International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

A Graph Mining Study for GCC International Airports

Hazem Bakkar[a*], Asma Qassim AlHamad[b], Lubna Abdelljawad[c]

[a]*British University in Dubai,Dubai, UAE*

[b,c]*Imam Abdulrahman Bin Faisal University,Dammam, SaudiArabia*

[a]*Email: hazem.bakkar@hotmail.com*

[b]*Email: aqalhamad@iau.edu.sa*

[c]*Email: lmabdelljawad@iau.edu.sa*

Abstract

In the last few years, Gulf Cooperation Council countries (GCC) allocated billions of dollars for expanding and upgrading its airports, therefore many studies and researches were held in order to analyze the crucial impact airports have on the economical development of these countries. This study aims to analyze airports network in GCC countries and reveal the facts hidden beneath it. The author prepared a genuine dataset about GCC airports network and represented it as a graph dataset. Using an open source software called Gephi [1], the author applied data mining techniques on the airports graph with the aid of several graphical metrics like degree centrality, betweenness and closeness centrality, and other types of metrics. The author was able to reveal some interesting facts about the airports in GCC countries. These facts showed that airports like Dubai International airport is considered an important continental hub for aviation, while other airports like Kuwait airport has a limited influence and importance in GCC airports network.

Keywords: Datamining; Graph mining; Dataset.

----

* Corresponding author.

1. Introduction

Recently, GCC countries allocated billions and billions of dollars for building new giant airports and expanding existing airports, these bold plans are now taking place on the ground and some of them is about to be accomplished like Dubai International Airport terminal 2 expansion [2]. Lately, Dubai declared a dedicated 7.8 billion dollar budget for upgrading and expanding its existing airports, while Qatar is building a giant airport with an estimated cost of 14.7 billion dollar. Kuwait is using the open-air traffic policy in its sky for all the airlines in order to boost its importance in aviation; also, it has a 6 billion dollar expansion plan for its old airport. In the other hand, Saudi Arabia announced a plan for upgrading and expanding its existing airports with a 20 billion dollars by the end of 2020. Figures released from the last Dubai Airport show expected that GCC countries will be able to transport more than 200 million passengers through its airports.

The previous massive budgets for developing aviation infrastructure in GCC countries encouraged the author to study airports network in these countries. Aiming to identify the most important airports in GCC countries and revealing some interesting relations and facts between GCC airports and other airports in other countries.

In our study, we will analyze a dataset that describes the airlines routs between Gulf countries and other countries in the world. In addition, we will discover and define the main characteristics of this dataset using several metrics and calculations.

2. Dataset

The dataset in this study is genuine; the author collected the airports' information from two authenticated resources:

- http://www.world-airport-codes.com .
- http://www.flightstats.com/go/Home/home.do .

The dataset was saved in tables then it was processed, the results of the processing operation were two tables, one table for nodes which represents the airports and their properties, the other table is for edges, which describes the links between airports in the form of source airport (departure) and target airport (destination). After creating nodes and edges tables our dataset was loaded into a graph visualization program called Gephi [1] for further processing and analysis. The graph consisted of 243 node and 680 edges.

The graph is un-directed and un-weighted one, because the links between the nodes (airports) are two ways links for all the nodes (airports), and all of the links have the same transportation values, because of that our graph is an un-weighted graph.

The nodes table contained the following attributes:

- Nodes, Id, Label, Airport, Latitude, longitude.
- Edges table contained the following attributes:

- Source, Target, type, Id, and weight, since our graph is un- weighted graph, so all the edges have a weight value=1.

Figure (1) shows our graph before preprocessing. While figure (2) shows the graph after applying Force Atlas (2) algorithm [3], Force Atlas (2) algorithm eliminates overlapping between nodes and dissuade hubs in the graph, while figure (3) shows another algorithm that can be applied on the airlines graph, this algorithm is called Fruchterman Reingold algorithm [4].

