Identifying and analysing the role of key airports and the structure of the air transportation network

Alexander Earl
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, England
line 5: email address or ORCID

Mohammod Ibrahim Bashir
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, England
line 5: email address or ORCID

Savio Siu Huen Fung
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, England
line 5: email address or ORCID

Sultan Ayub
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, England
line 5: email address or ORCID

***Abstract*—In this work, the air transportation network is analysed to identify the most significant airports using the degree, closeness, betweenness and eigenvector centrality measures. The key hubs were mostly located in North America as well as international airports in major cities across the globe. The varying impacts on the network after removing 50% of these airports is then discussed. The network is also identified as not having a core-periphery structure due to the node distribution after removing the top 10% of nodes by degree.**

***Keywords*—network analysis, centrality, core-periphery**

I. Introduction

Air travel is one of the most effective ways for a country to develop. This is through delivering proficient links between key cities, enabling the transportation of people and goods which is vital for a country's economic growth. To expand, a nation's importing and exporting activity can impact its GDP, rate of exchange, and even levels of inflation [1]. For instance, in 2021 the UK's exports were calculated at £625 billion, with imports at £654 billion. This indicates that the UK runs at a trade deficit, meaning the importation is greater than exportation [2].

Air transport networks are in control of transporting millions of individuals and commodities worldwide in an economical manner. This is by means of reducing the overall flight distance and thus time travelled between countries. In 2012, over 115 million passengers travelled using UK air carriers and over 736 million using US air carriers [3].

The research will involve identifying the most significant airports using centrality measures such as degree, closeness, betweenness and eigenvector. By identifying these airports, additional research will explore what happens if half of those airports are shut down due to a disruptive event.

It will also be identified whether the network has a core-periphery structure as this helps analyse the robustness of the network. If the network has a core-periphery structure, removing the core airports would disconnect the others reducing the connectedness of the air transport network.

One of the biggest challenges with air transportation networks is the inflexibility regarding processing troublesome affairs at airports such as pandemics, intense weather, and terrorist threats. In April 2010, a volcano erupted in Iceland, causing 10 million passengers to be delayed with an estimated cost of $1.7 billion. Events such as these adversely influence trade and tourism.

As a result of wanting to eliminate the financial costs associated with disrupted events, there is a growing demand for air transportation networks to have high levels of resilience.

It is important to note that some solutions may not be applicable in a real-world context. For example, network resilience can be improved by increasing the number of routes between airports. However, this solution may not be financially sustainable.

In this report, the global air transport network will be analysed to find information such as key hubs in the network and the effect of removing these core airports. In the remainder of the paper, literature related to the conducted analyses will be discussed and the dataset and network will be presented. Subsequently, the network analysis methodology and results will be provided followed by concluding remarks on the work.

II. Related Work

Reference [6] is a paper on the global analysis of airport and airport regions (nodes and edges) to find local and global components within the global air transport network.

Reference [7] is an analysis of Chinese airports, identifying which are critical airports. 6 core Chinese airports are identified; delays at these locations cascade to the other airports resulting in the operation of the entire network to be compromised. Air traffic dynamics were considered to identify the core airports.

Reference [3] mainly focused on the European Air Transport Network and how to increase the resilience of the

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©2022 IEEE

network to natural weather conditions by adaptively rerouting edges to desired nodes.

III. Dataset and network presentation

The dataset used for analysis of the global air transportation network was acquired from [8] and consists of two files. The nodes spreadsheet represents airports using its International Air Transport Association (IATA) code as well as its geographical location. The edges spreadsheet represents directed routes using the IATA code of the airline that operates flights between the source and destination airport. The node data was last updated in January 2017 and the edge data was last updated in June 2014. As a result, the original network had nodes with 0 degrees which affected the calculation of statistics. For example, the high number of isolated nodes resulted in the number of communities originally being calculated as 2840 whereas the true value was actually 26.

There were also nodes without geographic coordinates, but this only affected the network visualisation as shown in Fig. 1 and Fig. 2. Therefore, the network was modified to remove nodes without degrees and the latitudes and longitudes were added in manually to fix these issues.

