An Improved FN Algorithm for Community Division of Air Route Network

ZHAO Zheng*, *ZHANG Saiwen*, *XU Lipeng*, *HU Li*

College of Civil Aviation/College of Flight, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China

Abstract: Community division is an important method to study the characteristics of complex networks. The widely used fast-Newman(FN) algorithm only considers the topology division of the network at the static layer, and dynamic traffic flow demand is ignored. The result of the division is only structurally optimal. To improve the accuracy of community division, based on the static topology of air route network, the concept of network traffic contribution degree is put forward. The concept of operational research is introduced to optimize the network adjacency matrix to form an improved community division algorithm. The air route network in East China is selected as the object of algorithm comparison experiment, including 352 waypoints and 928 segments. The results show that the improved algorithm has a more ideal effect on the division of the community structure. The proportion of the number of nodes included in the large community has increased by 21.3%, and the modularity value has increased from 0.756 to 0.806, in which the modularity value is in the range of $[-0.5, 1)$. The research results can provide theoretical and technical support for the optimization of flight schedules and the rational use of air route resources.

Key words: air route network; community division; topological structure; traffic flow contribution

CLC number: V355   **Document code:** A   **Article ID:** 1005-1120(2020)04-0630-08

0 Introduction

The air route network has significant impact on the safety and efficiency of aircraft operation. The traditional research on air route network mostly focuses on regional assessment, and seldom takes into account the interaction between different regions, which cannot fundamentally solve the problems such as airspace congestion and flight delay. The air route network is a typical complex network[1], and the community network theory belongs to the frontier of complex network theory. Studying the characteristics of the community structure of the air route network is helpful to analyze the mutual influences between the various areas of the air route network. Therefore, scholars pay more and more attention to it.

The classical algorithms for community structure division are mainly as follows. In 2002, Newman and Girvan[2] proposed an important GN community division algorithm based on the idea of splitting. According to some connection properties, the key paths connecting the two communities were found out from all edges and eliminated one by one, and the network community structure with the best effect was divided. In 2004, Newman[3] improved the GN algorithm and proposed the fast-Newman (FN) community division algorithm, which is a kind of condensation algorithm. The core idea is that every independent node in the network is regarded as an initial community, and then the result of the maximum modularity increment is taken as the next direction of agglomeration. The latest community contains all the nodes. The community structure with the greatest value-added is taken as the final result of the community division. Yang et al.[4] developed an accurate and scalable algorithm for detecting overlapping communities in networks with

node attributes, which is called communities from edge structure and node attributes (CESNA). Su et al. [5] proposed the multi-level community discovery algorithm in 2018, which relies on a combination of user interests and cohesiveness in describing community structures. To sum up, the existing community division algorithms mainly focus on the topological characteristics of the air route network[6], without considering the impact of the traffic flow of the air route network on the overall operation state of the network. In view of the shortcomings, based on the static topology of the air route network, this paper puts forward the concept of network traffic flow contribution degree, and divides the community of the air route network by combining the static structure index and dynamic flow to improve the FN algorithm.

1　Related Concepts of Community Division

1. 1　Community structure

A community is a collection of units with some common characteristics[7]. The more commonly used definition of community structure is based on the relative connection frequency: the vertices in the network can be divided into groups, the connections within the groups are dense and the connections between groups are sparse. A simple schematic diagram of the community structure[8] is shown in Fig.1.

According to the formation process of community in complex network, the idea of community structure in network can be divided into four catego-
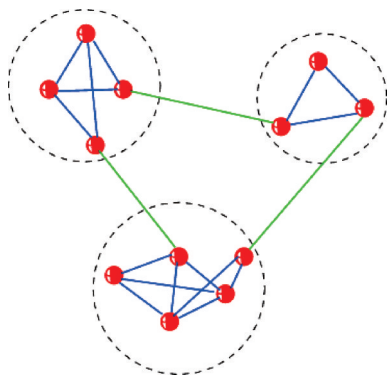

Fig.1　Schematic diagram of community structure

ries[9]: Cohesion process, splitting process, search process and other processes. The condensation process is based on vertices and forms a community through gradual condensation. On the contrary, in the splitting process, all the vertices in the network are regarded as a large community, and a small community is formed through gradual segmentation. The search process is an exploration process of establishing a gradual optimization goal, and the community structure is given directly by the final optimization results.

