University of Cape Town

The Structure and Evolution of the African Air Transport Network

Author:
Mr. David Snaddon

Supervisor:
Dr. Şebnem Er

Co-Supervisor:
Mr. Stefan Britz

Minor dissertation submitted in partial fulfilment of the requirements for the degree of M.Sc. Data Science

at the

Department of Statistical Sciences

June 25, 2025

Abstract

Mr. David Snaddon

The Structure and Evolution of the African Air Transport Network

This dissertation studies the structure and evolution of the African Air Transport Network (AATN) from 2015 to 2023. With respect to the network's structure, a power-law distribution appropriately characterises the networks degree distribution. The 'golden triangle' between Johannesburg O.R. Tambo, Cape Town, and Durban King Shaka plays a dominant role in the network's structure when considering airline seat capacity. Addis Ababa shows strong growth in node centrality, ending the period with the highest centrality across all observed centrality measures. Community detection reveals airport clusters that align with geographic regions, including African subregions and countries. k-core decomposition reveals a growing core of the network spread across the continent with a higher concentration in West Africa. Longitudinal trends of network-wide indicators detail the network's evolution. A gradual decrease in the average clustering coefficient and degree assortativity coefficient but an increase in the Gini coefficient and largest degree suggest that the network aligns to a growing airline hub-and-spoke structure. During this period, the COVID-19 pandemic occurs and significantly affects the network's evolution, particularly in measures related to the sizing of the network, calling for an investigation into the changes from pre-pandemic to the end of recovery phase. When the network recovers, it does not revert back to its pre-pandemic structure completely, adding new routes and not reintroducing some old ones. Moreover, a similar magnitude of variability during the period is found in the global air transport network.

Key words:
Network analysis; Africa; Air transport network

Acknowledgements

I would like to express my gratitude to my supervisors, Şebnem Er and Stefan Britz, for their regular guidance and feedback throughout this eventful year. I also thank David King for his invaluable support in arranging access to the data. Lastly, I extend my deepest thanks to my wife, Gracia Snaddon, for her unwavering support during her pregnancy and the first few months of raising our daughter, Astrid, who has been my source of motivation in completing this work.

Contents

List of Figures

List of Tables

List of Abbreviations

AATN African Air Transport Network
CCDF Complementary Cumulative Distribution Function
GACI Global Airport Connectivity Index
GDS Global Distribution System
IATA International Air Transport Association
ICAO International Civil Aviation Organisation
MIDT Marketing Information Data Transfer
OAG Official Airlines Guide
SAATM Single African Air Transport Market

Airport codes reference:
ABV Abuja
ADD Addis Ababa
AGA Agadir
ALG Algiers
BKO Bamako
CAI Cairo International
CMN Casablanca
CPT Cape Town
DAR Dar Es Salaam
DKR Dakar Leopold Sedar Senghor
DSS Dakar Blaise Diagne International
DUR Durban King Shaka International Airport
ERS Windhoek Eros Airport
GOM Goma
HLA Johannesburg Lanseria International Airport
JNB Johannesburg O.R. Tambo International
LAD Luanda
LOS Lagos
MBA Mombasa
NBO Nairobi Jomo Kenyatta International Airport
PLZ Port Elizabeth

WIL Nairobi Wilson Airport
WVB Walvis Bay
ZNZ Zanzibar

Chapter 1

Introduction

The air transport network in Africa plays a key role in the economic and social development of the continent, making its study highly valuable. Air transport networks comprise airports connected by scheduled passenger air routes. Each new route added to the network not only establishes a direct link between two cities. It also enhances onward connections, thereby opening up new economic opportunities by facilitating the movement of people and goods.

This dissertation presents an analysis of the structure and evolution of the African Air Transport Network (AATN) using network analysis techniques. Section 1.1 discusses Africa and why the analysis of its air transport network is important, as well as the efforts being made to promote the growth of the AATN. Section 1.2 then provides an explanation of network analysis and how this is applied to the AATN. Section 1.3 presents the research objectives and Section 1.4 provides the dissertation outline.

1.1 African Air Transport Network

Africa has a general lack of integrated transportation infrastructure, which has been identified as an obstacle to the continent's growth in intra-regional trade (Bassens et al., 2012). Air transport networks can be particularly useful to connect cities. This is especially true where ground transport infrastructure between them is inadequate or complicated by geographic obstacles and borders. However, this comes at a higher cost.

The population of Africa was estimated to be 1.48 billion in 2023, representing 18.3% of the global population (United Nations and Social Affairs, 2024). In contrast, the estimated GDP of Africa was $2.86 trillion US in 2023, accounting for approximately 2.73% of the global GDP (International Monetary Fund, 2024). According to Walsh et al. (2024), Africa's presence in the aviation industry accounted for 2.1% of global air passengers. This disparity highlights the apparent need for the African aviation industry and broader economy to grow.

Efforts are being made at various levels to stimulate and grow both components within the AATN and the AATN as a whole. At a continental level, the African Union initiated a flagship project, in January 2018, called the Single African Air Transport Market

(SAATM) (African Union Commission et al., 2019). SAATM aims to fully implement the Yamoussoukro Decision. This is done through a set of guidelines that include liberalising air service agreements between member states. It also involves adhering to International Civil Aviation Organisation (ICAO) standards and other best practices on safety. Furthermore, SAATM enablers, such as improving visa openness, are part of the implementation. SAATM is expected to improve air connectivity and travel convenience, reduce air travel costs, and stimulate economic growth and tourism on the continent. By 2021, 35 member states had committed to operationalising SAATM where it was determined that member states were only partially implementing the Yamoussoukro Decision, indicating challenges to overcome to achieve SAATM (African Union Commission, 2021). Bassens et al. (2012) note how air connectivity between African cities has been notably aided by the Yamoussoukro Decision and has made air transport a major driver in the globalisation of African cities.

At more localised levels, various initiatives exist to grow regions within the AATN. For example, Cape Town Air Access was launched in 2015 by the Provincial Government of the Western Cape, South Africa, to improve direct air access to the City of Cape Town (Nonyati, 2020). The initiative focuses on developing business cases, supported by data analysis, to present to airlines for both new and existing routes. It offers ongoing performance monitoring, route-stimulation initiatives, marketing and communication support for airlines, and policy and regulatory advocacy to address stakeholder needs with government entities (*Services – Air Access | Wesgro* 2025). This project has successfully facilitated several new airline routes, including intra-African connections, thus contributing to the AATN's growth (*Home – Air Access | Wesgro* 2021). With the need for growth in the AATN established and efforts to achieve it presented, the next focus is an exploration of network analysis and its application to the AATN.

1.2 Network Analysis

Network analysis is an interdisciplinary study of networks, where networks consist of nodes that are joined together by edges (Newman, 2010). A network can be more formally defined by:

$$G = (V, E), \tag{1.1}$$

where G is the network, V is the set of all nodes in the network, and E is the set of all edges in the network. The sets V and E satisfy the condition:

$$E \subseteq [V]^2, \tag{1.2}$$

where the edges consist of pairs of the nodes. For this study, nodes represent airports whilst edges represent scheduled passenger air routes. Network analysis techniques can reveal various features of the network. These include its size and efficiency, as well as identifying communities within it. Such techniques also help determine the network's

core, and identify key airports and routes that are important to the network. Ultimately, this contributes to an understanding of the network's structure.

Network analysis is an interdisciplinary study of networks, where networks consist of nodes that are connected by edges (Newman, 2010). A network can be more formally defined as:

$$G = (V, E), \tag{1.3}$$

where G is the network, V is the set of all nodes, and E is the set of all edges. The sets V and E satisfy the condition:

$$E \subseteq [V]^2, \tag{1.4}$$

meaning that each edge is a pair of nodes (Diestel, 2017). In this study, nodes represent airports, while edges represent scheduled passenger air routes. By applying network analysis techniques, it is possible to assess properties such as the network's size and efficiency, identify communities, determine the network's core, and highlight key airports and routes that contribute to the overall structure of the network.

Representing the air transport network for network analysis purposes results in a simplification of information, where only intra-African routes are included, and certain temporal aspects and unique flight details are omitted. However, this approach reveals several insights about the structure and evolution of the AATN. A more comprehensive explanation about the network assembly and the assumptions made are provided in Chapter 3.

Insights generated through network analysis on the AATN can inform various fields of study such as economic development for tourism, trade and investment (Nonyati, 2020); government policy, international relations and air route development; geography (Otiso et al., 2011); and epidemiology (Rocha, 2009). The research objectives of this investigation are presented next.

1.3 Research Objectives

This study covers the period from 2015 to 2023 to study the evolution of the AATN. During this period, the COVID-19 pandemic causes a major disruption in the AATN and the global network, which can substantially influence the future structure and evolution of the AATN.

The research objectives of this dissertation are divided into three components:

1. Illustrate and describe the structure of the African Air Transport Network
2. Detail the evolution of the African Air Transport Network
3. Evaluate the recovery of the African Air Transport Network after the COVID-19 pandemic

For the purposes of this dissertation, the analysis of the evolution of the AATN focuses on characteristics that can be analysed over time to understand how the network is evolving.

These characteristics can reveal much about the network's structure. At the same time, they reduce the information pertaining to a characteristic at a given point in time to a single value or simplified representation, making longitudinal analysis more manageable. In contrast, the analysis of the AATN's structure will focus on structural characteristics within the network for the most recent year, which may be too extensive and possibly not meaningful to analyse in terms of evolution. The analysis of the recovery of the AATN after the COVID-19 pandemic makes use of network analysis tools to understand how much the network has recovered, but also how the recovered network differs from that prior to the pandemic.

1.4 Dissertation Outline

Network analysis is used to study the structure and evolution of the AATN, aiming to contribute to the literature of network analysis applied to regional air transport networks and of the AATN in general. Chapter 2 reviews relevant literature on the application of network analysis to regional air transport networks and other influencing works related to network analysis and the AATN. Chapter 3 describes and presents the data source and preparation, where airline schedules data are transformed into networks.

Chapter 4 presents the methodology, which is followed by the results and discussion in Chapter 5. Both the methodology and the results and discussion chapters are presented in the structure of the three research objectives using network analysis techniques to first illustrate and describe the structure of the AATN, then to detail its evolution, and finally to evaluate the recovery of the AATN after the COVID-19 pandemic. Conclusions and recommendations are then made in Chapter 6.

Chapter 2

Literature Review

This literature review provides an overview of existing research on air transport networks. Section 2.1 examines applications of network analysis on regional air transport networks, identifying key concepts and methods. Section 2.2 broadens the scope to include wider applications and related studies, offering insights into addressing shortcomings in existing methodologies. In addition, it explores analytical methods that could enhance research efforts. To provide additional context, research beyond network analysis on the African Air Transport Network (AATN) is also considered. Finally, Section 2.3 concludes the chapter. This review guides the formulation of the research objectives and methodology for the dissertation.

2.1 Structure and evolution of regional air transport networks using network analysis

The use of network analysis to analyse the structure and evolution of a regional air transport network is a recurring topic. Table 2.1 provides a selection of literature that follows this theme. Some studies approach this from a global perspective (Guimerà et al., 2005; Song and Yeo, 2017), whilst others focus on intra-country regions such as Dai et al. (2018) for South East Asia and Cardillo et al. (2013) for Europe. Furthermore, several other studies focus on a country perspective, such as Xu and Harriss (2008), Jia et al. (2014), and Cheung and Gunes (2012) for the USA; Bagler (2008) for India; and Rocha (2009) for Brazil. Wang et al. (2011) addressed the structure of China's air transport network, while Wang et al. (2014) – a study by mostly the same group of authors – addressed the network's evolution.

TABLE 2.1: Studies of air transport networks for various regions using network analysis.

Reference	Region	Latest Year	Evolution Period	Period	Spatial Aggregation	Directed / Undirected
Guimerà et al. (2005)	World	2000		1 Week	City	Undirected
Song and Yeo (2017)	World	2017		N/A	Airport	Undirected
Dai et al. (2018)	South East Asia	2012	1979-2012	Annual	City	Undirected
Cardillo et al. (2013)	Europe	2011	2011	1 Day	Airport	Undirected
Wang et al. (2011)	China	2007/2008		5 Months	City	Undirected
Wang et al. (2014)	China	2012	1930-2012	Annual	City	Undirected
Xu and Harriss (2008)	USA	2005	2002-2005	Quarterly	City	Undirected
Jia et al. (2014)	USA	2010	1990-2010	Annual	City	Directed
Cheung and Gunes (2012)	USA	2011	1991-2011	Month	Airport	Directed
Bagler (2008)	India	2004		N/A	Airport	Both
Rocha (2009)	Brazil	2006	1995-2006	Annual	City	Both

There are various considerations to take when a network [1] is assembled to represent an air transport network, which are often constrained by the available data. The network is a simplified representation of the real-world network, which has time-dependent characteristics, such as the times of departures and arrivals and the dates that flights are available. Aggregations are performed over periods of time to determine which edges and nodes to include in the network. In Table 2.1, these time periods are indicated for the various studies. For example, Xu and Harriss (2008) aggregated over a quarterly period, where any routes that exist during the given quarter are included. The result is 16 different networks covering all of the quarterly periods from 2002 to 2005. Another consideration is the spatial aggregation, where some studies establish the nodes of the network as the cities in which the airports are located; therefore, if multiple airports are located in a single city, each airport and its connections are merged into a single node. Additionally, networks are assembled as either directed, with edges representing flights with direction between nodes, or undirected. An undirected network is simpler and enables more analysis techniques compared to a directed model (Bagler, 2008). Table 2.1 indicates whether the network is directed or undirected and which spatial aggregation level is applied to their networks.

Several characteristics are investigated in the various studies of the structure of the air transport networks. Among these, the degree distribution is found to be widely examined (Guimerà et al., 2005; Dai et al., 2018; Cardillo et al., 2013; Wang et al., 2011; Xu and Harriss, 2008; Jia et al., 2014; Cheung and Gunes, 2012; Bagler, 2008; Rocha, 2009). The degree of a node is simply the number of edges connected to it, and hence the degree distribution is concerned with the distribution of the degrees across all the nodes. It is typical for the degree distribution of a network – including different types of networks such as social networks and web page links – to follow a power-law distribution[2] (Newman, 2010). In the various studies, not all networks were determined to have a power-law degree distribution. However, all of them could generally be characterised as having a high frequency of nodes with lower degree and a low frequency of nodes with higher degree.

The distributions of other metrics at the node level were analysed, such as node betweenness centrality (Guimerà et al., 2005; Wang et al., 2011; Xu and Harriss, 2008); and node closeness centrality (Wang et al., 2011). Betweenness centrality distributions are also known to show power-law distributions (Newman, 2010).

Guimerà et al. (2005), Song and Yeo (2017), Wang et al. (2011), Cheung and Gunes (2012), and Rocha (2009) made use of centralities at the node level, which were used to highlight the best performing nodes within the air transport network among various centrality metrics such as degree, betweenness, and closeness centrality. This helps describe the structure of the network through key nodes of high importance.

A centrality measure of interest concerning key edges is the edge betweenness centrality. The studies reviewed in Table 2.1 do not include much analysis on edge betweenness

[1] Often referred to as a graph

[2] See Section 4.1.1 for a detailed description of the power-law.

centrality, although Xu and Harriss (2008) analysed the distribution of edge betweenness centrality of the USA network in the 4th quarter of 2005 and revealed a power-law distribution.

Geographic plots of the spatial distribution of nodes and edges allow for geographic characteristics of the network to be analysed. Guimerà et al. (2005), Wang et al. (2011), and Xu and Harriss (2008) did this while mapping the node degree to the size of the node plotted. Guimerà et al. (2005) and Wang et al. (2011) mapped node betweenness centrality to the plot, while Wang et al. (2011) mapped node closeness centrality. These plots are useful as they help one see the spatial distribution of nodes whilst also indicating their importance to the air transport network according to a given centrality measure.

Relationships between various attributes were analysed to help understand the structure of the air transport network. Guimerà et al. (2005) showed the relationship between node betweenness and node degree for the global air transport network in 2000, where there was a positive relationship between betweenness and degree, but low correlation. Xu and Harriss (2008) in contrast found a positive non-linear relationship with high correlation between betweenness and degree for the USA air transport network in the 4th quarter of 2005. Xu and Harriss (2008) also showed the relationship between a node's clustering coefficient and degree in the USA air transport network. Their findings indicated that cities with low degree tend to have a high clustering coefficient, meaning that neighbouring cities are highly interconnected. Conversely, cities with high degree have lower clustering coefficients. Xu and Harriss (2008) highlighted that this indicates a 'rich-club' effect, which is embedded in hub-and-spoke[3] network structures.

Some studies used community detection, where communities of nodes within an interconnected network were identified based on the connections that existed between them. Guimerà et al. (2005) performed community detection on the global air transport network for 2000 using simulated annealing for modularity maximisation. This revealed communities that were aligned with geographic regions around the world. Jia et al. (2014) performed community detection on the USA air transport network for 1991, 2002 and 2010 using Louvain Community Detection, which highlighted Alaska and Hawaii to have distinct communities, but for the rest of the USA it appeared to not highlight any distinct communities according to geographic region.

The structural analysis of transport networks focuses on characteristics and features that exist within a network. This approach has been described as being 'cross-sectional' (Dai et al., 2018). In contrast, the analysis of the evolution of these networks focuses on characteristics of the network that can be tracked over time. One of the ways this has been done is to use certain network metrics, which include number of nodes and edges in a network, average shortest path, average clustering coefficient, power-law exponent, degree assortativity coefficient, average degree, network density, and network diameter.

[3]In a hub-and-spoke structure, airlines design a network where passengers may change aircraft at hub airports, allowing for optimisation of economies of density (Brueckner et al., 1992).

Each of these metrics provides insight into the structure of a network; moreover, they can be tracked over a period of time.

Jia et al. (2014) presented the evolution of the number of nodes and edges in the USA air transport network from 1990 to 2010. In this period, there was a major increase in the number of cities and routes included in the network after 2001. The 9/11 terrorist attack may have prompted the construction of new airports to enhance the network's resilience. In contrast, Rocha (2009) observed a gradual decline in the number of airports and routes in the Brazilian air transport network from 1995 to 2006, while the number of passengers grew.

Dai et al. (2018) compared the average shortest path and clustering coefficient to those of randomly generated networks with the same number of nodes and edges as the actual South East Asian air transport network from 1979 to 2012. The study found a clear shift in 1996. From this point, the actual network's average shortest path became shorter and more efficient than the randomly generated network benchmark. In contrast, the average clustering coefficient of the randomly generated network stayed consistently lower than that of the actual network throughout the entire period. Dai et al. (2018) monitored the power-law exponent, describing the slope of the network's degree distribution, and the degree assortativity coefficient. Over the period, the power-law exponent decreased as more cities gained higher degrees. The degree assortativity coefficient, which measures the correlation between the degrees of connected nodes, went through a period of increasing slightly from -0.14 in 1979 to -0.01 in 1995 and then decreased to around -0.25 in 2012. When the degree assortativity coefficient decreases, it indicates that the degree of a node is becoming less correlated with the degrees of its connected cities, possibly due to stronger hub-and-spoke structures.

Another method of distilling the amount of information contained in a network is to reduce the network to its core, which can be tracked over time. For air transport network studies, there have been different definitions for determining the network core; for example, Wong et al. (2023) defined core edges as those that have existed consistently for more than five years. However, Dai et al. (2018) and Wang et al. (2014) made use of k-core decomposition to visualise the evolution of the network's core over time.

Cardillo et al. (2013) applied multi-layered network analysis to the European Air Transport Network, where each layer in the analysis may represent an airline. The characteristics analysed included degree distribution, average shortest path, clustering coefficient, size of the giant component, and the 'rich-club' coefficient. Samples of layers were made with increasing sample size to observe how these metrics related to sample size. A comparison was done where one layer consisted of major airlines and a second layer consisted of low-cost carriers. Low-cost carriers are known to use point-to-point[4] network models while major airlines are known to employ hub-and-spoke models. This distinction highlighted

[4]Point-to-point networks consist of direct flights without few layover options (Fu et al., 2019)

the 'rich-club' effect of major airlines. Additionally, the effect of low-cost carriers on 'small-worldness' was demonstrated.

The collection of past studies of network analysis applied to regional networks demonstrate nuances in how networks are assembled and which tools and techniques are applied to analyse the structure and evolution of the networks. The next section explores wider applications of network analysis and air transport network studies.

2.2 Broader applications of network analysis and studies of air transport networks

Continuing the topic of network analysis on air transport networks, Cheung et al. (2020) introduced the Global Airport Connectivity Index (GACI) as a singular measure of airport connectivity. This index combines five centrality measures: degree, flow-betweenness, closeness, eigenvector, and regional importance, which all exhibit a high degree of correlation. Airports were initially ranked according to each measure, and these rankings were then distilled into a single ranking through principal component analysis. Furthermore, a participation coefficient was calculated to evaluate connectivity beyond an airport's own region. By employing k-means clustering on both the participation coefficient and the GACI, hubs were classified into global and regional categories.

Arvis and Shepherd (2011) also developed an Air Connectivity Index. This was done on a country level, where Arvis and Shepherd (2011) defined connectivity as the importance of a country as a node within the global air transport system. Therefore, a gravity model was used to determine the connectivity of each country. These models were inspired by the gravity formula to which the analogous parameters of the masses of nodes and the distance between them were fed to determine the attractive force between the nodes. In this implementation, connectivity was analogous to mass.