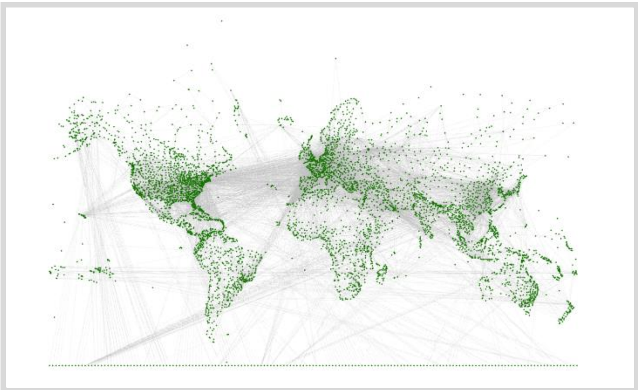

Fig. 1. Network created from original dataset using Gephi and GeoLayout plugin.

Fig 2. Network created from cleaned dataset using Gephi and GeoLayout plugin; isolated nodes have been removed and geographic coordinates added to all nodes without them.

IV. Network analysis methodology

One research problem is to identify key hubs in the air transportation network. A paper on the “Empirical analysis of airport network and critical airports”, showed that airports with higher degree or betweenness are the important nodes in the air transport network [7]. Degree distribution is the number of airports one single airport is connected to, these important nodes in an airport network point to them being key hubs.

Another measure to identify these key hubs is the clustering coefficient of an airport. An analysis on Australian airports showed nodes with a large clustering coefficient can show how “compact” a nodes system of connections to its neighbours are and how a larger clustering coefficient is more likely to be travelled in [9]. Airports that are more likely to be travelled in and have a more compact network compared to others indicate that they are “key hubs”.

An airport being a key hub points to it being important in the network, and that the removal of these nodes would prove to have negative effects on the overall air transportation network and literature shows that nodes with high degree and clustering coefficient are important and therefore key hubs.

Another research problem is discovering if the air transport network has a core-periphery structure. One piece of academic literature that shows how to accomplish this is the “Measurement and Analysis of Online Social Networks”[10]. In order to identify if the network has a core-periphery structure, some core nodes must be identified where removing them would result in the rest of the network consisting of small, disconnected clusters. According to the literature removing nodes with the highest degree-distribution, the Strongly Connected Components (SCC) begins to split them into smaller sizes, showing that those “cores” with a high degree distribution are keeping the remainder or “periphery” nodes connected.

This report also aims to use centrality measures in order to identify significant airports. Similarly, a report on identifying critical airports for controlling infectious diseases and outbreaks used multiple different measures of centrality such as page rank centrality and degree centrality to see which airports had the highest values, the report found that different measures still showed similarities in the airports with higher values, and this shows that they have significance in the network. Therefore, using the airport data, different types of centrality measures can be calculated and compared to identify significant airports.

V. Results and discussion

Table 1. Network statistics calculated using Gephi

Statistic	Value
Average Degree	19.756

Network Diameter	14
Modularity	0.660
Number of Communities	26
Average Clustering Coefficient	0.444
Average Path Length	4.146

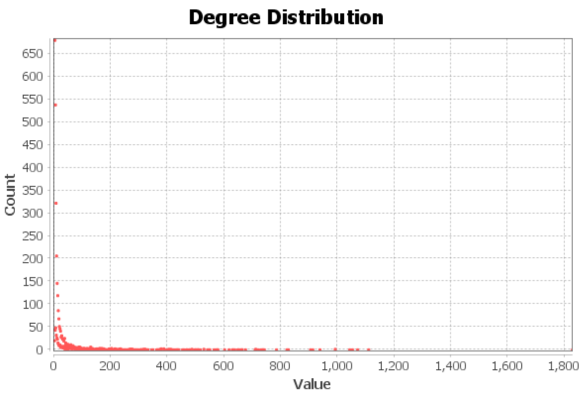

Fig. 3. Degree distribution of the network constructed using Gephi.

