Journal of Traffic and Transportation Engineering 9 (2021) 41-52
doi: 10.17265/2328-2142/2021.02.001

Probability Distribution of Average Length of Node Path and Its Evolution Trace of Aviation Network of China Based on Complex Network

Cheng Xiangjun, Yang Fang and Li Tao

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Abstract: In order to reveal the complex network feature and its evolution principle of aviation network of China, probability distribution and its evolution trace of average length of node path of aviation network of China were researched according to statistics data in years 1988, 1994, 2001, 2008 and 2015 of civil aviation of China. Floyd algorithm to calculate the path length between any two nodes of network was applied and average length of node path of aviation network was obtained according to this algorithm. It was discovered that average length of node path to other nodes had normal distribution function in each year. At meantime, the location parameter and scale parameter of normal distribution function had linear evolution trace. Airline rate was an index to describe the density of airline. It was found that average length of node path of aviation network of China evolved synchronously with airline rate and they had linear relationship.

Key words: Aviation network of China, average length of node path, probability distribution, evolution trace, airline rate.

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 in years 1988, 1994, 2001, 2008 and 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 average length of node path and its evolution trace of aviation network of China. Floyd algorithm [4] to calculate the path length between any two nodes of network was applied and average length of node path of aviation network was obtained according to this algorithm. Although the scale of aviation network of China was developed each year with increasing of airport amount and airline amount, but it was discovered that average length of node path to other nodes had normal distribution function in each year. At meantime, the location parameter and scale parameter of normal distribution function had linear evolution trace. It means that the probability distribution evolved along the linear trace and had positive correlation. Airline rate was an index to describe the density of airline. It was found that average length of node path of aviation network of China evolved synchronously with airline rate and they had linear relationship.

2. Probability Distribution of Average Length of Node Path 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 [4], $(v_i, v_j) \in E$. 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 length of path $d_{i,k}$ between node v_i and v_k

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

is defined as the quantity of edges of shortest path connecting these two nodes. The rank of v_i is defined as the serial number of descending order by node degree [4].

The path length matrix $D=(d_{i,j})_{n\times n}$ was obtained from adjacent matrix A by algorithm Floyd [4].

$$d_{i,j}=\begin{cases}\Omega_{i,j}, \text{the quantity of edges of shortest path between node } v_i \text{ and } v_j\\ 0, i=j\\ \infty,\ v_i \textit{ and } v_j \textit{ was not connected}\end{cases} \quad (2)$$

Since the aviation network of China was fully connected network in each year [3], so the path length matrix D could be simplified as the following:

$$d_{i,j}=\begin{cases}\Omega_{i,j}, \text{the quantity of edges of shortest path between node } v_i \text{ and } v_j\\ 0, i=j\end{cases} \quad (3)$$

According to the definition, the average path length $\overline{d}_i$ of node v_i to other nodes is:

$$\overline{d}_i=\frac{1}{n-1}\sum_{j=1}^{n}\Omega_{i,j} \quad (4)$$

The average path length $\overline{d}$ of network is:

$$\overline{d}=\frac{1}{n}\sum_{i=1}^{n}\overline{d}_i=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{i,j} \quad (5)$$

The data in Table 1 were calculated from statistics data [3]. The quantity of city with airport was the node number N. The quantity of airlines among cities was the edges number L of network.

2.1 Probability Distribution of Average Length of Node Path of Aviation Network of China in 1988

The adjacent matrix A_{1988} was obtained from statistic data [3]. The path length matrix D_{1988} was obtained from A_{1988} by algorithm Floyd [4]. The average path length of 85 nodes was calculated through D_{1988} and Eq. (4). The data in Table 2 were calculated by statistics data. The curve in Fig. 1 was drawn by the data in Table 2. It was similar to normal distribution function by observation. Let the quantity of nodes be N, the quantity of interval be m, the spacing of interval i be δ_i, the times of appearance in interval i be k_i, the frequency of appearance in interval i be p_i. Here $N_{1988}=85$, $m_{1988}=5$. It could be known from the probability density function of normal distribution in Table 2 and Fig. 1:

$$q_i=\frac{k_i}{m_{1988}},\quad \sum_{i=1}^{m_{1988}}q_i=1 \quad (6)$$

$$q_i=p_i\delta_i \Rightarrow p_i=\frac{q_i}{\delta_i},\quad \sum_{i=1}^{m_{1988}}p_i=1 \quad (7)$$

The normal distribution function is $N(\mu,\sigma)$, μ is the location parameter, σ is the scale parameter, its probability density function is $f(x)$.

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (8)$$

It could be obtained from the statistical properties of normal distribution [5]:

$$\hat{\mu}_{1988}=\frac{1}{N_{1988}}\sum_{i=1}^{N_{1988}}x_i=2.592 \quad (9)$$

$$\hat{\sigma}_{1988}=\sqrt{\frac{1}{N_{1988}-1}\sum_{i=1}^{N_{1988}}(x_i-\overline{x})^2}=0.493 \quad (10)$$

Let $\hat{\mu}_{1988}$ and $\hat{\sigma}_{1988}$ be location parameter and scale parameter to substitute in Eq. (8) respectively, the equation of fitting curve could be obtained:

$$\hat{y}_{1988}=0.81e^{-2.058(x-2.592)^2} \quad (11)$$

Table 1 The information of airport and airline of aviation network of China.

