's doing well or not. A simple use case of such an algorithm, would be in a classic Mario Bros game, the computer is trying to learn how to beat a level, and to do so, it can perform certain actions (use all the buttons in a controller) and its results can be measured by either the final score, the time it took to complete the level, the health-points left, or any mix logical or weighted mix of the previous. Some industry use cases include robot navigation, AI in games, among various others. An illustration of a robot making decisions based on reward and penalty is shown in figure 2.5

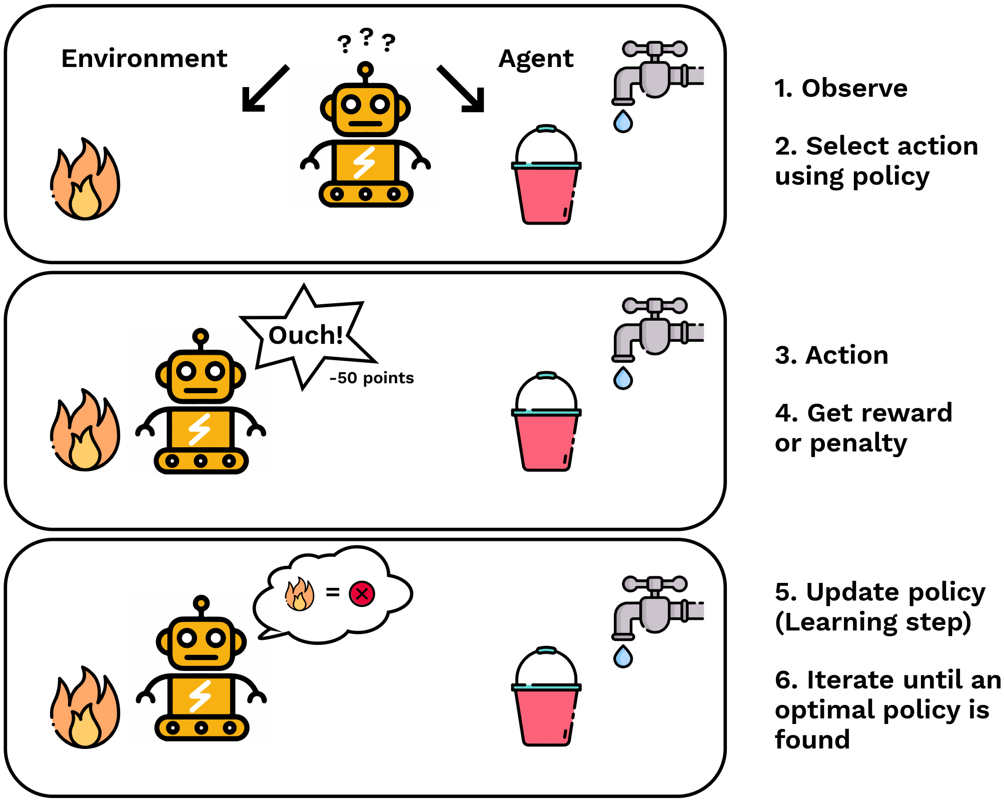


Figure 2.5 Example of Reinforcement Learning.

A helpful resource that aids in visualizing all of the 3 mentioned subcategories of ML: supervised learning, unsupervised learning, and reinforcement learning, as well as their use cases, are shown in figure 2.6.

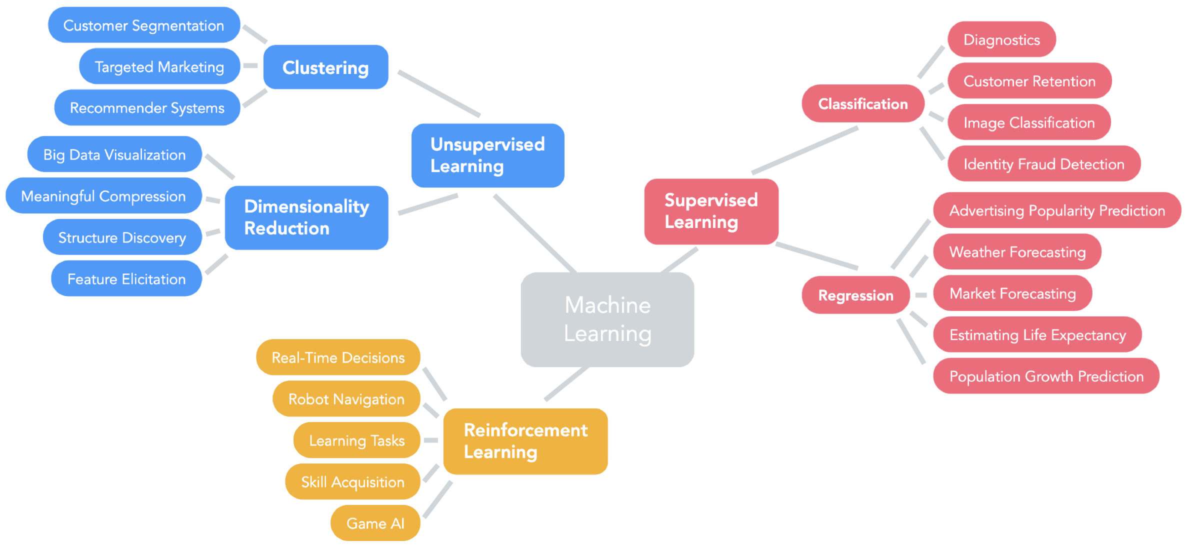


Figure 2.6 Machine Learning Categories.

Neural networks (NN) are the main tools used by ML algorithms for prediction. They have numerous key elements, but the most basic ones are neurons, weights, and biases, which would allow one to construct a simple NN. The first element, neurons, are functions that encapsulate the biases and weights within them, and when they receive inputs, they will process the data and very commonly use an activation function as well in order to limit the data to a particular range. The second element, weights, could be thought as the most fundamental elements of NNs, they are learned through trial and error in cycles called epochs, where the algorithm attempts minimizing the error of prediction through iteration. Once weights are assigned through learning (sometimes it can take very long through model training), they can be saved and loaded into a NN with the same architecture to predict with inputs. The third element, biases, are what the NN considers should be modified after finding the product of the weights with the data. The biases will be wrong, but through trial and error the model is able to learn the optimal biases [32].

If we think of a NN as a trivial linear function, the slope could be considered the weight of the model, the Y-intercept would be the bias, and the whole function would be the neuron. The previously mentioned elements can be observed in figure 2.7. It can be appreciated that the output layer has 4 neurons, which are represented by the circles, the lines connecting the neurons are the weights, and the bias is usually encapsulated within the neuron itself. Another important element that should be mentioned is that neurons are arranged in layers, where in figure 2.7 5 layers are present. One of them is the input layer, there are 3 hidden layers, and an output layer where the user gets the results.

Going back to the house prediction example. In this case, the input layer will have as many neurons as we have features for the prediction. So, if we have squared footage, number of bathrooms, latitude, and longitude, we will have 4 neurons in our input layer. The decision around number of hidden layers as well as the number of neurons, commonly known as hyper parameters is assigned via trial and error. The reason being, there is no way of knowing beforehand what values are appropriate hence the trial and error approach. Larger number of hidden layers and nodes allow for training of more complex models, but the trade-off is that it increases computational demand. The output layer, in this case, will consist of just one neuron, given we are making one house price prediction for every set of features. The edges in the NN will carry the weights, which will be automatically adjusted through trial and error with the objective of minimizing prediction error and maximizing accuracy. These can be measured because the data is labeled, so the model knows how far it is from making the right predictions, and the fact that the data is labeled makes the machine learning model one within the supervised learning category.

A term very often heard within ML is the tuning of hyper parameters. These parameters are adjusted through trial and error. This is achieved through either the modeler's prior experience or use case from literature or a mixture of both. It is common practice in ML algorithms to divide a dataset into training and validation data, usually a 75%-25% split is done. This allows to tune the hyperparameters using the training data, and then see the algorithms accuracy at generalizing the data that the model hasn't seen, using the

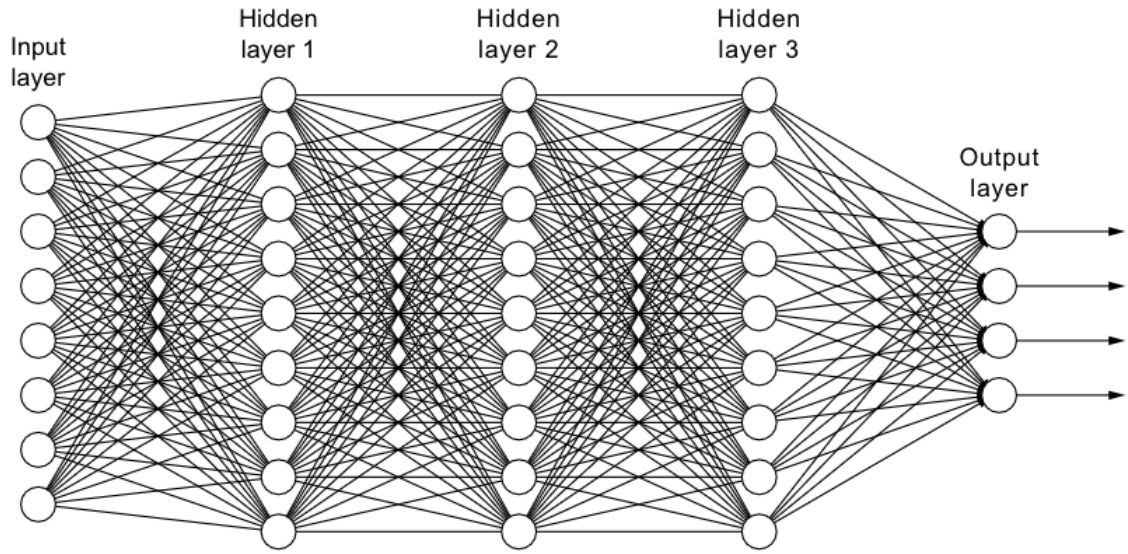


Figure 2.7 NN Architecture.
Retrieved from "*Probabilistic Deep Learning with Python*" in Manning, available in [33]

validation dataset.

Epochs is a hyperparameter that tells the model how many times the training data will pass through the neural network. In order to decide how many epochs to use in a model, it's common to use a method called early stopping, where you will train your model using the number of epochs that minimizes the error of the validation set. If training isn't stopped at this point, the validation error will start increasing, while the training error is still decreasing, a problem known as overfitting, where the model isn't good at generalizing anymore [34].

Batch size is a way of dividing the dataset into sub batches that the network is going to use to pass them and adjust the weights and biases. The learning rate will tell the model how big the adjustments will be every time it adjusts the weights and biases based on the loss function and the batch that is being passed. Having a very high learning rate could make the accuracy of the model vary greatly, given the model is "taking higher risks" when making the parameter adjustment during training. The aforementioned effect can be seen in figure 2.8

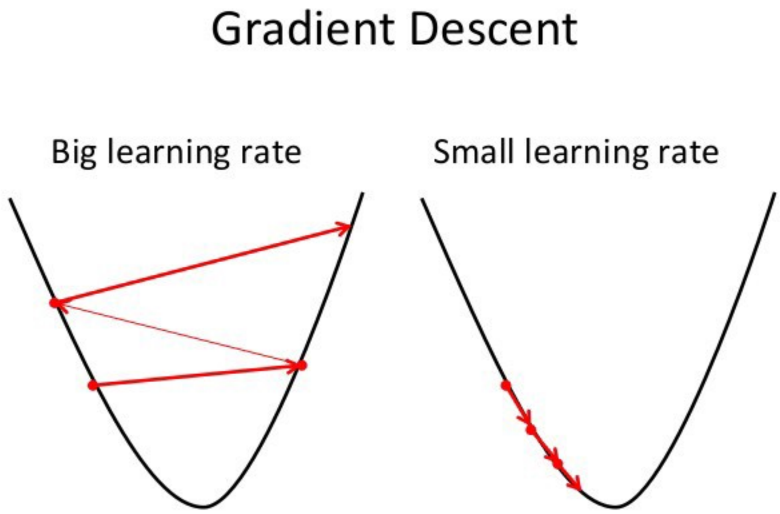


Figure 2.8 Learning Rate.
Retrieved from "*Introduction to Neural Networks and Their Key Elements (Part-B)*" in Manning, available in [34]

Activation functions are hyperparameters that transform the output of the neuron by constraining them to a set of values, because otherwise they could range all the way from negative infinity to positive infinity, thus complicating the training of the model. There are numerous activation functions, amongst which the most common ones we find: tanh, that constrains values between -1 and 1; sigmoid function, constraining values between 0 and 1; softmax function, which is used when dealing with a classification problem and constrains the classes into a probability distribution, where the sum of the probabilities is equal to 1; relu (rectified linear units), which increases in a linear fashion for all positive values, and is zero for any negative values [35].

Lastly, another important topic in ML and NNs, is preventing the model from overfitting. The first thing to know is how to detect overfitting, and it's quite simple. In order to do so, monitor the errors of your validation set, and they will usually tend to increase during the first epochs of training. Once the model has been trained through numerous epochs, there will be a point where the error will start increasing instead of decreasing, which means that the model is getting worse at generalizing. One should be cautious, because the error of the training set will always tend to decrease no matter the number of epochs, hence the importance of monitoring the validation set. See for example figure 2.9, where the model starts overfitting after around 100 iterations, where the accuracy starts trending downwards indefinitely. Arguably the easiest method to avoid overfitting is to do what is referred to as "early stopping", where training is stopped once the validation set accuracy starts decreasing indefinitely or the error starts increasing indefinitely.

Another common method of avoiding overfitting is L1 and L2 regularization, which works by adding an extra element to the loss function, which disincentivizes the model from using very high weights. L1 regularization is usually preferred over L2, because it tends to reduce the weights of features that are deemed as less important, converging to zero or sometimes even excluding them from the computations. The main difference between L1 and L2 regularizations, is that the noise introduced to the model with L1 is linear, whereas in L2 it's quadratic. The parameter that is introduced should be fine tuned through trial and error, hence being another hyperparameter of the model. Dropout is the last regularization method to be discussed, where overfitting is prevented by randomly dropping neurons with a determined probability, thus preventing the model from learning the noise in the training data and resulting in weights that generalize better [36].

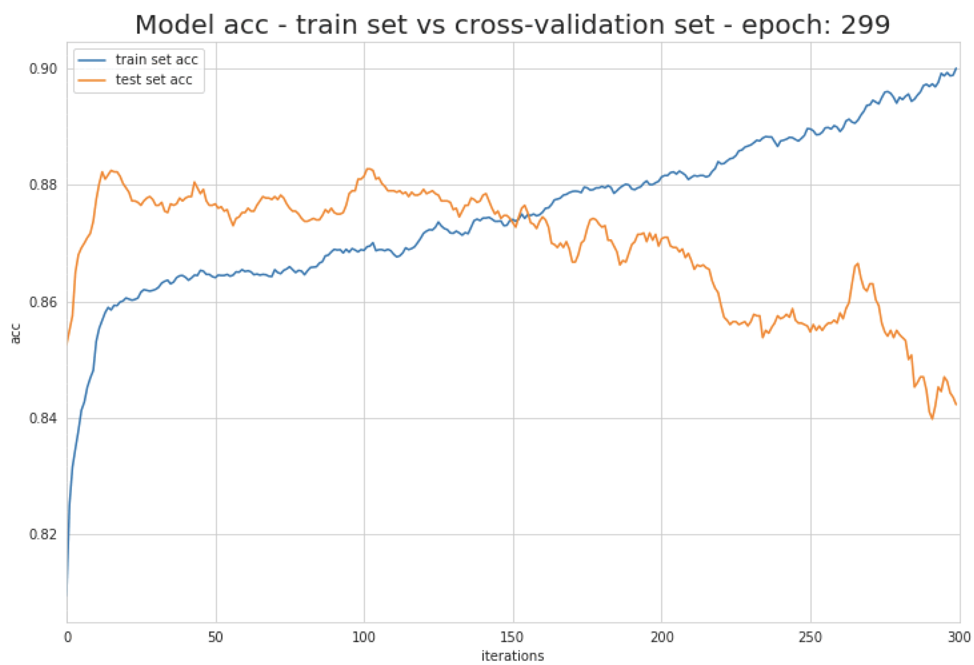


Figure 2.9 Learning Rate.
Retrieved from "*Preventing Deep Neural Networks from Overfitting*" in Towards Data Science, available in [36]

Chapter 3

The World Trade Web: Countries' Topologies and Their Effect on Income Level

Network analysis and machine learning have been reviewed in the previous two chapters, respectively. The former, will serve as the main tool of analysis for chapter 3 of this dissertation, titled "The World Trade Web: Countries' Topologies and Their Effect on Income Level"; the latter, for chapter 4, titled "The World Trade Web: A Deep Learning Approach to Link Weight Prediction". In the following chapter the topology of the world trade web is explored and analyzed, and relationships between network characteristics and income level are identified.

3.1 Introduction

Studying the world trade web (WTW) is of great importance, especially as a analytical tool for countries to design or refine their trade policies. International trade between countries is critically influential in shaping the world economy. The global gross domestic product (GDP) for 2017 (also known as Gross World Product, GWP) is 80.14 Trillion USD, out of which 16.3 Trillion USD (20%) comes from international trades. This does not account for all of the secondary effects that trade brings with it, like employment in factories, shipping and logistics companies, research and development, technological advances and transfer. Trade also enables the wide-spread of availability of products and services over the globe, which has been shown to accelerate globalization [2].

Approaching the study of the WTW using network analysis allows to integrally understand relationships among countries that couldn't be obtained without the use of network analysis, such as: assortativity, which would enable to identify if stronger countries tend to trade with weaker countries or vice versa; clustering, to see if there is a relationship between how strongly a country interacts with other countries and the strength of the trade relationships between the countries it trades with; disparity, to measure how well distributed a country's exports and imports are; structure of communities, to identify trade blocks. From an economic perspective, they are relevant because these first and higher order trade relationships play a role in the degree of dependency of countries on a given country or pool of countries.

The study of economic interactions using network analysis has recently showed to be highly insightful in numerous use cases, among the following: the analysis of globalization and regionalization in international trade [3]; understanding the potential and risks of economic systems [4]; empirically derive the structure of the world economy [5]; understand global interdependencies [6]; better understand the role of network characteristics in countries' incomes [7, 8]. Nonetheless, the amount of research one can find in the area is still scarce, being addressed by just a handful of authors. Network analysis can help understand how crises propagate through the WTW, where a shock to a highly central country is more likely to be transferred to the rest of the network [7]. Understanding the structures of communities within the WTW can better help identify trade blocks, where shocks that originate in one trade block would be less likely to impact other trade blocks. Moreover, network analysis has also recently proven useful in tangential applications to understand global and regional labor mobility, knowledge spillover, and the formation of geo-industrial clusters [37]. Additionally, dependency analysis, when applied with network analysis, can enrich the understanding on the role of countries' network characteristics in countries' incomes. Understanding the characteristics of the WTW can aid in comprehending the structure of the network and pinpoint specific channels of propagation of economic and financial disasters and shocks, thus enabling policy makers prevent and prepare for them. The previous implies a natural interdependency among the countries, since a reduction of a country's exports to another one can inhibit the latter's ability to manufacture exportable goods to its trading neighbors, a negative ripple effect [6, 8, 38–41]. Trade flows have been shown to be highly correlated with other country interactions such as flows of services, workers, and financial assets, hence being a relevant indicator for broader economic relations [42].

This chapter studies the topology of the World Trade Web by analyzing its unweighted and weighted network characteristics and attempts to answer numerous relevant questions. The data used for the analysis is the 2017 exports for 238 countries and territories for which there is available information [21]. This allows for the study of the trade interactions among countries, phenomenon that has, to this date, scarcely been studied with this technique. Using both, a directed unweighted and weighted network approach with the magnitude of exports as weights, this chapter's objective is to answer the following questions: Does the WTW still follow a degree power law distribution[1] when incorporating a richer dataset? How do the WTW's trading communities look like? Which continents are more susceptible to instability originating from their trading partners and spreading through trade? From a multivariate perspective, which countries are the most central

[1] A power law distribution describes a phenomenom where few items are clustered on one end of a distribution, representing at least 95% of the occurrences

in the WTW? Which countries have a substantially high dependency on others? Do countries trade with partners that are similar (in degree and weighted degree) to them? Does geographical proximity and trade agreements influence the relative intensity of trade among countries? What relationships can be found taking into consideration all of the countries' network characteristics? What actions can countries take to improve their per capita gross domestic product (PCGDP)? What continents are the major players in trade? How do the continental trade flows look like? This chapter also attempts to address several concerns that have been pointed out by authors that have done tangential works. Fagiolo, Reyes, and Schiavo [7] mention the lack of in-depth analysis into the topological characteristics of the individual continents and regions from the cross-sectional perspective, as well as the lack of studies in the role of geographical proximity in shaping the WTW and how fragile the network is. It is also mentioned that it is necessary to analyze if the topological properties of the WTW, from a weighted analysis perspective, can explain macroeconomic dynamics of growth and development.

The results obtained are insightful to understand the behavior and economic characteristics of countries with a determined network structure. Geographical proximity and trade agreements are found to have a crucial impact on the intensity of interactions among countries. The continent most susceptible to instability is North America, and the one least vulnerable is Europe. The results of Ordinary Least Squares Regression (OLS) show that in order to increase their income, countries should associate with more neighbors that are themselves weaker; reciprocate fewer of their trade links; trade more strongly with countries that are themselves stronger; and increase their export to GDP ratio.

The following sections are organized as follows: firstly, in section 3.2 related works in the field of unweighted network analysis and weighted network analysis will be discussed; following, in section3.3 the database that was used is described as well as how the network was constructed; afterwards, in section 3.4 the methodology to be employed for each one of the objectives and the results is covered; next section 3.5, goes through the conclusions and future work that could be done using these results as a foundation; finally, in section 3.6 the limitations of the work are discussed.

3.2 Literature Review

For network analyses, the entities (countries, people, etc.) are represented as nodes, and edges between any pair of entities represent their relationship (trades, friendship, etc.). The edges between nodes might be *binary* or *weighted*. A binary edge captures if the relation between two nodes exists or not (e.g., if two countries trade or if two people are friends) whereas weighted edges capture the magnitude or extent of the relationship (e.g., the value of trades between two countries, or the extent of friendship between two people). In this light, network analyses might fall into two categories of binary (or unweighted) and weighted depending on whether the underlying network is represented by binary or weighted edges. There are well-studied methodologies to analyze unweighted networks (see, for example, these comprehensive introductions [10, 43]). In contrast, the techniques to analyze weighted networks are still novel, questionable, and not necessarily well-established (see, for example, the various generalizations for a weighted clustering coefficient that have been proposed [11–14]). Nonetheless, numerous novel weighted network metrics have been developed for weighted analysis [11, 15–20]. We believe that the weighted network analyses of the World Trade Web is of practical importance by capturing not only the trades of two countries but also the magnitude of those trades.

Failing to use weighted links when working with the WTW results in a vast loss of information and possible insights that could be obtained from a weighted network analysis. Nonetheless, the bulk of the available literature on network analysis applied to the WTW approaches the analysis in an unweighted fashion (see [3, 5, 44–46]). With an unweighted approach, equal weights are applied to all trade links in the WTW. The previous can be argued inappropriate since it is, for example, giving the same importance to a trade of 10 billion dollars and one of 10 dollars. However, to be able to contrast results with the ones from the literature, unweighted analysis of the WTW is to also be performed in the current chapter, and results that debate other authors' findings were obtained.

Custom-weighing methods other than the actual magnitude of the trade flows have been used in the literature, which can be argued inappropriate. Fagiolo, Reyes, and Schiavo [7] use an arbitrary and custom weighting method that involves the addition of, for example, the exports of country "A" to country "B" and vice versa, over the gross domestic product (GDP) of the exporting country ("A"), divided by 2. The drawbacks of using such a weighting approach, is that this deflates trade magnitudes. E.g. if country "A" exports 1 billion dollars to country "B" and the economy of the former is 10 billion dollars, then the ratio of exports to GDP would be 0.1. In a similar way, if country "B" exports 1 dollar to country "A", and the size of the former is 10 dollars, then the ratio is also 0.1, so this is holding back the detection of the magnitude and potential of stronger trade links. This mechanism can be useful in some cases, where the researcher wants to analyze how important a partner is to a country in particular, instead of the entirety of the WTW. Weighing mechanisms like the previous could be arguably considered inappropriate for the analysis of the topology of the WTW depending on the objective of the analysis to be performed, hence this chapter uses the export magnitudes as weights. Breiger [47] also uses a custom weighting method, where trade flows are studied using weighted links on an undirected network. In this work, custom weighting that is depurated by average imports and exports is used, again having the same drawbacks as the custom weighing method used by Fagiolo, Reyes, and Schiavo [7]. Bhattacharya [48] and Bhattachayra, Mukherjee, and Manna [49] weigh each link using the difference between exports and imports (having the same drawbacks as the aforementioned studies) and an intertemporal comparison where constant United States Dollars (USD) aren't used was performed, rather current

USD, which can be argued inadequate for numerous applications in economics because inflation is not accounted for. Serrano, Boguñá, and Vespignani [6] use a rather peculiar weighing mechanism, where a trade link only exists if there's a bilateral trade imbalance between two countries, and it is weighted with the magnitude of said imbalance.