1. 2　Modularity function

Regarding the measurement standard of community structure, most adopt the modularity function Q defined by Newman[10]. Modularity refers to the expectation of the difference between the proportion of the edge connected to the nodes within the community structure and the proportion of any edge connected to the nodes in the network. The Q function is used to quantitatively describe the modular level of community division. The value of modularity is in the range of $[-0.5,1)$. The larger the Q value, the better the community division effect.

2　Improved Fast-Newman Algorithm

FN algorithm is not only a greedy algorithm for finding the largest module value, but also a condensation one for adding edges to the network. Its core idea is a bottom-up hierarchical clustering process, which detects the community structure by maximizing the modularization. It can greatly reduce the time complexity and can be used to analyze complex networks with millions of nodes.

The specific FN algorithm process is as follows.

Step 1　Consider a complex network of n nodes as n communities, the initial situation is

$$a_{ij} = \begin{cases} 1/2m & \begin{array}{l} \text{Connection between} \\ \text{nodes } i \text{ and } j \end{array} \\ 0 & \text{Else} \end{cases} \quad (1)$$

$$b_i = k_i/2m \quad (2)$$

where k_i is the degree of node i and m the total number of edges contained in the complex network.

Step 2   According to the principle of greedy algorithm, two communities with edge connections are merged in turn along the direction that makes Q increase or decrease fastest, and the Q value increment after the merger is calculated. The formula is

$$\triangle Q = 2(a_{ij} - b_i b_j) \tag{3}$$

Step 3   Replace the corresponding data and add the rows and columns of the i and j communities.

Step 4   Repeat Steps 2 and 3 until each node is merged into a community.

2.1   Traffic flow contribution

The general definition of the network traffic flow is as follows. Assume that all nodes have the ability to pass, and the initial traffic flow is 0. For every carrier passing, the traffic flow of the node increases by 1. In air route network, traffic flow generally refers to the number of aircraft passing a certain flight segment or a certain waypoint per unit time. The traffic flow of the air route network node is defined as "flow", and the flow of the ith node is expressed as "$flow_i$".

For the FN algorithm, only the topological structure division of the static level of the network is considered, which leads to the problem that the community division is often not the optimal community structure. In order to improve the accuracy of community division, based on the static topology of the route network, this paper adds the network dynamic traffic flow contribution as the basis of community division, and proposes an improved FN algorithm named flow-contribution fast-Newman (FC-FN) algorithm.

In the basic FN algorithm, Q is obtained by static topology calculation, but in the actual operation of the network, the traffic flow of each node in the network is different and not fixed, so the importance of each node to the network[11-12] is different.

Define the variable fc_i to represent the contribution of the i node to the Q value in the network.

2.2   Process description of FC-FN algorithm

It is assumed that the connections between the nodes in the network and the edges in the network are known. The network can be defined as $G(V, E)$, where $V = \{v_i | i = 1, 2, 3, \cdots, n\}$ is a collection of nodes; n the number of nodes; $E = \{e_{ij} | i, j = 1, 2, 3, \cdots, m, i \neq j\}$ the edge of the network; and m the number of edges.

In operational research, according to the relationship between total supply and total demand, the transportation problem can be divided into two categories: balanced transportation problem and unbalanced transportation problem. The latter is usually transformed into the former by adding virtual supply nodes or virtual demand nodes. Community division is a complex network problem. To analyze the influence of different "$flow_i$" on Q more effectively, the idea of "balance of supply and demand"[13] is introduced. Classify the node traffic flow, and increase the number of "virtual nodes" to the nodes with large traffic flow.

Step 1   The traffic flow of nodes is a discrete variable, and the median[14] of traffic flow of all nodes in the network is selected as the base of flow classification. This is a method to measure the trend of traffic flow concentration. The median is the representative value of all data determined by its location in the data set, which is not affected by the maximum or minimum of data distribution. When the individual data in a group of data changes greatly, the median can be used to describe the central tendency of this group of data. In the process of calculating the traffic flow of the air route network, data distortion may occur. Choosing the median can avoid errors caused by the extremely large or small individual data to the overall calculation.