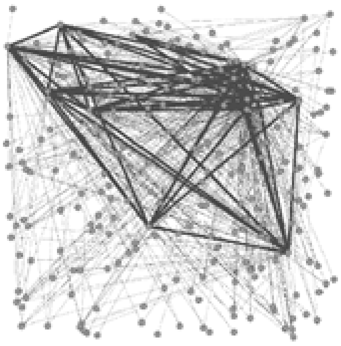

Figure 1: Graph before processing

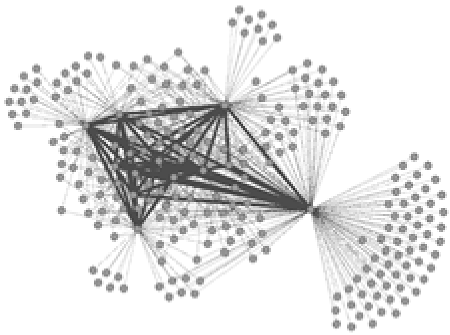

Figure 2: Force Atlas 2 Visualization Algorithm

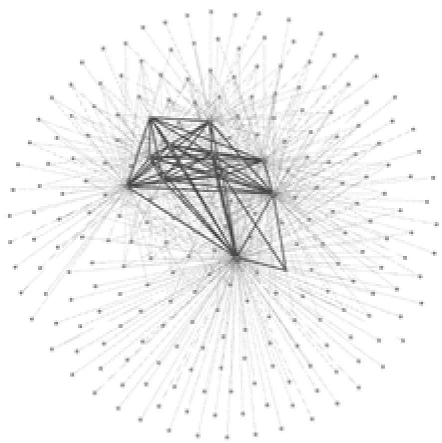

Figure 3: Fruchterman Reingold Visualization Algorithm

2.1 Gephi

The study implemented on the GCC Airlines network was performed using an open source software called Gephi 0.8.[1], Gephi is used for visualizing and analyzing large network graphs. Gephi uses a 3D render engine to display a real time graphs [5].

Gephi has the ability to analyze and perform several types of tests that help the researcher to discover the most important characteristics of a network topology.

In order to analyze and discover interesting patterns hidden in our graph; we used several metrics in our experiments. These metrics are calculated and visualized using Gephi [1]. The following are the metrics studied in our experiments:

3. Experiments

3.1 Degree Centrality

The number of links that are connected to a node identifies the degree of this node. Degree is calculated as follows:

$$K(i) = \sum a_{(i,j)}$$

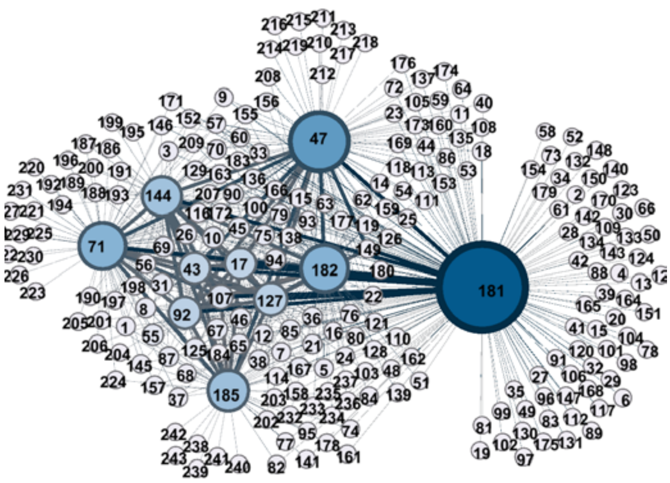

Figure 4: GCC Airports network degrees

Metric is useful for locating which nodes are central in a network with regard to its spreading ability and influence in the network.

In our experiment, we used Gephi [1] to identify the node with the highest degree, the result showed that Dubai

International airport (DXB Dubai, AE) has the highest number of links in all the international airports in GCC countries; Dubai airport has 180 links with other airports around the globe. Figure no. (4) Shows the nodes with their degrees, the values of the degrees are represented by the size of the node. Additionally, Figure (5) shows GCC airports sorted from the highest to the lowest ranked degrees.

