An Analysis of Global Aviation Networks

Faculty of Science, Department of Computer Science, University of Calgary
Authors: Anika Achari, Afnan Chaudhry, and Ejaaz Lakhani
UCID: 30096828, 30077054, 30025543

Abstract

Air transport is among the most important modes of transportation. Consequently, analysis of aviation networks is vital to improving the efficiency and connectivity of air travel in both the developed and developing world. Social Network Analysis (SNA) was applied to analyze a global aviation network. Databases relating to airports, global flight routes, and city population were acquired through Kaggle and GeoNames. In this network, nodes are represented by source and destination airports and flight routes between them act as links. The SNA methodology involved exploring centrality measures, degree distribution, and clustering coefficient measures. Comparisons to appropriate null models were also made. The results of analyzing each centrality measure allowed for the identification of important airline hubs around the world. Furthermore, the analysis was also concerned with identifying important airline hubs in North America and understanding the relationship between population size and airport influence.

It was revealed that airports in London and Paris had the highest degree, betweenness, and eigenvector centrality values. Additionally, it was observed that larger nodes are not associated with higher centrality values and have no major impact on how influential an airport is in the network. The results, although representative of a small number of airports in the network, are critical to understanding the dynamics and structure of the aviation industry. These results demonstrate which airports are the most influential and critical to the efficiency of global aviation. Although the work is in an infancy stage, more comprehensive analysis can lay the foundation for changes that should be implemented to improve aviation, combat airport mismanagement, and improve the performance of airports that are currently failing to fulfill their required role.

Keywords: Aviation network; social network analysis; centrality measures; degree distribution.

I. Introduction

Aviation is among the world's leading industries as it facilitates trade, encourages tourism, generates significant economic growth, creates millions of jobs, provides support for remote communities, and enables rapid responses in the wake of global disasters[1,2]. As an industry, it serves to connect peoples, cultures, and societies across the globe and is thus, critical to the continued success and development of the modern world[1]. Furthermore, air transportation has become increasingly important in both the developed and undeveloped world, given its inherent safety, ease of transport, speed, and affordability when compared to other modes of transportation[1]. Expectedly, the aviation sector has expanded rapidly around the world since 2001[3]. Global air transportation and travel was significantly impacted in the aftermath of the September 9/11 terrorist attacks, which resulted in the implementation of major changes and in some cases, entire restructurings of aviation networks around the world[3]. In the years following 2001, aviation networks expanded to include new airports and routes, which led to certain global traffic and topological properties displaying tremendous growth[3]. For example, in 2017, the total number of passengers that were annually transported through the global aviation network exceeded 4 billion, and this level of growth is expected to continue in the coming years, with Airbus predicting an annual growth of 4.4% in revenue for the next two decades[4]. Given that the aviation industry is multifaceted, aviation networks are highly dynamic; with this dynamism contributing directly to the functioning and evolution of the network[1]. Therefore, developing a better understanding regarding the dynamics and structure of aviation networks is a fundamental step that is required to optimize global aviation. This research aims to identify the most significant airline hubs in the world and in North America to gain insight how robust and well-connected aviation currently is, which will hopefully lay the foundation for improving the industry in the future.

II. Research Questions

The overarching objective of this research was to understand the dynamics and structure and of the global aviation network to identify which airports are significant. This objective was achieved through three, specifically designed questions:

- Question one: Which airports are considered to be global airline hubs?
- Question two: Which airport(s) is the most significant airline hub in North America?
- Question three: Does population size of a city impact the influence of the city's largest airport in the global aviation network?

These three questions were designed to guide our analysis. The first question was intended to identify airports that are imperative to aviation, as this would aid in identifying which nodes are highly connected. We utilized the centrality measures of degree centrality and betweenness centrality to address the first question to identify the most important global airline hubs. The second question was intended to identify the most significant hub in North America. Likewise, degree centrality and betweenness centrality measures were used to address the second research question. Addressing questions one and two was critical to understanding which nodes were highly prolific. Furthermore, addressing questions one and two was necessary to understanding why certain airports may operate at a higher efficiency when compared to others, which may result in changes being implemented to improve travel around the world.

