

Vulnerability of Regional Aviation Networks Based on DBSCAN and Complex Networks

Hang He^{1,*}, Wanggen Liu¹, Zhenhan Zhao¹, Shan He¹ and Jinghui Zhang²

¹Civil Aviation Flight University of China, Guanghan, 618307, China

²Webchain Pty Ltd., Sydney, NSW, 2118, Australia

*Corresponding Author: Hang He. Email: 43162164@qq.com

Received: 12 January 2022; Accepted: 08 March 2022

Abstract: To enhance the accuracy of performance analysis of regional airline network, this study applies complex network theory and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to investigate the topology of regional airline network, constructs node importance index system, and clusters 161 airport nodes of regional airline network. Besides, entropy power method and approximating ideal solution method (TOPSIS) is applied to comprehensively evaluate the importance of airport nodes and complete the classification of nodes and identification of key points; adopt network efficiency, maximum connectivity subgraph and network connectivity as vulnerability measurement indexes, and observe the changes of vulnerability indexes of key nodes under deliberate attacks and 137 nodes under random attacks. The results demonstrate that the decreasing trend of the maximum connectivity subgraph indicator is slower and the decreasing trend of the network efficiency and connectivity indicators is faster when the critical nodes of the regional airline network are deliberately attacked. Besides, the decreasing trend of the network efficiency indicator is faster and the decreasing trend of the maximum connectivity subgraph indicator is slower when the nodes of four different categories are randomly attacked. Finally, it is proposed to identify and focus on protecting critical nodes in order to better improve the security level of regional airline system.

Keywords: DBSCAN algorithm; regional airlines; key node; vulnerability

1 Introduction

Regional airlines are an important part of the national route network [1] and exert a vital role in the construction of a strong civil aviation country, but regional airlines are also a topic that has long plagued the development of China's air transportation industry. With the in-depth study of complex network characteristics, the study of node importance has become an important topic in complex networks. This study analyzes the topology and network performance of regional aviation, identifies the key nodes of regional aviation network effectively, and evaluates the node importance to improve the safety of regional aviation network.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Scholars both at home and abroad have proposed various evaluation methods, such as single-indicator evaluation. P. Bonacich proposed the degree centrality based on node degree value [2]. He proposed the mesoscopic centrality by considering from the propagation path perspective [3] and the proximity centrality considering the distance between nodes in the network [4]. There are also multi-indicator fusion evaluations. YANG et al. [5] considered that the importance of nodes depends on the degree and the mesonumber of two indicators, and Yu Tian applies the aggregation coefficient and network importance to evaluate the importance of nodes [6]. With the idea of multi-attribute information fusion, Hongming Mo proposed a node multi-attribute identification method for complex networks [7]. Xiping Zhang introduced the concept of m-order neighbor nodes and proposed a node importance method for complex networks based on the importance contribution of m-order neighbor nodes [8]. Feng Wang put forward the concept of m-order neighbor node importance contribution coefficient matrix to establish the evaluation model after considering the node's own attributes, the node's position in the network as well as the degree importance contribution and meso-importance contribution of m-order neighbor nodes [9]. Besides, an estimation mechanism for assessing critical nodes in IoT systems has been proposed. The estimation of critical nodes consists of two components, respectively, the utilization relationship between nodes and the impact on the system if the node is attacked, quantifying these two components in terms of node importance value and node risk value [10]. In summary, the existing literature is based on degree, proximity, mesonumber, and aggregation coefficients to evaluate the importance of network nodes or to propose a new method to optimize and innovate these indicators [11].

Clustering is an important component of unsupervised learning in the field of machine learning, and clustering algorithms have important applications in industry and have also received extensive attention and research in academia [12]. Sun et al. [13] first classified the heartbeats into five common Arrhythmia classes. The heartbeats classified as S and V by Rist are further classified by Morst to enhance the classification accuracy. Hassan et al. [14] proposed a simple oversampling method based on Multivariate Gaussian distribution and K-means clustering, called GK-Means. Clustering proves an effective data mining method to discover data distribution and hidden features from featureless online learning data [15]. The common clustering analysis methods include hierarchical methods, partitioning methods, grid-based methods, density-based methods, and model-based methods. This study adopts DBSCAN clustering algorithm. DBSCAN algorithm is a classical density-based clustering algorithm, which was first published in the journal science. DBSCAN algorithm does not need to pre-specify the number of clusters for clustering, and can discover any number and shape of clusters in noise-containing datasets [16], which has the advantages of strong interpretability, good visualization and easy understanding.

