Journal of Traffic and Transportation Engineering 9 (2021) 1-7
doi: 10.17265/2328-2142/2021.01.001

Probability Distribution of Edge in Adjacent Matrix of Aviation Network of China and Algorithm of Searching Non-overlap Community Structure Based on Complex Network

Cheng Xiangjun, Yang Fang and Wei Chong
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Abstract: In order to discover the probability distribution feature of edge in aviation network adjacent matrix of China and on the basis of this feature to establish an algorithm of searching non-overlap community structure in network to reveal the inner principle of complex network with the feature of small world in aspect of adjacent matrix and community structure, aviation network adjacent matrix of China was transformed according to the node rank and the matrix was arranged on the basis of ascending node rank with the center point as original point. Adjacent probability from the original point to extension around in approximate area was calculated. Through fitting probability distribution curve, power function of probability distribution of edge in adjacent matrix arranged by ascending node rank was found. According to the feature of adjacent probability distribution, deleting step by step with node rank ascending algorithm was set up to search non-overlap community structure in network and the flow chart of algorithm was given. A non-overlap community structure with 10 different scale communities in aviation network of China was found by the computer program written on the basis of this algorithm.

Key words: Air transportation, adjacent matrix, deleting step by step with node rank ascending algorithm, aviation network of China, network community structure.
Foundation: National Natural Science Foundation of China (71971017).

1. Introduction

Aviation network is typical complex network with small world characters [1, 2]. About certain nation's aviation network, there are some unknown features in the field of complex network. This paper faces to the aviation network of China through analyzing the passenger data [3] of civil aviation airlines among 203 airports in year 2015 to reveal the complex network feature. According to complex network theory, network system of airports and airlines of China was constructed with airports regarded as nodes and airline regarded as edges to study the probability distribution of node degree and clustering coefficient of aviation network of China. The probability distribution of edge between nodes in adjacent matrix of aviation network of China was studied through regression analysis and curve fitting. Power function of probability distribution of the quantity of edge in adjacent matrix divided into approximate area arranged by node rank ascending was found. According to the feature of rapid decline of adjacent probability with rising of node rank and the definition of network community, deleting step by step algorithm was set up to look for non-overlap community structure in network and the flow chart of algorithm was given. A non-overlap community structure with 10 different scale communities in aviation network of China was found by the computer program written on the basis of this algorithm.

Corresponding author: Yang Fang, Ph.D., lecturer, research fields: transportation management, traffic and transportation engineering.

2. Probability Distribution of Edge in Adjacent Matrix of Aviation Network of China

For network $G=(V,E)$, where $v_i \in V$ is the node of G. V is the set of node. E is the set of edge [2], $(v_i,v_j)\in E$. Matrix $A=(a_{i,j})_{n\times n}$ was constructed, where

$$a_{i,j}=\begin{cases}1,(v_i,v_j)\in E\\ 0, otherwise\end{cases} \quad (1)$$

Matrix A is called adjacent matrix of network G. The degree of v_i is defined as the quantity of edges connecting to this node. The rank of v_i is defined as the serial number of descending order by node degree [2].

2.1 Adjacent Matrix of Aviation Network of China

The adjacent matrix $A=(a_{i,j})_{203\times 203}$ of aviation network of China in year 2015 was gotten by dealing with statistic data [3]. For intuitively understanding the distribution features of matrix A, A was transformed into $A^{'}=(a^{'}_{i,j})_{203\times 203}$: the center as the origin point, the node was arranged by ascending order of rank from the origin point to outside; in the horizontal direction, the nodes with odd number rank were on the left side of the origin point, the nodes with even number rank were on the right side of the origin point; in the vertical direction, the nodes with odd number rank were above the origin point, the nodes with even number rank were below the origin point.

$$\text{Let } a^{'}_{i,j}=\begin{cases}\bullet, a_{i,j}=1\\ \square, a_{i,j}=0\\ \Delta, i=j\end{cases} \quad (2)$$

The matrix $A^{'}$ was printed in Fig. 1 to show the diagram of adjacent matrix with ascending node rank of aviation network of China.