Finally, the third question was intended to discover the importance of population size on connectivity. If a relationship between population size and connectivity can be established, it can be helpful in understanding why certain cities are currently being underserved by the aviation network. To address the third research question, we extracted ten airports with the highest eigenvector centrality values in the network, and then compared these values with the population size of each of the ten cities. We also compared the eigenvector values obtained with population data for the ten largest cities in the world, to determine, if population size impact the influence of the airport in the network. Each question is intended to help us better understand global aviation and lay the foundation for improving it.

III. Methodology

A. Data Collection

Databases relating to air transportation are widely available for the purpose of social network analysis and modelling air transport networks. The databases utilized for this work were obtained through Kaggle, which is an online platform for data science and machine learning. Three independent databases were downloaded from the platform including:

- **Route database**: Database of over 59,036 flight routes between 3,209 airports. Data is ISO 8859 encoded, and each entry contains information relating to code of the airline, unique identifier for airline, code of source airport, identifier for source airport, code of destination airport, identifier for destination airport, and number of stops made by each flight.

- **Airport, airline and route database**: Originally downloaded from Openflights.org under the Open Database license. Data was updated in January 2017 and contains over 10,000 airports across the globe. Each entry contains the airport ID (unique identifier for airport), name of airport, name of city served by airport and name of the country in which the airport is located.

- **World cities population and location database**: Downloaded from the GeoNames geographical databases public platform. Includes city population and geolocation data for the 15,000 cities around the world. The database was last updated in September 2017.

B. Data Preprocessing

Each of the three databases described were subjected to thorough pre-processing before applying Social Network Analysis. Pre-processing was applied to remove unnecessary values present in columns, including (a) missing values, (b) false values, (c) null values and (d) duplicate values. Missing values signify incomplete data, and their presence in a dataset can negatively impact the results. Several columns were also dropped from the Routes database and the Airport, airline, and route database. Following pre-processing that involved the removal of the specific aforementioned data issues (a-d), the three databases were merged to create a single database that is referred to as the "master" database in this paper. This was done using the VLOOKUP function in Excel. VLOOKUP was used to find common identifiers in the population database and other two databases. This common identifier was then used to merge the databases using the VLOOKUP function.

C. Network construction

The global aviation network was constructed using the "master" database. Airports in the database formed the nodes in the network, while flights connecting the airports formed the links. The constructed network was formed by utilizing both source and destination airports. In this network, there were 3209 nodes and 17968 edges. Network construction was achieved using NetworkX, a Python package for the study of the dynamics and functions of complex networks. The visualization (Figure 1) is ideal given that it displays all nodes and links in the aviation network and demonstrates the high level of connectivity among larger cities/airports around the world, while demonstrating that the smallest cities experience a lack of connectivity with the rest of the network, given their distance from the center.