Some authors have argued that there is evidence to symmetrize the network [7] which could potentially allow the researcher to simplify the analysis by reciprocating trade links that aren't currently being reciprocated and removing the directionality of the flows. However, this can be deemed inappropriate under particular circumstances. When analyzing the WTW, Fagiolo, Reyes, and Schiavo [7] use a metric of symmetry proposed by Fagiolo [50], and after getting positive results for this test, the data matrix is symmetrized by removing the directionality. Symmetrizing the network can be argued inadequate when there is a lower network density, which stands at 37% using the dataset from [21] and a reciprocity of 62%, and this doesn't account for magnitudes of the flows, which could arguably make the WTW less symmetrical.

Dependency analysis can be used to understand how the network characteristics of countries can impact their development. General characteristics of countries such as initial GDP conditions, physical capital, human capital, and other variables like degree of openness, geographical, and political characteristics have been used to explain GDP per capita growth rate [51–53]. Kali and Reyes [8], whose work is based on Harrison [51], Yanikkaya [52], and Irwin [53], perform a dependency analysis including network characteristics, such as export dependency and import dependency of the nodes, as independent variables to explain GDP per capita growth rates. A drawback with this exercise when the researcher's objective is to find the significance of variables that are network characteristics, is that variables that aren't network characteristics are included in the regression, such as human capital, physical capital, regime, climate, and access to water, which can reduce the significance of the variables that are network characteristics.

Also, the independent variables that are network characteristics are metrics that are complicated for countries to manipulate through public policy in order to improve their growth (such as centrality), which doesn't allow countries to take recommended actions to improve their growth rates. Arora and Vamvakidis [54] have shown that the gains from trade do not depend just on the degree of trade openness, but the number of trading partners, which is associated with higher growth rates [8]. The previous relationship is associated as a result of the countries with more trade partners being exposed to better technologies, more markets and competition.

For the weighted approach of this work, the volume of trade will be used as the weight of the edges of the WTW, and will be approached as a directed network, which means that unlike Fagiolo, Reyes, and Schiavo [55], the matrix won't be transformed as a result of a symmetry index, since it is a strong and unreliable assumption that the export flows from country "A" to country "B" are the same as the export flows from the latter to the former. Just by glancing at the matrix, one can notice that there's no strong symmetry, so no approach to remove directionality will be undertaken. As of today and to the best of our knowledge, no other works have been identified that analyze the WTW on a directed, weighted fashion for the year 2017 with the weighting mechanism hereby proposed, and with the rich database provided by COMTRADE[21].

3.3 Data

This chapter uses a dataset from the World Integrated Trade Solution (WITS), provided by the United Nations (UN) COMTRADE [21]. The data is from 2017, which is currently the most recent year with information on all of the countries reported by the database, and reports data con 238 countries. Note that we will refer to any independent territory reported by the database as a country. The raw dataset comes in the form of a weighted edgelist, which facilitates the processing of data in softwares like Gephi[2] and Python libraries like NetworkX[3]. Table 3.1 shows a sample weighted edgelist dataset.

Table 3.1 Dataset Sample Weighted Edgelist

Exporting Country 1	Importing Country 1	Trade Flow Magnitude 1
Exporting Country 1	Importing Country 2	Trade Flow Magnitude 2
...	...	...
Exporting Country m	Importing Country n	Trade Flow Magnitude n

Based on the nature of trade, a country's export is another country's import. Stemming from this, the first column in table 3.1 corresponds to the ISO3 code of the country that is exporting, the second one to the ISO3 code of the country that is importing, and the third one to the magnitude of the trade flow. Henceforth, countries can appear numerous times in both columns because they can be a source of exports and destination of imports to and from numerous countries.

We should also take into consideration that trade flows reported from this database come from official, legal trade, which fails to capture the underground economy. One should note that often trade happens informally between countries, so developed countries' data could end up being more reliable than that of developing countries due to the existence of rule of law and strong institutions.

There's a limited availability of literature on the WTW after the year 2010, which could be due to the database constructed by Gleditsch [56] not being updated anymore. This database has been used by numerous authorities in the field [29, 30, 55, 57], but hasn't been updated since 2011, and the URL to it doesn't work anymore. Also, the database

[2]Gephi is a software for visualizing and analyzing complex networks. The official documentation can be accessed here: https://gephi.org/users/

[3]NetworkX is a library that enables the analysis of complex networks. The official documentation can be accessed here: https://networkx.org

constructed by Gleditsch [56] consists of just 159 countries (according to the authors that have used it, since we couldn't access it ourselves, given that the links to it are broken), which indicates that this database is considerably less informative than COMTRADE's database [21], which consists of 238 countries. One can believe that given that Gleditsch's database isn't available anymore, recent authors like Chow [58] have recurred to using the OECD database [59], which currently consists of 64 OECD and non-OECD countries. The drawbacks of using this database are that it doesn't include most of the African countries, as well as small countries from southeast Asia and the Caribbean. Also, 64 countries is less than one third of the countries and territories reported by the database used in the present study [21]. To the best of our current knowledge, no other author has used such an in-depth analysis to obtain insights at the country and continent level, together with a dependency analysis for actual trade policy recommendation. Also, the database is publicly available on a reliable website that any researcher and/or reader can access and thus replicate the current experiment or work with the same data on similar research projects.

3.4 Network Statistics and Results

3.4.1 General Properties

In simple terms, a network is a collection of points joined by lines. Within the field of network analysis, the points are commonly referred to with the term "vertices" or "nodes", while the lines that connect said nodes or vertices are referred to as "edges". Edges can be labeled with additional information like an edge weight, which allow to capture more details of the system. There are 2 main type of networks: undirected and directed. The former does not take the directionality of the edges into consideration, while the latter does. A trivial example of an undirected network can be a communication network, where if one person has talked to the other person there has already existed some form of communication between them, even if the other person does not reply. On the other hand, a directed network can be a series of airports, where airplanes can fly from one airport to another one, but the opposite might not be true. Directionality of the edges conveys additional information as well, as it indicates the flow of the interaction between the nodes. Neighbor nodes are those that are adjacent to another node (connected by an edge). In the context of the WTW, a country is represented by a node, their trade relationship is the edge, the weight, or strength of said trade relationship is the magnitude of the trade flow, and it is a directed network where directionality works as follows: an outgoing edge represents an export from the source node, and an incoming edge represents an import from the target node. A country is considered the neighbor of another country if they're connected in either direction (import or export), given that the fact that they are interacting commercially turns them into trade partners, hence, neighbors.

Mathematically, a dense graph is a graph where the number of edges is close to the maximum number of edges that can exist within the network. The opposite is also true, where having few edges relative to the total possible edges is associated with a sparse graph. The graph density[4] for the WTW is 0.372, meaning that there exists 37.2% of the total possible connections. Note that there are 238 countries, and if every country was connected to every other country, the number of connections would be 56,406 (238x237). The density obtained differs from the 50% that Garlaschelli, Loffredo [57], and Fagiolo, Squartini, and Garlaschelli [30] found for the year 2000 using Gleditsch's database [56]; the 65% found by Fagiolo, Reyes, and Schiavo [55] for 2000. Hence, these results indicate that the WTW is considerably less dense and therefore more poorly connected than previously determined by other authors, showing a lot of opportunity for new trade partnerships and relationships between countries to occur.

[4] Graph density represents the actual number of edges as a proportion of the total possible edges if all the nodes were connected to each other.

The average nearest neighbor degree (ANND) is defined as the number of trade flows (number of outgoing + incoming edges) a country's neighbors has, on averaged over the number of neighbors. The ANND obtained is 259, significantly higher than the 120 obtained by Fagiolo, Reyes, and Schiavo [55] and Fagiolo, Squartini, and Garlaschelli [30] for the year 2000 and the 100 obtained by Fagiolo, Reyes, and Schiavo [29]. The results obtained indicate that countries are connected, on average, to significantly better connected neighbors than what was previously thought. The average nearest neighbor strength for the network is 314,447,654 thousand dollars.

In directed networks, the frequency of loops of length 2 is captured by a metric called "reciprocity", which captures the fraction of vertices that you point to that also point back at you. In the case of the WTW, a country's link is reciprocated when it is exporting to a country it also imports from. The reciprocity obtained in the WTW is 62%, which represents the proportion of the relationships between countries that are bidirectional in relation to the total edges, meaning that they are partners through both importing and exporting from and to each other. The previous result is significantly different from the ones found by Garlaschelli and Loffredo [60] and Fagiolo [50] of around 80%, and the 95% obtained by Garlaschelli and Lofredo [57], and the almost 100% obtained by Fagiolo, Reyes, and Schiavo[55]. This once again indicated that the probability of the network being considered symmetrical is substantially lower, arguably invalidating symmetrizing the network as done by Fagiolo, Reyes, and Schiavo [55].

In graph theory, the eccentricity of a vertex is the greatest distance between itself and any other vertex, or how far the node is from the node that is most distant (measured as total steps to reach the node travelling through edges) in the graph. Similarly, the diameter is the maximum eccentricity among all the vertices in the network. The diameter obtained for the WTW is 3, which means the maximum amount of trades flows that should be followed to get from one country to another one is just 3. The only other identified work that had computed this metric for the WTW is Chows's [58], where a value of 2 was obtained, but given that the OECD database was used, which only consists of 64 countries, the results obtained are significantly different.

The reason for the significant differences obtained are likely due to the fact the database hereby used considerable more complete, meaning that it includes even small islands from the Caribbean and Southeast Asia, which enriches the analysis and adds a considerable amount of countries, links, and relationships that didn't previously exist, thus reducing density, reciprocity, and numerous other metrics. Also, as previously mentioned, other databases like the one from the OECD are missing most of the African countries as well.

One of the most simple centrality measures in network analysis is degree centrality, which indicates the number of edges that are connected to a vertex. In the case of the WTW, the degree of a country is the sum of the number of countries it exports to and the ones it imports from. Figure 3.1 shows a heat map of the degree by country. One can observe developed countries show a higher degree, namely countries such as Canada, USA, Australia, Japan, New Zealand as well as most of Europe. Developing countries with a remarkable degree include Brazil, South Africa, India and China. Note that most of Africa is notably poorly connected. Another trend that can be observed is that geographically smaller countries tend to be less connected than large countries. The correlation between the size of countries (in land square kilometers) and their degree is 0.267, showing a low-moderate correlation, meaning that larger countries do tend to have a higher degree.

Statistically, a power law is defined as a functional relationship between two variables, where changes in one of the variables results in a proportional relative change in the other variable. In other words, one variable is changing as a power of another one. In the case of the WTW, it would follow a power law degree distribution if numerous countries had a low degree, and few countries had a very high degree (known is some domains as the 80-20 rule). By observing Figure 3.2a, one can note that the degree distribution of the world

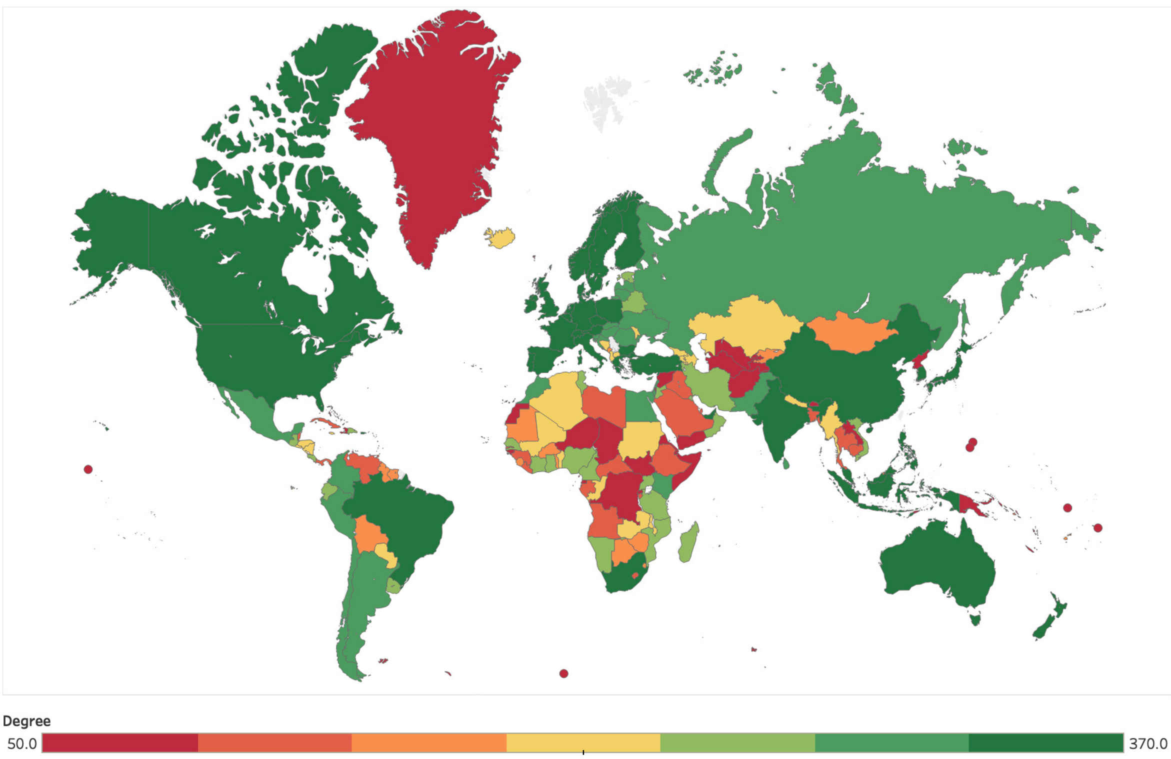
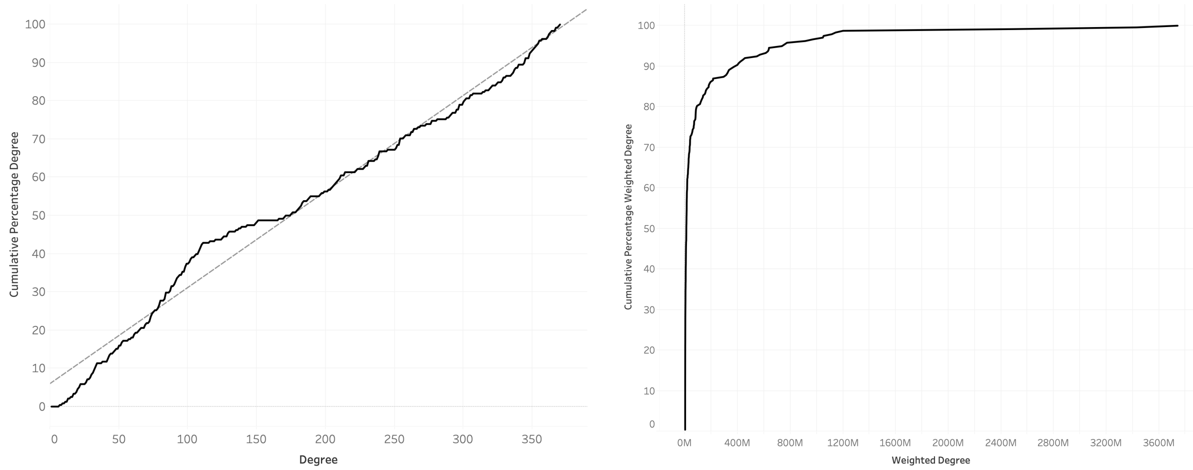


Figure 3.1 Heat Map of Degree by Country



(a) Binary **(b)** Weighted

Figure 3.2 Degree Distributions (Cumulative Percentage)

trade web doesn't seem to follow a power-law distribution. If one looks at the cumulative distribution, it's increasing linearly, whereas for power law one would expect for most of the observations to accumulate in the lower degrees, and for it to marginally increase afterwards. For power law to occur in this dataset, there should be numerous countries with a low degree, and very few countries with a high degree, which is not the case.

The average weighted degree of a network is the average magnitude of its edges. In the case of the WTW, the average weighted degree is the average magnitude of the export and import flows. The average weighted degree is 68,439,821 Thousand USD. When observing the network taking into consideration weighted edges, it can be shown in Figure 3.2b that the weighted world trade web seems to follow a power law weighted degree distribution, which means that numerous countries have very weak trade links, whereas very few countries have very strong ones. There is strong evidence of this, given 95% of the countries trade less than 20% of the USA's trade[5].

Figure 3.3. is a heat map of the difference between the number of exporting partners (out-degree) and the number of importing partners (in-degree). If a country obtains a

[5]The USA is the country that trades the most.

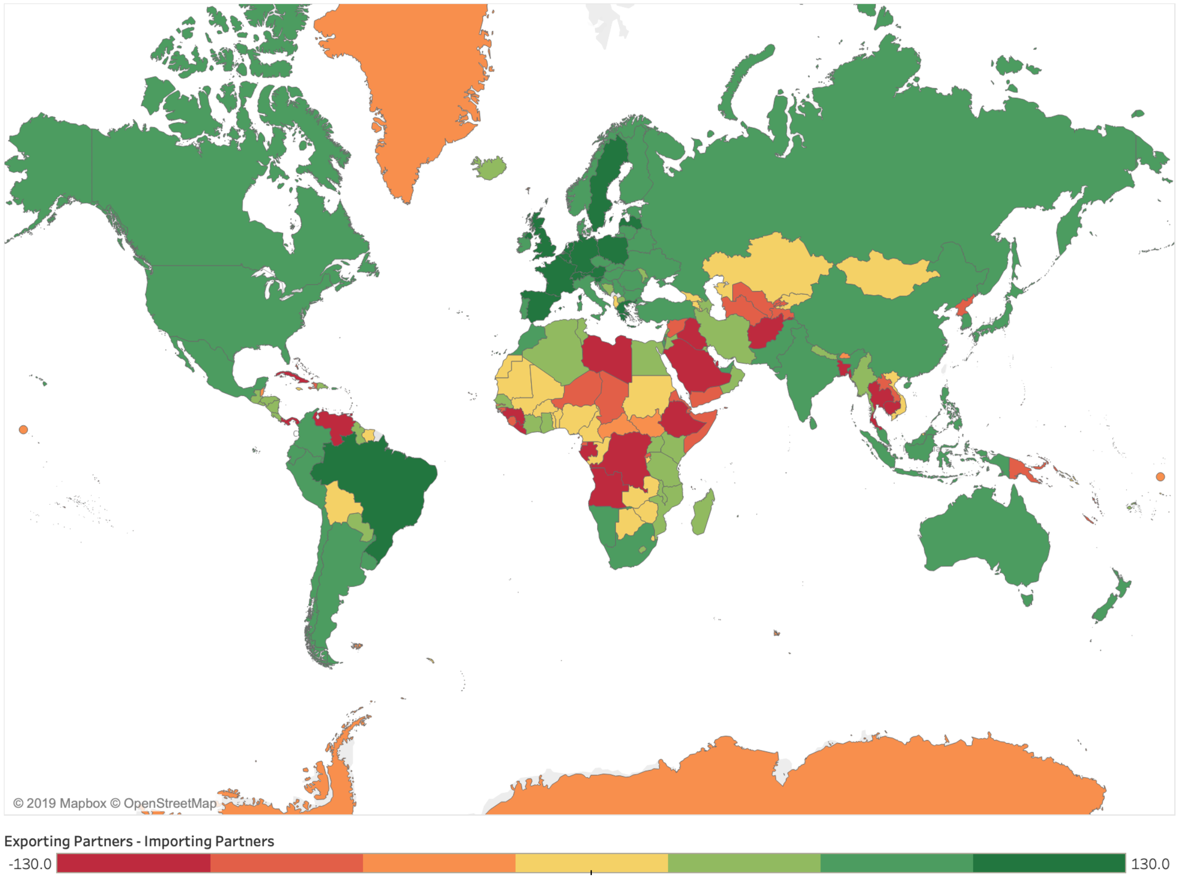


Figure 3.3 Heat Map of Number of Export Partners Minus Number of Import Partners

positive score, this indicates that it exports to more countries than it imports from. When observing this heat map, it can be noted that the main continent where countries import from more countries than the ones they export to is Africa, with a few exceptions. Some territories in the middle east also appear to have this characteristic. Most of the rest of the territories appear to have a positive result in this metric. Note that most small islands and territories are sink nodes, meaning that they only import. This can be reflected by observing the number of nodes with a null out-degree, which amounts to 96.

Figure 3.4 shows the net exports (total exports minus total imports) of countries, divided by their corresponding GDP. This allows to see the magnitude of their trade

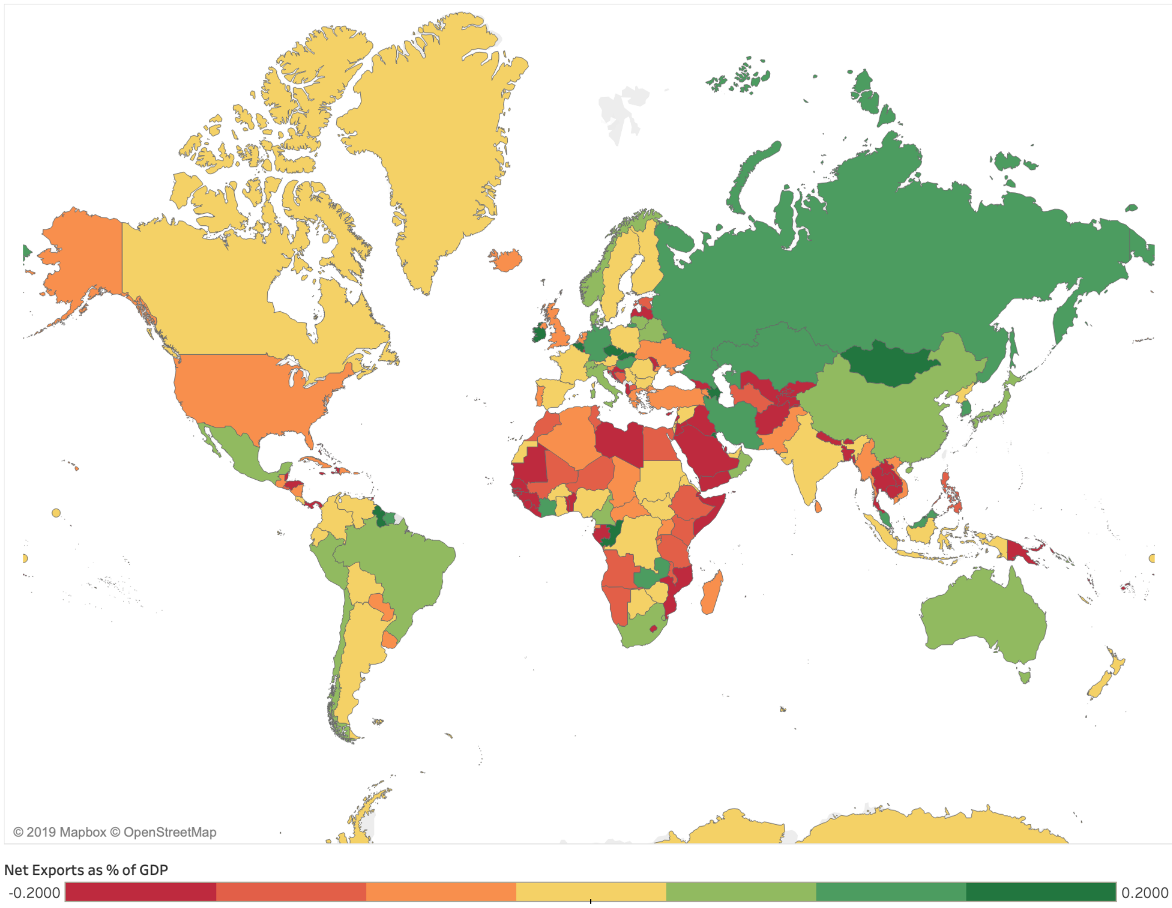


Figure 3.4 Heat Map of Net Exports as % of GDP

surplus or deficit in proportion to their GDP. Russia, Australia, Mexico, Brazil, Chile, South Africa, Eastern Asia, and other few countries appear to have large trade surpluses relative to the rest of the countries. Most of the African countries and small islands denote the opposite characteristic, having a large trade deficit, which is supported by Figure 3.3.

