Applications of Network Theory in Climate Analysis

By

Origene Tuyishimire (origene.tuyishimire@aims.ac.rw)
African Institute for Mathematical Sciences (AIMS), Rwanda
Supervised by: Doctor Philip Knight
University of Strathclyde, Scotland

June 2021

AN ESSAY PRESENTED TO AIMS RWANDA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN MATHEMATICAL SCIENCES

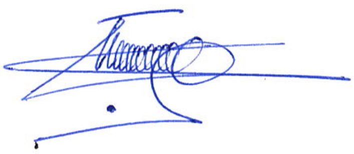
Abstract

Several statistical methods have been applied to analyze the climate system and remote connections in that system, but those methods are far from perfect when dealing with such a big and complex system. Therefore, we present another method called "Network Theory" that is specifically designed to analyze complex systems like climate. From Network Theory, we apply centrality measures and community detection algorithms that can uncover links and influential factors in the climate system. With Network Theory, we get a new way of extracting the deepest information in climate networks like how two distant regions are related and the influential factor behind that relationship. Using time-series precipitation data on global and local levels, Network Theory was able to highlight the importance of ocean weather systems on terrestrial weather. We were able to view the effects of trade winds originating from the equatorial eastern hemisphere mainly across the Pacific ocean as they pass around the globe. We were able to identify changes in climate network structure over time, which revealed patterns of climate change. Results obtained proved that Network Theory can be reliable when analyzing climate.

Keywords: Network Theory; Climate network; Centrality measures; Community detection; Trade winds.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.



Origene Tuyishimire, June 2021

Contents

List of Figures

1. Introduction

1.1 Background

Climate change has become an important topic mostly due to its impact on humanity and its surroundings. Several researchers have linked global issues like increasing trends of diseases, droughts, shrinking agricultural yields to climate change. Although many statistical methods have been used to predict how climate will change and the impacts, it is still not easily predictable. A climate event in one area can alter the climatic conditions of another geographical area. This shows that there is a climate teleconnection between geographical regions that are thousands of kilometers apart and Network Theory provides a new way of understanding this dynamic (Deza, 2015).

Networks are of big importance in human existence, there is a network of planets which earth is part of, the internet networks that connect the whole world, the social networks, the transport networks, biological, physical and so many others (Estrada and Knight, 2015). It is impossible to disregard the presence of networks in one's daily life. Following Estrada (2012), a network can be defined as "a diagrammatic representation of a system", in which different parts or locations are represented by 'nodes' which can have links or relationships, usually called 'edges', connecting them. Many areas of research have benefited from being analyzed in terms of networks, such as transportation, social relations, biology, and electricity. This provides cognizance of what is driving the system, how entities are connected, related, and affect each other (Spyrou and Escudero, 2018).

Although it is still a young application in climate science analysis, Network Theory has already proven to be a useful tool in offering predictive insight to researchers in this field. Usually, people look for specific areas and they see the dynamic and changes like the increase in temperature and precipitation, but in climate, we also need to see and understand the interactions between locations, which was missing in classical statistical methods. Understanding these interactions can unveil how one location affects another. Recently Donges et al. (2009a) observed the long-range connections in El Nino climate networks. Through this research, it was demonstrated that the El Nino-Southern Oscillation (ENSO) has a huge amount of power to affect the global climate system's stability.

When one takes different geographical areas (nodes) and adds edges linking the nodes based on their degree of interdependence applying tools from Network Theory, one can glean information of how the system's components are related mainly using the correlation network method. This can reveal hidden links between distant places and it has been proven that by grouping nodes linked in this way one can potentially improve the predictive potential of historical data (Donges et al., 2009b).

Network analysis tools used include centrality measures and community detection algorithms. Centrality measures such as betweenness, closeness, and eigenvector focus on examining the influence of each node on others, thus identifying the influential ones while Community detection help in the identification of sub-network clusters in a bigger Network, every sub-Network has patterns of connections, which indicates the level of resemblance between nodes (Estrada, 2012). This research was aimed at exploring ways of constructing climate networks from time-series data of mean monthly precipitation while focusing on discovering connections between distant regions and what lies behind those connections.

1.2 Problem Statement

Climate change has become a major area of research, many different approaches have been used to try to understand the global climate system and the changes it is undergoing. Various statistical methods such as linear correlation and principal component analysis (PCA) have been applied to measure the level of mutual dependence between climatic conditions of different regions, but these methods only look for relationships and don't look for the forces behind that relationships and how those forces are transferred. While these methods are still useful particularly in dealing with linear systems, the climate system has shown evidence of being seriously nonlinear: climate drivers are often not proportional to the climate conditions they give rise to, a small perturbation in initial conditions can result in an enormous difference in subsequent events which may even be disastrous and irreversible events.

This shows a need for a method that can handle linear systems and make use of discrete mathematics to deal with nonlinear systems such as the climate system. And with this we would be able to analyze the climate system, to be able to develop a better model and improve our forecasting, thus providing well-grounded information on what needs to be done. This essay focus on using the complex Network Theory method in analyzing the climate system by discovering hidden links between geographical regions that may be in great distance.

1.3 Aims and Objectives

The purpose of this research is to build climate Networks from climate data of precipitation monthly mean and use those Networks to investigate connections between geographical areas, examine climate Networks' strength and their effects on the global or local climate variability.

1.3.1 Specific Objectives.

- Building climate Networks of different geographical regions.
- Discover hidden climate links between distant geographical regions.
- Investigate why and how distant regions might be related using different measures.
- Use climate science to understand the driving force behind climate networks and their variability.

2. Literature Review

After introducing graph theory and the main idea of network theory and some of its uses, this section will briefly discuss previous applications of network theory in many fields of study, including climate research.

Network theory has been around since 1736 when Leonard Euler published his work titled "The Solution of a Problem Relating to the Theory of Position". In this work, Euler analyzed the Königsberg bridge problem, which was concerned with how one can pass once on each of the seven bridges over the river called Preger located in Kaliningrad, Russia now. As it is seen in Figure 2.1, Euler Reformulated the problem in a way that looked more like what we currently call a 'Graph'. From there he proved that there couldn't be such a path and this marked the birth of Graph Theory as a new branch of mathematics (Gross and Yellen, 2005).

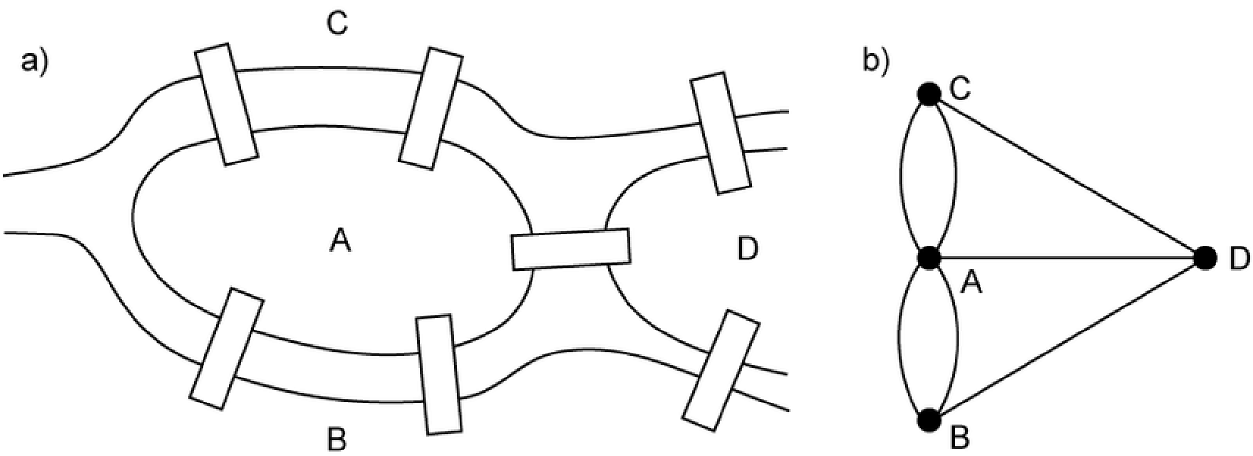


Figure 2.1: Königsberg bridge problem (a), and it's reformulation by Euler (b)
Source:researchgate.net, Accessed on the 5th May 2021

Since then, this new concept has been an integral part of solving complex problems as well as dealing with complex systems in different sciences fields, with climate science being the latest.

Among the first fields to use network theory was sociology, as per Barnes (1969) and Scott (1988). In social life, relationships are very important than personal attributes. however, both personal attributes and relationships are necessary for the understanding of a social structure, argued by Otte and Rousseau (2002). Thus network theory is preferred because it considers social structure to be a network of links that link people and channel resources, it focuses on the qualities of links rather than the qualities of individual members, and it considers communities to be networks of individual relationships that people cultivate, and sustain in their lives. In Social network analysis, social structures are examined using networks combined with a number of centrality measures and mutual information that explains the structure of social networks and explain the underlying dynamics and patterns seen in the structures (Otte and Rousseau, 2002).

In research methods and optimization, network theory has been used for quite a while, Likaj et al. (2013) used it to analyze the transport network of roads, railways, air transport, and water transport all from the manufacture to the consumer, with an aim of finding the shortest path with minimum possible transport costs. A network model of the transportation problem was constructed and thoroughly examined in order to decrease transport or shipment costs. To find an answer to the challenge of determining a Minimum Spanning Tree, the Kruskal algorithm was applied and the Dijkstra algorithm was used to determine the shortest path between two locations. The results produced showed that Algorithm Dijkstra is very effective in determining the shortest path from one node to another while utilizing the Kruskal method for the Minimal Spanning Tree, proved to be productive as it has shown a minimum spanning tree to reach the destination node from the manufacturer with the lowest overall cost (Likaj et al., 2013).

In computer networks, this theory has found applications in different aspects, such as cloud computing, resource allocation, and routing. The routing process involves choosing a path in which information will pass through going from node A to node B in an internet network. Like in network theory, computer science also uses the term "node" to refer to any device connected to a network. According to Malviya and Malviya (2018), when a node needs to share information with a particular node or to broadcast in an entire network, the shortest path available has to be chosen and a routing protocol called Open Shortest Path First (OSPF) is responsible for this choice.

With an aim of finding the shortest and cost-effective path for information sharing, OSPF employs a popular algorithm in graph theory called Dijkstra's algorithm. Since the departure node have information about all possible paths reading to the destination node(s), Dijkstra's algorithm is used to weigh the costs of using every possible path, and conclude on which one is effective, thus fostering fast communication while keeping costs low (Malviya and Malviya, 2018).

In climate science, network theory found its first application in the late 1990s and early 2000's when researchers like, Tsonis et al. (2006) wanted to know what networks have in common with climate, in their research they proposed one way that networks can be applied to climate is by taking a grid of oscillators and assume they represent the climate, each oscillator is then representing a dynamic system that behaves in its own ways. Collectively we are then interested in the behaviors of all the dynamical systems interacting and the network structure they form.

According to Tsonis and Roebber (2004), the climate system has a “small-word” architecture. this means that it is stable and a good information transfer. These characteristics allow it to fast respond to any fluctuations caused by El Nino or any other phenomenon, thus maintaining its stability.

Donges et al. (2009b) used this theory in their research, titled “The backbone of the climate network”, In this research, they applied mutual information approach in building networks, which enabled them to extract both nonlinear and linear relationships between time-series data, This approach was combined with the use of centrality measures and betweenness. The combination of these methods applied on the monthly average surface area temperature (SAT) data, enabled them to uncover that there are pathways of dynamic information flow and global energy in the global climate system, and this was termed “The backbone of the climate network”. Surface ocean currents play a significant role in the transfer of information and energy, through distributing heat all around the globe, they have much power to influence the global climate and it was apparent in this research.