Year	1988	1994	2001	2008	2015
The quantity of city with airport	85	122	130	150	203
The quantity of airline among cities	265	589	730	940	1,924

Table 2 Probability distribution of average path length of node of aviation network of China in 1988.

Interval	[1.5~2]	(2~2.5]	(2.5~3]	(3~3.5]	(3.5~4]
Spacing	0.5	0.5	0.5	0.5	0.5
Median	1.75	2.25	2.75	3.25	3.75
Times	6	29	34	13	3
Frequency	0.0706	0.3412	0.4	0.1529	0.0354
Probability	0.141	0.682	0.8	0.306	0.071

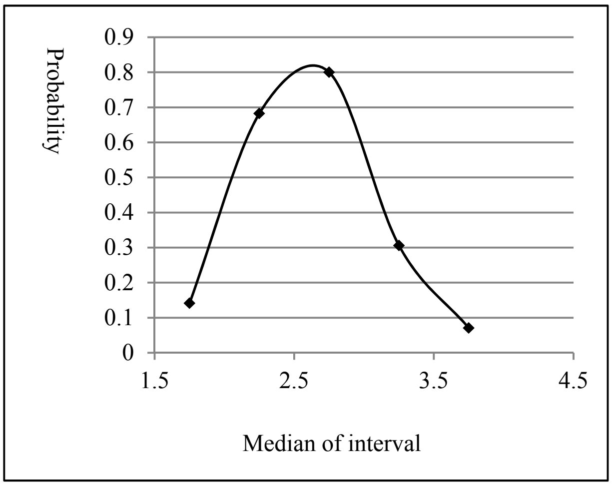

Fig. 1 Probability distribution diagram of average path length of node of aviation network of China in 1988.

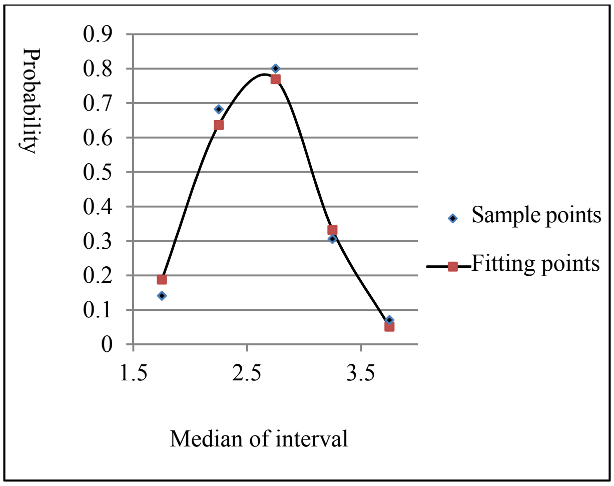

Fig. 2 Fitting effect diagram of probability distribution of average path length of node in 1988.

The fitting curve (Eq. (11)) and sample points $p_{i,1988}$ were drawn in Fig. 2. The effect of fitting was good in Fig. 2. The $\varepsilon_{i,1988}$ ($i = 1, 2, 3, 4, 5$) in equation (12) was calculated as residual to investigate the degree of approximation of fitting through the residual distribution [6].

$$\varepsilon_{i,1988} = p_{i,1988} - \hat{y}_{i,1988} \tag{12}$$

The residual $\varepsilon_{i,1988}$ was drawn in Fig. 3. All points in Fig. 3 were random fluctuation around $\varepsilon = 0$ and the range of fluctuation was very small. $|\varepsilon_{i,1988}| < 0.05 < \hat{\sigma}_{1988} = 0.493$. This illustrated that the fitting effect was very good.

2.2 Probability Distribution of Average Length of Node Path of Aviation Network of China in 1994

The adjacent matrix A_{1994} was obtained from statistic data [3]. The path length matrix D_{1994} was obtained from A_{1994} by algorithm Floyd [4]. The

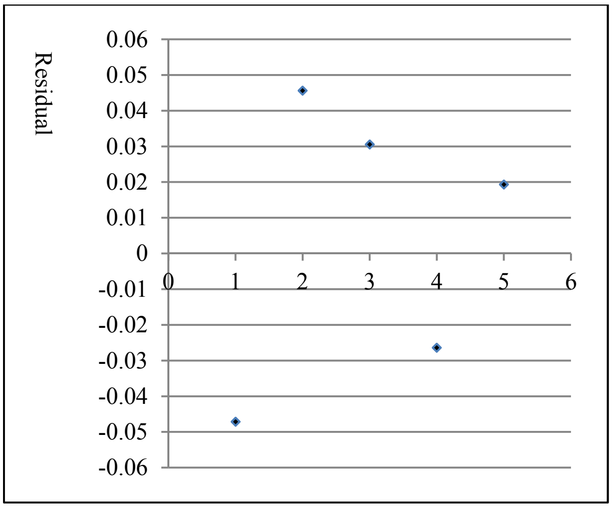

Fig. 3 Residual diagram of fitting curve of probability distribution of average path length of node in 1988.

average path length of 122 nodes was calculated through D_{1988} and Eq. (4). The data in Table 3 were calculated by statistics data. The curve in Fig. 4 was drawn by the data in Table 3. It was similar to normal distribution function by observation. The definition and relationship were same as Eqs. (6) and (7) in Eq. (1). Here $N_{1994}=122$, $m_{1994}=5$. It was calculated from the statistics data:

$$\hat{\mu}_{1994}=\frac{1}{N_{1994}}\sum_{i=1}^{N_{1994}}x_i=2.356 \quad (13)$$

$$\hat{\sigma}_{1994}=\sqrt{\frac{1}{N_{1994}-1}\sum_{i=1}^{N_{1994}}(x_i-\overline{x})^2}=0.395 \quad (14)$$