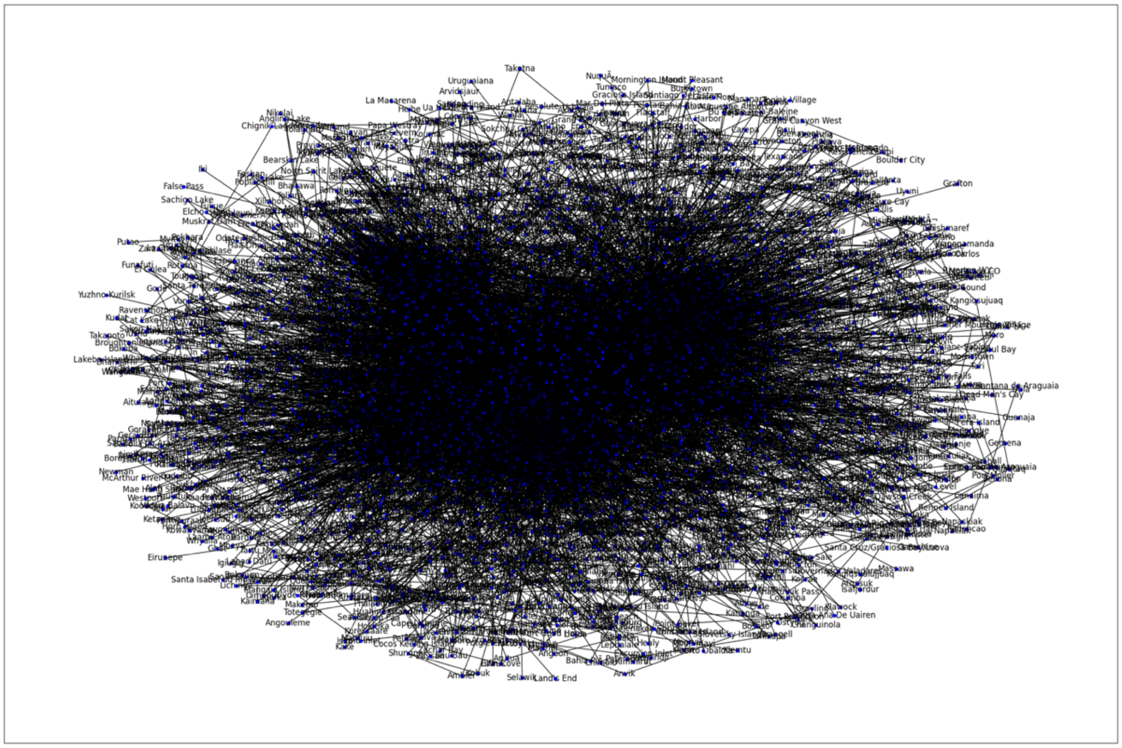


Figure 1. Global aviation network visualization using NetworkX library.

D. Social network analysis

Social network analysis (SNA) refers to the technique of utilizing networks and graph theory to gain further insight into social structures and patterns of relationships among people and groups[5,8]. In SNA, nodes in a specific network are represented by the people and groups while links demonstrate the relationship between nodes[5]. SNA is an essential analytical tool as it permits the study of human interactions and behaviour[5,8]. Consequently, SNA has become a widely employed method in research and business for understanding and interpreting relationships at the various levels in societal hierarchies[5,8]. Network analysis is conventionally based on the following fundamental concepts including network density, centrality, and betweenness[5,8]. Centrality measures especially, are among the most widely applied SNA measures, as they are vital for the identification of central players in a network[5,8]. Furthermore, centrality measures are applied for the purpose of quantifying the importance of each element in a network[5,8].

IV. Results

A. Overview of primary network statistics

Source and destination airports contained in the "master" database were utilized to construct the nodes of our global aviation network, while flight routes between airports served as links. The global aviation network constructed consists of 3209 nodes and 17968 edges. Connected components refers to the maximal set of nodes such that each pair of nodes is connected by a path. The global aviation network is comprised of four connected components. This demonstrates that there are groups of airports that are separate from all other groups of airports. Network density is the ratio of observed edges to the number of possible edges for a given network. Density is critical to understanding the level of connectivity that exists within a single network and is a vital measure to employ when comparing multiple networks. The network density for the global aviation network was 0.00349, indicating a low network density for the global aviation network.

A clustering coefficient measures the number of triangles in a graph. Specifically, a clustering coefficient is a measure of the tendency of nodes to cluster together, and for this particular network, the average clustering coefficient value was 0.51453, which indicates that this particular network is experiencing significant clustering of nodes. Average degree for the global aviation network was 11.1985, with a maximum degree of 346 and a minimum degree of 1. Path length could not be determined, given that this network is not connected.