In order to research the adjacent probability in different area of adjacent matrix, the adjacent matrix was divided into approximate zones from the origin point to outside. If the value of each $a^{'}_{i,j}$ in matrix $A^{'}$ occupied one area unit, then the same area in zoning means the same possible adjacent edges in adjacent matrix showing in zoning diagram of Fig. 2. Let the side length of square around the origin point be x_1, the side length of square successively outside be $x_2,\ldots,x_{m-1},x_m$. If $x_m^2-x_{m-1}^2=x_{m-1}^2-x_{m-2}^2=\ldots=x_2^2-x_1^2=x_1^2$, then

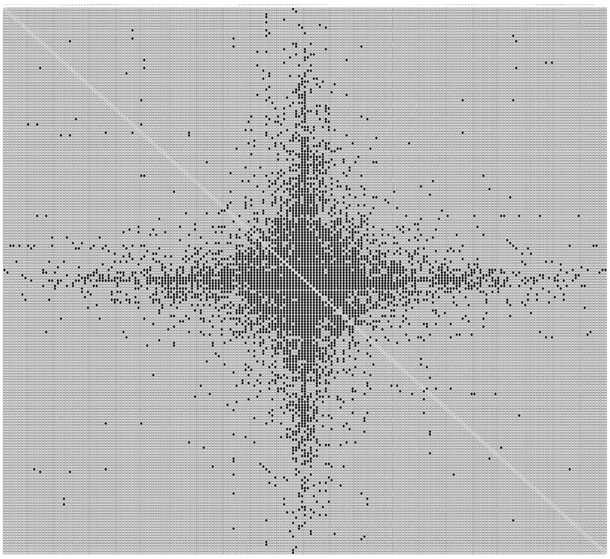

Fig. 1 Diagram of adjacent matrix with ascending node rank of aviation network of China.

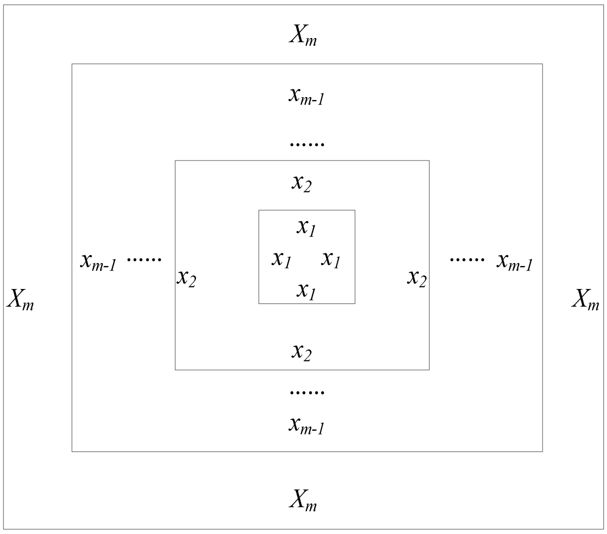

Fig. 2 Zoning diagram of adjacent matrix with ascending node rank of aviation network of China.

the annular region has the same area. That is $x_2=\sqrt{2}x_1, x_3=\sqrt{3}x_1, \ldots, x_m=\sqrt{m}x_1$.

Since $x_m=203$, let $m=10$, then $x_1=\frac{x_m}{\sqrt{m}}=\frac{203}{\sqrt{10}}=64.19\approx 64$.

2.2 Probability Distribution of Edge in Adjacent Matrix

The diagram of adjacent probability in adjacent matrix interval of aviation network of China was drawn in Fig. 3 using adjacent probability in Table 1. Let the serial number of interval be x and probability be y. Let $u=\ln x$, $v=\ln y$. Take u as abscissa and v as ordinate to draw associated scatter diagram in Fig. 4. The correlation coefficient r of scattered points in Fig. 4 was calculated by Eq. (3) [4].

$$r=\frac{L_{uv}}{\sqrt{L_{uu}L_{vv}}}=\frac{\sum_{i=1}^{n}(u_i-\bar{u})(v_i-\bar{v})}{\sqrt{\sum_{i=1}^{n}(u_i-\bar{u})^2\sum_{i=1}^{n}(v_i-\bar{v})^2}} \tag{3}$$