Let $\hat{\mu}_{1994}$ and $\hat{\sigma}_{1994}$ be location parameter and scale parameter to substitute in Eq. (8) respectively, the equation of fitting curve could be obtained:

$$\hat{y}_{1994}=1.01e^{-3.205(x-2.356)^2} \quad (15)$$

The fitting curve (Eq. (15)) and sample points $p_{i,1994}$ were drawn in Fig. 5. The effect of fitting was good in Fig. 5. The $\varepsilon_{i,1994}$ ($i=1,2,3,4,5$) in equation (16) was calculated as residual to investigate the degree of approximation of fitting through the residual distribution [6].

$$\varepsilon_{i,1994}=p_{i,1994}-\hat{y}_{i,1994} \quad (16)$$

The residual $\varepsilon_{i,1994}$ was drawn in Fig. 6. All points in Fig. 6 were random fluctuation around $\varepsilon=0$ and the range of fluctuation was very small. $|\varepsilon_{i,1994}|<0.05<\hat{\sigma}_{1994}=0.395$. This illustrated that the fitting effect was very good.

2.3 Probability Distribution of Average Length of Node Path of Aviation Network of China in 2001

The adjacent matrix A_{2001} was obtained from statistic data [3]. The path length matrix D_{2001} was obtained from A_{2001} by algorithm Floyd [4]. The average path length of 130 nodes was calculated through D_{2001} and Eq. (4). The data in Table 4 were calculated by statistics data. The curve in Fig. 7 was drawn by the data in Table 4. It was similar to normal

Table 3 Probability distribution of average path length of node of aviation network of China in 1994.

Interval	[1.1~1.6]	(1.6~2.1]	(2.1~2.6]	(2.6~3.1]	(3.1~3.6]
Spacing	0.5	0.5	0.5	0.5	0.5
Median	1.35	1.85	2.35	2.85	3.35
Times	3	29	59	27	4
Frequency	0.0246	0.2377	0.4836	0.2213	0.0328
Probability	0.049	0.475	0.967	0.443	0.066

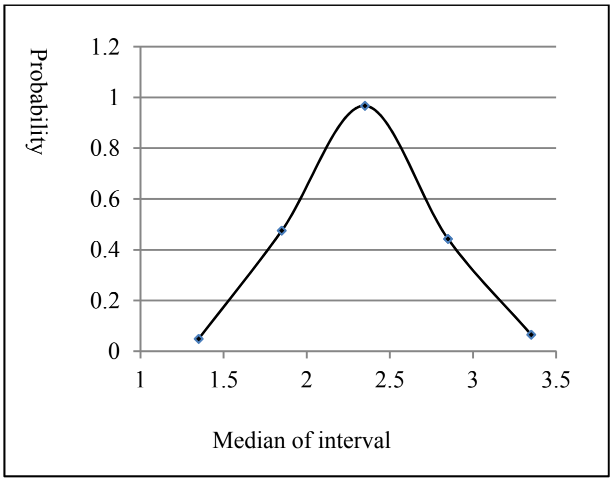

Fig. 4 Probability distribution diagram of average path length of node of aviation network of China in 1994.

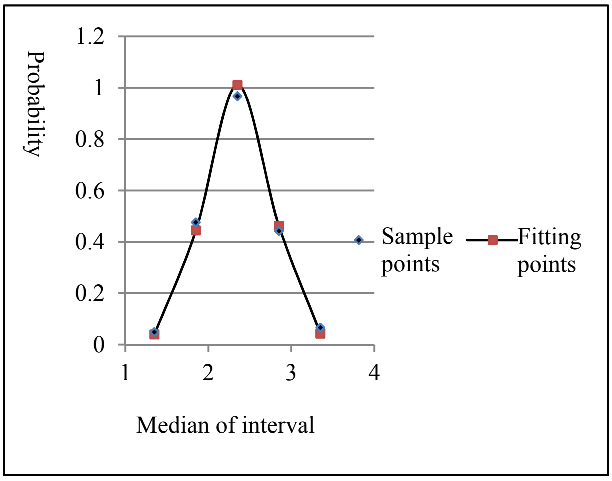

Fig. 5 Fitting effect diagram of probability distribution of average path length of node in 1994.

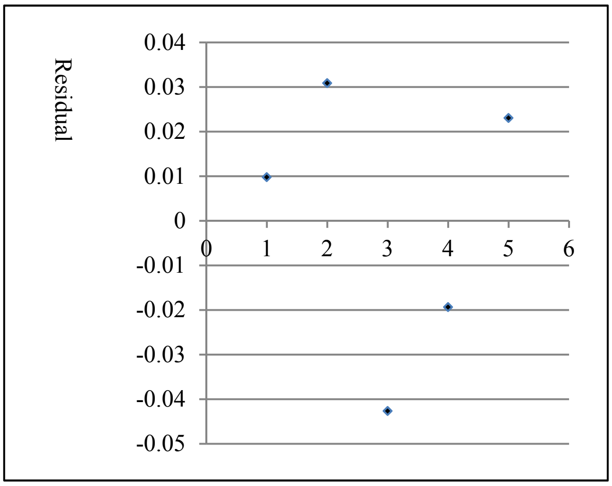

Fig. 6 Residual diagram of fitting curve of probability distribution of average path length of node in 1994.