In research of Steinhaeuser et al. (2009) titled "An Exploration of Climate Data Using Complex Networks", the cross-correlation between four variables namely, Air temperature, pressure, relative humidity, and precipitation was used to measure the level of similarity between regions. Considering that some climate phenomena might take time to occur and impact different regions, the cross-correlation was computed for lags or delays of -6 to 6 months. A weighted network is constructed, where the edge's weight is defined by the cross-correlation value between the two data points.

Using the "WalkTrap" algorithm for community detection Steinhaeuser et al. (2009) discovered 4 different communities. Community one was comprised of Tropical Wet-Dry parts of South America and Africa and Monsoon climate zones of South-East Asia. Similar to the first one, the second community was also comprised of Tropical Wet-Dry regions and small parts of Monsoon regions, like northern India to be specific. The third community was comprised of Sub-Arctic climate regions and Mediterranean regions, while The last one was comprised of largely Greenland and Antarctica or some mountains like the Andes, Rocky Mountains, and the Himalayas. Some locations in central Asia and South America were also included. The formation of these communities was attributed to climate regions that are defined primarily based on precipitation and temperature found in each region.

3. Methodology

This chapter looks at methods used to construct climate networks and analyzes the constructed networks. Among those methods, we look at the correlation coefficient matrix, Adjacency matrix, and Network centrality measures like betweenness, closeness, eigenvector, and others. Python programming provided an ideal environment for the computations.

3.1 Constructing Climate Networks

Mathematically, a network is defined as the pair $G = (V, E)$, where the set V contains vertices and set E contains Edges. It is built by connecting the related nodes using edges. The intensity of the relationship between nodes is defined by the correlation between those nodes, and the adjacency matrix is used to specify which level of relatedness is good enough to consider two nodes as adjacent. In this research, the used networks are unweighted and undirected, meaning that there can only be one edge between two nodes.

3.1.1 The Correlation Coefficient Matrix.

The first step of network construction is building a correlation matrix, Normally the correlation is used to show the strength of a linear relationship between two variables representing two geographical regions in this research. The correlation can be any number in the range of -1 to $+1$ inclusive. If the correlation between two geographical regions is equal or above 0.5, we consider those regions to be related and changing concurrently, but if the correlation is less than -0.5, the two regions are inversely related. i.e one will increase while the other is decreasing.

Given two variables x and y with measurement records (x_i, y_i) by indexing $i = 1, ..., n$ The correlation coefficient (r_{xy}) between x and y is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{3.1.1}$$

Where:

- $\sum_{i=1}^{n}$ is the summation
- $(x_i - \bar{x})$ is the subtraction of each value of x and the mean of x
- $(y_i - \bar{y})$ is the subtraction of each value of y and the mean of y

By computing the correlation between each region and others, one ends up making the correlation matrix, which is a matrix containing the linear correlations between all variables under study, and from this, one can easily see which regions are related and on what scale.

3.1.2 The Adjacency Matrix.

This is an important tool in network analysis, it checks whether or not there is a relationship between two regions. According to Bloch et al. (2019), An adjacent matrix is denoted by $A \in \mathbb{R}^{n\times n}$, where $A_{ij} = 1$ means there is a connection between two nodes, while $A_{ij} = 0$ means there is no connection between the two nodes. it can be computed by setting the threshold limit in the correlation matrix so that if the two regions correlate 0.5 or above, they are adjacent, otherwise, they are not adjacent. The threshold values of 0.5 are used to better get all the possible information.

The Adjacent matrix is then defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } r_{x_i y_j} \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{3.1.2}$$

3.2 Centrality Measures

Network centrality measures are measures that help in gleaning information out of networks. According to Bloch et al. (2019), they provide information related to the positioning of each node in a network and the influence of that node in a network, those information are related but also not limited to the paths from a particular node to others, the walks from that particular node to others, and the shortest possible path between two nodes in a network.

Some centrality measures like degree and closeness are normalized by multiplying or diving it by $(n-1)$ which put the value of that measure between -1 and 1, but normalizing is not important since we are only interested in the ranking of each node rather than the absolute value.

3.2.1 Degree Centrality.

Degree centrality is a centrality measure that counts how many neighbors a particular node has in a network, this is the number of edges directly linked to each node. So the node with many edges connected to it will be the most central. This shows which nodes can directly influence or be influenced by this node. The degree centrality of a node then denoted as

$$C^*_D(i) = \frac{\sum\limits_{j=1}^{n} A_{ij}}{n-1} \tag{3.2.1}$$

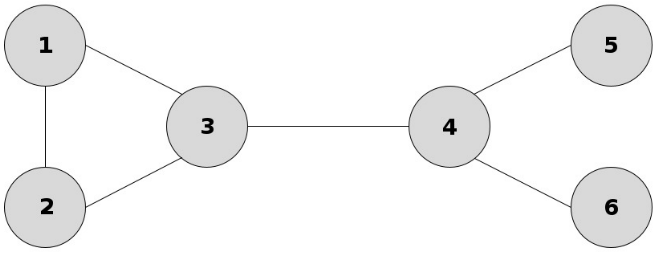


Figure 3.1: Simple network

Considering a simple network in figure 3.1 which consists of 6 nodes and 6 edges, one can calculate Degree centrality of each node as follows:

	1	**2**	**3**	**4**	**5**	**6**	**sum**		**n-1**		**Degree**
1		1	1	0	0	0	2		5		2/5
2	1		1	0	0	0	2		5		2/5
3	1	1		1	0	0	3	/	5	=	3/5
4	0	0	1		1	1	3		5		3/5
5	0	0	0	1		0	1		5		1/5
6	0	0	0	1	0		1		5		1/5

Table 3.1: Degree Centrality

3.2.2 Closeness Centrality.

Closeness centrality is a measure of the mean distance from a node to other nodes. this can also be defined as the average length of the shortest path between the node and all other nodes.

$$C_C(i) = \frac{n-1}{\sum\limits_{j=1}^{n} s(i,j)} \tag{3.2.2}$$

where:

- n: Total number of nodes
- $\sum\limits_{j=1}^{n} s(i,j)$: The sum of shortest paths from the node represented by i to all the nodes in a network represented by j.

Again considering the same simple network in figure 3.1,the closeness centrality of each node is calculated as follows:

	1	**2**	**3**	**4**	**5**	**6**	**n-1**		**sum**		**Degree**
1		1	1	2	3	3	5		10		5/10
2	1		1	2	3	3	5		10		5/10
3	1	1		1	2	2	5	/	7	=	5/7
4	2	2	1		1	1	5		7		5/7
5	3	3	2	1		2	5		11		5/11
6	3	3	2	1	2		5		11		5/11

Table 3.2: Closeness Centrality

The difference between Degree Centrality and Closeness Centrality is that in Degree centrality we look at the number of nodes, a specific node is connected to, while closeness centrality only focuses on the mean distance from a node to other nodes.

3.2.3 Betweenness Centrality.

Betweenness Centrality is a measure of how many times a particular node serves as a bridge along the shortest path between two other nodes. This is denoted by:

$$C_B(i) = \sum_{x \neq y \neq i} \frac{\sigma_{xy}(i)}{\sigma_{xy}} \tag{3.2.3}$$

where:

- σ_{xy} is the total number of shortest paths connecting nodes x and y.
- $\sigma_{xy}(i)$ is the total number of shortest paths connecting nodes x and y that pass through i.

By considering the same network in figure 3.1, One can calculate the betweennes centrality of **node 3** as follows:

	σ_{xy}	$\sigma_{xy}(i)$	$\sigma_{xy}(i)\ /\sigma_{xy}$
(1,2)	1	0	0
(1,4)	1	1	1
(1,6)	1	1	1
(1,5)	1	1	1
(2,4)	1	1	1
(2,6)	1	1	1
(2,5)	1	1	1
(4,6)	1	0	0
(4,5)	1	0	0
(5,6)	1	0	0
(Sum)			**6**

Table 3.3: Betweenness Centrality

Title 1	Title 2
(1,2)	0
(1,4)	1
(1,6)	1
(1,5)	1
(2,4)	1
(2,6)	1
(2,5)	1
(4,6)	0
(4,5)	0

Table 3.4: Betweenness Centrality

Thus it is seen that the betweenness of **node 3** is 6, and this same way should be used to calculate the betweenness centrality of each node in a network.

3.2.4 Eigenvector Centrality.

Eigenvector centrality is a measure of a node's relevance to its neighbors in a network. It assigns a score to each node in a network based on the fact that connections to high-scoring nodes contribute more to the node's score. This metric generalizes degree centrality by taking into account the relevance of the neighbors. It takes into account both the degree of a node and the degree of its neighbors. This means that a node with fewer connections might have a very high eigenvector centrality if those few connections are to other nodes that are very strongly linked (Zafarani et al., 2014).

3.2.5 Theorem (Perron-Frobenius Theorem)**.** *"Let $A \in \mathcal{R}^{n\times n}$ represent the adjacency matrix for a connected graph or $A : A_{i,j} > 0$ (i.e. a positive n by n matrix). There exists a positive real number (PerronFrobenius eigenvalue) $\lambda_{\max}$, such that $\lambda_{\max}$ is an eigenvalue of A and any other eigenvalue of A is strictly smaller than $\lambda_{\max}$. Furthermore, there exists a corresponding eigenvector $v = (v_1, v_2, \ldots, v_n)$ of A with eigenvalue $\lambda_{\max}$ such that $\forall v_i > 0$ " (Zafarani et al., 2014).*

So from Theorem 3.2.5, to compute the eigenvector centrality of a network one has to first compute the eigenvalues of Adjacency matrix A, then select the largest eigenvalue $\lambda_{\max}$ among all eigenvalues λ, he then computes the eigenvectors C_e associated to the selected highest eigenvalue which is $\lambda_{\max}$. According to Perron-Frobenius Theorem 3.2.5, all the elements of C_e must be positive.

Having an adjacent matrix A, and assuming that centrality C_e is an eigenvector matrix of A, the characteristic equation denotes that:

$$AC_e = \lambda C_e \tag{3.2.4}$$

or

$$(A - \lambda I)C_e = 0 \tag{3.2.5}$$

where:

- A is the Adjacent matrix
- C_e is the eigenvector of A
- λ is the eigenvalue of A

So the eigenvector centrality $C_e(i)$ of node i is the i^{th} element of the eigenvector corresponding to A's highest eigenvalue. This is equivalent to the summed centrality of i's neighbors:

$$C_e\left(i\right) = \frac{1}{\lambda}\sum_{j=1}^{n} A_{j,i} C_e\left(j\right) \tag{3.2.6}$$

By considering the same network in figure 3.1, and following the above Theorem 3.2.5 of Perron-Frobenius, the process of eigenvector centrality calculation will be as follow:

From our network, the corresponding Adjacency matrix is:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{3.2.7}$$

And from equation 3.2.5 the eigenvalues will be calculated using the determinant:

$$\det(A - \lambda I) = \begin{pmatrix} 0-\lambda & 1 & 1 & 0 & 0 & 0 \\ 1 & 0-\lambda & 1 & 0 & 0 & 0 \\ 1 & 1 & 0-\lambda & 1 & 0 & 0 \\ 0 & 0 & 1 & 0-\lambda & 1 & 1 \\ 0 & 0 & 0 & 1 & 0-\lambda & 0 \\ 0 & 0 & 0 & 1 & 0 & 0-\lambda \end{pmatrix} \tag{3.2.8}$$

By solving the determinant for λ one will get the eigenvalues $(2.28, 1.31, -1.89, -0.705, -1, 0)$, but following the theorem 3.2.5 only the highest eigenvalue $\lambda_{\max}$ is selected. i.e We will compute the eigenvector associated to **2.28**.