Network analysis is used in predictive modelling to forecast passenger demand when new direct air routes are established. Notable examples include studies by Wong et al. (2023) and Guo et al. (2019). Wong et al. (2023) divided the network into core and auxiliary networks, with the core network comprising routes that have existed for more than five years. The model also quantifies the Global Airport Connectivity Index (GACI) (Cheung et al., 2020) generated between two airports when a new direct route is introduced. On the other hand, Guo et al. (2019) demonstrated the application of gravity models to determine the stimulation of passenger demand. To predict demand for a new route, proxies such as population or GDP can substitute for the analogous masses, while passenger demand may represent the attractive force.

In their studies on the airline connectivity of African airports and the globalisation of African cities, Otiso et al. (2011) and Bassens et al. (2012) –with the latter building upon the former– utilised Marketing Information Data Transfer (MIDT) data, which is based on passenger origin-destination bookings. Although network analysis was not used, they

performed aggregations of passenger flows. Instead of focusing on an airport level, the data were aggregated to the city level, aligning with the focus on urban globalisation. Otiso et al. (2011) examined snapshots from 2001 and 2009, noting substantial growth in passenger flows over this period. They studied the relative importance of one city to another by analysing the ratio of origin-destination passengers between two cities against the total passenger volume of the related city, which indicated a city's relative importance within a region. Additionally, the 'hubness' of cities was analysed through the ratio of passengers passing through an airport on onward connections compared to total passengers, noting that airports serving as hubs for home airlines tend to exhibit higher 'hubness'. The flow of passengers making onward connections through specific cities, such as Casablanca and Johannesburg, was visualised in the study.

Bassens et al. (2012) used snapshots from 2003 and 2009 to observe the evolution of air passenger throughput in various African cities. The growth in intra-African passenger flows was visualised, highlighting some domestic routes in South Africa's 'golden triangle' of routes between Johannesburg O.R. Tambo, Cape Town, and Durban King Shaka, as well as in Nigeria and Northern Africa. Bassens et al. (2012) analysed the aggregated share of African passenger flows by continental regions, showing that Africa is the highest, followed by Europe, while the share of passenger flows increased for all regions except Europe. The continental region relative link change was determined for 61 African cities, highlighting, for example, the growth of Europe's relative air connectivity with Casablanca. The absolute passenger change between all cities and the five major African cities –Johannesburg, Cairo, Casablanca, Nairobi, and Lagos– was determined and ranked. A key distinction between the studies by Otiso et al. (2011) and Bassens et al. (2012) versus the air transport network studies in Section 2.1 is that these studies focused on actual passenger movements, while the air transport network studies focused on air transport infrastructure.

In their study on the impact of the COVID-19 pandemic, Bao et al. (2021) analysed the evolution of the global air transport network on a weekly basis from 1 January to 6 May 2020. They introduced a model that tested three strategies for removing connections from the network, assessing their effects on metrics such as unweighted average shortest path, average betweenness, and average t-core (indicative of the probability for cities to form triangles) over the period. The strategies employed were high degree, random, and a combination of both. The study identified cities with the most substantial decrease in betweenness, many of which were tourism hubs, as well as those with the least decrease, often capital cities. A plot of the correlation of the cities' betweenness rankings compared with that on the initial day demonstrated a substantial transformation in nodal centrality within the global air transport network during the time frame.

It should be noted that the studies referenced in Section 2.1 did not consider the time dependency of the transport network. For example, the waiting period at connecting airports between the arrival and departure of connecting flights was not considered. Wang et al. (2019) reviewed advancements in time-dependent network analysis, defining such networks as those with edge weights that change over time. Their findings suggested

that this field was still emerging, presented numerous challenges, and required substantial computational resources.

Resilience is a popular theme in transport network studies. Cheung and Gunes (2012) evaluated the resilience of the USA air transport network by progressively removing nodes to observe the size reduction of the giant component, simulating random airport failures and targeted attacks. Similarly, Xu and Chopra (2022) used a four-stage resilience cycle to evaluate the resilience of the Hong Kong Metro Rail, comprising preparedness, robustness, recoverability, and adaptation. In this cycle, the system is first prepared for disruption during the preparedness stage. The disruption then occurs during the robustness stage. Following this, the system must recover to normal operation in the recoverability stage. Finally, planning and changes occur during the adaptation phase to enhance future resilience. To assess preparedness, it was assumed that heterogeneously connected networks perform more resiliently in times of disruption. Consequently, preparedness was measured by how heterogeneous the network connections were. For this, a Gini coefficient of the flow-weighted betweenness of nodes was used.

2.3 Conclusion

The literature review examines a variety of studies on the application of network analysis to regional air transport networks, which focus particularly on their structure and evolution. The works highlight the considerations in assembling the network representation, with differing configurations such as representing the network as directed or undirected. Common key concepts in understanding network structure include analysing degree distributions and identifying nodes with high centrality. Additional methods, such as community detection and k-core decomposition, also prove useful in some cases.

When studying the evolution of these networks, common aspects that are tracked over time include the average shortest path, average clustering coefficient, degree assortativity, the number of nodes and edges, and the network's core. These aspects of structure and evolution inform the network analysis methodologies used in this dissertation.

Research on the AATN illustrates the region's importance; consequently, a method of applying network analysis techniques to this region can contribute an additional perspective to the existing literature. An approach of using a finer resolution of time period to analyse the effects of the COVID-19 pandemic to the air transport network is adopted, allowing for more sensitivity in tracking effects of events that occur over a short space of time.

Other potential avenues considered, but not implemented, for this study include time-dependent network analysis, although it is deemed infeasible for this dissertation, and resilience analysis, which could highlight geopolitical dependencies within the AATN, but is excluded due to time constraints. Although network analysis lends itself well to predictive modelling, it falls outside the theme of this dissertation. Following the literature

review, the next chapter discusses the data sources, processes, and considerations involved in assembling the networks.

Chapter 3

Data

This chapter outlines the data used in this study, utilising airline schedule data as the primary data source. Section 3.1 provides background information on the sources and details the data cleaning process. Section 3.2 then explains how the transport networks are assembled, which are required for the analyses conducted in the study. Lastly, Section 3.3 closes the chapter with a brief conclusion.

3.1 Airline Schedules Data

Two sources of data are used for the analysis. The primary source is the Official Airlines Guide (OAG) (*Airline Schedules Data* 2024), where data on global passenger airline schedules are used. The secondary source is an open-source dataset from Our Airports (Megginson, 2024), where the airports dataset is used to associate geographic coordinates with the airports in the schedules. The OAG airline schedule data was selected for its availability to the study and comprehensive coverage. The Our Airports dataset, although open source, was also chosen due to its accessibility and because it contains information on all the airports relevant to the study.

The airline schedules data spans from 2015 to 2023 for intra-African schedules and from 2016 to 2023 for global schedules. This range is limited by data availability, where earlier years could enhance understanding of the network's evolution. In the raw airline schedules data, each flight is detailed with its flight number, operational period, days of the week it will operate, origin and destination airports, departure and arrival times, and the operating airline.

The OAG schedules database is updated when airlines submit their planned schedules to OAG through a Standard Scheduled Message. OAG then in turn distributes the schedules to all industry stakeholders, such as Global Distribution Systems (GDS) and booking and reservation systems. OAG's system handles thousands of schedule changes daily from more than 900 airlines (*Airline Schedules Data* 2024), and the published schedules dataset is refreshed weekly. For this study, the dataset was generated on 18 March 2024.

In the raw airline schedules data, each flight is detailed with its flight number, effective dates of the schedule, days of the week it will operate, origin and destination airports,

departure and arrival times, and the operating airline. The fields used from the flight schedules data include departure and arrival airport, aircraft seat capacity, effective dates of the schedule, and the airport's city, country and region.

For this research, the schedules consider route segments and not routes with intermediary stops. There are cases where route segments are flown, but passengers may not be able to book tickets for the segment exclusively. For example, in 2023, Eurowings Discover (a German airline) flew a route segment between Windhoek Hosea Kutako International and Kruger Mpumalanga International Airport. Passengers could not book tickets for this segment (Pearson, 2022). Instead, passengers would book for trips between Frankfurt and Kruger Mpumalanga International Airport, which included a stopover in Windhoek. These segments cannot be identified within the schedules data, adding some inaccuracy to the analysis. For the analysis concerning the African Air Transport Network (AATN), only intra-African flight segments are considered, both international and domestic. In reality, the African network is interconnected with the global network, and airports located outside of Africa may influence the connectivity, such as shortest path, between African airports.

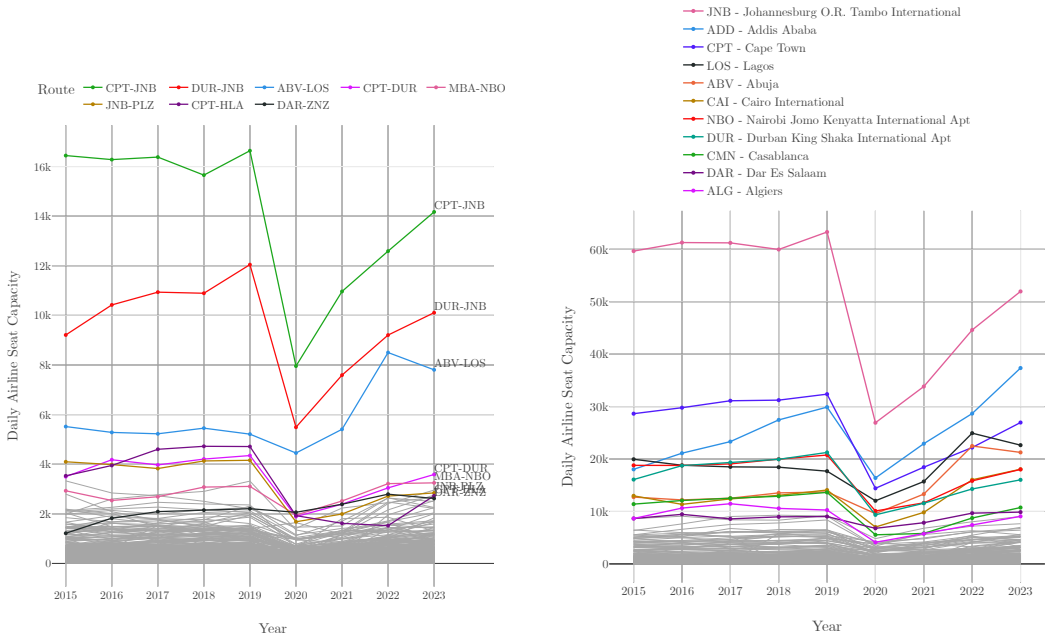
To exclude temporary routes and charter flights, flight segments with fewer than three flights in a calendar month are removed. Additionally, it is assumed that airlines have correctly published their schedules.

To determine which airports to include as part of Africa, the selection is based on the continental allocation of African countries according to both the International Air Transport Association (IATA) and the United Nation's M49 standard classification (Nations, 2024), which are found to be in alignment.

Figure 3.1 provides the trends of average daily intra-African seat capacity for top airport pairs (A) and top airports (B). An airport pair includes flights in either direction between the airports, and aggregations for airports include both flights arriving to and departing from the given airport. These insights can be extracted from the airline schedules data by conventional methods and without the need for network analysis tools and techniques. As can be clearly seen in (A) and (B), there is generally a sharp decline in airline seat capacity in 2020 due to the COVID-19 pandemic. Similarly to the findings of Bassens et al. (2012), the 'golden triangle' between Johannesburg, Cape Town, and Durban shows dominance in (A), with each airport pair of the triangle appearing in three of the top four airport pairs. In (B), it can be observed how Addis Ababa's intra-African seat capacity steadily grows over time despite setbacks caused by the COVID-19 pandemic.

3.2 Assembling the Network

In order for the airline schedules data to be prepared for network analysis, a network must be assembled in the form of nodes and edges. A network can be represented as an adjacency matrix, which is a square matrix in which the elements represent the relationships between the starting and ending nodes. An undirected network results in a symmetrical matrix.



(A) (B)

FIGURE 3.1: Daily airline seat capacity within the African Air Transport Network per year by (A) Top airport pairs (B) Top airports

As noted in Chapter 2, studies on air transport networks have varied in node representation: some used cities (Guimerà et al., 2005; Dai et al., 2018; Wang et al., 2011; Jia et al., 2014; Rocha, 2009), combining multiple city airports as single nodes, while others used airports (Song and Yeo, 2017; Cardillo et al., 2013; Cheung and Gunes, 2012; Bagler, 2008). In the AATN, cities have at most two long-term airports, as seen in Johannesburg, Nairobi, and Windhoek. During the study, Dakar shifted activities from Dakar Leopold Sedar Senghor to Dakar Blaise Diagne International. Temporary minimal routes are found at secondary airports in Durban, Monrovia, and Alexandria. Due to the rarity of multi-airport cities, the network uses airports as nodes to capture these nuances.

Given that network edges represent scheduled flights between airport pairs, as flights have a directional nature from origin to destination, air transport networks can be directed. However, most airline routes are reciprocal, as aircraft typically return to base. The annually aggregated AATN from 2015 to 2023 has an average reciprocity of 93.8%, indicating most routes are bidirectional. An undirected network captures this reciprocity but may introduce some inaccuracies from non-reciprocating routes. Bagler (2008) indicated that an undirected representation of the network allows for additional analyses to be performed on an air transport network. Furthermore, several of the reviewed studies considered undirected representations of their respective air transport networks (Guimerà et al., 2005; Song and Yeo, 2017; Dai et al., 2018; Cardillo et al., 2013; Wang et al., 2011; Xu and Harriss, 2008; Bagler, 2008; Rocha, 2009). Given the prevalence to represent networks as undirected and the high reciprocity of the AATN, this study primarily uses an undirected network representation of the AATN.

Airline schedules have temporal aspects, with routes flown between airports at various times and days. In assembling the network, routes are aggregated into undirected airport pairs for specific periods. This study primarily uses annual periods, supplemented with monthly periods for more detailed analyses. The annual aggregation choice aligns with past studies of network evolution, while monthly aggregation allows for a more sensitive analysis in certain scenarios. Shorter time period aggregations are possible, such as weekly, and can reveal more nuanced insights, however this study takes a longer term view on studying the evolution of the network over multiple years. Some temporal details, such as layovers, are not captured in this representation.

Figure 3.2 shows the 2023 AATN network with nodes as points and edges as faint blue lines geographically represented. The legend identifies the IATA region for each airport per OAG, dividing Africa into Central/Western, Eastern, North, and Southern regions. Table 3.1 summarises key details of the 2023 AATN.

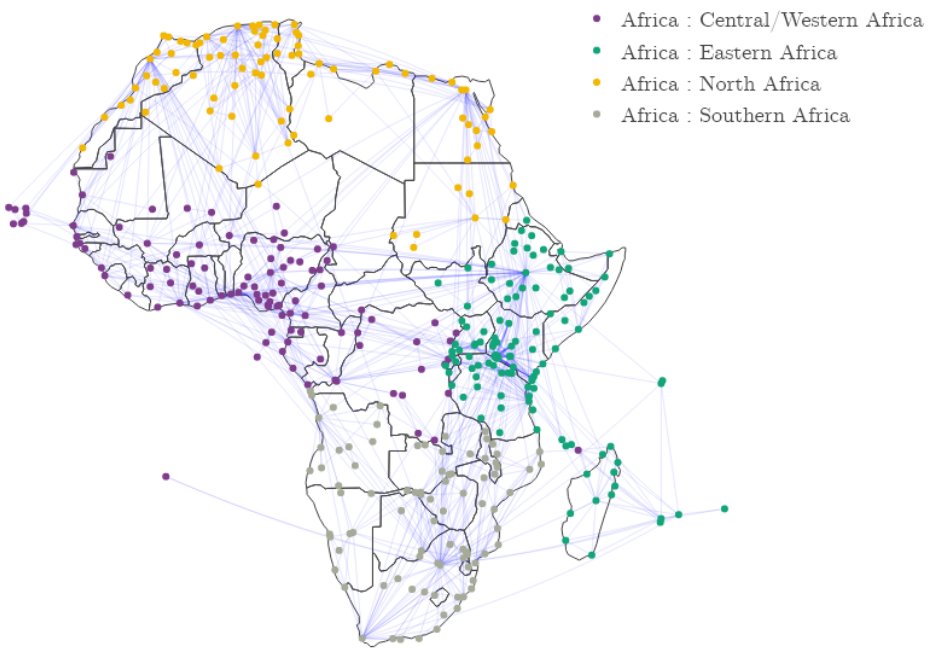


FIGURE 3.2: Region allocation of Airports according to IATA Regions. Airports and Routes according to the 2023 AATN.

3.3 Conclusion

As described in this chapter, the air traffic networks are assembled primarily from airline schedules data. The network representation of the AATN is not an exact representation of the actual network, where some aspects of the actual network are simplified. Some of the key design elements to the network models used in this study include limiting the

Table 3.1: Basic information about the 2023 AATN model

Attribute	**Value**
Year	2023
Number of Nodes	393
Number of Edges	1 124
Average Daily Seat Capacity	280 408
Average Daily Number of Flights	3 021.5
Number of Countries	57
Number of Components	1

network to include only intra-African routes, representing airports as nodes, aggregating active route periods to an annual level and also monthly level where required, and primarily using undirected edges. These assembled networks are then used to apply network analysis techniques as detailed in the methodology chapter that is addressed next.

Chapter 4

Methodology

In this chapter, the research methodology is presented, utilising the networks generated as detailed in Section 3.2. The methodology is divided into three main sections that align with the dissertation's objectives. Section 4.1 describes the application of network analysis tools and techniques to illustrate the structure of the African Air Transport Network (AATN). Section 4.2 addresses the methods for detailing its evolution. Further, Section 4.3 discusses the methods employed to evaluate the network's recovery after the COVID-19 pandemic. The chapter concludes with Section 4.4.

4.1 Network Structure

The AATN's structure considers how airports are interconnected topologically, mostly focusing on the year 2023 as a cross-section in time. First, the network's degree distribution is examined, followed by an analysis of key airports and routes. Next, community detection is explored within the AATN. Lastly, the network core is analysed using k-core decomposition. This section presents the methodologies for these components.

4.1.1 Degree Distribution

The degree distribution is a frequency distribution of the degrees of nodes. Newman (2010) described the degree distribution as one of the most fundamental properties of a network and a defining characteristic of its structure. It is common for a network's degree distribution to be heavy-tailed, particularly following a power-law distribution. Past studies of air transport networks, such as Dai et al. (2018) and Bagler (2008), have found that their respective networks follow power-law distributions. Networks with power-law distributions are referred to as 'scale-free'. An exploration into the AATN's degree distribution confirms whether this network possesses these structural characteristics as are found in other air transport networks and other types of networks.

A power-law distribution follows the formula:

$$p_k = Ck^{-\alpha}, \tag{4.1}$$

where p_k is the fraction of nodes given degree k, C is a constant and α is referred to as the exponent of the power-law.

Two common tools are used to help visualise power-law distributions: log-log histograms with logarithmic binning and Complementary Cumulative Distribution Functions (CCDFs). Logarithmic binning uses an increasing bin width, which typically follows an interval of $a^{n-1} \leq k \leq a^n$ for the n^{th} bin, where k is the degree and a is a selected constant to render a suitable visualisation. A CCDF, similar to a cumulative distribution function, can be expressed by:

$$P_k = \sum_{k'=k}^{\infty} p_{k'}, \tag{4.2}$$

where P_k is the probability that a node has degree of k or higher.

The AATN degree distribution is analysed for 2023, aggregated both for the entire year to compare with other air transport networks, and monthly to explore seasonal effects on the network. Degree distributions frequently exhibit a power-law behaviour predominantly in their tails, rendering it suitable to identify a power-law by focusing on the tail (Newman, 2010). A method is used to exclude the head of the distribution from the analysis.

The power-law exponent α indicates the slope of the distribution when plotted on log-log axes. α typically follows a range of $2 \leq \alpha \leq 3$. The value of α in the tail of a discrete distribution can be approximated from:

$$\alpha = 1 + N \left[\sum_i \ln \left(\frac{k_i}{k_{\min} - \frac{1}{2}} \right) \right]^{-1}, \tag{4.3}$$

where N is the number of nodes with degree k greater than or equal to $k_{\min}$. The sum is performed for $k \geq k_{\min}$, effectively ignoring the head of the distribution. Newman (2010) advises that the approximation requires $k_{\min}$ to be greater than about 6 to be accurate. An appropriate value for $k_{\min}$ can be determined by minimising the Kolmogorov-Smirnov distance, expressed as:

$$D = \max_{k \geq k_{\min}} |P(k) - S(k)|, \tag{4.4}$$

where $S(k)$ is the CCDF of the observed data, and $P(k)$ is the CCDF of the fitted power-law distribution, meaning the Kolmogorov-Smirnov distance is the maximum distance between the two CCDFs. The selected $k_{\min}$ is the value that results in the lowest Kolmogorov-Smirnov distance (Alstott et al., 2014; Clauset et al., 2009; Newman, 2010).

The error of α is estimated from:

$$\sigma = \frac{\alpha - 1}{\sqrt{N}}. \tag{4.5}$$

The power-law exponent and corresponding error are estimated for each of the degree distributions examined to evaluate their conformity to expectations and changes throughout the year.