B. Degree Centrality

Degree centrality is defined as the number of neighbors of a node[6,7,8]. It is among the most widely employed centrality measures given that it is the easiest centrality measure to compute[6,7,8]. With regards to this measure, higher values indicate that a given node is more central in the network[6,7,8]. As part of this work, we computed the degree centrality for the world's top thirty airports, however, the table provided (Table 1) displays degree centrality values for ten airports. Heathrow Airport in London, England demonstrates the highest degree centrality score at 0.107 (Table 1). Following Heathrow, Charles de Gaulle Airport in Paris has a degree centrality score of 0.087 (Table 1), followed by a degree centrality of 0.078 for Sheremetyevo International Airport in Moscow, Russia (Table 1).

Table 1. Degree centrality values for top ten airports.

Airport	**Degree centrality**
London	0.107
Paris	0.087
Moscow	0.078
Istanbul	0.076
Amsterdam	0.075
Frankfurt	0.074
Beijing	0.065
Atlanta	0.064
Chicago	0.063
Dubai	0.061

C. Betweenness centrality

Betweenness centrality is used to identify nodes which can be found on the shortest path between two other nodes[6,7,8]. This is another widely employed centrality measure that successfully identifies an individual's role in passing information from one part of a network to another[6,7,8]. Simply put, betweenness centrality is a measure that is vital to improving our understanding of the level of influence a given node has over the flow of information in a network. Similar to the previous centrality measure, Heathrow Airport in London, demonstrates the highest betweenness centrality value of 0.085 (Table 2). Likewise, Charles de Gaulle Airport in Paris displays the second highest betweenness centrality value of 0.074 (Table 2). It can be observed that the third highest betweenness centrality score can be attributed to the Ted Stevens Anchorage International Airport in Anchorage, Alaska, United States (Table 2).

Table 2. Betweenness centrality values for top ten airports.

Airport	Betweenness centrality
London	0.085
Paris	0.074
Anchorage	0.073
Moscow	0.056
Los Angeles	0.053
Tokyo	0.049
Dubai	0.047
Beijing	0.046
Istanbul	0.044
Chicago	0.044

D. Eigenvector centrality

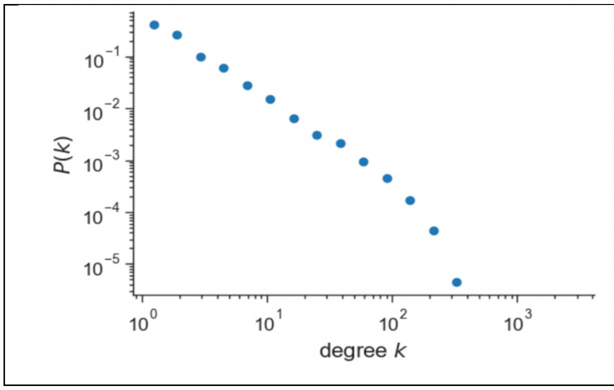
Eigenvector centrality measures the transitive influence of nodes and is utilized to measure the influence of a node in a network[6,7]. This measure assigns each node a centrality value a centrality based on the summation of its links to other nodes[6,7]. In a social network, this measure establishes the significance of a node based on the node's connections to other nodes with high eigenvector centrality values[6,7]. Heathrow Airport in London displays the highest eigenvector centrality value of 0.192 (Table 3), followed by an eigenvector centrality of 0.169 for Charles de Gaulle Airport in Paris (Table 3).

Table 3. Eigenvector centrality values for top ten airports.

Airport	Eigenvector centrality
London	0.192
Paris	0.169
Frankfurt	0.166
Amsterdam	0.164
Munich	0.146
Rome	0.145
Istanbul	0.131
Barcelona	0.123
Zurich	0.122
Brussels	0.120

E. Null model comparisons

To generate appropriate null models for the purpose of comparison, the Erdos-Renyi (E R) model was selected. ER models comprise of two input parameters: the total number of nodes a nd the probability of connection[9]. An Erdos-Renyi random model was generated using the Netwo rkX library in Python. An ensemble of 1000 random ER graphs was generated, after which the m ean and standard deviation for both was calculated.