Here, $n=10$. Using the data in Table 1, the value of correlation coefficient r was calculated, $r=-0.99$. The critical value of $r_{\big|{\alpha=1\% \atop f=8}}$ was 0.765 found in critical value table [4] at degree of freedom $f=n-2=8$ and level of significant α of 1%. Since $|r|=0.99>0.765=r_{\big|{\alpha=1\% \atop f=8}}$, the scattered points in Fig. 4 have significant linear correlation. Least square method [4] was used as an approach in Eq. (4) to fit the line with points in Fig. 4.

$$\begin{cases}\hat{\beta}_0=\bar{v}-\hat{\beta}_1\bar{u}=-0.43\\ \hat{\beta}_1=\frac{L_{uv}}{L_{uu}}=-1.864\end{cases} \tag{4}$$

The linear equation:

$$v=-0.43-1.864u \tag{5}$$

The point of fitting line (5) was drawn with the sample points in one diagram of Fig. 5.

To take t test [4] of Eq. (5), test hypothesis is: $H_0:\beta_1=0$. When the hypothesis is true, there is:

$$\hat{\beta}_1\sim N(0,\frac{\sigma^2}{L_{uu}}) \tag{6}$$

Here, $\hat{\beta}_1$ fluctuates near zero, statistic t is build:

$$t=\frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{L_{uu}}}}=\frac{\hat{\beta}_1\sqrt{L_{uu}}}{\hat{\sigma}} \tag{7}$$

Wherein:

$$\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}(v_i-\hat{v}_i)^2 \tag{8}$$

Statistic t was calculated by data: $t=-20.7$

To check the t distribution table [4], at significant level α of 0.01 and degree of freedom $f=n-2=8$ the value of $t_{\big|{\alpha=0.01 \atop f=8}}$ in table is 2.896.

So, $|t|=20.7>2.896=t_{\big|{\alpha=0.01 \atop f=8}}$, null hypothesis H_0 is refused. The linear correlation of Eq. (5) is significant.

The fitting curve Eq. (9) of probability distribution of node adjacent probability was derived from Eq. (5).

$$y=0.651x^{-1.864} \tag{9}$$

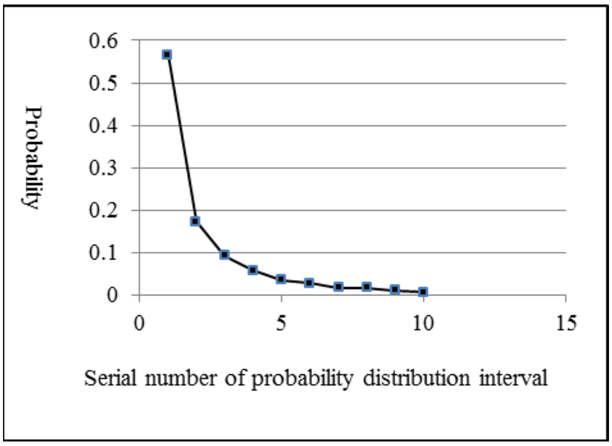

Fig. 3 Adjacent probability distribution diagram of adjacent matrix of aviation network of China.

Table 1 Adjacent probability in adjacent matrix interval with ascending node rank of aviation network of China.

m	1	2	3	4	5
Theory value of interval with ascending node rank	$[1, x_1)$	$[x_1, \sqrt{2}x_1)$	$[\sqrt{2}x_1, \sqrt{3}x_1)$	$[\sqrt{3}x_1, 2x_1)$	$[2x_1, \sqrt{5}x_1)$
Integer value of interval with ascending node rank	[1, 64)	[64, 91)	[91, 111)	[111, 128)	[128, 143)
Quantity of adjacent edge in adjacent matrix interval	2,178	664	352	222	136
Adjacent probability in adjacent matrix interval	0.56601	0.17256	0.09148	0.05769	0.03534
m	6	7	8	9	10
Theory value of interval with ascending node rank	$[\sqrt{5}x_1, \sqrt{6}x_1)$	$[\sqrt{6}x_1, \sqrt{7}x_1)$	$[\sqrt{7}x_1, \sqrt{8}x_1)$	$[\sqrt{8}x_1, 3x_1)$	$[3x_1, \sqrt{10}x_1]$
Integer value of interval with ascending node rank	[143, 157)	[157, 169)	[169, 181)	[181, 192)	[192, 203]
Quantity of adjacent edge in adjacent matrix interval	104	66	62	42	22
Adjacent probability in adjacent matrix interval	0.02703	0.01715	0.01611	0.01091	0.00572

Fitting points of curve Eq. (9) with sample points in Fig. 6 show a good fitting effect. It illustrates that the probability distribution of edge in adjacent matrix of aviation network of China is power function curve.

