MULTILAYER NETWORKS IN ECONOMY – A CASE STUDY ON LAND DEALS

UNIVERSITÁ DEGLI STUDI DI MILANO

FACOLTÁ DI SCIENZE POLITICHE, ECONOMICHE E SOCIALI

Laurea Magistrale in Data Science and Economics

Author: Nallapaneni Aditya Sai,
Matriculation Number: 933570

Supervisor: Prof. Gaito Sabrina Tiziana,
Dept. Of Computer Science,
Management and Quantitative Methods

Co-Supervisor: Prof. Silvia Salini,
Dept. Of Economics,
Management and Quantitative Methods

Academic Year 2019-2020

I declare that this dissertation entitled ***Multilayer Networks in an Economy: A case study on land deals*** *is a result of my research and any outside information used has been properly cited on the reference page. This dissertation has not been submitted for the candidature of any degree and has not been accepted for any degree whatsoever.*

Abstract

There are very complex systems around us. Understanding these complex systems is a major intellectual and scientific challenge. Network Science is a very promising approach to understand these complex systems. Networks help us to understand the complex relationships and processes in these systems. Thus, to understand the market of gold and copper and exchanges happening between countries we have used a network analysis approach. The first chapter of the dissertation comes up to analyze the gold and copper datasets which are taken from Land Matrix. Land Matrix is a public database consisting of countries related to land deals. We discuss results after doing analysis using python and a nice visualization in Gephi. In this present study, nodes are the name of all the countries which are part of the gold and copper market, and edges are the exchanges happening between them.

Most engineered and natural systems have sets of entities that interact with each other in complicated patterns which can involve a variety of different forms of relationships, as well as changes in time and other types of problems. Multiple subsystems and layers of connectedness are present in such systems, and it is critical to consider these ”multilayer” characteristics to improve our knowledge of complex systems. As a result, it is required to broaden ”conventional” network theory by establishing (and testing) a framework and related tools for comprehensively studying multilayer systems. The study of multilayer networks has become one of the most prominent directions in network science, with roots dating back several decades and originating in a variety of disciplines. In this work, the different models and metrics of multilayer networks are discussed. Also, this thesis covers how gold and copper are exchanging from countries.

Acknowledgement

I would like to thank my thesis supervisor Prof. Gaito Sabrina Tiziana and co-supervisor Prof. Silvia Salini for their guidance and support in finishing this study. I am thankful to my parents, friends who helped me in every sphere of life. Their contribution helped me to complete this study effectively.

Contents

List of Figures

Chapter 1

Introduction

1.1 Research background

A network is a set of relationships or in other words, a network consists of interacting components. A network is made up of two important components – the first is known as nodes which are simply all the individuals that are part of the network and the second are edges which is the relationship between the nodes. In this present study, nodes are the name of all the countries which are part of the gold and copper market, and edges are the exchange happening between them. We have used network properties to understand, analyze both networks. It helps us to understand the structure of the network.

Network Theory has been used over a long period to analyze and model complex physical, biological, and social systems. As more research is carried, data evolves and becomes more heterogeneous and complex, these networks become an oversimplification of the course being analyzed. The classical network representation used over the years gets less useful because of its heterogeneity and relations.

Therefore, a need to go beyond the traditional networks into a more advanced framework capable of hosting objects and their concerns arises, and Multilayer Network can do that.

Multilayer networks represent numerous types of complex systems that have multiple networks or have disparate interactions between entities. They are sets of nodes, edges, and layers where the model's implementation determines the layers' interpretation. The main difference between the types is the procedure or criteria of linking a node to a layer. In multilayer network graphs, nodes are partitioned into groups called layers, and a node in one layer can relate with a node in another layer through the edge.

Due to the ever-maturing research on complex systems and multilayer networks, there has been a need to move beyond just simple graphs to more complicated networks. Edges, for example, can be directed, have varying strengths (i.e., 'weights'), exist only between nodes belonging to different sets (e.g., bipartite networks), or be active only at certain periods. Recently, there has been a surging interest in investigating networks having numerous types of connections, sometimes known as "networks of networks." Such systems have been studied for decades in discipline. such as sociology and engineering, but the current explosion of efforts to establish frameworks for studying multilayer complex systems and to generalize a vast corpus of common tools from network research is a new phenomenon. In recent years, studying networks for several layers (or various types of edges) and networks of networks has become increasingly popular. Unfortunately, the sudden and massive influx of articles on multilayer networks has resulted in an equal influx of diverse nomenclature, and the lack of a consensus (or even widely acknowledged) language and mathematical framework for researching multilayer networks is particularly troublesome.

Furthermore, work on generalizing complex network concepts like degree, transitivity, centrality, and diffusion are still in their early stages. Discussed in this paper is the general definition of multilayer networks which can be used to represent most types of complex systems that consist of multiple networks. We analyze the literature to establish a natural mapping from each type of network to this multilayer-network representation, and we classify the various multilayer-network representations based on the sorts of constraints they put on this multilayer-network representation.

1.1.1 Models of Multilayer Networks

Network Theory has been used over a long period to analyze and model complex physical, biological, and social systems. As more research is carried, data evolves and becomes more heterogeneous and complex, these networks become an oversimplification of the course being analyzed. The classical network representation used over the years gets less useful because of its heterogeneity and relations.

Therefore, a need to go beyond the traditional networks into a more advanced framework capable of hosting objects and their concerns arises, and Multilayer Network can do that. Multilayer networks represent numerous types of complex systems that have multiple networks or have disparate interactions between entities. They are sets of nodes, edges, and layers where the model's implementation determines the layers' interpretation. The main difference between the types is the procedure or criteria of linking a node to a layer. In multilayer network graphs, nodes are partitioned into groups called layers, and a node in one layer can relate with a node in another layer through the edge.

1.1.2 Exponential Random Graph Model (ERGMs)

Multilayer networks are used in the analysis of weighted networks (Pilny & Atouba, 2018). ERGMs are an essential and quite flexible model in Network statistics. Initially, it was designed for binary edge encoding. The presence of advantages between nodes. Under the model, each network layer represents a specific dyadic category (Pilny & Atouba, 2018). The model layers are assumed to be generated by the ERGM process conditional on the nearest lower layer. An advantage of the model is the ability to adopt binary network statistics specifications, describing both ”across layer networks” and ”between layer networks”. Some of the most used statistics include density summaries such as the count of edges, homophily such as the count of advantages amidst nodes with the same characters, degree-based statistics, and several triangular-based statistics. The social networks analysis develops when more than a single relation exists between a standard set of actors. ERGMs determine any bond between pairs of individuals with a bunch of variables such as network structures. In social networks analysis, ERGMs describe the connectivity network structure (Van Der Pol, 2019). ERGMs are quite flexible and can be incorporated in any network statistics and were initially defined for the binary-edged net.

1.1.3 Models of Change Propagation

The framework is made up of 3 different tiers that lead to change propagation. These layers include the product tier, change tier, and social layer, and the model takes a fundamental interest in the interaction within the product, change, and social layer. Each layer has its unique and directed network made up of nodes connected via intra-layer edges. The product tier represents the system or product undergoing design. Here, the nodes represent both the hardware, software components, and the documents associated with the process. The interlayer edges represent the network’s technical interfaces, either physical connections or channels for energy flow. The change layer of the grid represents the change propagation. The node of this layer represents individual changes; edges represent the relationship. Finally, the social layer represents the networks and the edges represent the various relationships existing among the people in the organization. Inter-layer edges link the product, social, and change layers to complete the multilayer network.

1.1.4 Multilayer Stochastic Model

The model is an aggregation of the M networks where each is a subgraph of a specific network G. Each layer is an outcome of probabilistic removal of the links and nodes from G. The final network includes any links that appear in at least K layers. Two sets of results are usually achieved in this model (Valles-Catala et al., 2016). for a given graph structure the format is the entire state of all links. The second result is the appropriate scaling of the node and link selection activity in a layer. The Stochastic model is an example of a non-standard site bond percolation model (Valles- Catala et al., 2016). The links in this model are usually independent as the layer numbers approach infinity. The model is mostly applicable in wireless communication networks with numerous channel radios, countless social networks with overlapping membership, and transportation networks.

1.2 Thesis Problem

The thesis problem is to find out the multilayer networks metrics using given datasets. The datasets in this case are the gold and copper networks, which are used in the analysis. It is mainly focused on the following steps. The factors affecting the network for the exchange of gold and copper. The metrics used to analyze the datasets of copper and gold. Usage Multilayer Network models. Visualization of networks in Gephi.

1.3 Aim

This thesis aims to establish the measures of multilayer networks using the provided datasets. This thesis project helps us to visualize meaningful data from the datasets which contain information about gold and copper imports of countries. In particular, the weak connections of the trade networks can be identified and improvised.

1.4 Thesis Objectives

To establish the metrics of multilayer networks using the provided datasets. To identify the weak connections in the networks after analyzing the gold and copper networks using the social network analysis method.

1.5 Thesis Significance

By depicting the relationships between nodes and studying the emergent structures and connections, the researcher will be able to unveil intricate and hidden structures in textual sources, as well as social interactions and groupings. Also, the research will enable us to understand the factors affecting the network for the exchange of gold and copper globally. Through this, appropriate measures could be taken to ensure the smooth exchange of gold and copper within the mentioned zones; this would ensure that low costs are incurred during the gold and copper exchanges. Through the social network analysis method, the researcher will be able to identify the weak connections in the networks after careful analysis of the gold and copper networks and subject theme to corrections to smoothen the exchange process. The analysis of multilayer networks in this project will help in the description of complex systems that could be used in addition to the Gephi tool, Python, and Network X. Also, through the research on multilayer networks, it is easy to study the features of the gold and copper networks and this could be important in finding insights into ways of improving the network. The research is vital in the improvement of the transport network between the countries facing challenges in the gold and copper business. Another significance of this research is that by the application of big data network analysis in mineralogy, it is possible to find missing minerals such as gold and copper. Complex data sets provide important insights into social media connections, urban traffic patterns, and metabolic pathways, to name a few. Network analysis provides a new perspective on minerals, just as complex data sets provide important insights into social media connections, urban traffic patterns, and metabolic pathways.