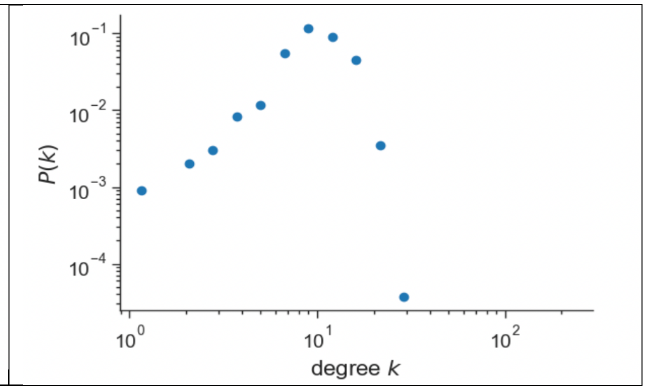


Figure 2. (Left) Degree distribution for global aviation network. Degree (k) is shown on x-axis and probability(P(k)) is shown on the y-axis.
Figure 3. (Right) Degree distribution for ensemble of 1000 random ER graphs. Degree (k) is shown on x-axis and probability(P(k)) is shown on the y-axis.

The graphs display the degree distribution for the global aviation network (Figure 2) an d an ensemble of 1000 randomly generated ER graphs. Note the difference in these distributions. The x-axis for both graphs displays the k-spectrum for the degree distribution. It can be observed that for the ensemble of ER networks, there is a peak in the graph, which is not observed in the d egree distribution for the global aviation network. For the global aviation network, the slope of th e degree distribution decreased with time, suggesting that more important airports have larger de grees. Analyzing degree distribution is imperative in this work to identifying airports that are im portant with regards to their topological positions. It is observed that the mean for clustering and shortest path are larger when compared to the standard deviation values for the ensemble of 1000 random ER graphs (Table 4).

Table 4. Mean and standard deviation for clustering and shortest path for ensemble of ER graphs.

	Mean	**Standard deviation**
Clustering (ER)	0.0034855924776271987	0.00023788386668038776
Short Path (ER)	3.6140273575954165	0.008364214314598014

V. Discussion

A. Interpretation of results

Centrality analysis was performed using the measures of degree, betweenness and eigenvector centrality. Degree centrality assessment revealed the cities of London, Paris, and Moscow displayed the highest degree centrality values, suggesting that the major airports in each of these cities has a large number of connections to other nodes, or airports, in the global aviation network. This measure of centrality is indicative of which nodes in a particular network are highly connected or comparatively more important than other nodes. Therefore, we can assume that London, Paris, and Moscow are the three most important and highly connected nodes in the network. An assessment of degree betweenness centrality revealed that London, Paris, and Anchorage have the highest betweenness centrality values within the network.

Degree centrality analysis established the importance of both London and Paris, and therefore, the high betweenness centrality scores for both of these cities is not surprising. Anchorage however, demonstrated the third highest betweenness centrality value in the network. Ted Stevens Anchorage International Airport is among the world's leading cargo hubs[10]. It serves as the home to hubs for UPS and FedEx[10]. Furthermore, Anchorage is equidistant from New York City and Tokyo, and thus, all air cargo traffic between North America and Asia passes through this airport[10]. In short, betweenness centrality revealed high scores for London, Paris, and Anchorage, suggesting that these are important 'bridges' in the aviation network.

Eigenvector centrality analysis revealed that London, Paris, and Frankfurt displayed the highest eigenvector centrality values in the global aviation network. Eigenvector centrality measures the influence of a node in a network and is often considered to be an extension of degree centrality[6,7,8]. Eigenvector centrality results indicate that London and Paris are the most influential, further suggesting that both London and Paris are highly connected nodes that each have several connections, thereby, making them the most influential nodes in the global aviation network. Note that London and Paris demonstrated consistently high scores for each centrality measure that we explored.

