# Homogeneity Analysis of Multi-airport System Based on Airport Attributed Network Representation Learning

*LIU Caihua*[1,2], *CAI Rui*[1], *FENG Xia*[1,2*], *XU Tao*[1,2]

1. College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, P. R. China;
2. Key Laboratory of Intelligent Application Technology for Civil Aviation Passenger Service, Beijing101318, P. R. China

(Received 10 December 2020; revised 15 June 2021; accepted 10 August 2021)

**Abstract:** The homogeneity analysis of multi-airport system can provide important decision-making support for the route layout and cooperative operation. Existing research seldom analyzes the homogeneity of multi-airport system from the perspective of route network analysis, and the attribute information of airport nodes in the airport route network is not appropriately integrated into the airport network. In order to solve this problem, a multi-airport system homogeneity analysis method based on airport attribute network representation learning is proposed. Firstly, the route network of a multi-airport system with attribute information is constructed. If there are flights between airports, an edge is added between airports, and regional attribute information is added for each airport node. Secondly, the airport attributes and the airport network vector are represented respectively. The airport attributes and the airport network vector are embedded into the unified airport representation vector space by the network representation learning method, and then the airport vector integrating the airport attributes and the airport network characteristics is obtained. By calculating the similarity of the airport vectors, it is convenient to calculate the degree of homogeneity between airports and the homogeneity of the multi-airport system. The experimental results on the Beijing-Tianjin-Hebei multi-airport system show that, compared with other existing algorithms, the homogeneity analysis method based on attributed network representation learning can get more consistent results with the current situation of Beijing-Tianjin-Hebei multi-airport system.

**Key words:** air transportation; multi-airport system; homogeneity analysis; network representation learning; airport attribute network

**CLC number:** TP399 **Document code:** A **Article ID:** 1005-1120(2021)04-0616-09

## 0 Introduction

Multi-airport system means multiple airports serving the same region, and the analysis of multi-airport system has become a hot research topic. The research on multi-airport system mainly focuses on three aspects: Analyzing the development of multi-airport system from multiple perspectives[1-3], airport selection of passengers in multi airport system[4-6], and the homogeneity of multi-airport system[7-10].

The homogeneity of multi-airport system refers to the phenomenon that the airports in the multi-airport system lose their unique characteristics and gradually converge, which is mainly manifested in the large overlap of route layout. Research on the homogeneity of multi-airport systems is of great significance to the synergetic development of multi-airport system, and it can help prevent multi-airport system from losing its unique characteristics, wasting resources, and restricting the development of multi-airport system.

Since there is no uniform standard for the homogen- eity of multi-airport system, researchers can only indirectly evaluate the homogeneity of multi-

*Corresponding author, E-mail address: xfeng@cauc.edu.cn.

**How to cite this article**: LIU Caihua, CAI Rui, FENG Xia, et al. Homogeneity analysis of multi-airport system based on airport attributed network representation learning[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2021, 38(4): 616-624.

http://dx.doi.org/10.16356/j.1005-1120.2021.04.008

airport system from the perspective of calculating similarity. Ref.[7] obtained the convergence coefficient of each airport in the Pearl-River-Delta multi-airport system via a comprehensive calculation of the proportion and growth rate of the passenger transportation capacity. Jiang et al.[8-9] analyzed the degree of homogeneity of the Yangtze-River-Delta multi-airport system by calculating the difference between the flight frequency, passenger throughput, and cargo and mail throughput of each airport in the proportion of the data in the entire region. These methods treated all airports in the multi-airport system as discrete individuals. They all lacked the consideration of the network formed by the multi-airport system and its navigable airports.

Ref.[10] considered the homogeneity of multi-airport system from two aspects: Airport attribute similarity and airport network similarity. Airport attributes include the number of infrastructure, passenger throughput, etc. Airport network is constrcuted by the airports in the multi-airport system and their navigable airports. However, some airport attributes are less related to the homogeneity of multi-airport system, which seriously affect the accuracy of the homogeneity calculation. In addition, the attribute information of airports and the airport network are calculated separately and then weighted fused directly, which leads to the loss of the characteristics of the two parts of the algorithm. In order to better integrate airport attributes and airport network information, the airport network and airport attributes should be calculated synchronously under the same framework to form an organic integration of these two parts.

