AFRICAN CENTER OF EXCELLENCE IN DATA SCIENCE

Social Ties and Public Health:

Use of Social Network Analysis to Control the Spread of Infectious Diseases in Africa

By

Steven Karera
Registration Number: 220000260

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Data Science in Biostatistics

University of Rwanda, College of Business and Economics

Supervisor: Dr. Eric Umuhoza

September 2022

Declaration

I declare that this dissertation entitled "**Social Ties and Public Health: Use of Social Network Analysis to Control the Spread of Infectious Diseases in Africa"** is the result of my own work and has not been submitted for any other degree at the University of Rwanda or any other institution.

Names
Steven Karera

Signature

Approval sheet

This dissertation entitled "**Social Ties and Public Health:** Use of Social Network Analysis to Control the Spread of Infectious Diseases in Africa**"** was written and submitted by **Steven Karera** in partial fulfillment of the requirements for the degree of Master of Science in Data Science majoring in **Biostatistics** is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 14 % which is less than 20% accepted by the African Centre of Excellence in Data Science (ACE-DS).

____
Supervisor

____
Head of Training

Abstract

Infectious diseases continue to be one of the biggest global public health concerns, despite major progress in microbiological research. For the past decades, more than 50 emerging or re-emerging infectious diseases have occurred all over the world. In Africa each year, almost half of the countries experience an emerging or re-emerging infectious disease. Researchers have been using Social Network Analysis techniques to model the spread of infectious diseases in veterinary medicine, and now various researchers are concentrating on how the overall state and structure of the spread of infection are affected by social networks formed by humans. A series of recent studies have indicated that many networks are revealed to obey the power-law degree distribution. This study aims at leveraging an anonymized snapshot of all active Facebook users and their friendship networks to measure the intensity of connectedness between locations (countries) in Africa, to better understand the structure of the spread of potentially infectious diseases and the effect of social ties on death rate due to the spread of communicable diseases.

This study adopts a quantitative analytical approach by applying three main methods of analysis: social network analysis, hypothesis testing on the distributions of network data, and assessing the effect of controlling crucial nodes in the community by fitting a simple one-parameter regression function (dependent variable is the proportion of death due to infectious diseases). The resulting Social Connectedness Index (SCI) network data in Africa is composed of 50 nodes and 2,450 edges. The estimated network parameters were: proportion of all potential connections between vertices (6%), 3.98 average length of the shortest paths between all combinations of vertices in the network, and the longest path length between any pair of vertices was 5. Results from the initial estimation revealed that the social network structure formed by the Facebook social connectedness index follows power-law distribution and is a scale-free network subset of small-world networks. Our results cast a light on how countries' social connectedness might affect the spread of infectious diseases and subsequently increase the death rate due to these diseases.

This aspect of the research only covered a basic set of tools for drawing inferences from complex network data. Therefore, future research should be conducted in more realistic network models for how networks change over time and affect the spread of infectious diseases.

“Social”, “Networks”, “Infectious Diseases”, “Spread”, “Network Data”, “Connectedness”, “Africa”

Dedication

I dedicate this thesis to my late parents, Athanase Karera and Jeanne Nyiramahingura who instilled in me the power of perseverance and commitment. A special appreciation to my Sister, Grace Duhuze Karera for pushing against all odds to ensure I excelled in my studies.

Acknowledgments

I would like to express my deep gratitude to the following people for their unwavering and timely support through the last 2 years of this study:

Dr. Eric Umuhoza for agreeing to supervise me on this journey. He demystified the process and offered me an opportunity of a lifetime to acquire unique skills through his timely, focused, and thoughtful guidance. I will miss our supervision meetings. I am eternally grateful.

Special thanks to my sister and friend, Dr. Marie Grace Duhuze Karera, for her encouragement and inspiration. Without your support and reassurance, this would have stayed a dream.

After all, I would not have been able to accomplish this feat without my fiance's constant word of encouragement through the low and high points. Her unwavering support helped keep the end in sight.

Table of Content

List of Tables

List of Figures

List of Acronyms & Abbreviations

AAAI*: Association for the Advancement of Artificial Intelligence*
AIDS*: Acquired immunodeficiency syndrome*
CDC: Centers for Disease Control and Prevention
COVID-19*: Coronavirus Disease of 2019*
DALYs*: Disability-adjusted life years*
DF*: Degrees of freedom*
DRC*: Democratic Republic of the Congo*
EID*: Emerging or Re-emerging Infectious Disease*
GBD*: Global Burden of Disease*
HBV*: Hepatitis B Virus*
HIV*: Human Immunodeficiency Virus*
KS*: Kolmogorov Smirnoff*
MVD*: Marburg Virus Disease*
RQ*: Research Question*
SARS*: Severe acute respiratory syndrome*
SCI*: Social Connectedness Index*
SNA*: Social Network Analysis*
USA*: United States of America*
WHO/AFRO*: World Health Organization Regional Office for Africa*
YLL*: Years of life lost*

1. Introduction

1.1. Background

Communicable diseases affect people worldwide and result from pathogens, including bacteria, viruses, and parasites. For humans, the spread of pathogens occurs in a variety of ways, such as: spread directly or indirectly from one person to another, waterborne or foodborne diseases, or aerosolization of infected particles in the environment and by insects and ticks (Taylor, 2011). A variety of disease-causing bacteria and viruses are transmitted in the mouth, nose, throat, and airways, and symptoms range from mild to severe and treatment depends on the cause of the disease.

Despite significant advances in microbiological research, infectious diseases persist to be one of the biggest global public health concerns. Infectious diseases account for approximately one-quarter of the more than 52 million people who die each year worldwide (Khabbaz et al., 2014). The Global Burden of Disease (GBD) survey estimates that over 31% of years of life lost (YLL) and years lived with disability (DALYs) are lost to premature death worldwide due to communicable diseases. The vast majority of these occurred in less developed countries, where 35% of DALYs were lost to infectious and parasitic diseases (Birmingham & Stein, 2003).