Alternative distribution fits are evaluated to verify if the power-law distribution aptly represents the AATN's degree distribution in 2023. The power-law fit is compared to an exponential distribution and an exponentially truncated power-law distribution. The probability density functions for these fits are given in Equations 4.6 and 4.7, respectively.

$$p_k = Ce^{-\lambda k} \tag{4.6}$$

$$p_k = Ck^{-\alpha}e^{-\lambda k} \tag{4.7}$$

In Equations 4.6 and 4.7 C, α, and λ are constants. The exponentially truncated power-law distribution is considered to be a nested distribution of the power-law distribution, as it retains characteristics of the power-law distributions, but is exponentially cut off at its tail (Clauset et al., 2009).

Analysing the degree distribution provides insights into the network's structure. The analysis can confirm whether the distribution conforms to expectations, notably with a heavy-tailed shape, which indicates that there are a large number of nodes with relatively small degree in the network and a small number of nodes with relatively large degree acting as hubs in the network. The method for identifying and characterising key airports and routes is presented next.

4.1.2 Key Airports and Routes

Network analysis is used to identify key airports and routes in the AATN by evaluating attributes that indicate their importance and structure. Tracking these attributes over time aids in understanding the AATN's evolution, using annual data from 2015 to 2023.

Key airports are identified by node degree and daily airline seat capacity. Section 5.1.2 shows that airports qualify as key by having a degree of 30 or higher, or capacity of over 15 000 seats per day in the 2023 AATN.

The following node measures and attributes are determined for the airports:

1. Degree
2. Daily airline seat capacity
3. Daily number of flights
4. Betweenness centrality
5. Eigenvector centrality
6. Closeness centrality
7. Clustering coefficient
8. Eccentricity

Degree, betweenness centrality, eigenvector centrality, and closeness centrality are all centrality measures designed to determine how important a node is to a network and are explained further below.

The daily airline seat capacity is a metric derived from the edge weight and is an indication of the utility of an airport. Therefore, relationship plots between the measures and daily airline seat capacity are generated, with daily airline seat capacity on the horizontal axis for comparison. The longitudinal trends from 2015 to 2023 are then examined to observe the evolution of the airports.

In assessing key routes in the 2023 AATN, the daily airline seat capacity, edge betweenness centrality and the great circle diameter of the airport pairs are investigated. Given that the daily airline seat capacity is an edge weight and is a measure of the strength of the edge, its relationship with edge betweenness centrality is investigated. Key routes are identified as those having edge betweenness centrality higher than 0.025 or daily airline seat capacity higher than 3 200 seats per day.

The daily airline seat capacity of the nodes and edges is shown in a combined plot, to geographically represent the relationship between the nodes and edges, and to comment on how these are related to key airports and routes. This is repeated for node and edge betweenness centrality.

The newly introduced network measures used to examine the key airports and routes in the AATN are discussed in more detail below.

4.1.2.1 Node Betweenness Centrality

Node betweenness centrality is the proportion of shortest paths between all node pairs that pass through a given node. The formula for node betweenness centrality is expressed by the following equation:

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}, \tag{4.8}$$

where $\sigma(s,t)$ represents all possible shortest paths starting at s and ending at t, and $\sigma(s,t|v)$ represents all these shortest paths that pass through node v excluding where $v \in \{s,t\}$. If $s = t$ then $\sigma(s,t) = 1$. V represents all possible nodes in a network (Brandes, 2008).

4.1.2.2 Eigenvector Centrality

Eigenvector centrality addresses a limitation of degree centrality, where nodes with low degree can still have a high centrality value by virtue of the node being closely connected to other nodes with high degree. The eigenvector centrality is determined as follows:

$$\lambda \mathbf{e}_i = \sum_j \mathbf{A}_{ij} \mathbf{e}_j, \tag{4.9}$$

where $\mathbf{A}_{ij}$ is the network's adjacency matrix value at nodes i and j, and $\mathbf{e}$ is the eigenvector for the largest positive eigenvalue λ (Bonacich, 1987). $\mathbf{e}_i$ gives the centrality value of node i.

4.1.2.3 Closeness Centrality

Closeness centrality is effectively the inverse of the average shortest path to every other node in the network, acting as an indicator of overall efficiency. Therefore, a node has high centrality when it is close to each of the other nodes in the network. The closeness centrality of a node is given by the following equation:

$$C(u) = \frac{n-1}{N-1} \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}, \tag{4.10}$$

where $d(v,u)$ is the shortest-path between v and u, N is the number of nodes in the network, and $n-1$ is the number of nodes that are reachable from u. Equation 4.10 can be applied over networks with more than one component (Wasserman and Faust, 1994).

4.1.2.4 Clustering Coefficient

The clustering coefficient measures the proportion of pairs of a node's neighbours that are also connected to each other and is calculated for a node by the following:

$$c_u = \frac{2T(u)}{\deg(u)(\deg(u)-1)}, \tag{4.11}$$

where $T(u)$ is the number of triangles through node u and $\deg(u)$ is the degree of node u (Saramäki et al., 2007).

4.1.2.5 Eccentricity

The eccentricity of a node is the length of the longest of the shortest paths from the given node to all other possible nodes within the network, thus providing an indication of how efficiently the node is connected. The eccentricity calculation requires that a node must be part of the largest component, therefore in years with multiple components, some nodes do not have an eccentricity value.

4.1.2.6 Edge Betweenness Centrality

Edge betweenness centrality is a similar concept to node betweenness centrality, which concerns the number of shortest paths which pass through a given edge. The formula for edge betweenness centrality can be expressed as (Brandes, 2008)

$$C_B(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)}, \tag{4.12}$$

where $\sigma(s,t|e)$ represents all these shortest paths that pass through edge e.

4.1.2.7 Great Circle Distance

The great circle distance between two locations assumes a spherical shape for the earth, and can be determined using the Haversine formula shown in the following equation:

$$d = 2 \cdot \arcsin^2 \left(\sqrt{\sin^2 \left(\frac{\lambda_B - \lambda_A}{2} \right) + \cos(\lambda_A) \cos(\lambda_B) \sin^2 \left(\frac{\mu_B - \mu_A}{2} \right)} \right) \cdot r, \tag{4.13}$$

where λ is the latitude and μ is the longitude for the given location A or B, while r is the radius of the sphere for which 6 371 009m is used (Gade, 2010).

4.1.3 Community Detection

Girvan and Newman (2002) described communities in networks as subsets of nodes that are densely connected by edges, but where connections between subsets are less dense. Community detection provides insight into the structure of a network, based on these relationships. This analysis applies community detection to the 2023 AATN; however, some intra-year analysis is also provided in Appendix A.

A common method for community detection that is performed on air transport networks is the Louvain method as demonstrated by Jia et al. (2014), Guo et al. (2019), and Diop et al. (2021). The Louvain method is used in this study for continuity.

The Louvain algorithm aims to maximise network modularity by optimising node community assignments (Blondel et al., 2008). In the first stage of the algorithm, every node is assigned its own community. Next, the communities are considered one by one in random order and are merged with a neighbouring community that would result in the highest increase in network modularity. However, if merging the communities does not result in increased network modularity, then the communities are not merged. This process is repeated until a minimum threshold in modularity increase is reached.

The modularity quality function used is given in Equation 4.14, where K_c is the sum of the degree of the nodes in community c and m is the total number of edges in the network. e_c is the actual number of edges in community c and γ is a resolution parameter (Reichardt and Bornholdt, 2006; Traag et al., 2019).

$$Q = \frac{1}{2m} \Sigma_c \left(e_c - \gamma \frac{K_c^2}{2m} \right) \tag{4.14}$$

The γ resolution parameter, absent in the original Louvain algorithm, is used by Traag et al. (2019) to adjust the number of detected communities, where higher values lead to more communities. Setting this parameter to 1 replicates the original algorithm's results. γ can be described as a threshold where communities should have a density of at least γ, while the density between communities should be lower than γ (Traag et al., 2019). Community detection plots are generated for increasing values of the resolution parameter, providing a range of community plots of different sizes of communities.

For community detection on the AATN, the network edges are weighted by airline seat capacity, since it provides a measure of magnitude of how connected an airport pair is by the flights between them. Therefore, busier routes generate a stronger community link than less busy routes.

Since the Louvain algorithm uses a random order of nodes to allocate them into communities, different runs of the algorithm can yield different results. However, the same result is often achieved. Given that the air transport network is a relatively small network for the Louvain algorithm to process, it is feasible to perform 100 runs of the algorithm to ensure repeatability. The most commonly occurring community allocation result from these runs is then used. This approach is applied in each case where community detection is performed on a given network at a given resolution.

Community detection provides insights into the community structure of the AATN at varying levels of resolution. The method for analysing the AATN's core is presented next.

4.1.4 K-core Decomposition

In past analyses of air transport networks, such as Guo et al. (2019), there have been attempts to determine how the core of a network is structured. The k-core decomposition method for determining the core of the network was used by Dai et al. (2018) and Wang et al. (2014), with the latter noting that k-core decomposition has not been widely used in other air transport network analyses.

k-core decomposition produces a sub-network of a given network, which represents its 'core' (Batagelj and Zaversnik, 2003). The k-core sub-network is the maximal sub-network where all nodes have degree of at least k. The core revealed by k-core decomposition is effectively the most interconnected set of nodes within the network, where the strength of the connection is not considered, but rather the number of connections each node has within the core.

k-core decomposition is performed on the 2023 AATN to reveal the sub-networks generated with increasing values of k starting at $k = 3$, until the maximum possible k value is reached. Community detection is applied to the sub-networks generated, using a resolution parameter $\gamma = 1$[1] and weighted by seat capacity, to additionally reveal some community structures within the sub-networks.

The number of nodes, number of edges, and the ratios of the number of edges and airline seat capacity compared to that of the original network are calculated to determine each sub-network's contribution to these. k-core decomposition is, therefore, used to analyse the core of a network, as an element of the network's structure.

This is the final analysis aimed at illustrating and describing the structure of the network; the next section focuses on methods for detailing the evolution of the network.

[1] In order to apply the original Louvain algorithm.

4.2 Evolution of the Network

Although the methods discussed in this section reveal much about the structure of the AATN, their focus is on how the structure evolves over time from 2015 to 2023 on an annual basis. The effects of intra-year seasonality could be insightful as airlines adjust their networks due to seasonal demand; however, this has not been included in the scope of this research. This section first presents a method for longitudinal analysis using network-wide indicators applied to the AATN, followed by examining the evolution of the core of the AATN.

4.2.1 Longitudinal Analysis of Network-Wide Indicators

Network-wide indicators are used to express a given characteristic of a network with a single value. This section considers a longitudinal analysis of these network-wide indicators over the years 2015 to 2023, using annual aggregations of the AATN.

Some of these network-wide indicators can be benchmarked against expected values from a randomly generated network with the same number of nodes and edges. A random network generator, where the probability of any new given pair of nodes being connected is equal, is used to generate 1 000 iterations of random networks for each year. The mean results of these metrics serve as benchmarks.

For average shortest path and network diameter metrics, the algorithms must be applied to a single component where all nodes are interconnected. For the AATN, a single component is found within the network for every year except for 2016, where there is an isolated network component between Namibian domestic airports. In this case, the largest component contains 98.97% of the nodes. For the random case, separated components are common, with the mean largest component containing between 99.18% and 99.68% of nodes over the 2015 to 2023 period. In general, the largest component is used to determine the average shortest path and the network diameter.

The metrics analysed include number of nodes and edges, daily airline seat capacity, mean degree, network density, power-law exponent of the degree distribution, average shortest path, diameter, average clustering coefficient, degree assortativity coefficient, Gini coefficient of node degree, and the degree of the largest degree node. The new concepts introduced are detailed below.

4.2.1.1 Mean Degree and Network Density

Two metrics that describe the proportionality between nodes and edges within a network are the mean degree of a network and the network density.

The formulae for mean degree and network density are provided in the following equations, respectively:

$$c = \frac{2m}{n}, \tag{4.15}$$

and

$$\rho = \frac{2m}{n(n-1)}, \tag{4.16}$$

where m is the number of edges in the network and n is the number of nodes in the network (Newman, 2010).

4.2.1.2 Power-law Exponent

The power-law exponent provides insight into the slope of a power-law degree distribution of a network and can assist in characterising the evolution of the network when evaluated over a period of time. The power-law exponent, for a given $k_{\min}$, is determined using Equation 4.3, as discussed in Section 4.1.1.

4.2.1.3 Average Shortest Path and Network Diameter

As a metric to characterise the efficiency of a network, the average shortest path is calculated by determining the average value of the lengths of the shortest paths between all possible node pairs in a network.

The average shortest path is determined by the following equation:

$$a = \sum_{\substack{s,t \in V \\ s \neq t}} \frac{d(s,t)}{n(n-1)}, \tag{4.17}$$

where V is the set of all nodes in the network, $d(s,t)$ is the number of edges on the shortest path from node s to node t and n is the number of nodes in the network (Dai et al., 2018).

Similar to the average shortest path, the diameter of the network is also a measure of the efficiency of a network. The diameter is the length of the largest shortest path between any two nodes within a network.

4.2.1.4 Average Clustering Coefficient

The average clustering coefficient takes the average of the clustering coefficient, described in Section 4.1.2.4, over all nodes in the network.

The average clustering coefficient is a measure of how interconnected a network is. Although transitivity is also a measure of network interconnectivity, in this study only the average clustering coefficient is used for continuity since it has been used more commonly in previous regional air transport network studies, as referred to in Chapter 2.

4.2.1.5 Degree Assortativity Coefficient

The degree assortativity coefficient measures the correlation between the degrees of connected nodes. It is determined using the following equation:

$$r = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b}, \tag{4.18}$$

where, given the distribution of nodes with degree x and y, e_{xy} is the fraction of all edges that join nodes with degree x and y. The terms a_x and b_y are the fraction of edges that start and end with a node with degree x and y respectively. Finally, σ_a and σ_b are the standard deviations of the distributions of a_x and b_y respectively (Newman, 2003).

4.2.1.6 Gini Coefficient and Largest Degree

The Gini coefficient and the degree of the largest degree node both relate to the degree distribution of the network. The former is a metric of inequality, with the equation for the Gini coefficient provided by the following equation:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \bar{x}}, \tag{4.19}$$

where x is the degree of a given node, $\bar{x}$ is the average degree and n is the number of nodes (Sen, 1997).

This concludes the methodology of the longitudinal analysis of network-wide indicators; the methodology of the evolution of the core is described next.

4.2.2 Evolution of the Core

Since k-core decomposition, as described in Section 4.1.4, is able to reduce the amount of information of a network down to a core, it is feasible to analyse the evolution of the core over a period of time. Therefore, k-core decomposition is used to generate sub-networks for the maximum possible k value for each year from 2015 to 2023, which are then examined.

As in Section 4.1.4, Louvain community detection is applied to each sub-network with resolution parameter $\gamma = 1$ and weighted by seat capacity to reveal some community structure within the sub-networks.

Having outlined the use of k-core decomposition to detail the AATN's evolution, methods for evaluating its post-COVID-19 recovery will now be discussed.

4.3 COVID-19 Pandemic Recovery

The analysis of the AATN's evolution during the study period reveals a substantial impact of the COVID-19 pandemic. Therefore, it is appropriate to investigate how the network

changes from pre-pandemic to post-pandemic. This section first investigates the commonalities and differences in the AATN between these periods. This is followed by quantifying the degree to which the network changes over the period using a similarity measure.

4.3.1 Pre-Pandemic vs End of Recovery: Commonality and Differences

Network analysis set operations[2] provide useful tools to visualise commonalities and differences between networks. The 2019 and 2023 AATNs are compared to understand the changes from pre-pandemic to post-pandemic. The visualisations include the intersection of the 2019 and 2023 AATNs, where only edges common to both networks are retained, as defined by Equation 4.20. The differences in pre-pandemic and 'end of recovery phase' are visualised by subtracting the common edges of the 2019 AATN from the 2023 AATN and vice versa, as defined by the equations 4.21 and 4.22 respectively.

Let e_{2019} and e_{2023} represent the sets of edges for the 2019 and 2023 AATNs, respectively. The intersection and differences between these networks are defined as follows:

$$e_{2019} \cap e_{2023}, \tag{4.20}$$

$$e_{2023} \setminus e_{2019}, \tag{4.21}$$

and

$$e_{2019} \setminus e_{2023}, \tag{4.22}$$

where these visualisations help understand changes in the structure of the AATN from pre-pandemic to post-pandemic.

4.3.2 Network Similarity

To compare the AATNs structure before, during and after the COVID-19 pandemic, a similarity measure, as used by Jia et al. (2014) is adopted. This similarity measure is used to compare similarities in a network's structure between different time states. The similarity measure first determines a similarity value for a given node by use of the following equation:

$$S_i(t) = \frac{|A_i(t_1) \cap A_i(t_2)|}{|A_i(t_1) \cup A_i(t_2)|}, \tag{4.23}$$

where A_i is the set of edges connected to node i at a given time state. The numerator in Equation 4.23 represents the intersection of the sets A_i in the two different time states, while the denominator is the union. The overall similarity measure of the network is the average similarity value of each node in the union of the two network states.

Trends of various metrics, which include the similarity measure, number of flights per day, number of seats per day, number of edges, and mean degree, are analysed. These

[2] Using set theory.

metrics are plotted on an annual time scale from 2015 to 2023 and are indexed against the period prior to the COVID-19 pandemic, specifically 2019, for comparison of changes in the various attributes. In addition, a plot with higher resolution on the time scale is prepared, where months in a given year are indexed against the same month in 2019. This approach highlights intra-year changes and accounts for recurrent seasonal flights. The trends of the similarity measure and number of edges indexed are used to compare the AATN with the global network, examining the years from 2016 to 2023.

4.4 Conclusion

The methodology chapter is structured to address the three research objectives separately, utilising various network analysis tools and techniques throughout the study.

In Section 4.1, the methods to illustrate and describe the structure of the AATN using a range of network analysis techniques are presented. These techniques include analysing the network's degree distribution, identifying and characterising key airports in the network, examining communities detected within the network, and examining the network's core.

Section 4.2 presents methods for detailing the evolution of the AATN by longitudinally analysing network-wide indicators and the network's core from 2015 to 2023.

Finally, Section 4.3 presents the methods used to evaluate the recovery of the AATN after the COVID-19 pandemic. This involves using network analysis operations to assess the commonalities and differences between the pre-pandemic and post-pandemic states of the network, and by comparing changes in the network structure relative to the period prior to the COVID-19 pandemic using network-wide measures.

Having outlined the methodology, the results and discussion chapter is presented next.

Chapter 5

Results and Discussion

The structure of the results and discussion mirrors that of the methodology, with three main sections aligned to the study's research objectives on the African Air Transport Network (AATN). Section 5.1 illustrates the network's structure, Section 5.2 details its evolution, and Section 5.3 evaluates the network's recovery after the COVID-19 pandemic. The chapter concludes with a summarised discussion of the key findings in Section 5.4.

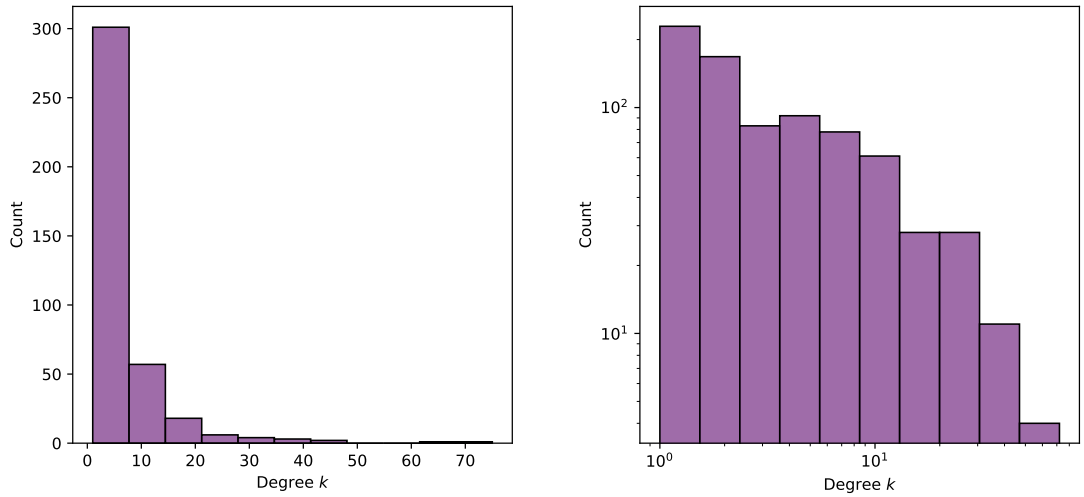
5.1 Network Structure

As outlined in Chapter 4, various features can be explored to analyse network structures. This section examines the degree distribution, analyses key airports and routes, applies Louvain community detection, and concludes with k-core decomposition analysis of the network's core.