Chapter 2

Data Exploration

2.1 Libraries Used

The entire project is based on networking which has been used to analyze the given datasets in Python using NetworkX. Python is an Open-Source Initiative development that has an open-source license for developing software. Another software that is used in this study is the Gephi tool. Gephi is a graph and network analysis program that is free to use. It makes use of a 3D render engine to show massive networks in real-time and speed up the research process. A multi-task and flexible architecture opens up new ways to interact with large data volumes and deliver valuable visual results. In the context of interactive network exploration and interpretation, we show some fundamental aspects of Gephi. It allows spatializing, filtering, traversing, modifying, and clustering of network data and provides quick and broad access to network data. Finally, we highlight crucial components of dynamic network visualization by demonstrating Gephi's dynamic characteristics.

On the other hand, NetworkX is a graph in a network analysis library written in Python. This is a code example that demonstrates how we utilized NetworkX to compare subgraphs and alter the graph parameters. If you have never used NetworkX before, it'll show you how to convert a weighted edge-list to a NetworkX graph and what you can do if you want to investigate a specific node in an extremely complex network graph, especially with the weight's attribution.

2.2 Datasets

The datasets show how the countries are interconnected to each other. To obtain a clearer understanding of the market connection we dealt with the In-degree and Out-degree distribution of the nodes; this will help give us an understanding of the country with the most inflow and the one with the most outflow. Usually, the import and export of gold is being done in all countries Two datasets have been used for this study; Gold and copper. The dataset consists of countries, which provide information on metallic resources like gold and copper and how there are connected in the market. This dataset is taken from Land Matrix which is an open-access platform that gives detailed information about deals in almost 100 countries. The Country focuses on Europe Asia, Africa, the USA, and South Africa. The dataset is connected to two Columns in from or too. The dataset shows that the country connected one to another.

2.3 Land Matrix

Gathering information about land deals is a difficult process. The information is hard to find and even harder to confirm through various sources. To overcome this problem, the Land Matrix database was introduced (landmatrix.org), which stores huge amounts of data on land deals from all countries and provides a systematic overview of large-scale land investments. The data presented on the Land Matrix platform now covers almost 100 countries and is constantly being developed. The Land Matrix database covers deals with agricultural production, timber extraction, carbon trading, industry, renewable energy generation, conservation, and tourism in low- and middle-income countries. However, due to limited coverage of certain sectors and to reduce bias in the dataset, we have used only a portion of the entire database for our analysis, focusing on mineral types. Such as gold and copper.

The Land Matrix Initiative (LMI) was created to address the lack of reliable data on large-scale land acquisitions. It is an independent global land monitoring initiative consisting of several global and regional partners. The GIGA is one of the key partners in this initiative and hosts the Land Matrix Database. It also focuses on the dynamics and impacts of large-scale land acquisitions. Large-scale land acquisitions (LSLAs) are playing an increasingly important role and have far-reaching impacts, transforming rural landscapes, economies, and societies. For instance, National Land Observatories (NLOs) include domestic as well as transnational deals, whereas the Global Observatory only includes transnational deals.

In the thesis, we have focused particularly on the mineral resources available in the Land Matrix database. For example, gold and copper. We have looked at the import and export of resources from different countries.

2.4 Data Analysis

For analyzing both the Gold and Copper Networks, the social network analysis method was used. To analyses and visualize the networks, the Network X package was used in python. Gephi is a network visualization and analyses software.

To proceed with the network analyses and visualization, first, we need to prepare the dataset in a particular format. We prepared two separate files for Nodes and Edges. Node files contain two columns, one with the ID and the other one with the name of the country in the network. The Edges file contains two important columns – Source and Target/Destination. The source is the column with the country name from which any transaction took place and the target is the column with the name of the county to which the transaction was made. The third column is also made in the edge file which defines the weight for each relationship/transaction.

Once these files were prepared they were saved in CSV format. The next step is to import these files in Gephi and Network X.

Below, we can observe the sample entries of balanced data.

deal_gold_network

source	destination	weight
AUS	LAO	2
AUS	PHL	1
AUS	MLI	1
AUS	BRA	3
AUS	MDG	1
AUS	ARG	2
AUS	EGY	1
AUS	DOM	1
BRA	PER	1

deal_copper_network

source	destination	weight
AUS	LAO	2
AUS	PHL	1
AUS	PER	1
AUS	DOM	1
AUS	COD	3
AUS	ZMB	1
CHN	LAO	1
CHN	PER	2
CHN	ZMB	3

Figure 2.1: Gold and Copper Network Dataset

2.4.1 Gephi Visualization

Gephi is a tool for data analysts and data scientists who are keen to explore and understand graphs. for each graph data, the user interacts with the representations, manipulate the structures, shapes, and colors to reveal hidden patterns. The goal is to help data analysts to make a hypothesis, intuitively discover patterns, isolate structure singularities or faults during data sourcing. This is software for Exploratory Data Analysis.

The following steps are to visualize both the networks in Gephi one by one.

Open Gephi. Create a new project. Click on the Data Laboratory tab. Open a new workspace. Click on the Node tab and choose the option import spreadsheet. Select the CSV file with the node name and import. Make it a directed graph by choosing the Directed option at the end of the process of importing the spreadsheet. Now, Click finish. To import the edges, click on the edges tab and click import spreadsheet. Now, choose the edges CSV and import.

Now, Go the Graph layout section where the graph gets visualized there. To customize the visualization, the layout tool, and appearance tool, and filters were used. To run the different network metrics, we use a statistic toolbar. Once the visualization is customized, go to the preview section and refresh to get the current customized view of the network. The visualizations were saved using the save option on the bottom left corner of the screen of Gephi. Similar steps were followed for both the networks in Gephi.

To analyze the networks using the python Network X package are used. Network X is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Chapter 3

Gold Network Analysis

The gold network dataset which consists of Source, Destination, and weight is modified into two different datasets. Such as gold node and gold edge, shown in the below fig.5 This makes it easier to find the network interactions in between the countries and makes better visualization in the Gephi tool. The Gold node consists of Id, Label, and Category. Whereas gold edge is about Source, Target, and weights which are initially defined. Moreover, we find all the measures that show how the connections are made and the results of various centrality measures, degree distributions. There is a total of 61 nodes in the gold node network i.e., the number of countries, and the gold edge network is a total of 86 edges which determine the connection between the countries. The data is balanced and there are no missing values. We dealt with the connections between the countries using different metrics like Indegree distribution, Outdegree distribution, betweenness centrality, closeness centrality, and connected components.

Gold node

Id	Label	Category
0	AUS(Australia)	Gold
1	LAO(Laos)	Gold
2	PHL(Philippines)	Gold
3	MLI(Mali)	Gold
4	BRA(Brazil)	Gold
5	MDG(Madagascar)	Gold
6	ARG(Argentina)	Gold
7	EGY(Egypt)	Gold
8	DOM(Dominican)	Gold
9	PER(Peru)	Gold
10	UKR(Ukraine)	Gold

Gold edge

Source	Target	Weight
0	1	2
0	2	1
0	3	1
0	4	3
0	5	1
0	6	2
0	7	1
0	8	1
4	9	1
10	1	1

Figure 3.1: Sample of Gold Network dataset

- Basic Information of the network
- Number of nodes (no. of countries in the network) – 61
- Number of edges (Connections between countries) – 86

We can observe in the below graph Fig.6 all the connections concerning their weights connected with blue nodes and edges. In this, there are only a few countries that have less distribution among the gold network. The connection between Canada and Brazil is strong. For instance, Greece is observed as an outlier where it doesn't contain any connections between the countries. Moreover, We visualize the networks in Gephi which depicts the nodes and edges understandably and clearly. Below, The thick edges/ lines show the weight of the connection.

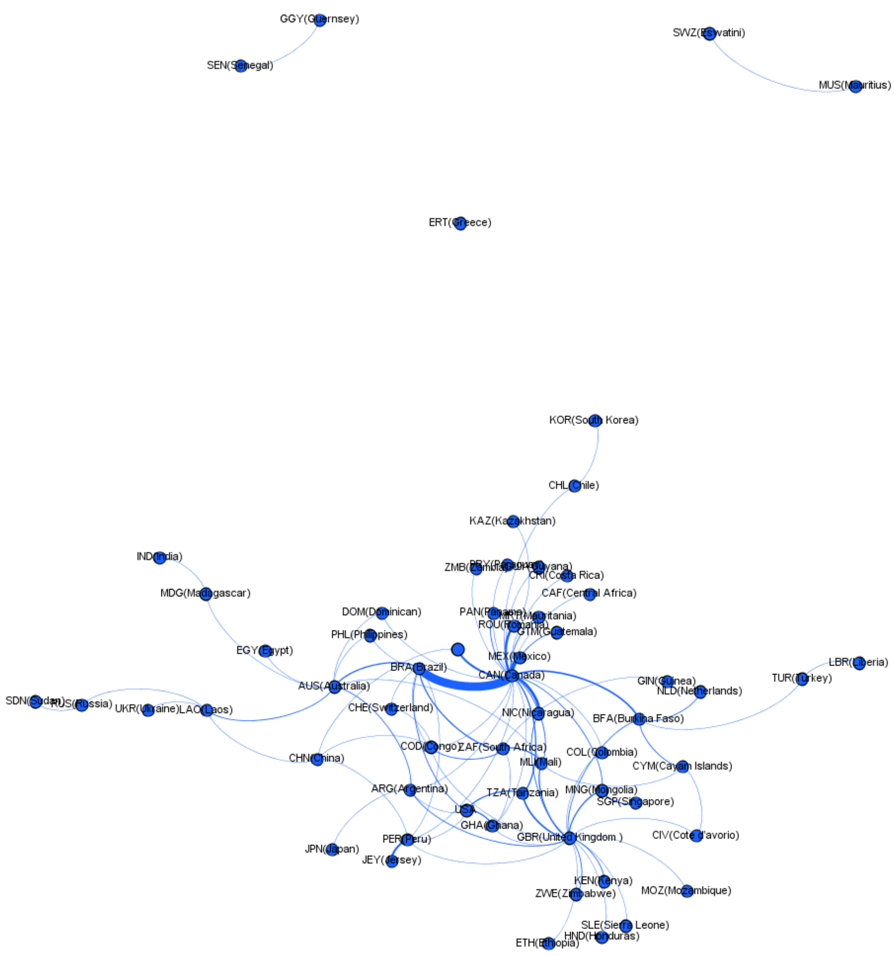

Figure 3.2: Visualization of the gold network

3.1 Degree Distribution

The first step in understanding a network is by looking at its node-level properties and understanding the role of each component. A node is characterized by its degree and the degree distribution is the frequency distribution of the number of links per node. To understand which country is a major player, this is understood by the degree distribution of the network. Moreover, the number of direct connections /edges of a node is known as the degree of that node. The degree distribution is the graphical representation of the whole dataset showcasing several nodes on the y axis and no. of connections on the x-axis. If the network size increases, the distribution becomes increasingly narrow. We usually find the degree distribution by Pdeg(k), defined by Pdeg(k)=fraction of nodes in the graph with degree k and degree distribution of a network P(k), tells us the probability that a randomly chosen node will have degree K. In our case we get to know that the average degree of the gold network is 1.41 and the average weighted degree is 2.89. We compute the degree centrality in python by using nx. degree(G), where G is the variable which is the collection of nodes in the network. In Addition, the node degree is the number of edges adjacent to the node. The weighted node degree is the sum of the edge weights for edges incident to that node. Below are the results depicting each countries degree distribution.