Recently, several scholars have introduced complex network topologies into the field of airport networks, and Yaru Dang constructs a Chinese air passenger flow network and adopts a deliberate attack with three centrality degrees to explore the network's resistance to destruction [17]. Some scholars have also investigated the evolution of aviation airport networks. Xinwei Yang analyzed the basic nature of Chinese airline networks and their evolution in recent years, and studied the destructiveness of the network structure through the maximum connectivity subgraph index [18]. Lingling Yan used five methods such as degree centrality and near centrality to rank the importance of nodes in Chinese airline networks, and used vulnerability indicators to confirm the effectiveness of the ranking methods [19]. Delong Wang combined the complex network topological characteristics with the traffic and economic characteristics of the airport itself in order to comprehensively evaluate the importance of the airport network [11]. Based on the complex network, the degree of influence of different influencing factors on flight punctuality, whether it is positive or negative, is studied. Moreover, the study of these factors will provide some references for the practical application of improving flight punctuality [20]. With the purpose of improving the objectivity of urban rail network vulnerability assessment, Chaoqun Ma et al. [21] integrated passenger demand characteristics into the calculation of network vulnerability, and

analyzed the characteristics of urban rail network and its vulnerability from the perspective of network passenger flow with Xi'an urban rail network as an example. Based on the disturbance events influencing the vulnerability of the subway system, a directed complex network of disturbance events is constructed, and it is concluded that only a few key nodes can directly affect the vulnerability of the subway [22]. The vulnerability of the global network is higher than that of the local network when the central node of the urban express network is deliberately attacked. The vulnerability of the global network and the local network are synchronized when the transit node is deliberately attacked. The impact of the sequential failure of individual nodes on the network is small, while the impact of the cascade failure of nodes on the network is larger, and the failure of nodes with high importance in successive deliberate attacks can cause the network to collapse rapidly [23]. He linked community detection in the field of complex networks with clustering algorithms in machine learning, explored new methods for community structure detection, and provided certain reference ideas for cross research of related disciplines [24].

The current aviation network performance studies are based on complex network topologies, and the selection of key nodes is relatively simple, lacking a perfect index system. Besides, the analysis of network performance uses a single index. In fact, the importance of airport network nodes is also related to the attributes of the airport and its region, and the comprehensive consideration of multi-layer indicators makes it reasonable to assess the importance of nodes. In addition, several indicators that can measure the network performance are selected to explore the degradation trend of the performance of regional airline network and propose optimization suggestions.

2 Research Theory and Data

2.1 Construction of Regional Airline Network Model

In this paper, airports with less than 2 million passengers per year are defined as feeder airport. Besides, the routes connecting feeder airports at one end or both ends are positioned as feeder routes. All feeder routes form a feeder airline network. Using Gephi visualization software, the undirected regional airline network diagram is drawn, as shown in Fig. 1 below.

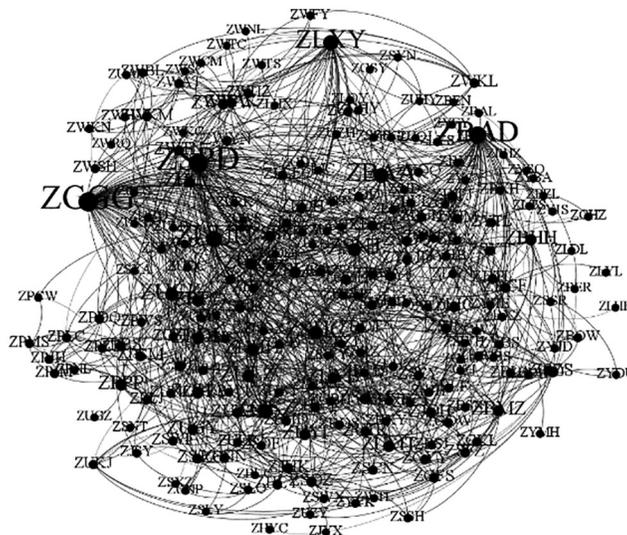


Figure 1: Regional airline network map (2020 data)

There are 161 nodes and 2290 edges in the figure. The nodes represent the regional airports and the edges represent the direct connection between two airport nodes. The size of a node in the graph depends on its degree value, indicating the higher the degree value, the larger the node.

2.2 Airport Node Importance Assessment Index

The identification of airport key nodes is important for the vulnerability analysis of the regional airline network. To establish a perfect index system of airport node importance, the selected indexes should be comprehensive and representative. They should not only reflect the nature of complex network topology, but also combine with the attributes of the airport and the region where the airport is located. Therefore, the selected indicators are shown in Tab. 1 below.

The greater the degree of a node in an airport network, the more routes the airport has, and the higher its importance. The number of nodes reflects the role and influence of the nodes in the network, which is a vital global geometric quantity.

Airport attributes are reflected by passenger throughput and landings. Passenger throughput is one of the most intuitive signs of an airport's capacity and busyness, and airports with higher throughput have more flights and are more important. City attributes include prosperity and population. A city's population reflects a city's consumption demand and spending power, and the higher the population of the city where the airport is located, the higher the demand for air travel, the higher the level of city GDP, the higher the spending power of residents, the higher the possibility of choosing air travel, and the higher the importance of the airport. Therefore, the selected indicators are presented in Tab. 1 below.

where the V1 node degree is calculated as $k_i = \sum_j a_{ij}$, i, j indicates a node, a_{ij} denotes the adjacency matrix variable, k_i Expression degree.

The formula for calculating the number of points is $B_i = \sum_{k \neq j \in G} \frac{D_{ij}(i)}{D_{kj}}$, G indicates network, $D_{ij}(i)$ is the number of shortest paths through node i and D_{kj} is the number of shortest paths between nodes k, j .