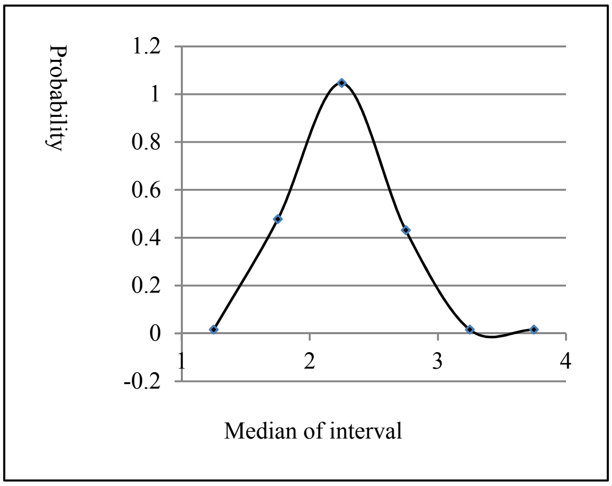

Fig. 7 Probability distribution diagram of average path length of node of aviation network of China in 2001.

Table 4 Probability distribution of average path length of node of aviation network of China in 2001

Interval	[1~1.5]	(1.5~2]	(2~2.5]	(2.5~3]	(3~3.5]	(3.5~4]
Spacing	0.5	0.5	0.5	0.5	0.5	0.5
Median	1.25	1.75	2.25	2.75	3.25	3.75
Times	1	31	68	28	1	1
Frequency	0.0077	0.2385	0.5231	0.2153	0.0077	0.0077
Probability	0.015	0.477	1.046	0.432	0.015	0.015

distribution function by observation. The definition and relationship were same as Eqs. (6) and (7) in Eq. (1). Here $N_{2001}=130$, $m_{2001}=6$. It was calculated from the statistics data:

$$\hat{\mu}_{2001}=\frac{1}{N_{2001}}\sum_{i=1}^{N_{2001}}x_i=2.27 \tag{17}$$

$$\hat{\sigma}_{2001}=\sqrt{\frac{1}{N_{2001}-1}\sum_{i=1}^{N_{2001}}(x_i-\overline{x})^2}=0.379 \tag{18}$$

Let $\hat{\mu}_{2001}$ and $\hat{\sigma}_{2001}$ be location parameter and scale parameter to substitute in Eq. (8) respectively, the equation of fitting curve could be obtained:

$$\hat{y}_{2001}=1.053e^{-3.48(x-2.27)^2} \tag{19}$$

The fitting curve (Eq. (19)) and sample points $p_{i,2001}$ were drawn in Fig. 8. The effect of fitting was good in Fig. 8. The $\varepsilon_{i,2001}$ ($i=1,2,3,4,5,6$) in Equation (20) was calculated as residual to investigate the degree of approximation of fitting through the residual distribution [6].

$$\varepsilon_{i,2001}=p_{i,2001}-\hat{y}_{i,2001} \tag{20}$$

The residual $\varepsilon_{i,2001}$ was drawn in Fig. 9. All points in Fig. 9 were random fluctuation around $\varepsilon=0$ and the range of fluctuation was very small. $|\varepsilon_{i,2001}|<0.07<\hat{\sigma}_{2001}=0.379$. This illustrated that the fitting effect was very good.

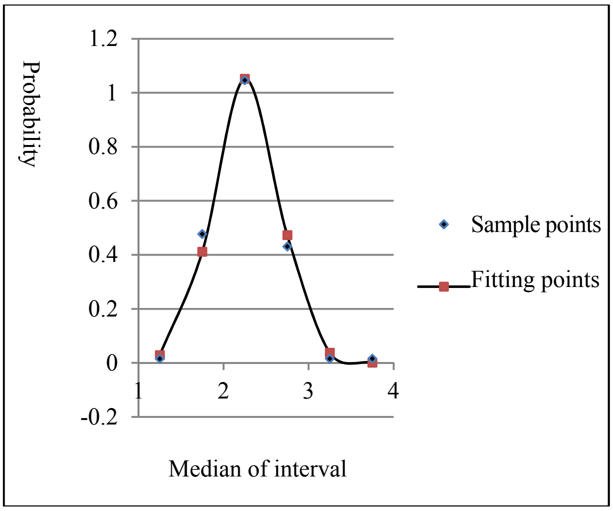

Fig. 8 Fitting effect diagram of probability distribution of average path length of node in 2001.

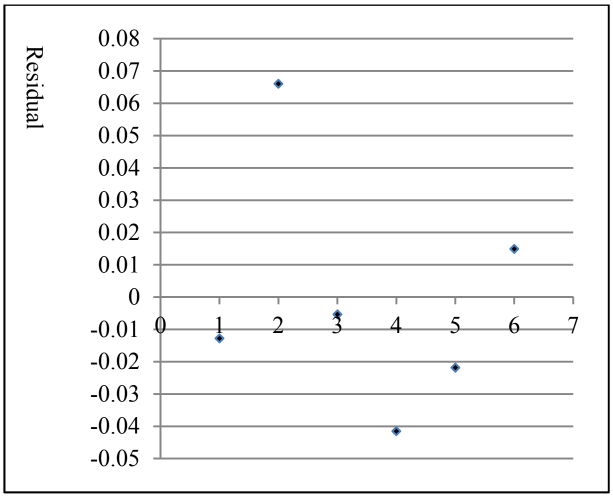

Fig. 9 Residual diagram of fitting curve of probability distribution of average path length of node in 2001.