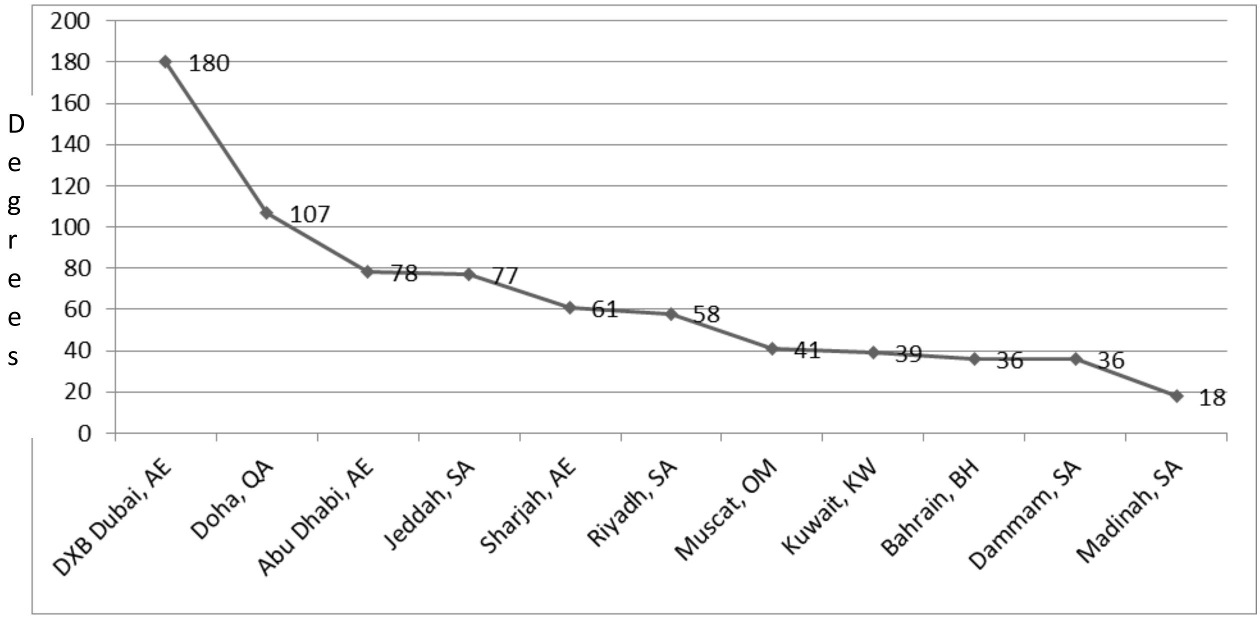

Figure 5: GCC Airports network degrees

Degree metric in this study reveals the high influence that Dubai International airport has in the Aviation transportation industry. So, if Dubai International airport stops operating for any reason a 180 flights will be effected directly or indirectly, and the whole network topology will be changed and maybe damaged. This fact emphasizes the importance of corporation between the nearby airports in case of any emergency shutdown of an important airport like Dubai International airport.

3.2 Betweenness Centrality

Betweeness centrality is defined as the number of short paths that cross through a certain node [6], [7]. Nodes that have high betweenness value usually indicates to high influence in spreading capabilities in a network [8]. In our study, we found that Dubai International airport has the highest value for betweenness metric in GCC airports. This means that Dubai International airport is a crucial connecting airport in the shortest pathways between many pairs of airports in the world. In addition, betweenness metric for Dubai International airport shows that this airport is playing an important role in connecting several geographically separated regions; usually a node with a high betweenness value indicates how important it is in connecting separated communities [8]. It is noted from figure (6) and figure (7) that the ranking order for GCC airports is changed compared to the degree centrality results showed in figure (4) and figure (5). For example, Abu Dhabi Airport has a higher degree than Jeddah Airport, but in betweenness metric Jeddah airport ranking as a connector between other airports is higher than Abu Dhabi airport, so Jeddah Airport can link more pairs of airports than Abu Dhabi airport.