3.4.2 Community Detection

The algorithm used for community detection was developed by Blonde [61] and is based on modularity optimization. This algorithm identifies the communities based on capturing the sets of highly and densely interconnected nodes. It has been shown that the aforementioned algorithm has a high quality of communities detected and has been proven by ad hoc modular networks. The method also allows for the use of weighted edges to account for strength in the relationship of the communities. A resolution parameter was introduced by Lambiotte, Delvenne, and Barahona [62] and is constructed through the connection between community detection and Laplacian dynamics, using extended versions of the current algorithms to test its efficiency. This resolution parameter allows for “fine tuning” of the communities; a higher parameter will result in less, bigger communities. On the other hand, a lower parameter will result in more, smaller communities. A resolution parameter of 0.6 was used for community detection, and the resulting communities can be observed in Figure 3.5.

Figure 3.5, with a resolution parameter of 0.6, detects 7 communities. The 1st one groups North America and the northern part of South America. The 2nd one most of South America with Africa, Ukraine, most of the Middle East, South Asia and Southeast Asia. The 3rd one groups Finland, Russia, and most of central Asia, mainly the “stans” (Uzbekistan, Afghanistan, Pakistan, etc.). The 4th one groups Most of Europe, Morocco, Tunisia, the United Kingdom, Iceland, Norway and Denmark. The 5th one groups China, Mongolia, Iran, Australia, New Zealand, and Papua New Guinea. The 6th one groups Denmark, Greenland, and the Faroe Islands. The 7th one groups Cyprus, Greece, Cayman Islands, and Pitcairn. Once again, the algorithm detects a strong interaction among countries that are geographically proximate.

One can observe that the algorithm can also capture some countries with trade agreements within the same community, given that they have stronger ties. The vast majority of the countries that are members of the European Union can be detected within the same community. The North American Free Trade Agreement (NAFTA) between Mexico, the USA and Canada can also be detected, given that they are all within the same community. Brazil, India, and South Africa are all detected within the same trade community and form part of "BRICS" economies (Brazil, Russia, India, China, and South Africa). Note that the interaction between "BRICS" is not yet influent enough to group Russia and China within the same community. Also, one can note that China and Australia are within the same community, and this could be due to the strong interactions that occur since Australia is well known to be a crucial supplier of raw materials for Chinese manufacturing facilities.

Community detection is of vital importance, because it identifies trade blocks that are formed by countries that have many and strong interactions among them. Identifying trade blocks can signal how the spread of different economic, financial, and in general stability shocks to a country, can propagate to other countries or regions. Shocks that affect a country or region have a higher probability of being mostly contained within the community, which will be affected the most by said shock, while other trade blocks should be affected to a lesser extent.

3.4.3 Unweighted Centrality Index

For unweighted centrality, 6 metrics are going to be taken into account: degree centrality, betweenness centrality, closeness centrality, harmonic centrality, eigenvector centrality, and page rank. Homologous weighted metrics are going to be used on section 3.4.4 for a measure of weighted centrality. A brief explanation for each one of the measures is going to be provided. Newman [10] offers further details on each one of these measures. Given that the metrics are highly correlated among them and interdependent, it is possible to create a country importance index using principal component analysis (PCA), which reduces dimensionality and summarizes information, while minimizing the loss of information.

Degree centrality is a rather simple measure and it's just the degree of the node, in this case the amount of countries a determined country is trading with, or the sum of its total outgoing edges and incoming edges.

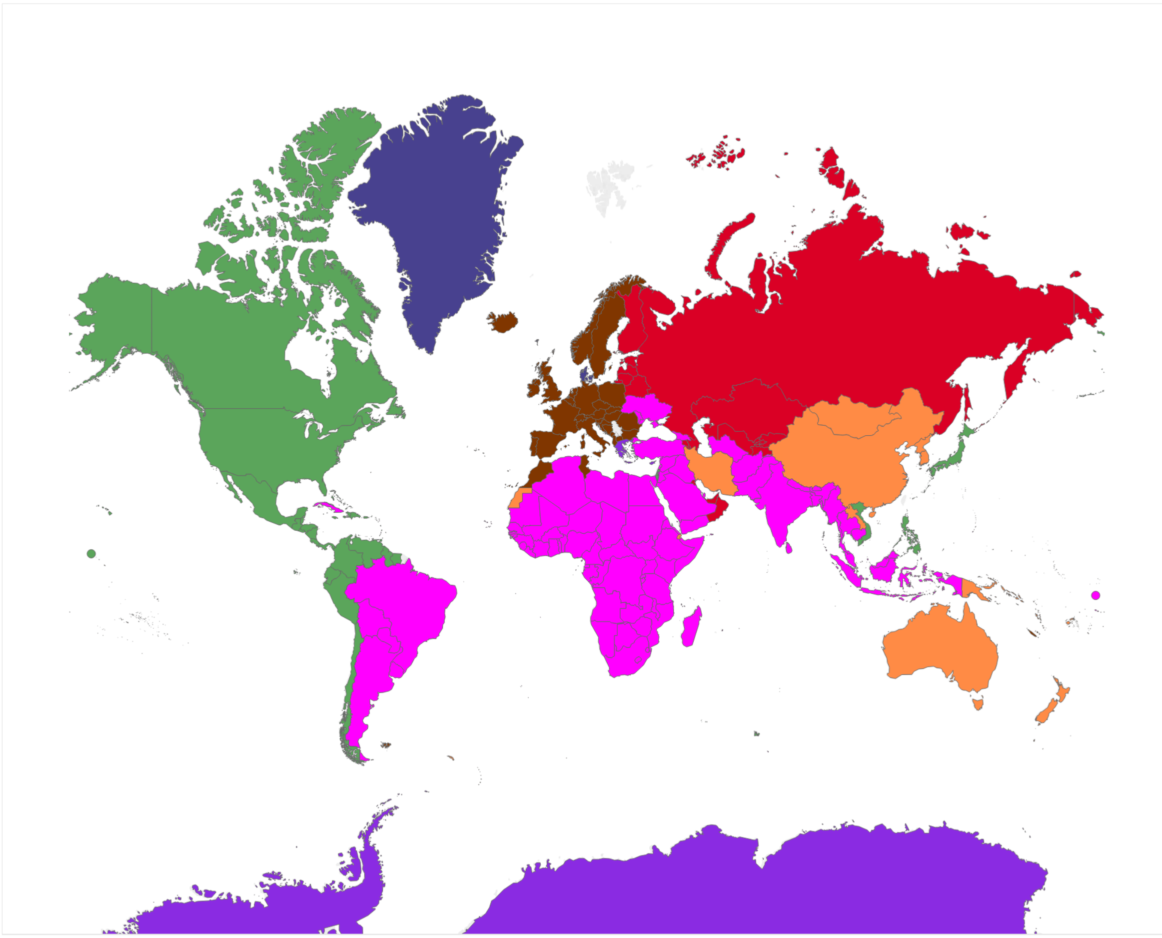


Figure 3.5 Community Detection with Resolution Parameter = 0.6

Eigenvector centrality also takes into account how important the neighbors of a node are. In this case, the importance of a node is not only going to be determined by the amount of neighbors or edges it has, but also by how well connected the neighbors themselves are. Vertices get a score proportional to the sum of the scores of its neighbors. Eigenvector centrality was constructed on the notion that, according to Bonacich [63], its creator, and contrary to traditional social network research from Mizruchi and Domhoff [64], and Mintz and Schwartz [65], power itself doesn't necessarily properly determine centrality. In a power hierarchy, one's power depends on the power of the partners you have power over. In other words, in unweighted networks, being connected to countries that are themselves well connected will results in a higher eigenvector centrality. Similarly, in weighted networks being more strongly connected to stronger countries will results in a higher eigenvector centrality.

Pagerank, introduced by Page [66] is based on another notion of centrality called the Katz centrality, and differs by diluting the scores that nodes get from receiving connections from prestigious vertices by the amount of outgoing edges that said prestigious vertex has. Pagerank takes eigenvector centrality as its foundation.

Harmonic Centrality, when computed with edges, uses well known Djikstra's algorithm [67]. The algorithm works by creating a tree of all the existing shortest paths from a node to all of the other possible nodes in the graph, it works with a signal travelling through the nodes and edges that is going to avoid using the edges with high degree, hence nodes with better connections will have less traffic of signals. Nodes with better connections will have a lower harmonic centrality score.

Closeness centrality is the measure of the mean distance from one vertex to the rest of the vertices, being an unconventional measure of centrality. This measure is useful, because in the case of a social network, if someone has a lower mean distance to other people, their actions will reach other people in the community faster than someone with a higher mean distance. The previous applies in the case of countries as well.

Betweenness centrality captures the extent to which a vertex lies on paths between other vertices, or how many shortest paths need to go through the vertex that is object of the analysis. The development of this centrality measure is attributed to Freeman [68].

Before performing PCA for dimensionality reduction and information summarization, one should first verify that all the variables are in the same direction (more is better or vice versa). In case one variable is not in the same direction as the rest, the inverse should be computed. Once all the variables are in the same direction, they should be normalized in order for them not to be unwillingly weighted based on their scale. Furthermore, one should perform 2 common tests to assess if the variables are significantly and sufficiently intercorrelated among them, and hence information reduction can be performed on them. The 2 tests are the Barlett test of sphericity and Kaiser-Meyer-Olkin (KMO). For the former, a value below 0.05 is desired, and the current dataset obtained a score <0.001, hence satisfactory. For the latter, a score below 0.5 is unacceptable, in the 0.50s miserable, 0.60s mediocre, 0.70s middling, 0.80s meritorious, and 0.90s marvelous [69]. The KMO score obtained is 0.7, which is considered a middling score, hence PCA can be performed to obtain an unweighted centrality index.

One should evaluate the component matrix, where factor loadings above 0.3 are significant, while the ones above 0.5 are highly significant. Having variables with factor loadings below 0.3 should make the researcher reconsider whether to include that variable or not. All of the factor loadings were considerably above 0.5, hence all of them will be retained.

Among the criteria commonly evaluated to decide the number of factors that should be extracted, one finds the latent root (eigenvalue) and the percentage of variance criterion. Factors with eigenvalues greater than 1 are usually considered significant [70, p. 107]. The scree plot suggest the extraction of just one factor, which accounts for 78% of the total

Table 3.2 Top 10 Most Central Countries Based on Unweighted Centrality Index

Centrality Rank	**Country**	**Centrality Index**	**Times Centrality Index is Surpassed by the USA**	**Centrality Relative to the USA**
1	United States of America	16.07	1.00	100.00%
2	China	12.24	1.31	76.16%
3	Germany	12.06	1.33	75.04%
4	Great Britain	11.00	1.46	68.47%
5	Netherlands	10.91	1.47	67.89%
6	France	10.61	1.51	66.02%
7	Canada	9.59	1.68	59.67%
8	Italy	9.58	1.68	59.64%
9	Spain	8.89	1.81	55.33%
10	Switzerland	8.84	1.82	54.99%

variance explained. These are desired results when one wants to create an index, because it denotes high homogeneity and interdependence of the principal component to be used in the construction of the index. The results of PCA allow to create a hierarchy of the most important countries taking into consideration all of the 6 centrality metrics altogether. Also, the score of the index can be interpreted as well to denote the magnitude of the difference in importance between the countries. The results of the previous exercise are shown in table 3.2.

The centrality rank isn't the only aspect worth noting on table 3.2, but one should also pay attention to the percent of centrality relative to the USA, which serves as a reference point to the magnitude of the difference in centrality between the countries, in this case, to what extent each country is less influential that the USA. It's worth noting that the USA is about 80% more central than Switzerland, and Spain. It is about 50% more central than Great Britain, Netherlands, France, Canada, and Italy. Also, it is around 30% more central than China and Germany. Table 3.2 allows to identify the countries whose economic and financial health and stability can impact in a greater magnitude than the rest of the less central countries, and quantifies the differences in proportion of such impact based on their centrality.

According to the centrality index and differing to what was found by Fagiolo, Reyes, and Schiavo [55] from the year 1981-2000, the situation 15 years later has significantly changed. The importance of China increased drastically, now being the second most central country. Russia is not a member of the most central countries anymore. Japan and France have lost relevance during this time frame, while Great Britain remains fairly constant. Germany has regained its centrality. It's important to note that Fagiolo, Reyes, and Schiavo [55] only used the weighted metric random walk betweenness centrality (more about this metric on the next section) to rank the countries, which arguably has numerous limitations when denoting centrality, since there's a great variety of centrality metrics, hence we chose to perform PCA to consolidate them into one and take all of them into consideration. Also, Chow [58] has significantly different results using the same metric as [10] to measure centrality, given that their results show Japan, Hong Kong, and Singapore as the 3rd, 4th, and 5th most central countries respectively for 2009, whereas this chapter's results show that none of them are even within the 10 most central countries. Table 3.6 shows a red-yellow-green color-coded heat world map to visualize the centrality index of the countries in the WTW.

In general terms, one can observe that the USA, Canada, China and most countries in

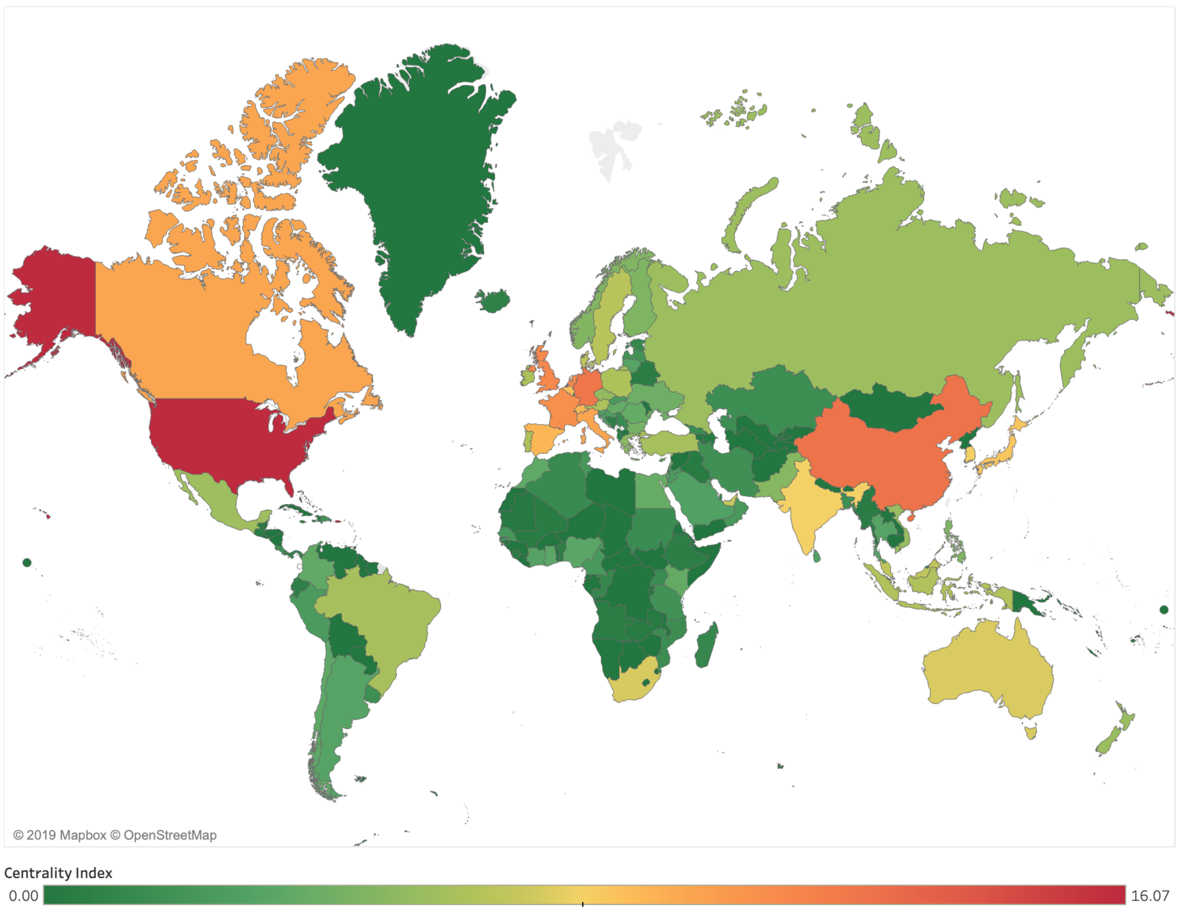


Figure 3.6 Red-Yellow-Green Color-Coded Heat Map (Centrality Index in Descending Order: Red, Orange, Yellow, Light Green, Dark Green)

Europe are the most central countries. On the other hand, Africa, the Middle East, and South America appear to be the least influential territories.

3.4.4 Weighted Centrality Index

For weighted centrality, 6 metrics are going to be taken into account: weighted degree, random walk betweenness centrality (RWBC), random walk closeness centrality (RWCC), weighted harmonic centrality, weighted eigenvector centrality, and weighted page rank. It is relevant to note that RWBC is the homologous of betweenness centrality (BC), but is used for weighted networks instead, since BC can't be calculated for a weighted network as such; the same relationship applies between RWCC and closeness centrality (CC). Given that the metrics are highly correlated among them and interdependent, this allows for the use of PCA as well to create a weighted centrality index.

RWBC was developed by Newman [25] and Fisher and Vega-Redondo [26] and can be intuitively explained using signals. What the algorithm does, is send signals through all of the edges, and each signal has a target node. The signal is going to perform a random walk, where nodes that have a higher weighted degree have a greater probability of being chosen as a route. Hence, those nodes that have the most traffic of signals going through them are going to be the ones with the highest RWBC. A higher RWBC means higher importance in the betweenness centrality aspect.

RWCC is also known as information centrality and was developed and tested by Stephenson and Zelen [27]. Robust statistical knowledge is necessary to fully comprehend the technicalities of the measure and providing an intuitive explanation is challenging. However, this metric has been further optimized and tested by Brandes and Fleischer [28] in order to be able to approximate this metric for large networks being less computationally intensive. A higher RWCC means higher importance in the closeness centrality aspect.

Once again, the Barlett test of sphericity and KMO were computed. For the former, the current dataset obtained a score <0.001, hence satisfactory. For the latter, the score obtained is 0.829, which is considered a meritorious score, hence PCA can be performed. When evaluating the component matrix, the RWCC obtained a factor loading that was below 0.3, and in general the correlations between this variable and the rest were significantly lower than the average, hence the variable was eliminated, and the exercise was recomputed without it. The KMO and Barlett test scores remained highly significant, and now all of the factor loadings were significant as well.

The scree plot suggested the extraction of just one factor, which accounts for 84% of the total variance explained. Once again, the results obtained were desired because it denotes high homogeneity and interdependence of the principal component to be used in the construction of the index. The results of PCA allow to create a hierarchy of the most important countries taking into consideration all of the 5 centrality metrics altogether. Also, the magnitude of the index can be interpreted as well to denote the magnitude of the difference in importance between the countries. The results of the previous exercise are shown in table 3.3.

Once again, one should also pay attention to the percent of centrality relative to the USA, which serves as a reference point to the magnitude of the difference in centrality between the countries. It's worth noting that the USA is about 5 times more central than South Korea, Canada, and Hong Kong. It is about 4 times more central than France, Great Britain, Japan, and the Netherlands. Also, it is more than twice as central as Germany, and around 30% more central than China. Table 3.3, once again, allows to identify the countries whose economic and financial health and stability can impact in a greater magnitude than the rest of the less central countries, and can also help quantify the differences in proportion of such impact according to their centrality.

Comparing the results to the unweighted centrality index computed in the previous index, one can see that in general and as expected, using the weights boosts the importance

Table 3.3 Top 10 Most Central Countries Based on Weighted Centrality Index

Centrality Rank	**Country**	**Centrality Index**	**Times Centrality Index is Surpassed by the USA**	**Centrality Relative to the USA**
1	United States of America	41.18	1.00	100.00%
2	China	31.88	1.29	77.43%
3	Germany	17.74	2.32	43.07%
4	Netherlands	10.15	4.06	24.65%
5	Japan	10.10	4.08	24.52%
6	Great Britain	9.88	4.17	24.00%
7	France	9.45	4.36	22.94%
8	Hong Kong	8.60	4.79	20.89%
9	Canada	7.56	5.44	18.37%
10	South Korea	7.43	5.54	18.04%

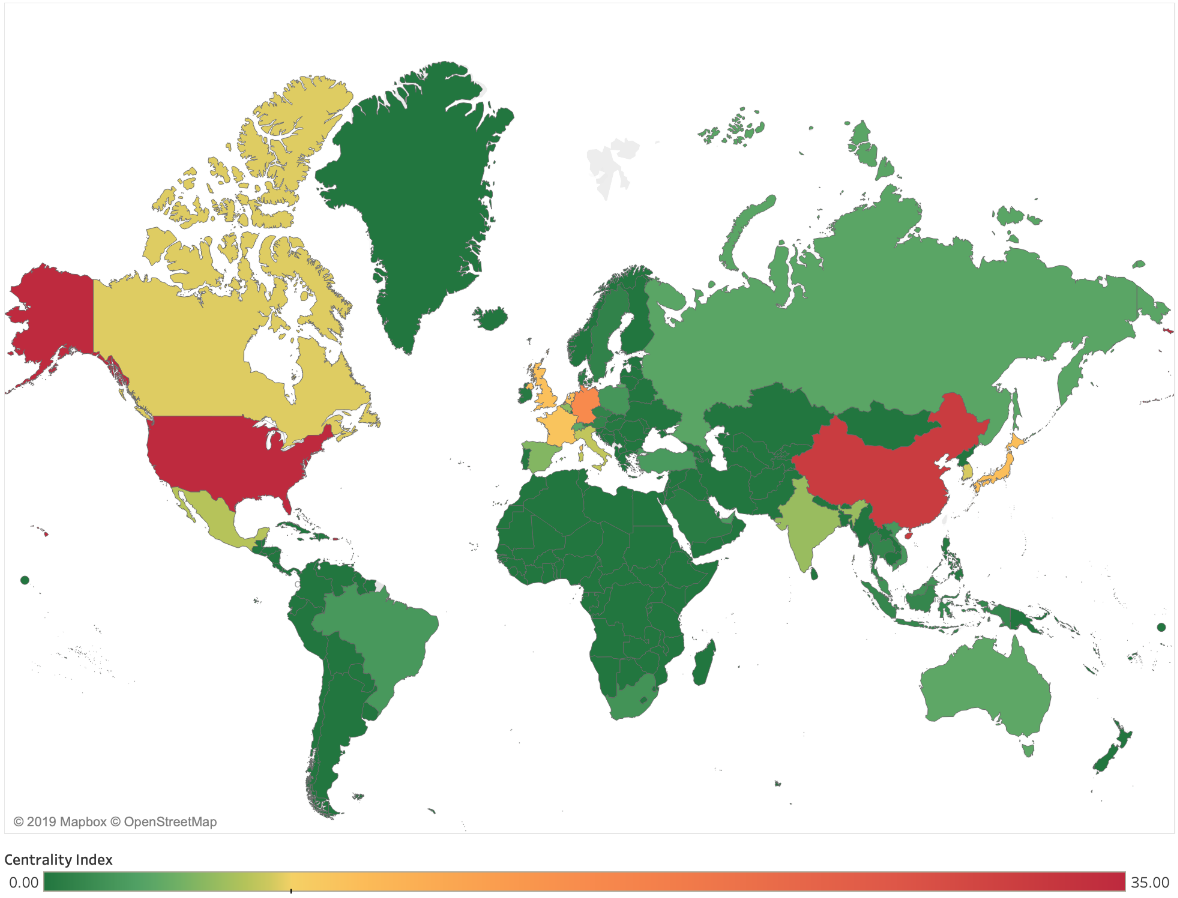
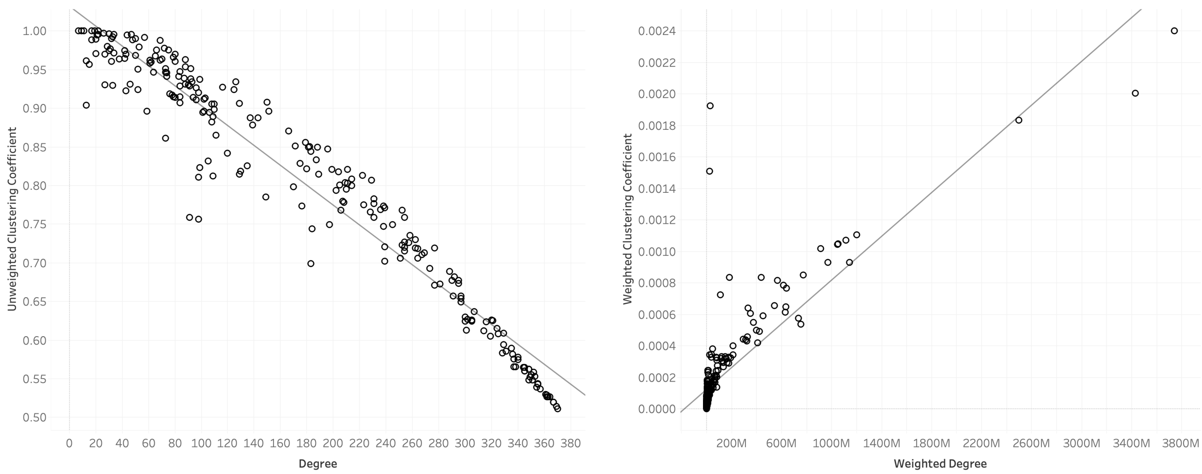


Figure 3.7 Red-Yellow-Green Color-Coded Heat Map (Centrality Index in Descending Order: Red, Orange, Yellow, Light Green, Dark Green)

of the USA and China relative to other countries significantly, because of their colossal trade flows. Japan wasn't included within the 10 most central countries when not using weights, but now is. France's centrality is reduced. Not only that, but also Japan and the Netherlands now become more central than Great Britain. Canada loses a couple of rankings, and Italy, Spain, and Switzerland aren't within the most central countries anymore, and are instead replaced by Hong Kong, Japan, and South Korea. Figure 3.7 shows a red-yellow-green color-coded heat world map according to weighted degree to visualize the centrality index of the countries in world trade.