2.4 Probability Distribution of Average Length of Node Path of Aviation Network of China in 2008

The adjacent matrix A_{2008} was obtained from statistic data [3]. The path length matrix D_{2008} was obtained from A_{2008} by algorithm Floyd [4]. The average path length of 150 nodes was calculated through D_{2008} and Eq. (4). The data in Table 5 were calculated by statistics data. The curve in Fig. 10 was drawn by the data in Table 5. It was similar to normal distribution function by observation. The definition and relationship were same as Eqs. (6) and (7) in Eq. (1). Here $N_{2008}=150$, $m_{2008}=6$. It was calculated from the statistics data:

$$\hat{\mu}_{2008}=\frac{1}{N_{2008}}\sum_{i=1}^{N_{2008}}x_i=2.296 \tag{21}$$

$$\hat{\sigma}_{2008}=\sqrt{\frac{1}{N_{2008}-1}\sum_{i=1}^{N_{2008}}(x_i-\overline{x})^2}=0.379 \tag{22}$$

Let $\hat{\mu}_{2008}$ and $\hat{\sigma}_{2008}$ be location parameter and scale parameter to substitute in Eq. (8) respectively, the equation of fitting curve could be obtained:

$$\hat{y}_{2008}=1.053e^{-4.37(x-2.296)^2} \tag{23}$$

The fitting curve (Eq. (23)) and sample points $p_{i,2008}$ were drawn in Fig. 11. The effect of fitting was good in Fig. 11. The $\varepsilon_{i,2008}$ ($i=1,2,3,4,5,6$) in equation (24) was calculated as residual to investigate the degree of approximation of fitting through the residual distribution [6].

$$\varepsilon_{i,2008}=p_{i,2008}-\hat{y}_{i,2008} \tag{24}$$

The residual $\varepsilon_{i,2008}$ was drawn in Fig. 12. All points in Fig. 12 were random fluctuation around $\varepsilon=0$ and the range of fluctuation was very small. $|\varepsilon_{i,2008}|<0.16<\hat{\sigma}_{2008}=0.379$. This illustrated that the fitting effect was very good.

2.5 Probability Distribution of Average Length of Node Path of Aviation Network of China in 2015

The adjacent matrix A_{2015} was obtained from statistic data [3]. The path length matrix D_{2015} was obtained from A_{2015} by algorithm Floyd [4]. The average path length of 150 nodes was calculated through D_{2015} and Eq. (4). The data in Table 6 were calculated by statistics data. The curve in Fig. 13 was drawn by the data in Table 6. It was similar to normal distribution function by observation. The definition and relationship were same as Eqs. (6) and (7) in Eq. (1). Here $N_{2015}=203$, $m_{2015}=5$. It was calculated from the statistics data:

$$\hat{\mu}_{2015}=\frac{1}{N_{2015}}\sum_{i=1}^{N_{2015}}x_i=2.136 \tag{25}$$

Table 5 Probability distribution of average path length of node of aviation network of China in 2008.

Interval	[1~1.5]	(1.5~2]	(2~2.5]	(2.5~3]	(3~3.5]	(3.5~4]
Spacing	0.5	0.5	0.5	0.5	0.5	0.5
Median	1.25	1.75	2.25	2.75	3.25	3.75
Times	2	33	73	40	1	1
Frequency	0.0133	0.22	0.4866	0.2667	0.0067	0.0067
Probability	0.027	0.44	0.974	0.533	0.013	0.013

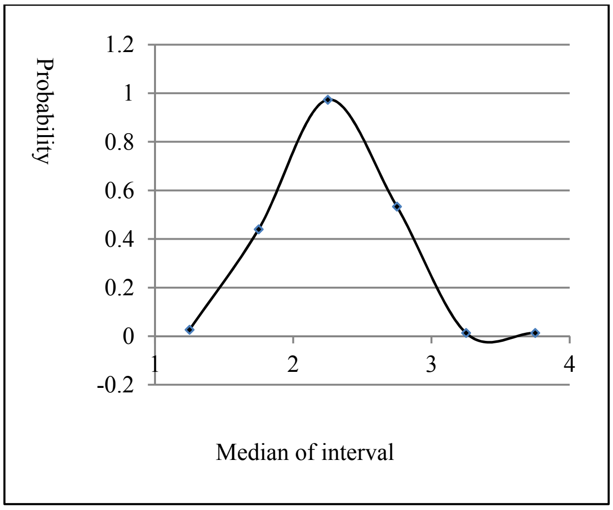

Fig. 10 Probability distribution diagram of average path length of node of aviation network of China in 2008.

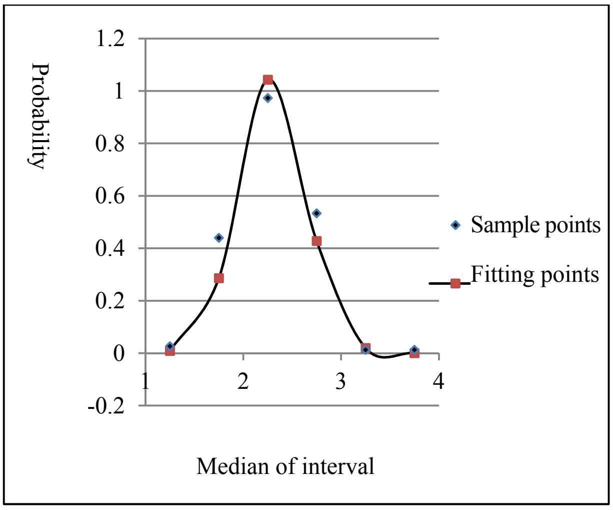

Fig. 11 Fitting effect diagram of probability distribution of average path length of node in 2008.