Then, using the equation 3.2.5, the eigenvectors will be calculated as follows:

$$\begin{pmatrix} 0-2.28 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0-2.28 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0-2.28 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0-2.28 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0-2.28 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0-2.28 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{3.2.9}$$

Where : $C_e = (v_1, v_2, v_3, v_4, v_5, v_6)^T$

After solving for the vector $\mathbf{C}_e$, the solution of will be:

$$C_e = (0.45, 0.45, 0.58, 0.41, 0.18, 0.18)^T \tag{3.2.10}$$

From this solution, one can see that the eigenvector associated with **node 3** is the highest, which indicates that node three is the most central among all. This measure is scalable and can be used on a network of much more nodes and indicate which ones are the most central.

3.3 Community Detection

Although a network represents a complex system, it is most of the times comprised of functional modules, also known as Communities in a network. These are clusters of nodes that are more connected between themselves than the rest of the network's nodes. The communities holds much information about the network like their strength and detecting them can help in gleaning valuable facts about that network (Khan and Niazi, 2017). A number of algorithms such as Louvain, OSLOM and infomap have been used over years in community detection of different networks but only Louvain algorithm is used in this research as it has proven to have a high community detectability performance according to (Fortunato and Hric, 2016).

According to Blondel et al. (2008), Louvain algorithm is an algorithm which works on maximizing the modularity level for every community within the network.

It is divided into two stages namely, modularity optimization and community aggregation that are iteratively replicated. It starts with every node in its own community, and then performs modularity optimization, in this phase a random order of the nodes is chosen and each node would be assigned to a new community until there is no significant change in modularity.

Modularity is defined as the measure of how dense are the connections within a community. It is a critical measure because it shows how similar are particular nodes in a network to the level at which a network can be split into structurally, functional distinct modules each comprised of nodes with similar characteristics. Modularity quantifies how often we can find same type nodes on both ends of an edge, by comparing the random graph with the real graph, to see if such connections are likely.

Modularity value ranges between -1 and 1 and a network with higher modularity level will be indicated by having many connections within a community while having fewer connections directed towards other communities. Modularity quantity Q of a community is also Denoted as:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta \left(C_{(i)}, C_{(j)} \right) \tag{3.3.1}$$

Where:

- A is the adjacency matrix

- m is the number of edges in the network
- k_i is the degree of vertex i
- $C_{(i)}$ is the community of vertex i
- $\delta(C_{(i)}, C_{(j)})$ is the Kronecker delta, it is 1 if if nodes i and j belong to the same community and 0 otherwise.

After the first step is completed, all nodes that are in a community are then combined to make a bigger node. Edges that connects the larger nodes are determined by the cumulative weight of the edges that previously connected the two communities and this step is called community aggregation.

$$\Delta Q = \left(\frac{\Sigma_{in} + 2k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right) - \left(\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right) \tag{3.3.2}$$

Where:

- Σ_{in} is the total score of all edges in the destination community of vertex i
- Σ_{tot} is the total score of all edges linking to nodes in the destination community of vertex i
- m is the number of edges in the network
- k_i is the degree of vertex i
- $k_{i,in}$ is the sum of the weights of the links between i and other nodes in in the destination community of vertex i .

After this step the two phases are therefore repeated until the maximum possible modularity is achieved and the process stops there (Blondel et al., 2008).

4. Results and Discussion

In this chapter, we look at the data used, present and discuss the results obtained from the computations and try to understand what they mean in regards to the weather and climate.

4.1 Data Processing

The Global Precipitation Climatology Project (GPCP) Monthly Analysis Product Dataset is used to construct networks (Adler et al., 2018). This dataset has a spatial resolution of $2.5°$ longitude $\times$ $2.5°$ latitude, and monthly mean values of precipitation from 1979 to January 2021 are recorded for each grid point. This spatial resolution gives 72 longitudinal points and 144 latitudinal points, which gives 10368 points, and these are what we consider as nodes of the network. for each grid point, we have 505 time-steps or values of precipitation amount. with this data, one could get the global picture of climate behavior with a focus on precipitation.

To look at a local level, a rather small dataset of annual Precipitation data over some selected African cities extracted from the CRU Dataset is used (Harris et al., 2020). This one contains yearly mean precipitation values from 1960 to 2019 for 46 cities across Africa, and this one gives a brief view of climate behaviors, similarities and differences between African cities that are listed below and the general climate status of their respective countries.

Index	Tag	City	Index	Tag	City	Index	Tag	City
0	C1	Abuja	16	C17	Dodoma	31	C32	Port_Harcourt
1	C2	Accra	17	C18	Juba	32	C33	Rabat
2	C3	Addis_Ababa	18	C19	Kampala	33	C34	Tripoli
3	C4	Algiers	19	C20	Kigali	34	C35	Windhoek
4	C5	Antananarivo	20	C21	Kinshasa	35	C36	Yamoussoukro
5	C6	Asmara	21	C22	Lilongwe	36	C37	Yaounde
6	C7	Bamako	22	C23	Maseru	37	C38	Abidjan
7	C8	Bangui	23	C24	Mbabane	38	C39	Alexandria
8	C9	Banjul	24	C25	Mogadishu	39	C40	Antsirabe
9	C10	Bissau	25	C26	Monrovia	40	C41	Bafata
10	C11	Brazzaville	26	C27	Nairobi	41	C42	Bamenda
11	C12	Bujumbura	27	C28	NDjamena	42	C43	Bata
12	C13	Cairo	28	C29	Niamey	43	C44	Beira
13	C14	Cape_Town	29	C30	Nouakchott	44	C45	Benghazi
14	C15	Conakry	30	C31	Ouagadougou	45	C46	Benoni
15	C16	Dakar						

Table 4.1: cities and their respective Indices and Tags

To uncover connections between regions based on the global dataset and cities in Africa, we compute the correlation matrix for each dataset. At this stage we start to get a general overview of both datasets, it is uncovered that some regions are related while others are not but this information is very general and that's why we apply a mathematical concept called Network Theory to help us understand how this region's climatic conditions are related, why they are related and what is the cause of that relation.

The introduction of network theory starts with computing adjacency matrices on both datasets, here we threshold the correlation matrix and assume that two regions are related if their correlation is between 0.5 and 0.8 on the global dataset. The correlations above 0.8 are not considered because the two regions with such a correlation are almost always geographically close to each other, and since our goal is to build a network of distant connected regions, it is better to remove such regions that overcloud the network while not giving necessary information. On the other hand, two regions whose correlation is between -0.5 and -1 indicate that when precipitation increase in one region it will decrease in another and vice versa. On a small dataset, cities whose correlation is between 0.5 and 1 are considered to be positively related or negatively related if the correlation is between -0.5 and -1. The adjacency matrices are computed, one for cities that are positively correlated, and another for cities that are negatively correlated, and from these adjacency matrices, we build different networks and apply different network measures to extract as much information as possible.

4.2 Analysis of Results

Among 46 cities, 34 showed signs of a significant relationship between one or more cities. By observing the bellow network 4.1, C41 which is Bafata from Guinea-Bissau looks to be related to many other cities, this is the same for other cities like Bissau and Banjul which are all from west Africa.

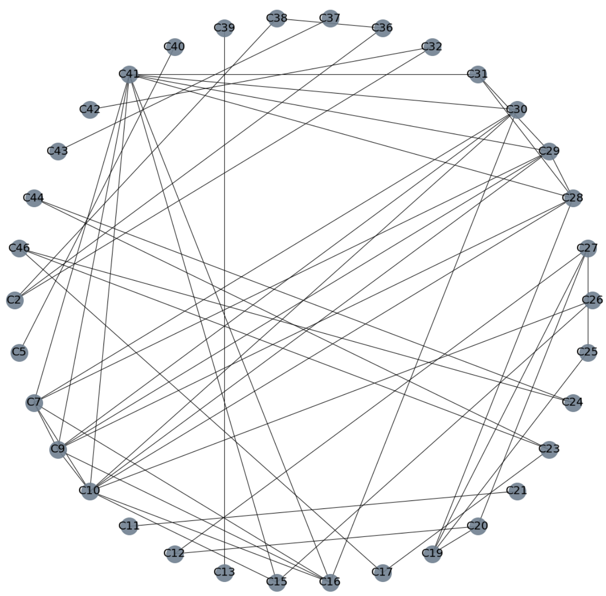


Figure 4.1: A circular network of positively correlated cities

Looking at the cities' geographical locations on map 4.2, almost all related cities are located in the tropical region which is also known as the equatorial region. This region is located at latitude $23.5°$ North of the equator down to $23.5°$ South of the equator. It has two climatic seasons, dry and wet, and the precipitation of this region is mainly defined by trade winds.

Trade winds, also known as easterlies are a type of wind that blows from east to the west but near the equator. They are mostly caused by warmth and evaporation in the equatorial atmosphere. As per the Hadley cell phenomenon, the equatorial region is closer to the sun, which makes it relatively warmer than poles, and as warm air from the equatorial region rise in the atmosphere, colder air sinks from the poles. These winds move to the west because as the earth rotates eastward the air in the north of the equator is deflected right, while the air in the south of the equator is reflected left and this is referred to as the Coriolis effect.

4.2.1 Network Analysis.

To analyze the network deeply, we visualize it on a geographical map 4.2, so we can consider the location of each city and the climate characteristics of that location. Here cities are by their indices as they are seen in the table