In order to solve the above problems, this paper proposes a multi-airport system homogeneity analysis algorithm based on the airport attributed network representation learning. The airport attribute information has also been changed from the number of runways and other infrastructures to the route region information, which is believed to be able to better calculate the degree of homogeneity of the multi-airport system. Airport attribute information and airport network information are no longer calculated separately. The airport network node vector and the attribute node vector are uniformly and synchronously embedded into a unified low dimensional space through network representation learning. The homogeneity of multi-airport system is calculated by averaging the similarity between the airport vectors obtained by the airport attributed network representation learning. Finally, the rationality of the algorithm is evaluated based on the data of Beijing-Tianjin-Hebei airport multi-airport system.

## 1 Airport Representation Learning with Attributed Network

The homogeneity of multi-airport system is calculated by analyzing the route layout. Airports and their navigable airports naturally form an airport graph, as shown in Fig.1. In Fig.1, the solid circles represent airports in the multi-airport system, and the hollow circles represent navigable airport nodes. The connection line represents the connection between airports based on the route. Each airport node has attribute information, which is represented by a rectangular box.

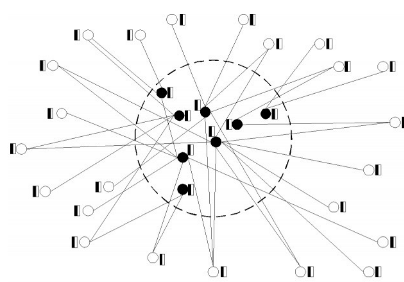


Fig.1 Airport attribute network diagram

Through graph representation learning, the node vector representation of airports considering the global route layout of multi-airport system can be obtained, and the homogeneity degree between airports can be calculated via the node vector. Therefore, the homogeneity algorithm of multi-airport system is expressed as a problem of airport network representation learning. In order to better learn the vector representation of airport nodes, the route region attribute is introduced into the airport network. So it is formulated as an attributed network representation learning problem. In order to get the unified representation of airport network and airport attribute vector, it is necessary to represent the airport structure vector and the airport attribute vector respectively, and optimize their fusion embedding representation through a unified loss function.

$N=\{G, A\}$ is an airport attribute network with $n$ airport nodes, $G \in \mathbf{R}^{n \times n}$ represents the airport network structure, which is a adjacency matrix with weight. The element $(i, j)$ of the matrix means that there is a route between the airport $i$ and the airport $j$. If there is no route between them, the value is set to 0. $A \in \mathbf{R}^{n \times m}$ is a matrix representing the attribute information of the airports in the network. The data in row $i$ of matrix $A$ represents the $m$-D attributes of the airport $i$.

The algorithm process diagram is shown in Fig.2. The model includes two branches: One constructs a learning model $H^{(G)}$ based on the node similarity of the node network structure to obtain the potential vector space representation of node structure, the other constructs a model $H^{(A)}$ to learn the potential vector space representation of airport node attribute information. The two learned vector representations are fused into a unified low dimensional space vector of airport features, represented as $H$.

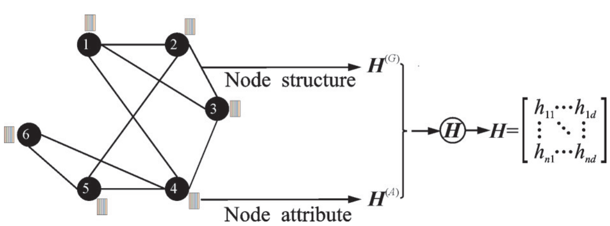


Fig.2 Network representation learning of airport attribute

### 1.1 Network structure vector representation

The network structure similarity matrix $S^{(G)}$ is constructed by the structural similarity between the airport nodes, and the structural similarity of the nodes is calculated by the cosine similarity of the nodes in the airport network structure matrix $G$, where element $s_{ij}^{(G)}$ represents the similarity between nodes $i$ and $j$. If the two airport nodes are similar, the hidden vectors should be similar in the network node vector space obtained by network representation learning[11]. Then, the vector representation learning model of airport network structure is

$$\min_{H^{(G)}} \frac{1}{2} \sum_{i,j=1}^{n} s_{ij}^{(G)} \times \left\| \frac{\boldsymbol{h}_i^{(G)}}{\sqrt{d_i^{(G)}}} - \frac{\boldsymbol{h}_j^{(G)}}{\sqrt{d_j^{(G)}}} \right\|_2 \tag{1}$$

where $n$ is the total number of airport network nodes. $\boldsymbol{h}_i^{(G)}$ and $\boldsymbol{h}_j^{(G)}$ are the hidden vector representations in network representation space, and $d_i^{(G)}$ and $d_j^{(G)}$ the normalizations for $\boldsymbol{h}_i^{(G)}$ and $\boldsymbol{h}_j^{(G)}$, which are the sums of row $i$ and row $j$ data in matrix $S^{(G)}$, respectively.