Table 6 Probability distribution of average path length of node of aviation network of China in 2015.

Interval	[0.9~1.4]	(1.4~1.9]	(1.9~2.4]	(2.4~2.9]	(2.9~3.4]
Spacing	0.5	0.5	0.5	0.5	0.5
Median	1.15	1.65	2.15	2.65	3.15
Times	1	38	132	29	3
Frequency	0.0049	0.1872	0.6502	0.1429	0.0148
Probability	0.01	0.374	1.3	0.286	0.03

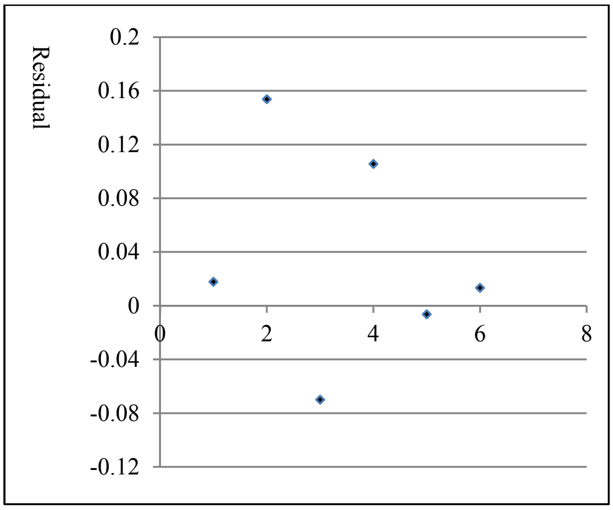

Fig. 12 Residual diagram of fitting curve of probability distribution of average path length of node in 2008.

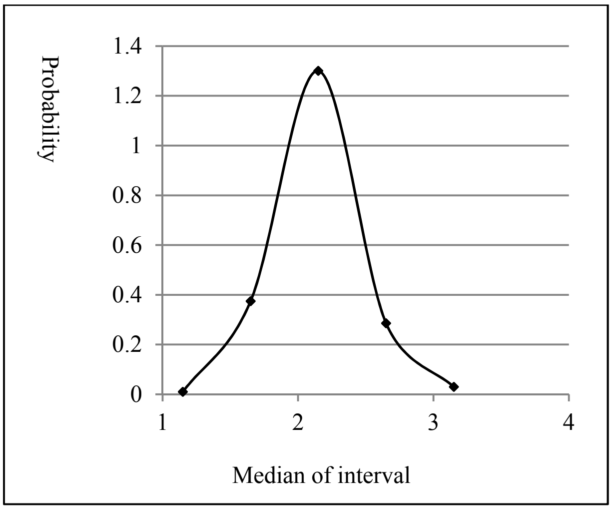

Fig. 13 Probability distribution diagram of average path length of node of aviation network of China in 2015.

$$\hat{\sigma}_{2015} = \sqrt{\frac{1}{N_{2015}-1}\sum_{i=1}^{N_{2015}}(x_i-\overline{x})^2} = 0.314 \quad (26)$$

Let $\hat{\mu}_{2015}$ and $\hat{\sigma}_{2015}$ be location parameter and scale parameter to substitute in Eq. (8) respectively, the equation of fitting curve could be obtained:

$$\hat{y}_{2015} = 1.27e^{-5.076(x-2.136)^2} \quad (27)$$

The fitting curve (Eq. (27)) and sample points $p_{i,2015}$ were drawn in Fig. 14. The effect of fitting was good in Fig. 14. The $\varepsilon_{i,2015}$ $(i=1,2,3,4,5)$ in equation (28) was calculated as residual to investigate the degree of approximation of fitting through the residual distribution [6].

$$\varepsilon_{i,2015} = p_{i,2015} - \hat{y}_{i,2015} \tag{28}$$

The residual $\varepsilon_{i,2015}$ was drawn in Fig. 15. All points in Fig. 15 were random fluctuation around $\varepsilon = 0$ and the range of fluctuation was very small. $\left|\varepsilon_{i,2015}\right| < 0.05 < \hat{\sigma}_{2015} = 0.314$. This illustrated that the fitting effect was very good.

2.6 The Evolution of Probability Distribution of Average Length of Node Path of Aviation Network of China

The probability distribution parameter and probability density function of average path length of nodes of aviation network of China in different years were in Table 7. The evolution of parameters and probability density in years 1988, 1994, 2001, 2008 and 2015 could be found in Table 7. In order to observe the evolution tendency of probability distribution of average path length of nodes, the curves of probability distribution in these 5 years were drawn in Fig. 16 together. In Fig. 16, the normal distribution curves from right to left were probability distribution in years 1988, 1994, 2001, 2008 and 2015 respectively. The curve became narrow from wide and the peak of curve became higher from lower. The change of location parameter illustrated the same condition. It was that the median of curve moved to left and became smaller. The change of scale parameter is also like this. It was that the value of scale parameter became smaller year by year and the curve became narrower. Since the area under the curve was invariant with value of 1, so the peak of curve became higher.