Table 1: Node importance index system

Level 1 indicators	Secondary indicators	Indicator description
Topology U1	V1 degree	Number of edges connected to the node
	V2 Median	The ratio of the number of paths passing through the node to the total number of shortest paths in the network among all shortest paths
Airport factors U2	V3 Throughput	The number of passengers entering and leaving the range of airplanes such as airlines
	V4 Landing and takeoff times	Number of flight landings and takeoffs per unit time
Urban Factors U3	V5 Prosperity	GDP level of the region to which the airport belongs
	V6 Population	Resident population in the area where the airport is located

2.3 Vulnerability Metrics

The commonly used vulnerability metrics include network efficiency, network connectivity, and maximum connectivity subgraph size. The network efficiency metric measures the effectiveness of the

network, the connectivity metric calculates the overall working performance of the network, and the maximum connectivity subgraph metric computes the connectivity of the network.

Network efficiency is expressed as:

$$E = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{1}$$

where i and j denote nodes, G denotes the network, N denotes the number of airport nodes, and d_{ij} indicates the shortest distance between nodes i and j .

The network connectivity is expressed as:

$$\gamma = \frac{e_G}{e_{G_{max}}} \tag{2}$$

The formula e_G denotes the number of edges of the network G , and $e_{G_{max}}$ signifies the maximum number of edges for the network.

The maximum connected subgraph dimensions are expressed as:

$$\bar{J} = \frac{N'}{N} \tag{3}$$

where N and N' refer to the maximum connected subgraph size before and after node damage, respectively, that is, the maximum number of connected subgraph nodes formed by the network.

2.4 Entropy Weight Topsis Method

The entropy topsis method is an assignment method based on the entropy method to improve the TOPSIS method model, which can determine the weights according to the amount of information reflected by the degree of variation of each index value, and can calculate the closeness to the optimal solution by the value of the difference between the evaluation object and the positive and negative ideal solutions.

Step 1: Standardization of indicators. x'_{ij} is the normalized value and x_{ij} is the original value of the i th indicator in the j th year.

Positive indicators:

$$x'_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \tag{4}$$

where, i denotes the year, j denotes the indicator, and i, j represent non-zero natural numbers.

Step 2: The normalization of the indicator is processed to calculate the weight of the i th indicator in the j th year.

$$x_{ij} = \frac{x'_{ij}}{\sum_{i=1}^n x'_{ij}} \tag{5}$$

Step 3: Calculate the information entropy of the indicators e_j :

$$e_j = -\frac{1}{\ln n} \sum_{j=1}^n (x_{ij} \times \ln x_{ij}) \quad 0 \leq e_j \leq 1 \tag{6}$$

Step 4: Calculate the coefficient of variation of each indicator and the weight of the indicator.

$$g_j = 1 - e_j \quad (7)$$

$$w_j = \frac{g_j}{\sum_{j=1}^m g_j} \quad (8)$$

Step 5: Construct the weighted normalized decision matrix V_{ij} , $V_{ij} = W_j Z_{ij}$, $Z = (Z_{ij})_{m \times n}$.

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \quad j = (1, 2, \dots, n) \quad (9)$$

Step 6: Determine the ideal solution and negative ideal solution. It indicates the larger the value of the element V in the decision matrix V_{ij} , the better the solution.

$$\text{Ideal solution: } V^+ = (V_1^+, V_2^+, \dots, V_m^+) = \{\max V_{ij} | j = 1, 2, \dots, m\} \quad (10)$$

$$\text{Negative ideal solution: } V^- = (V_1^-, V_2^-, \dots, V_m^-) = \{\max V_{ij} | j = 1, 2, \dots, m\} \quad (11)$$

Step 7: Calculate the distance of each node to the ideal solution S_i^+ and the distance to the negative ideal solution S_i^- .

$$S_i^+ = \sqrt{\sum_{j=1}^m (V_j^+ - V_{ij})^2} \quad (12)$$

$$S_i^- = \sqrt{\sum_{j=1}^m (V_j^- - V_{ij})^2} \quad (13)$$

Step 8: The relative proximity of each ne is calculated and ranked, indicating the larger the value of C_i , the more important the node is.

$$C_i = \frac{S_i}{S_i^+ + S_i^-} \quad (14)$$

2.5 DBSCAN Algorithm

DBSCAN is a representative of a typical clustering method that divides classes based on data density and can be effective in automatically discovering the number of target clusters without setting initial values and effectively discovering clusters of different shapes. Given a sample set of $D = (x_1, x_2, \dots, x_m)$, the parameters $(\epsilon, \text{MinPts})$ are adopted for describing the closeness of the sample distribution in the neighborhood. In the formula, ϵ describes the neighborhood distance threshold of a sample, and MinPts depicts the threshold of the number of samples in the neighborhood of a sample with distance ϵ . The clustering process of the DBSCAN algorithm can be expressed as follows.

- (1) An arbitrary selection of a data object point p from the data set.
- (2) If the selected data object point p is a core point for the parameters Eps and MinPts , find all data object points that are reachable from the density of p to form a cluster.

- (3) If the selected data object point p is an edge point, select another data object point.
- (4) Repeat steps (2) and (3) until all points are processed.

2.6 Indicator Data Sources

In this study, U1 indicators have been calculated and the data are obtained from the Civil Aviation Administration of China (CAAC) Civil Aviation Advance Flight Plan Management System. Besides, the domestic flight schedule for China's summer and autumn seasons in 2020 is selected. Specifically, a total of 2148 flight segments data of 161 airport nodes are included. U2 indicator data are obtained from Civil Aviation from Statistics, and U3 indicator is obtained from the official website of National Bureau of Statistics. In summary, the partial data of the importance index of regional airline network nodes are summarized in [Tab. 2](#) below.

