Analysis of China's aviation network and the key node

Tianyu Kong

Electrical & Computer Engineering, University of Toronto, Ontario, Canada

tianyu.kong@mail.utoronto.ca

Abstract. Aviation network plays a crucial role in the economic development.However, over these years, the epidemic of COVID-19 seriously influenced the aviation network, including airport shutdowns and a large number of flight cancellations, which causes serious economic losses. Our team examined the topology traits of China's aviation network in our research. The details of the aviation network, including airports as nodes and air routes as edges, are visualised in the study. Essential factors of the network are examined, including degree distribution, clustering coefficient, centrality, etc. Gephi is utilized to visualise the communities of the aviation network. The study analyses these factors and provides a case analysis of how the aviation network will behave if one important node is removed. In the study, it is also discussed the centrality, how the communities change and whether the network collapses to gain a clear understanding of the airports as nodes in the network.

Keywords: Aviation Network, Network Analysis, Node Analysis, Visualization with Gephi.

1. Introduction

The aviation network describes a network system composed of airlines connected in a certain way in a region[1]. Airport transport in China support around one percent of the country's GDP, which shows that air transport plays a crucial role in the economy[2]. However, COVID-19 has caused the airline sector significant economic harm. According to The International Air Transport Association (IATA), there is a $314 billion decline in airline passenger revenue in 2020, which is 55% less than in 2019[3], the key reason for which is the epidemic of COVID-19. In order to curb the epidemic of COVID-19, airports in China would shut down if the city suffers a surge in COVID-19 infections[4].

The project aims to explore the basic characteristics of China's aviation network, the trend of the connectivity of China's aviation network, and how the network will behave to recover by generating new connection communities when an airport is released from lockdown, based on the open source database of the world's aviation network offered by OpenFlights.

2. Related work

In Dorothy and Mehmet's study(2012), the complex network properties are examined, including the average shortest path, degree distribution, clustering coefficient, and betweenness centrality. This study shows that the American air transportation system displays a small-world feature and a distinctive partial power-law degree distribution[5].

Chinese aviation network analysis under actual node attack research by W Sun, G Tang, and P Yang gives an idea of analysing the node attacks to the network by removing the most important node in the network[6].

DOI: 10.54254/2755-2721/5/20230674

3. Research analysis

Our team did preliminary investigations and analyses on the research questions, and further research is planned to be done with our data source and tools. Our justification for each section of our research questions are represented below.

Since March 29, 2020, China had closed most of the international airlines[7]. In addition, the airports would shut down if numerous infections occur[6, 8], which means some nodes in the aviation network might shut down due to the epidemic.

For containment of epidemic recurrence, a flight circuit breaker mechanism, which requires airlines to be suspended where a certain number of passengers are detected as COVID-19 positive, has been implemented in China[9]. Thus, some connections in the aviation network could be closed, where our group is interested in how the aviation network would be changed to make up for the loss of connections.

In order to revitalize the aviation industry from lockdowns caused by the epidemic, China lifts the restriction on hub airport capacity and eliminates restrictions on airlines from aviation hubs to smaller aviation nodes to increase the utilization factor between nodes of the aviation network[10], which inspire our ideas on how the aviation network behave to recover.

4. Data source

For the final project, we chose to analyze an undirected and unweighted network of reoccurring flights from OpenFlights.org. This database includes sections of airports and routes.

The airport database contains more than 10,000 airports and other transport terminals all over the world since January 2017. Every piece of data contains specific information, including airport ID, airport name, location of the airport(city and country), IATA code, ICAO code, detailed position including latitude and longitude, timezone, etc.

Since June 2014, 67,663 routes on 548 international airlines connecting 3,321 airports have been recorded in the route database. The information in the database includes the source airport, destination airport, stops, etc.

5. Properties to be examined

Degree distribution is calculated to find out whether the Chinese aviation network displays the power-law behaviour.

Network density measures how interconnected each node city is, It reflects the overall distribution and tightness of the entire Chinese aviation network.