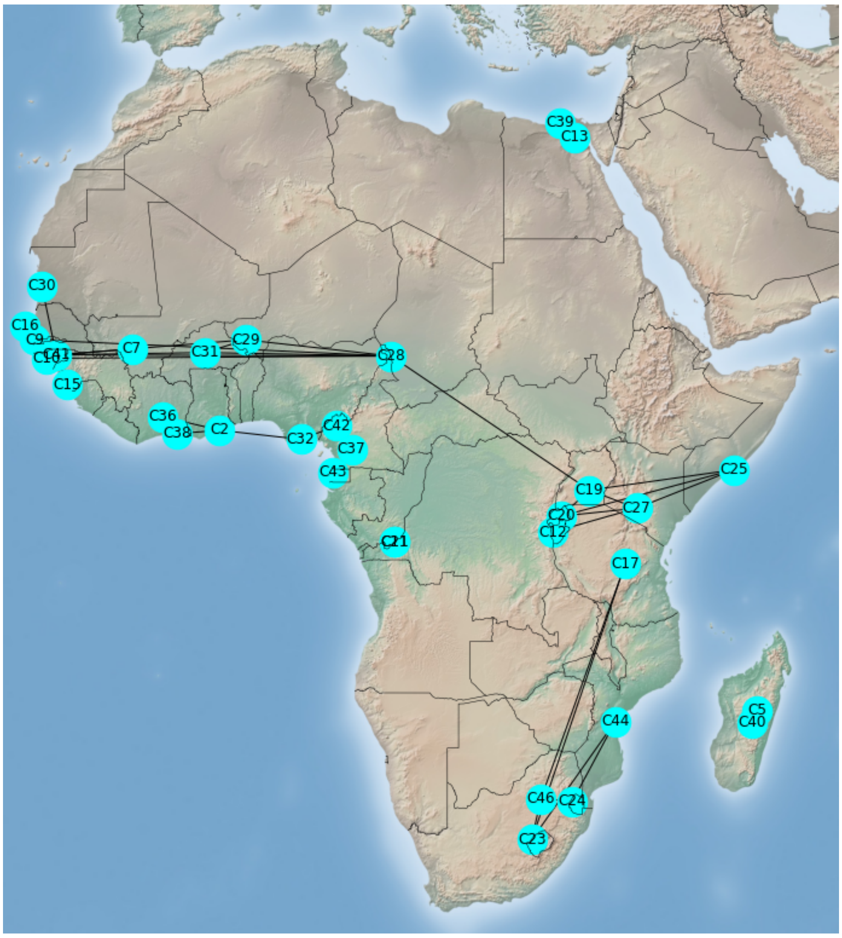
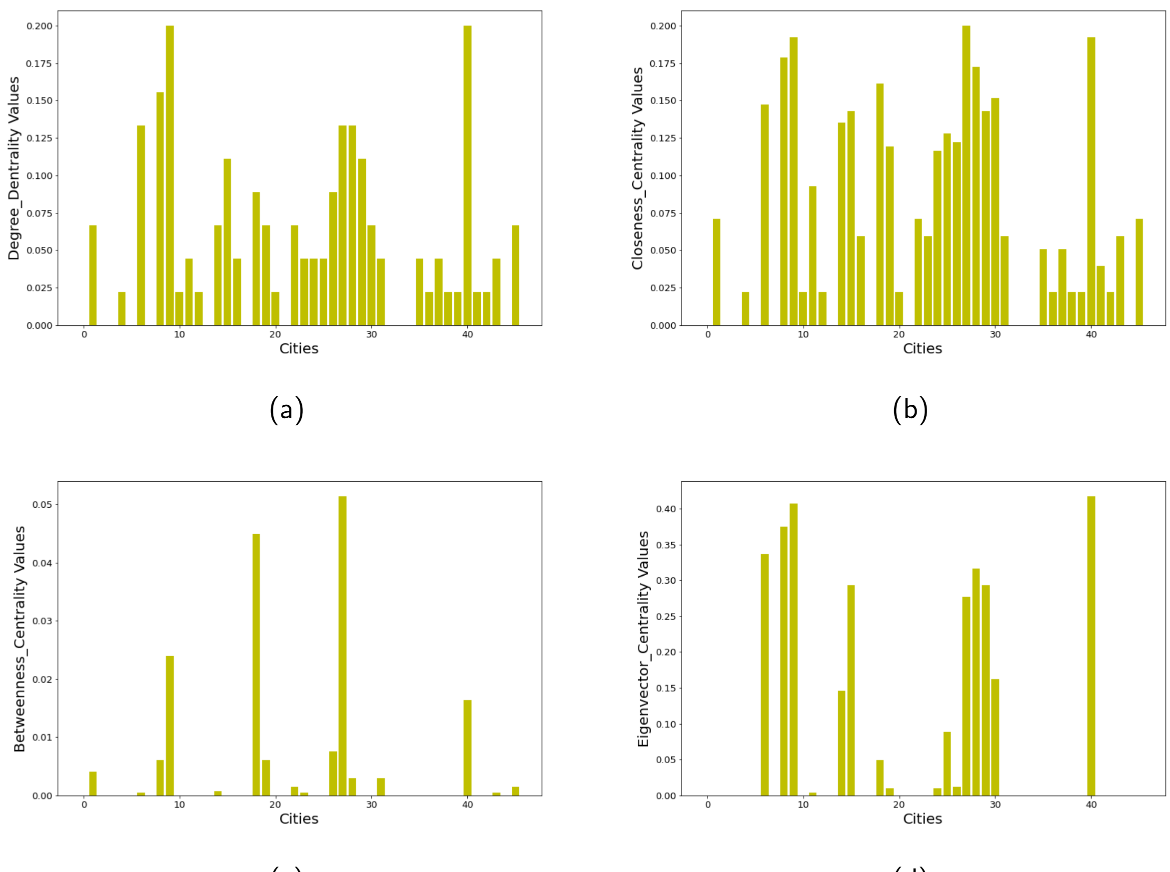


Figure 4.2: A geographical view of positively related cities

The connected cities are all in the equatorial region as it can be seen in 4.2, which means that their precipitation levels and change are affected by the trade winds as explained above. Although they are related, they have some different characteristics such as forests and industrial regions which makes them react differently to the trade winds and other atmospheric phenomena. This is why some cities are central in a network and others are not, and to evaluate how important some cities are to the network, centrality measures provide a big insight.

4.2.2 Cities Centrality Analysis.

In this research, centrality measures provided information about which regions are influential in network communication and stability. The result of centrality measures used is presented below bar plots so that one can easily see which cities stand out.



(a) (b) (c) (d)

Figure 4.3: Centrality measures observed on all 46 cities, 4.14a Degree Centrality, 4.14b Closeness centrality, 4.3c Betweenness centrality, 4.3d Eigenvector centrality

Looking at centrality measures bar plots 4.2.2, one can see that some cities stand out in all four measures used, and this indicates that those cities are of big importance to the network's stability and communication. Also, many cities from the northern part and southern part of Africa that are out of the equatorial region are not part of the network. This is a very important fact that highlights how unique the equatorial region can be just because of the trade winds. Focusing on the most influential cities, we have Bafata city (C41) and Niamey (C29) but to deeply understand network behavior, let us look at the 5 most influential cities.

Centrality Measures

Degree		Betweenness		Closeness		Eigenvector	
C41	0.200	C28	0.051	C28	0.199	C41	0.417
C10	0.200	C19	0.044	C41	0.192	C10	0.407
C9	0.155	C10	0.023	C10	0.192	C9	0.374
C29	0.133	C41	0.016	C9	0.178	C7	0.336
C28	0.133	C27	0.007	C29	0.172	C29	0.316

Table 4.2: Best 5 important cities

Generally, Bafata (C41) and Bissau (C10) all from Guinea Bissau, West Africa, are among the tops in all 4 measures, and this result shows that the positioning and climate characteristics of Guinea Bissau make it unique in the African climate system, but much deeper research is needed to find out why Guinea Bissau is unique.

Specifically, the Degree centrality measure again shows that two cities Bafata (C41) and Bissau (C10) from Guinea Bissau in West Africa, have the highest degree centrality 0.2 which is equivalent to nine connections followed by Banjul city from the Gambia (C9) with seven connections then Niamey (C29) from Niger and Ndjamena (C28) from chad with six connections. All of these cities are from central and west Africa, which takes us back to the trade winds. As (Richards et al., 2015) said, the prevailing trade winds blow westward, when these winds are blowing, its energy is being transferred in the atmosphere from one region to another and one city to another going further west, here the atmosphere transfers energy in the west following the wind's direction, which is why we see many cities in West Africa having many connections. Also, coastal regions of West African countries experience wind reversals due to the West Africa monsoon originating from the Atlantic ocean, and many cities near the ocean are affected by this (Richards et al., 2015).

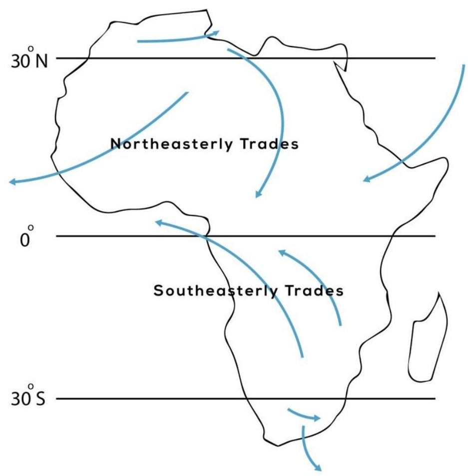


Figure 4.4: Trade winds directions over Africa both north and south of the equator
Source:researchgate.net, Accessed on the 8th June 2021

Closeness centrality shows that Ndjamena city (C28), Bafata city (C41) and Bissau (C10) have the highest closeness respectively, they are the best information spreaders, and thus they efficiently transfer trade winds from cities to cities. Betweenness centrality shows that Ndjamena city (C28) is the most important bridge in facilitating communication between other cities. This is understandable since Ndjamena is located in the central Africa region and so it acts as a hub of communication between eastern and western Africa cities. Kampala city (C19) of Uganda in East Africa comes second. The fact that these two cities with the highest betweenness centrality are from the east and central Africa region, means that their regions facilitate the communication and energy transfer of trade winds east to west.

As eigenvector centrality indicate the node that is connected to other powerful nodes, we uncover that Bafata city has the highest eigenvector, this is mostly due to its connection with some influential cities like, Ndjamena, and Niamey, which makes it more powerful. Generally, centrality measures enabled us to see how the trade winds affect the equatorial regions and how different cities react to these winds.

Richards et al. (2015) Argued that trade winds affect each region differently, mostly due to the geographical characteristics of each region, and from figure 4.4, we see that following the direction of trade winds, the regions in the west are on the receiving end of these wind's energy. We believe that it is for this reason west African cities have dominated in all centrality measures.

Network construction also enabled us to realize that some cities are affected differently to trade winds, and from the below circular network we could see a network of cities that are negatively correlated.

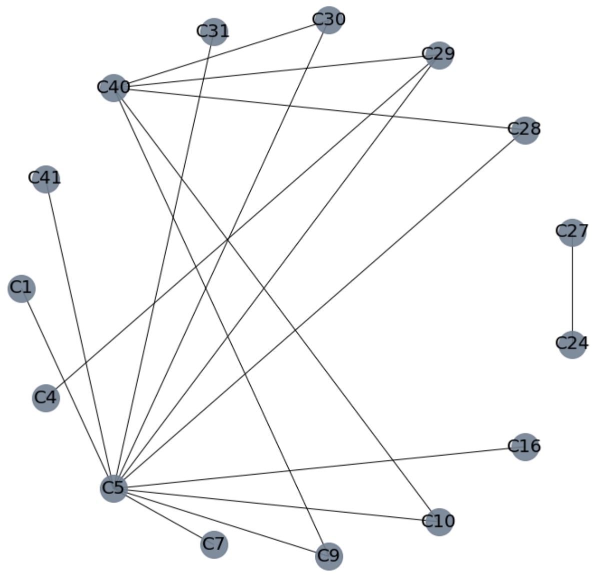


Figure 4.5: Negatively correlated cities

A city or region is said to be negatively correlated to another if the precipitation pattern of both cities differs. From the above figure 4.5 we discover that Antananarivo city (C40) and Antsirabe city (C5) which are both cities from Madagascar have precipitation which is negatively correlated to that of many cities in central and west Africa regions. To analyze why this happens, we put the cities on a geographical map to understand the climate characteristics of regions in which these cities are located.