Over the past 50 years, more than 50 emerging or re-emerging infectious diseases (EIDs) have emerged around the world. In Africa, an EID is conducted annually in almost half of the countries (Torti et al., 2020). The following table shows Africa's status for infectious diseases that are presented by the Africa Centres for Disease Control and Prevention.

Table 1.1: Infectious diseases from Africa CDC and most affected areas.

Disease	**Most affected areas**
Anthrax	Most countries of sub-Saharan Africa and Asia, several southern European countries, the Americas, and certain areas of Australia
Avian Influenza	Africa, Asia, Europe and the Middle East
Chikungunya	**Recent Outbreaks in Africa:** Congo, Kenya, Senegal and Sudan
Cholera	**Recent Outbreaks in Africa:** Ethiopia, Cameroon, Tanzania, Angola, Algeria, Mozambique, DRC, Somalia, Kenya, Zambia, South Sudan, Sierra Leone, Republic of Congo, Chad, Niger, Nigeria, Zimbabwe, and Guinea-Bissau
COVID-19	Worldwide pandemic
Crimean-Congo Haemorrhagic Fever	**Recent Outbreaks in Africa:** Uganda, Senegal, South Africa, and Mauritania
Dengue Fever	**Recent Outbreaks in Africa:** Burkina Faso, Côte d'Ivoire, Egypt and Cape Verde
Ebola Virus Disease	West Africa (Sierra Leone, Liberia & Guinea) and Democratic Republic of Congo
Hepatitis B Virus (HBV)	6.1% of the population of the WHO African region and 2% of the population of the WHO Eastern Mediterranean region (Which includes some of the north African Countries)
Hepatitis C Virus	2.3% of the WHO Eastern Mediterranean region (Which includes some northern African countries).
Hepatitis E Virus	**Recent Outbreaks in Africa:** Namibia, Nigeria, Niger, and Chad
HIV (Human	The WHO African region is the most affected region with 25.7

Immunodeficiency Virus)	million people living with HIV in 2017
Lassa Fever	**Recent Outbreaks in Africa:** Benin, Ghana, Guinea, Liberia, Sierra Leone, Togo, Burkina Faso, Mali, and Nigeria.
Malaria	The WHO AFRO Region represented 90% of cases and 91% of deaths
Marburg Virus Disease (MVD)	**Recent Outbreaks in Africa:** Angola, Democratic Republic of Congo, Kenya, and South Africa
Measles	**Recent Outbreaks in Africa:** DRC, Nigeria, Zambia, and Ethiopia
Meningococcal Meningitis	It is seen worldwide, but the greatest burden is found in sub-Saharan Africa's meningitis belt, which stretches from western Senegal to eastern Ethiopia
Monkeypox	**Recent Outbreaks in Africa:** Cameroon, Nigeria, and Central Africa Republic
Plague	**Recent Outbreaks in Africa:** The Democratic Republic of the Congo, and Madagascar.
Poliomyelitis (Polio)	**Recent Outbreaks in Africa:** Somalia, DRC, Nigeria, Madagascar, Cameroon, Equatorial Guinea, South Sudan, Somalia, Kenya, Ethiopia, Congo, and Angola
Rift Valley Fever	**Recent Outbreaks in Africa:** Kenya, Uganda, Niger, Mauritania, South Africa, Mauritania, and Madagascar
Tuberculosis	Worldwide
Yellow Fever	Thirty-four African countries are endemic to yellow fever.
Zika Virus	**Recent Outbreaks in Africa:** Cape Verde

According to the World Health Organization, the principles of best practice in infection control are based on extensive research and should be adopted to help prevent avoidable infections and to control existing ones (World Health Organization. Regional Office for Europe, 2021). This current study aims to strengthen health systems; to better inform the general public, health organizations, the research community, and government bodies on how to control the burden of disease outbreaks, and disease preparedness in a country. This will be achieved through an analysis of social networks to understand the death rate due to the spread of infectious diseases using the social connectedness index which measures the relative probability that two individuals across two locations are friends with each other on Facebook (Bailey et al., 2018).

1.2. Problem Statement

So far, the successful implementation of social network analysis techniques in the healthcare system, uses traditional data (i.e; data collected through questionnaires, interviews, or electronic patient records from central public authorities) to explain the network activities (Glass & Glass, 2008; Zelner et al., 2012). Current trends in epidemiology do not leverage the opportunities offered by the variety of data generated at diverse speeds, from several sources, and in very large volumes, like a measure of social connectedness derived from the use of Facebook (Grieve et al., 2013).

Most early studies as well as current work focus on determining and understanding the overall structure and effects on the spread of infection. The emphasis is on how it is affected by social networks formed by online communities to contribute to better control of the spread of infectious diseases.

1.3. Research Objective

This study aims at leveraging social connectedness between countries in Africa, and to better understand the effect of social ties on death due to the spread of infectious diseases.

1.4. Research Questions

The main research question this study seeks to answer can be phrased as follow:

How can social network analysis techniques be used to curb the spread of communicable diseases in Africa?

To better respond to the main research question the following specific sub-questions are proposed:

1. RQ1: How effective is the social connectedness index at describing the social interaction structure of African countries?
2. RQ2: What are the key nodes playing a central role in the entire network? In other words, what are the countries that quickly connect to other countries in the network?
3. RQ3: What effect do social ties have on death due to the spread of communicable diseases in Africa?

1.5. The rationale for the study

In this study, we investigate the impact of social connectedness on death due to the spread of infectious diseases using the Social Connectedness Index (SCI) as a proxy for real-world friendship networks. The findings of our study might be useful in strengthening health systems; better informing the general public, health organizations, the research community, and countries' disease preparedness.

2. Related Work

Barnes first introduced the concept of social network analysis (SNA) in the early 1919s in an anthropology article that emphasized the social aspect of relationships between nodes in a network (Barnes & Harary, 1983). Klovdahl defined a network as a cluster of nodes connected by different kinds of links (Klovdahl, 1985). The author assessed whether network data and network analysis can help assess an infectious disease outbreak and how an understanding of social network structure could help design appropriate containment strategies. The nodes can be communities, individuals, or other institutions; the connections can be asymmetrical or symmetrical and the structure of a network has consequences for its members and the network as a whole.