The network is a scale-free network as can be identified by the long tail on the degree distribution (Fig. 3) which seems to follow a power law. This means that most nodes have a low number of degrees, but a significant number of nodes have a higher degree than average. These nodes can be classified as hubs. The network also has properties of a small-world network since it has a small average path length of 4.146 and a high average clustering coefficient of 0.444. A strong community structure exists within the network as it has a high modularity value of 0.66.

Table 2. Airports with the highest degrees in the network and their clustering coefficients

IATA Code	Airport	Degree	Clustering Coefficient
ATL	Hartsfield-Jackson Atlanta	1826	0.2619
ORD	Chicago O'Hare	1108	0.2549
PEK	Beijing Capital	1069	0.2048
LHR	Heathrow	1051	0.3110
CDG	Paris Charles de Gaulle	1041	0.2180
FRA	Frankfurt	990	0.2200
LAX	Los Angeles	990	0.3118
DFW	Dallas/Fort Worth	936	0.2138
JFK	John F. Kennedy	911	0.3312
AMS	Amsterdam	903	0.2086

The top 10 airports with the highest degrees can be seen in Table 2 along with their clustering coefficient values. Clustering coefficient denotes how many of the neighbours of a node are connected to each other and therefore how well ‘clustered’ a node is. In the air transport network, a high clustering coefficient of an airport would indicate that the neighbouring airports are not dependent on that airport. All 10 airports are international airports which means they operate flights globally. This could be an explanation for the degree values which are greatly higher than average as well as lower than average clustering coefficient values. Half of the airports are also located in the United States which could identify the USA as the country with the biggest role in the air transportation network.

Table 3. Airports with the highest closeness centrality (1) in the network

IATA Code	Airport	Closeness Centrality
CDJ	Conceição do Araguaia	1
DUT	Unalaska	1
ERS	Windhoek Eros	1
KLN	Larsen Bay	1
KOO	Kongolo	1
SPB	San Luis Obispo County Regional	1
TKJ	Tok Junction	1
AOS	Amook Bay Seaplane	1
SYB	Seal Bay Seaplane	1
PVE	El Porvenir	1
SSB	St. Croix Seaplane Base	1
BLD	Bradley	1
GCW	Grand Canyon West	1
CKX	Chicken	1

All the airports with the highest closeness centrality in the network (a value of 1) are presented in Table 3. Closeness centrality measures how close a node is to all other nodes in the network and therefore is useful to identify how efficiently a node can spread information. In the context of the air transport network, this could identify airports most at risk of spreading epidemics globally. This could also help evaluate the impact on the connectedness of the network if these airports are closed due to disruptive events. Of the 14 airports, 10 are located in North America, 2 in South America and 2 in Africa.

Table 4: Airports with the highest betweenness centrality in the network

IATA Code	Airport	Betweenness Centrality
LAX	Los Angeles	1034522.3989
ANC	Ted Stevens Anchorage	820399.3482
CDG	Paris Charles de Gaulle	813854.2012
LHR	Heathrow	702368.5630
ORD	Chicago O'Hare	664992.4221
PEK	Beijing Capital	651405.3854
DXB	Dubai	634412.4886
FRA	Frankfurt	587555.2971
SEA	Seattle–Tacoma	566562.7216
GRU	São Paulo	521839.3744

Table 4 displays the top 10 airports with the highest betweenness centrality in the network. Betweenness centrality represents how many shortest paths between a pair of nodes a node is a part of. In the air transportation network, a high betweenness centrality identifies airports that most act as intermediaries between other airports. Only 4 of the 10 airports are located in the USA, the smallest proportion of all the different centrality measures.