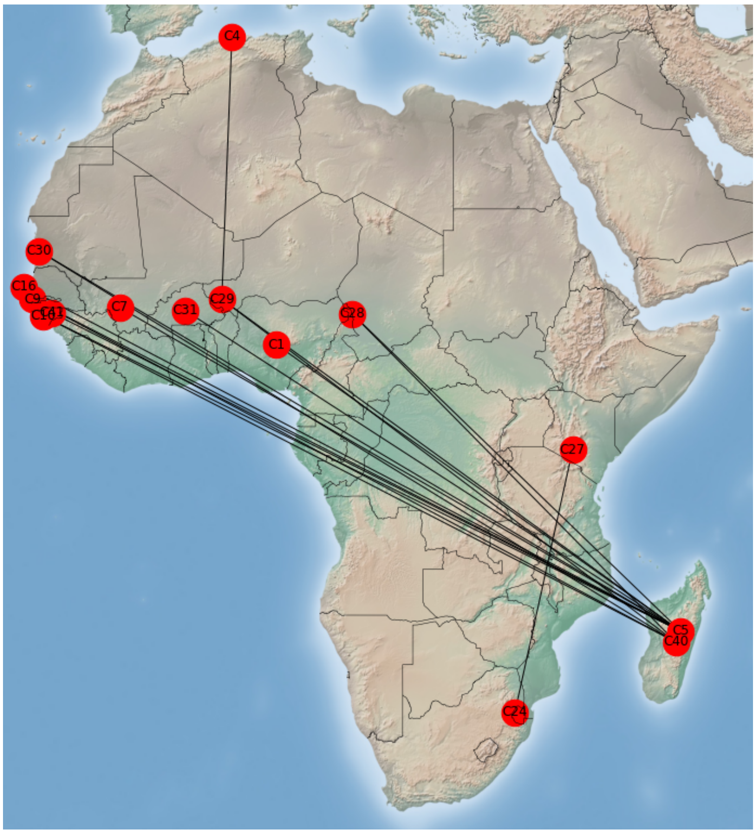


Figure 4.6: Negatively related cities

According to Moore (2013) Madagascar's weather is dependent on the southeastern trade winds and monsoon winds from the Indian Ocean. The east coast of Madagascar is directly exposed to trade winds, thus it receives mild rain throughout the whole year with the peak coming from December to March. This makes the country's climate slightly different from that of the central Africa region. Additionally, Madagascar risks experiencing many extreme events, with an estimated 3 to 4 tropical cyclones every year (USAID-CDC, 2021). A cyclone is a low-pressure center system of fast rotating storms and strong winds that originates from the ocean causing torrential rain in nearby regions. cyclones can have different names such as a hurricane, typhoon, or tropical cyclone based on where they originated from.

Tropical Cyclone (TC) that affects Madagascar originates from warm Indian ocean waters, it is formed when wet and warm air above the ocean rises above the ocean's surface. this leaves a low-pressure zone below which causes the air from high pressure to shift toward the low-pressure zone (Knutson et al., 2010). Contrary to cyclones that happen in a short period, monsoon is not a storm but rather a seasonal change in the direction of winds. this change causes torrential rain but also can cause droughts. The combination of tropical cyclones and monsoons over Madagascar sabotage its climate pattern's which end up making the country to be different from those in central and western Africa.

4.2.3 Community Analysis.

Although we have uncovered that some cities are related or not related, in such a network the structure of how vertices are connected carries much information. The fact that some cities can be tightly related in between their clusters and loosely related to other clusters, means much in the climate characteristics of each cluster. Using community detection we see that cities that are geographically near easily communicate between themselves, and communicate less with those that are distant.

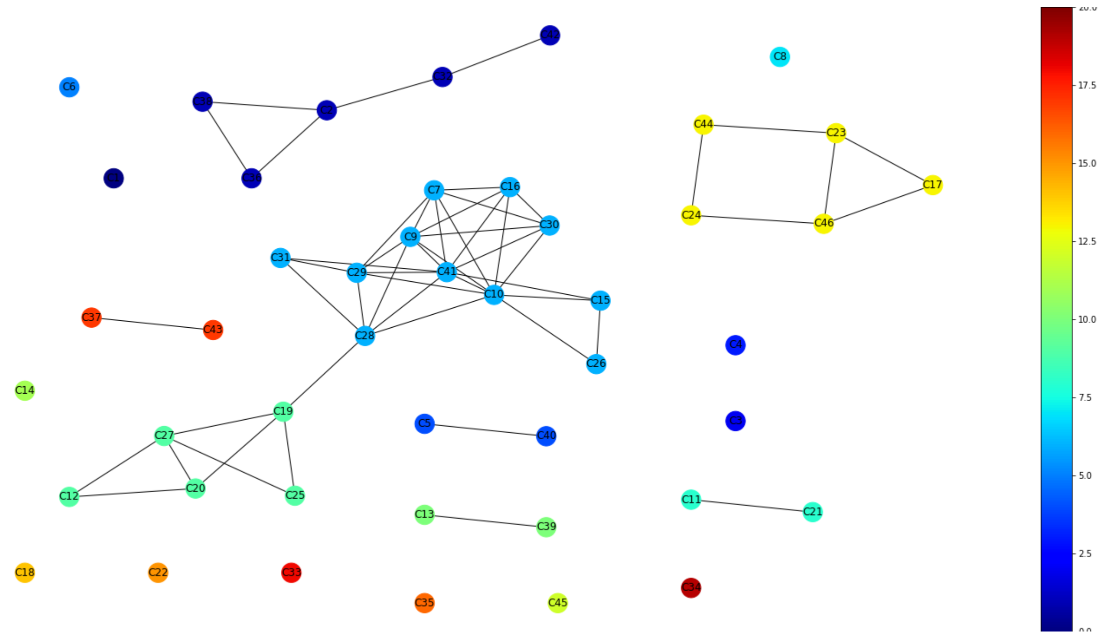


Figure 4.7: Clusters of cities

From the Figure 4.8 above we observe four major communities, these communities indicate easy and fast communication inside each cluster but also how clusters interact.

The first community consists of cities 18,26,29,11,24 which are Kampala city in Uganda, Nairobi city in Kenya, Kigali city in Rwanda, Bujumbura city in Burundi, and Mogadishu city in Somalia respectively. These cities are all in the East Africa region. this region is largely plateaued and has some highlands. As the climate of these countries is somehow similar they all fall in the same community and are tightly connected to each other, which means that any change in temperature or precipitation in one country, is highly likely to be felt by the neighbors also.

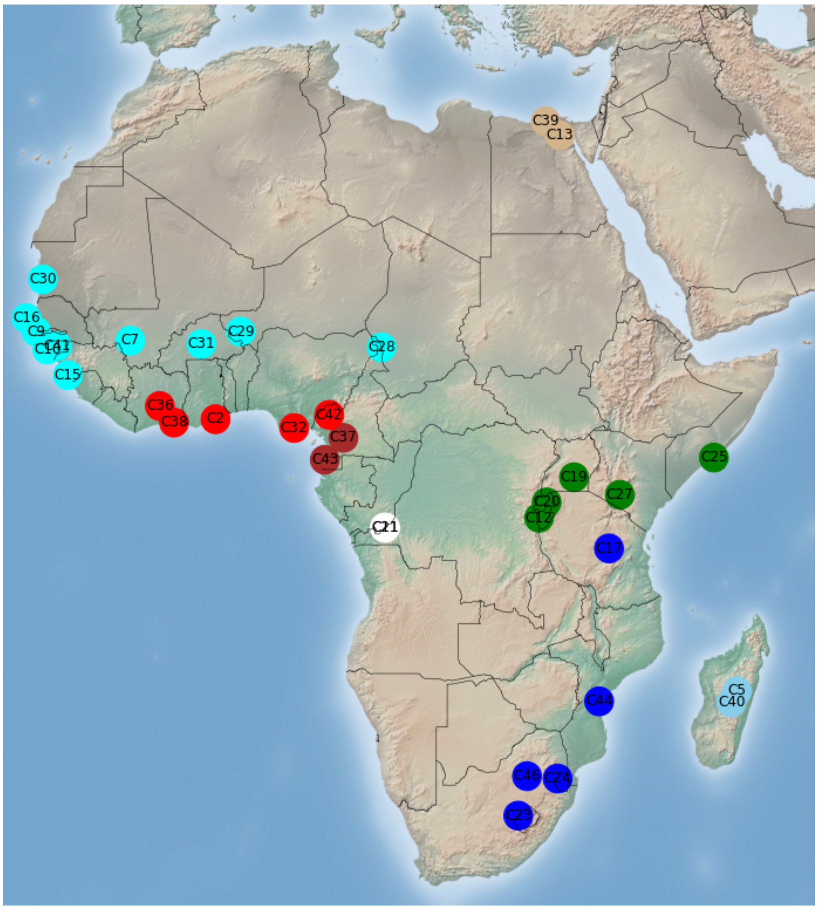


Figure 4.8: Geographical Clusters of cities

The second observed community is comprised of cities 41,31,1,37,35 which are Bamenda in Cameroon, Port Harcourt in Nigeria, Accra in Ghana, Abidjan, and Yamoussoukro both in Ivory coast respectively. Some of these cities are in the west Africa region while others are in the central, but all of these cities reside around the Gulf of Guinea.

The third observed community is located in central-west Africa and is comprised of cities 27,30,28,9,25,6 which are Ndjamena in chad, Ouagadougou in Burkina Faso, Niamey in Niger, Bissau in Guinea Bissau, Monrovia in Liberia, and Bamako in Mali. The fourth and last community has cities in the southern part of Africa those are city 43,22,23,45,16 which are Beira, in Mozambique Maseru in Lesotho, Mbabane in Swaziland, Benoni in South Africa, and Dodoma in Tanzania.

It is interesting to see that most of the communities are formed in longitudinal patterns or east to west. But again some intriguing facts are observed. Dodoma city (C17) in Tanzania, is in the same community as Southeast cities rather than being with east African cities that are geographically near. Referring to 4.4, this may be due to the fact that this city is below the equator in the southern hemisphere, and trade winds curve to the left in the southern hemisphere while curves right in the northern hemisphere (Richards et al., 2015), But also some cities like Kigali and Bujumbura are below the equator and not in the same community as Dodoma. So much deeper research is needed to conclude why Dodoma's climate is clustered with southern cities.

4.2.4 Global Network Analysis.

Although we have seen a network of some African cities, Africa is part of a much bigger network. The climate network is not restricted to Africa, which is why many of the climate phenomena happening in Africa are mainly influenced by another phenomenon somewhere in the ocean or another continent. Thus it is much important to understand the global climate as a network and analyze the driving force of this network and why some connections exist between distant geographical areas.

Using network theory techniques we build a network of 120 sample regions (nodes) that are randomly selected, and apply different measures on four different samples. The size of a node representing a region is defined based on how well each region scored in those measures. and this helps us in easily evaluating which regions are more important without going into complicated mathematical analysis. The first selected sample is analyzed using the Degree centrality measure, and from this measure, we see which regions have a high number of relations as it is seen in figure 4.9.