The average path length, which measures the effectiveness of the network, is the average of the shortest paths' edge numbers between any two nodes.

The clustering coefficient describes the local cohesion of a network.

Degree centrality directly describes the significance of degree connection between one node and every other node.

Closeness centrality shows the proximity degree between one node and every other node.

Betweenness centrality is a control index that shows measure of a single node city which becomes the intermediary of passenger flow linkage between other node cities.

6. Tools to be used

Gephiis a visualization and exploration tool for network graphs.

NetworkX is to create and manipulate complex networks.

NumPy and pandas are used for organizing and handling the data frame.

Matplotlib is to perform the visualisation, such as histogram, scatter plot, and heatmap.

NetworkX and Basemap to visualise the airport in the exact 2-D geographical position.

Python pandas is a tool to read airport and route datasets. With the Networkx python package, the topology and characteristics of the Chinese aviation network will be found. Matplotlib python package is used in the project, to visualise the degree distribution.

The study filtered out the airports and routes dataset in China with python, and input the data into a CSV document, which Gephi can read. Then the data need to be checked out by Gephi and Excel tools for data-cleaning work by deleting the wrong nodes because the data in the datasets of airports and routes do not match well. The study also used Gephi to collect the basic characteristics of the Chinese aviation network, including degree distribution, modularity report, average-shortest-path length, etc., convincing that the data calculated with python are correct. The geographic position of the network is also generated, which makes it evident for reading and analysing.

7. Network characteristics

7.1. Degree distribution

The degree distribution of a node in the network defines the probability distribution of that node's degree.

According to the following equation, P(k) is the cumulative degree distribution, N represents the number of cities at each node, and k represents the total number of routes at each node.

$$P(k) = \sum_{k'=k}^{N} p(k') \#(1)$$

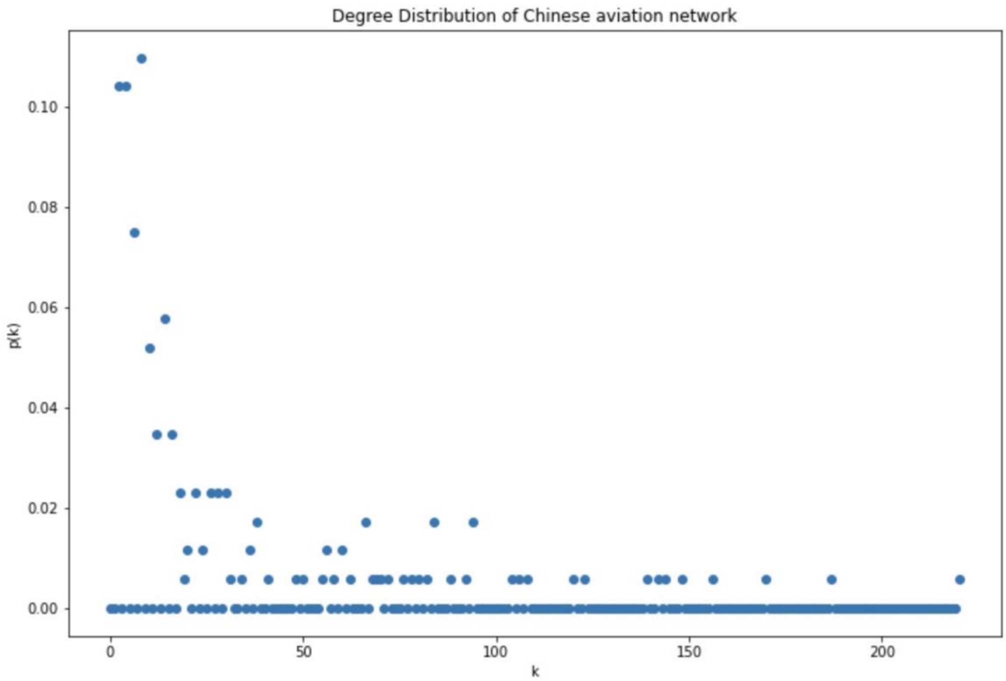

Figure 1 . Degree distribution with original coordinates.