A social network as a class of nodes connected by social relationships, Priss and Nery built on this in 2007 and proposed that the concept of the neighborhood used in social network analysis is graph theory, where neighboring points (nodes) are connected by a line (links) are all the points that a point has connected to form a neighborhood (Priss & Nery, 2007). Most studies have illustrated the idea that SNA is an analytical tool that allows one to study, characterize, and quantify the nature and extent of contacts between a group of nodes (Khabbaz et al., 2014; Stattner & Vidot, 2011). The literature has also shown that network analysis techniques provide a framework to analyze, understand, and control the spread of disease through the interactions between subjects or the environment.

Researchers have used SNA techniques for years and have demonstrated the need for studies to be conducted. As previously reported in Italy, researchers built a network from the movement data to study the movement patterns of cattle in Italy and the relationship between premises concerning the potential spread of bovine disease (Natale et al., 2009). Snchez-Matamoros provided a complete description and characterization of a complete and reliable network to better understand the potential risks related to equine movements in Spain (Sánchez-Matamoros et al., 2012).

In contrast to these studies, which used SNA techniques to model the widespread diseases (contagious) in veterinary medicine, various researchers are now focusing on how the overall state and structure of the spread of infections are influenced by social networks formed by humans.

In 2011, Erick Stattner presented the perspectives offered by social network techniques in epidemiology, arguing that only interactions between individuals or the environment enable disease transmission through social contacts (Stattner & Vidot, 2011). In 2008, Laura and her partner characterized the social contact network for potential transmission of influenza in children and adolescents (Glass & Glass, 2008). A recent study shows that neglecting individual interactions when studying the spread of an epidemic can lead to a dramatic underestimation of the severity of infection (Zino et al., 2018).

Table 1.2: Modeling the effect of the spread of infectious with and without social interaction

Modeling without social interaction variable	Modeling with social interaction variable
Dramatic underestimation of the severity of an infection.	Identifying vulnerable locations/communities and implementing contamination strategies to reduce the spread of this disease.
Public health is studied only through surveys and by aggregating metric estimates from healthcare practitioners. Which could be biased.	Estimates the features of the spread of the disease on a wide scale without the active participation of users
Inability to quantitatively compare rates of spontaneous and contagious infection, preventing the ability to characterize the relative importance of social transmission (Hill et al., 2010).	Reduces response time and improves overall productivity in public health surveillance. For example, Google Flu Trends, which models flu prevalence by analyzing geolocated search queries (Ginsberg et al., 2009)
	Ability to model and predict the occurrence of universal epidemics from everyday human interactions (Sadilek et al., 2012).

In recent years, the power-law degree distribution has been found to be followed by many networks (Tanimoto, 2009b), and the basic model for undirected networks provides an explanation for the fact that real-world network degree distributions tend to obey power-laws with exponents between

2 and 3 such networks as scaling-free (Tanimoto, 2009a). All of the literature also demonstrated a vector-borne disease that can only be spread between vectors and hosts and can therefore be modeled on a scale-free bipartite network(Bisanzio et al., 2010). The validity of this theory was confirmed by numerical simulations by (Gómez-Gardeñes et al., 2008), who had previously examined the spread of disease on scale-free bipartite graphs as a model for STDs in heterosexual populations.

3. Materials and Methods

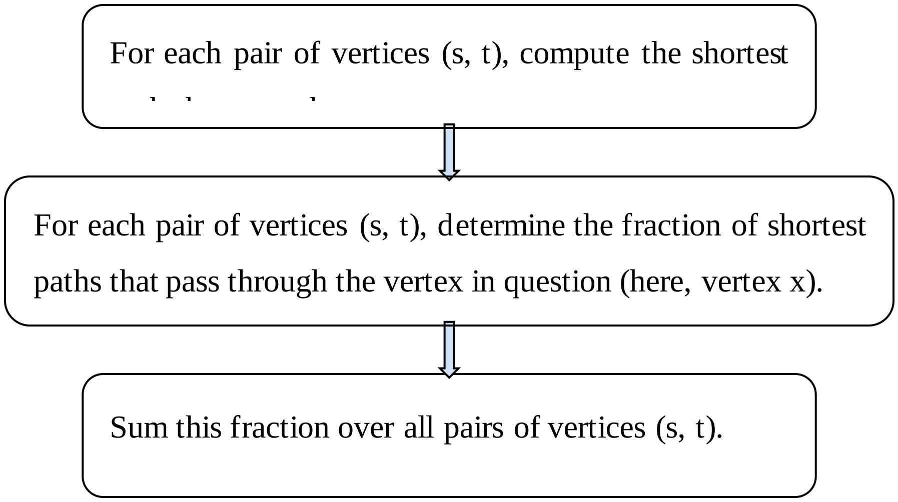
This section presents the approach used in my study. The section is further divided into four subsections: in 3.1 we define important concepts, in 3.2 we describe the data source, in 3.3 we present ethical considerations and 3.4 shows an adopted analytical approach.

3.1. Important definitions

Before jumping to the methods used in this study, we define the basics and key terms related to SNA used in this work:

- ***Nodes***: The entities (i.e; countries) to be linked in the network. Synonyms: vertices of a network.
- ***Edges***: The relationships/interactions(i.e; Facebook social connectedness) between the nodes. Synonyms: ties
- ***Adjacency matrix*****:** A square matrix where the column and row names are the vertices of the chart. This data format is accepted by various network analysis packages in R.
- ***Edge list:*** A block of data that contains at least two columns: a column of nodes that correspond to the source of a connection, and another column of nodes that contains the destination of the connection.
- ***Weighted network graph:*** An additional column in an edge list that describes the attributes of the edges. If the edges have a size attribute, the graph is considered weighted.
- ***Degree centrality:*** A local centrality measure that counts the number of links held by each node and indicates people who can quickly connect to the broader network.
- ***Closeness centrality:*** A global centrality measure that takes into account the entire network. This measure evaluates each node based on its proximity to all other nodes within the network. It measures the shortest paths between all nodes and then allocates each node a score based on its sum of shortest paths, additionally, it is useful for finding the nodes best positioned to influence the entire network the fastest.
- ***Betweenness centrality:*** A popular estimate that captures a node's role by allowing information to be communicated across all parts of the network.