(a) Binary **(b)** Weighted

Figure 3.8 Scatter Plots of Clustering Coefficient

3.4.5 Unweighted Clustering

For unweighted graphs, the clustering of a node is defined as the fraction of existing triangles over the total possible triangles that could be formed through that node. For the unweighted analysis of clustering, the correlation between the clustering coefficient and the degree was computed and is -0.97. The results obtained indicate that there is an extremely strong inverse relationship between the amount of countries a determined country interacts with (degree) and the amount of the triadic closures that it forms. This implies that countries with numerous trade interactions have partners that do not tend to trade among themselves, hence triangles are not formed. The results obtained in this section are in consensus with what has been found by the bulk of the other authors who have computed this metric for the WTW. Figure 3.8a shows the scatter plot of unweighted clustering coefficient and unweighted degree.

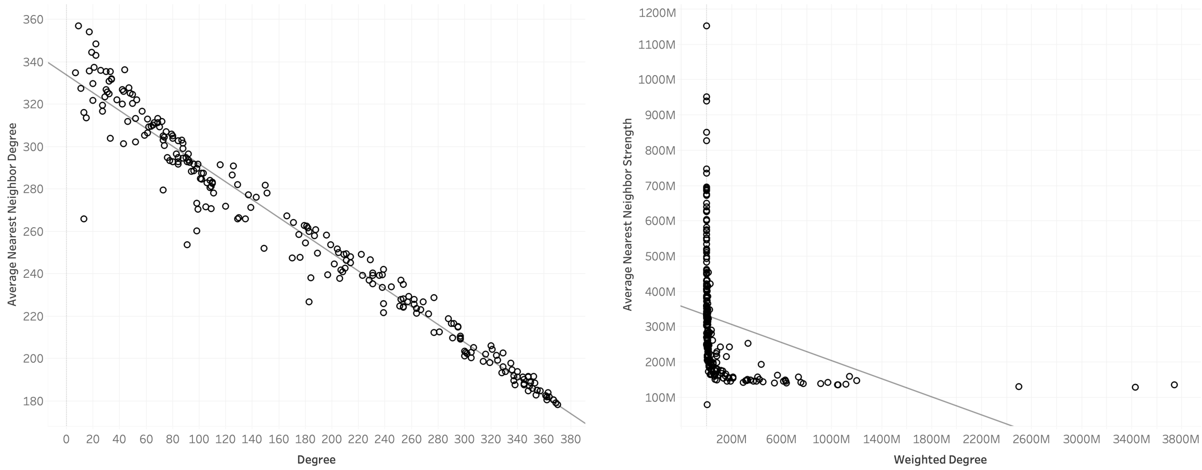
3.4.6 Weighted Clustering

For weighted graphs, the metric to be used is the directed weighted clustering coefficient for directed graphs, developed by Fagiolo[14], which is a weighted-directed adaptation of the well-known binary clustering coefficient. The directed weighted clustering coefficient is defined as the geometric average of the subgraph edge weights, and it measures the tendency of a network to form neighborhoods that are strongly connected.

For the weighted analysis of clustering, the correlation between the weighted clustering coefficient and the weighted degree was computed and is 0.84. The results obtained indicate that there is a very strong direct relationship between how strongly a country interacts with other countries and the strength of the trade relationships between the countries it trades with. Note that using weights, the result is the polar opposite than with the unweighted version in the previous section. The results hereby obtained are significantly different than the ones from other authors. Chow [58] finds a virtually null correlation for the same variables for the year 2009. Additionally, Fagiolo, Reyes, and Schiavo [29] find a positive and significant relationship between the variables, however said relationship is extremely weak by the looks of the scatter plot shown (no correlation was provided). The results of Fagiolo, Reyes, and Schiavo [7] show similar results to those from Fagiolo, Reyes, and Schiavo [29], but still no correlation was provided. Figure 3.8b shows the scatter plot of weighted clustering coefficient and weighted degree.

3.4.7 Unweighted Assortativity

To assess the assortativity of the network, two exercises can be performed. The first one is to compute the Degree Pearson Correlation Coefficient (DPCC) of the entire network, which is an algorithm developed by Newman [71] and then adapted for directed networks by Foster [72]. The second one is the correlation between average nearest neighbor degree



(a) Scatter Plot of ANND and Degree **(b)** Scatter Plot of ANNS and Strength

Figure 3.9 Assortativity

(ANND) and node degree (ND), as performed in other works [7, 29, 30, 57, 58].

The observed correlation between ANND and ND is -0.98, and the unweighted DPCC of the network is just -0.27, consistent with the previous correlation in direction but not in magnitude. The results hereby obtained are in consensus with what most of the studied authors have obtained, suggesting that from the unweighted perspective the network is highly disassortative. This suggests that countries that are well connected tend to interact with countries that aren't well connected themselves. Figure 3.9a shows the resulting scatter plot for ANND and ND.

3.4.8 Weighted Assortativity

A similar exercise to the one performed in the previous section is performed but including weight in the analysis. The homologous for ND is neighbor strength (NS), and for ANND it's average nearest neighbor strength (ANNS). A weighted version of the DPCC is computed as well, and a score of -0.06 is obtained. The correlation between ANNS and NS is -0.31. The results obtained suggest a weak disassortativity of the weighted network, considerably weaker than the ones obtained from the previous section. The correlation obtained is similar to the one achieved by Fagiolo, Squartini, and Garlaschelli [30], however the DPCC we obtained appears to be considerably lower. Similar results in the correlation between ANNS and NS are obtained by Fagiolo, Reyes, and Schiavo [7, 29], but no DPCC was provided. The results suggest a weakly disassortative weighted WTW, meaning that countries that are stronger tend to, to a low extent, trade with countries that are less powerful. Figure 3.9b shows the resulting scatter plot for ANNS and NS.

3.4.9 Disparity

To measure how well distributed a countries' exports are, the Herfindahl Hirschman Index (HHI) is going to be used [73, 74]. The previous index is well known for its ability to detect concentration and is mostly used in economics for market share concentration and monopolistic analysis.

The HHI can be extrapolated to any other field. In this case, if a country exports the totality of its exports to a single country, the HHI for that country will have a value of 1; on the other hand, if they export an equal percentage of their total exports to all of the countries they trade with, the value of the index is going to be equal to 1 divided by the number of partners, which serves as a reference value. The lowest possible HHI (total disparity) is 0.0042 (1/237) in case a country exports the same percentage of its total exports to each one of the other 237 countries. This serves as a reference value for the rest of the HHI values. Both the HHI of imports (HHIImp) and the one of exports (HHIExp) will be further analyzed.

Having a high HHI implies a high dependency and attachment to few countries, which could cause the economic shocks of one country to be easily transmitted to countries that are trading a high percentage of their total GDP with said country. This serves as a recommendation to those countries to diversify their portfolio of exports in order to decrease the dependency on their trading partners. For practicality, Antarctica will be ignored for the analysis, given that it's conformed by just 5 very small territories, and given their characteristics, have an outstandingly high HHIImp, and no exports. As a reference, the average HHIExp among all the countries is 0.167, and for HHIImp it's 0.2. Also, the reader should note that, as mentioned earlier, the entirety of the 238 countries included in the analysis import, but only 142 countries export. This happens mainly because small countries tend to just import and are mainly tourist destinations, hence their economies grow through tourism and not exports.

Figure 3.10 shows how the HHIExp is distributed by color coded continent, and one can use table 3.4 as support for the following analysis. The continent with the highest HHIExp is North America, with an average score of 0.30. Among the high outliers, one finds Bermuda (HHIExp = 0.67), Mexico (HHIExp = 0.64), and Canada (HHIExp = 0.57). The continent with the second highest concentration is Oceania, with a value of 0.20. Among the countries with the highest concentration one finds Solomon Islands (HHIExp = 0.44), Palau (HHIExp = 0.24), and Samoa (HHIExp = 0.17). The third continent with the highest concentration is Asia, with an average of 0.18. Two of the countries belonging to this continent have the highest concentration, which are Kuwait (HHIExp = 0.81), Mongolia (HHIExp = 0.74), and then there's Oman (HHIExp = 0.5). Africa has an average concentration very close to Asia, with a slightly lower value of 0.17. Among its most concentrated countries one finds Cabo Verde (HHIExp = 0.56), Swaziland (HHIExp = 0.49), and Lesotho (HHIExp = 0.45). The continent with the second lowest concentration is South America, with a value of 0.11, and it has no significant outliers worth mentioning. The continent with the lowest average is Europe, with a value of 0.10, and among the only outlier territories worth mentioning one can find Andorra (HHIExp = 0.39), and Albania (HHIExp = 0.3).

Figure 3.11 shows how the (HHIImp) is distributed by color coded continents, and table 3.4 serves as support for the following analysis. One can also get insights on how

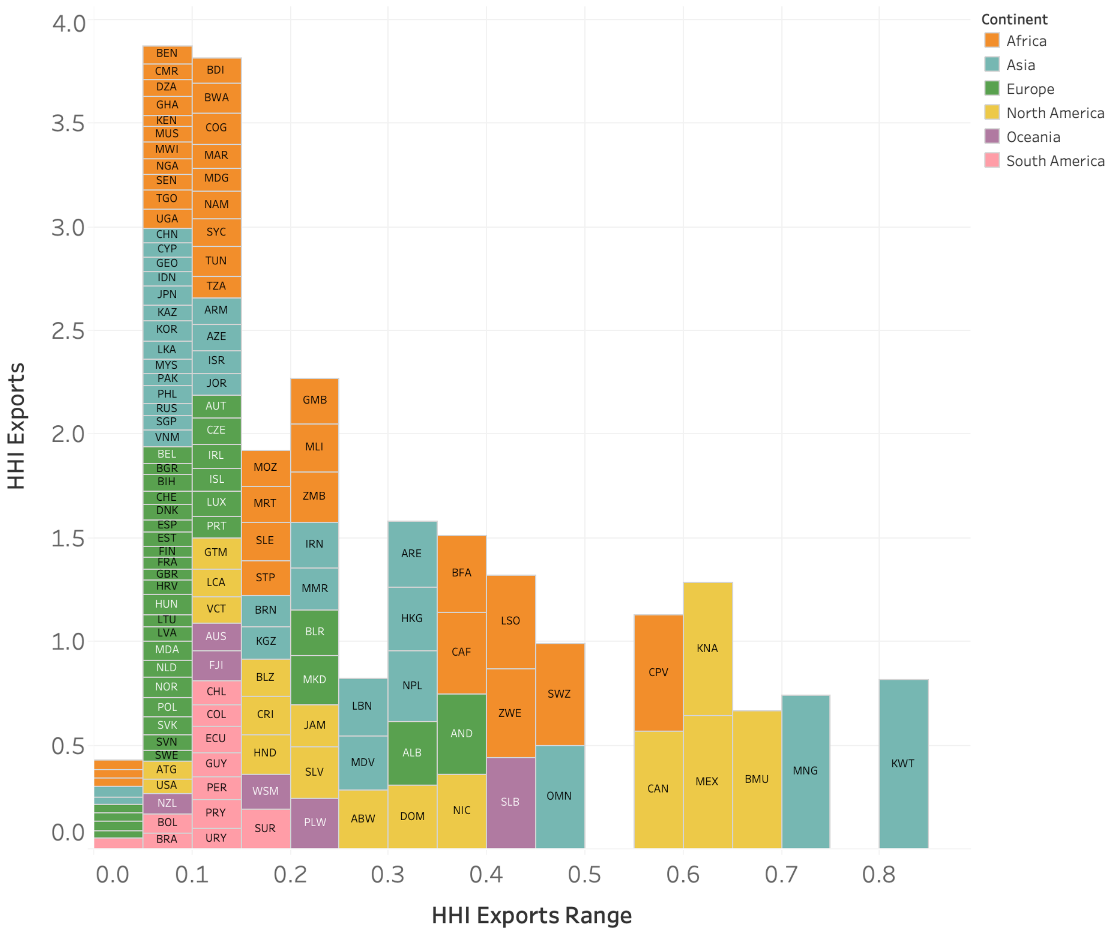


Figure 3.10 Structure of HHI Exports by Continent (Height of rectangle is proportional to the HHI for that country)

many countries are in each range of HHI values as well. Among the insights obtained from the visualization are that North America is the continent with the highest average concentration, having some small Caribbean countries with very high concentrations like Turks and Cacos Islands (HHIImp = 0.8), Saint Marteen (0.65), Saint Barthelemy (HHIImp = 0.55), but also some large economies like Canada (HHIImp = 0.41) and Mexico (HHIImp = 0.35). Oceania has an average of 0.21, being the second most dependent continent. Among the high outliers one finds Tokelau (HHIImp = 0.52), Nauru (HHIImp = 0.50), and Niue (HHIImp = 0.38). Africa is the third most dependent countries. Numerous countries of its countries have a relatively low HHIImp, and this can be reflected on their average score of 0.19, with only a few exceptions like Western Sahara (HHIImp = 0.93), Lesotho (HHIImp = 0.75), and Swaziland (HHIImp = 0.70). Asia has the same average score as Africa, and among the exceptions that have High dependency one can note North Korea (HHIImp = 0.89), Christmas Island (HHIImp = 0.75), and Bhutan (HHIImp = 0.65). One can note that in general, the countries and territories that appear to have a higher HHIImp are small territories and islands. South America has the second lowest HHIImp average, with a value of 0.15, with no significant outliers to be mentioned. Europe is the continent with the lowest average HHIImp, with a score of 0.12, meaning that they have their trade volumes more equally distributed and low dependency on singular countries. Among the only exceptions to the previous are European countries like Andorra (HHIImp = 0.45), Faroe Islands (HHIImp = 0.35), which is a self governing territory part of the Kingdom of Denmark, and Belarus (HHIImp = 0.375).

3.4.10 General Descriptive Network Statistics

The current section has as an objective to make an in-depth analysis into the topological characteristics of the individual continents and regions from the cross-sectional perspective and can also aid in determining how fragile the WTW is.

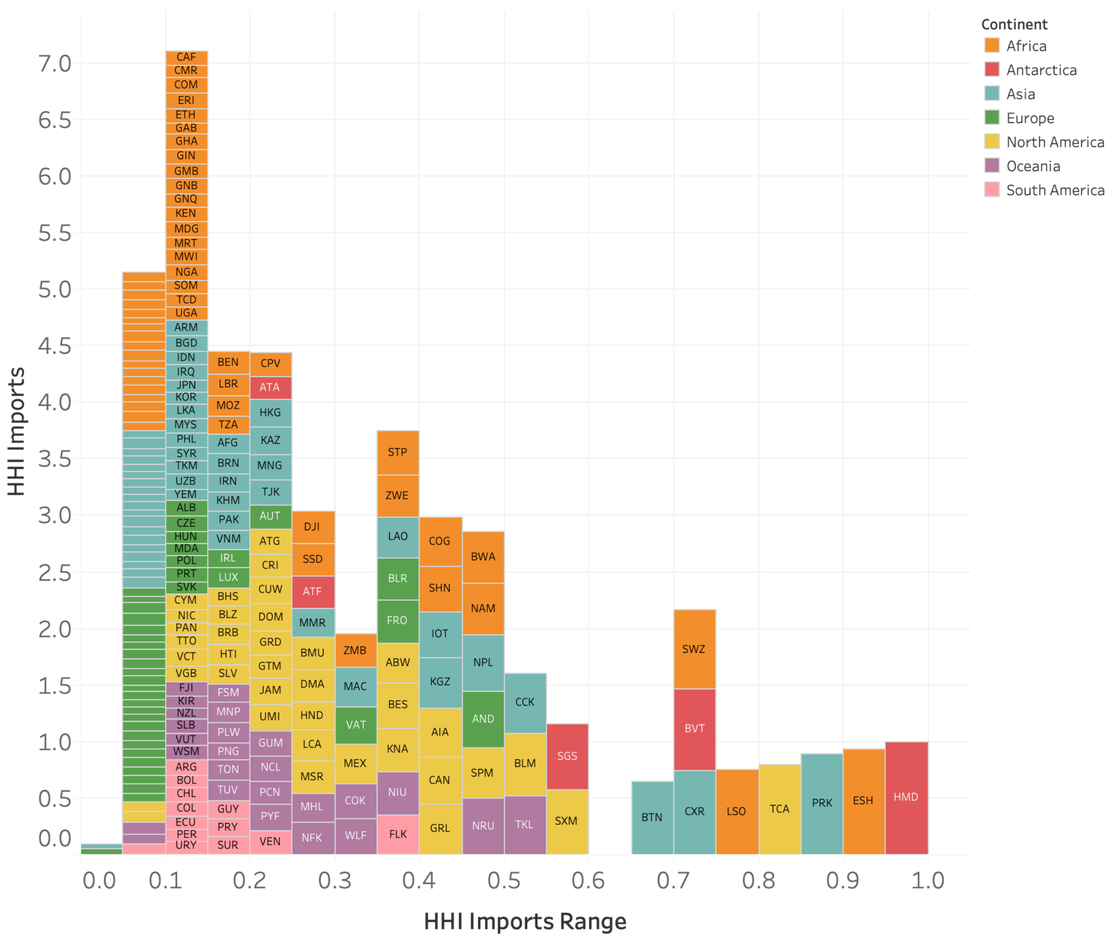


Figure 3.11 Structure of HHI Imports by Continent (Height of rectangle is proportional to the HHI for that country)

Table 3.4 HHI, Avg. Weighted Degree, and Avg. Weighted Degree/GDP by continent

	Africa	**Antica.**	**Asia**	**Europe**	**N.A.**	**Oceania**	**S.A.**
Number of Countries	54	5	53	40	37	25	13
Unweighted Centrality Index	-0.26	-7.87	1.39	3.70	-1.49	-4.32	0.89
Weighted Centrality Index	-1.16	-1.44	0.77	1.25	0.30	-1.13	-0.68
Degree	167	17	199	281	127	78	225
Weighted Degree (Millions)	15	108	230	296	138	21	68
HHI Exports	0.17		0.18	0.10	0.30	0.20	0.11
HHI Imports	0.19	0.56	0.19	0.12	0.27	0.21	0.15
ANND	265.98	336.25	252.19	215.53	275.92	300.76	237.41
ANNS (Millions)	306	679	274	205	352	516	246
PCGDP	2,530		14,350	33,095	13,985	13,427	8,568
%(Trade/GDP)	28.26%		27.26%	31.05%	11.60%	85.74%	4.59%
%(Exports/GDP)	14.94%		22.60%	38.42%	9.63%	5.45%	22.36%
%(Imports/GDP)	41.58%		34.19%	45.51%	38.72%	498.91%	19.37%

Antica. = Antarctica || N.A. = North America || S.A. = South America

Table 3.4 shows the averages per continent for miscellaneous metrics (except weighted degree, which is the sum of all the countries' weighted degrees), which consolidate relevant information for each one of the regions. The metrics include: for centrality, unweighted centrality index and weighted centrality index, degree, and weighted degree; for concentration, HHI of Exports and HHI of Imports; for characteristics of neighbors, ANND and ANNS; for income, PCGDP; for openness, total trade as a percentage of GDP, total exports as a percentage of GDP, and total imports as a percentage of GDP. Note that once again and for practicality purposes, the metrics of Antarctica will be once again ignored, given that it is formed by 5 countries and territories that are outliers in every metric and represent a very small percentage of the world's total population and production.

Regarding centrality, one can notice that in all of the metrics, the most central continent is by far Europe, and the least central one is Oceania. As mentioned in the previous section, the continent that is most diversified is Europe, and the one that is the most concentrated is North America. The continent that is on average connected to the best connected neighbors is Oceania (highest ANND), and the one connected to the worst connected neighbors is Europe. On the other hand, the continent connected to the strongest neighbors is Oceania (highest ANNS), and the one connected to the weakest is Europe. The continent with the highest average PCGDP is Europe, and the lowest is Africa.

Now, analyzing openness, the continent that trades the most as a percent of their GDP is Europe, and the lowest, by far, is South America. Notice that the second lowest is North America, so America in general trades a very low percentage of their GDP. Now, taking into consideration the percentage of exports over GDP, Europe is the continent that exports the most as a percent of its GDP, while Oceania is the one that does so the least. Now analyzing imports, the continent that imports the most as a percent of its GDP is Oceania, but this average is highly skewed upwards because of the Marshall Islands (strategic island for the USA during World War II, with a population of just 55,000), that imports 78 times its total GDP. Given the extreme outlier nature of the Marshall Islands, it will be excluded for this part of the analysis. When excluded, the average of Oceania drops from 498% to 42%, and is now surpassed by Europe as the continent that imports the most as a percentage of its GDP, and the continent that does so the least is South America.

Figure 3.12 shows the continental trade flows, and the following insights can be obtained from it and the data underlying it: Antarctica is a sink node, meaning that it just imports, with no outgoing flows; the continent with the highest influence in the WTW, followed by Asia and North America, while the rest of the continents render virtually insignificant to world trade; the continents with the highest intracontinental trade are, from highest to lowest, Europe, Asia, and North America. 64% of the entire trade in the world comes from just the interactions within and between North America, Europe, and Asia. South America, Africa, Oceania, and Antarctica are minor players in the WTW. Overall, the northern hemisphere is where close to two thirds of the total trade in the world takes place.

3.4.11 Correlations

The purpose of the current section is to analyze if the topological properties of the WTW, from a weighted analysis perspective, relate with the macroeconomic dynamics of growth and development. Given that this is atheoretical, exploratory analysis will be undertaken in order to look for relations between PCGDP and the different topological properties of the WTW. In the following section, a regression will be performed aiming to find statistical significance in dependence relationships between PCGDP and different topological properties, which would enable countries to fine tune their trade policies to maximize PCGDP.