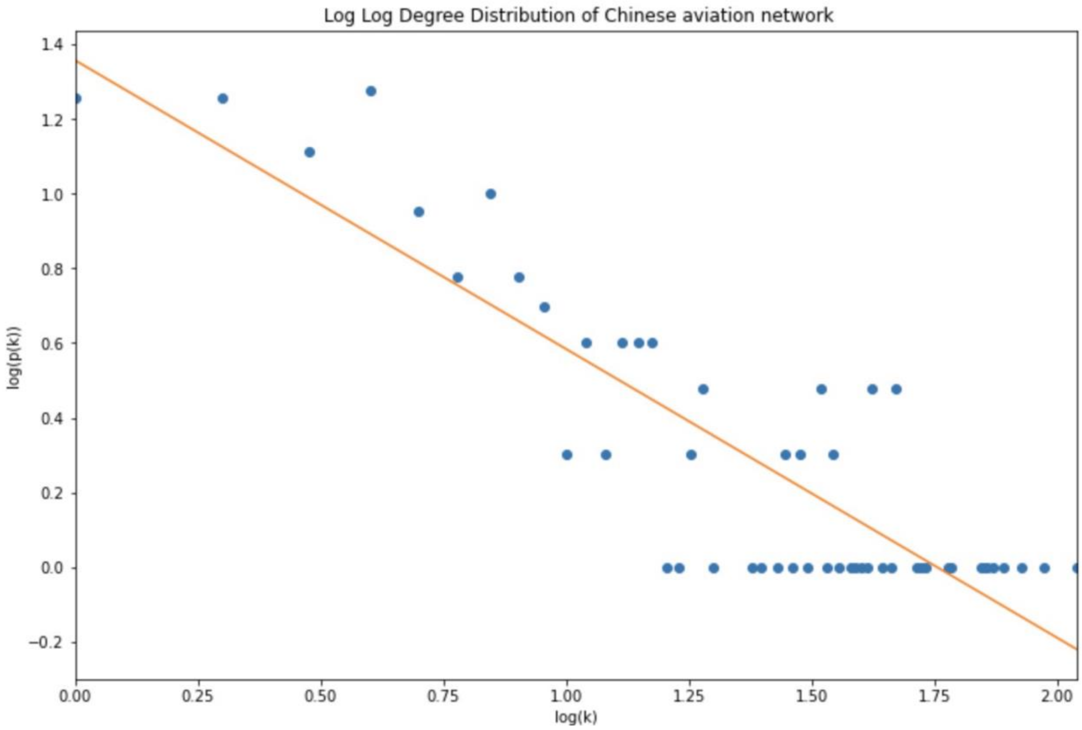

Figure 2 . Degree distribution with logarithmic coordinates.

Figure 1 and Figure 2 show the visualisation of the degree distribution of the aviation network of China. The original coordinates and double logarithmic coordinates are used to gain the degree distribution and information to tell whether the network is a power-law network. Figure 2 indicates the distribution is power-law with "long tail" attribute.

7.2. Density

With the network density, it is understandable how closely the network is connected. Because the graph is an undirected graph, the equation is the total number of edges divided by the total possible edges.

$$D = \frac{m}{\frac{n(n-1)}{2}} \tag{2}$$

From the equation shown above, the total number of nodes is represented by n, whereas the number of edges is represented by m.

By using pandas of python, the density of the Chinese aviation network is 0.09.

7.3. Average-Shortest-Path length

The average path length describes the average length of the shortest pathways from one node to another node in the network, which can reflect the efficiency of an aviation network. The equation is shown below.

$$d = \frac{1}{n}\sum_{i \neq j} d_{ij} \tag{3}$$

With pandas of python, it is found that the average-shortest -path length is around 2.19 for China's aviation network.