Table 5: Airports with the highest eigenvector centrality in the network

IATA Code	Airport	Eigenvector Centrality
ATL	Hartsfield-Jackson Atlanta	1
LHR	Heathrow	0.8393
ORD	Chicago O'Hare	0.7634
JFK	John F. Kennedy	0.7565
LAX	Los Angeles	0.7159
CDG	Paris Charles de Gaulle	0.6565
FRA	Frankfurt	0.6082
PEK	Beijing Capital	0.5518
DFW	Dallas/Fort Worth	0.5462
AMS	Amsterdam	0.5210

Table 5 presents the 10 highest airports by eigenvector centrality. Eigenvector centrality assigns scores to each node so that the final centrality value of a node is relative to the scores of its neighbours. The eigenvector centrality of an airport indicates the importance of the airport taking into consideration the importance of neighbouring airports. The 10 airports in table 5 are the same 10 airports in table 2. This suggests that there is a correlation between the degree of an airport and its significance in the network.

Table 6: Network statistics after removing 50% of highest degree nodes (ATL, PEK, CDG, LAX, JFK):

Statistic	Value
Average Degree	18.110
Network Diameter	14
Modularity	0.673
Number of Communities	40
Average Clustering Coefficient	0.436
Average Path Length	4.210

After removing 50% of the highest degree nodes, the average degree of the network decreased. All other statistics increased except network diameter which stayed the same.

Table 7: Network statistics after removing 50% of highest closeness centrality nodes (CDJ, ERS, KOO, TKJ, SYB, SSB, GCW):

Statistic	Value
Average Degree	19.791
Network Diameter	14
Modularity	0.650
Number of Communities	28
Average Clustering Coefficient	0.444
Average Path Length	4.141

After removing 50% of the highest closeness centrality nodes, the modularity of the network and average path length decreased. Average degree and number of communities increased while network diameter and average clustering coefficient stayed the same.

Table 8: Network statistics after removing 50% of highest betweenness centrality nodes (LAX, CDG, ORD, DXB, SEA):

Statistic	Value

Average Degree	18.565
Network Diameter	14
Modularity	0.673
Number of Communities	45
Average Clustering Coefficient	0.438
Average Path Length	4.253

After removing 50% of the highest betweenness centrality nodes, the average degree and average clustering coefficient of the network decreased. All other statistics increased except network diameter which stayed the same.

Table 9: Network statistics after removing 50% of highest eigenvector centrality nodes (ATL, ORD, LAX, FRA, DFW):

Statistic	Value
Average Degree	18.118
Network Diameter	14
Modularity	0.669
Number of Communities	66
Average Clustering Coefficient	0.434
Average Path Length	4.210

After removing 50% of the highest eigenvector centrality nodes, the average degree and average clustering coefficient of the network decreased. All other statistics increased except network diameter which stayed the same.

In all scenarios, network diameter stayed the same which suggests the removed airports were not essential to the shortest path between the two most distant airports. The number of communities also either stayed the same or increased. This could be because these airports tend to operate more flights internationally which limits the formation of communities.

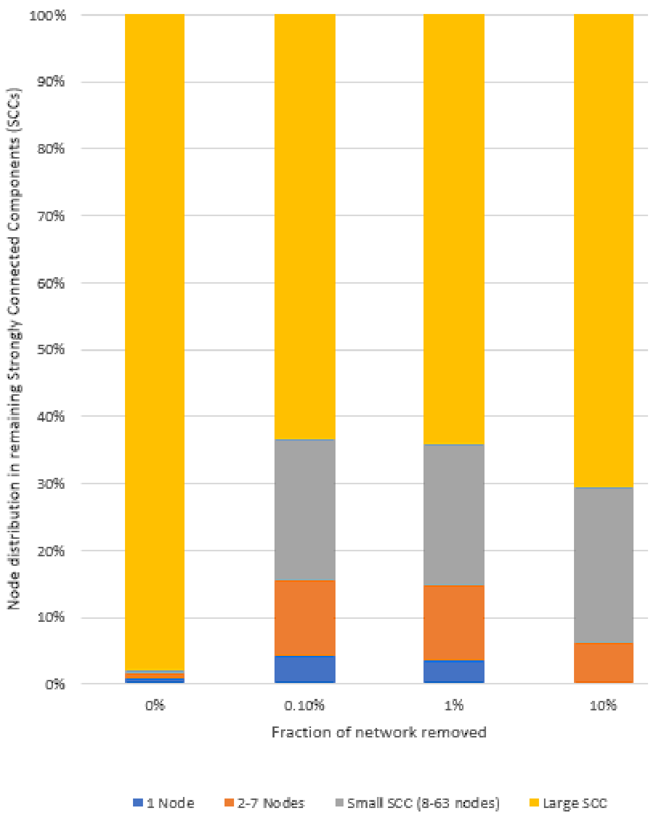

Fig. 4. Node distribution of the network after removing the highest degree nodes from the network