Table 2: Importance index of regional airline nodes

Airport	degree	Median	Throughput/million passengers	Landing and takeoff times	Prosperity/ billion yuan	Population/ 10,000 people
ZSLG	21	3.356	192	11826	3277	460
ZHXF	11	1.355	190	73847	4601	526
ZBCF	14	1.729	189	12733	1763	403
ZULZ	31	3.094	186	23629	2157	425
ZPDL	28	3.758	177	13734	1484	334
ZWAK	16	1.822	171	14831	1315	271
ZYYJ	17	1.226	166	5740	727	55
ZUMT	22	2.829	165	17845	1092	66
ZWTN	15	1.334	160	11063	406	253
ZGZH	16	1.108	157	10991	3176	416
ZHLY	13	0.942	154	180286	5128	706
ZHSY	10	0.434	152	11369	1915	321
ZSZS	17	1.911	152	20250	1512	116
ZSJG	13	0.480	149	4432	4494	836
ZWYN	12	0.689	148	8056	1266	56
ZHES	11	0.552	143	8925	1117	346
ZUYI	20	1.646	138	12536	493	100
ZSLQ	13	0.888	138	9498	5262	662
ZPTC	22	1.555	137	8420	716	64

3 Node Importance Assessment

3.1 Node Clustering Based on DBSCAN Algorithm

Based on the implementation process of DBSCAN clustering, the key parameters Eps and Minpts are set. A point p is arbitrarily selected from the sample set of clustered data X . If the condition of the point

satisfies the determination of the core object, then all data points from that point density can become a cluster, while the data points that do not belong to any cluster are marked as noise points [25].

Fig. 2 provides the aggregation effect when different Eps and MinPts are combined respectively. By comparison, it can be observed that Minpts = 10 has only one class, Minpts = 8 clustering has poor accuracy of only 2 classes, Minpts = 7 has more clustering noise points, and Minpts = 3 clustering result has higher accuracy. By observing the clustering results with different Minpts values, the number of noise points represents the accuracy. Therefore, the key parameters selected in this paper in the DBSCAN algorithm are: Eps = 0.15, Min Pts = 3.