DOI: 10.54254/2755-2721/5/20230674

7.4. Clustering coefficient

The clustering coefficient shows the propensity for nodes to group collectively., from which the local cohesion of the network could be reflected. The local clustering coefficient and average clustering coefficient will be examined in the study.

The local clustering coefficient represents how close the neighbours nodes tend to be a complete graph. The equation is to divide the number of edges between ki nodes by the number of ei edges which ki nodes are fully connected. The local clustering coefficient is shown as :

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \tag{4}$$

Table 1 displays the findings for the top 10 local clustering coefficients and the associated airports.

Table 1. Top 10 local clustering coefficients and their related airports.

Code	Airports	Clustering Coefficient
AEB	Baise Youjiang Airport	1
AKU	Aksu Airport	1
AOG	Anshan Air Base	1
AVA	Anshun Huangguoshu Airport	1
BPX	Qamdo Bangda Airport	1
HJJ	Zhijiang Airport	1
HZH	Liping Airport	1
JIQ	Qianjiang Wulingshan Airport	1
JIU	Jiujiang Lushan Airport	1
JUH	Jiuhuashan Airport	1

The table above shows the local clustering coefficient of those airports is 1, and their related airports are all small.

The average clustering coefficient shows the aggregation of the whole network, which is shown:

$$\mathrm{C} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{C_i} \tag{5}$$

DOI: 10.54254/2755-2721/5/20230674

Using the aforementioned equation, the average clustering coefficient for China's aviation network is 0.66.

8. Centrality analysis

8.1. Degree centrality

The degree centrality is a useful metric that expresses one node's degree as its degree centrality, demonstrating the significance of the degree of connectivity from one node to every other node.

$$\mathrm{Degree_i} = \sum_{j=i}^{N} a_{ji} \tag{6}$$

8.2. Closeness centrality

The closeness centrality can be computed as the total length of the shortest pathways connecting one node to every other node. In Figure 3, closeness centrality shows the proximity degree connecting one node to every other node. The equation of the closeness centrality is provided below, where d represents the length from nodes i to j.

$$C(i) = \frac{1}{\sum_{j \neq i} d_{ij}} \tag{7}$$

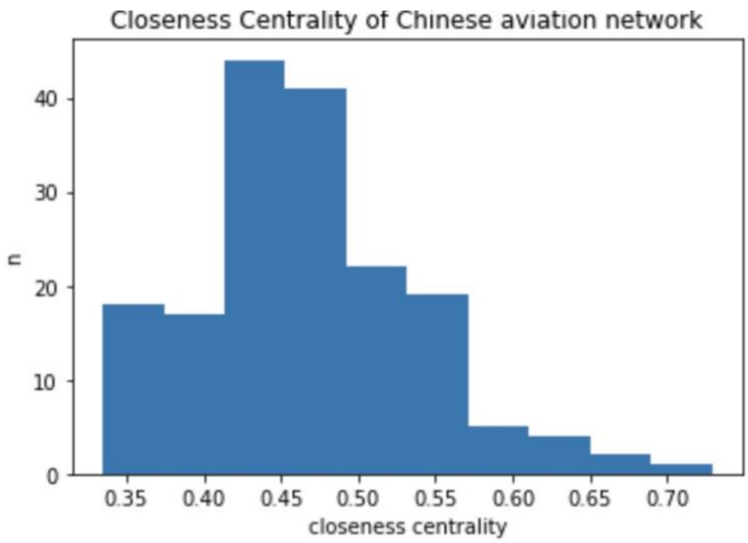

Figure 3. Results of closeness centrality.