Figure 4 shows the node distribution of the network's SCCs including the initial state and after removing up to the top 10% of the highest degree nodes. In the network, the large SCC contain 98% of the nodes including the top 10% of nodes by degree. Therefore, it can be concluded that the network does not have a core-periphery structure as removing these nodes does not significantly affect the connectedness of the graph.

VI. Conclusions and Perspetives

To conclude this report, the research problems that were covered will be summarised and perspectives on how the report could be extended will be discussed. Calculating various network statistics on the air transport network has given insight into it about the structure, calculating the degree distribution of the nodes showed that the network is a scale free network, and that the network consists of a significant number of hubs, which display a high degree distribution. This may mean that the network has a heavy reliance on these 'hub' nodes and that without them it would cause disruptions in the network especially with most of these nodes being in the US, as this means the air transport network is heavily relying on this country.

Using different methods of centrality, closeness centrality, betweenness centrality and eigenvector centrality, it was uncovered that for some of these measures the top valued airports were common and this shows significance in these airports. To further show the significance of airports with high centrality measures, after removing 50% of the airports with the highest centrality values, the number of communities remained the same as well as the network diameter meaning overall high centrality nodes did not disrupt the network.

Finally, when evaluating if the air transport network has a core periphery structure, after highlighting nodes with high degree and progressively removing them it was found that the networks connectivity was not significantly changed, therefore the air transport network doesn't have a core periphery structure, in other calculations however it was found that the network has a strong community structure due to its modularity.

References

[1] "How Importing and Exporting Impacts the Economy", *Investopedia*, 2022. [Online]. Available: https://www.investopedia.com/articles/investing/100813/interesting-facts-about-imports-and-exports.asp#:~:text=A%20high%20level%20of%20imports,productivity%20over%20the%20long%20run. [Accessed: 01- May- 2022].

[2] "E1 Trade", *House of* Commons, 2022. [Online]. Available: https://researchbriefings.files.parliament.uk/documents/SN02815/SN02815.pdf. [Accessed: 01- May- 2022].

[3] S. Dunn and S. Wilkinson, "Increasing the resilience of air traffic networks using a network graph theory approach", *Transportation Research Part E: Logistics and Transportation Review*, vol. 90, pp. 39-50, 2016. Available: 10.1016/j.tre.2015.09.011.

[4] I. Diop, C. Cherifi, C. Diallo and H. Cherifi, "Revealing the component structure of the world air transportation network", *Applied Network Science*, vol. 6, no. 1, 2021. Available: 10.1007/s41109-021-00430-2.

[5] W. Cong, M. Hu, B. Dong, Y. Wang and C. Feng, "Empirical analysis of airport network and critical airports", *Chinese Journal of Aeronautics*, vol. 29, no. 2, pp. 512-519, 2016. Available: 10.1016/j.cja.2016.01.010.

[6] "OpenFlights.org: Flight logging, mapping, stats and sharing", *Openflights.org*, 2022. [Online]. Available: https://openflights.org/. [Accessed: 01-May- 2022].

[7] M. Hossain and S. Alam, "A complex network approach towards modeling and analysis of the Australian Airport Network", *Journal of Air Transport Management*, vol. 60, pp. 1-9, 2017. Available: 10.1016/j.jairtraman.2016.12.008.

[8] A. Mislove, M. Marcon, K. Gummadi, P. Druschel and B. Bhattacharjee, "Measurement and analysis of online social networks", *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07*, 2007. Available: 10.1145/1298306.1298311 [Accessed: 01 May 2022]