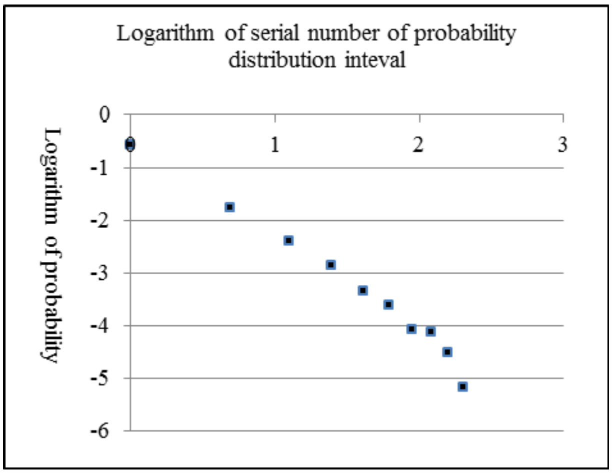

Fig. 4 Adjacent probability logarithm diagram of adjacent matrix of aviation network of China.

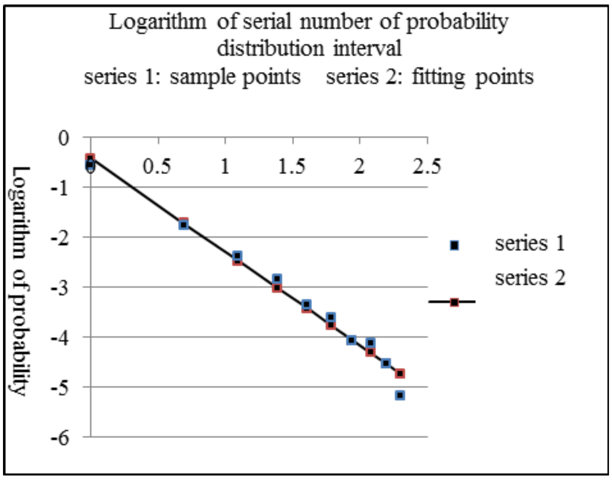

Fig. 5 Effect diagram for fitting line of logarithm adjacent probability.

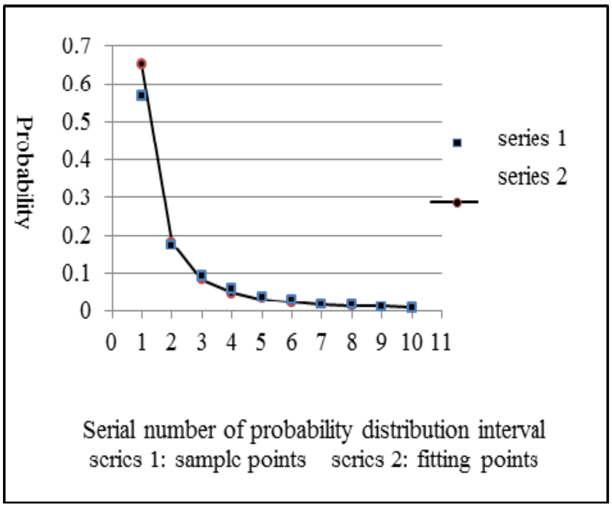

Fig. 6 Effect diagram for fitting curve of adjacent probability.

3. Non-overlap Community Structure of Aviation Network of China

The strictest definition of community structure of complex network was given by Palla et al. [5]: community structure was also called clique, a clique was a fully connected sub-graph consisting of three or more than three modes, in which any two nodes had connection. Based on this definition of community structure, overlap among communities is permitted. Overlap means that single node may belong to not only one community but also several communities and these different communities are connected with these overlap nodes.

Observing Fig. 1, it was found that there are dense connection zone with more spars connection zones. Community in aviation network will be found according to the above definition of clique. The finding of fully connected sub-graph in aviation network will contribute to discovering the structure of aviation network. But this definition of community will lead to the possibility of overlap community structure. The overlap community structure of aviation network will disturb the function analysis of full connected airport group. So, based on the definition of clique, an algorithm of searching non-overlap community structure is needed to find the maximum scale fully connected community structure of aviation network of China.