The Laplacian matrix and trace of the matrix are introduced to transform Eq.(1) by referring to the related knowledge of spectral clustering[12]. The transformed model and the constraint condition are

$$\max_{H^{(G)}} Z_G = \mathrm{Tr}(H^{(G)^{\mathrm{T}}} L^{(G)} H^{(G)}) \tag{2}$$

$$H^{(G)^{\mathrm{T}}} H^{(G)} = E \tag{3}$$

$$L^{(G)} = D^{(G)^{-\frac{1}{2}}} S^{(G)} D^{(G)^{-\frac{1}{2}}} \tag{4}$$

where $D^{(G)}$ is a diagonal matrix, $d_i^{(G)}$ the element in the row $i$ and column $i$, and $E$ the identity matrix. By solving the objective function $Z_G$, the representation vector $H^{(G)}$ of airport node structure representation is obtained.

### 1.2 Node attribute vector representation

Similar to the representation learning process of the airport node network structure, the airport node attribute information matrix $A$ is obtained via the cosine similarity calculation of the node attribute information $S^{(A)}$. In order to minimize the divergence degree between the node similarity calculated by Euclidean distance and the node similarity in similarity matrix, the objective function is constructed as

$$\max_{H^{(A)}} Z_A = \mathrm{Tr}(H^{(A)^{\mathrm{T}}} L^{(A)} H^{(A)}) \tag{5}$$

$$H^{(A)^{\mathrm{T}}} H^{(A)} = E \tag{6}$$

$$L^{(G)} = D^{(G)^{-\frac{1}{2}}} S^{(G)} D^{(G)^{-\frac{1}{2}}} \tag{7}$$

where $D^{(A)}$ is a diagonal matrix, and each element on its diagonal is the sum of all elements in the corresponding row of the matrix $S^{(A)}$. By solving the objective function $Z_A$, the representation vector $H^{(A)}$ with node attribute characteristics is obtained.

### 1.3 Node vector fusion representation

Neither network vector nor attribute information vector of airport node can individually express the characteristics and contents of the entire airport attribute network. These two need to be merged into a unified vector space.

In order to make the node attribute information representation vector $H^{(A)}$ and node network structure feature vector $H^{(G)}$ better complement with each other in the whole node vector representation, the principal component analysis method is used to express the correlation between them, as shown in Eq.(8). Similarly, in order to make the features and contents expressed by $H^{(G)}$ and $H^{(A)}$ embodied in $H$, the correlations between $H^{(G)}$ and $H$, $H^{(A)}$ and $H$ should also be reflected in the objective function

$$p_{GA}=\mathrm{Tr}(H^{(A)^{\mathrm{T}}}H^{(G)}H^{(G)^{\mathrm{T}}}H^{(A)}) \tag{8}$$

$$p_{GH}=\mathrm{Tr}(H^{(G)^{\mathrm{T}}}HH^{\mathrm{T}}H^{(G)}) \tag{9}$$

$$p_{AH}=\mathrm{Tr}(H^{(A)^{\mathrm{T}}}HH^{\mathrm{T}}H^{(A)}) \tag{10}$$

On this basis, the sum of $Z_G$, $Z_A$ and their correlation are obtained. The vector spaces $H^{(G)}$ and $H^{(A)}$ are fused by maximizing the solution. However, the direct summation operation will lead to a large proportion of node network structure characteristics or node attribute information in the overall node representation. Therefore, two weight adjustment parameters $\lambda_1$ and $\lambda_2$ are introduced to trade-off the attribute information representation $H^{(A)}$ and the association between $H^{(A)}$ and $H^{(G)}$ in the objective function. The objective function is solved as in Eq. (11), and the constraint conditions are as in Eq. (12). The overall node vector space $H$ is obtained by solving the objective function $Z$ in the entire network.