- Average Degree – 1.410

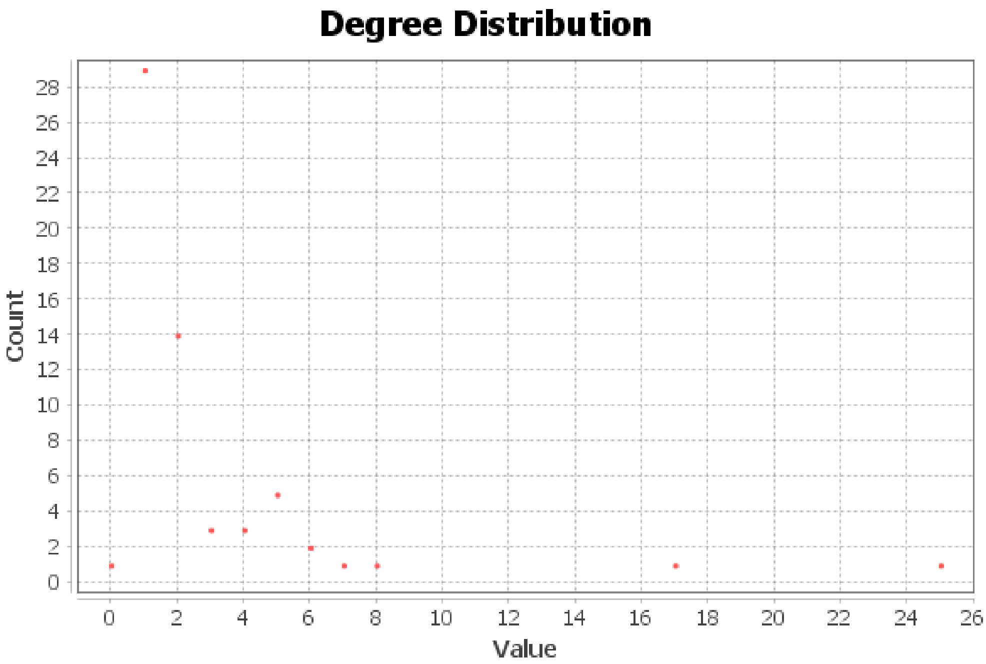

Figure 3.3: Degree Distribution

- Average Weighted Degree – 2.86

Results:
nx.degree(G)
Degree View: ({'AUS': 8, 'LAO': 4, 'PHL': 2, 'MLI': 5, 'BRA': 7, 'MDG': 2, 'ARG': 5, 'EGY': 1, 'DOM': 2, 'PER': 6, 'UKR': 1, 'RUS': 2, 'SDN': 1, 'CHN': 4, 'ERI': 2, 'COD': 6, 'CAN': 25, 'GUY': 1, 'NIC': 2, 'COL': 2, 'PRY': 1, 'MEX': 1, 'CRI': 1, 'PAN': 1, 'GTM': 1, 'CHL': 2, 'MNG': 3, 'ZMB': 1, 'TZA': 3, 'MRT': 1, 'ROU': 2, 'CAF': 1, 'BFA': 5, 'GHA': 4, 'HND': 1, 'GBR': 16, 'CIV': 2, 'ETH': 1, 'MOZ': 1, 'SLE': 1, 'KEN': 1, 'ZWE': 1, 'CHE': 2, 'ZAF': 5, 'GIN': 2, 'IND': 1, 'USA': 5, 'JPN': 1, 'JEY': 1, 'KOR': 1, 'SGP': 1,

'TUR': 2, 'LBR': 1, 'NLD': 2, 'KAZ': 1, 'CYM': 3, 'GGY': 1, 'SEN': 1, 'MUS': 1, 'SWZ': 1})

Furthermore, weights depict the total number of user-to-user links between two countries. These weighted degrees were binned into 0 to 10 intervals for the P(w) calculation. The degree distribution plot over here shows that most of the nodes have degrees below 9, with maximum nodes having degree value one. GBR (United Kingdom) and CAN (Canada) have the 16 and 25 highest degree values respectively. There is one node ERT(Greece) that has zero-degree distribution as it tends to have no connections. It shows that Canada is the highest number of connections followed by the United Kingdom and Australia.

To get a clearer understanding of the connection of the market we are going to analyze the In-degree and Out-degree distribution of the nodes which will help us in understanding which country has more inflow and which has more outflow.

3.1.1 In-Degree Distribution

The degree of a node in a network is the number of connections it has to other nodes and the degree distribution is the probability distribution of these degrees over the entire network. The degree of a vertex is the number of edges connecting to it. For vertex, the number of head ends adjacent to a vertex is called the indegree of the vertex. In-degree distribution our gold network is defined as the total number of incoming edges. In the following visualization below - BRA (Brazil) and PER (Peru) have the highest number of incoming edges followed by Congo, Argentina, Mali, and Burkina Faso. The larger the size of the node darker the color means that has a high In-degree. Although from the degree distribution it seemed like Canada has the highest connections but it has become clear that it has zero In-degree distribution meaning that there is no inflow in the gold market. Similarly, UK and Australia have low In-degree distribution. The major players in the inflow sector are Brazil and Peru followed by Congo, Argentina, Mali, and Burkina Faso. Below Figure.8 shows the gold network In-degree distribution.

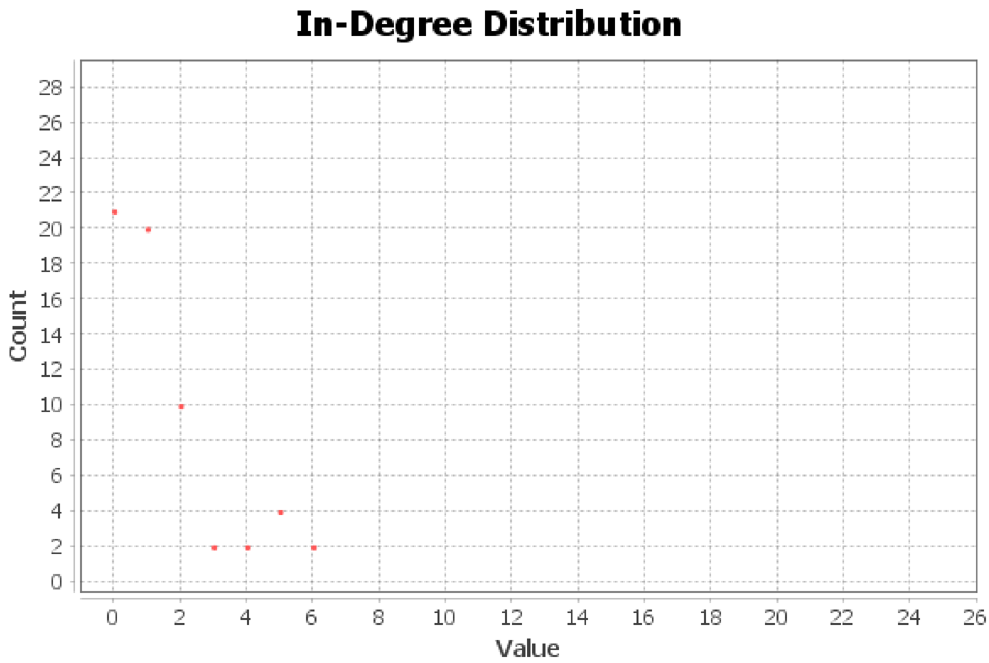

Figure 3.4: Graphical Visualization of In-Degree Distribution

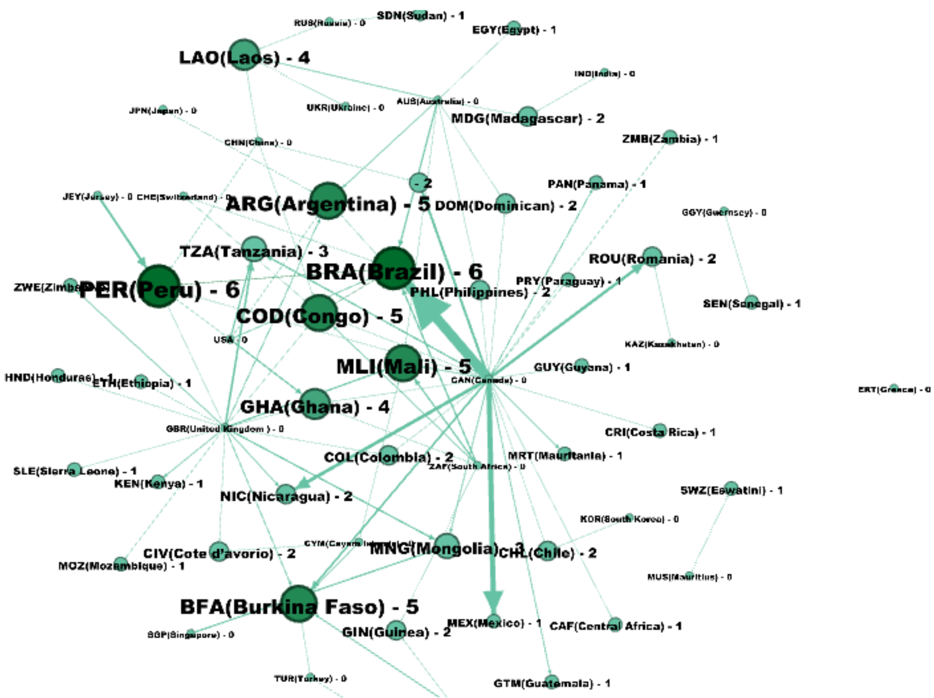

Figure 3.5: Plotting nodes for the In-Degree Distribution

3.1.2 Out-Degree Distribution

The Out-degree of a vertex V written by deg+(v), is the number of edges with v as an initial vertex. To find the out-degree of a vertex we count the number of edges from the vertex.