The betweenness of a node x in a graph $H := (X, E)$ with X nodes is calculated as follows:



The in-between can be represented more compactly as:

$$Betweenness(x) = \sum_{s \neq x \neq t \in X} \frac{\sigma_{st}(x)}{\sigma_{st}} \quad \text{………. \textbf{\textit{(Equation 1)}}}$$

where σ_{st} is the number of shortest paths from node s to node t and $\sigma_{st}(x)$ is the number of those paths that pass through x.

- ***Global transitivity:*** estimates the probability that a vertex's adjacent vertices are connected. It is also known as the clustering coefficient.
- ***Average shortest path length:*** The average number of steps along the shortest paths for all possible combinations of network hubs. It is a metric of the efficiency of connections or mass transport in a network
- ***Smallworldness:*** Refers to an ensemble of networks in which the mean shortest path between nodes increases sufficiently slowly depending on the number of network nodes. The "small-worldness" index was proposed by (Humphries & Gurney, 2008).

3.2. Data source

To achieve the objective of this study, data on social connectedness and estimates of mortality rates from infectious diseases were considered. The data used to measure the intensity of connectedness between countries comes from the Humanitarian Data Exchange. The Social Connectedness Index (SCI) is a measure of the social connectedness between different regions. In particular, it measures the relative likelihood that two people in two locations are friends with each other on Facebook. Further details on the underlying data and construction of the index have been described elsewhere (Bailey et al., 2018). The resulting SCI network data in Africa consists of 50 nodes and 2,450 edges. In addition to the social connectedness data, estimates of the mortality rate from infectious diseases were obtained from the World Bank's 2020 Global Health Estimates.

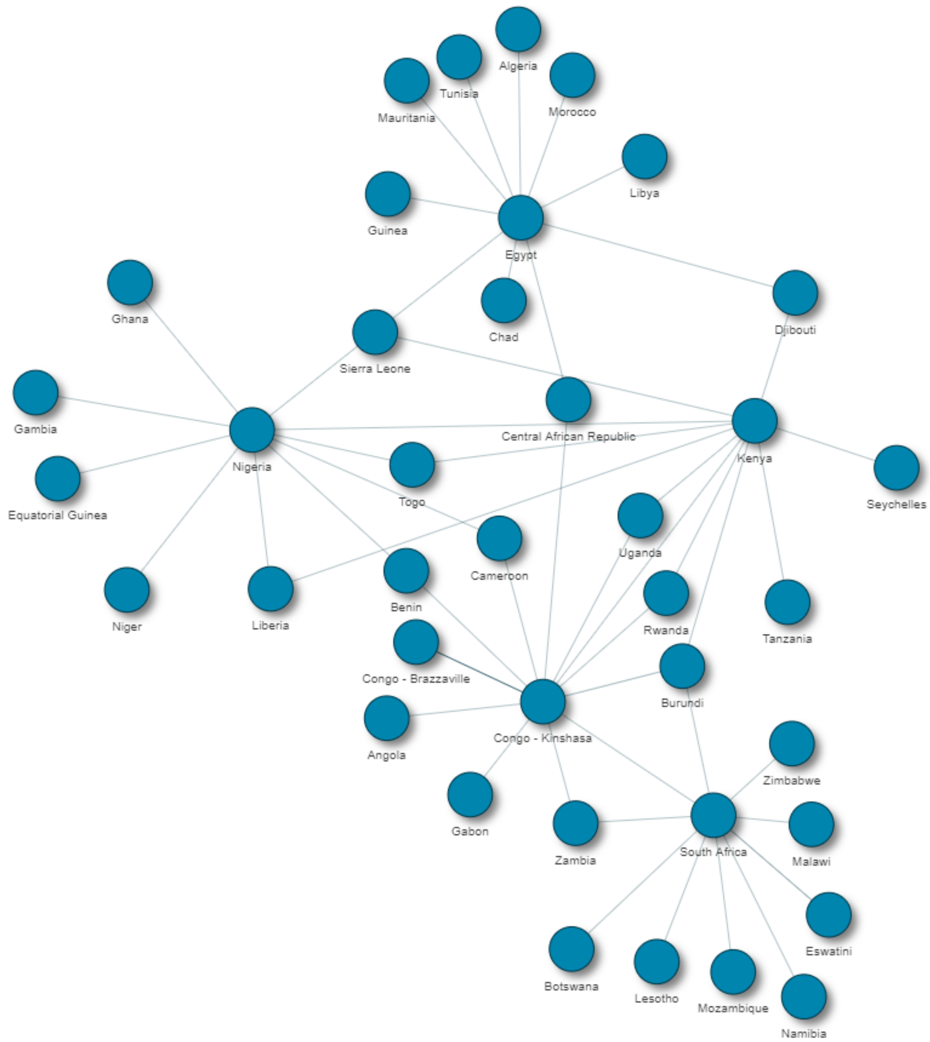


*Figure 2.1: Social network connectivity of African countries: The best-connected countries in different regions. (**North**: Egypt, **South**: South Africa, **West**: Nigeria, **East**: Kenya, **Central**: Congo-Kinshasa)*

3.3. Ethical consideration

The dataset includes an anonymized snapshot of all active Facebook users and their friend networks to measure the intensity of connectivity between locations. The data is made publicly available by Facebook's Data for Good team, which has developed privacy protection products to help solve some of the world's biggest humanitarian problems. The index provides researchers with connectedness scores, but not the number of connections between two places or any of the underlying data. The dataset uses sampling, differential privacy noise, and normalization to protect privacy.