Define the traffic flow base of node classification as $flow_{base}$. In order to ensure that the increased "virtual node" can have common characteristics with the original node, "virtual connection" between the nodes should be added as the same as the original connection. The formula for calculating the base flow of node classification is

$$flow_{base} = \text{MEDIAN}(flow_i) \quad i = 1, 2, \cdots, n \tag{4}$$

Step 2   The ROUND function is used to

round up the number of "virtual nodes" needed to calculate node i, which is the traffic flow contribution fc_i. The calculation formula is

$$fc_i = \mathrm{ROUND}\left(\frac{\mathrm{flow}_i}{\mathrm{flow}_{\mathrm{base}}}\right) \quad i = 1, 2, \cdots, n \quad (5)$$

Considering the situation that the traffic flow of the network node is less than half of the $\mathrm{flow}_{\mathrm{base}}$ of the node, $fc_i = 0$, but the actual running flow $\mathrm{flow}_i \neq 0$. To ensure the effectiveness of the calculation, make $fc_i = 1$. The formula for calculating fc_i is

$$fc_i = \begin{cases} \mathrm{ROUND}\left(\dfrac{\mathrm{flow}_i}{\mathrm{flow}_{\mathrm{base}}}\right) & \mathrm{flow}_i \geqslant \dfrac{\mathrm{flow}_{\mathrm{base}}}{2} \\ 1 & \mathrm{Else} \end{cases} \quad (6)$$

where $i = 1, 2, \cdots, n$.

Step 3   Optimize the network adjacency matrix. Define matrix $\boldsymbol{l} = (l_{ij})$ as the initial adjacency matrix of the network. The formula of l_{ij} is

$$l_{ij} = \begin{cases} 1 & \begin{array}{c} \mathrm{Connection\ between} \\ \mathrm{nodes}\ i\ \mathrm{and}\ j \end{array} \\ 0 & \mathrm{Else} \end{cases} \quad (7)$$

A new adjacency matrix $\boldsymbol{l}' = (l'_{ij})$ is obtained by adding the traffic flow contribution. The formula of l'_{ij} is

$$l'_{ij} = \begin{cases} fc_i * fc_j & l_{ij} = 1 \\ 0 & \mathrm{Else} \end{cases} \quad (8)$$

Step 4   The result of network community division includes k communities. A symmetric matrix $\boldsymbol{e} = (e_{ij})$ with k dimensions is defined, where the matrix element e_{ij} represents the proportion of the number of edges connected between the i community and the j community to the total number of edges contained in the network. The formula for calculating the sum of elements on the diagonal of a matrix is

$$\mathrm{Tr}(\boldsymbol{e}) = \sum_i e_{ii} \quad (9)$$

This formula represents the proportion of the number of connected edges between the nodes within the community in the total number of edges in the network.

Assuming that a_i is the sum of the elements in each row or column, which represents the proportion of the edges connected to the nodes in the i community in all edges, the Q function can be expressed as

$$Q = \sum_i (e_{ii} - a_i^2) = \mathrm{Tr}(\boldsymbol{e}) - \|\boldsymbol{e}^2\| \quad (10)$$

where $\|\boldsymbol{e}^2\|$ represents the sum of all the elements in the matrix $\boldsymbol{e}^2$. The upper limit of the Q value is 1. When the value of Q tends to 1, it indicates that the community structure of the entire network is more obvious, and the corresponding division effect of the community structure is better. If the proportion of edges within the community is not greater than the expected value of the proportion of any edges connecting nodes, then $Q \leqslant 0$.

Calculate the Q value of the modularity function under the FC-FN algorithm and record as Q', the formula of Q' is

$$Q' = Q * \left(1 + \frac{\mathrm{flow}_{\mathrm{base}}}{\max(\mathrm{flow}_i)}\right) =$$
$$\left(1 + \frac{\mathrm{flow}_{\mathrm{base}}}{\max(\mathrm{flow}_i)}\right) \sum_i (e_{ii} - a_i^2) =$$
$$\left(1 + \frac{\mathrm{flow}_{\mathrm{base}}}{\max(\mathrm{flow}_i)}\right) (\mathrm{Tr}(\boldsymbol{e}) - \|\boldsymbol{e}^2\|) \quad (11)$$

3   Case Study and Verification

3.1   Overview of air route network in East China

The air route network is mainly composed of waypoints and flight segments. The waypoints represent the nodes in the network, and the flight segments connecting the waypoints represent the edges in the network.