8.3. Betweenness centrality

Betweenness centrality is a control index, showing the measure of a single node city which becomes the intermediary of passenger flow linkage between other node cities in Figure 4. The equation of betweenness centrality is shown below.

$$C_B(x) = \sum_{i \neq j \neq x} \frac{n_{ij}(x)}{n_{ij}} \tag{8}$$

DOI: 10.54254/2755-2721/5/20230674

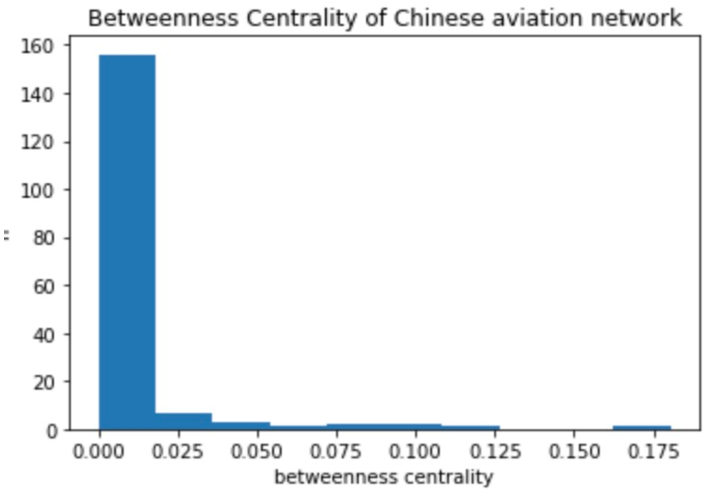

Figure 4. Results of betweenness centrality.

Table 2. The data of centrality analysis (China).

Airports	Degree centrality	Airports	Betweenness centrality	Airports	Closeness centrality
PEK	1.28	PEK	0.18	PEK	0.73
CAN	1.09	CTU	0.12	CAN	0.68
CTU	0.99	URC	0.09	CTU	0.66
XIY	0.91	CAN	0.09	XIY	0.64
CKG	0.86	XIY	0.08	CKG	0.63
SZX	0.84	KMG	0.07	SZX	0.63
KMG	0.83	CKG	0.05	PVG	0.62
PVG	0.81	PVG	0.05	KMG	0.60
SHA	0.72	HRB	0.05	CSX	0.60
CSX	0.70	SZX	0.04	SHA	0.59

Table 2 shows the centrality analysis for China. To understand the data better, our team also calculates the data of America for comparison as presented in Table 3, which will be mentioned in discuss session.

DOI: 10.54254/2755-2721/5/20230674

Table 3. The data of centrality analysis (America).

Airports	Degree centrality	Airports	Betweenness centrality	Airports	Closeness centrality
ATL	0.56	ANC	0.30	DEN	0.49
ORD	0.54	DEN	0.15	ORD	0.49
DEN	0.54	ORD	0.12	LAS	0.48
DFW	0.5	SEA	0.12	MSP	0.48
MSP	0.42	ATL	0.09	SEA	0.47
DTW	0.41	MSP	0.08	LAX	0.46
LAS	0.41	DFW	0.08	PHX	0.46
CLT	0.40	BET	0.07	SLC	0.46
IAH	0.37	FAI	0.07	ATL	0.45
PHL	0.32	LAS	0.06	DFW	0.45

9. Discussion

9.1. Structural Analysis

From the results which we showed in Results, the aviation network in China is a power-law network, which is relatively resilient to random attacks than targeted attacks.

The density of the Chinese aviation network(0.09) is relatively high, compared with another large aviation network, America, whose density is 0.02 using the dataset and equation mentioned before. It means China’s aviation network is connected closely. The average clustering coefficient of the Chinese aviation network(0.66) is higher than America’s(0.47), so the Chinese aviation network will be affected less while suffering from attacks in the network.

9.2. Centrality Analysis

From the results which we showed in Results, Compared with the American aviation network, the degree centrality and the closeness centrality of the Chinese aviation network are relatively high, and the betweenness centrality is relatively low. Therefore, we could conclude that the importance and proximity degree of one node in the Chinese aviation network is high.

9.3. Important Nodes Analysis

Figure 5 is the graph of the Chinese aviation network, supported by OpenFlights’ database and generated by Gephi. From this figure, Beijing Capital Airport(PEK) is the most important node in China. The data which are got in Results also prove the analysis. The study planned to remove PEK in order to show the degree to which extend the aviation network is affected.