$$\max_{H^{(G)},H^{(A)},H} Z=Z_G+\lambda_1 Z_A+\lambda_2 p_{GA}+p_{GH}+p_{AH} \tag{11}$$

$$H^{(G)^{\mathrm{T}}}H^{(G)}=E,\ H^{(A)^{\mathrm{T}}}H^{(A)}=E,\ H^{\mathrm{T}}H=E \tag{12}$$

If Eq. (11) is directly solved in the solution space of the whole network, it has high requirements on the computing performance and time of the computer, and it may not get the expected solution. In order to solve this problem, we change the global optimal solution to the local optimal solution, and solve the equations by Lagrange extremum. The specific optimization method is as follows.

Firstly, calculating the second-order partial derivative of $Z$ in Eq.(11), we get

$$\nabla^2_{H^{(G)}}Z=L^{(G)}+\lambda_2 H^{(A)}H^{(A)^{\mathrm{T}}}+HH^{\mathrm{T}} \tag{13}$$

Then, let the second derivative be equal to 0, we have

$$\begin{cases}(L^{(G)}+\lambda_2 H^{(A)}H^{(A)^{\mathrm{T}}}+HH^{\mathrm{T}})H^{(G)}=\alpha_1 H^{(G)}\\(\lambda_1 L^{(A)}+\lambda_2 H^{(G)}H^{(G)^{\mathrm{T}}}+HH^{\mathrm{T}})H^{(A)}=\alpha_2 H^{(A)}\\(H^{(G)}H^{(G)^{\mathrm{T}}}+H^{(A)}H^{(A)^{\mathrm{T}}})H=\alpha_3 H\end{cases} \tag{14}$$

where $\alpha_1, \alpha_2, \alpha_3$ are the eigenvalues. Eqs. (12, 13) are updated iteratively until the loss difference of Eq. (11) is less than the set threshold. Finally, the first $D$ eigenvectors in the result are taken as the final vector space solution $H$.

## 2 Homogeneity Calculations

After obtaining airport node representation vector $H$, the homogeneity degree within the multi-airport system is calculated. Taking any two airports $i$ and $j$ as an example, if the representation vectors of airports $i$ and $j$ are $\boldsymbol{h}_i=[x_{i1}, x_{i2}, \cdots, x_{id}]$ and $\boldsymbol{h}_j=[y_{j1}, y_{j2}, \cdots, y_{jd}]$, respectively, the homogeneity degree between airports $i$ and $j$ can be expressed as

$$T_{ij}=\frac{1}{2}+\frac{1}{2}\times\frac{\boldsymbol{h}_i\cdot\boldsymbol{h}_j}{\|\boldsymbol{h}_i\|\times\|\boldsymbol{h}_j\|}=\frac{1}{2}+\frac{1}{2}\times\frac{\sum_{k=1}^{d}(x_{ik}\times y_{jk})}{\sqrt{\sum_{k=1}^{d}(x_{ik})^2}\times\sqrt{\sum_{k=1}^{d}(y_{jk})^2}} \tag{15}$$

Based on the above-mentioned homogeneity degree between airports, the homogeneity degree of one airport in the multi-airport system can be calculated, which is the average value of the homogeneity degree between other airports and this airport, as shown in Eq. (16). The degree of homogeneity of the multi-airport system is to re-average the average homogeneity degree of a single airport in the multi-airport system

$$\bar{t}_i=\frac{\sum_{j\in[1,i)\cup(i,n)}T_{ij}}{n_0-1} \tag{16}$$

$$T_{A\_G}=\frac{\sum_{i\in[1,n]}\bar{t}_i}{n_0} \tag{17}$$

where $T_{ij}$ is the element in the matrix $T$, $1\leqslant i,j\leqslant n_0$, which represents the homogeneity degree among $n_0$ airports; $\overline{t_i}$ the average homogeneity degree of a single airport in the multi-airport system; and $T_{A\_G}$ the homogeneity degree of the multi-airport system.