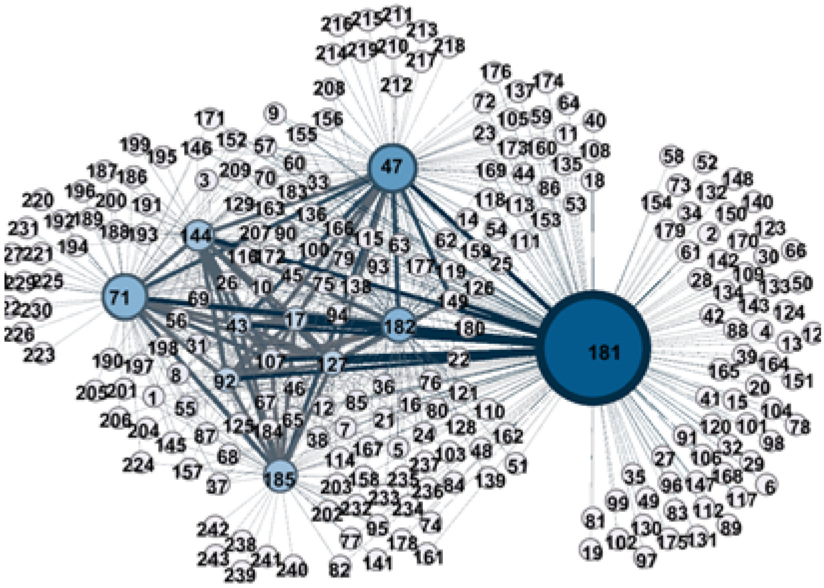

Figure 6: GCC Airports network Betweenness Graph

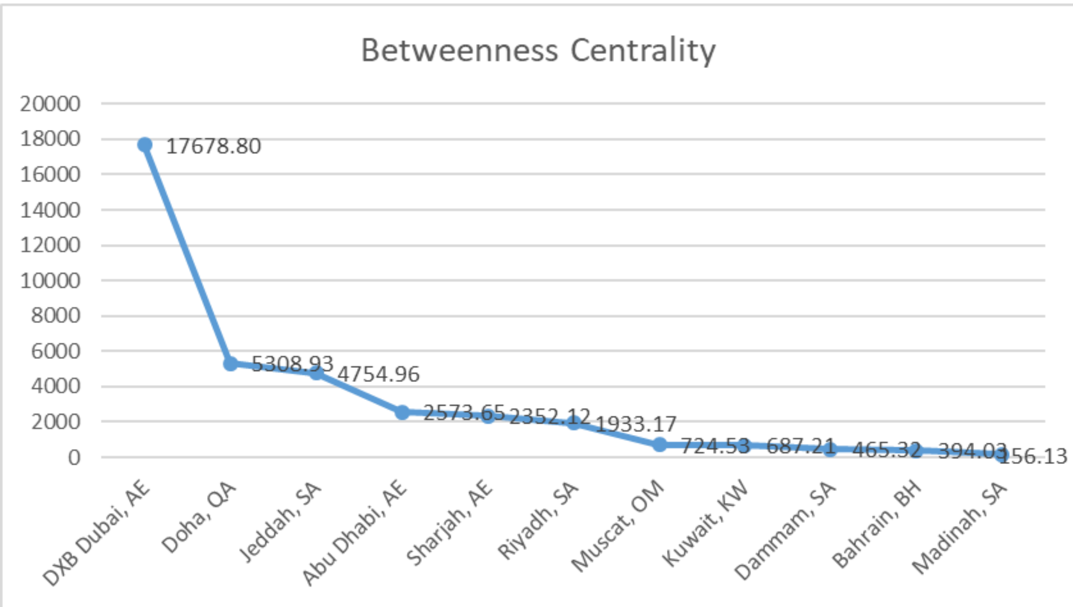

Figure 7: GCC Airports network Betweenness values

3.3 Closeness Centrality

Closeness of a node is defined as the inverse of farness of a node [9], which is defined as the sum of distances between this node and all other nodes in the network. In other words, closeness centrality metric indicates how long a piece of information will take to spread from a given node to all other reachable nodes using the shortest paths in the network [8]. Closeness centrality $C_C(i)$ is calculated by the following equation [9]:

$$C_C(i) = \frac{1}{Farness} = \frac{1}{\sum_{t \in V/i} d_G(i,t)}$$

where $d_G(i,t)$ is the shortest distance between node i and target node t in a network G = (V,E).