5.1.1 Degree Distribution

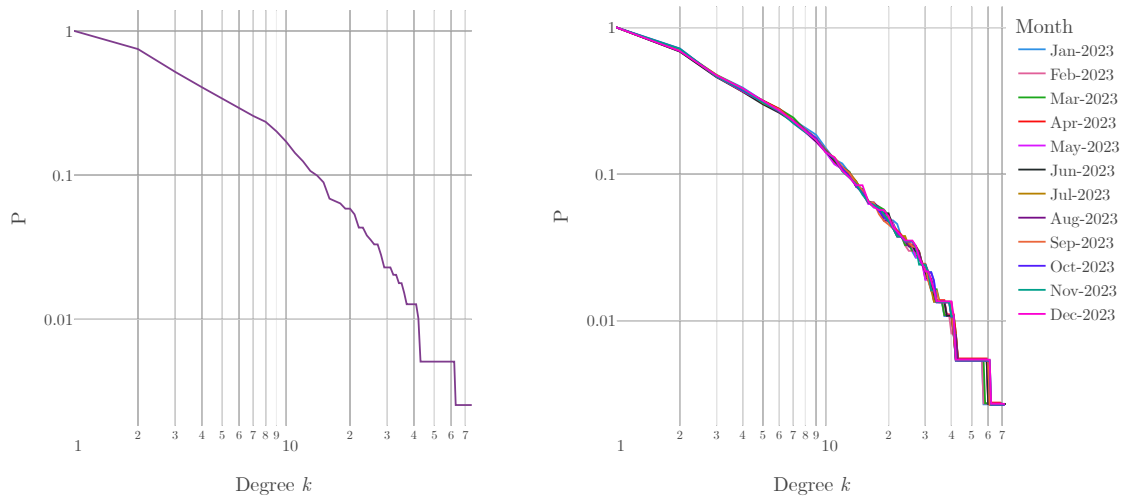
The results of the degree distribution analysis include various visualisations of the ATTN's degree distribution: Figure 5.1 provides the degree distribution of the 2023 AATN. Plot (A) visualises the distribution in linear scales, while plot (B) uses log-log scales with logarithmic binning, where both plots include 10 bins. Next, the Complementary Cumulative Distribution Function (CCDF) plots of the 2023 AATN and the monthly cases of the AATN in 2023 are shown in Figure 5.2. The cumulative value P indicates the fraction of nodes with degree higher than degree k, where the CCDF is plotted on log-log scales. Following this, Figure 5.3 displays a fitted CCDF of the power-law distribution fit generated for the 2023 AATN. In determining the fit, minimising the Kolmogorov-Smirnov distance (see Section 4.1.1) yields $k_{\min} = 8$. CCDFs are also fitted for two other heavy-tailed distributions, namely the exponentially truncated power-law and exponential distributions. Finally, the power-law exponent and its error are presented for each month of the 2023 AATN in Table 5.1. The resulting power-law exponent α of the entire 2023 AATN is found to be 2.624 with an error σ of 0.169.

The analysis of the degree distribution of the AATN provides insight into the structure of the network and assists in describing it. From Figure 5.1, it is evident that the degree



(A) Linear scales (B) Log-log scales

FIGURE 5.1: Degree distributions of the 2023 AATN. (A) and (B) show degree distributions of the same network, but with different scales.



(A) (B)

FIGURE 5.2: CCDFs of the AATN for the year 2023. (A) shows the AATN aggregated by year. (B) shows the AATN aggregated by month.

distribution of the 2023 AATN is heavy-tailed, with many nodes having low degree and fewer nodes having high degree. As discussed in Chapter 2, a heavy-tailed degree distribution is commonly found across many different types of networks. Furthermore, the shape of the distribution in (B), when plotted on log-log scales, suggests that the distribution could take the form of a power-law distribution. The CCDFs in Figure 5.2 provide further evidence that the distribution is a power-law distribution by exhibiting a fairly straight line from end to end, and particularly in the tail of the distribution. From Figure 5.2 (B), it visually appears that there is not much variation in the CCDFs for the different months.

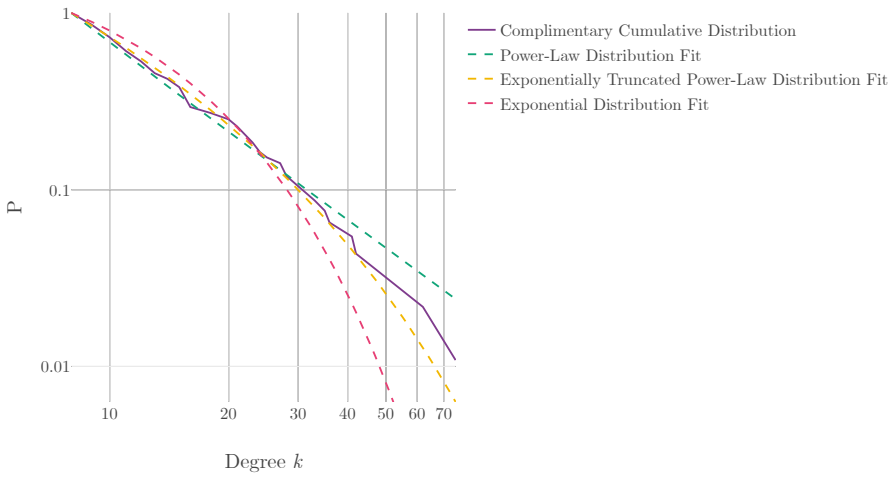


FIGURE 5.3: Fitted curves of the 2023 AATN degree distribution for $k_{\min} = 8$.

TABLE 5.1: Resulting exponent of power-law α together with error values, for the degree distributions by month.

Month	$k_{\min}$	α	σ
Jan-2023	9	2.75	0.211
Feb-2023	9	2.71	0.213
Mar-2023	7	2.53	0.162
Apr-2023	9	2.68	0.214
May-2023	8	2.56	0.185
Jun-2023	2	1.88	0.055
Jul-2023	2	1.88	0.055
Aug-2023	2	1.89	0.055
Sep-2023	9	2.70	0.209
Oct-2023	9	2.71	0.211
Nov-2023	9	2.68	0.210
Dec-2023	2	1.87	0.054

In Figure 5.3, it appears that the power-law distribution yields a better fit than the exponential distribution. However, the exponentially truncated power-law distribution appears to yield an even better fit than the power-law distribution. This can be expected, as the fitted function contains an additional degree of freedom. The power-law fit does, however, appear to provide a good fit to the distribution and can therefore be an appropriate fit to characterise the AATN's degree distribution. Additionally, the power-law exponent can be a useful measure to characterise the slope of the power-law distribution and, therefore, also the structure of the AATN. Since the degree distribution of the 2023 AATN follows a power-law distribution, it can be regarded as 'scale-free'.

The resulting power-law exponent ($\alpha = 2.624$) of the 2023 AATN fits within the expected range of $2 \leq \alpha \leq 3$ for network degree distributions (Newman, 2010). For the monthly cases, the power-law exponent α fluctuates between 1.87 and 2.75. For lower values of

α, the $k_{\min}$ values are lower, indicating that the inclusion of more of the head of the distribution may impact the results given that $k_{\min}$ goes below 6.

Further investigations into the distributions of other centrality measures can be made to provide further evidence of their heavy-tailed nature. Closeness centrality; however, is expected to follow a Gaussian distribution. With respect to the power-law exponent, future work could benefit from adapting the process of determining $k_{\min}$ to track changes in the power-law exponent throughout the months of the year, to make it less sensitive to subtle changes in the network. Methods to achieve this could include employing an alternative distance function in determining $k_{\min}$. Additionally, future research could include an analysis of the 'rich-club' coefficient, as explored in Cardillo et al. (2013), which could assist further in describing the network's structure.

The analysis of the AATN's degree distribution helps describe its structure, affirming the power-law distribution as a suitable fit for characterising the network's degree distribution. The analysis of key airports and routes is presented next.

5.1.2 Key Airports and Routes

This subsection presents the results and discussion first for the analysis of key airports, followed by the analysis of key routes, and finally geographic visualisations of nodes and edges and their importance in the network.

Figure 5.4 shows the relationships of various measures with daily airline seat capacity for airports within the 2023 AATN. The measures include: degree, betweenness centrality, eigenvector centrality, closeness centrality, daily number of flights, clustering coefficient, and eccentricity. The key airports identified are highlighted in the plots.

As described in Section 4.1.2, key airports are identified as airports with degree of 30 or higher or a daily airline seat capacity higher than 15 000 seats per day in the 2023 AATN. Figure 5.4 (A) shows a relationship between node degree and daily airline seat capacity for the 2023 AATN. The figure highlights the prominence of the key airports from the rest of the airports in terms of daily airline seat capacity and node degree. The key airports identified include: Johannesburg O.R. Tambo International, Addis Ababa, Cape Town, Lagos, Abuja, Cairo International, Nairobi Jomo Kenyatta International, Durban King Shaka International, Casablanca, Dar Es Salaam, and Algiers

Some further investigation into Figure 5.4 (A) shows that there is some correlation between airline seat capacity and degree; in fact, key airports could also be distinguished from one criterion alone, being a daily airline seat capacity higher than 9 050. The only airports which do not meet the criteria of having degree of 30 or higher are Cape Town and Durban King Shaka International, where Cape Town has a degree of 27 and Durban King Shaka International has a degree of 11.

Plots (B), (C), and (D) in Figure 5.4 reveal a correlation between airline seat capacity and node centrality in general. Johannesburg O.R. Tambo International, Cape Town, and

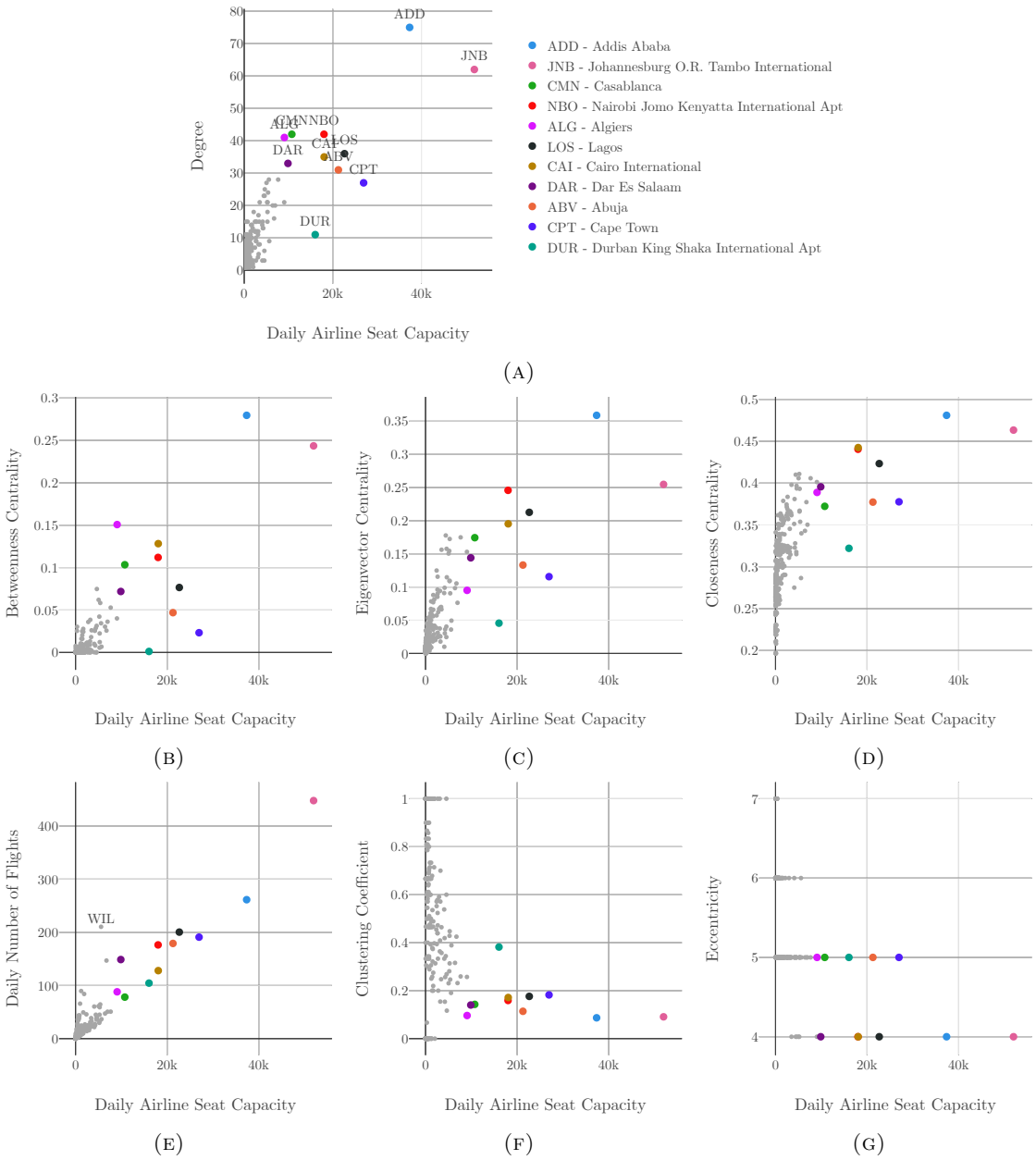


FIGURE 5.4: Relationships between various measures of key airports in the 2023 AATN and daily airline seat capacity. The measures include: (A) degree, (B) betweenness centrality, (C) eigenvector centrality, (D) closeness centrality, (E) daily number of flights, (F) clustering coefficient, and (G) eccentricity.

Durban King Shaka International show relatively low centrality for their seat capacity; this is likely due to much of the airports' seat capacity comprising of routes on the 'golden triangle', as referred to in Section 3.1.

Figure 5.4 (E) reveals a high correlation between the number of flights and airline seat capacity. However, Nairobi Wilson Airport [1] is an outlier due to numerous flights that are operated with small aircraft.

[1] Annotated as WIL on Figure 5.4 (E)

The clustering coefficient appears to be inversely proportional to the node degree, as shown in Figure 5.4 (F), where the clustering coefficient is a measure of how interconnected a node's neighbours are with each other. These findings are similar to those of Xu and Harriss (2008), where low degree cities were found to have high clustering coefficients. The three key airports with the highest clustering coefficients in 2023 are Durban King Shaka International, Cape Town, and Lagos respectively, with Durban visually standing out as an outlier, due to high seat capacity on the 'golden triangle' routes. However, the three airports with the lowest clustering coefficients are Addis Ababa, Johannesburg O.R. Tambo International and Algiers respectively. A lower clustering coefficient suggests that an airport can serve as a hub, as passengers between connected airports are less likely to have direct routes between them.

From Figure 5.4 (G), it appears that the eccentricity of airports in the 2023 AATN ranges between four and seven and between four and five for the key airports. The plot suggests that there is an inverse relationship between airline seat capacity and eccentricity, indicating that airports with higher airline seat capacity are more efficiently connected to the network.

Further analysis of the key airports is provided in Figure 5.5, which presents trends of the node measures. This, together with Figure 3.1 (B), which provides the trend of daily airline seat capacity, provides insights into the AATN's structure and evolution.

In terms of node centrality, Addis Ababa shows a gradual increase in degree, betweenness, eigenvector and closeness centrality from 2015 to 2023 as indicated in Figure 5.5 (A), (B), (C), and (D). By 2018, Addis Ababa has the highest centrality in the AATN in all four measures. Johannesburg O.R. Tambo International and Algiers both show a major increase in centrality in 2023, coinciding with the introduction of a route connecting these two distant airports as indicated in Figure 5.6 (B). The introduced route has the highest edge betweenness centrality in 2023, influencing both nodes' betweenness centrality and closeness centrality.

From 2015 to 2023, the largest movements in centrality are observed in four cities. Addis Ababa experiences the highest growth in both degree and betweenness centrality, while Cape Town sees the greatest increase in eigenvector and closeness centrality. Conversely, Nairobi Jomo Kenyatta International has the most severe reduction in degree and betweenness centrality, and Casablanca experiences the largest decrease in eigenvector and closeness centrality.

The closeness centrality calculation requires that the given nodes be part of the largest network component. As can be seen in Figure 5.5 (D), in 2016 there is more than one network component resulting in some nodes receiving a closeness centrality of zero.

The plot in Figure 3.1 (B) shows Johannesburg O.R. Tambo International consistently ranking highest in intra-African seat capacity. Cape Town moves from second in 2015 to third in 2023, while Addis Ababa rises, with a sharp trajectory, from fifth to second during the same period, overtaking Cape Town amid the COVID-19 pandemic. A decline

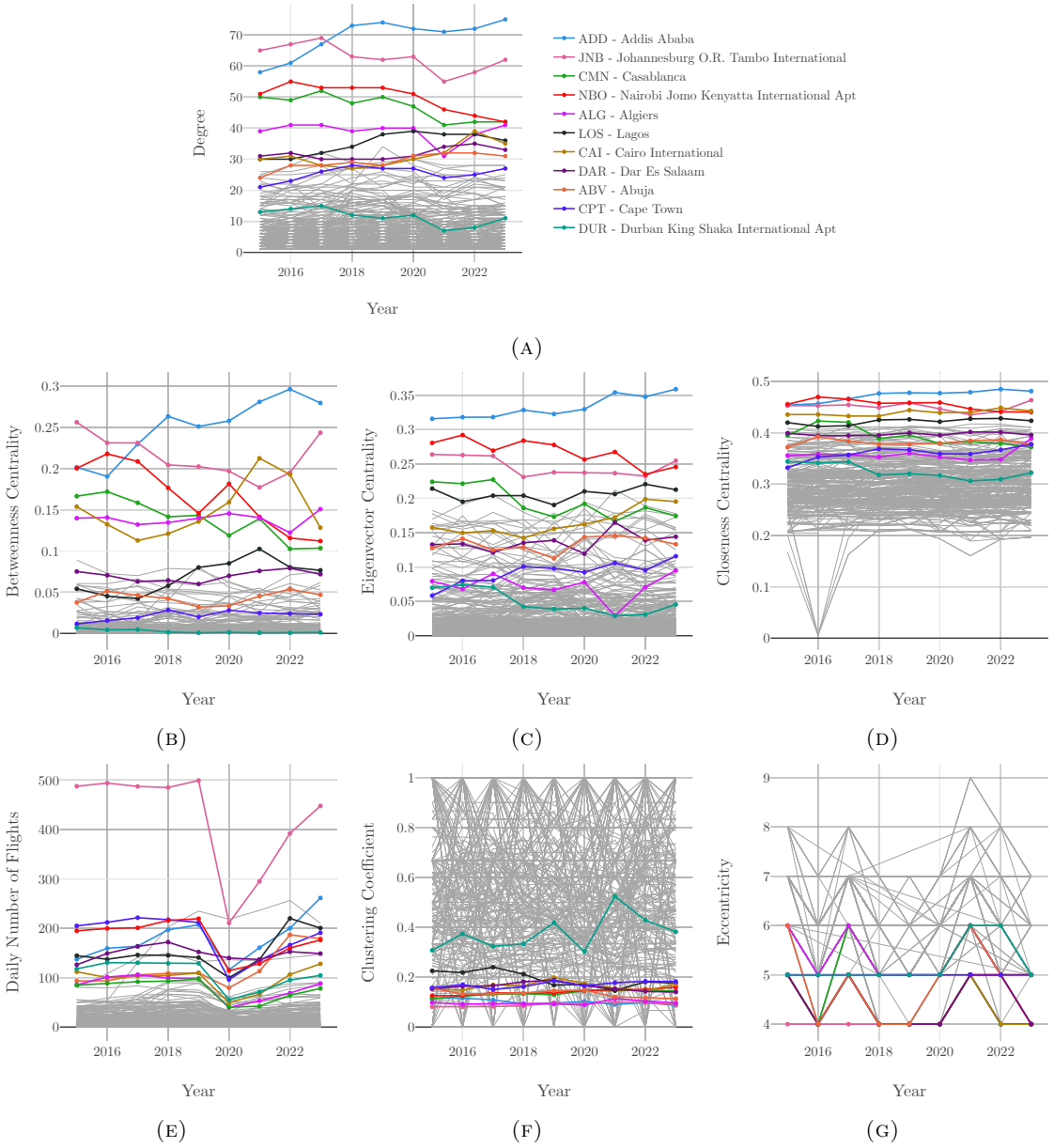


FIGURE 5.5: Trends of various measures for key airports in the AATN aggregated annually from 2015 to 2023. The measures include: (A) degree, (B) daily number of flights, (C) betweenness centrality, (D) eigenvector centrality, (E) closeness centrality, (F) clustering coefficient, and (G) eccentricity.

in 2020 due to COVID-19 is generally followed by a recovery in key airports. By 2023, only Addis Ababa, Lagos, Abuja, Cairo International, and Dar Es Salaam are fully recovered. Lagos sees a net decrease in seat capacity from 2015 to 2019. Further analysis of AATN's pandemic recovery is provided in Section 5.3.

For the daily number of flights, the trends are quite similar to daily airline seat capacity, as the differences would arise from a difference in the size of the aircraft. In Figure 5.5 (E), Nairobi Wilson Airport appears to be the busiest airport by number of flights in the 2020 AATN.

Figure 5.5 (F) displays trends in clustering coefficients of key airports. Durban King Shaka International shows the largest net increase from 2015 to 2023, while Lagos experiences the greatest decrease.

The eccentricity trends plot in Figure 5.5 (G) shows a minimal change in eccentricity over the period. The maximum range of eccentricity across the period is two, observed for Abuja and Casablanca. Among the key airports, eccentricity values range between four and six. Notably, no airport has an eccentricity below five in 2021, indicating that the network is less efficient during the pandemic recovery.