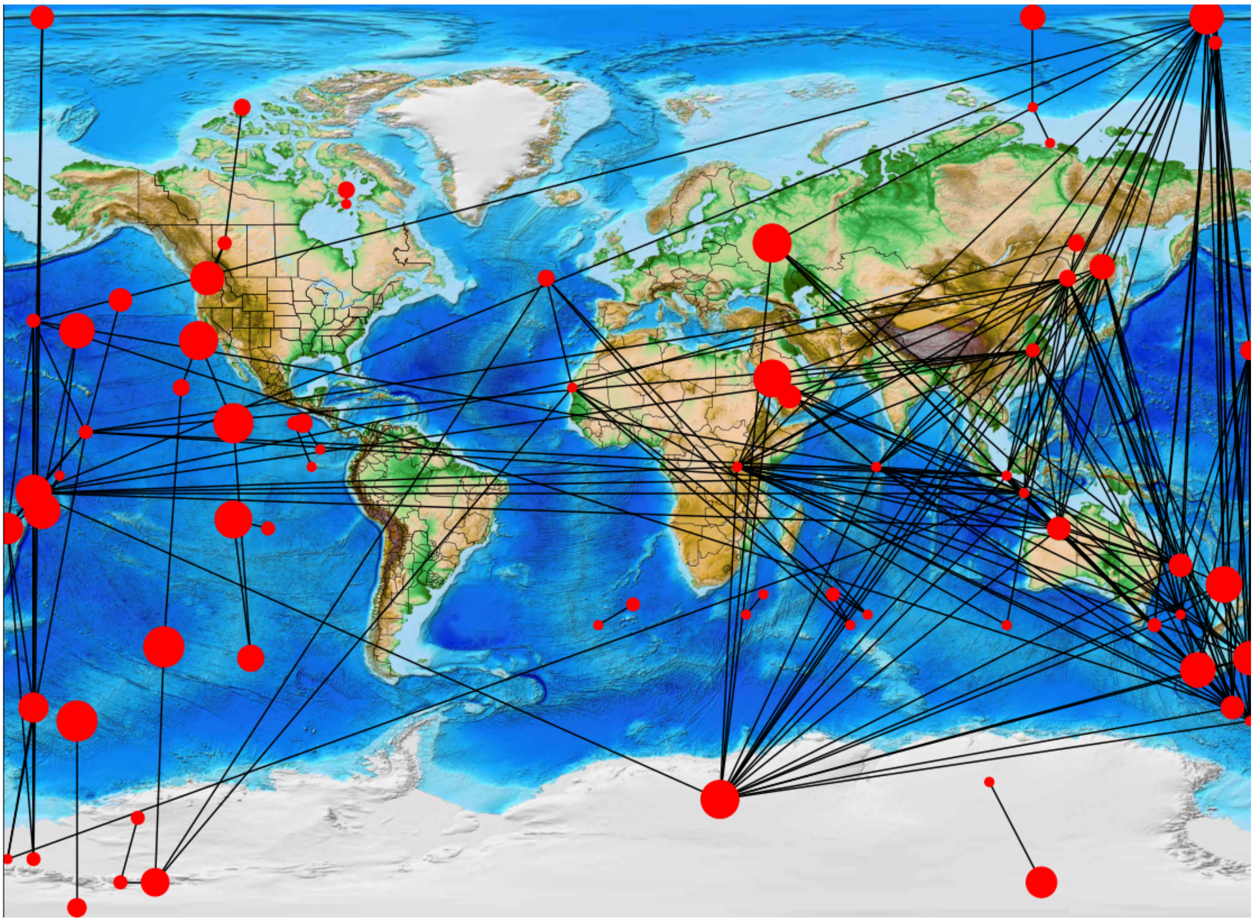


Figure 4.9: Regions with highest degree centrality

From the sample in the figure 4.9, we realize that many of the regions with the highest number of direct connections are geographically located in oceans, but particularly in the pacific ocean and few in the Indian ocean. This indicates that everything happening in the oceans can directly or indirectly affect the climate variability on the land surface all around the world. Also, we are able to see that many of the connections around the globe pass through or along the equatorial region which indicates how crucial this region is to the global climate. The oceans absorb a big percentage of solar radiation, particularly around the tropical region where solar radiations are perpendicular to earth and have high intensity. Oceans accumulate much heat but don't store it, it rather distributes this heat all around the globe, Through the process of evaporation. When ocean water molecules are heated, it results in the free movement of molecules. As this is a continuous process, it increases the humidity and temperature of sea surface air to form rain and sometimes storms that are transported above the land surface by trade winds. The tropical region is rainier because it receives much heat which results in higher evaporation and this is what we observe from figure 4.9, but generally, most of the rain we get on land surface is originated from oceans, which is in line with what Deza (2015) found.

To understand which regions facilitate the communication by serving as bridges in the climate network, we evaluate using the betweenness centrality measure as a measure of shortest paths in a network and the results are presented in figure 4.10 below.

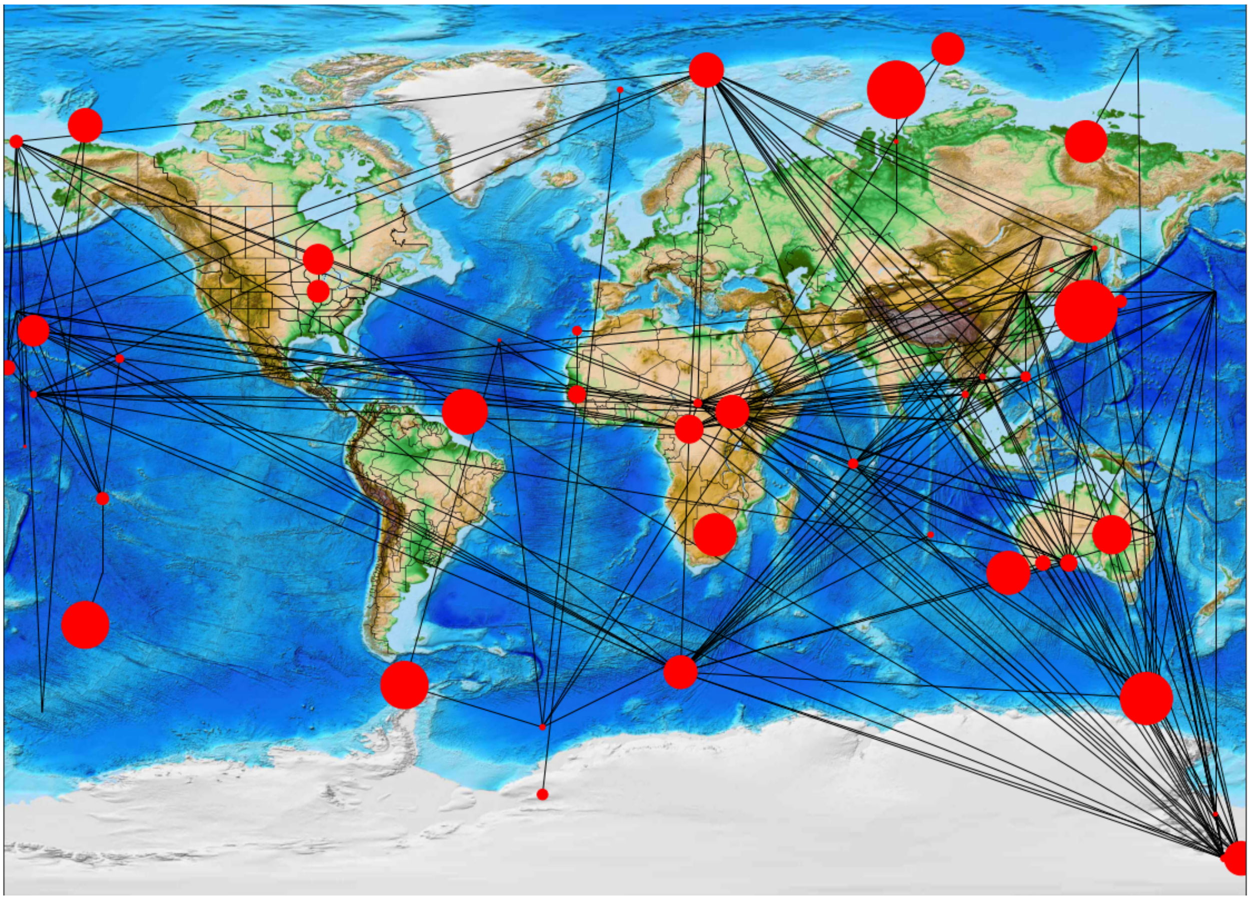


Figure 4.10: Regions with highest betweenness centrality

From figure 4.10, we see that much of the bridge regions are still in the oceans although this time they are not condensed in one region or ocean. This distribution shows that global climate communication is not limited to one place. It rather covers the whole world with some regions like oceans being very active in deciding the stability or variability of climate and the land surface experiencing those changes but also facilitating the communication. Even though the communication facilitator regions are everywhere in the world, from figure 4.10 we can see many connections passing through the equatorial region of Africa, South America, and South Asia. These regions are known for having very dense forests like the Amazon forest, The Congo forest, and the Borneo forest in southeast Asia. These forests also have a big role to play in the stability and communication of the global climate which is why many hot spots are geographically in the equatorial region.

Using closeness centrality measure we evaluate which regions are able to spread information efficiently, and from figure 4.11 we are able to see that globally all regions spread information efficiently, except for the regions in the north and south poles. Due to sun angle, both poles never get direct sunlight which results in always being cold, they also experience fewer daylight hours in winter. In such cold air, there is little to no evaporation capacity. No evaporation results in an increase of ice which act as a reflector of sunlight back to space.

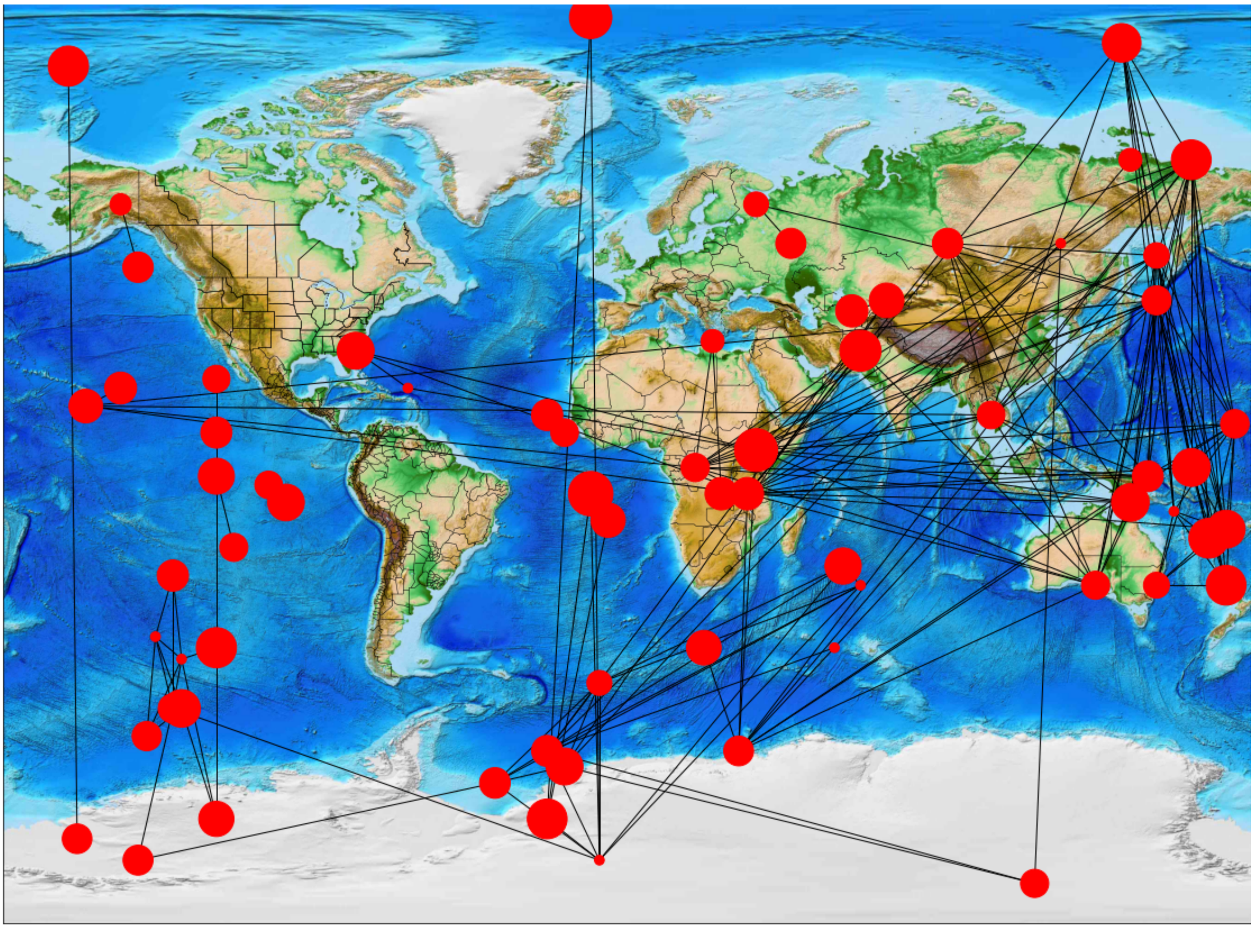


Figure 4.11: Regions with highest closeness centrality

Polar regions have a very important job to play in regulating the global climate. With its capacity to reflect much of the sunlight back to atmosphere, it balances the global temperature and if poles ever lose their capacity to reflect sunlight, the world will be facing much more global warming.