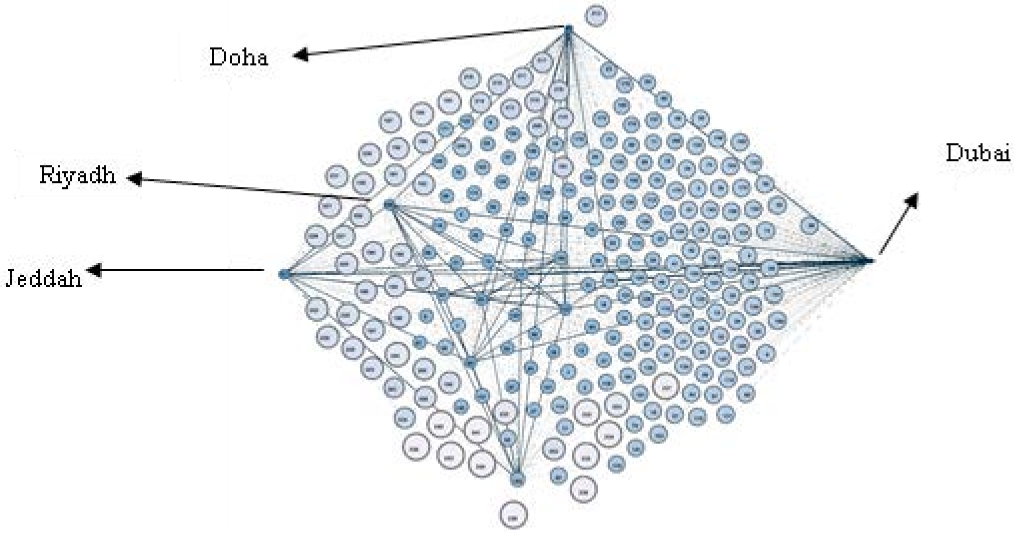

Figure 8: GCC Airports network Closeness Graph

The value of Closeness Centrality metric tends to become smaller and smaller whenever the nodes is closer to the center of a network, this means that the time taken for a piece of information or a vehicle to move from a node to another node will be shorter whenever the closeness centrality is smaller. In our study, the results for the calculations of closeness centrality metric are given figure (8) and figure (9). It is noticed that Dubai Airport has the smallest closeness value, which means that travelling through Dubai airport will save the traveler time when he travels between two airports both of them are connected to Dubai airport and both of are not directly connected to each other. In addition, it is noted that Sharjah airport and Abu Dhabi airport have the highest closeness value even though they both achieved a good results in betweenness and degree calculations; this is because Sharjah and Abu Dhabi airports are not connected to Dubai airport from an aviation prospect. it is known that these three airports are geographically close to each other and this eliminates the need for connecting them by air transportation. We perform an additional experiment on closeness centrality for Sharjah and Abu Dhabi airports, by connecting both of them to Dubai Airport and we repeated our closeness centrality calculations, we found that Sharjah and Abu Dhabi airports' closeness values became smaller and achieved the fourth and sixth ranks respectively. This additional experiment showed the great transportation value gained by an airport when it is connected to an important airport like Dubai International airport.

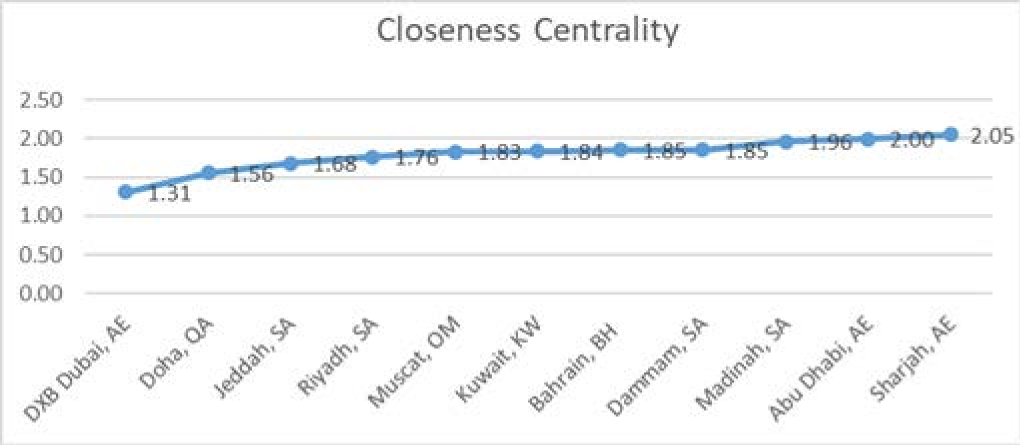

Figure 9: GCC Airports network Closeness Values

3.4 Eigenvector Centrality

The main idea behind Eigenvector centrality is that the importance of a node is not only determined by the node itself, it is also affected by the importance of its neighbouring nodes [8]. for example, in social life a person who is connected to important individuals will make himself important, and this example is also applicable to nodes in airport network.

In our study, we found that Dubai airport achieved the highest value in Eigenvector calculation; this means that Dubai airport is connected to many important airports in the world. Therefore the importance of Dubai International airport is also determined by the importance of its neighbouring airports. Doha airport also achieved a high value in Eigenvector and it is competing Dubai airport, it is also noted that Madinah airport has the lowest Eigenvector value, which makes it the least important airport in GCC airports. The results for Eigenvector calculation is detailed in figure (10) and figure (11).