3.1 Deleting Step by Step with Node Rank Ascending Algorithm to Search Non-overlap Community Structure

The adjacent probability between nodes declined rapidly with ascending order of node rank in aviation network of China according to Fig. 1 and Eq. (9). The edges were dense near the origin point. The node with less rank had more probability to be in clique. On the basis of this feature, an algorithm for searching community structure from the node with smaller rank to bigger ones could be constructed. For non-overlap among each community, all the nodes of one community would be deleted in network after finding of this community according to the definition. Then the other cliques would be searched by the definition till all cliques are found. Here the algorithm for searching non-overlap community structure through ascending order of node rank by deleting all nodes of clique after finding the clique till all clique be found was constructed. It was called deleting step by step with node rank ascending algorithm. The flow chart of this algorithm was shown in Fig. 7.

3.2 Non-overlap Community Structure of Aviation Network of China

Computer program was written in C language according to Fig. 7. Adjacent matrix A of aviation network of China in year 2015 was put into program as original data to run in result of community structure of aviation network of China with 10 different scale non-overlap cliques in Table 2.

Table 2 Non-overlap community structure of aviation network of China.

No.	Serial number in community	City with airport	Rank of node	No.	Serial number in community	City with airport	Rank of node	No.	Serial number in community	City with airport	Rank of node
C-1	1	Beijing	1	C-2	1	Ha'erbin	14	C-5	2	Xi'ning	39
	2	Shanghai	2		2	Dalian	15		3	Zhuhai	40
	3	Guangzhou	3		3	Zhengzhou	16	C-6	1	Changchun	33
	4	Xi'an	4		4	Nanjing	18		2	Ningbo	34
	5	Chongqing	5		5	Shenyang	20		3	Haila'er	43
	6	Chengdu	6		6	Huhehaote	24		4	E'erduosi	46
	7	Shenzhen	7		7	Shijiazhuang	25	C-7	1	Yantai	38
	8	Kunming	8		8	Hefei	35		2	Baishan	71
	9	Hangzhou	9		9	Nanchang	36		3	Yanji	102
	10	Haikou	10	C-3	1	Sanya	22	C-8	1	Lijiang	41
	11	Tianjin	11		2	Guilin	28		2	Lasa	48
	12	Changsha	12		3	Ji'nan	30		3	Yibin	69
	13	Xiamen	13		4	Wenzhou	37	C-9	1	Jieyangchaoshan	47
	14	Qingdao	17	C-4	1	Wulumuqi	26		2	Liuzhou	54
	15	Guiyang	19		2	Lanzhou	29		3	Zhanjiang	63
	16	Wuhan	21		3	Yinchuan	32	C-10	1	Hami	107
	17	Taiyuan	23		4	Mianyang	44		2	Ku'erle	110
	18	Fuzhou	27		1	Nanning	31		3	Kelamayi	134