The following visualization and graph depict the Out-degree for Gold Network that is the number of out-going links of a node. As we said earlier that out-degree distribution refers to the number of arcs incidents. That is the number of arcs directed away from the vertex. It is clear from the visualization that Canada has the highest number of exports followed by the United Kingdoms. The remaining nodes have out-degree less than 9.

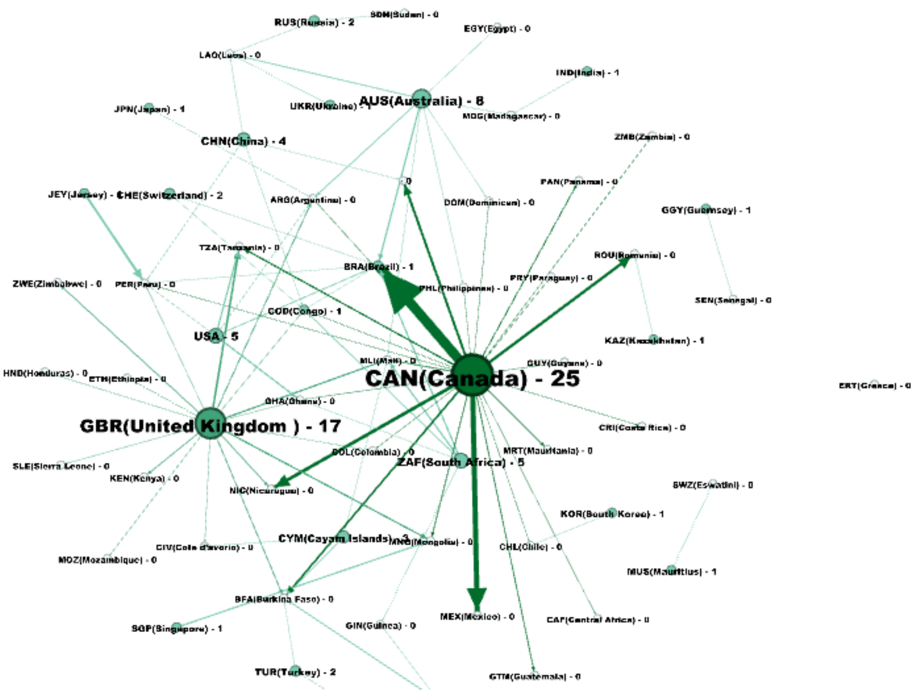

Figure 3.6: Graphical visualization of Out-degree node

3.2 Graph Density in Gold Network

Given the network size (number of nodes) and graph order, the density of a graph is a measure of how many ties exist between countries (number of links). It measures how close is the network to complete. Intuitively, the density of a network is the ratio of the number of links L to the number of possible links in a network with N nodes and it's given by 2L/N(N-1). A network property's density is important to consider for two reasons. The first benefit of density is that it can help us understand how connected the network is in comparison to how connected it could be. Second, when two networks with the same number of nodes and the same type of relationships are compared, it can tell us how the networks differ. In a gold network, the density is as you would expect, this will return a value between 0 and 1, with higher values indicative of a dense graph where most available connections have been completed. A complete graph has all possible edges and density equals 1. The Graph Density for the following gold network is

- Directed – 0.023
- Undirected – 0.047

The graph density is very low for both the directed and undirected networks. About 2 to 4 percent of nodes are connected, based on the output value of 0.023 and 0.047.

3.3 Connected Components in Gold Network

A connected component is a maximal subgraph of a network such that there exists at least one path from each member of that subgraph to each other member. A graph is connected if there exists a path between any pair of nodes in it. Connectivity represents how nodes are connected via a sequence of edges in a network. Two nodes are adjacent if they are connected via an edge. Two edges are incident if they share an endpoint. An edge in a graph can be traversed when one starts at one of its end-nodes, moves along the edge, and stops at its other end-node. A node vi is connected to node vj (or reachable from vj) if it is adjacent to it or there exists a path from vi to vj. A path is a sequence of links. The path length is the number of edges visited in the path and the graph is connected if there is a path between all pairs of nodes. When we talk about connected components we talk about node-generated subgraphs (subsets of vertices and all edges that are between them). The order of a network is the number of nodes while its size is the number of links. Moreover, a connected component or simply component of an undirected graph is a subgraph in which each pair of nodes is connected via a path. The main point here is reachability. In connected components, all the nodes are always reachable from each other. However, it seems that the import and export of Gold are working in almost all countries. We also discussed strongly and weakly connected components for the gold network which are important to find the nodes and edges with are highly connectivity. 4 weakly connected components mean few of the countries are not reachable to other countries in the network.

3.3.1 Strongly and weakly connected components

A strongly connected component (SCC) is the portion of a directed graph in which there is a path from each vertex to another vertex. It is a maximal subset of nodes that each node can reach from all the other nodes along with a directed graph. Strongly connected components usually have a size that is smaller than the weakly connected components. Moreover, we can find all strongly connected components in O(V+E) time. Here we observed in the gold network that the total number of strongly connected components is 57. For instance, in the movement of gold from one country Canada to Mexico there is strong connectivity but from Mexico, there are no other connections that tell us it's a directed graph.

Moreover, A graph is said to be weakly connected if there exists a path between any pairs of nodes, without following the edge directions. In the case of an undirected graph, a weakly connected component is also a strongly connected component. Hence, if a graph G doesn't contain a directed path (from u to v or from v to u for every pair of vertices u, v) then it is weakly connected. The elements of such a path matrix of this graph would be random. The algorithm used here is DFS(depth-first search) likewise, a DFS starting at a vertex v first visits v, then some neighbor w of v, then some neighbor x of w that has not been visited before, etc. the algorithm works in this way. In addition, we acquired 4 weakly connected components for the gold network. Where the Y-axis represents a count of the nodes and the X-axis represents the size of the nodes. In the Below figure.11, we can observe weakly connected components of the gold network.

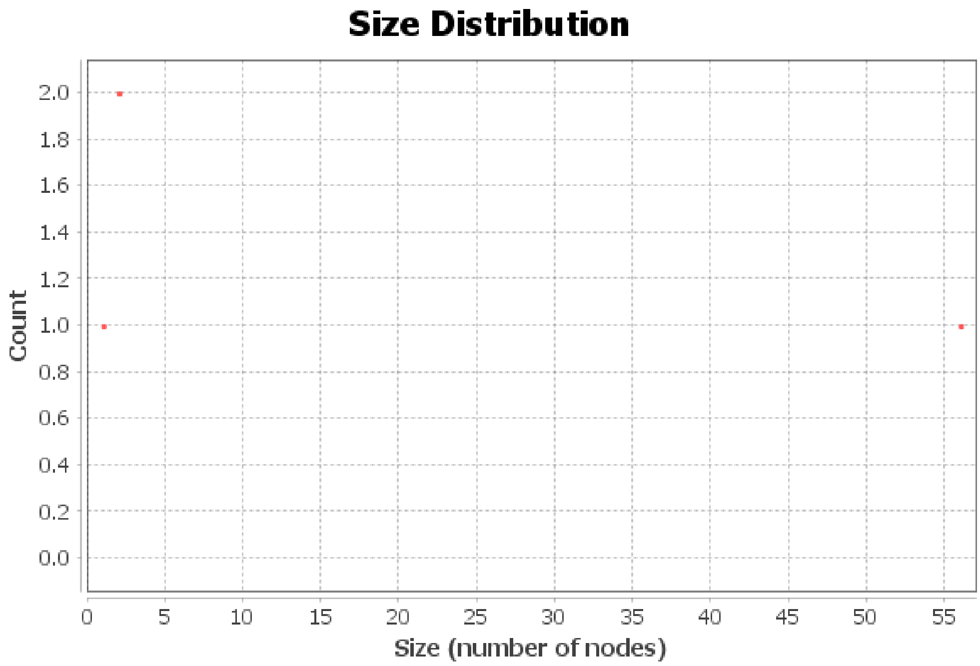

Figure 3.7: Weekly connected components of an undirected network

3.4 Centrality Measures in Gold Network

Centrality is a fundamental topic in network analysis. basically, we want to associate to each node a score that defines the importance of the node in the graph. We call geometric those measures assuming that importance is a function of distances. Geometric centrality depends only on how many nodes exist at every distance. In our case, considering weighted graphs, in which arcs have weights. At that point, one can equivalently consider a nonnegative square matrix M on the reels: the associated graph as a weighted arc for each nonzero entry. In this, we discuss various centrality measures such as, Betweenness centrality and Closeness centrality

3.4.1 Betweenness Centrality

The main aim is to measure the probability that a random shortest path passes through a given node. The intuition behind betweenness is that if a large fraction of shortest paths pass through x, then x is an important junction point of the network. We have computed the betweenness centrality results for each node in Figure12 below. The Figure represents the name nodes along with their Betweenness centrality score. These nodes are sorted according to the centrality score, The top or important node tends to be larger, and the low important node leads in a smaller size. Brazil (BRA) and Congo (COD) are the best connector" with 4 and 2 values.