The air route network in East China in October 2017 is selected as the research object. The airspace in the East China region borders the South-Central region to the south and the North China region to the north, including the busy airport clusters such as the Yangtze River Delta. It is one of the busiest airspaces in my country, and the data is researchable. On the basis of deduplication and correction of the original data, 352 waypoints and 928 segments in East China route network are selected for research. The topological structure is shown in Fig.2, in which the circle represents the waypoint and the link between the circles represents the segments.
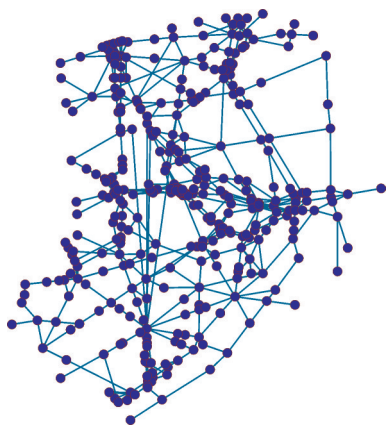
Fig.2   Topological structure diagram of East China air route network

3. 2   Community division of air route network in East China

FN algorithm and FC-FN algorithm are used to divide the community of the air route network in East China.

The division situation is analyzed from two aspects: the results of community division and the modularity function.

3. 2. 1   FC-FN algorithm processing

Due to the large amount of data, select the processing process of the route points "RUPUD" (serial number 17) and "OF" (serial number 18) under the FC-FN algorithm as a demonstration.

Step 1   Calculate the traffic flow base of node classification.

$$\text{flow}_{\text{base}} = \text{MEDIAN}(\text{flow}_i) = 190 \quad i = 1, 2, \cdots, 352$$

Step 2   Calculate the traffic flow contribution of node, and add "virtual node" and "virtual connection" as shown in Fig.3.

$$fc_{17} = 5$$
$$fc_{18} = 13$$

Step 3   Optimize the adjacency matrix accord-
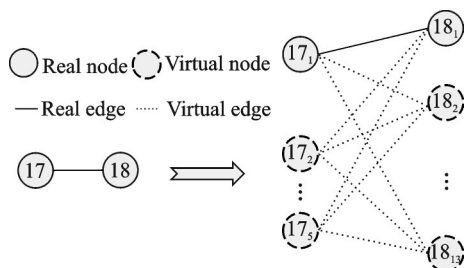

Fig.3   Schematic diagram of adding "virtual node" and "virtual connection" between nodes 17 and 18

ing to the traffic flow contribution.

$$l'_{17,18} = fc_{17} * fc_{18} = 65$$

The adjacency matrix of the network is the main influence factor of the algorithm, which represents the matrix of the adjacency relationship between vertices. For an undirected simple graph, the adjacency matrix must be symmetric and the main diagonal must be zero. Fig.4 shows the transformation of the adjacency matrix with fc_{17} and fc_{18}.
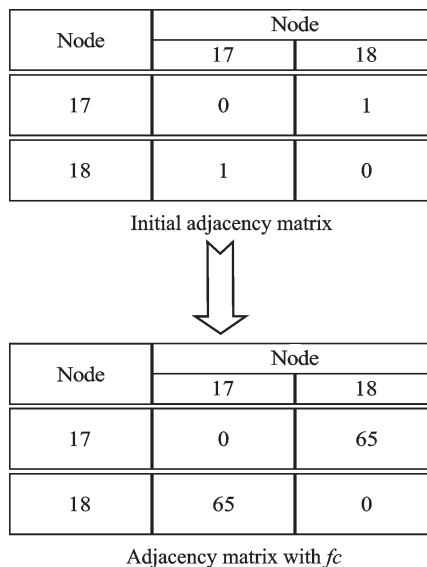
Node	Node	
	17	18
17	0	1
18	1	0

Initial adjacency matrix

Node	Node	
	17	18
17	0	65
18	65	0

Adjacency matrix with fc

Fig.4   Schematic diagram of adjacency matrix transformation with fc (nodes 17 and 18)

3. 2. 2   Analysis of community division results

To facilitate the visual display of the results of community division, each node is colored, and different colors correspond to different communities.

The FN algorithm divides the route network in East China into 11 communities, as shown in Fig.5, and the specific community contains the node information shown in Table 1. The FC-FN algorithm divides the route network in East China into 8 communities, as shown in Fig.6, and the specific community contains the node information shown in Table 2.