## 3 Experiment and Analysis

### 3.1 Dataset

The experiment uses the data of Beijing-Tianjin-Hebei multi-airport system, including two parts: (1) The airport network constructed by the airports in the multi-airport system and their navigable airports. The data are derived from the flight data of Beijing-Tianjin-Hebei multi-airport system in 201X. The data example is shown in Table 1. The first column represents the airports within the the multi-airport system, and the second column represents the airports with flights within the multi-airport system. The airport is represented by three character code: HDG, Handan Airport; NAY, Beijing Nanyuan airport; PEK, Beijing Capital Airport; SJW, Shijiazhuang Airport; TSN, Tianjin Airport; CAN, Guangzhou Baiyun Airport; BAV, Baotou Airport; CGD, Changde Airport; SHA, Shanghai Hongqiao Airport; HAK, Haikou Airport, etc. (2) The attributes of airport nodes in the airport network include the province and location information of each airport. The province information of the airport comes from the publicly released airport introduction, and the codes and administrative regions of Chinese provinces issued by the National Bureau of statistics [13] (Table 2).

In this paper, the three-character code is replaced with the digital code, and the administrative region is represented by numbers. All the nodes of the airport network constructed by Beijing-Tianjin-Hebei airport multi-airport system and its navigable airports are replaced by numbers. The seven administrative regions are North China, Northeast China, East China, Central China, South China, Southwest China and Northwest China, coded with numbers 1—7. The experimental data of node attribute information are shown in Table 3.

Based on Tables 1,3, the airport node network structure matrix and the airport node attribute information matrix are constructed respectively. The algorithm proposed in Section 2 is used to carry out the experiment. The airport node attribute information and network structure association value is set as 0.1, the weight of airport node attribute information is set as 0.4, and the dimension of airport node representation vector is set as 32.

**Table 1 Airport network related data example**

| Airport | Airports with flight |
|---|---|
| HDG | CAN |
| NAY | BAV |
| PEK | CGD |
| SJW | SHA |
| TSN | HAK |

**Table 2 Examples of province and administrative region in China**

| Province | Province code | Region |
|---|---|---|
| Beijing | 11 | North China |
| Liaoning | 21 | Northeast China |
| Shanghai | 31 | East China |
| Henan | 41 | Central China |
| Chongqing | 50 | Southwest |

**Table 3 Experimental data of airport attribute information**

| Airport code | Province code | Region code |
|---|---|---|
| 1 | 11 | 2 |
| 9 | 50 | 7 |
| 20 | 44 | 4 |
| 21 | 43 | 5 |
| 22 | 61 | 6 |

### 3.2 Baselines

The algorithm in Ref. [10] is chosen as the benchmark algorithm. It considers the homogeneity of multi-airport system from two aspects: Airport attribute similarity and airport network similarity. In this method, the node vector representation of the airport network is obtained through the network representation learning algorithm node2vec [14-15]. The dimension of the airport node vector is set to 32. The homogeneity degree between airports in Beijing-Tianjin-Hebei multi-airport system is calculated by fusing the airport node vectors and the similarity computed by the airport attributes.

The homogeneity degree is expressed by the color depth. In order to more intuitively express the homogeneity degree between airports in Beijing-Tianjin-Hebei multi-airport system, the confusion matrix is used to express the result, as shown in Table 4 and Fig.3.

**Table 4 Homogeneity of Beijing-Tianjin-Hebei airport group in 201X by baseline method**

| Airport | Beijing | Tianjin | Shijiazhuang | Nanyuan | Handan | Tangshan | Qinhuangdao | Zhangjiakou |
|---|---|---|---|---|---|---|---|---|
| Beijiing | 1.000 | 0.651 | 0.536 | 0.367 | 0.422 | 0.409 | 0.415 | 0.377 |
| Tianjin | 0.651 | 1.000 | 0.705 | 0.492 | 0.497 | 0.466 | 0.481 | 0.423 |
| Shijiazhuang | 0.536 | 0.705 | 1.000 | 0.626 | 0.591 | 0.556 | 0.561 | 0.491 |
| Nanyuan | 0.367 | 0.492 | 0.626 | 1.000 | 0.524 | 0.470 | 0.484 | 0.421 |
| Handan | 0.422 | 0.497 | 0.591 | 0.524 | 1.000 | 0.778 | 0.782 | 0.640 |
| Tangshan | 0.409 | 0.466 | 0.556 | 0.470 | 0.778 | 1.000 | 0.873 | 0.720 |
| Qinhuangdao | 0.415 | 0.481 | 0.561 | 0.484 | 0.782 | 0.873 | 1.000 | 0.681 |
| Zhangjiakou | 0.377 | 0.423 | 0.491 | 0.421 | 0.640 | 0.720 | 0.681 | 1.000 |