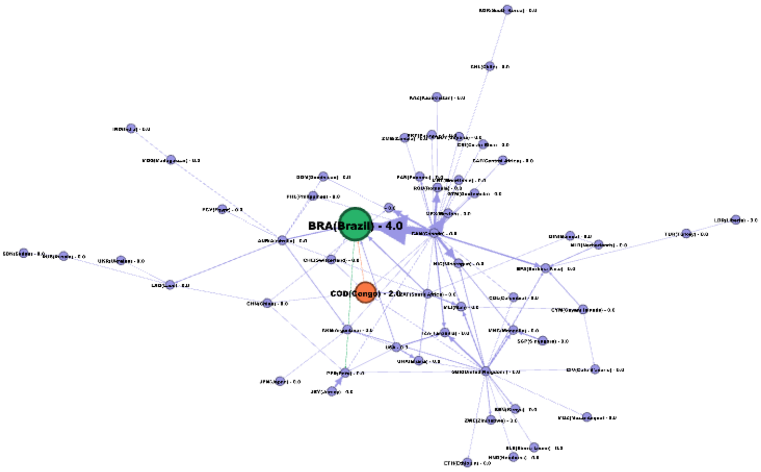

Figure 3.8: Betweenness Centrality of Gold network

According to the betweenness centrality, a node that acts as a bridge between the majority of the nodes turns itself into a more powerful node in the network. In the Figure above we can see those two nodes (i.e., Brazil and Congo) have a high number of neighbors, which means that they have played an important role in the gold network, Another interesting thing can be seen in all the remaining nodes (i.e., countries) have zero betweenness centrality score (i.e., they are not acting as a bridge). In the previous statement, it is appearing that there are only two countries which are more contributing to the import and export of Gold.

3.4.2 Closeness centrality

The intuition behind closeness is more central nodes have smaller distances, and thus have a smaller denominator, resulting in a larger centrality. Moreover, in a connected graph, the closeness centrality of a node is a measure of centrality in a network calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the closer it is to all other nodes.

In this section, we have presented the results of closeness centrality on the Gold network for each node. The Figure below represents the names of nodes along with their Closeness centrality score and the color of the node. The yellow color node in the graph represents a high score and the purple color node represents a low score. The score is normalized i.e., it can be maximum 1 and minimum 0. Where centrality score 1 means node is close to all other nodes in the network. According to closeness centrality, a node that is close to all other nodes in the network turns itself into a more powerful node in the network.

We observe that node Canada (CAN) has the highest centrality score with a value of 0.4929921773142112, followed by Mali (MLI), Peru(PER), Congo(COD) with centrality scores of 0.42025562656293414. which means that they have played a more important role in the import and export of Gold. It is appearing that Canada leads the market in the exchange of Gold, controlling the market followed by the above-listed countries. Below are the results from python analysis of the gold network.

{'AUS': 0.35853976531942633, 'LAO': 0.27864775239498896,
'PHL': 0.361064693244211, 'MLI': 0.42025562656293414,
'BRA': 0.41683891415185337, 'MDG': 0.262929161234246,
'ARG': 0.41347731000546745, 'EGY': 0.2602598296481115,
'DOM': 0.361064693244211, 'PER': 0.42025562656293414,
'UKR': 0.2154251531120923, 'RUS': 0.21725079000287273,
'SDN': 0.17679719462302745, 'CHN': 0.3265680665011335,
'ERI': 0.3464269354099862, 'COD': 0.42025562656293414,
'CAN': 0.4929921773142112, 'GUY': 0.32450118000429096,
'NIC': 0.3560499058380414, 'COL': 0.3560499058380414,
'PRY': 0.32450118000429096, 'MEX': 0.32450118000429096,
'CRI': 0.32450118000429096, 'PAN': 0.32450118000429096,
'GTM': 0.32450118000429096, 'CHL': 0.32866145154280746,
'MNG': 0.361064693244211, 'ZMB': 0.32450118000429096,
'TZA': 0.361064693244211, 'MRT': 0.32450118000429096,
'ROU': 0.32866145154280746, 'CAF': 0.32450118000429096,
'BFA': 0.38262079433341767, 'GHA': 0.3715303365266519,
'HND': 0.32450118000429096, 'GBR': 0.4037101294541572,
'CIV': 0.2864312091657987, 'ETH': 0.28326622342916,
'MOZ': 0.28326622342916, 'SLE': 0.28326622342916,
'KEN': 0.28326622342916, 'ZWE': 0.28326622342916,
'CHE': 0.3051856335754641, 'ZAF': 0.32450118000429096,
'GIN': 0.251329345297441, 'IND': 0.2059083792798312,
'USA': 0.3224602920797356, 'JPN': 0.28804037326223575,
'JEY': 0.2913135593220339, 'KOR': 0.24414850686037126,
'SGP': 0.26158768592182635, 'TUR': 0.275651540003645,
'LBR': 0.21362994350282485, 'NLD': 0.27864775239498896,
'KAZ': 0.24414850686037126, 'CYM': 0.3088625689197468,
'GGY': 0.01694915254237288, 'SEN': 0.01694915254237288,

'MUS': 0.01694915254237288, 'SWZ'

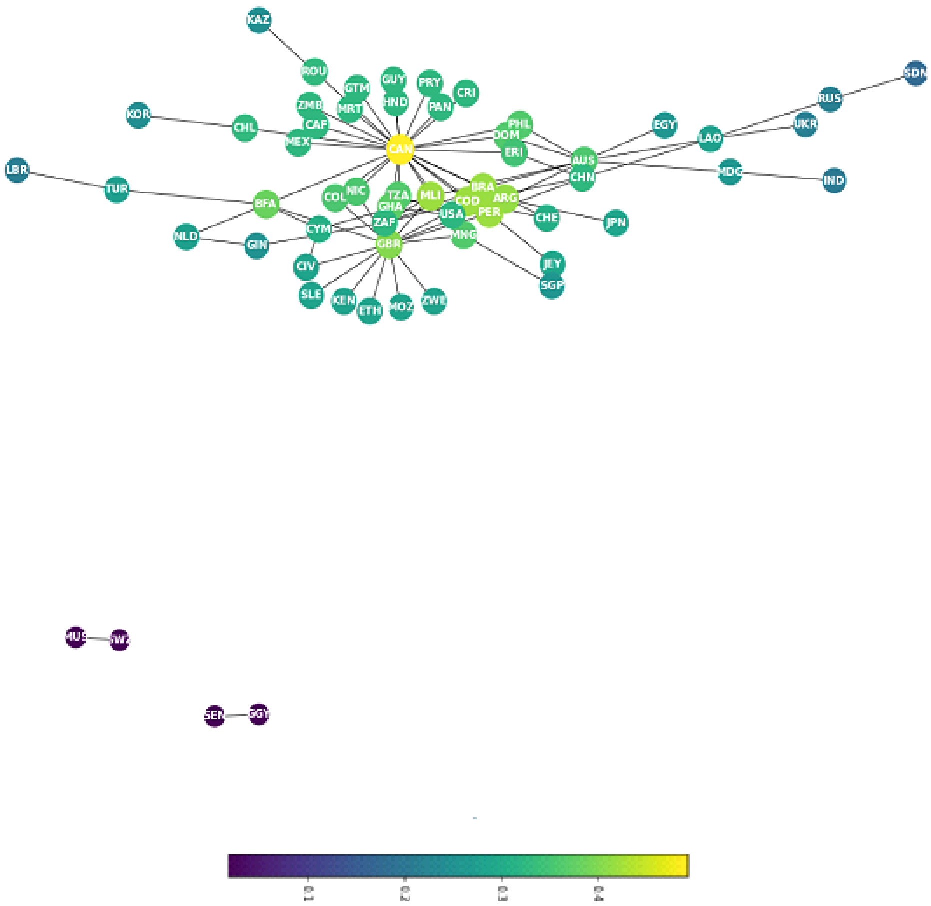

Figure 3.9: Closeness Centrality of Gold network

Chapter 4

Copper Network Analysis

The copper network dataset which consists of Source, Destination, and weight is modified into two different datasets. Such as copper node and copper edge, shown in the below figure.14. This data consists of a total of 53 entries which depicts the import and export of copper between the countries. To find the network interactions between the countries Gephi tool makes better visualization for the networks. Firstly, The Gold node consists of Id, Label, and Category, and the gold edge is about the Source, Target, and weights. Moreover, we find all the metrics that show how the connections are made and the results of various metrics. Second, there is a total of 37 nodes in the copper node network, and in the copper edge network, there are a total of 53 edges that determine the connection between the countries. The data is balanced and there are no missing values. The average degree is calculated by measuring the number of edges compared to the number of nodes. The resulting average degree for the copper network is value 1.42. We dealt with different metrics like Indegree distribution, Outdegree distribution, betweenness centrality, closeness centrality, and connected components. Below is the sample data of copper node and copper edge networks.

- Number of Nodes – 37
- Number of Edges – 53
- Average Degree – 1.432

copper node

Id	Label	Category
0	(AUS)Australia	Copper
1	(PER)Peru	Copper
2	(DOM)Dominican	Copper
3	(COD)Congo	Copper
4	(ZMB)Zambia	Copper
5	(LAO)Laos	Copper
6	(CHN)China	Copper
7	(MMR)Myanmar	Copper
8	(PHL)Philippines	Copper