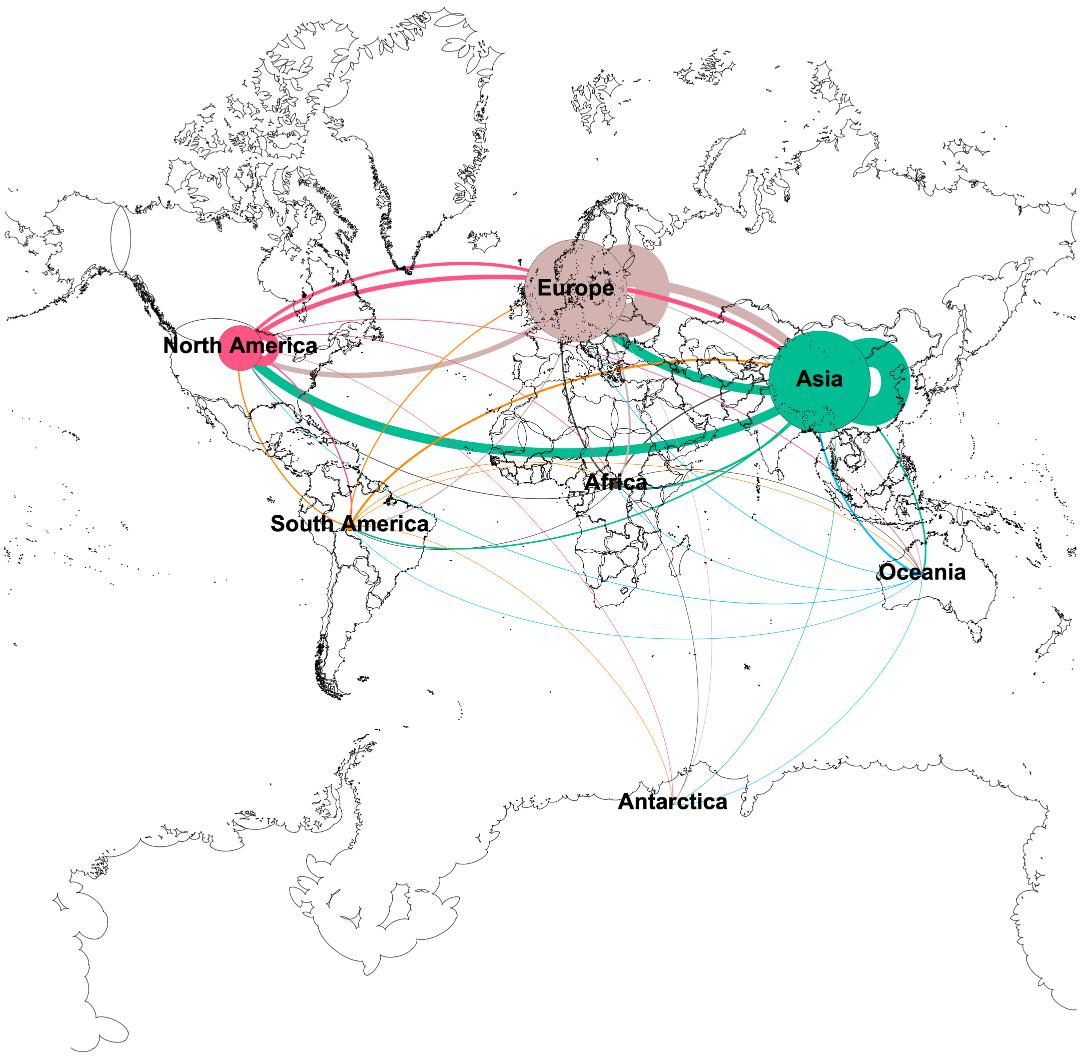


Figure 3.12 Continental trade flows. Nodes represent continents, edges represent flows. Edge colors are the same as the exporting continent's node. Node size is proportional to the weighted degree. Edge thickness is proportional to the magnitude of the trade flow. Self loops are attached to the right of each node and account for intracontinental trade and are also proportional to the magnitude of the total flows.

Table 3.5 shows the correlation matrix of numerous metrics, characteristic of the WTW. The abbreviation of some variables had to be modified for space efficiency. “WCent” is weighted centrality index (computed in section 3.4.4); “WCluster” is the weighted clustering coefficient; “Recip” is reciprocity; “WEigenv” is weighted eigenvector centrality; “Exp/GDP” is the percentage of total exports divided by the corresponding GDP; “Imp/GDP” is homologous of the previous one but for imports; “WDeg” is weighted degree; “Land” is the territory size in squared kilometers.

Table 3.5 Correlation Matrix and 1-Tailed Significance of Relevant Indicators[6]

Pearson Correlation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1.PCGDP	-	.36	.16*	.11*	**.42***	**.52***	**.49***	**.41**	.08	.36	.23	0	.36	.07
2.WCent	.36	-	.15*	.13*	.39*	**.46***	**.47**	**.92**	.19	**.96**	0.10	.08*	**.98**	**.51**
3.HHIImp	.16*	.15*	-	**.48**	.38	.37	**.40***	.25*	.17*	.07*	.03*	.12	.16*	.04*
4.HHIExp	.11*	.13*	**.48**	-	**.42**	**.42**	**.42***	.25*	.25*	.03*	0	.01	.15*	.03*
5.ANNS	**.42***	.39*	.38	**.42**	-	**.94**	**.96***	**.54***	**.69***	.38*	.32*	.10	**.41***	.23*
6.ANND	**.52***	**.46***	.37	**.42**	**.94**	-	**.97***	**.61***	**.55***	**.44***	.32*	.10	**.47***	.24*
7.Degree	**.49**	**.47**	**.40***	**.42***	**.96***	**.97***	-	**.63**	**.65**	**.45**	.34	.12*	**.48**	.27
8.WClust	**.41**	**.92**	.25*	.25*	**.54***	**.61***	**.63**	-	.28	**.86**	.23	.07*	**.96**	**.47**
9.Recip	.08	.19	.17*	.25*	**.69***	**.55***	**.65**	.28	-	.19	.31	.11*	.19	.17

[6] Pearson Correlation Absolute Values $>$ 0.4 and their significance in Bold, negative correlations with asterisk superscript. PCGDP = Per Capita Gross Domestic Product | WCent = Weighted Centrality | HHIImp = Herfindahl Hirschman Index of Imports | HHIExp = Herfindahl Hirschman Index of Exports | ANNS = Average Nearest Neighbor Strength | ANND = Average Nearest Neighbor Degree | WClust = Weighted Clustering Coefficient | Recip = Reciprocity | WEig = Weighted Eigenvector Centrality | E/G = Exports to Gross Domestic Product Ratio | I/G = Imports to Gross Domestic Product Ratio | WDeg = Weighted Degree |

10.WEig	.36	**.96**	.07*	.03*	.38*	**.44***	**.45**	**.86**	.19	-	.13	.05*	**.92**	**.52**
11.E/G	.23	.10	.03*	0	.32*	.32*	.34	.23	.31	.13	-	**.52**	.14	.12*
12.I/G	0	.08*	.12	.01	.10	.10	.12*	.07*	.11*	.05*	**.52**	-	.07*	.26*
13.WDeg	.36	**.98**	.16*	.15*	**.41***	**.47***	**.48**	**.96**	.19	**.92**	.14	.07*	-	**.48**
14.Land	.07	**.51**	.04*	.03*	.23*	.24*	.27*	**.47**	.17	**.52**	.12*	.26*	**.48**	-
1-Tailed Significance	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1.PCGDP	-	0	.03	.10	**0**	**0**	**0**	**0**	.18	0	0	.49	0	.20
2.WCent	0	-	.04	.06	0	**0**	**0**	**0**	.01	**0**	.13	.17	**0**	**0**
3.HHIImp	.03	.04	-	**0**	0	0	**0**	0	.02	.20	.38	.08	.03	.31
4.HHIExp	.10	.06	**0**	-	**0**	**0**	**0**	0	0	.35	.49	.46	.05	.35
5.ANNS	**0**	0	0	**0**	-	**0**	**0**	**0**	**0**	0	0	.12	**0**	0
6.ANND	**0**	**0**	0	**0**	**0**	-	**0**	**0**	**0**	**0**	0	.11	**0**	0
7.Degree	**0**	**0**	**0**	**0**	**0**	**0**	-	**0**	**0**	**0**	0	.08	**0**	0
8.WClust	**0**	**0**	0	**0**	0	**0**	**0**	-	0	**0**	0	.22	**0**	**0**
9.Recip	.18	.01	.02	0	**0**	**0**	**0**	0	-	.01	0	.10	.01	.02
10.WEig	0	**0**	.20	.35	0	**0**	**0**	**0**	.01	-	.06	.30	**0**	**0**
11.E/G	0	.13	.38	.49	0	0	0	0	0	.06	-	**0**	.05	.09
12.I/G	.49	.17	.08	.46	.12	.11	.08	.22	.1	.3	**0**	-	.22	0
13.WDeg	0	**0**	.03	.05	**0**	**0**	**0**	**0**	.01	**0**	.05	.22	-	**0**
14.Land	.20	**0**	.31	.35	0	0	0	**0**	.02	**0**	.09	0	**0**	-

There are numerous interesting and complex correlations that one can find from the previous correlation matrix. PCGDP is highly and negatively correlated with ANNS, and this was expected, given the results of section 3.4.8, where countries with a higher Weighted Degree have a lower ANNS. The relationship would be second order, given that countries with a higher weighted degree tend to also have a higher PCGDP, as observed on the able, so the correlation between weighted degree and ANNS, and PCGDP and ANNS would be expected to be very similar, which is the case. A similar and expected relationship is observed between ANND and PCGDP, which was expected based on the results of section 3.4.7, where countries with a higher ANND have a lower ND, and according to the correlation matrix, countries with higher ND have higher PCGDP. Hence, there is an inverse relationship between PCGDP and ANND. One can also note that countries that have a higher PCGDP also tend to be more clustered.

Now, analyzing the correlations with weighted centrality, one can see obvious very high positive correlations with weighted eigenvector centrality and weighted degree, given that these metrics were used to compute the weighted centrality. However, it is interesting to observe that countries with higher weighted centrality index also have a higher degree. Another result worth noting is that countries with more territory have a higher centrality, and this could be due to the fact that more territory has a high probability to come with more natural resources and borders, as well as access to the ocean, hence more trade and centrality.

It is worth noting as well that countries that have concentrated export destinations also have concentrated import origins, so they tend to be highly dependent on both flows. Additionally, countries with strong and well-connected neighbors tend to have more concentrated exports. Also, having fewer trade connections is related to countries with poor export diversification.

Countries whose neighbors are strong on average tend to be less clustered, a lower reciprocity of trade links, tend to be themselves poorly connected, and trade less, but their neighbors are well connected. Also, countries that are connected with neighbors that themselves are well connected are less clustered, reciprocate less trade links, tend to connect stronger with weak neighbors, trade less, and are better connected.

Countries that are better connected tend to cluster, trade and reciprocate, and connect more strongly with stronger neighbors. Countries that are more clustered tend to connect more strongly to strong neighbors, trade more, and have more territory. Countries who connect more strongly with stronger neighbors tend to be larger and trade more. It is also interesting to note that countries that export a high percentage of their GDP, tend to also import a high percentage of it. Finally, larger countries tend to trade more.

3.4.12 Cluster Analysis

Cluster analysis (CA) is a renowned and acknowledged method to group objects (countries in this case) based on analogous characteristics that they possess [75, p. 418]. To group the countries based on their network characteristics (the same variables from the previous section are used), CA was used.

Mahalanobis distance D2, a multivariate distance measure that equally weighs each variable [75, p. 432], was used to remove outliers (with a Chi-squared distribution transformation to statistically determine outliers). Outlier removal is necessary because CA is sensitive to them, creating additional clusters that include outliers only and thus unnecessarily complicating the analysis. These outliers included countries that have large economies and are highly involved in international trade, such as Mexico, China, USA, Germany, and Russia. However, these are later added to the cluster that is the most similar to them.

Hierarchical methods are useful to determine the optimal number of clusters [75, p. 444]. Hence, a dendogram was used to determine the optimal number of clusters, and the clustering algorithm used for the computation was Ward's method, using squared Euclidean distances. The optimal number of clusters is 3.

It is commonly accepted by the research community to use a combination of both hierarchical and non-hierarchical methods which counterweighs their benefits and weakness [76]. Hence, once the number of clusters was determined with the hierarchical method, this information served as an input for the non-hierarchical method, since the latter needs the number of clusters as an input [75, p. 444].

A K-means non-hierarchical procedure followed, which is a method that minimizes the intra cluster distance and maximizes the inter cluster distance. This is desirable because the members of each cluster should be highly homogeneous, and the clusters should be heterogenous [75, p. 444]. Each of the 3 clusters is labeled based on its main characteristics. Table 3.6 shows the final cluster centers. The first cluster (left column, 54 member countries) is labeled "Low Income Poorly Connected", given that it groups countries with a low PCGDP, small territories, weak trade (low exports and imports as a percent of GDP, as well as low weighted degree), few trading partners (low degree), connected to strong and well connected neighbors (high ANND and ANNS), weakly connected to stronger neighbors (low weighted eigenvector centrality), concentrated import origins, weak neighborhoods (low weighted clustering coefficient), and little reciprocating with their partners. The third cluster (right column, 22 member countries) is labeled "High Income Well Connected", given that it groups countries with high PCGDP, large territories, strong trade (high exports and imports as a percent of GDP, and high weighted degree), numerous trading partners (high degree), connected to weak and poorly connected neighbors (low ANND and ANNS), strongly connected to stronger neighbors (high weighted eigenvector centrality), diversified import origins (low HHIImp) and moderately diversified export origins (moderate HHIExp), strong neighborhoods (high weighted clustering) and reciprocate most of their trade links (high reciprocity). The second cluster (middle, 98

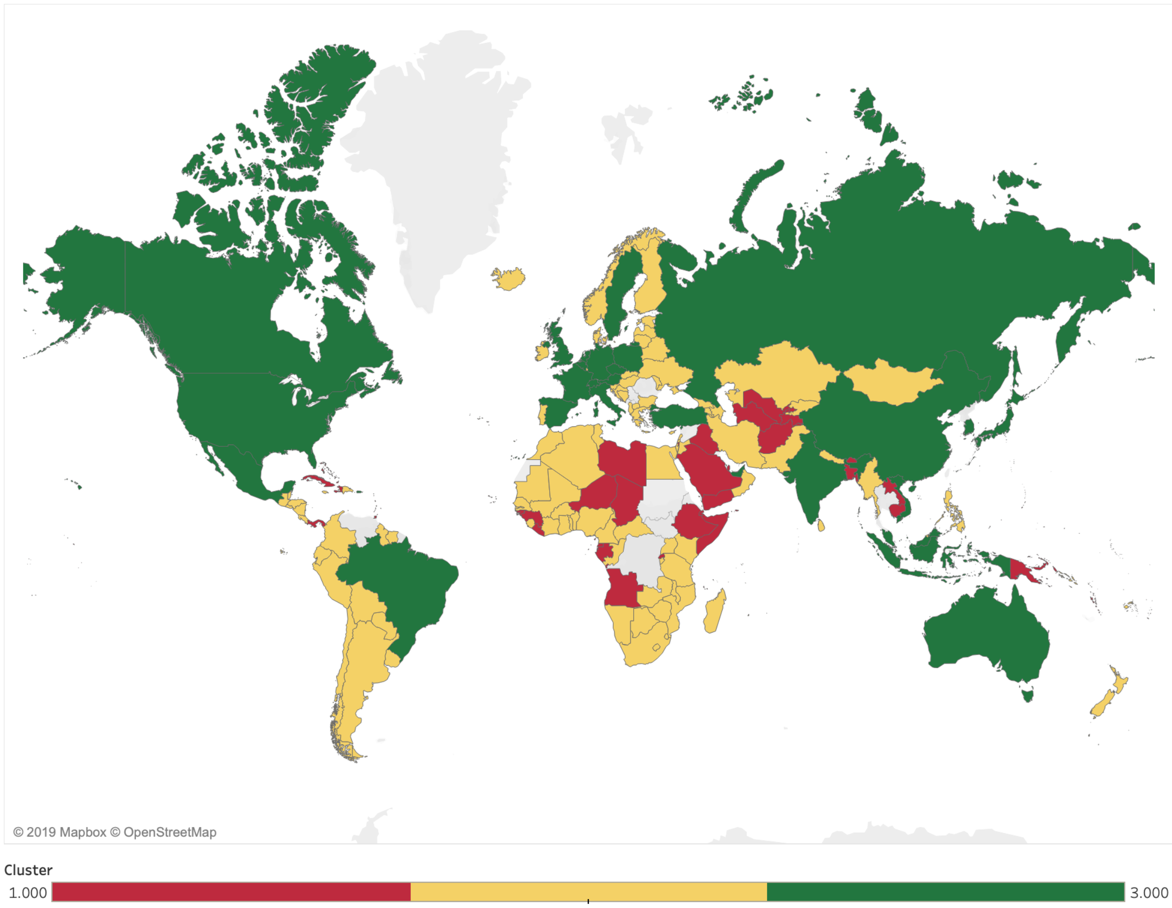


Figure 3.13 Cluster Map

1 (Red) = Low Income Poorly Connected, 2 (Yellow) = Low Income Moderately Connected, 3 (Green) = High Income Well Connected

member countries) is a middle ground between cluster 1 and 3 (in the middle for most network characteristics), with the exception that the PCGDP is virtually the same as the cluster of Low Income Poorly Connected countries, and is thus named accordingly. Note that at the end of the procedure, the outliers were incorporated to the cluster of High Income Well Connected countries, given that it was the one that resembled their characteristics the most. Figure 3.13 aids in visualizing the clusters.

Table 3.6 Final Cluster Centers

Cluster	Low Income Poorly Connected	Low Income Moderately Connected	High Income Well Connected
Land	0.03	0.04	0.14
Exports/GDP	0	0.22	0.37
Imports/GDP	0.14	0.13	0.13
WCent	0.02	0.05	0.51
Degree	0.15	0.61	0.94
WDeg	0.01	0.04	0.51
WEigenv	0.01	0.03	0.43
ANNS	0.42	0.15	0.02
ANND	0.78	0.35	0.06
WClust	0.09	0.11	0.64
HHIImp	0.17	0.16	0.07
HHIExp	0.04	0.22	0.12
PCGDP	0.1	0.11	0.3
Reciprocity	0.04	0.88	0.91

[a]

[a]For more information on abbreviations, see table 3.5

Table 3.7 Regression Results and Accuracy

$\beta 0$	$\beta 1$		$\beta 2$		$\beta 3$		$\beta 4$		R^2
Unstd.	Unstd.	Std.	Untd.	Std.	Unstd.	Std.	Unstd.	Std.	.195
32471**	0.00005479**	(.34)	(15985)**	(.27)	70734**	.25	14014*	.18	-

*Significant with 95% confidence | **Significant with 99% confidence | Unstd. = Unstandardized Coefficients | Std. = Standardized Coefficients | Numbers in parenthesis are negative

3.4.13 Multiple Regression

Numerous regression models with PCGDP as the dependent variable were attempted but, as expected given the numerous highly correlated metrics, variance inflating factors (VIFs) which indicate multicollinearity, tended to be extremely high. Given the previous circumstances, numerous variables had to be eliminated in order to solve the undesirable multicollinearity problems and to just retain a handful of significant and insightful variables, that could help policy makers to fine-tune their trade policies in order to improve their PCGDP. The final regression model to be used is shown in 3.1.

$$PCGDP = \beta_0 + \beta_1 ANNS + \beta_2 Reciprocity + \beta_3 Weigenv + \beta_4 ExpGDP \tag{3.1}$$

"PCGDP" is per capita GDP, β_0 is the constant of the equation, "ANNS" is average nearest neighbor strength and β_1 is its coefficient, "Reciprocity" is self-explanatory and β_2 its coefficient, "WEigenv" is the weighted eigenvalue and β_3 its coefficient, and "ExpGDP" is the percent of total exports over GDP and β_4 its coefficient. The resulting coefficients, t-values, and significance are shown in table 3.7.

As observed in table 3.7, the R^2 is 0.195, and a score this low was expected, given that PCGDP also depends on numerous other hard economic variables like quality of education, efficiency of government spending, rule of law, efficiency of institutions, technology, among others. The purpose of the current exercise is to see if the topological properties of the

WTW included in (1) are statistically significant explaining PCGDP, hence the ability to make accurate predictions with the independent variables is not of interest for the current study.

There is a highly significant inverse relationship between ANNS and PCGDP, meaning that to increase a country's PCGDP, it should associate with countries that are weaker. The previous supports the core-periphery model of dependency, where weaker countries are more strongly connected with developed, stronger countries. There is also a highly significant inverse relationship between PCGDP and the reciprocity, which would suggest that having a high reciprocity with your trading partners tends to decrease your PCGDP, so one should seek less symmetry (more research on how exactly this happens should be undertaken). Another result is that there is a direct and highly significant relationship between weighted eigenvector centrality and PCGDP, which means that countries with higher income per capita tend to associate more strongly with stronger partners, so if a country desires to increase their PCGDP they should focus on trading more with countries that are strong. The results also show a high significant direct relationship between PCGDP and the percent of total exports over GDP, which means that in order to increase the income of their population, countries should focus on exporting more, which has been the model of growth during the previous couple of decades for China.

All of the research questions that were formulated at the end of the introductory section will be addressed in the following section. Possible future work that could be done using the present research as a foundation is proposed as well.

3.5 Conclusions and Future Work

The first conclusion is that when analyzing the WTW from a binary perspective, it doesn't seem to follow a power-law distribution. However, when analyzing it from a weighted perspective, it does appear to follow a power law distribution, where numerous countries have weak trade links, and few countries have extremely strong trade links.

A structure of the communities was displayed in Figure 3.7, with evidence that countries that are geographically closer tend to form more and stronger relationships, which is expected because of transportation costs and trade agreements that tend to happen among countries that are geographically close. Also, one can observe the communities and infer that trade agreements also increase the intensity of trade among communities.

The continent most susceptible to instability spread through their trading partners is North America, which has by far the highest HHI both for exports and imports. However, on average, they trade the least percentage of their total GDP (next to South America), minimizing the percentage of their economy that depends on foreign countries. The previous information can be observed in table 3.3. The most central continent is Europe, which also has the lowest HHI of export and imports, which helps it partially shield from instability originating from other countries. However, they are one of the continents that on average trades the most as a percentage of its GDP, which could facilitate the transmission of instability from other continents and territories.

According to the Unweighted Centrality Index computed in section 3.4.3 using PCA, the most central countries, in descending order, are the following: USA, China, Germany Great Britain, Netherlands, France, Canada, Italy, Spain and Switzerland. However, when one takes into consideration the weights, which was addressed in section 3.4.4 constructing a Weighted Centrlaity Index using PCA, the picture changes. The most central countries in descending order are the following: USA, China, Germany, Netherlands, Japan, Great Britain, France, Hong Kong, Canada, and South Korea. Comparing the results between the unweighted centrality index and the weighted one, one can see that in general and as expected, when using the weights the importance of the USA and China are boosted relative to other countries, mainly because of their massive trade flows. Also, the ranks between the countries are shifted. Spain and Switzerland are no longer within the 10 most central countries, and are instead replaced by Hong Kong, Japan, and South Korea. This can be observed in tables 3.2 and 3.3. The results obtained are significantly different from that of other authors, mainly because of the database used, as well as because 6 centrality metrics were incorporated using PCA for the centrality indices.

Among the large countries (by the size of their GDP, with the goal of excluding small Islands) that have a substantially high dependency on others based on their HHIImp and HHIExp are: Mexico (HHIExp = 0.64, HIIImp = 0.35), Canada (HHIExp = 0.67, HHIImp = 0.41), Kuwait (HHIExp 0.81), Mongolia (HHIExp = 0.74), Andorra (HHIExp = 0.39, HHIImp = 0.45), Albania (HHIExp = 0.3), Belarus (HHIImp = 0.375), Lesotho (HHIExp = 0.74, HHIImp = 0.75), Swaziland (HHIExp = 0.49, HHIImp = 0.70), and North Korea (HHIImp = 0.89).

Using an unweighted approach in section 3.4.7, one can observe that countries with high degree tend to associate with neighbors that themselves have low degree, pointing to strong disassortativity both through the correlation of ANND and ND, as well as the DPCC. However, this drastically changes when using the edge weights, as the correlation between ANNS and NS is significantly lower, as well as the DPCC. However, both metrics still point to disassortativity, regardless of it being weaker than from the unweighted approach.

Taking into consideration numerous network metrics and their relationships (correlations), there are several key insights that one finds. Countries that have a high concentration on their imports (HHIImp) tend to also have a high concentration on their exports (HHIExp), making them more vulnerable through both concentrations. Countries with higher PCGDP tend to associate with countries that are themselves weakly and poorly connected, pointing to a core-periphery structure. Also, countries with higher PCGDP tend to be better connected and more clustered. Having more land (more territory) is associated with being more central. Countries whose neighbors on average are stronger and also better connected, tend to be less clustered and to reciprocate their links to a lesser extent. Lastly, countries that tend to associate more strongly with countries that are themselves strongly tend to cluster more heavily and have more territory.

According to the results in section 3.4.13, actions that countries could take to improve their PCGDP include associating with more neighbors that are themselves weaker; reciprocate fewer of their trade links; trade more strongly with countries that are themselves stronger; and increase their export to GDP ratio.

This research provides the foundation for a critical latent line of investigation regarding public policy, specifically trade policy. With the numerous insights hereby conveyed, trade policy can be tuned at the country and continent levels in order to minimize the magnitude of dependencies while still being able to benefit of the gains of trade. The current research describes to fine detail the binary and weighted characteristics of the WTW using the most complete database that has been identified to date, and provides information that is helpful towards determining what actions can be taken to improve income per capita.