Centrality measure analysis helped in addressing the first research question pertaining to which airports are considered to be global airline hubs. Our analysis would suggest that London and Paris are the two most important global airline hubs given their level of connection and influence within the network. The second research question pertained to identifying the most significant airline hubs in North America. Atlanta was found to have the highest degree centrality of all North American cities, Los Angeles had the highest betweenness centrality, and furthermore, no North American cities possessed high eigenvector centrality scores. Considering that this question only pertained to identification of significant hubs in North America, we did not expect to see any North American cities with high eigenvector centrality values, as this would indicate that an airport is influential in the entire network, when, in reality, they would only be influential in North America. Thus, we identified Atlanta and Los Angeles as being the two most significant hubs in North America.

The third question pertained to identifying if a relationship exists between population size and the influence of that city's largest airport in the global aviation network. To address this question, we utilized our previous analysis of eigenvector centrality. Although London was shown to have the highest eigenvector centrality score, followed by Paris, it is important to consider that London and Paris are not the largest cities with regards to population size within the top ten cities (Table 3). Istanbul is the largest city in Europe; however, it demonstrates the seventh-highest eigenvector centrality (Table 3). Furthermore, we can observe that Frankfurt has an eigenvector centrality of 0.166 (Table 3) and Amsterdam has an eigenvector centrality value of 0.164 (Table 3). However, Amsterdam has a larger population at 1,166,000 when compared to Frankfurt at 791,000. Likewise, Zurich has a higher eigenvector centrality value of 0.122 (Table 3), when compared to Brussels, which has a value of 0.120 (Table 3), despite being a smaller city with regards to population (Table 3). Furthermore, the world's four largest cities with regards to population size including Tokyo (Population of 37,435,191), Delhi (29,399,141), Shanghai (Population of 26,317,104) and Sao Paulo (Population of 21,846,507) are not listed among the top ten cities with regards to eigenvector centrality, suggesting that larger cities do not have a larger eigenvector centrality, meaning, that larger cities, including some of the world's largest cities, do not have the most influence in the network. Therefore, it is not true that larger cities have airports that are more influential or significant in the global aviation network.

Our approach to conducting SNA was successful, given that all three research questions were answered using specific SNA measures. We would like to acknowledge that our results deviated from what was expected. Although we were successful in identifying the most important global airline hubs and in North America based on the data we had acquired, they were not the results that were expected, and this surprised us. It also taught us that aviation networks are highly complex and rely on numerous different factors, many of which are never within our control. We also learned that the network we constructed is extremely large and it is impossible to conduct analysis that accounts for all nuances in it. Our approach to analyzing this network taught us that in the future, employing community detection to partition the network will serve a greater purpose and allow us to answer more targeted questions relating to aviation.

B. Limitations

Although this work does provide insight into the structure of a global aviation network, further analysis is required if the ease of aviation travel and global connectivity is to be improved in the future. Our analysis relied on aviation data that was initially collected in 2012 and was updated last in 2017. Aviation is a fast-growing industry that will only continue to expand as the population of the world increases[1,2]. Therefore, analysis that relies on data collected over five years ago, does not accurately reflect the structure and dynamics of the current global aviation network. Furthermore, our approach to conducting SNA did not include community detection analysis. Community detection may have provided greater insight into the differences and similarities between airports and airliners in each continent. It may have also aided in understanding why certain airports are more influential than others.

C. Conclusions and Future Directions

Although this work does provide insight into the dynamics and structure of global aviation, it is imperative to note that such analysis must be considerably strengthened if global aviation is to be successfully improved. Aviation has successfully established global connectivity; however, significant strides are still required to reduce inequalities between individuals and countries. Currently, a manifold of obstacles plagues third world and developing nations. For example, many airliners in Asia, Latin America and Africa have struggled with mismanagement, lack of planning and aviation accidents, leading to their inability to maintain their share of the aviation market.