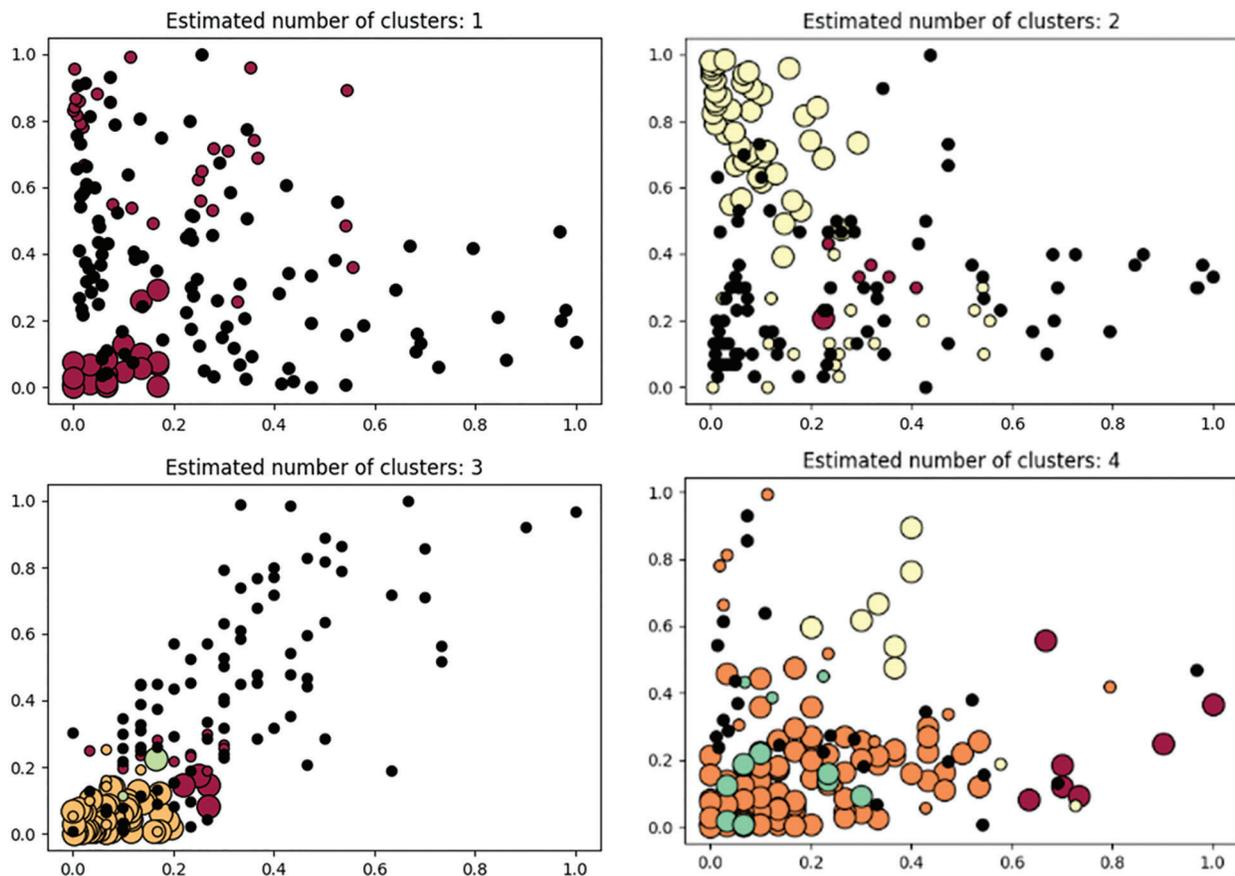


Figure 2: DBSCAN cluster classification results

The clustering results are shown in Fig. 3 below, and the algorithm divides the nodes into 4 clusters. There are 8 airports in the first cluster, respectively, ZHLY, ZPDL, ZULZ, ZGCD, ZUMT, ZSZS, ZBCF and ZSFY. They are basically consistent with the rankings calculated by the entropy-weighted TOPSIS method, demonstrating the accuracy of the algorithm.

Category number	Sample size	Airport Name									
0	8	ZHLY	ZPDL	ZULZ	ZGCD	ZUMT	ZSZS	ZBCF	ZSFY		
1	108	ZHXF	ZSLG	ZGWZ	ZHNY	ZBHD	ZBMZ	ZLAK	ZWAK	ZWHZ	ZLYA
		ZUTR	ZPBS	ZGCI	ZSJU	ZHES	ZUWX	ZYQQ	ZUDX	ZBLL	zbcd
		ZYJM	ZUZH	ZGBS	ZUPS	ZBXZ	ZBUH	ZLQY	ZPLC	ZWKC	ZLLN
		ZLGM	ZYDD	ZBEN	Zbuc	ZYHE	ZWSC	ZYBA	ZYJX	ZWHM	ZULB
		ZGSY	ZWTC	ZUDC	ZLJC	ZWTL	ZYLD	ZUGZ	ZLGL	ZLHB	ZUWS
		ZUKD	ZYDU	ZHSN	ZLDL	ZYMH	ZYFY	ZJYX	ZBAR	ZBSN	ZBDT
		ZSWF	ZSJG	ZSDY	ZYSQ	ZGHY	ZYYJ	ZLDH	ZYDQ	ZSSM	ZUKJ
		ZYMD	ZUNZ	ZSAQ	ZPZT	ZWfy	ZSGS	ZBXH	ZWAT	ZLJQ	ZPJM
		ZYYK	ZYBS	ZLTS	ZSJH	ZGMX	ZSJD	ZBYZ	ZSLO	ZYJD	ZYTN
		ZUBD	ZLYS	ZLZY	ZWNL	ZUQJ	ZWCM	ZUJZ	ZURK	ZWRQ	ZBHZ
		ZPNL	ZWKN	ZLHX	ZUHY	ZUNP	ZBER	ZUAL	ZWTS		
2	9	ZUYB	ZUXC	ZUGU	ZJQH	ZGFS	ZBZL				
		ZWTN	ZUNC	ZUBJ							
3	12	ZBLF	ZGLG	ZLHZ	ZBAL	ZBZJ	ZBCZ				
		ZPWS	ZYAS	ZBdH	ZUBZ	ZGHC	ZSTX				

Figure 3: Final clustering results

3.2 Entropy Weight TOPSIS Method to Determine Node Importance Ranking

The entropy weight TOPSIS comprehensive evaluation method was applied to calculate the above 161 nodes, and the entropy weights of the six secondary indicators were 0.103667, 0.238827, 0.113405, 0.255894, 0.131915 and 0.156294. The specific ranking is shown in Tab. 3 below.

Table 3: Key node ranking results

Airport	Si ⁺	Si ⁻	Ci	Sort	Airport	Si ⁺	Si ⁻	Ci	Sort
ZHLY	0.0665	0.1535	0.6978	1	ZBHD	0.1331	0.0587	0.306	13
ZSRZ	0.0985	0.0972	0.4968	2	ZUXC	0.1472	0.0640	0.3031	14
ZHXF	0.1043	0.0793	0.4318	3	ZBMZ	0.1544	0.0665	0.3012	15
ZPDL	0.1387	0.0924	0.3997	4	ZLAK	0.1338	0.0557	0.294	16
ZULZ	0.1304	0.0826	0.3879	5	ZSZS	0.1423	0.0523	0.2688	17
ZSLG	0.1380	0.0859	0.3836	6	ZUGU	0.1383	0.0485	0.2597	18
ZGWZ	0.1217	0.0710	0.3684	7	ZBCF	0.1467	0.0511	0.2585	19
ZGCD	0.1234	0.0639	0.3412	8	ZWAK	0.1458	0.0508	0.2583	20
ZUMT	0.1402	0.0709	0.3359	9	ZSWF	0.1604	0.0543	0.2529	21
ZHNY	0.1335	0.0616	0.3157	10	ZSLQ	0.1528	0.0513	0.2512	22
ZUYB	0.1429	0.0649	0.3122	11	ZJQH	0.1507	0.0479	0.2411	23
ZYCY	0.1315	0.0586	0.3081	12	ZSJG	0.1603	0.0493	0.2354	24

Throughput is ZSLG, ZHXF, ZBCF, ZULZ, and ZPDL. Landing and takeoff times are ZHLY, ZSRZ, ZGWZ, ZHXF, and ZYCY. Prosperity is ZSWF, ZSLQ, ZHLY, ZHXF, and ZSJG, indicating the high economic level of the regions where they are located. Besides, the population is ZHNY, ZGFS, ZBHD, ZSWF, and ZSJG. Based on the data of each index and the comprehensive evaluation results, ZHLY, ZULZ, ZPDL and ZUMT rank relatively high, indicating that they are more important in all aspects of

the regional airline network, and the key nodes with higher importance generally have more routes opened, more influence and higher annual throughput.