3.4. Methods

This study adopts a quantitative analytical approach by applying three main analytical methods: (1) social network analysis; (2) hypothesis tests on the distributions of network data and; (3) Analysis, testing, and evaluation of the effect of controlling crucial nodes in the community by fitting a simple regression function with one parameter (dependent variable is the proportion of deaths due to infectious diseases).

3.4.1. Social Network Analysis

To characterize and understand the structure of online social interaction between African countries, SNA techniques are applied to a network built from Facebook social connectedness data, with each country as a node and the index of social connectedness as a marginal weight looked at. Two nodes are called neighbors if they are linked by an edge, and the *degree* k_i is the number of neighbors it has. The structure of the network is first assessed by node-link diagrams with standard measures of centrality since network data is often quite complex and graphs help human pattern recognition abilities to better understand social structures (Correa & Ma, 2011).

3.4.2. Distributions of network data: hypothesis testing

In this study, we test whether SCI data from Facebook obey a power-law closure distribution. The distribution can be described as the 80-20 rule (known as Pareto's principle, Zepf's law) governing various domains. Social network theories assume that a small number of vertices meet a large number of edges (Khan et al., 2015). The Kolmogorov-Smirnov (KS) test was performed to test whether the generated data are from the power-law distribution with selected parameters and the observed data are from the same distributions. In addition, we estimated the Small-worldness Index to determine whether Facebook's social ties form a network with “Smallworld" characteristics using the following procedure:

1. Checking if a network’s smallworldness is higher than one;

2. Inspecting that the network has transitivity substantially higher than comparable random networks, and
3. That its average shortest path length is similar or higher (but not many times higher) than that computed on random networks.

3.4.3. The effect of social ties on death due to the spread of communicable diseases

To demonstrate the effectiveness of social connectedness on the further and faster spread of infectious diseases, a simple linear regression analysis was used to assess the effect of social connectedness on death due to the spread of infectious diseases. Our regression model is represented by:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \text{………. } \textbf{\textit{(Equation 2)}}$$

Where:

Y_i = Death rate due to the spread of infectious diseases in a country

X_i= Facebook's social connectedness index for a country

β_0 *and* β_1 = Model coefficients

ε_i = The error term which is assumed to be random.

4. Results

The main focus of this study is to describe the online social interaction structure of African countries to better understand the effect of social ties on death due to the spread of infectious diseases. The scope of the study is detailed in the following matrix (Table 4.1).

Table 4.1: Scope of the study.

Research questions	**Indicators**	**Measurement techniques**	**Important variables**
How effective is the social connectedness index at describing the social interaction structure of African countries?	• Network density • Graph diameter • Average path length • Smallworldness	Counts, proportion, average, and euclidean distance	Proportion of all potential edges between vertices that exist in the network graph, and path length between any pair of vertices
What are the key nodes playing a central role in the entire network?	• Degree centrality, • Closeness centrality, • Betweenness centrality	Counts	Top connected countries: Countries that quickly connect other countries in the network
What effect do social ties have on death due to the spread of communicable diseases in Africa?	• Social connectedness • Death rate	Rate, average	Average country's social connectedness, and death rate due to the spread of infectious diseases

4.1. Social Network Analysis

SNA is commonly used to find nodes that act as a bridge from one part of a diagram to another. This section briefly introduces the social connections between African countries. As mentioned in Section 3.2, the SCI provides the numerical estimate of the intensity of social ties across countries. Figure 4.1 depicts the visual representation of a simplified network diagram of the best-connected countries across Africa. It can be observed that Central African countries (such as Equatorial Guinea, and the Central African Republic) have stronger ties to the continent (using the legend in which the navy color represents a high degree of centrality). As shown, Equatorial Guinea has the highest betweenness centrality (419.83) and is more central in the network. This implies that Equatorial Guinea has more control over the network, it serves as a bridge from one part of a chart to another. Betweenness centrality estimates the frequency with which a country acts as a bridge on the shortest path between two other countries (Equation 1).