Plenty of research can be inspired from the input of this paper, for example: How is reciprocating less trade links associated with a higher PCGDP? Is it because they export to more countries they don't import from? Or the other way around? Why is associating more strongly with stronger partners associated with a higher PCGDP? Is it because those partners aid with technology transfers to the less developed country that associates with them? Is it just because it allows for more exports to larger countries? Or is it because you get catapulted to more trade partners when associating with them? What can be done to increase the connectivity and exports of Africa? Or is Africa condemned by certain circumstances to have trade deficits and be poorly connected? Is it just a matter of time before this improves? Which countries have been able to decrease their dependency (measured by HHIImp and HHIExp) through time? How did they achieve this? Will their methods work for other countries and help them be less vulnerable?

3.6 Limitations and Discussion

One of the main limitations in this chapter, is that the weighting mechanism for the edges in the WTW is subject to discussion, depending on the objective that the researcher has. In the literature, there are authors that use other weighing mechanisms, such as the ratio of a particular trade flow to GDP of the exporting country. There can be extensive debate when it comes to the advantages and disadvantages of each weighing mechanism based on what the objective that is to be attained is. Additionally, this analysis is performed from the cross sectional perspective, and a similar analysis could potentially yield significantly different results if made from the times series perspective, having either one country as the object of analysis, or the entirety of the WTW. Finally, the validity and usefulness of PCGDP to explain prosperity has been an object of extensive discussion in the scientific community within economics. Countries are diverging away from having PCGDP as one of the main economic performance metrics, and are gravitating gradually towards other complementary indicators that could better measure quality of life, such as education, healthcare, sustainability, and happiness.

In the next chapter, we use a deep neural network for link weight prediction in the WTW. The predictive model is based on the gravity model of trade, as defined by the prominent economist Walter Isard [22]. We aim at being able to predict the trade flows between countries more accurately and with a lower variance than other works that use techniques based around Ordinary Least Squares Regression, specifically the ones of Rose [23] and Head [24].

Chapter 4

The World Trade Web: A Deep Learning Approach to Link Weight Prediction

4.1 Introduction

Studying the world trade web (WTW) is of great importance and countries can use it as a tool to design their trade policy. The global gross domestic product (GDP) for 2017 (also known as Gross World Product, GWP) is 80.14 trillion USD, out of which 16.3 trillion USD (20%) comes from trade. This does not account for all of the indirect benefits that trade brings with it like employment in factories, shipping and logistics companies, research and development, technological advances and transfers. Additionally, trade enables the availability of products and services in regions which otherwise wouldn't have access to them. Any small island can be taken as a trivial example of a territory that just imports but doesn't have any exports, where imports are of crucial importance for the basic functioning and well-being of the country. Needless to say, trade has played a crucial role in the ever-increasing important interactions among countries that in turn accelerate globalization [2]. It is also well known in the economic literature that there are gains of trade that result from specialization in production due to division of labor, agglomeration, economies of scale, scope, an increase in the total production possibilities, and trade through markets by selling one type of product for other more valuable goods [77]. Furthermore, understanding the characteristics of the WTW aids in comprehending the structure of the trade network and pinpoint specific channels of propagation of economic and financial disasters and shocks, thus enabling policy makers prevent and prepare for them. Naturally, this implies an interdependency among countries, since a reduction of a country's exports to another one can inhibit the latter's ability to manufacture exportable goods to its trading neighbors, a negative ripple effect [40, 41, 78–81]. Trade flows have been shown to be highly correlated with other country interactions such as flows of services, workers, and financial assets, hence being a relevant indicator for broader economic relations [42].

Surveying existing approaches for link weight prediction in the WTW, one finds that economists have addressed this issue with a method known in the literature as "The Gravity Model of Trade", which was originally theoretically proposed by Walter Isard in 1954 [22], then further expanded and popularized by Dutch economist Jan Tinbergen in 1963 [82]. The bulk of the empirical work in the existing literature uses econometric approaches with variations of the aforementioned model to predict the magnitude of trade links between countries. Using econometric approaches in this context is useful because it allows to interpret the statistical significance, direction, and magnitude of the impact of each one of the variables in the model. However, econometric approaches have been used for numerous years, which begs the question on whether the accuracy of the predictions can be improved with newer algorithms and techniques. This leads anyone within the computer science field to wonder if this problem has been addressed using deep neural networks (DNNs), techniques which usually provide an adequate solution to the modelling and prediction of complex problems like the one hereby addressed. To the best of our current knowledge, no approach has been taken where deep learning is used for link weight prediction in the WTW, which motivates the current chapter. The question that naturally arises and that is to be answered in this chapter is the following: can DNNs beat the accuracy of existing econometric models for link weight prediction in the WTW? If so, how can this be useful for public policy decision-making?

Forecasts of exports and imports are central providers of economic forecasts, which can aid in public policy decision-making. More accurate forecasts result in better GDP forecasts, which can aid in commercial policy, particularly the optimization of tariffs, quotas, and subsidies. Additionally, predicted trade flows that are significantly lower than than the observed trade flows, could arguably provide some evidence towards the existence of informal trade between two countries. The results obtained in this chapter enable the identification of such estimation errors, which could potentially shed some light as to which countries are transacting informally. The consequences of informal trade can be substantial, by reducing the tax collection by the state, thus reducing the tax base, generating unfair competition for enterprises, and endangering intermediate and final consumers with products that aren't inspected at a point of entry. Henceforth, it's in the countries' best interest to identify and mitigate informal trade.

The use of DNNs for link weight prediction in the WTW showed an improvement in performance of as much as 8.2% over multiple regression, while also reducing the standard deviation of the standard errors by as much as 13.8%. However, it should be taken into account that when using nonlinear activation functions in DNNs, the interpretation of the parameters is lost (known as the black box of deep learning). Therefore, using DNNs is more accurate in link weight prediction, but it comes with the cost of the loss of interpretation of the parameters. If one wishes to identify the impact that each one of the variables has on the trade between countries, multiple linear regression should be used.

In domains where the cost of the loss or error is extremely high (say for example, cancer detection), using DNNs would prove to be more convenient given the increase in accuracy and reduction in loss, even when it is just a moderate 8.2%, like in this specific use case. Slight accuracy improvements like the one hereby achieved could make a vast difference in humanity's wellbeing in certain domains. This improvement in the prediction accuracy of trade flows is relevant because it allows for better GDP forecasts, which can signal public policy decision makers to actions that can be taken regarding tariffs, subsidies, and quotas to mitigate potential external shocks that could arise from hindered trade with relevant trade partners. Predictions with significant error could potentially signal to informal trade routes between countries.

The rest of this chapter is organized as follows: firstly, the existing theory and empirical works for link weight prediction are explored; furthermore the approach to be taken is desribed; follwowing, the datasets used and the structure of the models are described; next, the results obtained and their interpretation are showed; then, the hardware and software used in the experiment are described, and access to the code for reproducibility purposes is provided; finally, the last section goes through the conclusions and discussion.

4.2 Literature Review

In order to determine the features that should be included for link weight prediction within the WTW, one should first go to theoretical resources that build on the factors that incentivize trade between countries. A solid ground to begin looking for such resources is within the theoretical work and empirical findings of works in economics. The base theories, as well as the diverse adaptations stemming from said theories that have been made by several authors, should serve as a foundation in determining variables to include in our DNN for prediction.

In the literature related to trade, one finds two main theoretical approaches: location theory and trade theory. Location theory is commonly applied on research when one wants to understand the factors that incentivize international companies on where they locate their foreign operations. Location theory tries to find the optimal location of production, which is in function of the cost of the different factors of production (commonly capital and labor), as well as the transportation costs to consumers, which is usually viewed in terms of individual industries. On the other hand, trade theory attempts to explain the patterns of international production and trade, which are in function of the relative endowments of factors of production necessary to produce determined goods, as well as comparative advantage. Trade theory has traditionally viewed this in terms of aggregates like total exports, imports, national income, and so forth. Small differences aside, there is a considerable overlap between trade theory and location theory. They address similar questions and make similar assumptions on their theoretical approaches. However, none of them is able to predict specific countries or regions where production of a determined good will be located [22, 83, 84].

It should be rather simple to determine why, in general terms, most of the traditional factors considered in location theory (except for transportation costs, which can be approximated by distance between countries) wouldn't work for what is intended in the current research, since those approaches are mostly intended to be studied on a per-industry basis, comparing and contrasting directly with other countries to understand why production of a determined industry takes place in one country or region based on their capital, labor, and cost structure. Also, finding data that is industry-specific for each one of the countries can prove to be a daunting task. Trade theory, on the other hand, could be more useful in guiding the modelling of this chapter, given that it uses aggregate terms and could ease the comparison between countries, where data is readily available on trustworthy government sources from each one of the countries as well as international organizations.

The most common theory that is the basis for the bulk of the authors that address this problem is currently known as "The Gravity Model of Trade" (GM), developed by prominent economist Walter Isard [22]. This theory mainly emphasizes the role of distance in trade, where shorter distances bring lower transportation costs, hence more trade, and vice versa. One of the main critiques that is made to previous approaches in pure trade theory is that there is a two-country abstraction with either zero or fixed transport costs assumed, where distance is implicitly neglected. It is argued that trade encounters spatial resistances of different magnitudes for each pair of regions or nations based on their geographical proximity. It is important to point out that using a variable like distance could be misleading, given that it is sensible to the country sizes. For example, two countries can share a border, but the distance between their centers can be significantly larger than another pair of countries, due to country size. The most important takeaway from Isard [22] is that bilateral trade between two countries is inversely proportional to the geographic distance between them. Empirical evidence for the GM is strong, specifically for the role of distance in trade links [85–88].

Expanding on the theoretical proposition of Isard [22], a remarkable Dutch economist, Jan Tinbergen [82], used an analogy involving Newton's Universal Law of Gravitation (this is why the model is currently known as the gravity model of trade) applied to bilateral aggregate trade flows between countries, where said flows are proportional to the size of the countries, and inversely proportional to their distance. Note that the size of the countries can be measured by gross national product (GNP) or gross domestic product (GDP), but the bulk of the literature frequents the latter when accounting for country size. The findings with both metrics are consistent, hence they are commonly used interchangeably within this context.

One particularly classical, valuable, and early empirical contribution that uses the (now known as) GM as its foundation, dates to 1963 [89]. The model used to estimate the flows of trade is shown in 4.1.

$$a_{ij} = cc_i c_j \frac{e_{ii}^{\alpha} e_{jj}^{\beta}}{(1+\gamma r_{ij})^{\delta}} \tag{4.1}$$

Where:

a_{ij} = estimate of the value of exports from country i to country j

e_{ij} = national income of the country of export i

e_{jj} = national income of the country of export j

r_{ij} = distance of transportation

α,β= national income of the country of export i

γ= transportation cost coefficient per nautical mile

δ= isolation parameter

c_i = export parameter of the country of export

c_j = import parameter of the country of import

c = a constant

As a general note, in 1963 (when equation 4.1 was estimated) the term "ordinary least squares regression" hadn't been popularized, but the authors mentioned that to estimate the parameters, the logarithmic residual sum of squared errors is to be minimized. Also, note that in the 1950s and 1960s economists used electromechanical desk calculators to estimate the parameters of regressions, and these computations generally took at least 24 hours. In this case, the analysis was made using an Elliot 803 data processing machine. The starting points for the estimation were α= β= 0.5; γ= 0.01; δ= 2; $lnc = 4$. The results were the following:

α= 0.518

β= 0.504

γ= 0.00157

$\delta = 1.817$

lnc = -3.818

The values of α and β indicate that an increase in countries' income of 1% is expected to increase exports and imports by around 0.5%. Furthermore, the fact that $\alpha + \beta \approx 1$ indicates a static nature of the model. The values of γ and δ indicate that an increase in the distance of transportation will decrease the magnitude of trade between countries. The previous work, besides being old, is a classic and the foundation of most of the empirical work that was undertaken in the following years. One will note that numerous models from various empirical works build on the foundation of this econometric model and estimations.

Well known American economist Paul Krugman [86] is among the empirical contributors to support the idea that country size is proportional to the magnitude of trade links. The model developed explains how trade flows are directly proportional to country size and additionally, that trade barriers have the opposite effect on trade flows. Krugman also builds on the causes of trade between economies that are similar in their factor endowments, which challenges the classic theoretical propositions of Heckscher, Ohlin, and Samuelson [90–93], whose model is commonly referred to in trade theory as the Heckscher-Ohlin-Samuelson model (HOMM). The primary work behind the HOMM was performed by Eli Hecksher in 1919 [90], further expanded by Bertil Ohlin in 1933 [91], with some final expansions performed by Paul Samuelson in 1949 [93] and 1953 [92]. The main idea behind the HOMM is that countries produce and trade based on their relative abundance of factors of production (mainly capital and labor). It establishes that countries with a relative abundance in capital will produce and export goods that are intensive in capital, and similarly countries that have a relative abundance in labor will produce and export goods that are intensive in labor. Nonetheless, even though HOMM is a reasonable and logical theoretical proposition, little evidence has been found to support it. According to the HOMM, developed countries should trade heavily with developing countries. However, empirical evidence shows that this is far from the case. As Steffan Linder [94] states, previous theories overlook one main factor when attempting to understand trade between countries: quality of products demanded. Linder argues that since countries with similar incomes require products of similar quality, this leads them to be more prone to trading with one another.

Judging by its nature, one can argue that the GM that takes just country size and distance into consideration for prediction of the magnitude of trade between countries can be and overly simplistic approach. Hence, as acknowledged by Dueñas and Fagiolo [95], country-specific explanatory variables that capture whether countries share a border, language, religion, trade agreements, access to the sea, among others can be used to expand this basic model. These findings are relevant, because it is identified identify that in order to properly predict the weighted properties of the WTW, one must fix the binary structure equal to the observed one, meaning that the existing models fail to predict non-existing links between countries (zeros in the adjacency matrix). The bulk of the available literature has indeed fixed the binary structure of the WTW when working with trade flows, given the increase in complexity that arises when trying to predict the existence or absence of a trade link.

There have been remarkable works [23, 24] that construct databases that are a precious input for link weight prediction in the WTW and are an extension of the GM, even though the focus is rather on tangential issues. For instance, Rose [23] studies the effect of multilateral trade agreements on international trade. To accomplish this, quantitative, dyadic binary, and categorical variables are used. The quantitative variables are: real trade (dependent variable); distance between the trading countries; their GDP; per capita GDP; and land area. The dyadic variables are binaries like: whether both countries are in the general agreement on tariffs and trade (GATT) or the world trade organization (WTO), or just one is in GATT/WTO, whether they belong to the generalized system of preferences (GSP), if they have a regional free trade agreement, a currency union, common language, land border, common colonizer, currently colonized, were ever colonized, and another binary variable that captures whether the two countries remained part of the same nation during the sample (e.g., Guadeloupe and France). The categorical values that can take 3 values are: if none, one or both countries are landlocked and whether none, one or both countries are islands. Rose compiles the database from prestigious sources like the international monteray fund (IMF), the Penn World Table, World Bank, the United Nations (UN), among others, and is conveniently available for public access at http://faculty.haas.berkeley.edu/arose/. Due to the ease of replicability, the default model shown in Table 1 [23] is used as a benchmark, given that the author reports the root mean squared error (RMSE) of the predictions, which can be easily compared to the RMSE of running a DNN with the same database and verify if this metric can be decreased with machine learning. In a similar fashion to Rose's work, another paper that serves as a benchmark for our paper is Head et al. [24], the database compiled by Rose is extended. The ordinary least squares (OLS) (1) model reported in Table 2 [24] can be easily replicated, given that the RMSE is reported and the database used is publicly available in http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8.

To the best of my knowledge, there has just been one attempt to estimate the informal trade by countries around the world [96], where an econometric approach of multiple indicators and multiple causes (MIMIC) is used. Informal trade is estimated by the information redundancy procedure, parting from the fact that in commercial exchanges between two countries, there are two customs declarations made, one by the source country and another by the destination country. The discrepancies between what has been declared by an exporter and what has been declared by an importer allow to obtain a proxy of informal trade. The downside of the previous method is that it does not account for the exchange of goods of services that happens through unconventional methods, not being declared or passed through customs. In our approach, predicted trade flows that are significantly lower than than the observed trade flows, could arguably provide some evidence towards the existence of informal trade between two countries (not transacting through customs). The results obtained in this chapter enable the identification of such estimation errors, which could potentially shed some light as to which countries are transacting informally.

4.3 Approach

The approach to be taken is partially based on the DNN architecture proposed by Hou et al. [97], where DNNs are used for link weight prediction in other tangential domains and applications such as airports, scientific collaborations, United States congress committees, and social networks like forums and others.

Two experiments are to be run: the first one aims at using a DNN with the goal of reducing the RMSE reported by the default model shown in Table 1 [23]; the second one has the same goal as the previous, but applied to the OLS (1) model reported in Table 2 [24].

The variable to be predicted in each one of the DNNs is the magnitude of the trade relationship between the countries, and the features or independent variables follow the specifications of each one of the models as outlined in Rose [23] and Head et al. [24] respectively.

The model contains the following fully connected layers:

- An input layer, with an input shape of i, where i corresponds to the quantity of features (this size is mandatory for the functioning of the model).

- Two hidden layers with layer size of 19 of exponential linear units (ELUs), which tend to converge cost to zero faster and produce more accurate results.

- An output layer with a layer size of 1 (this size is mandatory for the functioning of the model) and a linear regression unit, which is usually suggested for the last layer in applications that don't involve classification.

The layer size is directly proportional to the number of observations in the dataset. Larger datasets require a model that is more discriminative. Empirical work usually sets the layer size using 4.2, where n is the layer size and d is the width of the layers [97].

$$d = log_2(n) \tag{4.2}$$

The decision on the number of layers is also directly proportional to the complexity of the relationship between the inputs and output. Empirical work usually sets the layer size to 4, which is generally considered a good balance regarding the tradeoff between learning speed and prediction accuracy [97]. Other layer sizes were attempted but the results didn't change significantly, hence we stick to 4 layers.

Backpropagation is used, which performs propagation of the errors from the output later back to each one of the earlier layers [98]. A minibatch of size 32 is used, and the decision was made through model tuning. The optimal minibatch size of 32 obtained for our dataset is in line with the generic baseline recommendation in the literature [99], [100].

The Adam optimizer is used, which is an algorithm for first-order gradient-based optimization of stochastic objective functions that is based on adaptive estimates of lower order moments. Adam has been widely used in the literature pertaining machine learning due to its computational efficiency, little memory requirements, as well hyper-parameters that need little to no tuning [101].

The data frames are randomly shuffled before training commences. Then, it is split into 3 subsets:

- 70% trining

- 15% validation

- 15% test

Given the nature of the dependent variable, where it's not categorical (it's not a classification problem), performing the training, validation, and test splits can result in subsets where the mean of the dependent variable varies considerably among the subsets. The function to perform the train, test, and validation split from Scikit-Learn (python library) shuffles the dataset randomly before performing the split, which can result in the differences in the mean of the subsets mentioned previously. To mitigate this issue, a loop of 100 initialization seeds are tested to identify the one that minimizes the difference between the average mean of the generalization sets and the training set. Once the seed that minimized this difference was identified, it was fixed for the train, test, and validation split of the dataset.

RMSE is used as a prediction accuracy metric. The main reason for choosing RMSE over any other metric is because it enables comparison of performance to the one obtained by other authors. The main fallback of RMSE is that it's scale sensitive, whereas other performance metrics like mean average percentage error (MAPE) aren't scale sensitive, which allows for comparison even across dependent variables with different scales. However, when using MAPE it is common for it to colossally inflate when calculating the MAPE in observations where the actual value is a decimal very close to zero, given that the denominator will be miniscule, thus inflating the MAPE. The previous can be mitigated by eliminating observations where the dependent variable is close to zero. However, justifying such a decision is not simple given the amount of information loss that could result from such a procedure. Hence, in line with what has been used by other authors, RMSE is used.

Before tuning hyperparameters, it should be acknowledged that there are numerous sources of randomness when running a DNN that complicate the reproducibility of results and compromise the validity of hyperparameter tuning if these sources of randomness are not controlled beforehand. Among these sources one finds: the initialization of weights and biases (which is done by default following a probabilistic distribution), the train, test, and validation split from Scikit-Learn (which has a shuffle parameter which is defaulted to True), and the fitting of the model with Keras (which has a shuffle parameter defaulted to True). These sources of randomness can be solved by fixing the initialization seed of the probabilistic distribution used for the initialization of weights and biases, and disabling the automatic shuffling by changing the shuffling parameters to "False", respectively. After the previous is performed, one can get reproducible results, which then allow to attribute the variations in model performance to the actual tuning of hyperparameters, instead of the randomness in the procedures previously outlined.

When training large models, it is common a phenomenon for the training error to decrease steadily as more epochs are computed, but the validation set error can start to increase after a determined number of epochs. A model with better generalization (lower validation and test set error) can be obtained by running the model just for the number of epochs that minimize the validation set error. The strategy previously described is known as early stopping and was used to determine the number of epochs that the model had to be run in order for the model to better generalize [102, p. 246-251]. The total number of epochs for the DNN based on Rose was 983, and for the one based on Head, 300.

Due to the size of the datasets and hardware limitations, grid search was not feasible. Hence, manual tuning of the model was necessary. When doing manual tuning, one can reduce the computational intensity over grid search by only using the combinations that make sense. For example, if reducing the batch size from 32 to 16 yields significantly worse results, then one immediately decides to not try all of the possible combinations of the rest of hyperparameters and a batch size of 16, which depending on the number of hyperparameters and complexity of the model can save a significant amount of time.

On the other hand, grid search is automated and would still go ahead and try all of the combinations of batch size 16 and the rest of the hyperparameters, which in this case would be an extra 18,432 combinations.

The following hyper-parameter values were attempted before determining the final values to be used (in bold the optimal parameter based on validation loss minimization):

- Batch Size: 16, **32**, 48, 64, 128, 256, 512, 1024
- Width of Layers: 10, 14, **19**, 22
- Number of Layers: **4**,5,6
- Optimizer: **Adam**, Adadelta, SGD, Adamax, RMSProp, Adagrad,Nadam
- Adam Learning Rate: **0.001 (default)**, 0.01, 0.05, 0.1
- Dropout: **0**, 0.1, 0.05, 0.1
- L1 Regularization: **0**, 0.01, 0.05, 0.1
- L2 Regularization: **0**, 0.01, 0.05, 0.1

Notice that for activation functions, numerous combinations between layers were attempted using exponential linear units (ELU), rectified linear units (RELU), scaled exponential linear units (SELU), softplus, softsign, tanh, sigmoid, hard sigmoid, exponential, and linear. The best performing structure was using ELU for all of the layers. Note that no regularization was needed given the training error virtually converging before allowing for the model to overfit on the validation set. Table 4.1 shows the variables included in each model, as well as the number of observations included after cleaning the data according to the specifications of each one of the corresponding papers.

After detecting the number of epochs where each model's training error converges and knowing that it's good at generalizing, the entire database is used for training with the optimal number of epochs detected, as well as the hyper-parameters found to be adequate. The training loss obtained through this method is to be used as the final performance metric of the model.

4.4 Data

Two datasets are used. The first one is compiled by Rose [23] and is available in http://faculty.haas.berkeley.edu/arose/. It contains 234,597 trade flows of 177 conuntries from 1948 to 1999. The second one is compiled by Head et al. [24] and is available in http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8. It contains 592,923 trade flows for 238 countries from 1948 to 2006. For both, each observation is formed by the magnitude of the trade transaction between two countries, one being the importer and the other one being the exporter, together with characteristics of each one of the countries including quantitative variables, their dyadic binary relationships, and categorical variables (that can take 3 values). Both models are shown in table 4.1. Note that the dependent variable is the one to be predicted (also known as output in machine learning) and the independent variables are known as features in machine learning.