4 Vulnerability Analysis of Feeder Airlines

Vulnerability was first started in disaster research and gradually introduced into transportation and social fields in the 1970s. Till the present, there has been no unified definition. Vulnerability of the transportation system refers to the vulnerability of the transportation system to a certain event, consequently leading to a significant reduction in service capacity. Combined with the characteristics of regional aviation, its vulnerability can be defined as the extent to which network connectivity is affected when some airports and routes cannot operate normally due to natural disasters or terrorist attacks, etc [26]. The vulnerability of a regional airline network can be measured by the maximum connectivity subgraph size, network efficiency, and connectivity.

4.1 Deliberate Attack

According to the entropy weight TOPSIS method ranking of nodes' importance in the previous section, the degree to which nodes fail under deliberate attacks and affect the change in vulnerability of the feeder airline network can be observed as follows.

Step1 Attack and remove the nodes in order from the highest to the lowest according to the node importance ranking.

Step2 Calculate the vulnerability index after removing the nodes

Step3 Plot and explore and draw conclusions.

Using Sublime Text software, write Python programs for node attacks and follow the steps to damage the nodes in the airline network with the aim to destroy the connectivity of the network. The default initial performance is 100%, and the key nodes are attacked intentionally. In addition, the relationship between the decreasing trend of vulnerability and the damaged nodes is analyzed, as shown in Fig. 4 below.

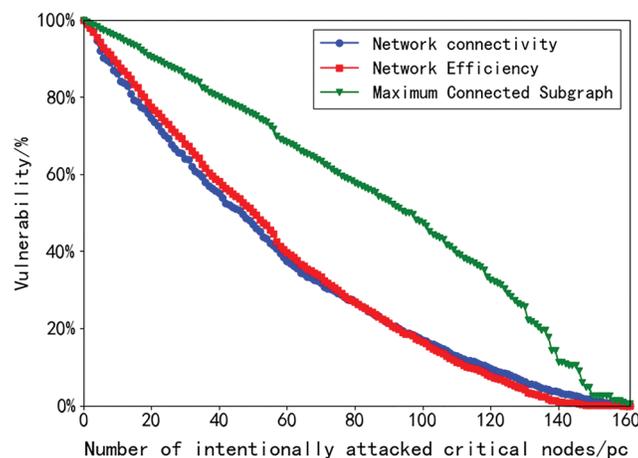


Figure 4: Vulnerability change under deliberate attack

The change of vulnerability after each node is attacked in Fig. 3 is not consistent. Besides, the nodes with greater importance cause more damage to the network vulnerability. The network efficiency and connectivity indexes change similarly, when the attack reaches the 79th node, the network efficiency decreases faster compared to the maximum connectivity subgraph size, which tends to decrease more slowly and finally

converges to approximately 3%. On the whole, the network efficiency and connectivity drop faster and then slower. Meanwhile, the maximum connectivity subgraph size is almost constant throughout the paralysis rate, presenting a linear downward trend. When the network is deliberately attacked, the feeder airline is vulnerable, and when the critical nodes are attacked, the network effectiveness and overall working performance are significantly affected.

4.2 Random Attack

Based on the previous DBSCAN clustering of nodes into four classes of nodes class_0-class_3, the degree of vulnerability change of the feeder airline network is found by observing the four classes of nodes that fail under random attacks and influence the feeder airline network, as follows.

Step1: Randomly attack the nodes in each category according to category 1, category 2, category 3 and category 4 as the attack objects, respectively, and remove the corresponding nodes.

Step2: Calculate the vulnerability index after removing the nodes

Step3: Plot, analyze, and draw conclusions.

Using Sublime Text software, write Python programs for node attacks. The nodes in the airline network are damaged sequentially based on the steps to destroy the connectivity of the network. The default initial performance is 100%. Random attacks are performed on nodes in the four categories. Besides, the relationship between the decreasing trend of vulnerability and the damaged nodes is analyzed, as shown in Fig. 5 below.