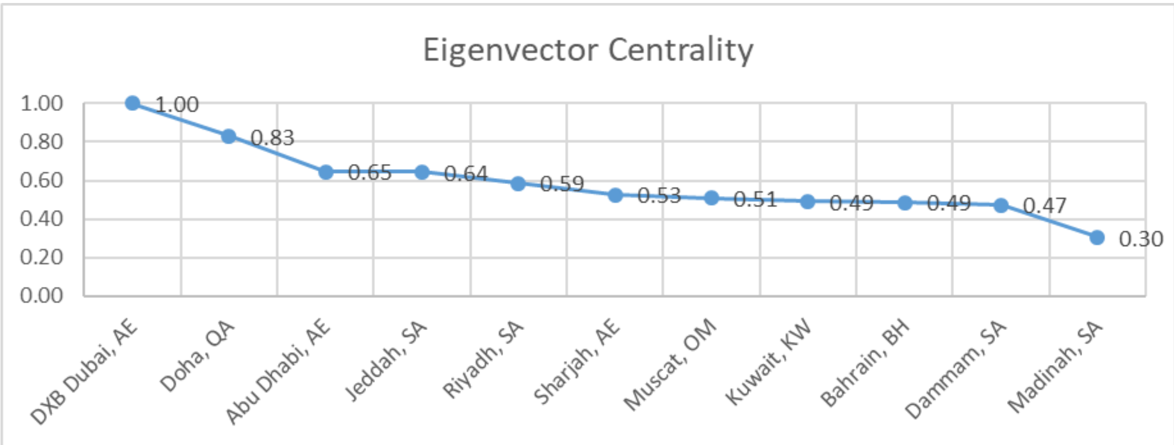

Figure 10: Eigenvector Centrality Values

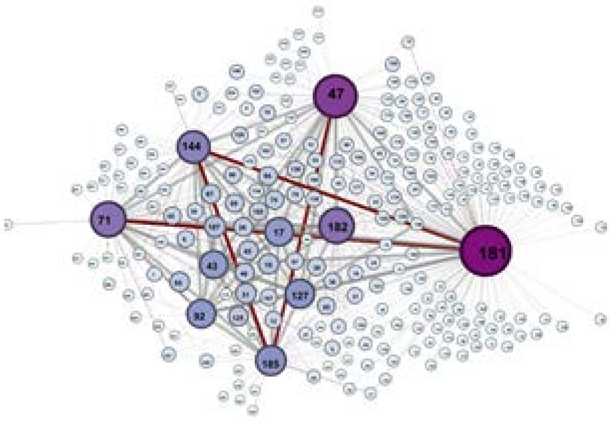

Figure 11: Eigenvector Centrality Graph

3.5 Page rank

Page rank metric was firstly used to evaluate the importance of a web page and to give this page a rank [10], later on the successful application of page ranking algorithm extended its application to other domains. Page ranking algorithm is now used to evaluate the importance of elements in Gene research [11], [12], transportation and traffic [13] and many other domains. Page ranking principle is built on probability calculations. Page ranking calculates the probability of a random walker or visitor using the links in a given network, this walker jumps to a random node with another probability, thus, the page ranking metric will be equal to the probability of this walker to visit a random node (i) at a certain time (t). In our study, we found that Dubai airport is achieving the highest value in page ranking metric, this means that Dubai airport is the most visited airport in all the airports in GCC countries. Also Doha Airport is still competing Dubai airport in terms of the most visited airport, while Madinah airport is the least visited airport in GCC airports. Results are shown in figure (12) and figure (13).

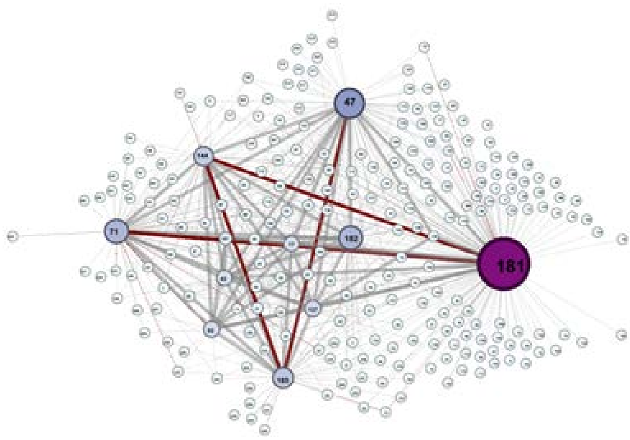

Figure 12: Page Ranking graph

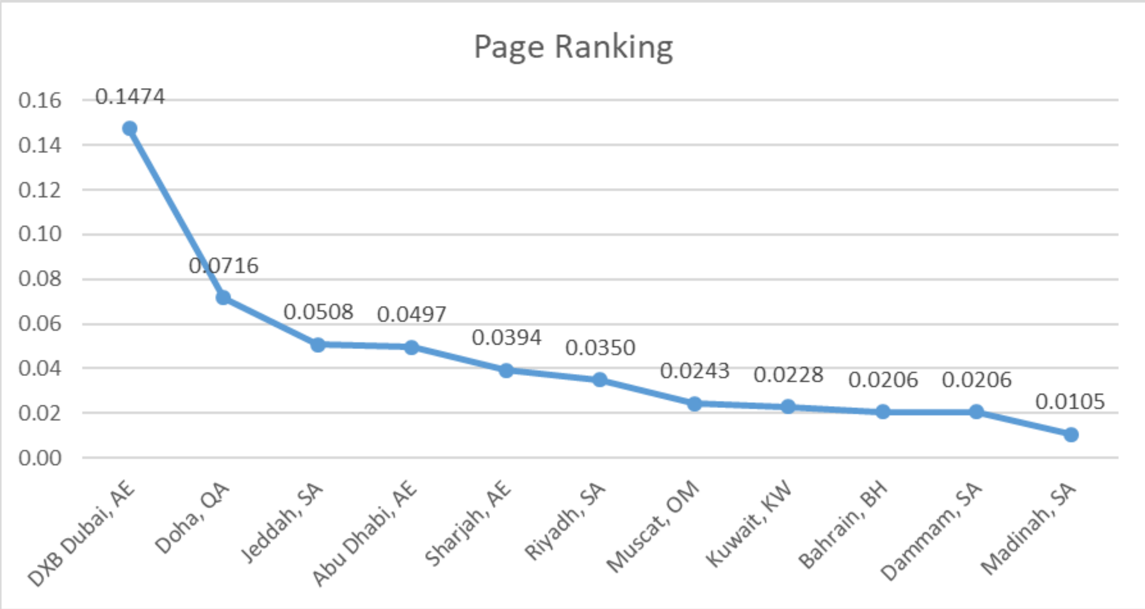

Figure 13: Page Ranking Values

3.6 K-shell index

K-shell index is one of the wide used methods in identifying the location of nodes in a network [14], [15]. K-shell index of a node is obtained using a method called K-decomposition, this method is detailed as follows:

- We start by removing the nodes that have one connection to other nodes, and assign them to 1-shell.
- We remove the nodes with degree of 2 and assign them to 2-shell.
- We repeat the previous step until all the nodes have been assigned to a shell.

The nodes of the highest K-shell represent the centre of the network and these nodes are considered the most influence nodes in this network.

In our study, the graph was clustered using K-decomposition method, the result of the clustering was 10-shell index with 17 central nodes and 113 edges. These remaining nodes are forming the core of the network, and these nodes have the highest spreading capabilities. K-shell index major pitfall that it gives all the remaining nodes the same shell index, which raises the need to use another metric to identify the most influence node in the network. In our study, 17 airports remained after performing the K-decomposition method. This information does not reveal the winner airport as a super spreader, so we will use the degree metric to identify the highest influence airport in the remaining nodes. Table (1) shows the results of combining degree metric with K-shell index; also figure (14) shows the network shape with 1-shell index while figure (15) shows the same network with 10-shell index.

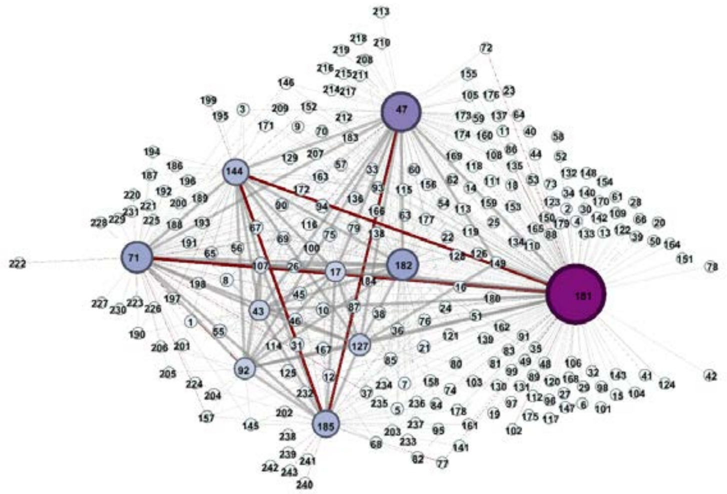

Figure 14: (1)- Shell graph

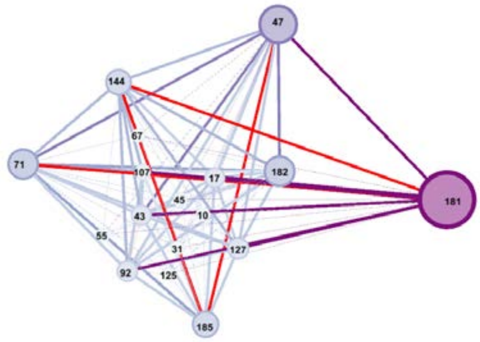

Figure 15: (10)- Shell graph

Table 1: K-shell index sub graph degrees and Degrees Centrality results

No.	***Label***	Node degree in sub graph – 10 shell index	Degree	Airport
1	181	14	180	DXB Dubai, AE
2	47	15	107	Doha, QA
3	182	14	78	Abu Dhabi, AE
4	71	16	77	Jeddah, SA
5	144	16	61	Riyadh, SA
6	185	14	58	Sharjah, AE
7	127	16	41	Muscat, OM
8	92	16	39	Kuwait, KW
9	17	16	36	Bahrain, BH
10	43	15	36	Dammam, SA
11	107	12	18	Madinah, SA