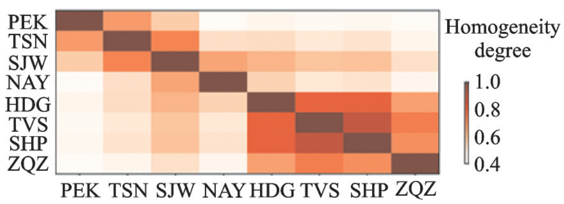


Fig.3 Homogeneity degree of Beijing-Tianjin-Hebei multi-airport system based on baseline algorithm[10] shown in heat map

## 3.3 Experimental results

The experimental results are shown in Table 5. According to Eq.(14), the average homogeneity degrees of Beijing Capital Airport, Tianjin Airport, Shijiazhuang Airport are 0.623 0, 0.617 5, 0.629 5, respectively. Similarly, the average homogeneity degree of other five airports in the multi-airport system and each airport can be obtained. On this basis, according to Eq.(15), the average homogeneity degree of each airport in the airport group is calculated again, and the homogeneity degree of the multi-airport system can be obtained. Therefore, the homogeneity degree of the development of Beijing-Tianjin-Hebei multi-airport system in 201X is 0.583 7. All the homogeneity degrees range from 0 to 1, and that of multi-airport system is greater than 0.5. Therefore, the development of Beijing-Tianjin-Hebei airport group in 201X has obvious homogeneity characteristics.

**Table 5 Homogeneity of Beijing-Tianjin-Hebei airport group in 201X by the proposed method**

| Airport | Beijing | Tianjin | Shijiazhuang | Nanyuan | Handan | Tangshan | Qinhuangdao | Zhangjiakou |
|---|---|---|---|---|---|---|---|---|
| Beijiing | 1.000 0 | 0.898 6 | 0.736 3 | 0.692 9 | 0.515 6 | 0.508 4 | 0.554 5 | 0.455 3 |
| Tianjin | 0.898 6 | 1.000 0 | 0.731 1 | 0.696 6 | 0.498 8 | 0.497 6 | 0.503 4 | 0.496 2 |
| Shijiazhuang | 0.736 3 | 0.731 1 | 1.000 0 | 0.621 6 | 0.547 9 | 0.636 8 | 0.719 5 | 0.413 4 |
| Nanyuan | 0.692 9 | 0.696 6 | 0.621 6 | 1.000 0 | 0.718 5 | 0.619 8 | 0.601 2 | 0.512 6 |
| Handan | 0.515 6 | 0.498 8 | 0.547 9 | 0.718 5 | 1.000 0 | 0.543 7 | 0.531 2 | 0.514 9 |
| Tangshan | 0.508 4 | 0.497 6 | 0.636 8 | 0.619 8 | 0.543 7 | 1.000 0 | 0.568 7 | 0.501 4 |
| Qinhuangdao | 0.554 5 | 0.503 4 | 0.719 5 | 0.601 2 | 0.531 2 | 0.568 7 | 1.000 0 | 0.505 7 |
| Zhangjiakou | 0.455 3 | 0.496 2 | 0.414 3 | 0.512 6 | 0.514 9 | 0.501 4 | 0.505 7 | 1.000 0 |

It can be seen from Table 5 that the three large-scale airports in the Beijing-Tianjin-Hebei multi-airport system have a great degree of homogeneity. As a super large airport in the world, Beijing Capital Airport has great development advantages in Beijing-Tianjin-Hebei region. From the information of navigable airport, province, administrative region and transportation scale, Tianjin Airport and Shijiazhuang Airport rank second and third in the multi-airport system, showing a high homogeneity degree with the Beijing Capital Airport, with the degree of homogeneity being 0.898 6 and 0.736 3, respectively. The homogeneity degree between Tianjin Airport and Shijiazhuang Airport is 0.731 1, both of which are alternate airports of Beijing Capital Airport and regional hub airport, and the homogeneity degree is also obvious. The more intuitive confusion matrice of homogeneity degree among airports of Beijing-Tianjin-Hebei airport group is shown in Fig.4, and the depth of color represents different degrees of homogeneity.