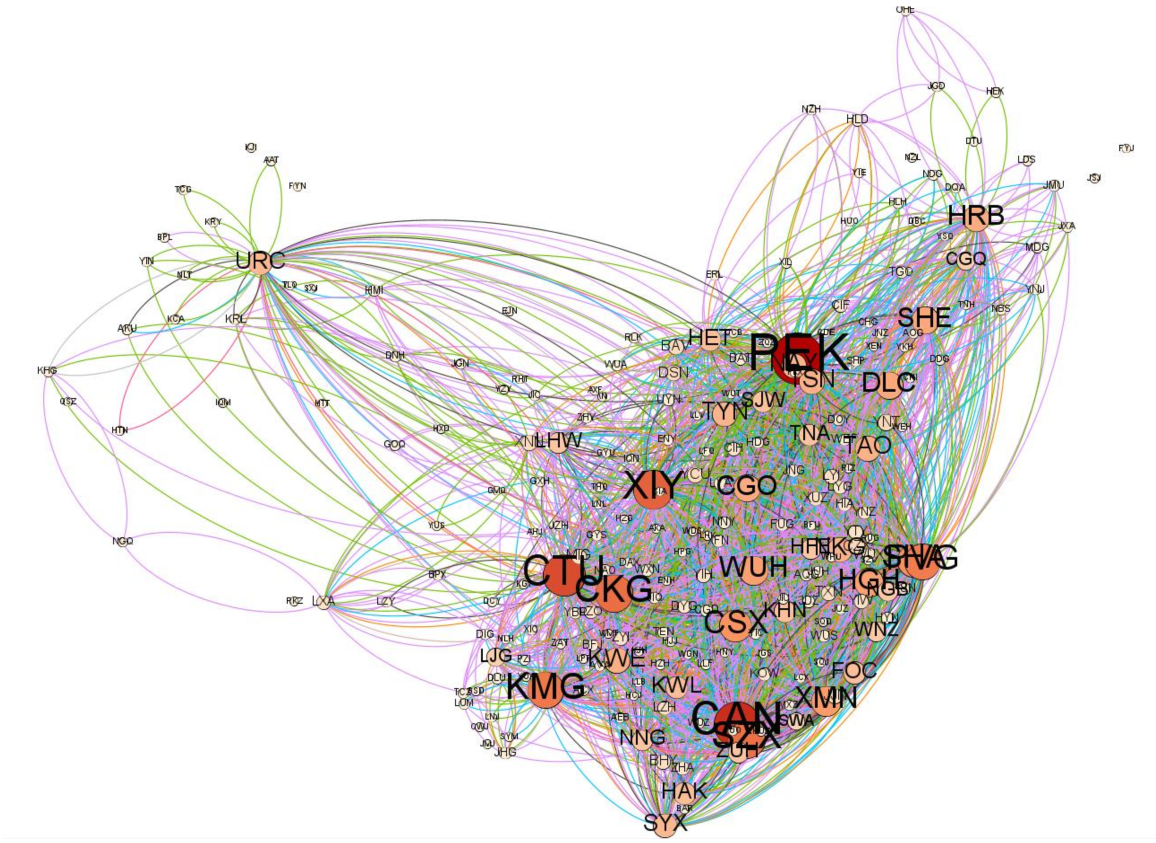

Figure 5. Graph of airports and airlines in China.

When PEK is removed, the number of large communities in the network decreases from 6 to 5, but the number of the largest communities increases from 39 to 55. From the data, we conclude that when PEK is removed, some communities associated with PEK get merged and become a larger community. Thus, PEK can be seen as a significant intermediary between communities in the Chinese aviation network, which confirms the results from results in Figure 6 and Figure 7.