Using eigenvector centrality measure we discover that the most powerful regions are the ones in the east going west, the influences of nodes are minimal as we move west. this takes us back to Trade winds, which blow east to west. based on how powerful these trade winds are, they influence the climate very much, but as winds move far into the west they lose power and influence. which is why we are seeing the influential ones concentrated in the east. In South America and parts of Africa, Trade winds are interrupted or diverted by the high amount of rising air above landmasses. This is why we don't see many connections in south America when analyzing the figure 4.12.

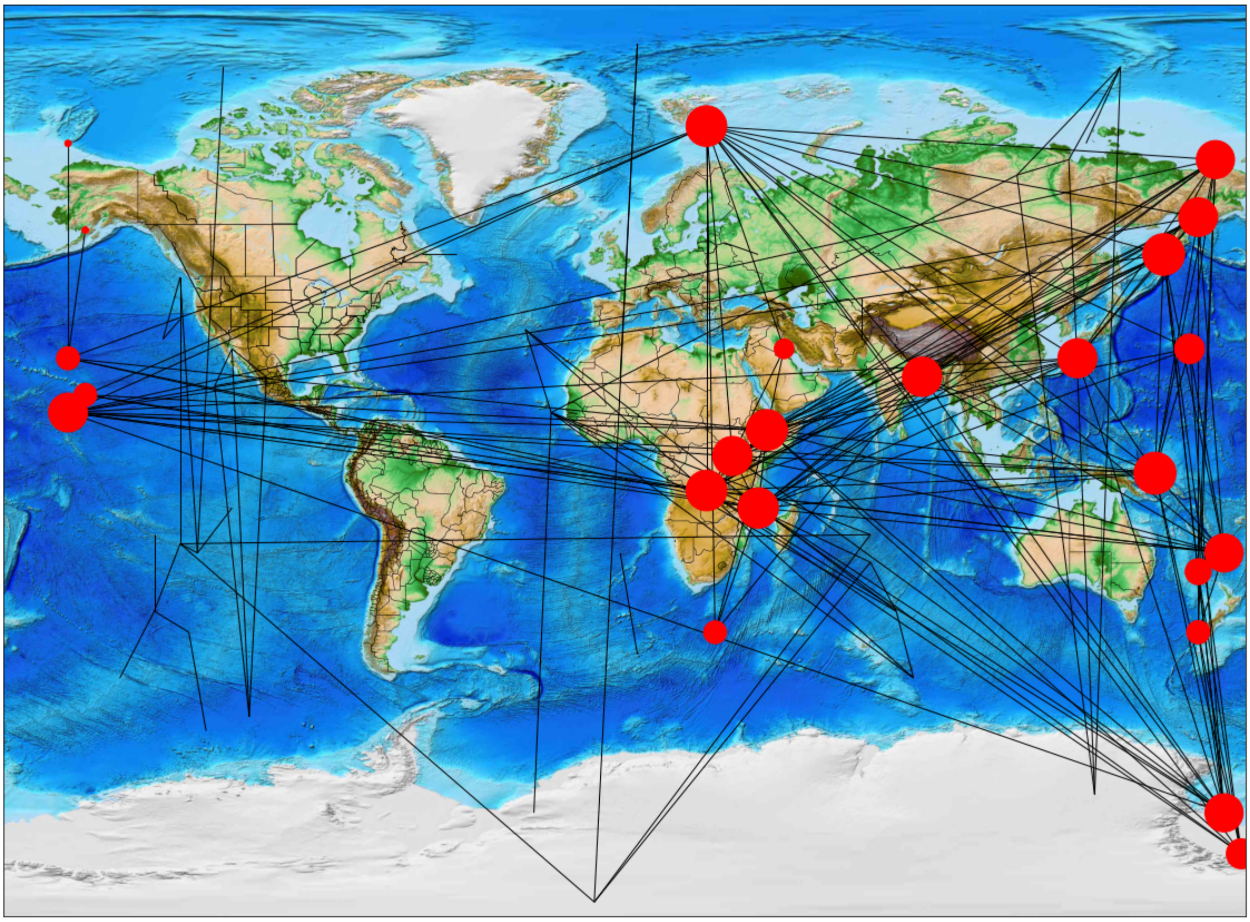


Figure 4.12: Regions with highest eigenvector centrality

The use of the above measures provided us with much insight into climate and what drives it. using only 1% of the data points, we were able to extract enough information and all the measures were reading to some phenomena that happen in the oceans such as trade winds and El Niño, and it explained how influential those phenomena are to the global climate network.

4.2.5 Global Climate Network Overview.

To acquire a broad picture of the global climate system, we raised the sample size to 500 nodes, which is 5% of the total number of nodes. This may seem like a little amount, but it provides a clear picture of the climate system in terms of precipitation. Looking at figure 4.13, we can see that the network is quite dense in the Pacific Ocean and its surrounding regions, and it gets less dense as we travel west to Australia, South Asia, and Africa. As one moves west from Africa, the majority of the nodes are found in the ocean rather than on land, this is attributed to the fact that there is many hills and tall buildings on land that exert friction forces to winds, thus halting efficient energy transfer above land surface. By combining all of these patterns, we can see that the worldwide network of precipitation is heavily reliant on the Pacific Ocean and other seas, as well as winds that transfer energy throughout the world.

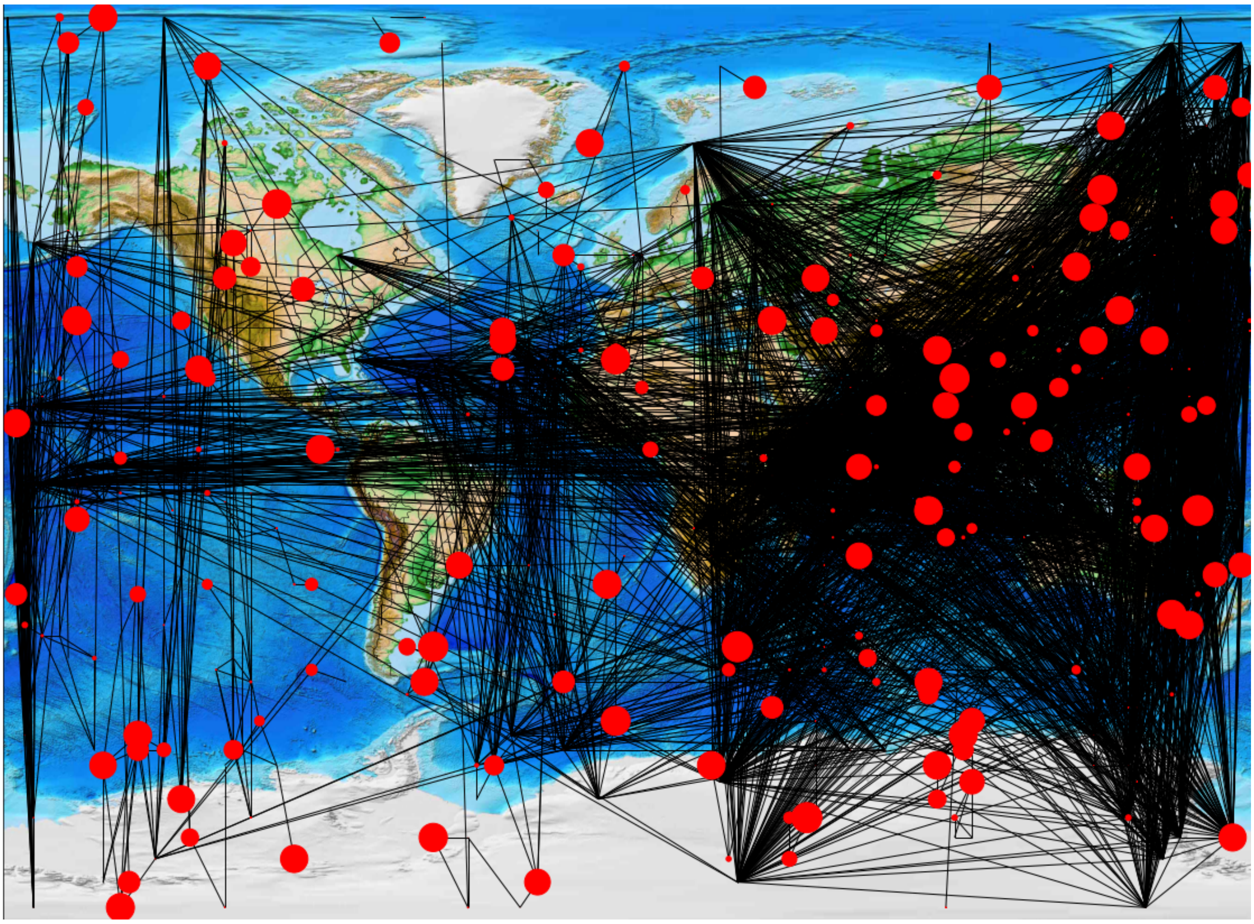


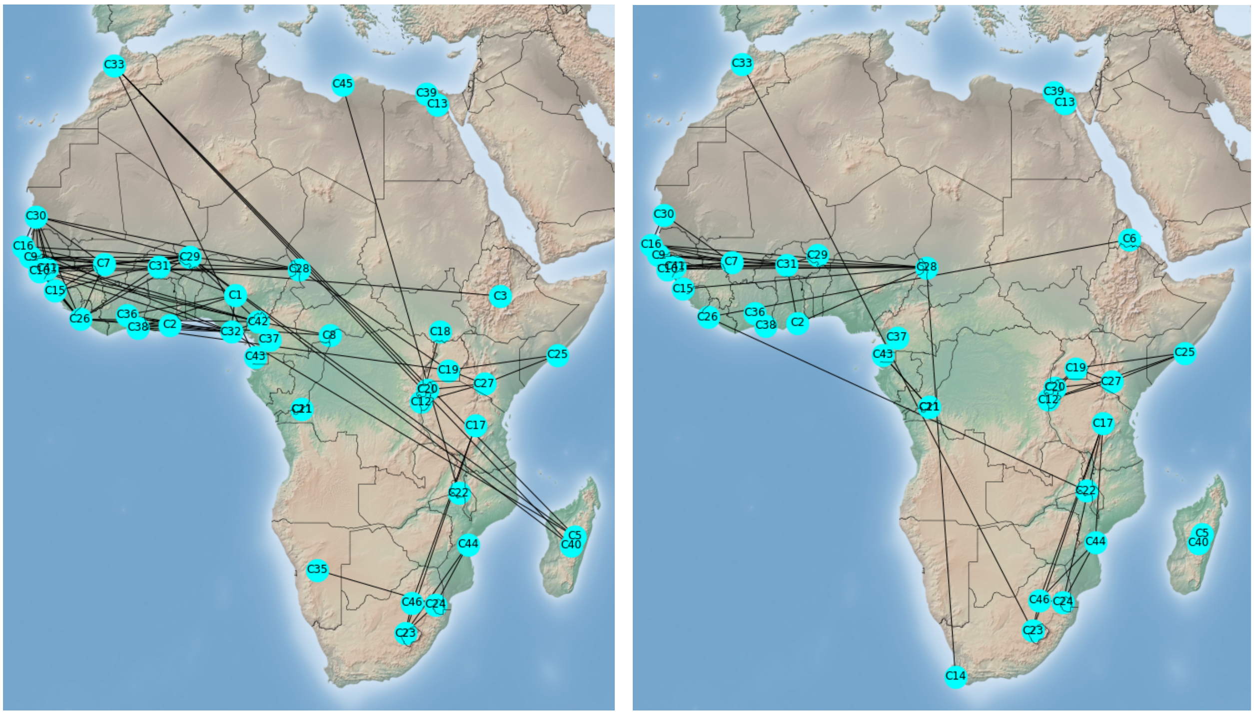
Figure 4.13: Global climate analysis

Constructing such networks provided us with an understanding of the global climate network and important phenomena that decide the behavior of the global climate network. Based on how powerful these phenomena are, a small perturbation on El Niño or Trade winds can easily affect the global climate through the paths explained above. Using network communities provides us with insight into the climate characteristics of particular regions, and a change in a network's community structures might indicate a change in the climate. Also, these phenomena play a big role in stabilizing the global climate as they distribute information such as heat and other energy around the globe. all these findings align with what Donges et al. (2009a) discovered, which calls for a deep analysis of these phenomena that serve as the backbones of the global climate network.