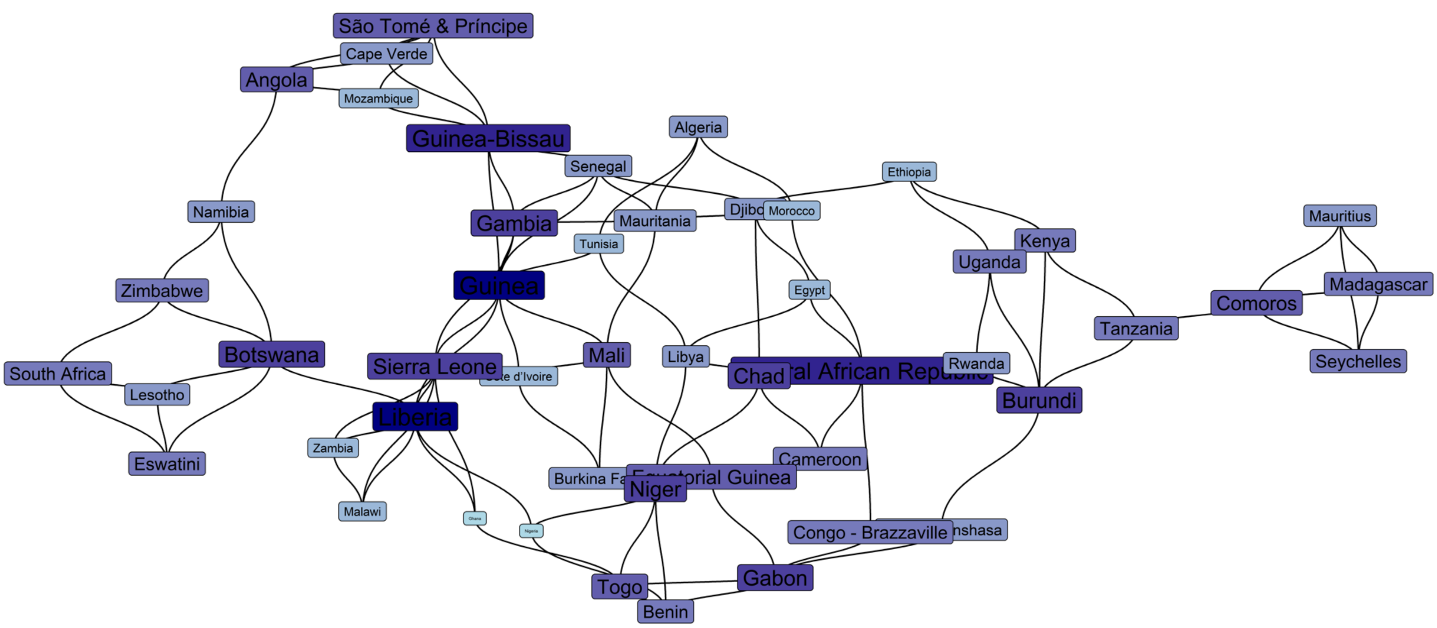


Figure 4.1: Social-tie Network of African Countries on Facebook: scale fill gradient: high = Navy, low = light blue

The estimated graph density was 6%. This is the fraction of all potential edges between vertices that exist in the network graph. The average length of the shortest paths between all combinations of vertices in the network was 3.98 and the longest path length between each pair of vertices was 11.

Table 4.2: Table of network parameters

Graph density	Graph diameter	Average path length
0.61	5	3.98

4.2. Network Data Distribution

Figure 4.2 illustrates the degree distribution of nodes in the Facebook social connection network. We can see that the graph's pattern shows a distribution model with a strong tail. Degree distributions tend to be right-skewed; that is, only a few nodes in most networks have the most connections. The existence of hubs that are orders of magnitude larger than most nodes is a feature of power-law networks. The parameter (gamma, γ) of a power-law distribution was estimated using the poweRlaw library in R and scaling-free properties of the network were tested. The power distribution is characterized by the exponent and the probability that a node has the degree k (k links) is given by:

$$\mathbf{P}(k) \sim k^{-\gamma} \text{ ………. } \textbf{\textit{(Equation 3)}}$$

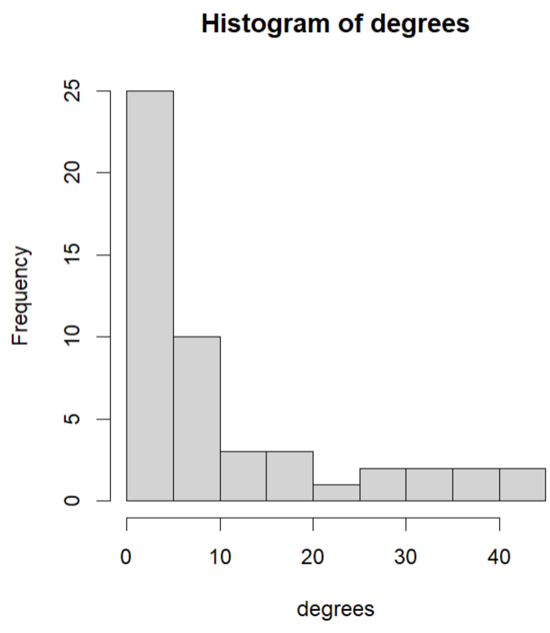


Figure 4.2: Node Degree Distribution

The study notes that the parameter is 2.63, which is greater than two. For scale-free networks, the exponent is between 2 and 3 (Barabási & Oltvai, 2004). We could not reject the hypothesis that our data follow a power-law distribution. The Kolmogorov-Smirnov test (Min D = 0.138, P-value = 0.93) shows that Facebook's social connectedness data follow the power-law degree distribution.

Table 4.3: Power-law distribution/Scale-free network

Gamma (γ)	Minimum $\mathbb{D}$/Kolgomorov-Smirnov (P-Value)
2.63	0.138 (0.93)

Like most scale-free networks (Amaral et al., 2000), the network formed by facebook social connectedness index meets small-world properties as indicated in table 3.4.

- The network's small worldness index (4.18) is higher than one;
- The network's transitivity (0.34) is higher than comparable random networks (0.06), and
- The network's average shortest path length (3.98) is higher than that computed on random networks (2.95).

Table 4.4: Smallworldness

Small-worldness index	Global transitivity	Average transitivity in the random networks	Average path length in random networks	Average path length
4.18	0.34	0.06	2.95	3.98

4.3. Effect of social ties on death due to the spread of infectious diseases

The results of our estimation show the empirical relationship between the social connectedness index and the death rate due to infectious diseases.

Table 4.5: Summary of Fitted Model

coefficients:				
Intercept	-56.634	22.623	-2.503	0.0158 *
Scaled_SCI	9.052	2.021	4.479	4.63e-05 ***
*F-statistic: 20.06 on 1 and 48 DF, p-value: 4.629e-05 ****				
*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*				

The regression output shows that the scaled variable of social connectedness is statistically significant (p-value < 0.0001). Then we find that our parameter of interest (scaled SCI) is 9.052. We interpret this as a 1% point increase in the SCI, resulting in an approximately 9.1% point increase in the death rate from infectious diseases. This significant impact of social connectedness draws attention to other policy tools that can foster social connections between citizens of neighboring nations as a strategy to stem the spread of disease(s).

5. Discussion

We examined the structure of online social interaction between African countries. Social Network Analysis (SNA) techniques were applied to a network built from Facebook social connectedness data, considering each country as a node and the index of social connectedness as an edge weight. Our sets of equations were developed to assess the structural properties of smallworldness and scale-free networks to understand how nations are socially connected and how social networks affect the overall state and pattern of infectious disease spread. According to Valente and Punpuang, people are more influenced by the people they are directly connected with than by imaginary connections (Valente & Pumpuang, 2006).

The low density of the resulting network (6%) and long-distance (5 pairs of nodes to go from one node to another) indicate weak transmission through the community. The results lead to a similar conclusion, with the density of networks decreasing as the network size increases (De, 2004). The study results show two things. First, the measurements of the global centrality of the networks showed that countries from the central region are best suited to affect the entire network fastest(**RQ 2)**. Second, countries identified as critical nodes in the network have higher death rates due to the spread of infectious diseases. Our findings shed light on how countries' social connectedness might affect the spread of infectious diseases and subsequent increased mortality from these diseases (**RQ 3)**. The results are in direct agreement with previous results (Glass & Glass, 2008; Priss & Nery, 2007).