Copper egde

Source	Target	Weight
0	5	2
0	8	1
0	1	1
0	2	1
0	3	3
0	4	1
6	5	1
6	1	2
6	4	3
6	3	4

Figure 4.1: Sample of copper network

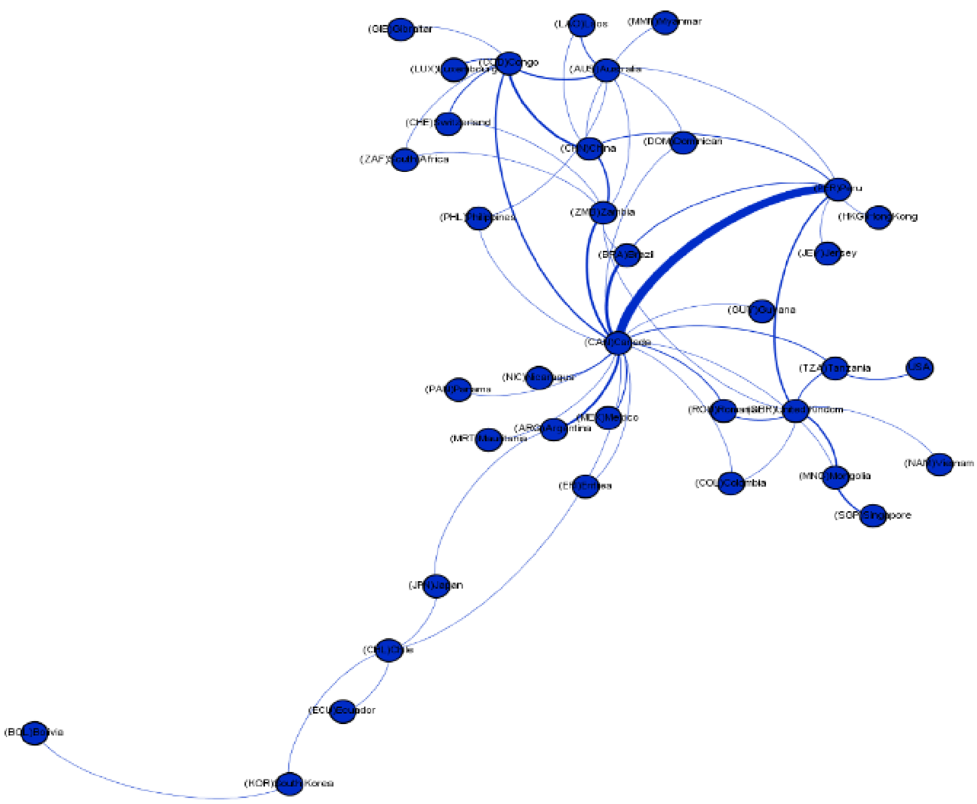

Figure 4.2: Visualization of the Copper Network

4.1 Degree Distribution of Copper network

Firstly, the degree sequence of a directed graph is the list of its in-degree and out-degree pairs. the degree distribution captures only a small amount of information about a network. But that information still gives important clues into the structure of a network. As we discussed earlier that degree distribution in a network is the number of connections it has to other nodes and degree distribution is the probability distribution of these degrees over the whole network. Time complexity is given by O(V + E) where V and E are the numbers of vertices and edges in the graph respectively. The degree distribution of a network P(k), tells us the probability that a randomly chosen node will have degree k. In the copper network, we can observe that Canada has a high degree distribution with many connections. Whereas Australia and Peru countries have similar distribution and also vary over time. Below are the results of all the countries with their degree distribution values. The dark color represents a high degree distribution as shown in the below figure.16. Thus, It is evident from the visualization that Canada has the highest number of connections in the network. The maximum number of nodes acquiring degree value is below 9.

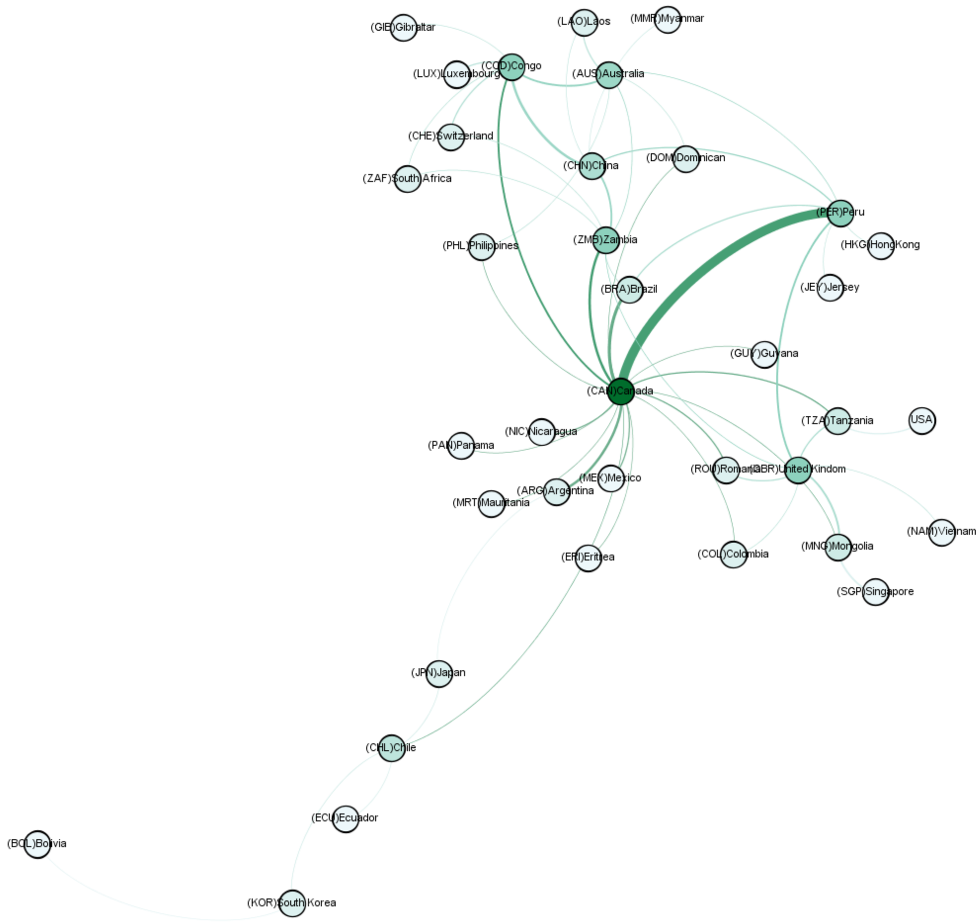

Figure 4.3: Copper network degree distribution

Degree View:
({'AUS': 6, 'LAO': 2, 'PHL': 2, 'PER': 7, 'DOM': 2, 'COD': 7, 'ZMB': 7, 'CHN': 5, 'MMR': 1, 'CAN': 18, 'GUY': 1, 'ERI': 1, 'BRA': 3, 'COL': 2, 'ARG': 2, 'MEX': 1, 'NIC': 1, 'PAN': 1, 'CHL': 4, 'MNG': 3, 'TZA': 3, 'MRT': 1, 'ROU': 2, 'ECU': 1, 'KOR': 2, 'BOL': 1, 'JPN': 2, 'GBR': 7, 'NAM': 1, 'HKG': 1, 'CHE': 2, 'ZAF': 2, 'JEY': 1, 'SGP': 1, 'USA': 1, 'LUX': 1, 'GIB': 1})

4.1.1 In-Degree Distribution

As we discussed earlier that for a directed graph G=(V(G), E(G)) and a vertex x1∈V(G) That is, the number of arcs directed towards the vertex x1. If we zoom in on a node in the network, we can observe some edges coming into the node and some edges going out from the node.

There are different colors for each network represented in the below figure.17 which determines the In-degree distribution between the copper network. The In-degree of node i is the total number of connections onto node I and is the sum of the ith row of the adjacency matrix. We might observe that the true degree distribution of a directed network is a joint distribution of In and out- degrees. A joint distribution can investigate the correlation of the In and out-degrees of vertices. For instance, if vertices with high out-degree tend to have also high In-degree. In our case, Countries like Zambia, Congo, Peru result in high In-degree distribution. whereas Tanzania, Mongolia, and Chile countries also have second high In-degree distribution with a value of 7, and Canada which appeared to have the largest number of connections in the degree distribution has 0 In-degree value.

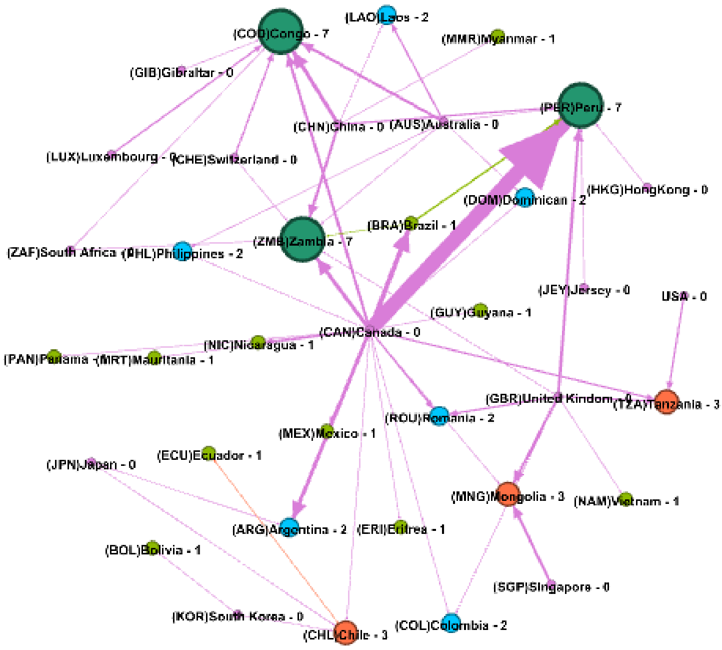

Figure 4.4: In-Degree Distribution of Copper Network

4.1.2 Out-Degree Distribution

The Out-degree representation is a bit opposite to the in-degree representation. In General, The Out-Degree of x1 refers to the number of arcs incident from x1. That's the number of arcs directed away from the vertex x1. A node is important if it has many neighbors, especially in the directed case, if any other nodes link to it, or if it links to many other nodes. As we see below figure.18 Canada has the highest out-degree value of 18. It means that Canada plays a major role as an exporter. Whereas Peru has a high Indegree distribution with a value of 0.