In table 4.1, log stands for natural logarithm. Distance stands for the distance in kilometers between the center of the 2 countries that are trading. Land border is a binary variable that determines whether the two interacting countries share a land border. GATT stands for General Agreement on Tariffs and Trade, and WTO stands for World Trade Organization. Regional FTA is a binary variable that indicates whether the interacting countries have a regional free trade agreement in place. Colonial relationship is a binary variable that shows whether two interacting countries have had a colonial relationship throughout history. GDP P/c stands for gross domestic product per capita. Log population origin is the natural logarithm of the population of the exporting country. Log population destination is the natural logarithm of the population of the importing country. GSP

Table 4.1 Structure of the Models

	Model Based on Rose [25]	Model Based on Head et al. [26]
Dependent Variable	Log Real Trade*	Log Real Trade*
Independent Variables	Log Distance*	Log Distance*
	Land Border**	Land Border**
	Both Countries in GATT/WTO**	Both Countries in GATT/WTO**
	Regional FTA**	Regional FTA**
	Common Language**	Common Language**
	Same Currency**	Same Currency**
	Colonial Relationship**	Colonial Relationship**
	Currently Colonized**	Currently Colonized**
	Log GDP*	Log Population Origin*
	Log GDP P/C*	Log Population Destination*
	Log Product Land Area*	Log GDP/Population Origin*
	One Country in GATT/WTO**	Log GDP/Population Desintation*
	GSP Membership**	ACP**
	Common Colonizer**	Same Legal System**
	Common Country**	
	Number Landlocked***	
	Number Islands***	
Observations	234,597	592,923

*For quantitative variables | ** for dyadic binary variables | *** for categorical variables, log is natural logarithm stands for generalized system of preferences, which provides tariff reduction for least developed countries. Number landlocked determines whether 0, 1 or 2 of the interacting countries are landlocked or not. Number islands determines whether 0, 1, or 2 of the interacting countries are islands. ACP stands for Africa, Caribbean and Pacific countries.

4.5 Results

Recall that two models were run. The first one followed the specifications used in the default model shown in Table 1 of Rose [25]. The second one, the ones used in the OLS (1) model reported in Table 2 of Head et al. [26]. The goal is to use deep learning to reduce the RMSE reported by these papers (increase the accuracy of their models), as well as the one obtained when trying to replicate their original experiments with the corresponding specifications. Note that when attempting to replicate their experiments, higher RMSEs than the ones reported in the original papers were obtained for both of them (nevertheless very close to the ones reported by the authors), even when following their specifications and using their datasets. The reason for the previous is unknown, it could have been because of differences in the software used for processing, or for exact details in the data cleaning process that were performed that weren't reported on the papers.

Early stopping and is used to determine the number of epochs that the model had to be run in order for the model to better generalize [102, p. 246-251]. Figure 4.1 and 4.2 show the early stopping process for both DNNs performed according to Rose's specification and Head's specifications respectively. For the former, the lowest validation error was obtained on epoch 983; for the latter it was 300. One could argue that the validation errors could have been further improved if run for more epochs. However, the decrease in loss was marginal relative to the computation time required for the decrease, hence it was decided to stop the training at 1,000 and 300 epochs respectively. Recall that 70% of the data was used for the training set, 15% for the validation set, and 15% for the test set.

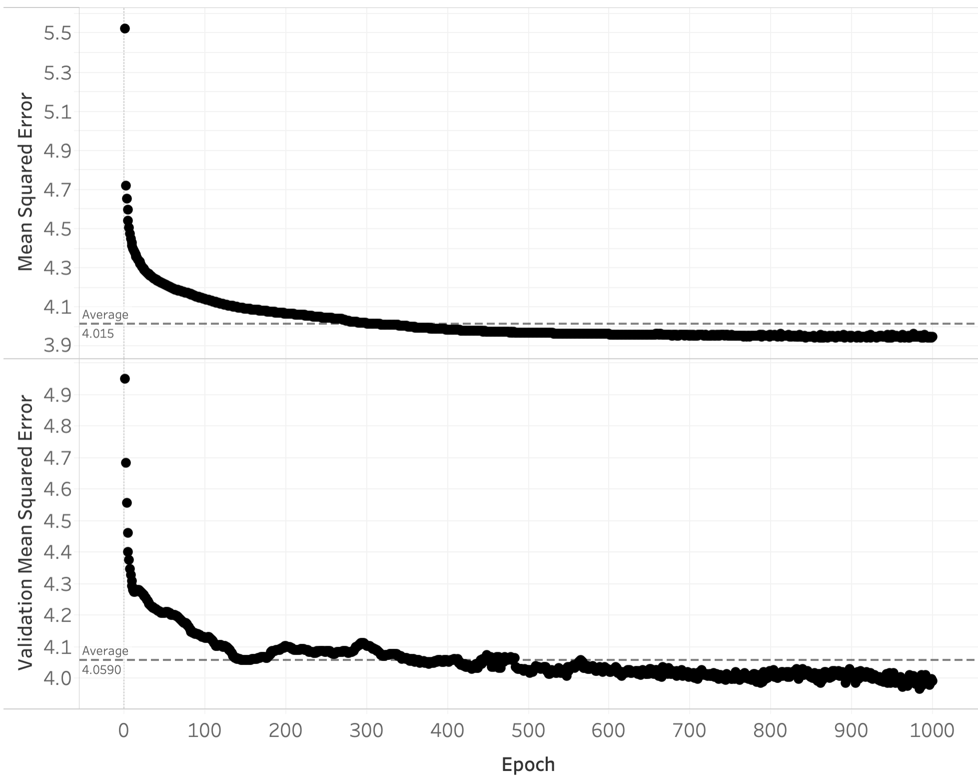


Figure 4.1 Rose Training and Validation Errors

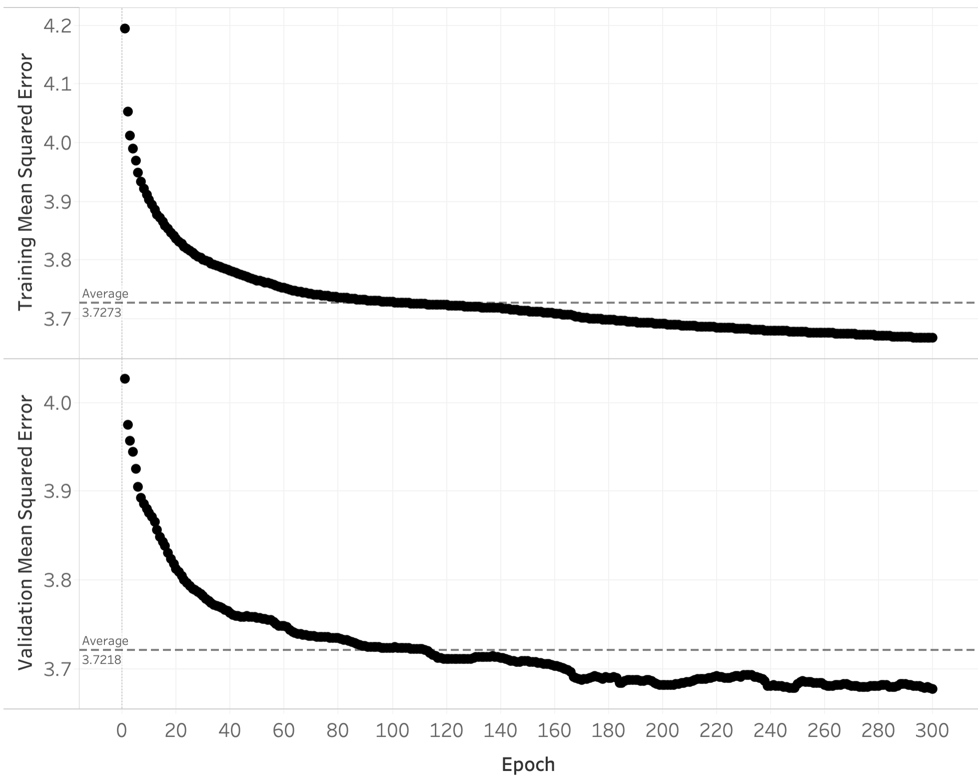


Figure 4.2 Head Training and Validation Errors

Following the running of both models, the test set error was reviewed in order to verify that generalization is adequate, and the error obtained was in between the losses of the training set and validation set, which is usually ideal and further strengthens the conviction that the model generalizes well. Note that using K-fold validation was considered and attempted but wasn't feasible due to computational limitations. Once the number of optimal epochs is determined, the model is retrained using all the data (no train, validation, test split) for the respective optimal number of epochs. The previously described workflow followed to determine the number of epochs to train for and how to split and train the data comes from Algorithm 7.2 [102, p. 273].

Figures 4.3 and 4.4 show the behavior of the errors when performing the original regressions as specified by Rose and Head respectively. The takeaways from figure 3 include: average predictions of larger trade flows tend to be more accurate; the average residuals for smaller trade flows tend to be overestimated and the larger ones underestimated; the absolute percentage error (APE) is not an adequate metric to keep track of when using logarithms, given that when the observed trade flow is closer to zero, the APE will tend to infinity, given that it's in the denominator of the computation of the APE, hence it skyrockets near zero values. From figure 4 the takeaways are similar: average predictions of larger trade flows tend to be more accurate, but as the trade flows get extremely large, the accuracy variates greatly; smaller trade flows tend to be underestimated and larger ones overestimated, with notably less variation than Rose's regression; average APE is once again verified as inadequate when using logarithms.

Figures 4.5 and 4.6 are homologous to 4.3 and 4.4, but correspond to the DNNs run based on each one of the respective models after using all of the data as training for the corresponding optimal number of epochs. In figure 4.5, it can be observed that the behavior of the 3 measures of error are very similar to the ones from figure 4.3, but when looking closely at the graphs one can note that the volatility is significantly lower with the

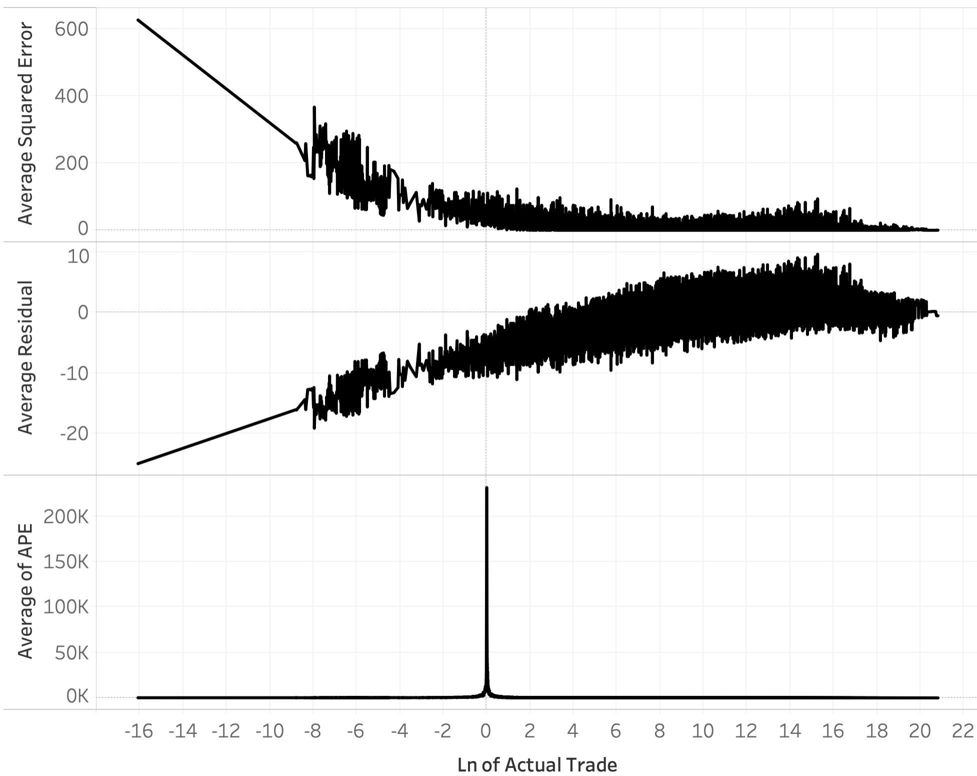


Figure 4.3 Rose Regression Errors Behavior

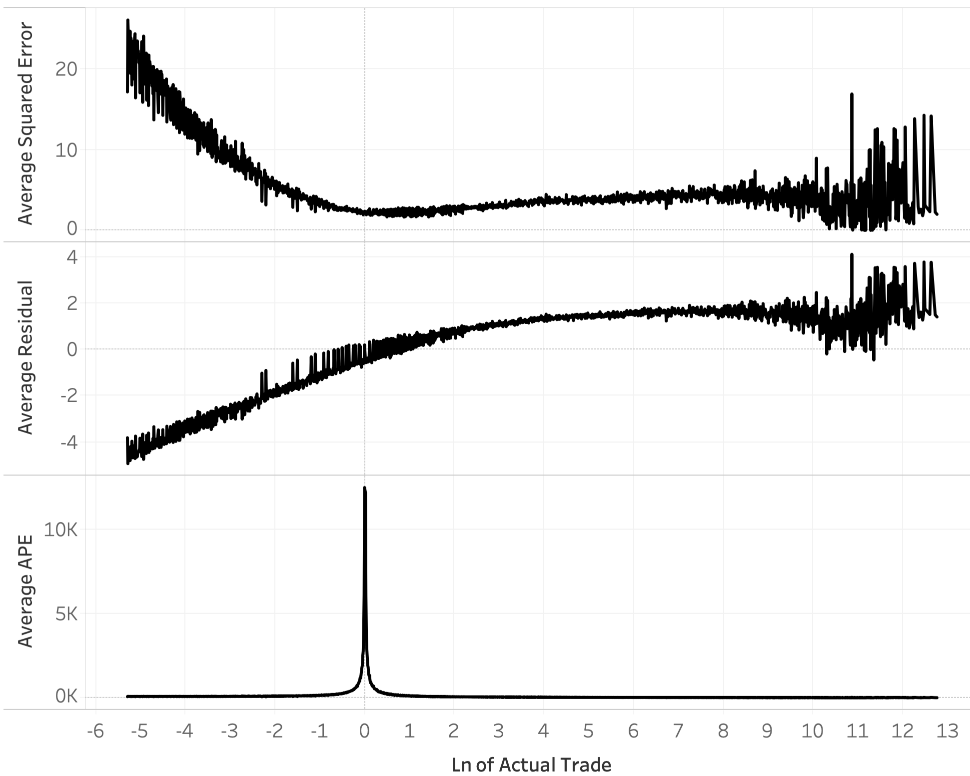


Figure 4.4 Head Regression Errors Behavior

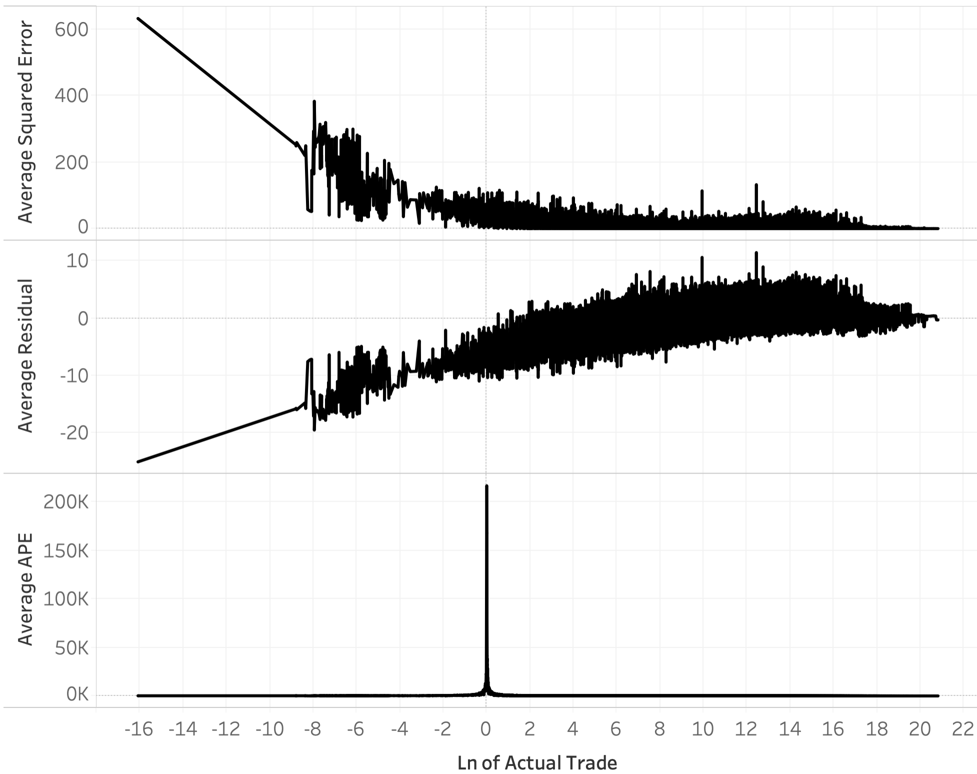


Figure 4.5 Rose DNN Errors Behavior

DNN than with the regression. To verify the previous statement, the standard deviation of the standard errors was computed for both the regression based on Rose [23] and the DNN based on the previous, and the regression is 16% more volatile in its squared errors. The same comparison was done between the regression based on Head et al. [24] and its corresponding DNN, and it was found that the regression is 4.5% more volatile in its squared errors. Volatility was also reviewed for the residuals, and it was found that Rose's regression is 8.23% more volatile than its corresponding DNN, and for Head it's 8.7%. Overall, the errors of the DNNs are considerably more stable than those of the multiple regression models.

Table 4.2 shows the performance comparison of the models run. Recall that there are 3 different results for each one of the two base models: the ones reported in the original papers, the ones obtained with our attempt to replicate the original paper's experiment

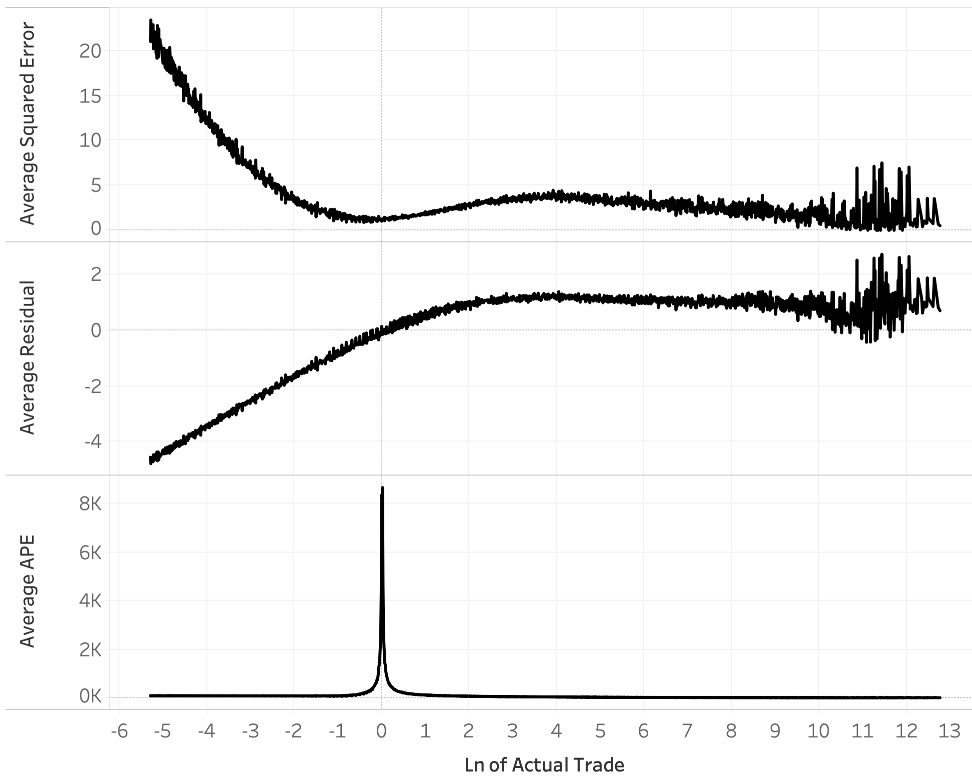


Figure 4.6 Head DNN Errors Behavior

Table 4.2 Performance Comparison

	Model Based on Rose [25]	Model Based on Head et al. [26]
Metric	RMSE	
Results Reported in Original Paper	1.98	1.88
Results Obtained When Attempting to Replicate Original Experiments	2.13	2.08
Results Obtained with Deep Learning	1.98	1.91

(where our RMSE was marginally higher than the one reported by the papers, even when following all of the corresponding specifications and using the original datasets), and the ones obtained with the DNNs. Note that the RMSE reported by Rose's paper is the same as the one obtained with deep learning. However, when comparing the RMSE obtained with our attempt to replicate the original experiment and the one obtained with the DNN, the RMSE was reduced by 7%. When looking at Head et al.'s RMSE reported in the original paper, it's lower than the one obtained with the DNN approach. However, when comparing the DNN approach with our attempt to replicate the original experiment's regression, the RMSE is reduced by 8.2%.

4.6 Reproducibility

In order to make the experiment reproducible, this section mentions the details of the implementation. The purpose is to allow other users to be able to not only replicate the experiment, but further modify the architecture outlined in this chapter and possibly obtain substantial improvements.

- Programming Language: Python 3.7.3
- Deep learning package: Tensorflow 1.13.1
- Operating system: Windows 10 Home 64-bit (10,0, Build 17763)
- Computer make and model: ASUSTek Computer Inc. TUF Gaming FX504GE_FX80GE
- Memory: 16,384 MB of RAM
- Processor: Intel Core i7-8750H CPU @ 2.20 GHz (12 CPUs)
- Threads used: 10
- Numpy version: 1.16.4
- Sklearn version: 0.21.2
- Pandas version: 0.24.2

The dataset as well as the code for implementation is publicly available on GitHub at: https://github.com/malg95/Link-Weight- Prediction-WTW/tree/master

4.7 Conclusions and Discussion

The use of deep neural networks (DNNs) for link weight prediction in the world trade web showed an improvement in performance of as much as 8.2% over multiple regression, while also reducing the standard deviation of the standard errors by as much as 13.8%. However, it should be taken into account that when using nonlinear activation functions in DNNs, the interpretation of the parameters is lost. Therefore, using DNNs is more accurate in link weight prediction, but it comes at the cost of the loss of interpretation of the parameters. If one wishes to identify the impact that each one of the variables has on the trade between countries, multiple linear regression should be used. In domains where the cost of the loss or error is extremely high (say for example, cancer detection), using DNNs would prove to be superior given the increase in accuracy and reduction in loss, even when it is just a moderate 8.2% in this specific use case, it could make a vast difference in humanity's wellbeing in certain domains.

The improvement in accuracy can results in better GDP forecasts, which can aid countries to better tune their public policy in regards to tariffs, quotas, and subsidies. This can mitigate potential trade hindering that could originate from external shocks like changes in the size of main trade partners, changes in regional or free trade agreements, and similar factors. Additionally, predicted trade flows that are significantly lower than than the observed trade flows, could arguably provide some evidence towards the existence of informal trade between two countries. The results obtained in this chapter enable the identification of such estimation errors, which could potentially shed some light as to which countries are transacting informally. The consequences of informal trade can be substantial, by reducing the tax collection by the state, thus reducing the tax base, generating unfair competition for enterprises, and endangering intermediate and final consumers with products that aren't inspected at a point of entry. Henceforth, it's in the countries' best interest to identify and mitigate informal trade.

4.8 Limitations

One of the main limitations of this chapter is that the impact of each individual independent variable on the dependent variable, as well as their statistical robustness, is lost when using ML instead of econometric approaches due to non-linearity and the black box effect in ML. Whether this tradeoff is worthwhile is subject to debate. In domains where the cost of loss or error is extremely high, say, for example, false negatives in cancer detection, the tradeoff might be worthwhile. Additionally, the predictive capabilities of this model for the future can be limited by the fact that numerous years are pulled together. This leaves an additional avenue of exploration for predictive capabilities of a cross sectional model instead of a time series based model.

Chapter 5

Conclusion

The World Trade Web (WTW) contains a wealth of information, that upon rigorous analysis can provide insights related to what network characteristics are associated with more prosperity, as well as how crisis can propagate along the network. Literature that attempts to describe the topology of the WTW is abundant, but insights on commercial actions countries could take to improve their well-being are scarce. Using network analysis to study the WTW has proven to be insightful in numerous occasions, such as the analysis of globalization and regionalization in international trade [3]; understanding the potential and risks of economic systems [4]; empirically derive the structure of the world economy [5]; understand global interdependencies [6]; better understand the role of network characteristics in countries' incomes [7, 8]. The main contribution of this dissertation is that in chapter 3, based on Fagiolo, Reyes, and Schiavo's [7] suggestion, we first provide an in-depth analysis of the topological characteristics of individual countries and regions from the cross-sectional perspective, as well as analyzing the role of geographical proximity in shaping the WTW to determine how fragile the network is. Knowing the fragility of the network is relevant to better understand the spread of financial crises, supply chain perturbations, among other trade and economic phenomena. Furthermore, in chapter 4 we improve the prediction accuracy of the trade links in the WTW relative to works that use econometric based approaches like Rose [23] and Head [24]. This contributes to the importance of machine learning in the field of economics, and could be critical when extrapolated to uses where the cost of error could be fatal. Predictions that are significantly lower than actual trade flows could provide initial evidence of informal trade flows, where it's in countries' best interest to mitigate these flows. Additionally, better accuracy in the prediction of flows can improve GDP forecasts, and in turn aid in public policy decision-making regarding tariffs, quotas, and subsidies.