The identification of key routes follows a similar approach to the identification of key airports, where Figure 5.6 (A) shows the relationship between daily airline seat capacity and edge betweenness centrality for the 2023 AATN. Unlike its node-level counterpart, provided in Figure 5.4 (B), there appears to be no correlation between daily seat capacity and edge betweenness centrality. Therefore, both routes with high daily seat capacity and routes with high edge betweenness centrality are identified as key routes. The qualifying criteria for inclusion is set to daily seat capacity being higher than 3 200 seats or edge betweenness centrality being higher than 0.025. It appears that routes with high daily seat capacity tend to be domestic routes and routes with high betweenness centrality tend to be international routes.

The key routes identified include: Cape Town - Johannesburg O.R. Tambo; Durban King Shaka - Johannesburg O.R. Tambo; Abuja - Lagos; Cape Town - Durban King Shaka; Mombasa - Nairobi Jomo Kenyatta; Algiers - Johannesburg O.R. Tambo; Addis Ababa - Mombasa; Windhoek Eros - Walvis Bay; Johannesburg O.R. Tambo - Walvis Bay; and Algiers - Cairo.

Trends of daily airline seat capacity for top routes are provided in Figure 3.1 (A) and similarly for edge betweenness centrality in Figure 5.6 (B). A plot of the relationship between edge betweenness centrality and the great circle diameter is shown in Figure 5.6 (C).

Figure 3.1 (A) highlights the prominence of the ‘golden triangle’ of routes between Cape Town, Durban King Shaka International, and Johannesburg O.R. Tambo International, which constitute three of the top four routes in 2023 according to daily seat capacity. By 2023, these routes do not fully recover to pre-pandemic volumes, whereas Abuja-Lagos grows substantially from pre-pandemic levels and is the route with the third highest seat capacity.

The trends of edge betweenness centrality of routes provided in Figure 5.6 show quite volatile patterns. Many of the top routes within the period are discontinued during the period. The Algiers - Johannesburg O.R. Tambo International route – introduced in 2023 – has the highest edge betweenness centrality, coinciding with a reduction in edge betweenness centrality for both Algiers - Cairo International and Cairo International - Johannesburg O.R. Tambo International.

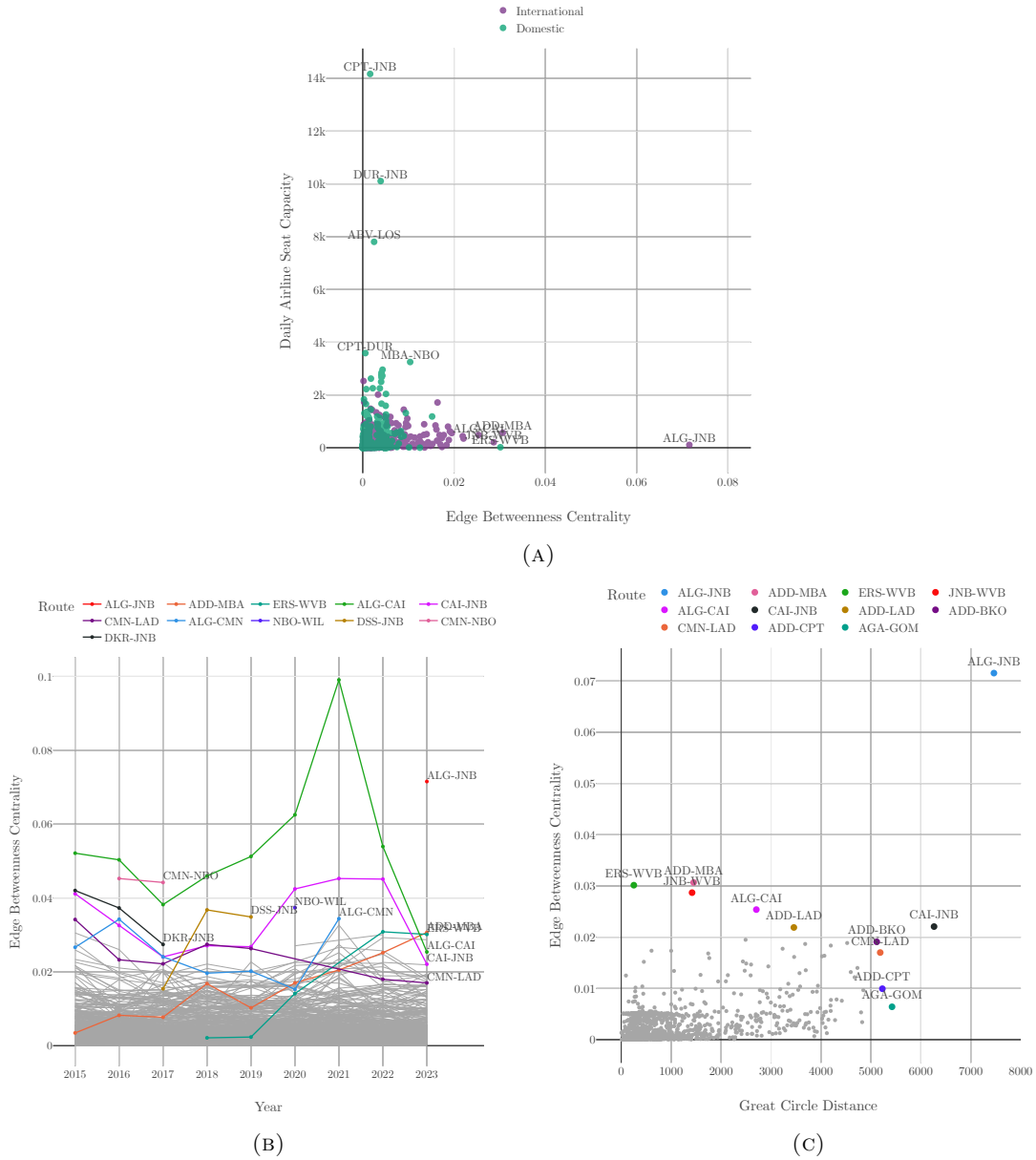
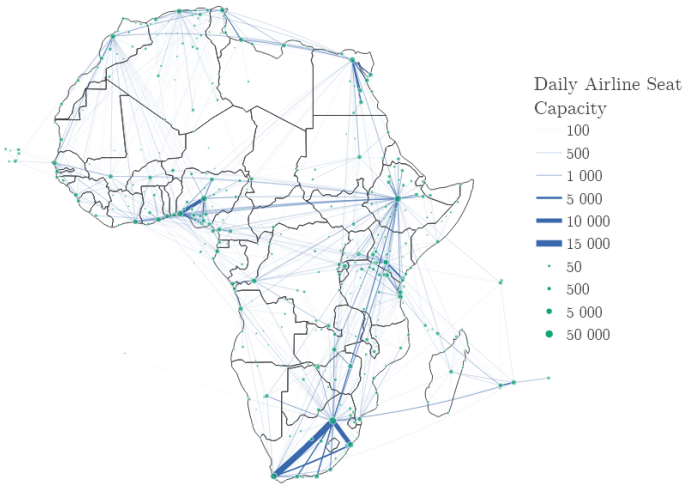


FIGURE 5.6: Plots for key routes in the African Air Transport Network showing: (A) daily airline seat capacity vs edge betweenness centrality for the 2023 AATN, (B) edge betweenness centrality vs year, and (C) edge betweenness centrality vs great circle distance of edges for the 2023 AATN.

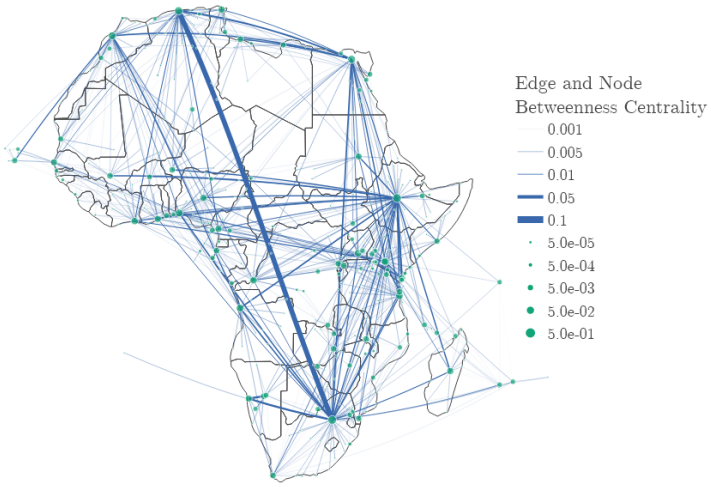
As discussed in Section 5.1.3, the Namibian domestic network is somewhat separate from the rest of the AATN since domestic flights from Windhoek are generally operated from Windoek Eros Airport and international flights are generally operated from Windhoek Hosea Kutako International. This creates a perceived 'chokepoint' through Walvis Bay for the shortest paths terminating at Namibian domestic airports, resulting in a high edge betweenness centrality for Windhoek Eros Airport - Walvis Bay and Johannesburg O.R. Tambo International - Walvis Bay.

The relationship between edge betweenness centrality and great circle diameter in Figure 5.6 (C) yields a Pearson correlation coefficient of 0.4433, indicating a weak positive linear

relationship, with geographically distant airports tending to have a higher betweenness centrality.



(A)



(B)

FIGURE 5.7: Geographic plots of nodes and edges scaled according to (A) daily airline seat capacity, and (B) edge and node betweenness centrality.

Combined geographic plots of the 2023 AATN are provided in Figure 5.7, where (A) indicates the daily airline seat capacity of the routes and airports, and (B) indicates both the node betweenness centrality and the edge betweenness centrality.

Figure 5.7 (A) highlights the highest capacity routes such as the 'golden triangle' and Abuja - Lagos, whilst also highlighting some hub airports and their influence on seat capacity, such as Addis Ababa, Johannesburg O.R. Tambo International, and Cairo International.

In comparison, Figure 5.7 (B) shows that routes with high edge betweenness centrality are quite different to routes with high airline seat capacity, as already established with Figure 5.6 (A). The plot also highlights nodes with high betweenness centrality and the edges which contribute to their betweenness. These nodes are Addis Ababa, Johannesburg O.R. Tambo International, Algiers, Cairo International, Nairobi Jomo Kenyatta International, and Casablanca.

The methods used in this section can be applied to other airports of particular interest, to understand their importance in the network and evolution thereof. The node eccentricity measure appears to offer limited insight; given that the closeness centrality is also analysed, which considers the average shortest path to all other nodes, future work could exclude the eccentricity measure.

Overall, the investigation into key airports and routes within the AATN provides useful findings into the network's structure and evolution. When considering seat capacity, the results are heavily influenced by the aforementioned 'golden triangle'; Johannesburg O.R. Tambo and Cape Town show the highest and third highest daily seat capacity for the 2023 AATN. There appears to be a positive correlation between a node's centrality and its daily seat capacity; however, the same cannot be said for routes, where there appears to be no correlation between edge betweenness centrality and seat capacity. By 2018, Addis Ababa had the highest centrality in the AATN in all four centrality measures and showed steady growth throughout the period.

Following this investigation into key airports and routes, the analysis of the structure of the AATN continues with community detection.

5.1.3 Community Detection

Louvain community detection is used to analyse the communities within the 2023 AATN, given the seat capacity that exists between the airports. This provides a snapshot of the community structure of the AATN. The algorithm uses only the list of airports and their relationships, measured by seat capacity, without considering geographic locations.

The Louvain algorithm aims to maximise the modularity of a network based on its partitioned communities. As described in Section 4.1.3, the implementation uses a resolution parameter γ that modifies the modularity quality formula. Lower γ values detect larger communities, and higher values detect smaller communities. Figure 5.8 shows the inverse relationship between modularity and the number of communities detected across a range of γ values, while Figure 5.9 shows community detection results for the 2023 AATN for a set of these values.

Despite the algorithm not being provided explicit geographic information, but rather the relationship of seat capacity between airports, the communities allocated appear to align with geographic boundaries. Referring to Figure 5.9 the following observations are made:

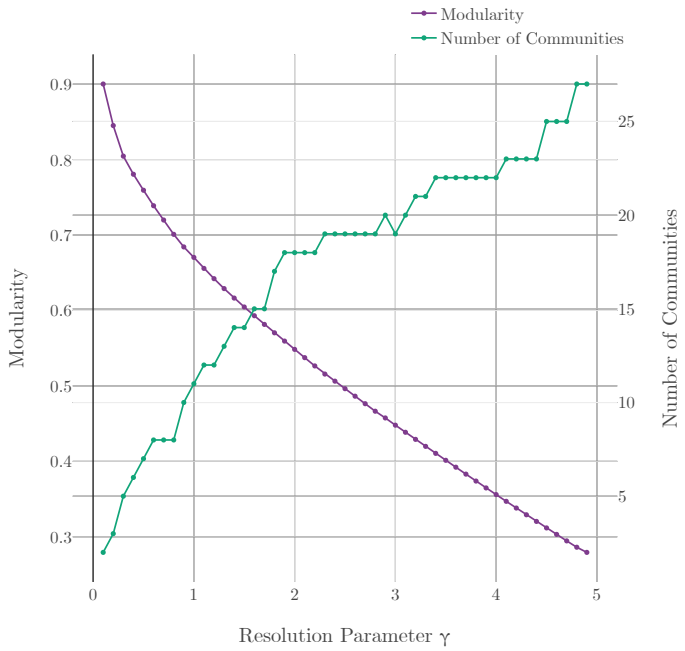


FIGURE 5.8: Both modularity and number of communities vs the resolution parameter γ applied to the 2023 AATN.

1. As mentioned previously, Windhoek has two airports: one favouring international routes and one favouring domestic routes. Consequently, some domestic airports do not have international connections through Windhoek, the country's main international gateway. At $\gamma = 0.1$, domestic airports that connect to Windhoek Eros (domestic airport) form their own small community, while the two international airports, Windhoek Hosea Kutako International and Walvis Bay, along with Swakopmund and Sesriem, are part of a larger community with the rest of the continent.

2. From $\gamma = 0.2$, there is a divide between South-East and North-West, with exceptions of Chad and Garoua in Cameroon.

3. At $\gamma = 0.3$, the communities align well with African subregions,[2] except for the Republic of Congo, Democratic Republic of Congo, Wau in South Sudan, Morocco, Malawi, Mayotte, St Helena, Chad, and Garoua in Cameroon.

4. By $\gamma = 0.5$, Nigeria is allocated its own community, and the Indian Ocean islands of Mauritius, Reunion, Seychelles, Madagascar, Mayotte and the Comoros island of Anjouan are also allocated their own community.

5. As γ is increased, individual countries tend to be allocated their own communities, which is likely attributed to their domestic air transport networks being generally isolated from international connections.

[2]See Figure 3.2 for African subregions

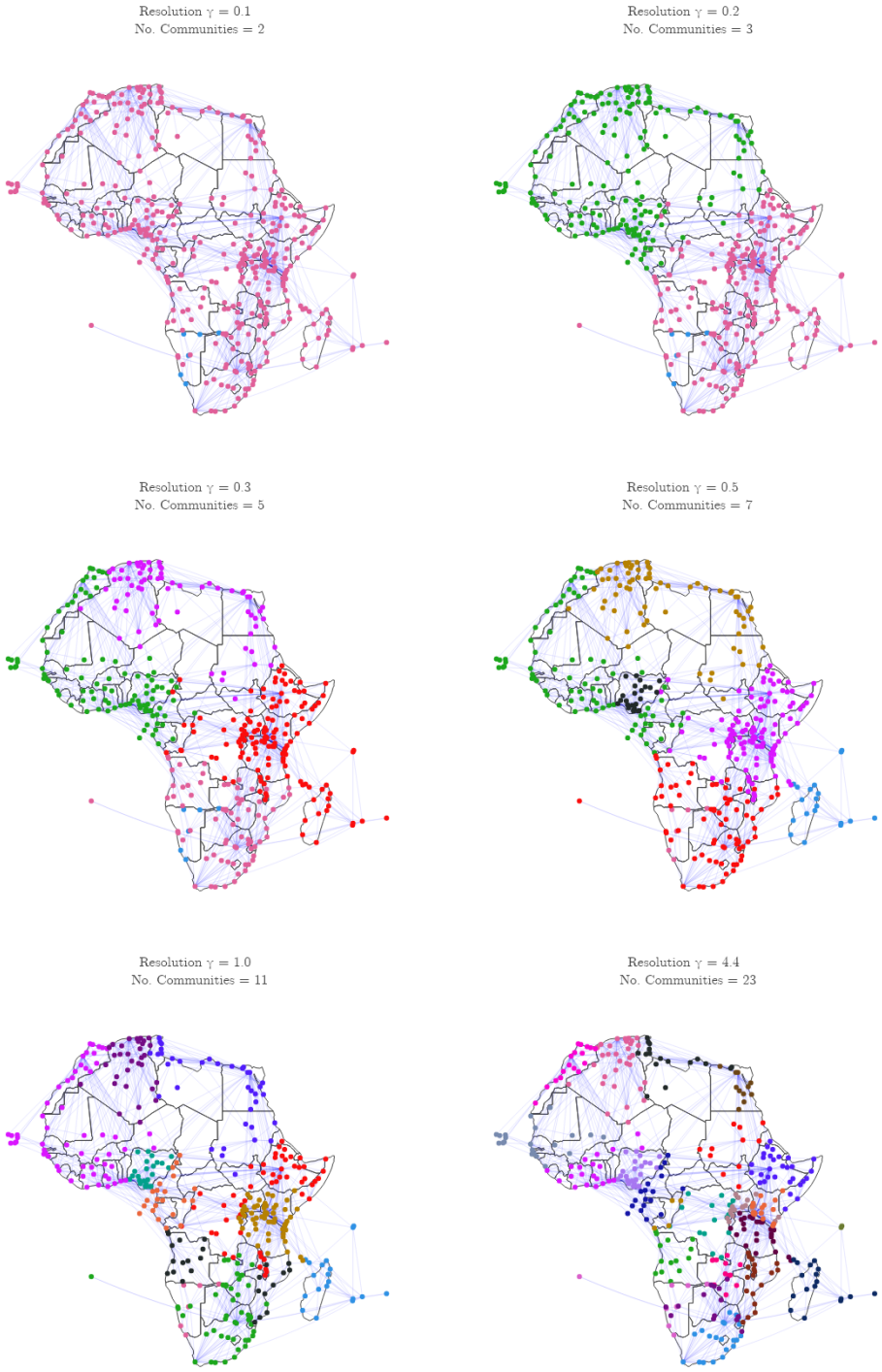


FIGURE 5.9: Louvain Community Detection applied to the 2023 AATN with increasing resolution parameter γ. The algorithm is weighted by seat capacity for each route.

6. At $\gamma = 1.0$, the result is the same as that achieved with the original Louvain algorithm, which does not include the resolution parameter. 11 communities are revealed

at this resolution.

7. At $\gamma = 4.4$, 23 communities are detected. Of the 57 countries, only six (Comoros, Egypt, Kenya, Namibia, South Africa and Zimbabwe) are allocated more than one community, each having two.

8. Communities generally appear to be geographically clustered, where airports tend to be in community with nearby airports.

The results presented in this section are for an aggregated network for 2023. Within the year, subtle changes in the network structure occur due to flight seasonality and network evolution. For detailed intra-year analysis reflecting this, see Appendix A.

While an airport can belong to multiple communities, Louvain community detection reveals dominant non-overlapping communities, which provides valuable insights into the network's structure. Further exploration into community detection can take multiple directions, including using newer or more bespoke community detection algorithms and exploring other attributes of the communities that are identified. This may include how they compare with each other and where connectivity can be improved, both within and between them.

Community detection reveals notable insights into the structure of the AATN, showing that communities often align with geographic regions, including African subregions and countries. The resolution parameter allows for exploration of the network's communities at varying levels, offering a detailed understanding of the network's community structure. The final aspect of the network structure we will analyse is the network's core.

5.1.4 K-core Decomposition

k-core decomposition is applied to the AATN to reveal insights into the structure of the AATN, providing an understanding of which nodes within the network are the most interconnected.

Figure 5.10 shows the resulting core networks identified from the 2023 AATN using k-core decomposition, where the results for increasing values of k are shown. A higher value of k results in a smaller core network. Community detection is also applied to these core networks with a resolution $\gamma = 1$ and weighted by airline seat capacity.

Table 5.2 relates to Figure 5.10 and shows the number of nodes, number of edges, the ratio of the number of edges in the core sub-network to the number of edges in the AATN, and the ratio of airline seat capacity in the core sub-network to that of the entire network.

The highest value of k that can be applied to the network is 8. In this case, the core network is sparsely distributed across the continent, except in West Africa, where there is a high concentration of airports included in the core.