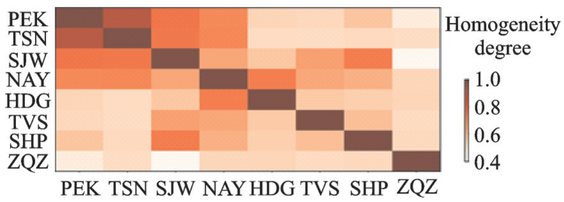


Fig.4 Homogeneity degree of Beijing-Tianjin-Hebei airport group based on learning method of attribute network representation shown in heat map

### 3.4 Experiment analysis

The homogeneity degree of Beijing-Tianjin-Hebei multi-airport system calculated by the baseline method[10] (represented as M1) is 0.516 6, and that calculated by the attributed network representation learning method (represented as M2) is 0.583 7. Obviously, the homogeneity calculated by M1 is less than that calculated by M2. From the point of the homogeneity among airports, the average homogeneity of a single airport obtained by M1 is also lower than that obtained by M2. Moreover, the homogeneity degree between airports obtained by M1 is relatively concentrated, while the homogeneity degree between airports obtained by M2 is quite different, ranging from 0.413 4 to 0.898 6.

The results of M2 are more reasonable: (1) The navigable airport numbers of the eight airports are 150, 88, 52, 54, 6, 9, 12 and 3. Taking the homogeneity with Beijing Capital Airport as an example, the homogeneity degrees of Tianjin Airport, Shijiazhuang Airport, Nanyuan Airport should be decreased but higher than those of the other four airports. However, there is little diffrence between the homogeneity degrees obtained by M1 algorithm, and the homogenization degrees of Tangshan Airport, Qinhuangdao Airport are higher than that of Nanyuan Airport, which is unreasonable. (2) The homogeneity degree obtained by M2 between Handan Airport and Nanyuan Airport is relatively large, because these two airports are the only airports that are not accessible to North China. (3) The homogeneity degrees among several small-scale airports calculated by M1 is higher than those by M2, because indicators such as infrastructure in M1 occupy a relatively large proportion. The number of these infrastructures is similar in small airports, but they cannot reflect the airport homogeneity degree. From the perspective of route layout, although the navigation directions have a certain degree of overlap, the proportion is quite different. The only same navigable airport among the four airports is Shanghai Pudong Airport. Both Handan Airport (six navigable airports) and Zhangjiakou Airport (three navigable airports) have no routes in Shandong, Xi'an, Neimeng directions, so the homogeneity degree should not be so high. In summary, the result of M2 is consistent with the actual situation, but the result of M1 is quite different. Therefore, adding the attribute information to the airport node has a great positive influence on the homogeneity calculation of the multi-airport system.

## 4 Conclusions

The attribute information which affects the route layout is added to each airport node in the airport network, and a multi-airport system homogeneity analysis algorithm based on attributed network representation learning is proposed. In this algorithm, the vector representations of the structural characteristics of the airport node network and the vector representation of the attribute information feature of the airport node are fused into a uniform airport node vector space, and then the homogeneity degree of the airport group is calculated via the similarity of these airport vectors. The experimental results show that the homogeneity degree of multi-airport system calculated by the proposed method with airport node attribute information outperforms the other fusion methods.

## References


[1] GUARINI M R, CHIOVITTI A, ROCCA F. Multi-criteria spatial decision analysis for the development of the Italian minor airport system[J]. Journal of Advanced Transportation, 2018(2): 1-33.

[2] ZIETSMAN D, VANDERSCHUREN M. Analytic hierarchy process assessment for potential multi-airport systems: The case of cape town[J]. Journal of Air Transport Management, 2014, 36: 41-49.

[3] ZHOU Huiyan, SHI Lina, ZHANG Xu, et al. On coordinated development of China's regional multi-airport compound system[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2011, 13(2): 48-52. (in Chinese)

[4] JUNG S Y, YOO K E. A study on passengers' airport choice behavior using hybrid choice model: A case study of Seoul metropolitan area, South Korea[J]. Journal of Air Transport Management, 2016, 57: 70-79.

[5] LOO B P Y. Passengers' airport choice within multi-airport regions (MARs): Some insights from a stated preference survey at Hong Kong International Airport[J]. Journal of Transport Geography, 2008, 16(2): 117-125.