The results of the social network structure of African countries formed by the Facebook social connectedness index are consistent with previous studies showing the distribution of power-law in a social network (Tanimoto, 2009b). Furthermore, the results of the first estimation indicated that the social network structure formed by the Facebook social connectedness index is a scale-free network that is a subset of smallworld networks (**RQ 1)**. A popular explanation is that infectious diseases can be modeled on a scale-free network and neglecting individual interactions when studying the spread of an epidemic can lead to a dramatic underestimation of the severity of infection (Zino et al., 2018). Furthermore, in this study, we find that the intensity of a country's social connectedness has a significant impact on the death rate from infectious diseases.

6. Conclusion

The social network analytical approach presented here is the innovative and new approach to characterizing and understanding the structure of online social interaction between communities at continental scale (Africa). Our analysis focuses on determining and understanding the overall structure and impact on the spread of infection. The focus is on the influence of social networks formed by online communities to help better control the spread of infectious diseases.

Overall, the study results demonstrate a strong effect of social ties on the death rate due to the spread of communicable diseases. Importantly, our results provide evidence on how countries can improve their disease preparedness to help prevent avoidable infections by monitoring their connected countries. Our results on distributions of social connectedness network data are broadly consistent with the fact that various structures of network data follow a power-law degree distribution known as the Pareto principle or Zepf's Law.

This aspect of the research only covered a basic set of tools for drawing inferences from complex network data. These tools are: (1) estimating centrality measures and describing network parameters, (2) measuring and evaluating social connectedness data distribution, and (3) building an empirical analytical approach to estimate the relationship between social connectedness index and death rate due to infectious diseases. Therefore, future research should be conducted in more realistic network models for how networks change over time and affect the spread of infectious diseases.

References

1. Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149–11152. https://doi.org/10.1073/pnas.200327197
2. Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*, *32*(3), 259–280. https://doi.org/10.1257/jep.32.3.259
3. Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272
4. Barnes, J. A., & Harary, F. (1983). Graph theory in network analysis. *Deepblue.lib.umich.edu*. https://deepblue.lib.umich.edu/handle/2027.42/25206
5. Birmingham, M., & Stein, C. (2003). *GLOBAL BURDEN OF DISEASE: PART A. The Burden of Vaccine-Preventable Diseases* (B. R. Bloom & P.-H. Lambert, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/pii/B9780121072582500032?via%3Dihub
6. Bisanzio, D., Bertolotti, L., Tomassone, L., Amore, G., Ragagli, C., Mannelli, A., Giacobini, M., & Provero, P. (2010). Modeling the Spread of Vector-Borne Diseases on Bipartite Networks. *PLoS ONE*, *5*(11), e13796. https://doi.org/10.1371/journal.pone.0013796
7. Correa, C. D., & Ma, K.-L. (2011). Visualizing Social Networks. *Social Network Data Analytics*, 307–326. https://doi.org/10.1007/978-1-4419-8462-3_11
8. De, P. (2004). Sexual network analysis of a gonorrhoea outbreak. *Sexually Transmitted Infections*, *80*(4), 280–285. https://doi.org/10.1136/sti.2003.007187

9. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014. https://doi.org/10.1038/nature07634

10. Glass, L. M., & Glass, R. J. (2008). Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Public Health*, *8*(1). https://doi.org/10.1186/1471-2458-8-61

11. Gómez-Gardeñes, J., Latora, V., Moreno, Y., & Profumo, E. (2008). Spreading of sexually transmitted diseases in heterosexual populations. *Proceedings of the National Academy of Sciences*, *105*(5), 1399–1404. https://doi.org/10.1073/pnas.0707332105

12. Grieve, R., Indian, M., Witteveen, K., Anne Tolan, G., & Marrington, J. (2013). Face-to-face or Facebook: Can social connectedness be derived online? *Computers in Human Behavior*, *29*(3), 604–609. https://doi.org/10.1016/j.chb.2012.11.017

13. Hill, A. L., Rand, D. G., Nowak, M. A., & Christakis, N. A. (2010). Infectious Disease Modeling of Social Contagion in Networks. *PLoS Computational Biology*, *6*(11), e1000968. https://doi.org/10.1371/journal.pcbi.1000968

14. Humphries, M. D., & Gurney, K. (2008). Network “Small-World-Ness”: A Quantitative Method for Determining Canonical Network Equivalence. *PLoS ONE*, *3*(4), e0002051. https://doi.org/10.1371/journal.pone.0002051

15. Khabbaz, R. F., Moseley, R. R., Steiner, R. J., Levitt, A. M., & Bell, B. P. (2014). Challenges of infectious diseases in the USA. *Lancet (London, England)*, *384*(9937), 53–63. https://doi.org/10.1016/S0140-6736(14)60890-4

16. Khan, B., Dombrowski, K., Curtis, R., & Wendel, T. (2015). Estimating Vertex Measures in Social Networks by Sampling Completions of RDS Trees. *Social Networking*, *04*(01), 1–16. https://doi.org/10.4236/sn.2015.41001

17. Klovdahl, A. S. (1985). Social networks and the spread of infectious diseases: The AIDS example. *Social Science & Medicine*, *21*(11), 1203–1216. https://doi.org/10.1016/0277-9536(85)90269-2

18. Natale, F., Giovannini, A., Savini, L., Palma, D., Possenti, L., Fiore, G., & Calistri, P. (2009). Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine*, *92*(4), 341–350. https://doi.org/10.1016/j.prevetmed.2009.08.026

19. Périssé, A., & Nery, J. A. da C. (2007). The relevance of social network analysis on the epidemiology and prevention of sexually transmitted diseases. *Cadernos de Saude Publica*. https://doi.org/10.1590/S0102-311X2007001500004