The bulk of the thesis consists of two research chapters. Chapter 3 delves into the topology of the WTW and then finds relationships between the network characteristics of countries and their income. We find numerous empirical insights. Countries that are geographically closer tend to form more and stronger relationships, which could be due to lower transportation costs and trade agreements that tend to happen between countries that are geographically closer and could provide evidence that trade agreements increase the intensity of trade among communities. The continent most susceptible to instability originating from its trade partners is North America, and the most central continent is Europe. The most central countries in the WTW are USA, China, Germany, Great Britain, and the Netherlands. The countries that have the highest dependency on other countries are Mexico, Canada, Kuwait, Mongolia, and Andorra. The exploratory econometric analysis that uses countries' network characteristics as an input suggests that there is evidence that countries with higher PCGDP tend to associate with more neighbors that are themselves weaker, reciprocate fewer of their trade links, and trade more strongly with countries that are themselves stronger, and have a higher export to GDP Ratio. This could signal to what actions some countries could take from the public policy perspective in order to achieve a higher income.

Chapter 4 builds on the gravity model of trade and uses a deep neural network with the purpose of link weight prediction in the world trade web, that is, predicting the magnitude of the trade interactions between countries. The inputs for prediction are characteristics of the interacting countries such as their gross domestic product and land area, as well as their bilateral relationship traits, which include variables like distance between them and dyadic binary variables such as whether they are in the same continent, share borders, language, ethnicity, trade agreements, legal system, among others. The results of using deep learning are favorable, improving the performance obtained by traditional methods like ordinary least squares regression by as much as 8%. This contributes to the use of ML to better understand global trade. Our results can arguably be used as evidence of the existence of informal trade between countries, taking into consideration the predicted trade flows that are significantly higher than the observed flows. These errors are obtainable with our methodology, and they have the potential of shedding some light in pointing regulators to where the informal flows are sourcing and directing to. This is relevant, because informal trade can bring with it undesirable consequences like lower tax collection by the government, a smaller base to collect taxes from, unfair competition for businesses, and the absence of safety inspections of products for intermediate and final consumers. Mitigating informal trade is in the countries' and citizen's best interest. The improvement in trade forecasts can also have a chain effect on the accuracy of GDP forecasts, which can aid in public policy decision-making, particularly for tariffs, trade, and subsidies.

The key contribution of this thesis is towards the area of the value in the use of network analysis in the field of economics. This thesis shows how network characteristics of countries can be insightful and actionable insights on commercial actions to take to improve the countries' incomes. Main players in the network are identified not only at the country level, but also at the continental level, which could aid in better understanding how crisis and pandemics can propagate along the network. Another achievement is that we improve the accuracy of traditional approaches on link weight prediction on the WTW, showing that DNNs can prove useful in this context and aid in the potential identification of informal trade routes, as well as the optimization of tariffs, quotas, and subsidies.

5.1 Discussion

Someone with the power of doing public policy in their respective country should take the results of research like the one provided in chapter 3 with extreme caution, given that policies and systems that work in one country might not work in others due to numerous factors. Among these factors that make policy results vary, one can find differentials in rule of law, safety, climate, culture, economy, fiscal policy, monetary policy, commercial policy, and so forth. One should really delve into the specific situation of every specific country with a panel of experts in order to evaluate whether a certain policy could potentially benefit their population. For example, protecting a rising industry from foreign competition in a country might prove useful if said industry shows signs of increased productivity, quality, and competitiveness at the international level through time. Once this industry has the opportunity to catch up to international standards, then the protection can be waived and it should be able to compete with other international competitors. While this policy might work in some cases like the one previously described, it might not work in other countries where protection is provided to an industry, but it does not show any signs of improved productivity and quality. If a government continues protecting an industry like this, it could lead to scarcity of products from said industry at the national level, low quality, and higher costs, which could have a ripple effect on social costs and development.

An interesting remark is that the results stemming from the Herfindahl-Hirschman Index analysis show that the countries that have the highest dependency on other countries are Mexico, Canada, Kuwait, Mongolia, and Andorra. Something worth noting is that 2 of these 5 countries, Mexico and Canada, are members of the North American Free Trade Agreement (NAFTA) and have a border with the most influential player in the WTW, the USA. Based on the results of the dependency analysis, it could be good for their economies to trade more strongly with the USA. However, from the risk management perspective, this results in lower diversification, which makes their economies more dependent on the USA, hence more prone to suffering from economic shocks originating from that country. This suggests that there is a trade-off between how much a country can benefit from trading with a strong country, and the increase in risk to exogenous shocks originating from having a high percent of their trade relying on one particular partner.

In the realm of grouping similar objects, multivariate exercises such as cluster analysis that are performed using numerous indicators prove to be challenging because they can sometimes group countries that are notoriously different into a particular cluster. Additionally, naming the clusters is subjective and dependent on the researcher, and proves to be a daunting task when countries that are themselves heterogeneous are grouped within the same cluster. This results in countries like Norway, Iceland, and New Zealand being included in a cluster labeled "Low Income Moderately Connected Countries", for example. However, this is not uncommon in multivariate exercises such as regression analysis, for example, where there could be some particular outliers and observations where the estimation error is considerably higher than the average observation.

Lastly, in my attempt to predict trade I use a feed forward deep neural network. This is just one particular architecture within deep learning that generally allows to improve prediction accuracy over other econometric methods like linear regression. However, there are numerous other deep learning architectures that are more complex that could allow for better prediction accuracy, but come at the cost of a higher computing power and more taxing hyperparameter tuning. Among these architectures, one finds convolutional neural networks, that are usually common for machine learning involving the processing of images. Such architectures are yet to be explored in this context to see if they result in more accurate and less volatile predictions.

5.2 Limitations and Future Work

One of the main limitations of the work done in chapter 3 is that the data comes from official, reported trade interactions between countries. It is well known that there is vast informal trade between countries, which is not accounted for in these numbers. The real topology of the WTW when incorporating these unaccounted trade flows could be significantly different from the one obtained using the data we are able to access. However, at the time of writing and to the best of the author's knowledge, the database used is the most comprehensive public data base. It is important to note that external factors such as COVID-19 can be sources of exogenous shocks to the WTW stemming from unstudied forces and hence impact the WTW's stability, dynamics, insights, which have the power to potentially weaken the conclusions obtained from this chapter. This adds to the relevance of constantly updating works like this one, in order to incorporate new phenomena that develop over time and improve on the integrity and validity of the conclusions. Better understanding trade networks can empower countries to build more resilient supply chains, as has been recently exposed by the COVID-19 pandemic. Additionally, the weighting mechanism used is subject to discussion depending on the research goals. Other researchers have used varying weighing mechanisms in other papers depending on their goals, and there can be an extensive discussion as to the advantages and disadvantages of each weighting mechanism. Moreover, PCGDP shouldn't be the ultimate goals of countries when it comes to their economic performance. Other metrics should be taken into account to get a comprehensive status of their economy, such as metrics related to health, quality of life, and education.

There is plenty of future work to be done stemming from the findings of this chapter, such as: determining how exactly does reciprocating less trade links associate with a higher PCGDP? Is it because of exporting to more countries and not importing from them or the other way around? Do technology transfers to less developed countries explain why associating more strongly with stronger partners is associated with a higher PCGDP? How could the connectivity and exports of Africa be improved, given that Africa is a big net importer? Which countries have decreased their dependency on other countries through time?

For the work presented in chapter 4, we were able to improve the prediction accuracy relative to other state of the art econometric methods. However, one of the main limitations of our approach is that we lose interpretation and direction of the impact of the distinct variables on the magnitude of the trade links. This is an advantage that econometric methods have over DNNs, and is hence a tradeoff. The questions remains on whether it's possible to improve these predictions using more complex DNNs or CNNs that would require a higher computational power than the one we had at our disposal. Having more computational resources could enable the user to use L1 and L2 regularization methods despite their slow down of the learning process of the algorithm, hence being able to do more training epochs and obtaining a better accuracy. Additionally, we have performed an analysis that pools numerous years in the analysis, and different results could be obtained if the researcher controls for the years and decides to perform an analysis from the cross sectional perspective and where the object of study is either one country in particular, or the entire WTW.

Bibliography

[1] World Bank. "Trade (% of GDP". In: https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS?most_recent_value_desc=true. World Bank, 2019.

[2] Paul Krugman, Richard N Cooper, and TN Srinivasan. "Growing world trade: causes and consequences". In: *Brookings papers on economic activity* 1995.1 (1995), pp. 327–377.

[3] Sangmoon Kim and Eui Hang Shin. "A Longitudinal Analysis of Globalization and Regionalization in International Trade: A Social Network Approach". In: *Social Forces* 81 (Dec. 2002), pp. 445–471. DOI: 10.1353/sof.2003.0014.

[4] Frank Schweitzer et al. "Economic Networks: What do we know and what do we need to know?" In: *Advances in Complex Systems* 12.04n05 (2009), pp. 407–422.

[5] Roger J Nemeth and David Smith. "International Trade and World-System Structure: A Multiple Network Analysis". In: *Review (Fernand Braudel Center)* 8 (Apr. 1985), pp. 517–560.

[6] M Serrano, Marián Boguñá, and Alessandro Vespignani. "Patterns of Dominant Flows in the World Trade Web". In: *Journal of Economic Interaction and Coordination* 2 (Nov. 2007), pp. 111–124. DOI: 10.1007/s11403-007-0026-y.

[7] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. "The evolution of the world trade web: a weighted-network analysis". In: *Journal of Evolutionary Economics* 20.4 (2010), pp. 479–514.

[8] Javier Reyes and Raja Kali. “The Architecture of Globalization: A Network Approach to International Economic Integration”. In: *Journal of International Business Studies* 38 (May 2007), pp. 595–620. DOI: 10.1057/palgrave.jibs.8400286.

[9] Matthew O. Jackson. *Social and Economic Networks*. Princeton, NJ, USA: Princeton University Press, 2008. ISBN: 0691134405, 9780691134406.

[10] Mark Newman. *Networks*. Oxford university press, 2018.

[11] A. Barrat et al. “The architecture of complex weighted networks”. In: *Proceedings of the National Academy of Sciences* 101.11 (2004), pp. 3747–3752. ISSN: 0027-8424. DOI: 10.1073/pnas.0400087101. eprint: https://www.pnas.org/content/101/11/3747.full.pdf. URL: https://www.pnas.org/content/101/11/3747.

[12] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. “Basic notions for the analysis of large two-mode networks”. In: *Social networks* 30.1 (2008), pp. 31–48.

[13] Tore Opsahl. “Clustering in two-mode networks”. In: *Conference and Workshop on Two-Mode Social Analysis (Sept 30-Oct 2, 2009)*. 2009.

[14] Giorgio Fagiolo. “Clustering in complex directed networks”. In: *Physical Review E* 76.2 (2007), p. 026107.

[15] Marc Barthelemy et al. “Characterization and Modeling of Complex Weighted Networks”. In: *Physica A* 346 (Feb. 2005), pp. 34–43. DOI: 10.1016/j.physa.2004.08.047.

[16] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. “Weighted Evolving Networks: Coupling Topology and Weight Dynamics”. In: *Physical review letters* 92 (July 2004), p. 228701. DOI: 10.1103/PhysRevLett.92.228701.

[17] Luca Dall’Asta et al. “Vulnerabilty of weighted networks”. In: *J Stat Mech* P04006 (Apr. 2006). DOI: 10.1088/1742-5468/2006/04/P04006.

[18] Jari Saramäki et al. "Generalizations of the clustering coefficient to weighted complex networks". In: *Physical Review E* 75 (Mar. 2007), p. 027105. DOI: 10.1103/PhysRevE.75.027105.

[19] Jukka-Pekka Onnela et al. "Intensity and coherence of motifs in weighted complex networks". In: *Physical Review E* 71.6 (2005), p. 065103.

[20] Andrea De Montis et al. "The Structure of Interurban Traffic: A Weighted Network Analysis". In: *Environment and Planning B: Planning and Design* 34.5 (2007), pp. 905–924. DOI: 10.1068/b32128. eprint: https://doi.org/10.1068/b32128. URL: https://doi.org/10.1068/b32128.

[21] United nations Comtrade. *World Integrated Trade Solution*. Retrieved from https://wits.worldbank.org/countryprofile-dataavailability.aspx?lang=en.

[22] Walter Isard. "Location theory and trade theory: short-run analysis". In: *The Quarterly Journal of Economics* (1954), pp. 305–320.

[23] Andrew K Rose. "Do we really know that the WTO increases trade?" In: *American Economic Review* 94.1 (2004), pp. 98–114.

[24] Keith Head, Thierry Mayer, and John Ries. "The erosion of colonial trade linkages after independence". In: *Journal of international Economics* 81.1 (2010), pp. 1–14.

[25] M.E. J. Newman. "A Measure of Betweenness Centrality based on Random Walks". In: *Social Networks* 27 (Oct. 2003), pp. 39–54. DOI: 10.1016/j.socnet.2004.11.009.

[26] E Fisher and Fernando Vega-Redondo. "The linchpins of a modern economy". In: *AEA Annual Meeting, Chicago, IL*. Citeseer. 2006.

[27] Karen Stephenson and Marvin Zelen. "Rethinking centrality: Methods and examples". In: *Social networks* 11.1 (1989), pp. 1–37.

[28] Ulrik Brandes and Daniel Fleischer. “Centrality measures based on current flow”. In: *Annual symposium on theoretical aspects of computer science*. Springer. 2005, pp. 533–544.

[29] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. “On the topological properties of the world trade web: A weighted network analysis”. In: *Physica A: Statistical Mechanics and its Applications* 387.15 (2008), pp. 3868–3873.

[30] Giorgio Fagiolo, Tiziano Squartini, and Diego Garlaschelli. “Null models of economic networks: the case of the world trade web”. In: *Journal of Economic Interaction and Coordination* 8.1 (2013), pp. 75–107.

[31] Arthur L Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229.

[32] Irfan Danish. “Introduction to Neural Networks and Their Key elements”. In: Towards Data Science, 2019.

[33] Oliver Durr, Beate Sick, and Murina Elvis. “Probabilistic Deep Learning with Python”. In: Manning, 2019.

[34] Irfan Danish. “Introduction to Neural Networks and Their Key Elements (Part-B) — Hyper-Parameters”. In: Towards Data Science, 2019.

[35] Irfan Danish. “Introduction to Neural Networks and Their Key Elements (Part-C) — Activation Functions Layers”. In: Towards Data Science, 2019.

[36] Piotr Skalski. “Preventing Deep Neural Network from Overfitting”. In: Towards Data Science, 2018.

[37] Jaehyuk Park et al. “Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters”. In: *Nature communications* 10.1 (2019), p. 3449.

[38] Kristin Forbes. “Are Trade Linkages Important Determinants of Country Vulnerability to Crises?” In: *NBER Working Paper* 8194 (Apr. 2001). DOI: 10.7208/chicago/9780226185057.003.0003.

[39] Kristin Forbes and Tilak Abeysinghe. “Trade Linkages and Output-Multiplier Effects: A Structural VAR”. In: (July 2002).

[40] Reuven Glick and Andrew K Rose. “Contagion and trade: why are currency crises regional?” In: *Journal of international Money and Finance* 18.4 (1999), pp. 603–617.

[41] Raja Kali and Javier Reyes. “Financial contagion on the international trade network”. In: *Economic Inquiry* 48.4 (2010), pp. 1072–1101.

[42] Maurice Obstfeld and Paul R Krugman. *International economics: theory and policy*. Addison-Wesley, 2003.

[43] Matthew O Jackson. *Social and economic networks*. Princeton university press, 2010.

[44] David Snyder and Edward Kick. “Structural Position in the World System and Economic Growth, 1955-1970: A Multiple-Network Analysis of Transnational Interactions”. In: *American Journal of Sociology - AMER J SOCIOL* 84 (Mar. 1979). DOI: 10.1086/226902.

[45] Marc Tiri, Ludo Peeters, and Gilbert Swinnen. “Knowledge, innovation and competitiveness: Dynamics of firms, networks, regions and institutions”. In: (Feb. 2015).

[46] Diego Garlaschelli and Maria Loffredo. “Fitness-Dependent Topological Properties of the World Trade Web”. In: *Physical review letters* 93 (Nov. 2004), p. 188701. DOI: 10.1103/PhysRevLett.93.188701.

[47] Ronald Breiger. “Structures of economic interdependence among nations”. In: *Continuities in structural inquiry* (1981), pp. 353–380.

[48] Kunal Bhattacharya et al. “The international trade network: weighted network analysis and modelling”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.02 (2008), P02002.

[49] K Bhattacharya, G Mukherjee, and SS Manna. “The international trade network”. In: *Econophysics of markets and business networks*. Springer, 2007, pp. 139–147.

[50] Giorgio Fagiolo. “Directed or undirected? A new index to check for directionality of relations in socio-economic networks”. In: *arXiv preprint physics/0612017* (2006).

[51] Ann Harrison. “Openness and growth: A time-series, cross-country analysis for developing countries”. In: *Journal of development Economics* 48.2 (1996), pp. 419–447.

[52] Halit Yanikkaya. “Trade openness and economic growth: a cross-country empirical investigation”. In: *Journal of Development economics* 72.1 (2003), pp. 57–89.

[53] Douglas A Irwin and Marko Terviö. “Does trade raise income?: Evidence from the twentieth century”. In: *Journal of International Economics* 58.1 (2002), pp. 1–18.

[54] Vivek Arora and Athanasios Vamvakidis. “How much do trading partners matter for economic growth?” In: *IMF Staff papers* 52.1 (2005), pp. 24–40.

[55] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. “The Evolution of the World Trade Web: A Weighted-network Analysis”. In: *Journal of Evolutionary Economics* 20 (Aug. 2010), pp. 479–514. DOI: 10.1007/s00191-009-0160-x.

[56] Kristian Skrede Gleditsch. “Expanded trade and GDP data”. In: *Journal of Conflict Resolution* 46.5 (2002), pp. 712–724.

[57] Diego Garlaschelli and Maria I Loffredo. “Structure and evolution of the world trade network”. In: *Physica A: Statistical Mechanics and its Applications* 355.1 (2005), pp. 138–144.

[58] William Chow. “An anatomy of the world trade network”. In: *Consultado en* (2013).

[59] Organization for Economic Cooperation and Development. Retrieved from https://stats.oecd.org/Index.aspx?QueryId=75506.

[60] Diego Garlaschelli and Maria I Loffredo. “Patterns of link reciprocity in directed networks”. In: *Physical review letters* 93.26 (2004), p. 268701.

[61] Vincent D Blondel et al. “Fast unfolding of communities in large networks”. In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[62] Renaud Lambiotte, J-C Delvenne, and Mauricio Barahona. “Laplacian dynamics and multiscale modular structure in networks”. In: *arXiv preprint arXiv:0812.1770* (2008).

[63] Phillip Bonacich. “Power and centrality: A family of measures”. In: *American journal of sociology* 92.5 (1987), pp. 1170–1182.

[64] Mark S Mizruchi and G William Domhoff. *The American corporate network, 1904-1974*. Sage Publications Beverly Hills, 1982.

[65] Beth A Mintz and Michael Schwartz. *The power structure of American business*. University of Chicago Press, 1985.

[66] Lawrence Page et al. *The PageRank citation ranking: Bringing order to the web.* Tech. rep. Stanford InfoLab, 1999.

[67] Edsger W Dijkstra. “A note on two problems in connexion with graphs”. In: *Numerische mathematik* 1.1 (1959), pp. 269–271.

[68] Linton Freeman. “The development of social network analysis”. In: *A Study in the Sociology of Science* 1 (2004).

[69] Henry F Kaiser and John Rice. “Little jiffy, mark IV”. In: *Educational and psychological measurement* 34.1 (1974), pp. 111–117.

[70] Joseph F. hair. In: *The Oxford Handbook of Innovation.* Pearson Prentice Hall, 2006, pp. 533–544.

[71] Mark EJ Newman. “Mixing patterns in networks”. In: *Physical Review E* 67.2 (2003), p. 026126.

[72] Jacob G Foster et al. “Edge direction and the structure of networks”. In: *Proceedings of the National Academy of Sciences* 107.24 (2010), pp. 10815–10820.

[73] AO Hirschman. “The paternity of an Index, ın American Economic Review”. In: (1964).

[74] Orris Clemens Herfindahl. *Copper costs and prices: 1870-1957*. Baltimore, Published, P, 1959. Published for Resources for the Future by . . ., 1959.

[75] J.F. Hair. “Cluster Analysis”. In: *Multivariate Data Analysis*. Always learning. Prentice Hall, 2010. Chap. 8, pp. 415–474. ISBN: 9780138132637. URL: https://books.google.ca/books?id=JlRaAAAAYAAJ.

[76] Glenn W Milligan. “An examination of the effect of six types of error perturbation on fifteen clustering algorithms”. In: *Psychometrika* 45.3 (1980), pp. 325–342.

[77] Paul R Krugman. “Increasing returns, monopolistic competition, and international trade”. In: *Journal of international Economics* 9.4 (1979), pp. 469–479.

[78] Raja Kali and Javier Reyes. “The architecture of globalization: a network approach to international economic integration”. In: *Journal of International Business Studies* 38.4 (2007), pp. 595–620.

[79] M Ángeles Serrano, Marián Boguñá, and Alessandro Vespignani. “Patterns of dominant flows in the world trade web”. In: *Journal of Economic Interaction and Coordination* 2.2 (2007), pp. 111–124.

[80] Kristin J Forbes. “Are trade linkages important determinants of country vulnerability to crises?” In: *Preventing currency crises in emerging markets*. University of Chicago Press, 2002, pp. 77–132.

[81] Tilak Abeysinghe and Kristin Forbes. "Trade linkages and output-multiplier effects: A structural VAR approach with a focus on Asia". In: *Review of International Economics* 13.2 (2005), pp. 356–375.

[82] Jan J Tinbergen. "Shaping the world economy; suggestions for an international economic policy". In: (1962).

[83] Stefan Nickel and Justo Puerto. *Location theory: a unified approach*. Springer Science & Business Media, 2006.

[84] Paul R Krugman. "On the relationship between trade theory and location theory". In: *Review of international economics* 1.2 (1993), pp. 110–122.

[85] Thomas Chaney. "The gravity equation in international trade: An explanation". In: *Journal of Political Economy* 126.1 (2018), pp. 150–177.

[86] Paul Krugman. "Scale economies, product differentiation, and the pattern of trade". In: *The American Economic Review* 70.5 (1980), pp. 950–959.

[87] Anne-Célia Disdier and Keith Head. "The puzzling persistence of the distance effect on bilateral trade". In: *The Review of Economics and statistics* 90.1 (2008), pp. 37–48.

[88] James E Anderson and Eric Van Wincoop. "Gravity with gravitas: a solution to the border puzzle". In: *American economic review* 93.1 (2003), pp. 170–192.

[89] Pentti Pöyhönen. "A tentative model for the volume of trade between countries". In: *Weltwirtschaftliches Archiv* (1963), pp. 93–100.

[90] Eli Filip Heckscher. *The effect of foreign trade on the distribution of income*. 1919.

[91] Bertil Ohlin. *Interregional and international trade*. Harvard University Press, Cambridge, 1935.

[92] Paul A Samuelson. "Prices of factors and good in general equilibrium". In: *The Review of Economic Studies* 21.1 (1953), pp. 1–20.

[93] Paul A Samuelson. "International factor-price equalisation once again". In: *The economic journal* 59.234 (1949), pp. 181–197.

[94] Staffan Burenstam Linder. *An essay on trade and transformation*. Almqvist & Wiksell Stockholm, 1961.

[95] Marco Duenas and Giorgio Fagiolo. "Modeling the international-trade network: a gravity approach". In: *Journal of Economic Interaction and Coordination* 8.1 (2013), pp. 155–178.

[96] Mehdi Abid. "Estimating the Size of the Informal Trade Across the World: Evidence from a MIMIC Approach". In: *Journal of the Knowledge Economy* (2019), pp. 1–52.

[97] Yuchen Hou and Lawrence B Holder. "Deep learning approach to link weight prediction". In: *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2017, pp. 1855–1862.

[98] D Ruhmelhart, G Hinton, and R Wiliams. "Learning representations by back-propagation errors". In: *Nature* 323 (1986), pp. 533–536.

[99] Yoshua Bengio. "Practical recommendations for gradient-based training of deep architectures". In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 437–478.

[100] Dominic Masters and Carlo Luschi. "Revisiting small batch training for deep neural networks". In: *arXiv preprint arXiv:1804.07612* (2018).

[101] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[102] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.