DOI: 10.54254/2755-2721/5/20230674

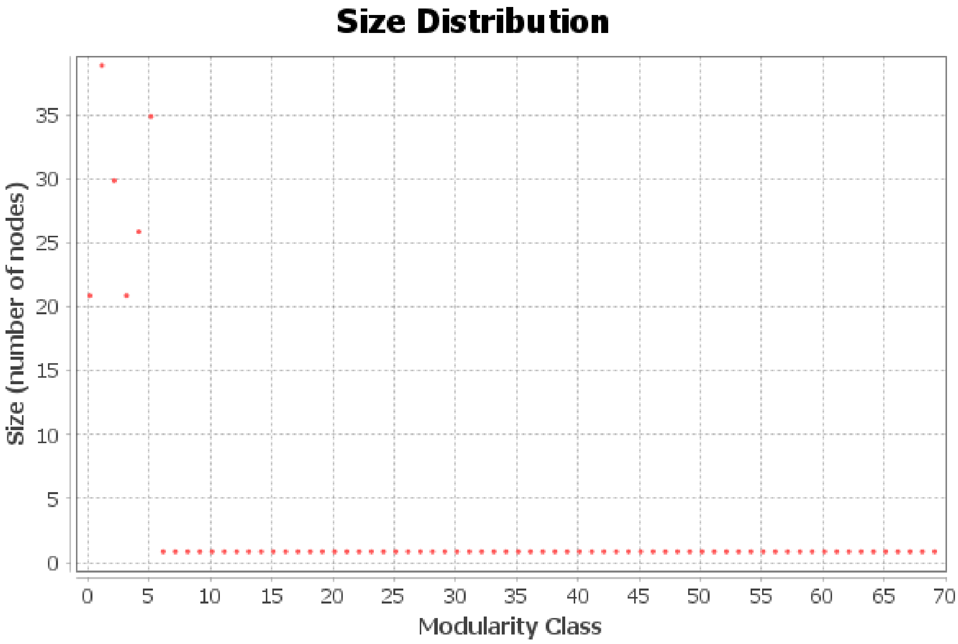

Figure 6. Modularity report of the chinese aviation network.

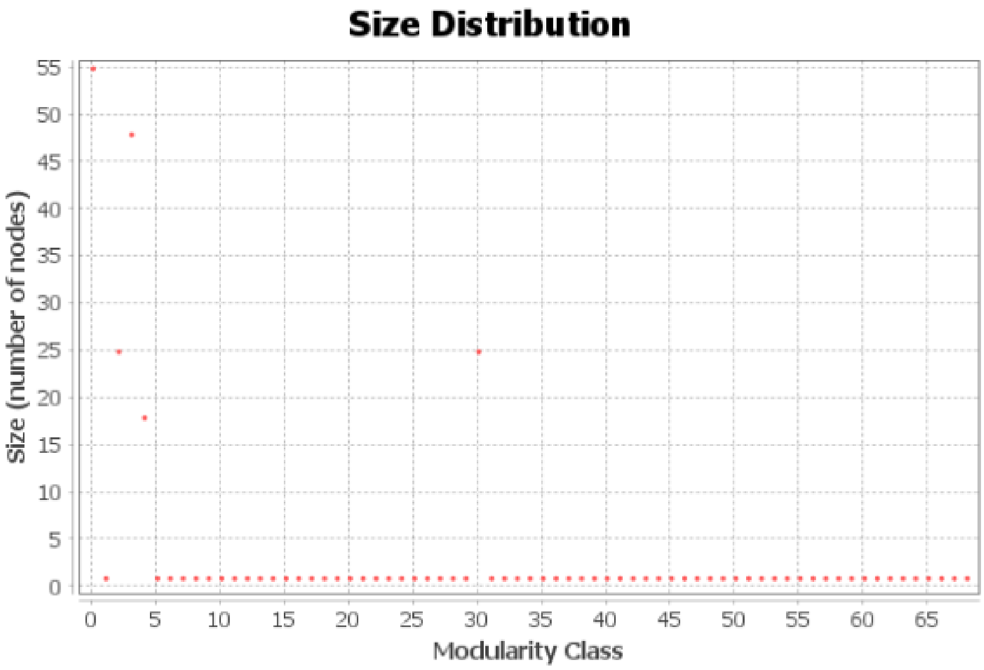

Figure 7. Modularity report of the chinese aviation network(PEK Removed).