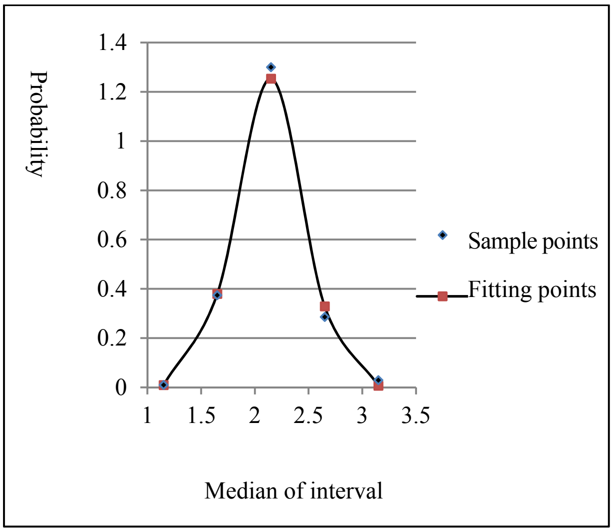

Fig. 14 Fitting effect diagram of probability distribution of average path length of node in 2015.

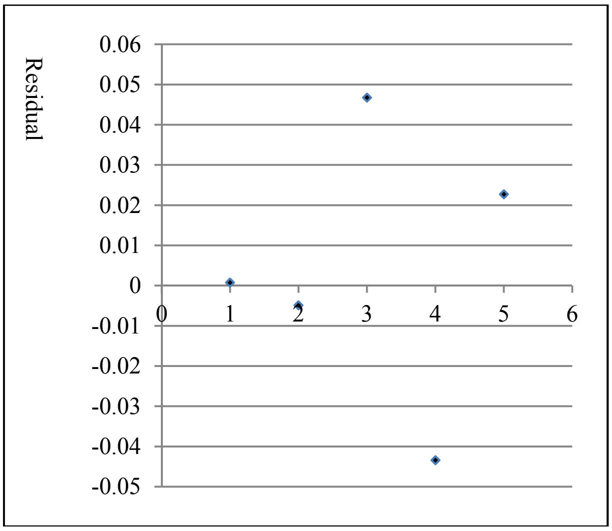

Fig. 15 Residual diagram of fitting curve of probability distribution of average path length of node in 2015.

Table 7 The evolution of probability distribution parameter and probability density function of average path length of nodes of aviation network of China.

Year	1988	1994	2001	2008	2015
Location parameter	2.592	2.356	2.27	2.296	2.136
Scale parameter	0.493	0.395	0.379	0.379	0.314
Probability density function	$0.81e^{-2.058(x-2.592)^2}$	$1.01e^{-3.205(x-2.356)^2}$	$1.053e^{-3.48(x-2.27)^2}$	$1.053e^{-4.37(x-2.296)^2}$	$1.27e^{-5.076(x-2.136)^2}$

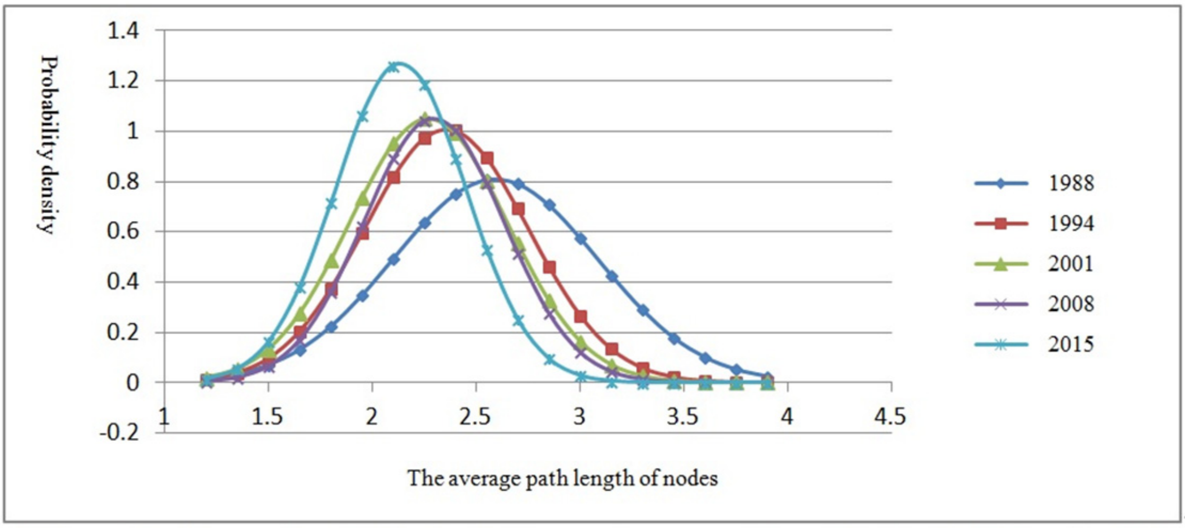

Fig. 16 The evolution diagram of probability distribution curve of average path length of nodes of aviation network of China.