[6] HU Yihong, ZHU Daoli. Research on airport group in Yangtze river delta based on complex network and passenger selection theory[J]. Shanghai Management Science, 2011(3): 22. (in Chinese)

[7] YANG Yiying, CHEN Ziyi, YANG Xinsheng. Analysis on positioning of collaborative development of airport group around Pearl River Delta Area[J]. Aeronautical Computing Technology, 2016, 46(1): 53-55. (in Chinese)

[8] JIANG Y, WANG L, JIANG X, et al. Spatial-temporal evolution of multi-airports' homogeneity in China Yangtze River Delta[J]. Procedia-Social and Behavioral Sciences, 2013, 96: 1402-1411.

[9] JIANG Yonglei, YANG Zhongzhen, WANG Lu, et al. Homogeneity analysis of airports in China Yangtze River Delta Region[J]. Editorial Committee of Economic Geography, 2013, 33(2): 122-127. (in Chinese)

[10] FENG Xia, CAI Rui, LIU Caihua, et al. Homogeneity cascade analysis of airport group based on fusion of airport attributes and airport network[J]. Journal of Beijing Jiaotong University, 2020, 44(2): 105-110. (in Chinese)

[11] HUANG X, LI J, HU X. Label informed attributed network embedding[C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Cambridge, United Kingdom: ACM, 2017: 731-739.

[12] CHUNG F R K. Spectral graph theory, regional[M]. [S.l.]: American Mathematical Society, 1997.

[13] Ministry of Administration of the People's Republic of China. Administrative region code [EB/OL]. (2020-02-06)[2020-04-28].http://www.mca.gov.cn/article/sj/xzqh/1980/2019/202002281436.html.

[14] GROVER A, LESKOVEC J. Node2vec: Scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S. l.]: ACM, 2016: 855-864.

[15] PEROZZI B, ALRFOU R, SKIENA S, et al. DeepWalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]: ACM, 2014: 701-710.


**Acknowledgements** This work was supported by the Natural Science Foundation of Tianjin (No. 20JCQNJC00720) and the Fundamental Research Fund for the Central Universities(No.3122021052).

**Authors** Dr. **LIU Caihua** received the B.S. degree in software engineering and the Ph.D. degree in computer application technology from Nankai University, China, in 2010 and 2017, respectively. She is a lecturer at the School of Computer Science and Technology, Civil Aviation University of China. Her research interests include computer vision and machine learning.

Prof. **FENG Xia** is a professor at the School of Computer Science and Technology, Civil Aviation University of China. Her research interests mainly focus on data mining and machine learning.

**Author contributions** Dr. **LIU Caihua** designed the study, complied the models and wrote the manuscript. Miss **CAI Rui** contributed to data analysis, result interpretation and manuscript revision. Prof. **FENG Xia** contributed to the discussion and revision of the study. Prof. **XU Tao** contributed to the discussion and background of the study. All authors commented on the manuscript draft and approved the submission

**Competing interests** The authors declare no competing interests.

(Production Editor: SUN Jing)

# 基于机场属性网络表示学习的机场群同质化分析

刘才华[1,2]，蔡 蕤[1]，冯 霞[1,2]，徐 涛[1,2]

（1.中国民航大学计算机科学与技术学院，天津300300，中国；

2.民航旅客服务智能化应用技术重点实验室，北京 101318，中国）

**摘要：**现有机场群研究较少从航线网络分析角度进行机场节点之间的同质化分析，航线网络中机场节点的属性信息考虑不足。为了解决此问题，文中提出了一种基于机场属性网络表示学习的机场群同质化分析方法。首先，构建包含属性信息的机场群航线网络，如果机场之间有航班，则为机场之间添加一条边，并为每个机场节点添加航线区域属性信息。其次，分别进行机场属性以及机场网络向量表示，通过网络表示学习方法将机场属性和机场网络表示向量嵌入到统一的机场表示向量空间，得到融合机场属性以及机场网络特性的机场特征向量。通过计算机场群内机场向量的相似度，可以方便地计算机场之间以及机场群的同质化程度。京津冀机场群数据集的实验结果表明，与目前其他算法相比，本文提出的基于属性网络表示学习的同质性分析方法可以得到更符合京津冀机场群现状的同质化计算结果。

**关键词：**航空运输；机场群；同质化分析；网络表示学习；机场属性网