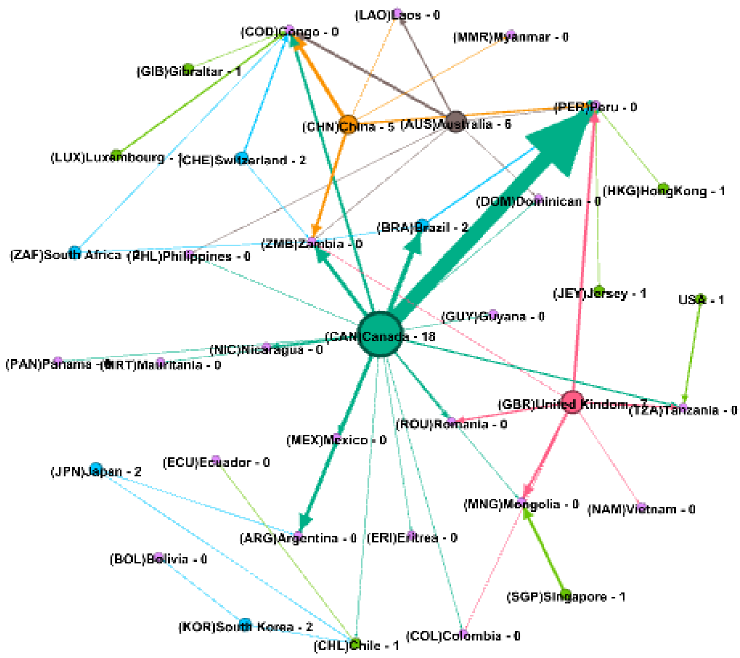

Figure 4.5: Out-degree distribution of Copper network

4.2 Graph Density in Copper Network

Density refers to the "connections" between participants. In our case density is defined as the number of connections countries has divided by the total possible connections where a country could have. As we discussed earlier the density of a network is a fraction of all possible links that are present. The density is not so high in our network it stands from 4 to 8 percent of nodes are connected in the network. For a directed, it is 4 percent and for undirected it is 8 percent. Intuitively, a directed graph has a total number of possible edges is given by $|v| \times |v-1|$ and in an undirected graph total no of possible edges is $|v| \times |v-1|/2$. In addition, the density of a network is the fraction between 0 and 1 that tells us what portion of all possible edges are realized in the network. Below is the density value of the copper network.

- Directed – 0.040
- Undirected – 0.080

4.3 Connected Components in Copper Network

4.3.1 Strong and weakly connected components

A graph is said to be strongly connected if every vertex is reachable from every other vertex and a weakly connected graph refers to all vertices that are connected by some path. Strongly and weakly connected components in a copper network are different from another network. As we can observe there is only one weakly connected component in the network i.e. (there is no existing path between any two pairs of vertices). whereas strongly connected components are of 37. This tells us there are many paths between the nodes in the network. This is done by the algorithm DFS (i.e., Depth-first search). In the below figure we observe there is only one weak node. This tends our network is strongly connected and doesn't vary. For a directed graph the DFS works until it finds the first vertex that still has a neighbor that has not been visited before. It continues with this neighbor until it backtracks again. Eventually, it will visit all vertices reachable from v. Then a new DFS is started at some vertex that is not reachable from v until all vertices have been visited. This algorithm is used to find the strong and weak ties for our network. In the below figure.19 we can observe the results depicting both strong and weak ties of the network.

Results:

Number of Weakly Connected Components: 1
Number of Strongly Connected Components: 37

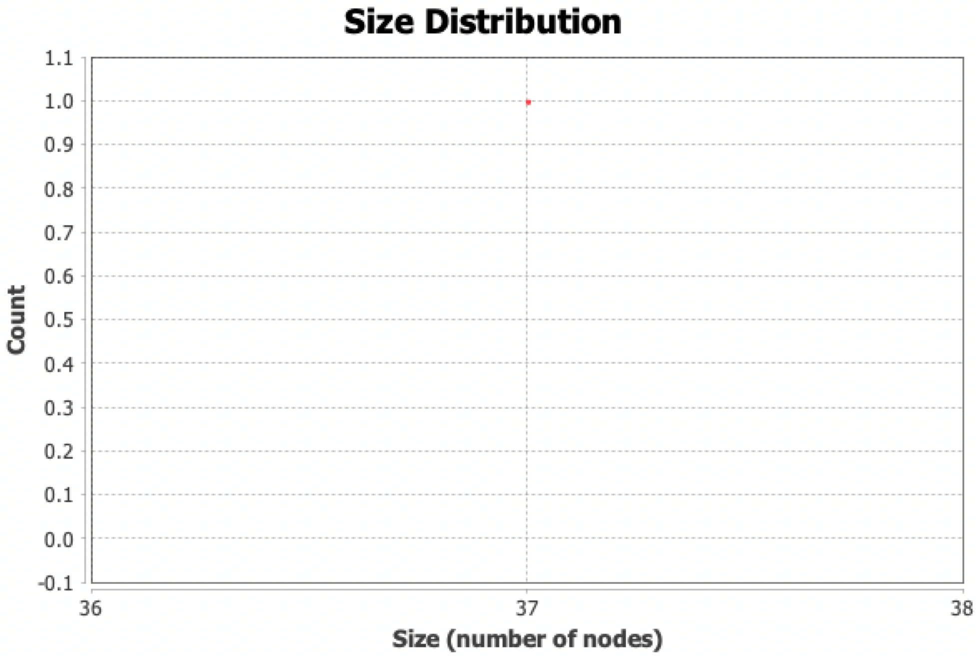

Figure 4.6: Strong and weak components in Copper network

4.4 Centrality Measures in Copper Network

4.4.1 Betweenness Centrality

Betweenness centrality measures the extent to which a vertex lies on paths between other vertices. Intuitively, the Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that goes through v. for weighted graphs the edge weights must be greater than zero which gives an infinite number of equal length paths between pairs of nodes. In the Figure below we can see that two nodes from Chile country are acting as a bridge between the remaining nodes. Moreover, the algorithm which has unweighted shortest paths between all pairs of nodes in a graph receives a score based on the number of shortest paths that pass through the node. Nodes that have the shortest paths between other nodes will have higher betweenness centrality scores.

Another interesting thing is that majority of the nodes (i.e., countries) have zero betweenness centrality score (i.e., not acting like a bridge). We can see that the node Chile has a high degree with a 3 betweenness score. In addition, the size of the edge represents the weight between countries i.e., how many times the Gold traveled between these countries. Although, the countries Canada and Peru have no importance in this network, however, the exchange of copper is comparatively more gold in many countries. It appears that Chile plays important role in the Copper network.

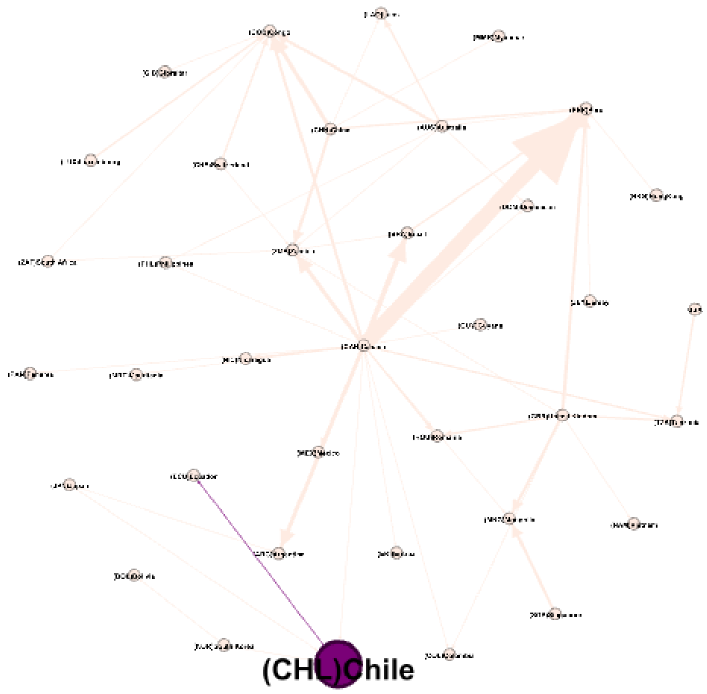

Figure 4.7: Betweenness centrality of the copper network

4.4.2 Closeness Centrality

The closeness centrality of a node measures its average inverse distance to all other nodes and the sum of distances depends on the number of nodes in the graph, closeness is normalized by the sum of minimum possible distances n-1. Nodes with a high closeness score have the shortest distances to all other nodes. In this section, we have presented the results of closeness centrality on the copper network for each node. The results below represent the nodes along with their Closeness centrality color where the score is normalized. According to closeness centrality, there are more chances that a node will be important if it is close to all other nodes in the network. In the below results we can observe that only one node Canada has high closeness centrality score apart from that Peru, Zimbabwe, Domenica all other nodes were scored with similar closeness centrality.

These are the values of closeness centrality measured using the network in python.

Results:
{'AUS': 0.3870967741935484, 'LAO': 0.2903225806451613,
'PHL': 0.4044943820224719, 'PER': 0.47368421052631576,
'DOM': 0.4044943820224719, 'COD': 0.4675324675324675,
'ZMB': 0.47368421052631576, 'CHN': 0.37894736842105264,
'MMR': 0.27692307692307694, 'CAN': 0.6206896551724138,
'GUY': 0.3870967741935484, 'ERI': 0.3870967741935484,
'BRA': 0.4444444444444444, 'COL': 0.4044943820224719,
'ARG': 0.3956043956043956, 'MEX': 0.3870967741935484,
'NIC': 0.3870967741935484, 'PAN': 0.3870967741935484,
'CHL': 0.4235294117647059, 'MNG': 0.41379310344827586,
'TZA': 0.41379310344827586, 'MRT': 0.3870967741935484,
'ROU': 0.4044943820224719, 'ECU': 0.3, 'KOR': 0.3050847457627119,
'BOL': 0.23529411764705882, 'JPN': 0.3050847457627119,
'GBR': 0.3956043956043956, 'NAM': 0.2857142857142857,
'HKG': 0.32432432432432434, 'CHE': 0.34285714285714286,
'ZAF': 0.34285714285714286, 'JEY': 0.32432432432432434,
'SGP': 0.29508196721311475, 'USA': 0.29508196721311475,
'LUX': 0.32142857142857145, 'GIB': 0.32142857142857145}

The yellow node in the figure.21 represents the node with high centrality score, followed by nodes having ocean color which are Peru, Congo, and Zambia.