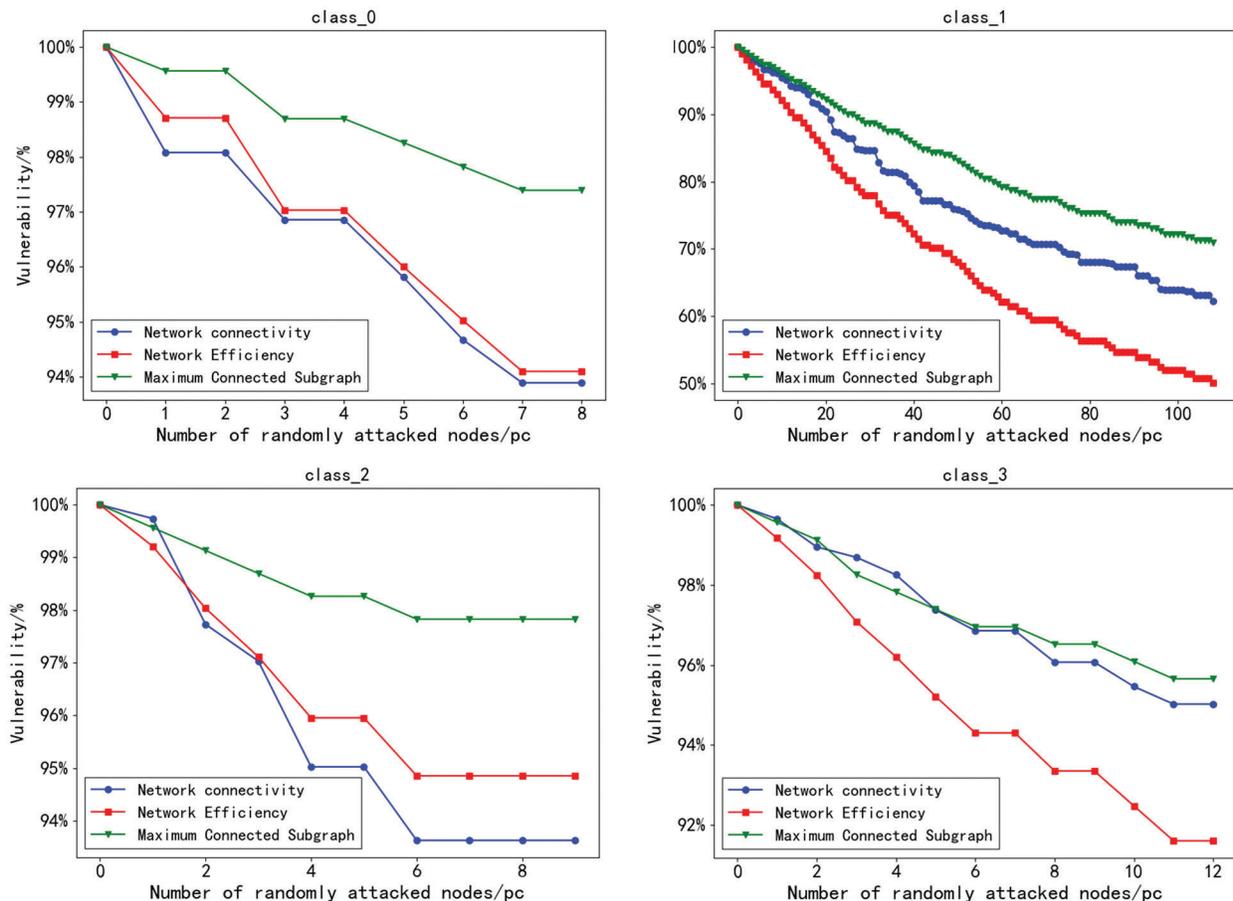


Figure 5: Vulnerability change under random attack

According to Fig. 5, when the nodes in class_0 are attacked first by random attacks, the maximum connectivity subgraph size indicator has a slow decreasing trend, and when the attacks reach 1, 3 and 8 nodes, the network efficiency and connectivity are significantly influenced. When the nodes in class_1 are attacked first, the paralysis rate of the maximum connectivity subgraph and network connectivity is almost constant with a linear decreasing trend, and when 40 nodes are attacked, the network efficiency decreases to 75%, showing that the airport node importance exerts a greater impact on the network effectiveness. Compared with deliberate attacks, the vulnerability index decreases slowly, and the regional airline network under random attacks exhibits some robustness and better stability than deliberate attacks. In addition, the stability of the whole network can be improved after protecting critical nodes.

5 Conclusion

To conclude, by combining the complex network theory and the related research of DBSCAN clustering algorithm, we have performed an applied study on the regional airline network based on the previous research results, which mainly includes the following aspects.

- (1) This study links the indicators in complex networks with the DBSCAN algorithm in machine learning, explores new node clustering methods, and provides certain reference ideas for the cross research of related disciplines.
- (2) Based on complex network theory, complex network indicators such as node degree and number of point mediators of regional airline network, as well as vulnerability indicators such as network efficiency, maximum connectivity subgraph and network efficiency can reflect the problems of insufficient connection between airports and topological waste, which can thus provide a basis for further planning and adjustment of subsequent routes.
- (3) After a comprehensive assessment of 161 regional aviation nodes, it was found that airports such as Luoyang, Rizhao and Xiangyang, which do not have high node degrees and node meshes, can also be applied as key nodes for key prevention due to higher landing and takeoff times and prosperity.
- (4) Through deliberate attacks on airport nodes and random attacks on four types of classified nodes, it is found that the maximum connectivity subgraph index has less impact, the network efficiency and connectivity indexes change in a similar trend, and that the network stability is improved after protecting the critical nodes.

During the research of this study, some areas were also found to be improved.