The network in China shows itself to be highly repairable, because there will not be more single nodes outside the network, which generate new communities with other nodes, even without the biggest airport in China. However, in general, the average-shortest-path length increases. Without PEK, there will be an increase of about 3.5%, negatively affecting the convenience of the aviation network.

10. Conclusion

The distribution of the aviation network in China is power-law with a “long tail” attribute, which is more resilient to random attacks than the targeted attacks, if there are targeted attacks on a small amount of higher degree nodes, the network will collapse immediately. In the epidemic of COVID-19, if the big airports in China’s aviation network are closed, the whole network will suffer a much bigger loss than the situation that a random epidemic occurs in other small cities.

China’s aviation network is more connected and dense since the network’s density is 0.09, which is relatively high in our comparison. There are a number of airports that share the same local clustering coefficient(1), which means the big airports are fully connected to the other domestic airports. The average clustering coefficient is 0.66, which means the Chinese aviation network is relatively stable. If there are random attacks on the network, China’s aviation network is expected to be affected relatively low.

Every airport in China has its way of connecting the network even though the biggest airport, PEK, is removed. Thus, the Chinese aviation network is relatively resilient and robust, due to its highly repairable attribute shown in our analysis. However, for the Chinese network, the increase in the average-shortest-path length is relatively high, which may be an important consideration in real-life transportation. Under the influence of COVID-19 and some blocking policies on the airports, people who want to travel in China may spend more time on airplane travel.

References

[1] D. Bing, “Reliability Analysis for Aviation Airline Network based on Complex Network,” Journal of Aerospace Technology and Management, vol. 6, no. 2, pp. 193–201, 2014.
[2] I. Šebeščáková, L. Melníková, and Ľ. Socha, “The assessment of the contribution of aviation to the national economy,” eXclusive e-JOURNAL, 2013. [Online]. Available: https://exclusiveejournal.sk/files/0/sebescakova-melnikova-socha.pdf. [Accessed: 02-Oct-2022].
[3] IATA, Ed., “Covid-19 puts over half of 2020 passenger revenues at risk,” IATA, 14-Apr-2020. [Online]. Available: https://www.iata.org/en/pressroom/2020-press-releases/2020-04-14-01/. [Accessed: 11-Aug-2022].
[4] X. Ren, “Pandemic and lockdown: A territorial approach to covid-19 in China, Italy and the United States,” Eurasian Geography and Economics, vol. 61, no. 4-5, pp. 423–434, 2020.
[5] D. P. Cheung and M. H. Gunes, “A complex network analysis of the United States Air Transportation,” 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012.
[6] W. Sun, G. Tao, and P. Yang, “Chinese aviation network analysis under actual node attack,” 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), 2021.
[7] CAAC, “Notice on Further Reducing International Passenger Flights during the Epidemic Prevention and Control Period,” CAAC, 26-Mar-2020. [Online]. Available: http://www.caac.gov.cn/en/XWZX/202003/t20200326_201748.html. [Accessed: 11-Aug-2022].
[8] Z. J. Cheng, Z. Zhan, M. Xue, P. Zheng, J. Lyu, J. Ma, X. D. Zhang, W. Luo, H. Huang, Y. Zhang, H. Wang, N. Zhong, and B. Sun, “Public health measures and the control of covid-19 in China,” Clinical Reviews in Allergy & Immunology, 2021.
[9] CAAC, “Civil Aviation Administration adjusts international passenger flight circuit breaker measures,” CAAC, 16-Dec-2020. [Online]. Available: http://www.caac.gov.cn/XXGK/XXGK/TZTG/202012/t20201216_205607.html. [Accessed: 11-Aug-2022].
[10] M. Hou, K. Wang, and H. Yang, “Hub Airport slot re-allocation and subsidy policy to speed up air traffic recovery amid covid-19 pandemic --- case on the Chinese airline market,” Journal of Air Transport Management, vol. 93, p. 102047, 2021.