As described in Section 4.1.4, k-core decomposition does not consider the strength of connections. As one can see in Figure 5.10, Durban King Shaka International and Cape

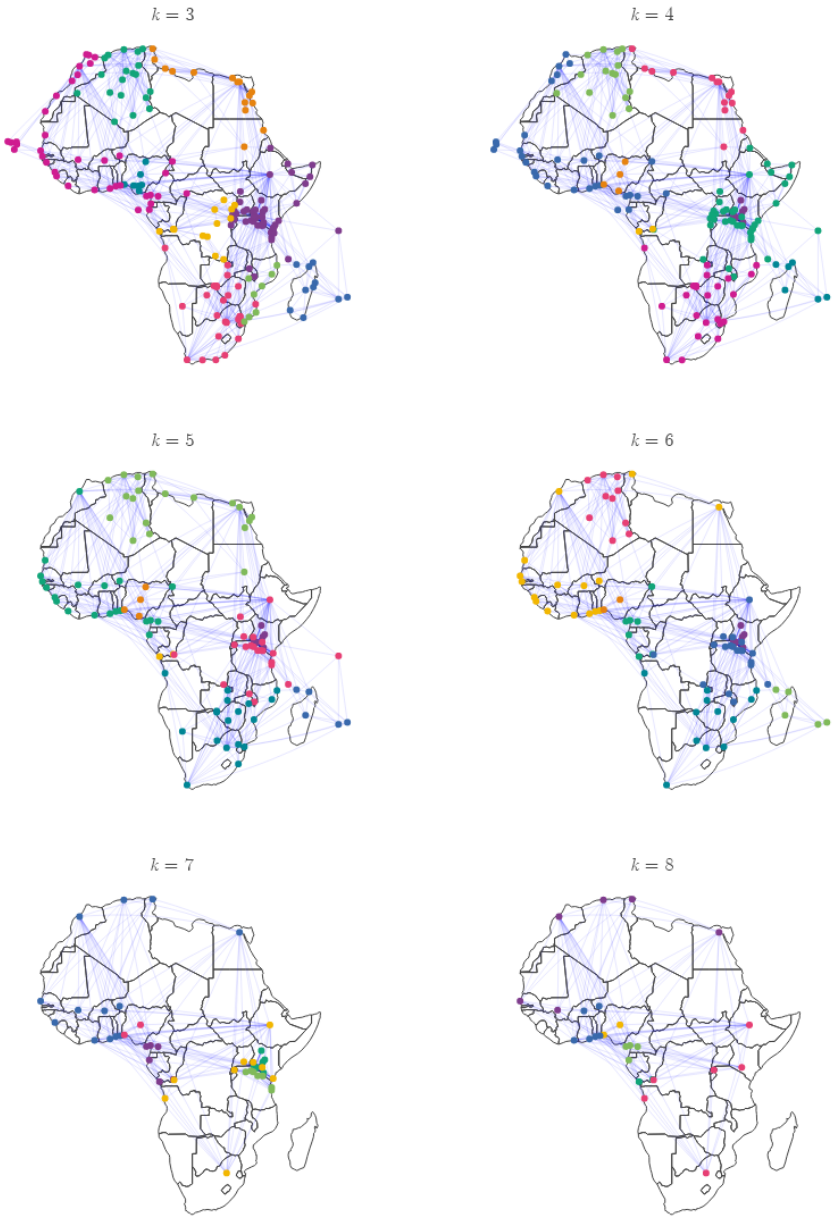


FIGURE 5.10: Resulting k-core decomposition of the 2023 AATN, with increasing k.

Town – which constitute two nodes of the 'golden triangle' – are excluded from the core at $k = 6$ and $k = 7$, respectively. However, the k-core definition of a core is still fairly robust: Of the 25 nodes that are excluded when increasing k from 7 to 8, only Conakry and Entebbe can be included in the $k = 8$ core set by adding one more connection to a node in that set. For Cape Town to be included in the $k = 7$ core, it needs three additional

TABLE 5.2: Summary of k-core sub-networks for given k value for the year 2023.

k	No. Nodes	No. Edges	Core-to-Total Edge Ratio	Core-to-Total Airline Seat Capacity Ratio
3	188	834	74.20%	84.02%
4	147	721	64.15%	75.45%
5	103	560	49.82%	66.16%
6	85	475	42.26%	50.39%
7	51	302	26.87%	30.84%
8	26	157	13.97%	19.99%

connections to nodes that are within that core. Durban King Shaka International requires two additional connections to nodes in the $k = 6$ core to be retained in the core.

With reference to the results in Section 5.1.2, the route with the third highest seat capacity, which is Abuja - Lagos, is included in the $k = 8$ core. Out of the 11 key airports identified, only Dar Es Salaam, Cape Town and Durban King Shaka International are not included in the $k = 8$ core. For Dar Es Salaam to be upgraded from the $k = 7$ core to the $k = 8$ core, it needs to be connected to an additional three nodes in the $k = 8$ core.

The $k = 8$ core sub-network appears to have influence on the AATN's airline seat capacity, where in Table 5.2 it can be seen to account for 19.99% of airline seat capacity over 13.97% of the AATN's edges. However, when considering that the 'golden triangle' between Johannesburg O.R. Tambo International, Cape Town, and Durban King Shaka International accounts for 9.94% of airline seat capacity in the AATN between just three nodes, the influence appears to be quite low.

In summary, k-core decomposition provides a robust definition of a core of the network, based on interconnectivity, which helps describe the structure of the AATN. The core of the AATN is distributed across the continent and also includes a cluster of airports in West Africa. This concludes the results and discussion of the structure of the AATN, with the evolution of the network considered next.

5.2 Evolution of the Network

The evolution of the network is examined by analysing how its structure changes over time. We first consider longitudinal trends in network-wide indicators, followed by investigating changes in the network's core structure.

5.2.1 Longitudinal Analysis

The longitudinal analysis tracks network-wide indicators to identify trends of various AATN characteristics over a period from 2015 to 2023. Figure 5.11 presents trends for a selection of network metrics, where the AATN is aggregated at an annual level. In some

cases, the network is compared to the average of 1 000 randomly generated networks with the same number of nodes and edges for the given year.

The analysis period includes the COVID-19 pandemic, where the network captures most nodes and edges from the 2019/2020 Northern Hemisphere winter season, resulting in a high count for 2020, as Figure 5.11 (A) and (B) show. However, there is a noticeable decline in nodes and edges in 2021 due to the pandemic. Section 5.3 provides a closer investigation of the effects of the COVID-19 pandemic and subsequent recovery.

5.2.1.1 Number of Nodes and Edges, and Seat Capacity

The number of nodes in the AATN, shown in Figure 5.11 (A), increases from 2015 to 2017, followed by a slight reduction from 2017 to 2019. After the COVID-19 pandemic, there is notable recovery up to 2023, reaching 393 nodes. In comparison, the number of edges in the AATN, as Figure 5.11 (B) shows, increases from 2015 to 2019, then decreases from 2019 to 2021, and subsequently recovers from 2021 to 2023, ending the period with 1 124 edges.

For context, consider South East Asia, which is also a developing region. According to Dai et al. (2018), in 2012, the South East Asian Air Transport Network had 237 nodes and 602 edges, with nodes representing cities aggregated annually. Xu and Harriss (2008) showed that the U.S. Air Transport Network had 272 nodes and 6 566 edges in Q4 2005, aggregated quarterly, also with nodes representing cities. Both networks are smaller than the 2023 AATN in terms of number of nodes. However, while the South East Asian Air Transport Network has fewer edges, the U.S. Air Transport Network has substantially more. One should note that the studies referenced here had different configurations in assembling their respective networks and, therefore, the comparison is not direct.

Figure 5.11 (C) illustrates the trend of daily airline seat capacity within the AATN from 2015 to 2023. Seat capacity is a measure that can be determined without the use of network analysis tools and techniques, but it is used as a measure of the strength of the network. Similar to the seat capacity of several routes and airports in Figure 3.1, the aggregate seat capacity of the AATN increases steadily from 2015 to 2019, sharply decreases in 2020 due to the COVID-19 pandemic, and recovers rapidly by 2023, reaching 97.9% of the capacity of 2019.

5.2.1.2 Mean Degree and Network Density

The mean degree shown in Figure 5.11 (D) and the density shown in 5.11 (E) are metrics used to express the number of edges in relation to the number of nodes. Both metrics show an increase from 2015 to 2019, followed by a decrease from 2019 to 2023. In 2023, the network density is 0.01459, meaning 1.459% of potential node pairs in the AATN are connected. In 2023, the average node degree is 5.72. Inferring from previous studies (Dai et al., 2018; Xu and Harriss, 2008), the South East Asian Air Transport Network had a mean degree of 5.08 and a network density of 0.02153 in 2012, while the U.S. Air

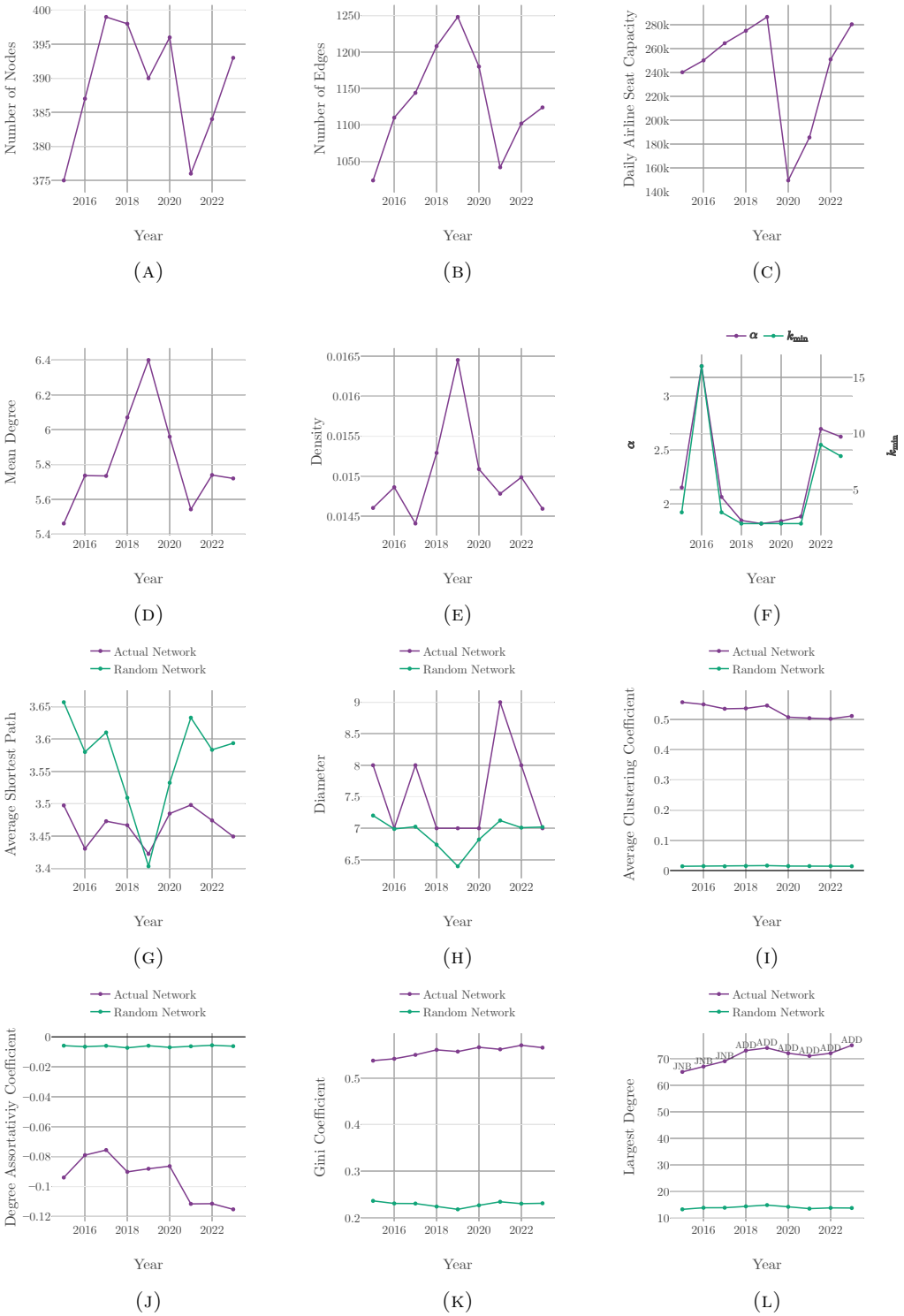


FIGURE 5.11: Longitudinal trends of network-wide indicators for the AATN, aggregated by year from 2015 to 2023: (A) number of nodes, (B) number of edges, (C) daily airline seat capacity, (D) mean degree, (E) density of network, (F) power-law exponent α of degree distribution, (G) average shortest path, (H) diameter of network, (I) average clustering coefficient, (J) degree assortativity coefficient, (K) Gini coefficient of node degrees, and (L) degree of largest degree node.

Transport Network had a mean degree of 48.279 and a network density of 0.17815 in Q4 2005, indicating they were denser than the 2023 AATN.

5.2.1.3 Power-Law Exponent

The power-law exponent (α) in Figure 5.11 (F), along with the $k_{\min}$ derived by the exclusion of the head of the degree distribution, is heavily influenced by erratic $k_{\min}$ values ranging from 2 to 16. This variability in $k_{\min}$ values hinders drawing meaningful conclusions from the trend of the power-law exponent. As described in Section 5.1.1, future work could explore methods to adapt the method of determining the power-law exponent of the degree distribution in the tail to make the results less sensitive to subtle changes in the network while tracking its evolution.

5.2.1.4 Average Shortest Path and Network Diameter

Figure 5.11 (G) shows the average shortest path of the network – which is the average of the number of edges along the shortest path between each node pair – from 2015 to 2023. There is no clear trend and the values are relatively consistent, ranging between 3.42 and 3.5. The average shortest path is slightly shorter than the random network case in general, except in 2019. For comparison, the average shortest path was found to be 3.12 in 2012 for the South East Asia Air Transport Network (Dai et al., 2018) and 1.90 in Q4 2005 for the U.S. Air Transport Network (Xu and Harriss, 2008).

Similar to the average shortest path, the network diameter (Figure 5.11 (H)) is a metric of the efficiency of the network and provides the longest shortest path within the network. There is no clear trend in the given period, except that the diameter reaches a high of 9 in 2021, likely due to disruption from the COVID-19 pandemic. The random case appears to perform a bit better showing higher efficiency than the actual case. The network diameter contributes little additional insight into network efficiency. Since the average shortest path already reflects changes in the network's efficiency with greater sensitivity, future studies might consider excluding the network diameter as a measure.

5.2.1.5 Average Clustering Coefficient

Figure 5.11 (I) shows the average clustering coefficient gradually decreasing over time to 0.51 by 2023, even through the COVID-19 period. In comparison, Dai et al. (2018) found that the average clustering coefficient was 0.21 for the South East Asia Air Transport Network in 2012, while Xu and Harriss (2008) reported 0.73 in Q4 200 5for the U.S. Air Transport Network. The observed coefficients are substantially higher than the random networks'.

Given that the actual network has a higher average clustering coefficient and lower average shortest path than a randomised network, the AATN exhibits 'small-world' properties (Dai et al., 2018).

5.2.1.6 Degree Assortativity Coefficient

The degree assortativity coefficient, shown in Figure 5.11 (J), measures the correlation between the degrees of connected nodes. It is negative at -0.12 in 2023, indicating disassortative mixing, where high-degree nodes connect to low-degree ones, reflecting the hub-and-spoke model generally observed in air transport networks. This coefficient gradually decreases over time, including during the COVID-19 pandemic. For the random case, it is more stable, higher, and close to zero throughout the period.

5.2.1.7 Gini Coefficient and Largest Degree

The Gini coefficient and largest degree from Figures 5.11 (K) and (L) both increase over time, far exceeding those of the average random network. The Gini coefficient indicates rising disparity between high and low degree nodes. The trend of the largest degree shows overall growth, with a slight dip occurring from 2019 to 2021, which coincides with the period affected by the COVID-19 pandemic. In contrast, the Gini coefficient does not exhibit the same dip during this time. The largest degree node is Johannesburg O.R. Tambo International (2015-2017) and Addis Ababa (2018-2023).

The network exhibits growth in number of edges and airline seat capacity leading up to the COVID-19 pandemic. However, trends of the AATN network-wide indicators associated with sizing the network are distinctly influenced by the COVID-19 pandemic. Despite this, there are several characteristics whose trajectories appear to be unaffected by the impact of the COVID-19 pandemic, namely the average clustering coefficient, degree assortativity coefficient, and Gini coefficient. This indicates that over the period affected by the COVID-19 pandemic, although there is a substantial reduction in the size of the network, many of its structural properties remain stable. The combined trends of declining clustering coefficient, declining degree assortativity coefficient, increasing Gini coefficient, and increasing largest degree suggest a growth of hubs connecting to small spokes within the network.

5.2.2 Evolution of the Core

k-core decomposition creates core sub-networks with reduced information, which can be meaningfully tracked over time. Figure 5.12 depicts the k-core sub-networks of the AATN from 2015 to 2023. Table 5.3, linked to Figure 5.12, shows the maximum k of the k-core decomposition, the number of nodes and edges, the ratio of the total core's number of edges to that of the entire network, and the ratio of the core's airline seat capacity to that of the entire network. Community detection, using a resolution parameter $\gamma = 1$, is applied to the core sub-networks to highlight sub-communities within the core.

From 2015 to 2017, the core network expands in size without a change in maximum k. In 2015, it includes airports in Eastern, Western, and Central Africa and Casablanca, growing by 2020 to encompass more across the continent. In 2018 and 2019, a different core emerges, centred around safari airports in Tanzania and Kenya, with a maximum k of

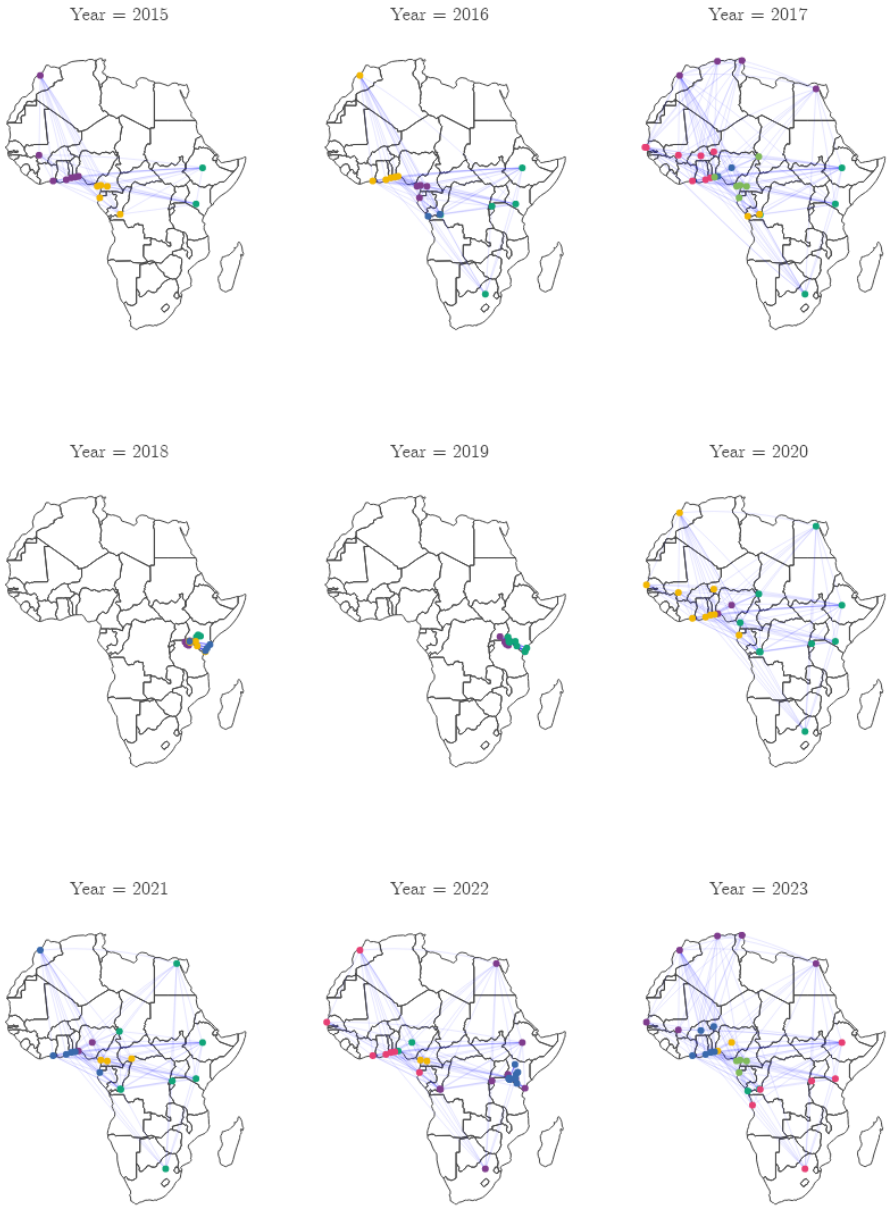


FIGURE 5.12: Evolution of the annual core network of the AATN. The core network is identified using k-core decomposition.

10, excluding some nodes from the k=9 core. From 2020 to 2023, maximum k fluctuates, leading to varying sub-network sizes. In 2022, the core sub-network reaches its largest size with 186 edges, a core-to-total edge ratio of 16.88%, and a maximum k of 9, matching the maximum k of 2015, indicating core growth over the years.

The core-to-total seat capacity ratio is found to exceed the core-to-total edge ratio, except

in 2018 and 2019, where the cores are centred on safari airports, indicating that they may not truly represent the AATN's core. The community detection process appears to cluster airports that are geographically nearby.

TABLE 5.3: Summary of k-core sub-networks for the evolution of the network.