- (1) The selection for parameters Eps and Minpts in DBSCAN algorithm needs to be optimized, and the classification of clusters needs to be verified.
- (2) Due to the epidemic, the passenger volume of regional airports is small and the disparity is small, leading to the poor entropy weight method and clustering results.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] P. W. Zhang and H. Sun, "Study on the current situation of regional airline development and support strategies," *Journal of Civil Aviation Flight Academy of China*, vol. 26, no. 1, pp. 11–15, 2015.
- [2] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 2010.
- [3] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

- [4] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Social Network*, vol. 1, no. 3, pp. 215–239, 1979.
- [5] Y. Yang, Y. Liu and M. Zhou, "Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing subway," *Safety Science*, vol. 79, no. 79, pp. 149–162, 2015.
- [6] Y. Tian, "Network node importance assessment algorithm based on network structured metrics," *Mobile Communication*, vol. 42, no. 8, pp. 62–66, 2018.
- [7] H. M. Mo, "A Multi-attribute based method for identifying the importance of complex network nodes," *Journal of Lanzhou College of Arts and Sciences (Natural Science Edition)*, vol. 32, no. 5, pp. 68–74, 2018.
- [8] X. P. Zhang, Y. S. Li, G. Liu and L. Wang, "A node importance assessment method for complex networks with node importance contribution," *Complex Systems and Complexity Science*, vol. 11, no. 3, pp. 26–32+49, 2014.
- [9] F. Wang, L. W. Xu and Y. F. Zheng, "Critical node evaluation of complex networks based on m-order neighbor nodes," *Journal of Fuzhou University (Natural Science Edition)*, vol. 47, no. 2, pp. 237–243, 2019.
- [10] B. Che, L. Liu and H. Zhang, "KNEMAG: Key node estimation mechanism based on attack graph for IOT security," *Journal of Internet of Things*, vol. 2, no. 4, pp. 145–162, 2020.
- [11] C. F. Wang and D. L. Wang, "Importance assessment of airport network nodes based on TOPSIS model," *Mathematical Practice and Understanding*, vol. 51, no. 1, pp. 79–87, 2021.
- [12] Y. Z. Wang, "Research on multi-density clustering algorithm," *Ph.D. dissertation*, University of Jilin, Jilin, 2020.
- [13] L. Sun, Y. L. Wang, Z. G. Qu and N. N. Xiong, "BeatClass: A sustainable ECG classification system in IoT-based eHealth," *IEEE Internet of Things Journal*, vol. 99, pp. 1, 2021. <https://doi.org/10.1109/JIOT.2021.3108792>.
- [14] M. M. Hassan, A. S. Eesa, A. J. Mohammed and W. K. Arabo, "Oversampling method based on Gaussian distribution and k-means clustering," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 451–469, 2021.
- [15] Y. Xue, J. Qin, S. Su and A. Slowik, "Brain storm optimization based clustering for learning behavior analysis," *Computer Systems Science and Engineering*, vol. 39, no. 2, pp. 211–219, 2021.
- [16] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492, 2014.
- [17] Y. R. Dang and S. Z. Song, "Destructiveness analysis of Chinese air passenger network based on centrality," *Complex Systems and Complexity Science*, vol. 10, no. 1, pp. 75–82, 2013.
- [18] X. Z. Yang and Z. X. Wang, "Structural characteristics and survivability of China's passenger route network," *Journal of Henan University of Science and Technology (NATURAL SCIENCE EDITION)*, vol. 41, no. 2, pp. 21–26 + 5-6, 2020.
- [19] L. L. Yan, Z. Q. Chen and Q. Zhang, "Analysis of important nodes in China's aviation network based on degree and clustering coefficient," *Journal of Intelligent Systems*, vol. 11, no. 5, pp. 586–593, 2016.
- [20] P. Wei, M. D. Lu, X. H. Zhang and Y. L. Teng, "Analysis of flight punctuality rate based on complex network," *Journal of Quantum Computing*, vol. 3, no. 1, pp. 13–23, 2021.
- [21] C. Q. Ma, S. Zhang, Q. Chen, R. Cao and L. Ren, "Characteristics of rail transit network and its vulnerability from the perspective of passenger flow characteristics," *Journal of Transportation Engineering*, vol. 20, no. 5, pp. 208–216, 2020.
- [22] X. L. Wang, J. Y. Li, M. K. Lian and Z. J. Sun, "Vulnerability analysis of metro system based on complex network model," *Urban Rail Transit Research*, vol. 24, no. 8, pp. 47–50, 2021.
- [23] N. Y. Mou, Q. P. Kang and C. F. Jia, "Vulnerability assessment of urban express network under the impact of emergency events," *Chinese Journal of Security Science*, vol. 30, no. 12, pp. 125–132, 2020.
- [24] H. He, Z. Zhao, W. Luo and J. Zhang, "Community detection in aviation network based on k-means and complex network," *Computer Systems Science and Engineering*, vol. 39, no. 2, pp. 251–264, 2021.
- [25] D. Cheng, L. Zhang, X. Y. Zhao, F. Li, X. Y. Wang *et al.*, "Prediction of new energy vehicle users' residence based on clustering algorithm," *Automotive Practical Technology*, vol. 46, no. 10, pp. 11–13, 2021.
- [26] X. H. Ren and L. Yang, "Vulnerability analysis of air cargo network in China," *Journal of Safety and Environment*, vol. 20, no. 3, pp. 840–848, 2020.