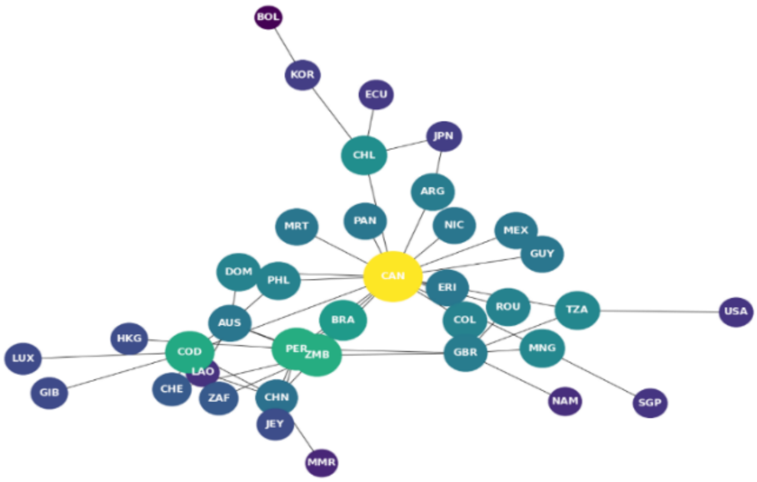

Figure 4.8: Closeness centrality of the copper network

Chapter 5

Comparison of Networks

Table 5.1: Comparison between gold and Copper Networks

No.	Property	Gold Network	Copper Network
1.	Number of Nodes	61	37
2.	Number of Edges	86	53
3.	Average Degree	2.869	2.811
4.	Highest Degree node	Canada	Canada
5.	Highest Indegree nodes	Brazil and Peru followed by Congo Argentina, Mali, and Burkina Faso.	Zambia, Congo, Peru
6.	Highest Out-degree nodes	Canada, United Kingdom	Canada
7.	Graph Density	Directed- 0.023 Undirected - 0.047	Directed- 0.040 Undirected- 0.080
8.	Betweenness Centrality	Brazil, Congo	Chile
9.	Closeness Centrality	Canada followed by Mali, Peru, Congo	Canada followed by Peru, Congo, Zambia

Chapter 6

Conclusion and Future Work

The multilayer network case study on the countries (i.e., on lands) tells us about the connections between the networks (i.e. Layers) and their importance by performing analysis and by visualizing the connections which vary in time and conditions. By evaluating various metrics for each of the networks gives us results that tend to further enhance communications of those countries and Analysis of these networks enabled us to find the countries who are leading in importing and exporting Gold and Copper. Finally, we deduced the following quantitative results from the extensive data analysis.

- Highest imports of countries.
- Highest exports of countries.
- The pattern of transport of Gold and Copper.
- The trade relationship between countries.
- Reachability of minerals between countries.
- Inter-connectivity between countries.
- Network density.

The data analysis can be devised from the graph theory to better understand the metrics like density, degree distribution, connected components, and centrality measures. Furthermore, we have compared the Gold and Copper network obtained through these approaches.

As result we have inferred that the network of copper is denser than the network of Gold, the relations of countries in transferring the Copper are more than Gold. In the case of the degree distribution, both networks follow the same pattern i.e., only Canada has the highest degree value. Although in a gold network the value is 25, and in a copper network the value is 18. The major exporter in both networks is Canada although there is another country - GBR (United Kingdom) leading in Copper Network.

In the Gold Network, there are two disconnected components from the networks-Senegal-Guernsey and Eswatini-Mozambique. These countries have no connections with the rest of the giant network. While there are no such components in the copper network. Similarly, it is found that there are more weakly connected components in the gold network than the Copper network. Likewise, the Gold network has more strongly connected components than the Copper network. The countries in the copper network are reachable to all other countries, while in the gold network few of the countries are not reachable to all other countries. While Brazil and Congo act as the bridges in the gold network, Chile acts as the bridge in the Copper Network.

If we look at the centrality scores Canada has the highest scores in both the networks. Followed by Peru and Congo and a few more countries. To conclude, the study upon the copper and gold from various countries tells us how networks are being connected and how they are interrelated with importing and exporting.

6.1 Future Scope

This work gives us insights for intelligent data analytics of different datasets, a similar approach used in this project with a dataset that contains information regarding importing and exporting minerals from countries. Land matrix helps us to provide relevant and accurate data, it is an open-access platform to find detailed information about deals in almost 100 countries.

The tool used in this project is named Gephi, apart from Gephi, there are several different software tools such as Graphviz and Neo4j. In addition, the data analysis of the dataset is implemented in python. We can also use different software platforms to get a better understanding of social networks.

Bibliography

[Lan,] Land matrix. https://landmatrix.org.

[Al-Shawi and Laurent, 2016] Al-Shawi, M. and Laurent, A. (2016). *Designing for Cisco Network Service Architectures (ARCH) Foundation Learning Guide: CCDP ARCH 300-320*. Cisco press.

[Barabási, 2016] Barabási, A.-L. (2016). Network science. *Cambridge, U.K.:Cambridge Univ. Press*.

[Berlingerio et al., 2013] Berlingerio, M., Pinelli, F., and Calabrese, F. (2013). Abacus: frequent pattern mining-based community discovery in multidimensional networks. *Data Mining and Knowledge Discovery*, 27(3):294–320.

[Boccaletti et al., 2014] Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics reports*, 544(1):1–122.

[Boccaletti et al., 2006] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308.

[Boorman and White, 1976] Boorman, S. A. and White, H. C. (1976). Social structure from multiple networks. ii. role structures. *American journal of sociology*, 81(6):1384–1446.

[Borg and Groenen,] Borg, I. and Groenen, P. J. F. Modern multidimensional scaling:theoryandapplications.springerscience&business media.

[Bródka et al., 2018] Bródka, P., Chmiel, A., Magnani, M., and Ragozini, G. (2018). Quantifying layer similarity in multiplex networks: a systematic study. *Royal Society open science*, 5(8):171747.

[Bródka et al., 2011] Bródka, P., Skibicki, K., Kazienko, P., and Musiał, K. (2011). A degree centrality in multi-layered social network. In

2011 International Conference on Computational Aspects of Social Networks (CASoN), pages 237–242. IEEE.

[Cao et al., 2016] Cao, S., Lu, W., and Xu, Q. (2016). Deep neural networks for learning graph representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

[Carchiolo et al., 2011] Carchiolo, V., Longheu, A., Malgeri, M., and Mangioni, G. (2011). Communities unfolding in multislice networks. In *Complex Networks*, pages 187–195. Springer.

[Casiraghi et al., 2017] Casiraghi, G., Nanumyan, V., Scholtes, I., and Schweitzer, F. (2017). From relational data to graphs: Inferring significant links using generalized hypergeometric ensembles. In *International Conference on Social Informatics*, pages 111–120. Springer.

[Cavallari et al., 2017] Cavallari, S., Zheng, V. W., Cai, H., Chang, K. C.-C., and Cambria, E. (2017). Learning community embedding with community detection and node embedding on graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 377–386.

[Chakraborty and Narayanam, 2016] Chakraborty, T. and Narayanam, R. (2016). Cross-layer betweenness centrality in multiplex networks with applications. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 397–408. IEEE.

[Cherven, 2015] Cherven, K. (2015). *Mastering Gephi network visualization*. Packt Publishing Ltd.

[Clauset et al., 2004] Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.

[Comer, 2015] Comer, D. (2015). Internetworking with tcp/ip volume i: Principles protocol and architecture 6th ed. *Pearson*.

[De Domenico et al., 2015a] De Domenico, M., Lancichinetti, A., Arenas, A., and Rosvall, M. (2015a). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, 5(1):011027.

[De Domenico et al., 2015b] De Domenico, M., Nicosia, V., Arenas, A., and Latora, V. (2015b). Structural reducibility of multilayer networks. *Nature communications*, 6(1):1–9.

[Easley et al., 2012] Easley, D., Kleinberg, J., et al. (2012). Networks, crowds, and markets: Reasoning about a highly connected world. *Significance*, 9(1):43–44.

[Gardner and Dorling, 1998] Gardner, M. W. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)–a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627–2636.

[Harrer and Schmidt, 2012] Harrer, A. and Schmidt, A. (2012). An approach for the blockmodeling in multi-relational networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 591–598. IEEE.

[Ikechukwu, 2017] Ikechukwu, U. I. (2017). A survey on bandwidth management techniques via the osi model network and application layers. *Global Journal of Computer Science and Technology*.

[Kourtellis et al., 2013] Kourtellis, N., Alahakoon, T., Simha, R., Iamnitchi, A., and Tripathi, R. (2013). Identifying high betweenness centrality nodes in large social networks. *Social Network Analysis and Mining*, 3(4):899–914.

[Kurose, 2012] Kurose, J. F. (2012). Keith w. ross computer networking: A top-down approach.

[Mccabe, 2007] Mccabe, J. (2007). Network analysis, architechture and dedign 3rd ed. *Morgan KAufamaan Publishers*.

[Nagy et al., 2020] Nagy, L., Ruppert, T., and Abonyi, J. (2020). Analytic hierarchy process and multilayer network-based method for assembly line balancing. *Applied Sciences*, 10(11):3932.

[Newman, 2010] Newman, M. E. J. (2010). Networks: An introduction. *Oxford: Oxford University Press*.

[Pilny and Atouba, 2018] Pilny, A. and Atouba, Y. (2018). Modeling valued organizational communication networks using exponential random graph models. *Management Communication Quarterly*, 32(2):250–264.

[Platt, 2019] Platt, E. L. (2019). *Network Science with Python and NetworkX Quick Start Guide: Explore and Visualize Network Data Effectively*. Packt Publishing Ltd.

[Porter et al., 2009] Porter, M. A., Onnela, J.-P., Mucha, P. J., et al. (2009). Communities in networks. *Notices of the AMS*, 56(9):1082–1097.

[Tanenbacum and WEtherall, 0011] Tanenbacum, A. and WEtherall, D. (20011). Computer networks 5th ed, prentice hall press.

[Valles-Catala et al., 2016] Valles-Catala, T., Massucci, F. A., Guimera, R., and Sales-Pardo, M. (2016). Multilayer stochastic block models reveal the multilayer structure of complex networks. *Physical Review X*, 6(1):011036.

[van der Pol, 2019] van der Pol, J. (2019). Introduction to network modeling using exponential random graph models (ergm): theory and an application using r-project. *Computational Economics*, 54(3):845–875.

[Wehrmann et al., 2018] Wehrmann, J., Cerri, R., and Barros, R. (2018). Hierarchical multi-label classification networks. In *International Conference on Machine Learning*, pages 5075–5084. PMLR.

[White and Donohue, 2014] White, R. and Donohue, D. (2014). *The art of network architecture*. Pearson Education.