Year	Max k	No. Nodes	No. Edges	Core-to-Total Edge Ratio	Core-to-Total Airline Seat Capacity Ratio
2015	9	14	73	7.13%	8.31%
2016	9	17	102	9.19%	12.58%
2017	9	26	175	15.30%	19.00%
2018	10	20	124	10.26%	2.40%
2019	11	18	106	8.49%	2.20%
2020	10	20	126	10.68%	14.18%
2021	8	19	108	10.36%	14.82%
2022	9	33	186	16.88%	19.53%
2023	8	26	157	13.97%	19.99%

Given the drastic changes in the core networks revealed for 2018 and 2019, it could be beneficial for the k-core decomposition approach to be adapted to be less sensitive to subtle changes in the network when analysing its evolution.

This completes the detailing of the evolution of the AATN. Next, the effects and recovery of the COVID-19 pandemic are examined.

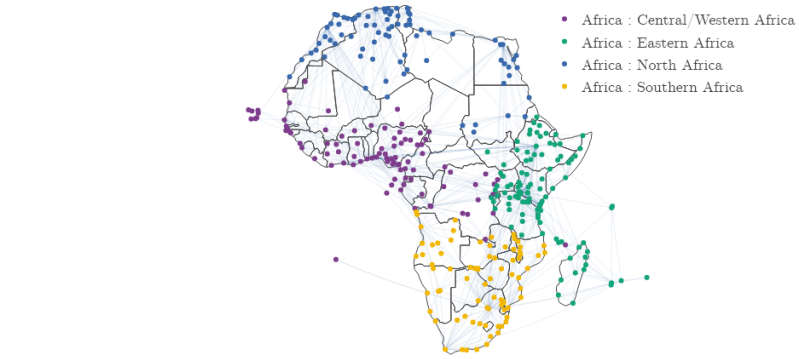
5.3 COVID-19 Pandemic Recovery

As outlined in the methodology, the study examines the evolution of the AATN during a period that includes the COVID-19 pandemic, which substantially impacts the network. This analysis begins with a comparative assessment of the commonalities and differences between the pre-pandemic phase and the recovery phase's conclusion, before quantifying how the network is altered throughout this time frame.

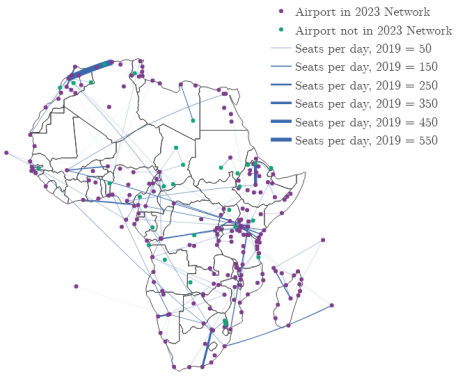
5.3.1 Pre-Pandemic vs End of Recovery: Commonality and Differences

Network analysis operations, using set theory, are used to reveal the intersection and differences between the 2019 AATN – representing the pre-pandemic network – and the 2023 AATN, representing the end of the recovery phase. The results from these operations are shown in Figure 5.13, where (A) shows the intersection, (B) the edges that exist in 2019 but not 2023, and (C) the edges that exist in 2023 but not 2019. In (B) and (C) the edges are weighted by daily airline seat capacity to indicate the prominence of the route that is added or removed.

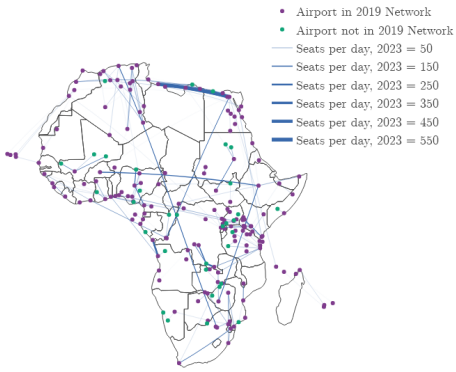
Figure 5.13 (A) shows a large number of edges and nodes that are common to both the 2019 AATN and the 2023 AATN. In comparison, Figure 5.13 (B) shows a fair amount of routes that exist in the 2019 AATN, but not in the 2023 AATN. Of these routes, the three



(A)



(B)



(C)

FIGURE 5.13: Commonality and differences in the AATN in the years 2019 vs 2023: (A) intersection of 2019 and 2023 networks, (B) edges that exist in 2019 but not 2023, and (C) edges that exist in 2023 but not 2019

with the highest daily airline seat capacity are Algiers - Casablanca (829.7), Addis Ababa - Axum (443.5), and Johannesburg Lanseria International - Port Elizabeth (302.9). In Figure 5.13 (C) there are several routes added to the network by 2023 that did not exist in 2019. Of these routes, the three with the highest daily airline seat capacity are Benghazi - Cairo International (508.4), Cairo International - Tripoli Mitiga (481.1), and Luanda - Pointe-Noire (241.7).

Now that the commonalities and differences between pre-pandemic and end of recovery phase have been visually represented and interpreted, the next section aims to quantify the degree to which the AATN changes over the period.

5.3.2 Network Similarity

To compare the AATN over time, the similarity measure described in Section 4.3.2 is used. Figure 5.14 illustrates this by comparing time states of the AATN with 2019.

FIGURE 5.14: Similarity measure of the AATN compared with 2019: (A) similarity measure vs other metrics compared with the year 2019, (B) similarity measure for Africa vs Global network compared with the year 2019, (C) similarity measure vs other metrics compared with same month in the year 2019, and (D) similarity measure for Africa vs global network compared with same month in the year 2019.

Figure 5.14 (A) aggregates the AATN annually, indexing the similarity measure, daily

number of flights, daily airline seat capacity, the number of edges, and mean degree to 2019. (B) also aggregates annually, using the similarity measure and the number of edges, and includes global network comparisons. (C) aggregates monthly, indexing the similarity measure, daily number of flights, daily airline seat capacity, the number of edges, and mean degree to the same month in 2019. (D) does the same for the similarity measure and the number of edges with global network comparisons.

Comparing Figures 5.14 (A) and (C), the latter shows a drastic drop in similarity from April to June 2020, hitting a low of 0.4262 in June, which then begins to recover, unlike the annual case in (A). From September 2020 in (C), the similarity stabilises between 0.5393 and 0.6345, indicating that although the number of edges continues to recover, the actual edges differ from 2019. Similarly, (A) exhibits flatter, declining similarity as edges recover. Before 2019, both plots show increasing similarity with network growth.

Figure 5.14 (B) and (D) show that the AATN's evolution mirrors that of the global network both in similarity and in number of edges throughout the period, suggesting that the AATN's magnitude of variation is not unique. This is the final section of the evaluation of the recovery of the network after the COVID-19 pandemic, and a conclusion of the chapter is presented next.

5.4 Discussion

In this chapter, the first section on network structure analyses the degree distribution of the network, revealing that a power-law distribution appropriately characterises the network's degree distribution. Key airports and routes are identified and examined, with key airports identified based on their degree and daily airline seat capacity, and key routes identified based on their daily airline seat capacity and edge betweenness centrality. A notable influence by the 'golden triangle' of routes between Johannesburg O.R. Tambo, Cape Town, and Durban King Shaka is revealed, with three of the top four highest seat capacity routes residing between these airports.

Although node centrality correlates with airline seat capacity, there is no similar correlation at the route level between airline seat capacity and edge betweenness centrality. Notably, Addis Ababa exhibits the highest centrality in the AATN by 2018 across all four centrality measures, with consistent growth throughout the period. The application of Louvain community detection on the AATN reveals community structures that appear to align with geographic regions, both in African subregions and individual countries. k-core decomposition highlights an interconnected core of the network that is distributed across the continent, with a marked concentration in West Africa.

Moving to the section on network evolution, where the progression from 2015 to 2023 is detailed, trends of network-wide indicators reveal a distinct impact of the COVID-19 pandemic, particularly on indicators associated with network sizing. However, other structural measures such as the average clustering coefficient, degree assortativity coefficient,

and Gini coefficient remain stable. The network demonstrates 'small-world' properties considering its higher average clustering coefficient and lower average shortest path compared to that of the average random network. Combined trends of a declining average clustering coefficient and degree assortativity coefficient, alongside an increasing Gini coefficient and largest degree, suggest a growing airline hub-and-spoke model within the network. Meanwhile, an evolving and growing network core is revealed.

The impact of the COVID-19 pandemic on the network and its subsequent recovery are explored in detail. The analysis shows that post-recovery, the network does not simply revert to its pre-pandemic state. Some new routes are introduced, while others are not reintroduced. Changes in the network are quantified using a similarity measure, revealing that the magnitude of the variations in the AATN's structure are consistent with trends observed globally. This concludes the results and discussion chapter, with final conclusions and recommendations to follow in the next chapter.

Chapter 6

Conclusion and Recommendations

This chapter concludes the dissertation on the structure and evolution of the African Air Transport Network (AATN). Section 6.1 summarises the main findings in relation to the research objectives. Section 6.2 reflects on the study's limitations, and Section 6.3 offers recommendations and identifies avenues for future research.

6.1 Conclusion

The three research objectives for this study entail illustrating and describing the structure of the AATN, detailing its evolution, and evaluating its recovery after the COVID-19 pandemic using network analysis tools and techniques. Through these objectives, the research enriches an understanding of the AATN's structure and evolution, contributing to both the specific literature on the AATN and the broader field of air transport networks.

The structure of the AATN is illustrated and described using a variety of analyses. Firstly, the network's degree distribution follows a power-law, featuring few highly connected hubs and many less connected nodes. The network notably exhibits 'small-world' properties. Node centrality within the network is correlated with airline seat capacity at airports, although there is no similar correlation between seat capacity and edge betweenness centrality for routes. The high seat capacity on the 'golden triangle' of routes between Johannesburg O.R. Tambo, Cape Town, and Durban King Shaka plays a dominant role in the network structure in terms of seat capacity. By 2018, Addis Ababa stands out as a leading node, exhibiting the highest centrality in four centrality measures, with consistent growth throughout the analysis period. Furthermore, applying Louvain community detection shows community structures within the AATN that correspond to geographic regions, including subregions and countries in Africa. Lastly, k-core decomposition reveals an interconnected core distributed throughout the network, with a notable concentration in West Africa.

The evolution of the AATN from 2015 to 2023 is detailed by several key findings. Prior to the COVID-19 pandemic, the network experiences growth in the number of edges and seat capacity, indicating expansion. However, the pandemic notably impacts these measures,

particularly affecting the network's size dynamics. Despite these changes, structural measures such as the average clustering coefficient, degree assortativity coefficient, and Gini coefficient remain stable. The trends of a decline in the average clustering coefficient and degree assortativity coefficient, combined with an increase in the Gini coefficient and highest degree, indicate a growing implementation of the hub-and-spoke model of airlines. In addition, there is evidence of an evolving and expanding network core.

The evaluation of the AATN's recovery post-COVID-19 indicates a slightly altered structure compared to its pre-pandemic state. Although some new routes are introduced, others are not reinstated. These changes in the network are quantified using a similarity measure, which reveals that the magnitude of the variations in the AATN's structure align with global trends.

6.2 Limitations

The limitations associated with this study include those imposed by the way the network representation is set up, the data set, the analytical methods used, and the time constraints for the research.

In assembling a network, several decisions must be made about how the network is going to represent the real world. For the AATN it is generally decided to represent the network with an annual time aggregation, with undirected edges, and using nodes to represent airports. With respect to time aggregations, the information associated with the time dependency between connecting flights is lost. With the undirected representation, only 93.8% of flights in the network are actually reciprocal, introducing some inaccuracy to the network. Various studies select different parameters when assembling their networks, limiting the ability to directly compare the results with those of other studies.

Furthermore, in assembling the network, only intra-African flights are considered, where in reality airports outside of Africa may play a role in the connectivity between African airports. It is assumed that on all route segments, passengers may board at the starting airport and disembark at the ending airport; however, it is known that this is not always the case. The data starts in the year 2015, due to data availability, where earlier years may have been beneficial.

In the analysis of the network structure, the community detection algorithm reveals dominant non-overlapping communities; however, it is possible that a given airport may belong to multiple communities, to which the memberships are not identified.

6.3 Future Work and Recommendations

The recommendations and areas for future work include expanding the study's scope and excluding certain elements in future analyses of the AATN's structure and evolution. In addition, this section identifies areas of network analysis that warrant further development

to improve their usefulness. Finally, further applications are suggested, which could be beneficial to various stakeholders.

Areas where the scope of the study could be expanded include a greater focus on the effects of the seasons of the year on the network, considering the roles of airports outside of Africa on the connectivity between African airports, performing supplementary analyses such as investigating the 'rich-club' coefficient, investigating the distribution of centrality measures other than the degree, and further exploring the communities that are identified by community detection.

In future work of the AATN, elements such as analysing the eccentricity of airports seem less beneficial compared to using closeness centrality, which offers a more sensitive and comprehensive measure by considering distances to all network nodes. For similar reasons, the diameter as a network-wide indicator appears less insightful than the average shortest path. Future work may consider excluding these two elements from the analysis.

Advancements in time-dependent network analysis could substantially enhance the study of air transport networks, providing a more accurate representation of reality. Time dependency greatly impacts the connectivity between hub airports and their spokes. Additionally, refining the applications of k-core decomposition and determining the power-law exponent can reduce sensitivity to subtle changes in network structure during its evolution.

There are opportunities for further application of network analysis on the AATN. The analysis used on the key airports and routes section can be applied to any airport of interest, helping to benchmark and reveal insights into their centrality and evolution within the AATN. Studying the resilience of the AATN can also provide insight by revealing geopolitical dependencies within the network.

While numerous aspects of the AATN remain unexplored, this study lays a foundation for applying network analysis techniques to examine the structure and evolution of the AATN, thus enhancing the existing literature and understanding of the subject.

Appendix A

Intra-year Community Detection

The community detection analysis performed in Section 5.1.3 is performed on an aggregated network for the year 2023. However, there are slight changes in the network during the course of the year due to seasonality of demand in the network and general evolution of the network within the year.

The Louvain community detection algorithm is applied to monthly aggregates of the network to demonstrate some of these nuances. Figure A.4 shows a facet wrap of community detection results for each month of the year, using a resolution parameter γ of 1, to produce the same results as the original Louvain algorithm.

For easier comparison, figure A.3 is also provided, showing the community detection result for $\gamma = 0.3$, resulting in fewer communities. There are some slight changes between the months that can be observed, including the following:

1. In January, some North African countries have formed their own community, by February these are included with other West African countries. Egypt and Sudan are included with East African countries.
2. In February, some Indian Ocean islands including Madagascar, Reunion, Mauritius, Seychelles and Mayotte are included with other East African countries, but in March they are included with Southern African countries.
3. From March to April East African countries are included with Southern African countries.
4. From January to May, Nigeria is detected to have its own community.
5. In November and December, airports in Niger are identified to be in community with East African countries, where between January and October they are included with West African countries.

Figure A.1 and Figure A.2 show heat maps of the modularity and number of communities respectively mapped to the resolution and month for the community detection results. These provide a high level view of changes in the communities of the network within the year.

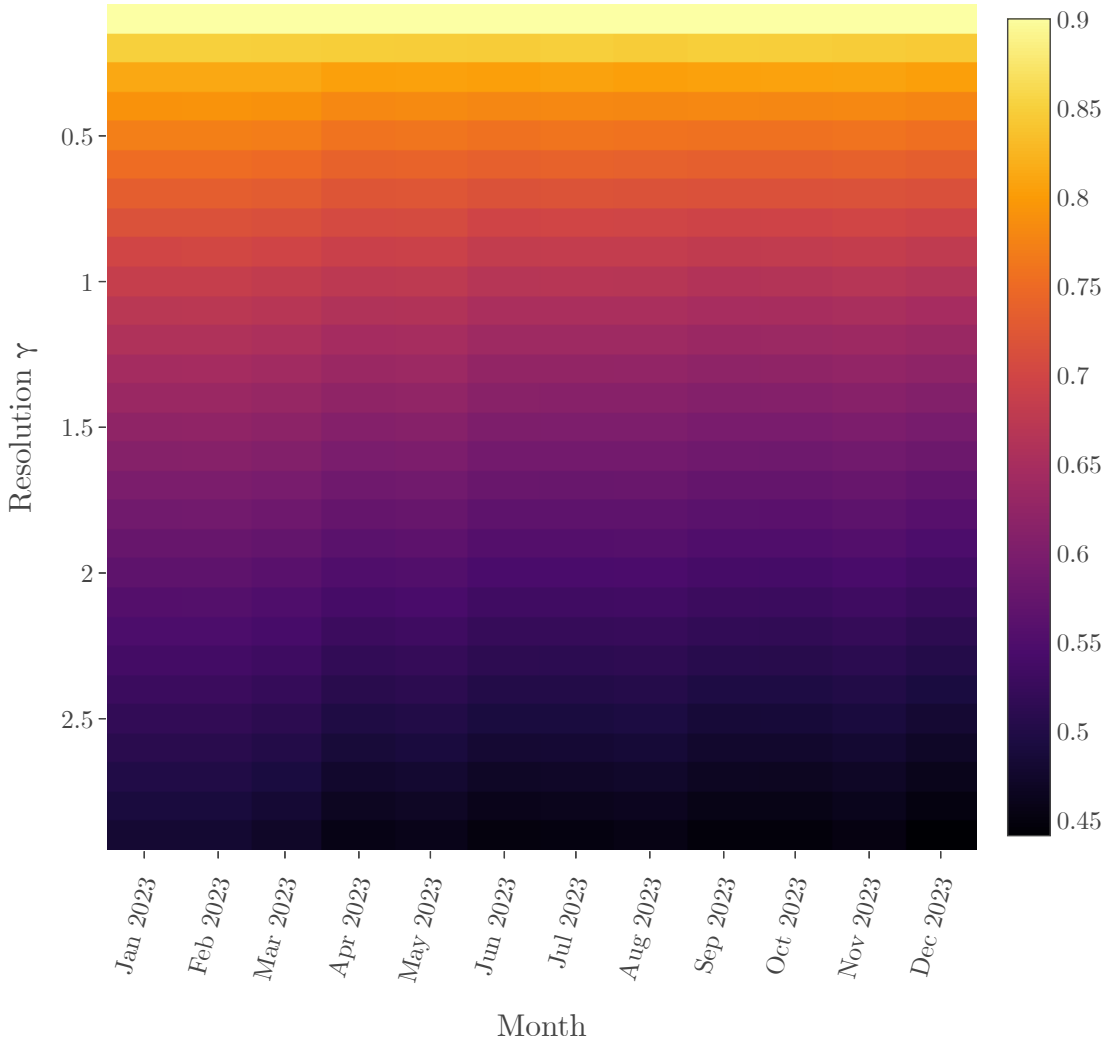


FIGURE A.1: Heat map of the modularity mapped against the resolution and month for the Louvain community detection results applied to the African Air Transport Network.

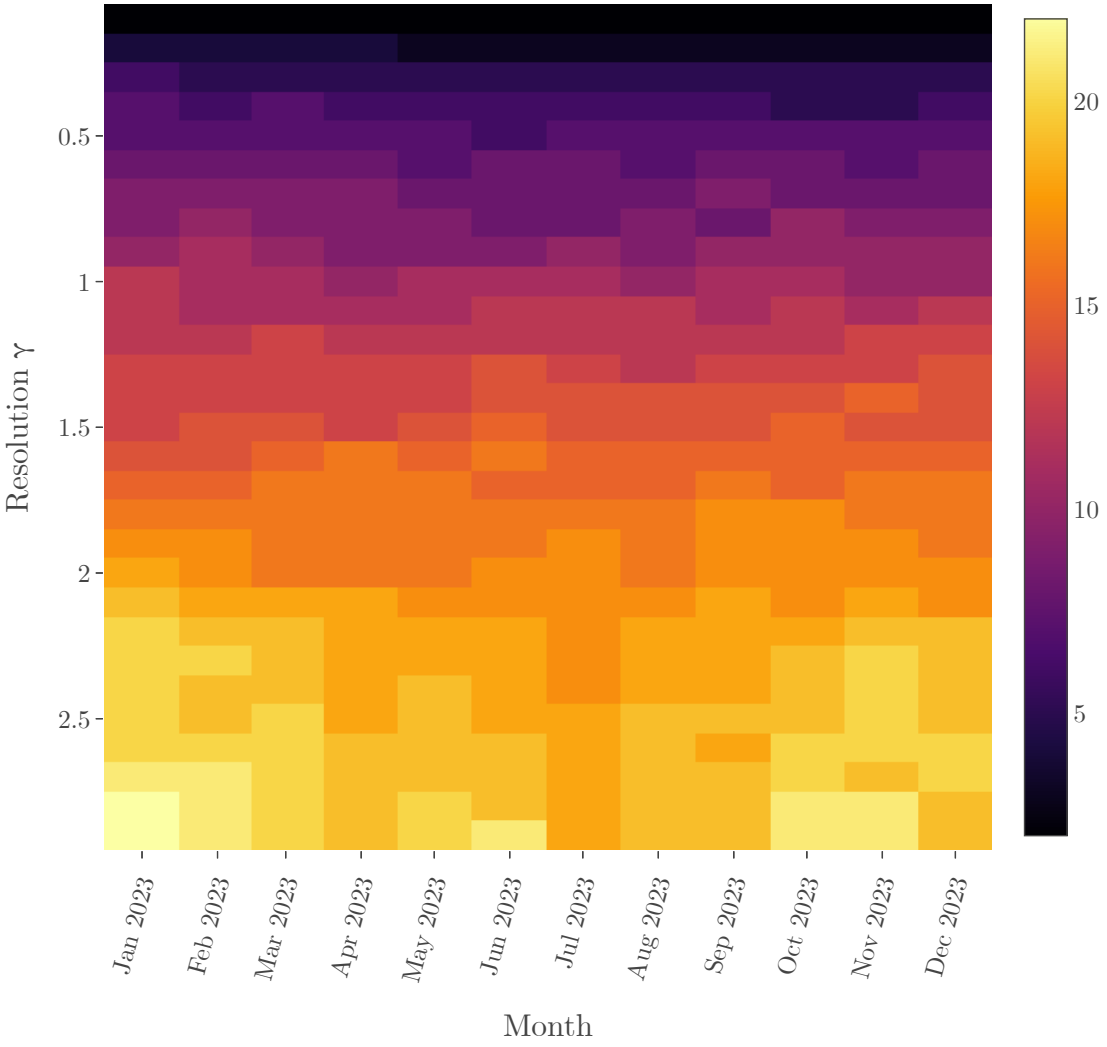


FIGURE A.2: Heat map of the number of communities mapped against the resolution and month for the Louvain community detection results applied to the African Air Transport Network.