4. Conclusion and Recommendations

This study revealed some important facts regarding air transportation in GCC countries; for instance, the study emphasized the fact that Dubai International airport is the most important airport in GCC countries. Doha airport is competing Dubai International airport, this was clearly noticed in all the results of the metrics used in this study, Doha airport was always the second whenever Dubai International airport is the first. Regarding Sharjah and Abu Dhabi airports, both of them can achieve more influence and importance in the network if they both secure a constant transportation link with Dubai International airport. This link cannot be an air transportation mean, because these airports are geographically close to each other. Such a link can give these three airports a tremendous expansion in their destinations list, also this link will give them the ability to back up each other in the case of emergency shut down or interruption in operation for any of them. Closeness centrality test showed the remarkable value gained by Abu Dhabi and Sharjah airport when the author assumed that there is a link between these two airports and Dubai International airport. Other GCC airports showed a very limited importance and influence in the airports network such as Kuwait airport, Madinah airport, Bahrain airport and Salalah airport. The metrics used in the study showed that the mentioned airports have a limited node degree, closeness and betweenness centralities. This means that these airports need more development and expansion in their links to other airports in order to achieve global importance and gain economic revenues from aviation transportation. Finally, this study showed that an important airport like Dubai International airport should be professionally operated and protected against any emergency circumstances that could interrupt its operation. This kind of interruption will directly affect the transportation traffic in many other neighboring airports, which will eventually create considerable financial losses for passengers, airlines companies and affected airports.

5. Future work

This study has many extendable ideas that can give the researcher wide range of ideas to use for conducting a genuine research, for example, this study shall be extended in order to cover more countries, and eventually all the countries in the world. Another future enhancement for this study is to consider the geographical closeness of airports locations, this will put into consideration the airports that can be linked by transportation means other than air transportation. Air transportation will be considered economically ineffective in such a case, for example, Dubai International airport and Sharjah airport; these two airports can be linked by busses and taxi vehicles and cannot be linked by air transportation because both of them are close to each other. Another future enhancement for this study is to consider the time element, because air traffic in season times can encourage airport management to initiate an exceptional temporary links to other airports, so it is important to test the airports network structure in season times.

References

[1]M. Bastian, S. Heymann and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks", in Proceedings of the Third International ICWSM Conference, San Jose, 2009, pp. 361-362.

[2]G. demand, "GCC states spending $100B in airport expansions to meet passenger demand", Al Bawaba, 2012. [Online]. Available: https://www.albawaba.com/business/gcc-states-spending-100b-airport-expansions-meet-passenger-demand-915412. [Accessed: 10- Jan- 2019].

[3]M. Jacomy, T. Venturini, S. Heymann and M. Bastian, "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software", PLoS ONE, vol. 9, no. 6, p. e98679, 2014. Available: 10.1371/journal.pone.0098679.

[4]T. Fruchterman and E. Reingold, "Graph drawing by force-directed placement", Software: Practice and Experience, vol. 21, no. 11, pp. 1129-1164, 1991. Available: 10.1002/spe.4380211102.

[5]"Gephi - The Open Graph Viz Platform", Gephi.org, 2009. [Online]. Available: https://gephi.org/.

[Accessed: 10- Jan- 2019].

[6]L. Freeman, "Centrality in social networks conceptual clarification", Social Networks, vol. 1, no. 3, pp. 215-239, 1978. Available: 10.1016/0378-8733(78)90021-7.

[7]N. Friedkin, "Theoretical Foundations for Centrality Measures", American Journal of Sociology, vol. 96, no. 6, pp. 1478-1504, 1991. Available: 10.1086/229694.

[8]S. Pei and H. Makse, "Spreading dynamics in complex networks", Journal of Statistical Mechanics: Theory and Experiment, vol. 2013, no. 12, p. P12002, 2013. Available: 10.1088/1742-5468/2013/12/p12002.

[9]G. Sabidussi, "The centrality index of a graph", Psychometrika, vol. 31, no. 4, pp. 581-603, 1966. Available: 10.1007/bf02289527.

[10]S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine", Computer Networks and ISDN Systems, vol. 30, no. 1-7, pp. 107-117, 1998. Available: 10.1016/s0169-7552(98)00110-x.

[11]J. Chen, B. Aronow and A. Jegga, "Disease candidate gene identification and prioritization using protein interaction networks", BMC Bioinformatics, vol. 10, no. 1, p. 73, 2009. Available: 10.1186/1471-2105-10-73.

[12]G. Iván and V. Grolmusz, "When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks", Bioinformatics, vol. 27, no. 3, pp. 405-407, 2010. Available: 10.1093/bioinformatics/btq680.

[13]B. Jiang, S. Zhao and J. Yin, "Self-organized natural roads for predicting traffic flow: a sensitivity study", Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 07, p. P07008, 2008. Available: 10.1088/1742-5468/2008/07/p07008.

[14]P. Erdős and B. Bollobás, "Graph theory and combinatorics", in Graph theory and combinatorics : proceedings of the Cambridge Combinatorial Conference in honour of Paul Erdős, 1984.

[15]S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt and E. Shir, "A model of Internet topology using k-shell decomposition", in Proceedings of the National Academy of Sciences of the United States of America, Boston, 2007, pp. 1150-11154.