Our analysis focused primarily on the identification of global airline hubs; however, continent-specific, and even country-specific analysis would likely be more advantageous. Improving aviation relies heavily on identifying regions across each continent that are suffering from a lack of connectivity. Our analysis revealed that London is arguably, the most important hub for global aviation. Thus, if Heathrow Airport were to experience any minor disturbance with regards to operation or congestion, travel around the globe would suffer, especially in rural areas, that depend heavily on aviation to support critical services such as medical transport. To avoid disruptions in global aviation, analysis must be focused on identifying regions across each continent that are currently under-served by major airports will improve connectivity in rural areas, and possibly even developing nations, that are currently not able to benefit in similar ways from the aviation industry. Furthermore, strengthened analysis may even lead to improvements to airports that are operating at low efficiency.

References

1. Luis E.C. Rocha, Dynamics of air transport networks: A review from a complex systems perspective, Chinese Journal of Aeronautics, Volume 30, Issue 2,2017, Pages 469-478, ISSN 1000-9361, https://doi.org/10.1016/j.cja.2016.12.029.

2. Sri Addepalli, Gerardo Pagalday, Konstantinos Salonitis, Rajkumar Roy, Socio-economic and demographic factors that contribute to the growth of the civil aviation industry, Procedia Manufacturing, Volume 19,2018, Pages 2-9, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2018.01.002.

3. Siozos-Rousoulis, L., Robert, D. & Verbeke, W. A study of the U.S. domestic air transportation network: temporal evolution of network topology and robustness from 2001 to 2016. J Transp Secur 14, 55–78 (2021). https://doi.org/10.1007/s12198-020-00227-x

4. Grewe, V., Gangoli Rao, A., Grönstedt, T. et al. Evaluating the climate impact of aviation emission scenarios towards the Paris agreement including COVID-19 effects. Nat Commun 12, 3841 (2021). https://doi.org/10.1038/s41467-021-24091-y

5. Ghali, N., Panda, M., Hassanien, A.E., Abraham, A., Snasel, V. (2012). Social Networks Analysis: Tools, Measures and Visualization. In: Abraham, A. (eds) Computational Social Networks. Springer, London. https://doi.org/10.1007/978-1-4471-4054-2_1

6. Diallo, S. Y., Lynch, C. J., Gore, R., & Padilla, J. J. (2016). Identifying key papers within a journal via network centrality measures. Scientometrics, 107(3), 1005–1020. https://doi.org/10.1007/s11192-016-1891-8

7. Landherr, A., Friedl, B. & Heidemann, J. A Critical Review of Centrality Measures in Social Networks. Bus Inf Syst Eng 2, 371–385 (2010). https://doi.org/10.1007/s12599-010-0127-3

8. Das, K., Samanta, S. & Pal, M. Study on centrality measures in social networks: a survey. Soc. Netw. Anal. Min. 8, 13 (2018). https://doi.org/10.1007/s13278-018-0493-2

9. Engel, A., Monasson, R. & Hartmann, A.K. On Large Deviation Properties of Erdös–Rényi Random Graphs. Journal of Statistical Physics 117, 387–426 (2004). https://doi.org/10.1007/s10955-004-2268-6

10. https://www.cnn.com/travel/article/anchorage-airport-cold-war-covid/index.html

11. Guo, W., Toader, B., Feier, R. et al. Global air transport complex network: multi-scale analysis. SN Appl. Sci. 1, 680 (2019). https://doi.org/10.1007/s42452-019-0702-2

12. Chernega, O., & Dorofeyeva, K. (2015). Diagnostics of world aviation network development state and features. Economic Annals-XXI, 7-8(2), 8-11. http://ea21journal.world/index.php/ea-v153-02/

13. Diop, I.M., Cherifi, C., Diallo, C. et al. Revealing the component structure of the world air transportation network. Appl Netw Sci 6, 92 (2021). https://doi.org/10.1007/s41109-021-00430-2

14. K. Xue, "Complex Network Variability Analysis with Impact of Major Events on Aviation Networks," *2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, 2021, pp. 32-37, doi: 10.1109/ICHCI54629.2021.00014.