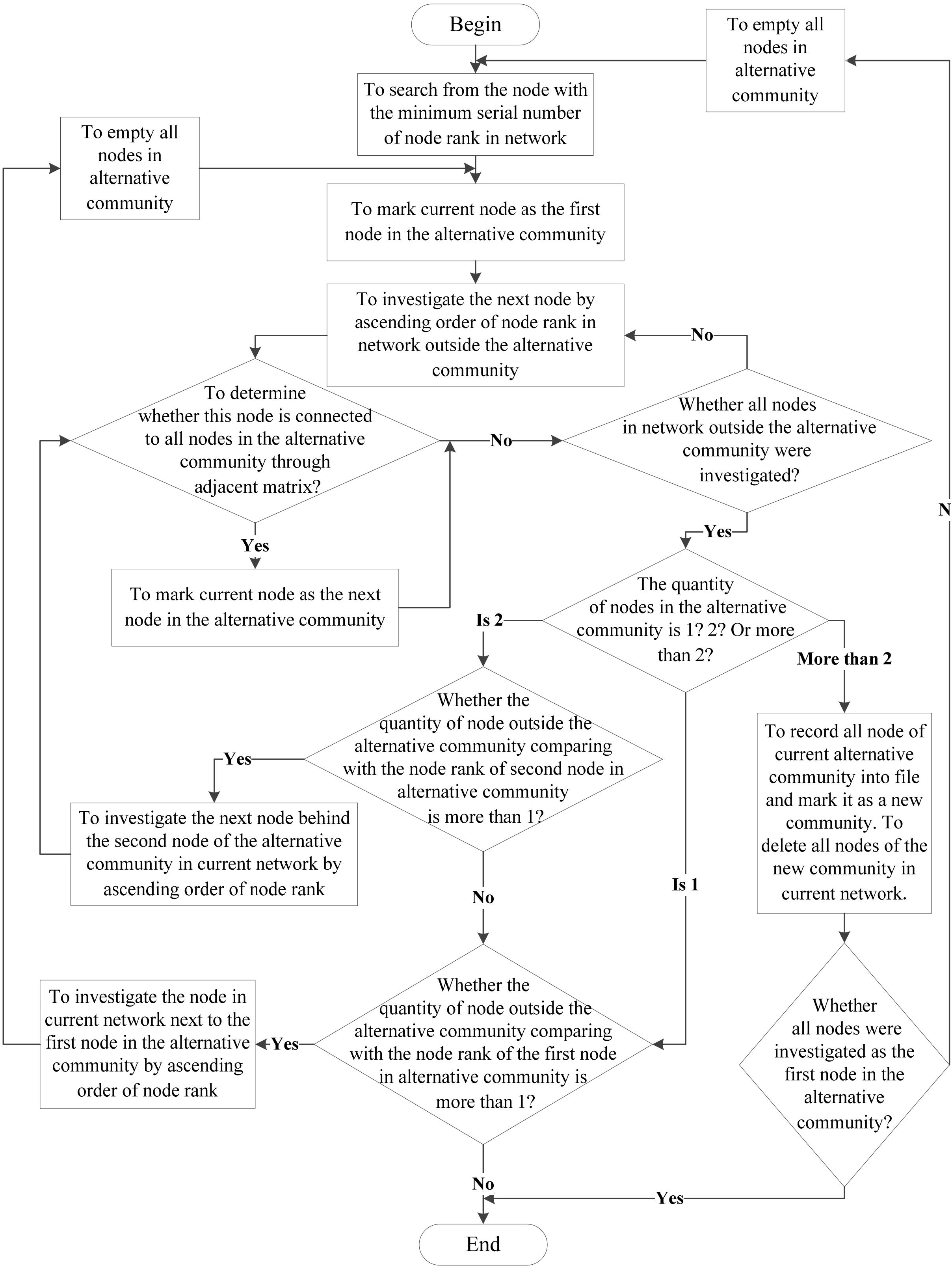

Fig. 7 Flow chart of deleting step by step with node rank ascending algorithm for searching non-overlap community structure.

4. Conclusion

According to statistic data, the basic data among 203 airports of civil aviation airlines of China in year 2015 were gotten. On the basis of complex network theory, network system of airports and airlines of China was constructed with airports regarded as nodes and airline regarded as edges to study the probability distribution of node degree and clustering coefficient of aviation network of China. The probability distribution of edge among nodes in adjacent matrix of aviation network of China was studied through regression analysis and curve fitting. Power function of probability distribution of the quantity of edge in adjacent matrix divided into approximate area arranged by node rank ascending was found. According to the feature of rapid decline of adjacent probability with rising of node rank and the definition of network community, deleting step by step with node rank ascending algorithm was set up to search non-overlap community structure in network and the flow chart of algorithm was given. A non-overlap community structure with 10 different scale communities in aviation network of China was found by the computer program written on the basis of this algorithm.

Reference

[1] Rocha, L. E. C. 2017. "Dynamics of Air Transport Networks: A Review from a Complex Systems Perspective." *Chinese Journal of Aeronautics* 30 (2): 469-78.

[2] Sun, X. J., and Si, S. K. 2016. *Complex Network Algorithms and Applications*. National Defence Industry Press.

[3] Division of Development and Planning of Civil Aviation Administration of China. 2017. *Civil Aviation Statistical Yearbook of China*. Civil Aviation Press of China.

[4] He, X. Q., and Liu, W. Q. 2006. *Application of Regression Analysis*. China Renmin University Press.

[5] Derenyi, I., Palla, G., and Vicsek, T. 2005. "Clique Percolation in Random Networks." *Physics Review Letter* 94 (16): 160202-4.