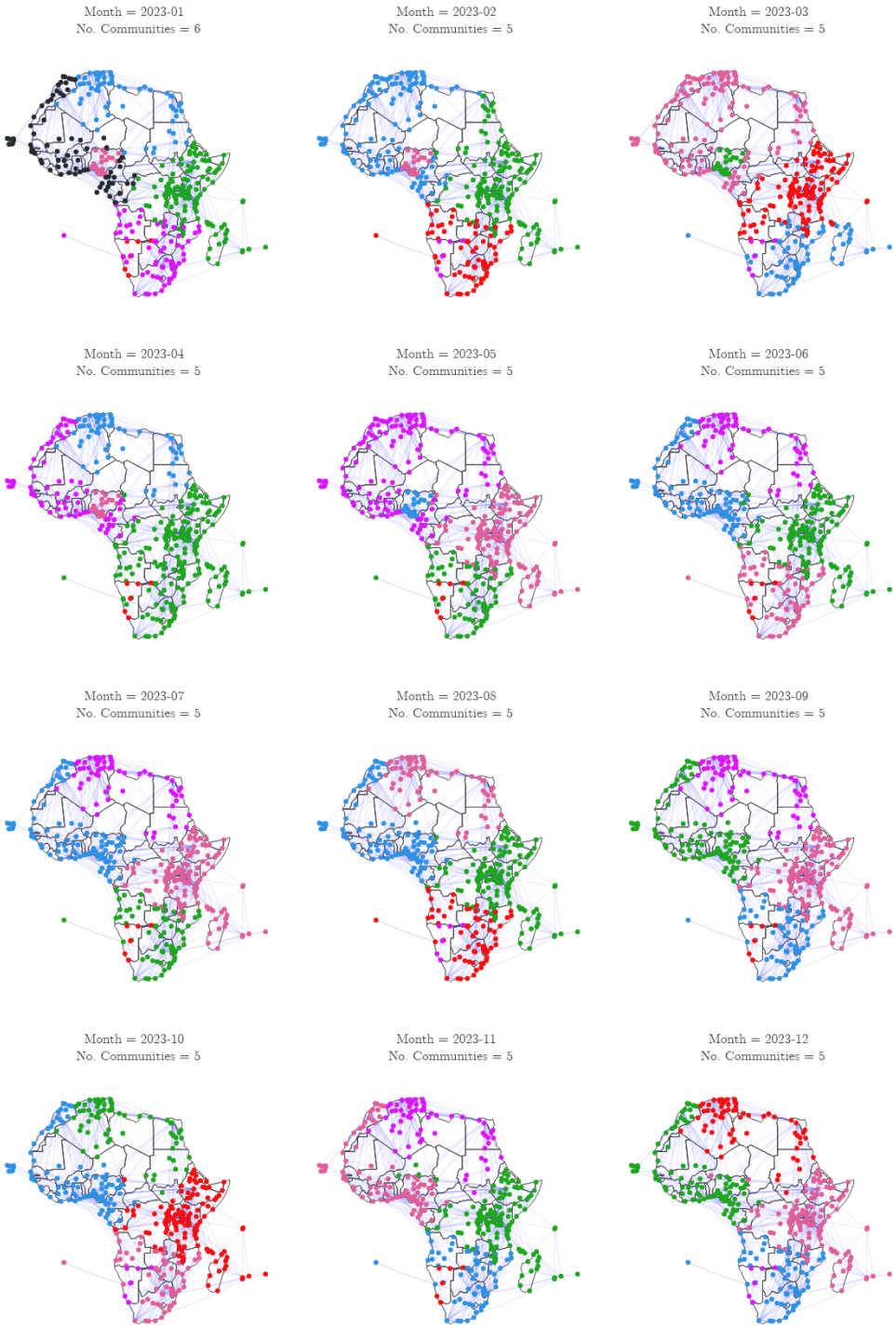


FIGURE A.3: Louvain Community Detection applied to the African Air Transport Network per month in 2023 for resolution parameter $\gamma = 0.4$. The algorithm is weighted by seat capacity.

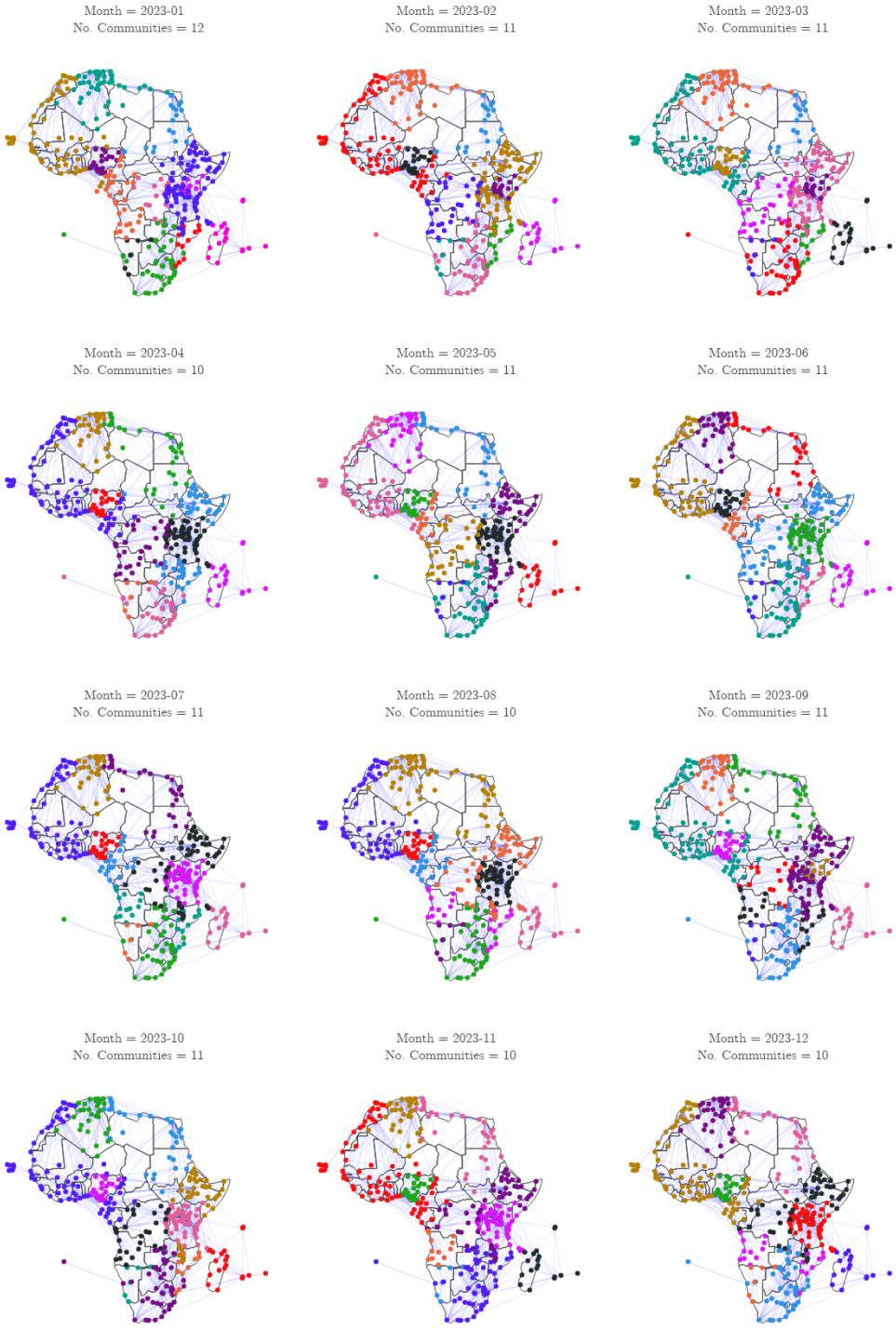


FIGURE A.4: Louvain Community Detection applied to the African Air Transport Network per month in 2023. The resolution parameter used is $\gamma = 1$, effectively applying the original Louvain algorithm. The algorithm is weighted by seat capacity.

Bibliography

African Union Commission (Dec. 2021). *AUC Information Paper on SAATM Implementation with Data Tables*. (Visited on 10/21/2024).

African Union Commission, African Civil Aviation Commission, and International Air Transport Association (2019). *The SAATM Handbook*. (Visited on 06/24/2025).

Airline Schedules Data (Oct. 2024). *Airline Schedules Data: An Insider's Guide | Data, Technology and Product | OAG*. https://www.oag.com/blog/airline-schedules-data-insiders-guide. (Visited on 10/23/2024).

Alstott, Jeff, Ed Bullmore, and Dietmar Plenz (Jan. 2014). "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions". In: *PLoS ONE* 9.1, e85777. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0085777. arXiv: 1305.0215 [physics]. (Visited on 07/28/2024).

Arvis, Jean-Francois and Ben Shepherd (June 2011). *The Air Connectivity Index: Measuring Integration in the Global Air Transport Network*. SSRN Scholarly Paper. Rochester, NY. (Visited on 04/05/2024).

Bagler, Ganesh (May 2008). "Analysis of the Airport Network of India as a Complex Weighted Network". In: *Physica A: Statistical Mechanics and its Applications* 387.12, pp. 2972–2980. ISSN: 0378-4371. DOI: 10.1016/j.physa.2008.01.077. (Visited on 10/04/2024).

Bao, Xiaoge et al. (Nov. 2021). "The Impact of COVID-19 on the Worldwide Air Transportation Network". In: *Royal Society Open Science* 8.11, p. 210682. DOI: 10.1098/rsos.210682. (Visited on 03/08/2024).

Bassens, David et al. (Dec. 2012). "African Gateways : Measuring Airline Connectivity Change for Africa's Global Urban Networks in the 2003-2009 Period". In: *South African Geographical Journal = Suid-Afrikaanse Geografiese Tydskrif* 94.2, pp. 103–119. DOI: 10.10520/EJC128548. (Visited on 03/13/2024).

Batagelj, V. and M. Zaversnik (Oct. 2003). *An O(m) Algorithm for Cores Decomposition of Networks*. DOI: 10.48550/arXiv.cs/0310049. arXiv: cs/0310049. (Visited on 11/05/2024).

Blondel, Vincent D. et al. (Oct. 2008). "Fast Unfolding of Communities in Large Networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2008/10/P10008. (Visited on 05/12/2024).

Bonacich, Phillip (1987). "Power and Centrality: A Family of Measures". In: *American Journal of Sociology* 92.5, pp. 1170–1182. ISSN: 0002-9602. JSTOR: 2780000. (Visited on 11/04/2024).

Brandes, Ulrik (May 2008). "On Variants of Shortest-Path Betweenness Centrality and Their Generic Computation". In: *Social Networks* 30.2, pp. 136–145. ISSN: 0378-8733. DOI: 10.1016/j.socnet.2007.11.001. (Visited on 03/08/2024).

Brueckner, Jan K., Nichola J. Dyer, and Pablo T. Spiller (1992). "Fare Determination in Airline Hub-and-Spoke Networks". In: *The RAND Journal of Economics* 23.3, pp. 309–333. ISSN: 0741-6261. DOI: 10.2307/2555865. JSTOR: 2555865. (Visited on 02/02/2025).

Cardillo, Alessio et al. (Feb. 2013). "Emergence of Network Features from Multiplexity". In: *Scientific Reports* 3.1, p. 1344. ISSN: 2045-2322. DOI: 10.1038/srep01344. (Visited on 03/21/2024).

Cheung, Dorothy P. and Mehmet Hadi Gunes (Aug. 2012). "A Complex Network Analysis of the United States Air Transportation". In: *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 699–701. DOI: 10.1109/ASONAM.2012.116. (Visited on 03/08/2024).

Cheung, Tommy K. Y., Collin W. H. Wong, and Anming Zhang (Jan. 2020). "The Evolution of Aviation Network: Global Airport Connectivity Index 2006–2016". In: *Transportation Research Part E: Logistics and Transportation Review* 133, p. 101826. ISSN: 1366-5545. DOI: 10.1016/j.tre.2019.101826. (Visited on 03/08/2024).

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman (Nov. 2009). "Power-Law Distributions in Empirical Data". In: *SIAM Review* 51.4, pp. 661–703. ISSN: 0036-1445. DOI: 10.1137/070710111. (Visited on 07/03/2024).

Dai, Liang, Ben Derudder, and Xingjian Liu (Apr. 2018). "The Evolving Structure of the Southeast Asian Air Transport Network through the Lens of Complex Networks, 1979–2012". In: *Journal of Transport Geography* 68, pp. 67–77. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2018.02.010. (Visited on 04/22/2024).

Diestel, Reinhard (2017). *Graph Theory*. Vol. 173. Graduate Texts in Mathematics. Berlin, Heidelberg: Springer. ISBN: 978-3-662-53621-6 978-3-662-53622-3. DOI: 10.1007/978-3-662-53622-3. (Visited on 06/25/2025).

Diop, Issa Moussa et al. (Nov. 2021). "Revealing the Component Structure of the World Air Transportation Network". In: *Applied Network Science* 6.1, p. 92. ISSN: 2364-8228. DOI: 10.1007/s41109-021-00430-2. (Visited on 03/08/2024).

Fu, Xiaowen et al. (Apr. 2019). "Exploring Network Effects of Point-to-Point Networks: An Investigation of the Spatial Patterns of Southwest Airlines' Network". In: *Transport Policy* 76, pp. 36–45. ISSN: 0967-070X. DOI: 10.1016/j.tranpol.2019.01.004. (Visited on 02/16/2025).

Gade, Kenneth (July 2010). "A Non-singular Horizontal Position Representation". In: *The Journal of Navigation* 63.3, pp. 395–417. ISSN: 1469-7785, 0373-4633. DOI: 10.1017/S0373463309990415. (Visited on 11/05/2024).

Girvan, M. and M. E. J. Newman (June 2002). "Community Structure in Social and Biological Networks". In: *Proceedings of the National Academy of Sciences* 99.12, pp. 7821–7826. DOI: 10.1073/pnas.122653799. (Visited on 05/21/2024).

Guimerà, R. et al. (May 2005). "The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles". In: *Proceedings of the National Academy of Sciences* 102.22, pp. 7794–7799. DOI: 10.1073/pnas.0407994102. (Visited on 04/23/2024).

Guo, Weisi et al. (June 2019). "Global Air Transport Complex Network: Multi-Scale Analysis". In: *SN Applied Sciences* 1.7, p. 680. ISSN: 2523-3971. DOI: 10.1007/s42452-019-0702-2. (Visited on 04/10/2024).

Home – Air Access | Wesgro (Oct. 2021). https://www.wesgro.co.za/air-access/home. (Visited on 11/14/2024).

International Monetary Fund (2024). *World Economic Outlook Database*. (Visited on 10/20/2024).

Jia, Tao, Kun Qin, and Jie Shan (Nov. 2014). "An Exploratory Analysis on the Evolution of the US Airport Network". In: *Physica A: Statistical Mechanics and its Applications* 413, pp. 266–279. ISSN: 0378-4371. DOI: 10.1016/j.physa.2014.06.067. (Visited on 03/08/2024).

Megginson, David (2024). *OurAirports: Open Data Downloads*. (Visited on 05/01/2024).

Nations, United (Oct. 2024). *Standard Country or Area Codes for Statistical Use (M49)*. (Visited on 10/20/2024).

Newman, M. E. J. (Feb. 2003). "Mixing Patterns in Networks". In: *Physical Review E* 67.2, p. 026126. DOI: 10.1103/PhysRevE.67.026126. (Visited on 07/31/2024).

Newman, Mark (Mar. 2010). *Networks: An Introduction*. Oxford University Press. ISBN: 978-0-19-159417-5. DOI: 10.1093/acprof:oso/9780199206650.001.0001. (Visited on 05/30/2024).

Nonyati, Sibusiso (2020). *Impact of Air Connectivity on Tourism, FDI and Trade: Insight from the Western Cape*.

Otiso, Kefa M. et al. (Apr. 2011). "Airline Connectivity as a Measure of the Globalization of African Cities". In: *Applied Geography* 31.2, pp. 609–620. ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2010.12.002. (Visited on 03/12/2024).

Pearson, James (Feb. 2022). *Eurowings Discover Adds Kruger National Park Using Airbus A330s*. https://simpleflying.com/eurowings-discover-kruger/. (Visited on 11/15/2024).

Reichardt, Jörg and Stefan Bornholdt (July 2006). "Statistical Mechanics of Community Detection". In: *Physical Review E* 74.1, p. 016110. DOI: 10.1103/PhysRevE.74.016110. (Visited on 05/30/2024).

Rocha, Luis E. C. da (Apr. 2009). "Structural Evolution of the Brazilian Airport Network". In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.04, P04020. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2009/04/P04020. (Visited on 04/23/2024).

Saramäki, Jari et al. (Feb. 2007). "Generalizations of the Clustering Coefficient to Weighted Complex Networks". In: *Physical Review E* 75.2, p. 027105. DOI: 10.1103/PhysRevE.75.027105. (Visited on 08/01/2024).

Sen, Amartya (1997). "On Economic Inequality". In: *On Economic Inequality*. Expanded ed. Oxford: Clarendon Press. ISBN: 0-19-829297-X.

Services – Air Access | Wesgro (Feb. 2025). https://www.wesgro.co.za/air-access/services. (Visited on 02/16/2025).

Song, Min Geun and Gi Tae Yeo (Sept. 2017). "Analysis of the Air Transport Network Characteristics of Major Airports". In: *The Asian Journal of Shipping and Logistics* 33.3, pp. 117–125. ISSN: 2092-5212. DOI: 10.1016/j.ajsl.2017.09.002. (Visited on 04/10/2024).

Traag, V. A., L. Waltman, and N. J. van Eck (Mar. 2019). "From Louvain to Leiden: Guaranteeing Well-Connected Communities". In: *Scientific Reports* 9.1, p. 5233. ISSN: 2045-2322. DOI: 10.1038/s41598-019-41695-z. (Visited on 05/16/2024).

United Nations, Department of Economic and Population Division Social Affairs (2024). *World Population Prospects 2024, Online Edition*. (Visited on 10/20/2024).

Walsh, Willie, Yvonne Makolo, and Kamil Al Awadhi (2024). *Focus Africa Presentation*. (Visited on 10/21/2024).

Wang, Jiaoe, Huihui Mo, and Fahui Wang (Oct. 2014). "Evolution of Air Transport Network of China 1930–2012". In: *Journal of Transport Geography*. Changing Landscapes of Transport and Logistics in China 40, pp. 145–158. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2014.02.002. (Visited on 10/06/2024).

Wang, Jiaoe et al. (July 2011). "Exploring the Network Structure and Nodal Centrality of China's Air Transport Network: A Complex Network Approach". In: *Journal of Transport Geography* 19.4, pp. 712–721. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2010.08.012. (Visited on 10/06/2024).

Wang, Yishu et al. (Dec. 2019). "Time-Dependent Graphs: Definitions, Applications, and Algorithms". In: *Data Science and Engineering* 4.4, pp. 352–366. ISSN: 2364-1541. DOI: 10.1007/s41019-019-00105-0. (Visited on 04/07/2024).

Wasserman, Stanley and Katherine Faust (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge: Cambridge University Press. ISBN: 978-0-521-38707-1. DOI: 10.1017/CBO9780511815478. (Visited on 11/04/2024).

Wong, Collin WH, Tommy King Yin Cheung, and Anming Zhang (July 2023). "A Connectivity-Based Methodology for New Air Route Identification". In: *Transportation Research Part A: Policy and Practice* 173, p. 103715. ISSN: 0965-8564. DOI: 10.1016/j.tra.2023.103715. (Visited on 10/08/2024).

Xu, Zengwang and Robert Harriss (Oct. 2008). "Exploring the Structure of the U.S. Intercity Passenger Air Transportation Network: A Weighted Complex Network Approach". In: *GeoJournal* 73.2, pp. 87–102. ISSN: 1572-9893. DOI: 10.1007/s10708-008-9173-5. (Visited on 04/04/2024).

Xu, Zizhen and Shauhrat S. Chopra (July 2022). "Network-Based Assessment of Metro Infrastructure with a Spatial–Temporal Resilience Cycle Framework". In: *Reliability Engineering & System Safety* 223, p. 108434. ISSN: 0951-8320. DOI: 10.1016/j.ress.2022.108434. (Visited on 03/08/2024).