Compared with Figs.5 and 6, it can be found that there are some nodes in the community divided by FN algorithm that are not ideal. The community divided by FC-FN algorithm has close internal connections and few external connections, and there are no obvious outliers.
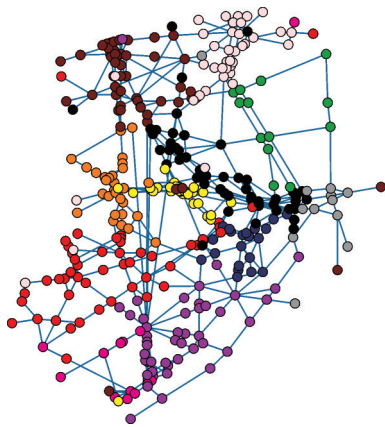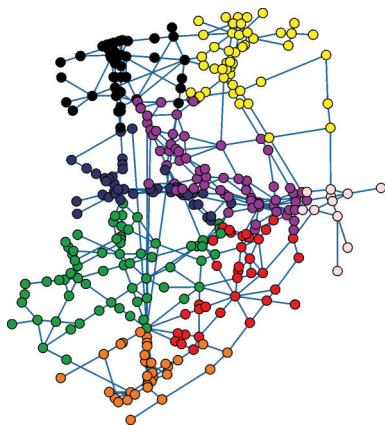
Fig.5　Community division graph of FN algorithm

Table 1　Result of community division (FN)

Serial number(Color)	Total number of nodes
1(Blue)	21
2(Green)	13
3(Yellow)	30
4(Orange)	29
5(Red)	50
6(Pink)	38
7(Magenta)	13
8(Purple)	46
9(Gray)	12
10(Brown)	52
11(Black)	48

Fig.6　Community division graph of FC-FN algorithm

Table 2　Result of community division (FC-FN)

Serial number(Color)	Total number of nodes
1(Blue)	46
2(Green)	66
3(Yellow)	45
4(Orange)	34
5(Red)	36
6(Pink)	11
7(Black)	45
8(Purple)	69

According to the results of community division, communities with more than or equal to 45 internal nodes are selected as large communities.

As shown in Table 1, the number of nodes within each community divided by FN algorithm is relatively small. Large communities have communities of Nos.5, 8, 10 and 11, accounting for 36.36% of the total number of communities, with a total of 196 nodes, accounting for about 55.6% of the total number of nodes. As shown in Table 2, the FC-FN algorithm divides a large number of nodes within each community, among which large communities have communities of Nos.1, 2, 3, 7 and 8, accounting for about 62.5% of the total number of communities, and the total number of nodes is 271, accounting for about 76.9% of the total number of nodes. The data comparison is shown in Table 3.

Table 3　Comparison of large organizations

Algorithm	Serial number of large communities	Total number of nodes	Proportion of nodes/ %
FN	5, 8, 10, 11	196	55.6
FC-FN	1, 2, 3, 7, 8	271	76.9

3.2.3　Analysis of modularity function

Calculate the Q value of the modularity function of the route network in East China under the FN algorithm and FC-FN algorithm respectively. The results are shown in Fig.7. When the FC-FN algorithm guarantees substantially the same number of iterations, the modularity of the air route network in East China is increased from 0.756 to 0.806, which is closer to 1. It shows that the community structure divided under the FC-FN algorithm is stronger than that of under the FN algorithm.
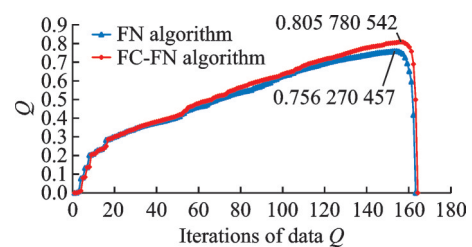

Fig.7　Schematic diagram of community structure

4 Conclusions

Based on the static topology of the route network, this paper proposes the concept of traffic flow contribution, and considers the impact of traffic flow on the network community division from a dynamic perspective. This method also introduces a solution to the problem of "balance between supply and demand", classifies the flow of nodes, and increases the number of "virtual nodes" for nodes with large flows, optimizes the network adjacency matrix, and forms an improved FC-FN algorithm. The steps of the algorithm and the optimization formula are given. The FC-FN algorithm ensures the efficiency of the operation, and the modularity value of the community is increased from 0.756 to 0.806 (the value of modularity is in the range $[-0.5,1)$).