20. Sadilek, A., Kautz, H., & Silenzio, V. (2012). Modeling Spread of Disease from Social Interactions. *Proceedings of the International AAAI Conference on Web and Social Media*, *6*(1), 322–329. https://ojs.aaai.org/index.php/ICWSM/article/view/14235

21. Sánchez-Matamoros, A., Martínez-López, B., Sánchez-Vizcaíno, F., & Sánchez-Vizcaíno, J. M. (2012). Social Network Analysis of Equidae Movements and Its Application to Risk-Based Surveillance and to Control of Spread of Potential Equidae Diseases. *Transboundary and Emerging Diseases*, *60*(5), 448–459. https://doi.org/10.1111/j.1865-1682.2012.01365.x

22. Stattner, E., & Vidot, N. (2011, May 1). *Social network analysis in epidemiology: Current trends and perspectives*. IEEE Xplore. https://doi.org/10.1109/RCIS.2011.6006866

23. Tanimoto, S. (2009a). A remark on the BA model of scale-free networks. *ArXiv:0902.1570 [Physics]*. https://arxiv.org/abs/0902.1570

24. Tanimoto, S. (2009b). Power laws of the in-degree and out-degree distributions of complex networks. *ArXiv:0912.2793 [Physics]*. https://arxiv.org/abs/0912.2793

25. Taylor, M. B. (2011). Water- and foodborne viruses : current developments. *CME : Your SA Journal of CPD*, *29*(5), 207–209. https://doi.org/10.10520/EJC63980

26. Torti, C., Mazzitelli, M., Trecarichi, E. M., & Darius, O. (2020). Potential implications of SARS-CoV-2 epidemic in Africa: where are we going from now? *BMC Infectious Diseases*, *20*(1). https://doi.org/10.1186/s12879-020-05147-8
27. Valente, T. W., & Pumpuang, P. (2006). Identifying Opinion Leaders to Promote Behavior Change. *Health Education & Behavior*, *34*(6), 881–896. https://doi.org/10.1177/1090198106297855
28. World Health Organization. Regional Office for Europe. (2021). Infection prevention and control: guidance to action tools. In *apps.who.int*. World Health Organization. Regional Office for Europe. https://apps.who.int/iris/handle/10665/341107
29. Zelner, J. L., Trostle, J., Goldstick, J. E., Cevallos, W., House, J. S., & Eisenberg, J. N. S. (2012). Social Connectedness and Disease Transmission: Social Organization, Cohesion, Village Context, and Infection Risk in Rural Ecuador. *American Journal of Public Health*, *102*(12), 2233–2239. https://doi.org/10.2105/ajph.2012.300795
30. Zino, L., Rizzo, A., & Porfiri, M. (2018). Modeling Memory Effects in Activity-Driven Networks. *SIAM Journal on Applied Dynamical Systems*, *17*(4), 2830–2854. https://doi.org/10.1137/18m1171485

Appendices

Appendix 1: Data sources

- ***The Humanitarian Data Exchange**: Facebook Social Connectedness Index (https://data.humdata.org/dataset?q=sci+data)*
- ***World Development Indicators and Other World Bank Data**: Death rate due communicable diseases (SH.DTH.COMM.ZS)*

Appendix 2: R Packages used

- *countrycode*
- *dplyr*
- *ggraph*
- *Hmisc*
- *igraph*
- *poweRlaw*
- *qgraph*
- *rio*
- *splineTimeR*
- *tidyverse*
- *visNetwork*
- *WDI*
- *writexl*

Appendix 3 Table of centralities measures

Country	Degree	Closeness	Betweenness
Angola	23	0.013	31.594
Burkina Faso	28	0.013	41.672
Burundi	44	0.016	165.058
Benin	33	0.014	74.113
Botswana	44	0.017	106.940
Congo - Kinshasa	30	0.014	41.888
Central African Republic	23	0.013	23.369
Congo - Brazzaville	22	0.013	26.910
Côte d’Ivoire	17	0.012	6.508
Cameroon	45	0.018	76.928
Cape Verde	19	0.013	8.243
Djibouti	21	0.013	11.657
Algeria	29	0.014	25.340
Egypt	19	0.013	6.098
Ethiopia	18	0.013	5.848
Gabon	35	0.015	68.817
Ghana	14	0.012	9.548
Gambia	15	0.012	12.480
Guinea	30	0.014	43.922
Equatorial Guinea	18	0.013	14.043
Guinea-Bissau	16	0.012	14.561
Kenya	20	0.013	48.025
Comoros	12	0.011	5.474
Liberia	14	0.012	20.977
Lesotho	26	0.014	33.148
Libya	34	0.015	102.624
Morocco	26	0.014	24.569
Madagascar	22	0.013	14.679
Mali	15	0.012	8.432
Mauritania	19	0.012	83.842
Mauritius	19	0.013	54.434
Malawi	19	0.013	35.215
Mozambique	14	0.012	11.316
Namibia	12	0.011	11.481
Niger	22	0.013	173.101
Nigeria	17	0.012	98.738
Rwanda	13	0.011	8.982
Seychelles	11	0.011	2.581
Sierra Leone	11	0.011	3.876
Senegal	13	0.012	12.552
São Tomé & Príncipe	11	0.011	2.853
Eswatini	12	0.011	31.357
Chad	11	0.011	2.233
Togo	13	0.011	10.667
Tunisia	12	0.012	7.748
Tanzania	12	0.012	8.869
Uganda	11	0.011	9.117
South Africa	12	0.011	10.973
Zambia	13	0.011	25.518
Zimbabwe	11	0.011	7.083

Appendix 4: Digital receipt

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	steven karera
Assignment title:	DISSERTATION 2020
Submission title:	Social Ties and Public Health: Use of Social Network Analysis...
File name:	Social_ties_and_public_health_PC_StevenK_1.docx
File size:	497.33K
Page count:	20
Word count:	4,504
Character count:	24,944
Submission date:	13-Jul-2022 10:26PM (UTC+0300)
Submission ID:	1641257594

Appendix 5: Plagiarism Rate