In his research titled " The backbone of the climate network ", Donges et al. (2009a) argues that trade winds can be considered as a cornerstone of global climate. This is due to the role these winds play in stabilizing the climate system when distributing hot and moist air, which when cools results in Precipitation. The increase in trade winds results in an increase and sometimes torrential rain. These winds can also be aggressive as they sometimes steer hurricanes towards land surfaces and by considering it's power to affect global climate, it is safe to say that any climate analysis that doesn't consider trade winds and other oceanic originated phenomena, is missing some important information.

4.2.6 Analysis of Climate Network Structure .

To explore patterns of climate change, we establish two networks with distinct time periods and analyze structure change to gain insight into how the climate structure has developed through time. Figure 4.14a shows that from 1960 to 1989, the Africa climate network was thick, with many links across the continent from east and south Africa to the west and certain portions of the north, passing via the central Africa region. However, as time went by, the structure of this network evolved. From 1990 to 2019, this network lost many of its links that ran through the central Africa region and strengthened its links between coastal regions beginning in east Africa, traveling through the south, and reaching west Africa as can be seen in figure 4.14b



(a) (b)

Figure 4.14: Network Structure of Africa in different time periods; 4.14a Africa climate network in years 1960 to 1989, 4.14b Africa climate network in years 1990 to 2019.

This change in structure may indicate changes in atmospheric circulation as the main driver of the climate network. As Ma et al. (2016) said, atmospheric circulation is a critical component of the Earth's climate system, influencing precipitation, moisture, and energy transmission. The tropical winds circulation is widely expected to weaken as a result of global warming driven by an increase in greenhouse gas concentration. As trade winds weaken, their power to cross above a continent will also be decreased since there are some friction forces above a hilly continent like Africa, and this might be the reason we are observing more connections around continent between coastal regions and lose connections in the central Africa region in the second figure 4.14b, Although we can't conclude that this structure change is entirely a result of weakened trade winds. Much deeper research on the relationship between global warming and atmospheric winds circulation is needed to conclude if the former affects the latter.

Interestingly Madagascar had a similar climate to some countries in west and north Africa in the years 1960 to 1989 but since 1990 this uniformity decreased and from the network structure in figure 4.14b, we can see that the climate characteristics of Madagascar look unique compared to different other parts of Africa.

5. Conclusion and Recommendations

5.1 Conclusion

Continuing researches with new techniques are still needed to understand and analyze the climate system. In this study, Network Theory was applied to two different datasets of precipitation data, The goal was to analyze climate, specifically, precipitation characteristics of different regions on a local and global scale, and results obtained are promising as they align with findings of other researchers like, Donges et al. (2009a), Steinhaeuser et al. (2009) and Deza (2015).

In Africa, the precipitation patterns are well observed around the equator and this is attributed to trade winds that are concentrated in the equatorial region and blow east to west. On a global scale, results showed that most of the precipitation experienced on land surface is originated or is much related to what happens in oceans such as the Pacific and Indian oceans. The climate network structure showed signs of changing over years and this continuous changes was attributed to the effects of global warming.

Despite some constraints faced, we were able to obtain conclusive results, and even if those results are not entirely new discoveries, they prove that network theory may be dependable in the field of climate science if much deeper researches are done. This rather serves as a foundation for those who may want to use this method is much more advanced analysis such as climate change and climate modeling.

5.2 Recommendations

In the future, We recommend using many network measures such as mutual information, assortativity, communicability, and others so that they can extract much more insight. To thoroughly analyze the climate system, we recommend the use of multi-layer networks in order to analyze multiple climate variables such as temperature, humidity, precipitation, wind speed, and others, since the changes in one variable may lead to a bigger change in another variable, it is therefore important to analyze many variables collectively.

Researchers are encouraged to use multiple networks of different time-spans to be able to analyze network structure change over time, which may indicate climate change patterns. We also would like to recommend the exploration of ways of using network theory in climate modeling as this is yet to be done.

ACKNOWLEDGEMENTS

I am very grateful to The Almighty God for strengthening me, providing me with the knowledge, capacity, and chance to embark on this research project and successfully complete it. I would also want to express my heartfelt gratitude to my supervisor, Dr. Philip Knight, for his great inspiration, patience, guidance, and relentless efforts during the research and preparation of my thesis. I also want to thank my fellow students as well as the whole AIMS–Rwanda community for their assistance in making this research a success. Thank you very much.

References

Coriolis Effects. https://www.nationalgeographic.org/encyclopedia/coriolis-effect/. Accessed: 2021-05-22.

Flight Environment, prevailing winds. https://www.weather.gov/source/zhu/ZHU_Training_Page/winds/Wx_Terms/Flight_Environment.htm. Accessed: 2021-05-22.

What Are Trade Winds? https://scijinks.gov/trade-winds/. Accessed: 2021-05-22.

Adler, R. F., M. R. Sapiano, G. J. Huffman, J.-J. Wang, G. Gu, D. Bolvin, L. Chiu, U. Schneider, A. Becker, E. Nelkin, et al. (2018). The global precipitation climatology project (gpcp) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere 9*(4), 138.

Barnes, J. A. (1969). Graph theory and social networks: A technical comment on connectedness and connectivity. *Sociology 3*(2), 215–232.

Bloch, F., M. O. Jackson, and P. Tebaldi (2019). Centrality measures in networks. *Available at SSRN 2749124*.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment 2008*(10), P10008.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology 92*(5), 1170–1182.

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

De Meo, P., E. Ferrara, G. Fiumara, and A. Provetti (2011). Generalized louvain method for community detection in large networks. In *2011 11th international conference on intelligent systems design and applications*, pp. 88–93. IEEE.

Deza, J. I. (2015). Climate networks constructed by using information-theoretic measures and ordinal time-series analysis.

Dijkstra, H. A., E. Hernández-García, C. Masoller, and M. Barreiro (2019). *Networks in Climate*. Cambridge University Press.

Donges, J. F., Y. Zou, N. Marwan, and J. Kurths (2009a). The backbone of the climate network. *EPL (Europhysics Letters) 87*(4), 48007.

Donges, J. F., Y. Zou, N. Marwan, and J. Kurths (2009b). Complex networks in climate dynamics. *The European Physical Journal Special Topics 174*(1), 157–179.

Estrada, E. (2012). *The structure of complex networks: theory and applications*. Oxford University Press.

Estrada, E. and P. A. Knight (2015). *A first course in network theory*. Oxford University Press, USA.

Fortunato, S. and D. Hric (2016). Community detection in networks: A user guide. *Physics reports 659*, 1–44.

Gozolchiani, A., S. Havlin, and K. Yamasaki (2011). Emergence of el niño as an autonomous component in the climate network. *Physical review letters 107*(14), 148501.

Gross, J. L. and J. Yellen (2005). *Graph theory and its applications*. CRC press.

Gupta, N., A. Singh, and H. Cherifi (2016). Centrality measures for networks with community structure. *Physica A: Statistical Mechanics and its Applications 452*, 46–59.

Hagberg, A., D. Schult, and P. Swart (2018). Networkx reference release 2.2. *Tech. rep.*.

Harris, I., T. J. Osborn, P. Jones, and D. Lister (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific data 7*(1), 1–18.

Khan, B. S. and M. A. Niazi (2017). Network community detection: A review and visual survey. *arXiv preprint arXiv:1708.00977*.

Knutson, T. R., J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. Srivastava, and M. Sugi (2010). Tropical cyclones and climate change. *Nature geoscience 3*(3), 157–163.

Likaj, R., A. Shala, M. Mehmetaj, P. Hyseni, and X. Bajrami (2013). Application of graph theory to find optimal paths for the transportation problem. *IFAC Proceedings Volumes 46*(8), 235–240.

Linhares, C. D., J. R. Ponciano, F. S. Pereira, L. E. Rocha, J. G. S. Paiva, and B. A. Travençolo (2020). Visual analysis for evaluation of community detection algorithms. *Multimedia Tools and Applications*, 1–23.

Ma, E. Network analysis made simple — scipy 2019 tutorial. https://www.youtube.com/watch?v=ED4NZ-4EWRw&list=RDCMUCkhm72fuzkS9fYGlGpEmj7A&index=23. Accessed 2021-05-11.

Ma, J., G. R. Foltz, B. J. Soden, G. Huang, J. He, and C. Dong (2016). Will surface winds weaken in response to global warming? *Environmental Research Letters 11*(12), 124012.

Malviya, D. K. and N. Malviya (2018). Survey on implementation of graph theory in routing protocols of wired computer network.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon (2002). Network motifs: simple building blocks of complex networks. *Science 298*(5594), 824–827.

Moore, G. (2013). Impact of the high topography of madagascar on the structure of the findlater jet. *Geophysical Research Letters 40*(10), 2367–2372.

Otte, E. and R. Rousseau (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science 28*(6), 441–453.

Que, X., F. Checconi, F. Petrini, and J. A. Gunnels (2015). Scalable community detection with the louvain algorithm. In *2015 IEEE International Parallel and Distributed Processing Symposium*, pp. 28–37. IEEE.

Richards, R., A. Emiliano, and R. Tejeda (2015). The effects of the trade winds on the distribution of relative humidity and the diurnal air temperature cycle on oceanic tropical islands. *J Climatol Weather Forecasting 3*(137), 2.

Scott, J. (1988). Social network analysis. *Sociology 22*(1), 109–127.

Spyrou, L. and J. Escudero (2018). Weighted network estimation by the use of topological graph metrics. *IEEE Transactions on Network Science and Engineering 6*(3), 576–586.

Steinhaeuser, K., N. V. Chawla, and A. R. Ganguly (2009). An exploration of climate data using complex networks. In *Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data*, pp. 23–31.

Steinhaeuser, K., A. R. Ganguly, and N. V. Chawla (2012). Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Climate dynamics 39*(3-4), 889–895.

Tsonis, A. A. and P. J. Roebber (2004). The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications 333*, 497–504.

Tsonis, A. A., K. L. Swanson, and P. J. Roebber (2006). What do networks have to do with climate? *Bulletin of the American Meteorological Society 87*(5), 585–596.

USAID-CDC (2021). Annex f: Climate change analysis.

Zafarani, R., M. A. Abbasi, and H. Liu (2014). *Social media mining: an introduction*. Cambridge University Press.