The FC-FN algorithm has higher requirements on the accuracy and rationality of traffic flow contribution. Different traffic flow conditions will have different effects on traffic flow contribution and get different community divisions, so it can provide theoretical reference for flight season change and air route resource adjustment.

Future work is worth exploring in the following areas：

(1) Data aspect：Traffic flow data has an important impact on the calculation of traffic flow contribution. The method of judging the validity of data can be used as the research direction.

(2) Algorithm aspect：The number of iterations of the FC-FN algorithm is basically consistent with FN algorithm. Methods to reduce the number of iterations, improve the operation efficiency and the processing ability of the algorithm for massive data can be used as the research direction.

References

[1] CAI K, ZHANG J, DU W, et al. Analysis of the chinese air route network as a complex network[J]. Chinese Physics B,2012,21(2)：608-614.

[2] GIRVAN M, NEWMAN E. Community structure in social and biological networks[J]. Proceedings of the National Academy of Science of the Nanjed States of America, 2002, 99(12)：7821-7826.

[3] NEWMAN M. Fast algorithm for detecting communi-ty structure in networks[J]. Physical Review E, 2004, 69(6)：066133.

[4] YANG J, MACAULEY J, LESKOVEC J. Community detection in networks with node attributes[C]// Proceedings of the 13th International Conference on Data Mining. Dallas：IEEE,2013：1151-1156.

[5] SU C, GUAN X, DU Y, et al. A fast multi-level al-gorithm for community detection in directed online social networks[J]. Journal of Information Science, 2018,44(3)：392-407.

[6] LI S, XU X, WANG F, et al. Topological structure of US flight network based on complex network theory[J]. Transactions of Nanjing University of Aeronautics and Astronautics,2015,32(5)：555-559.

[7] NEWMAN M, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2)：026113.

[8] FORTUNATO S, CASTELLANO C. Community structure in graphs[J]. Encyclopedia of Complexity and System Science, 2009. DOI: https://doi. org/10.1007/978-0-387-30440-3_76.

[9] STEINHAEUSER K, CHAWLA N. Identifying and evaluating community structure in complex networks[J]. Pattern Recognition Letters, 2010, 31(5)：413-421.

[10] NEWMAN M. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(23)：8577-8582.

[11] ZHAO Z, GUO Q, YU K, et al. Identifying influential nodes for the networks with community structure[J]. Physica A：Statistical Mechanics and Its Applications,2019,551：123893.

[12] JOSE E, RAMIREZ M, CLAUDIO M, et al. Quantifying the resilience of community structures in networks[J]. Reliability Engineering and System Safety, 2018,169：466-474.

[13] WANG Y, WANG J. Integrated reconfiguration of both supply and demand for evacuation planning[J]. Transportation Research Part E,2019,130：82-94.

[14] LEYS C, LEY C, KLEIN O, et al. Detecting outliers：Do not use standard deviation around the mean, use absolute deviation around the median[J]. Journal of Experimental Social Psychology, 2013, 49(4)：764-766.

Author　Dr. **ZHAO Zheng** is currently a master supervisor at Nanjing University of Aeronautics and Astronautics. His research interests focus on airspace and airport system assessment technology.

（Production Editor：WANG Jing）

一种航路网络社团划分的改进 FN 算法

赵　征，张赛文，徐礼鹏，胡　莉

（南京航空航天大学民航学院/飞行学院，南京211106，中国）

摘要：社团划分是研究复杂网络特性的重要方法，目前应用较普遍的fast-Newman算法只考虑了网络静态层面的拓扑结构划分，而忽视了动态交通流量需求，导致划分结果仅为结构化最优的社团结构。为了提高社团划分准确性，以航路网络静态拓扑结构为基础，提出了网络交通流贡献度的概念；同时，引入运筹学理念优化网络邻接矩阵，形成一种改进的社团划分算法。选取华东地区航路网络352个航路节点，928条航段进行算法对比实验验证。结果表明，大型社团包含节点数占比提升21.3%，模块度值由0.756提升至0.806，而模块度的取值范围为[−0.5,1)，因此，改进的算法对社团结构划分有更理想的效果。研究成果可以为航班时刻优化、航路资源合理利用提供理论和技术支撑。

关键词：航路网络；社团划分；拓扑结构；交通流量贡献度