3. The Linear Evolution Trace of Probability Distribution Parameter of Average Path Length of Nodes of Aviation Network of China

The evolution trace of probability distribution parameters of average path length of nodes of aviation network of China was discussed here. Let the location parameter of normal distribution be x. Let the scale parameter of normal distribution be y. The x was regarded as abscissa and y was regarded as ordinate. The location parameters and scale parameters in Table 7 were drawn in Fig. 17 as scattered points. The correlation coefficient r of scattered points in Fig. 17 was calculated.

$$r = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (29)$$

Here, $n = 5$. Using the data in Table 7, the value of correlation coefficient r was calculated, $r = 0.984$. The critical value of $r\big|_{\substack{\alpha=1\% \\ f=3}}$ was 0.959 found in critical value table [6] at degree of freedom $f = n - 2 = 3$ and level of significant α of 1%. Since $|r| = 0.984 > 0.959 = r\big|_{\substack{\alpha=1\% \\ f=3}}$, the scattered points in Fig. 17 have significant linear correlation. Least square method [6] was used as an approach in Eq. (30) to fit the line with points in Fig. 17.

$$\begin{cases} \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = -0.503 \\ \hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}} = 0.384 \end{cases} \quad (30)$$

The linear equation:

$$y = 0.384x - 0.503 \quad (31)$$

The points of fitting line (Eq. (31)) was drawn with the sample points in one diagram of Fig. 18. The fitting effect was good in Fig. 18. To take t test [6] of Eq. (31), test hypotheses is: $H_0 : \beta_1 = 0$. When the hypotheses is true, there is:

$$\hat{\beta}_1 \sim N(0, \frac{\sigma^2}{L_{xx}}) \quad (32)$$

Here, $\hat{\beta}_1$ fluctuate near zero, statistic t is build.

$$t = \frac{\hat{\beta}_1}{\sqrt{\dfrac{\hat{\sigma}^2}{L_{xx}}}} = \frac{\hat{\beta}_1 \sqrt{L_{xx}}}{\hat{\sigma}} \quad (33)$$

where:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (34)$$

Statistic t was calculated by data in Table 7 and Eqs. (31), (33) and (34): $t = 17.4$.

To check the t distribution table [6], at significant level α of 0.01 and degree of freedom $f = n-2 = 3$ the value of $t_{\substack{\alpha=0.01\\ f=3}}$ in table is 4.541. So, $|t| = 17.4 > 4.541 = t_{\substack{\alpha=0.01\\ f=3}}$, null hypotheses H_0 is refused. The linear correlation of Eq. (31) is significant.

The linear relationship between location parameter and scale parameter in Fig. 18 illustrated that the evolution of probability distribution of average path length of nodes of aviation network of China had linear trace.

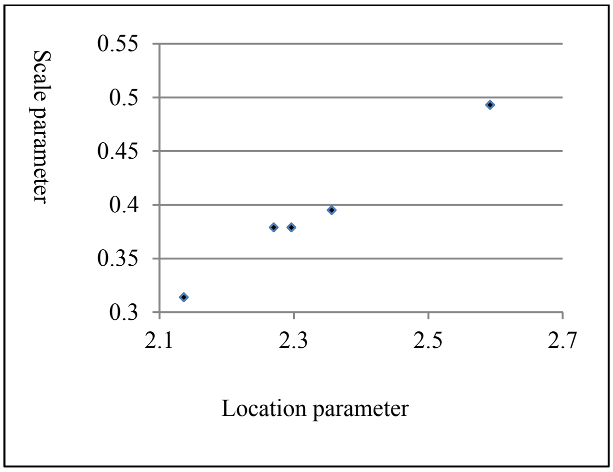

Fig. 17 The scattered points diagram of parameter relationship of probability distribution of average path length of nodes.

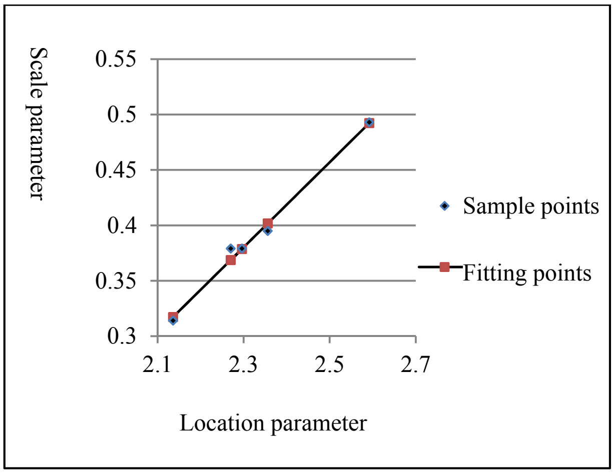

Fig. 18 The fitting effect diagram of parameter linear relationship of probability distribution of average path length.

4. The Linear Evolution Trace between Airline Rate and Average Path Length of Aviation Network of China

The airline rate R is an index to illustrate the density of airline. It is defined as following:

$$R = \frac{L}{C_N^2} = \frac{L}{\dfrac{N!}{2!(N-2)!}} = \frac{2L}{N(N-1)} \tag{35}$$

In the formula: L—the amount of airlines among cities; N—the amount of cities with airport.

The index R means the ratio of actual airlines with all possible airlines. It could be known from Eqs. (4), (5), (9), (13), (17), (21) and (25) that the average path length of aviation network was the same value of location parameter of probability distribution of average path length of aviation network in the same year. The average path length of aviation network in different year was in Table 9. The airline rate in each year calculated by the data in Table 1 was written in Table 9.

Let the average path length of aviation network be x and the airline rate be y. The x was regarded as abscissa and y was regarded as ordinate. The scattered points diagram of relationship between the average path length of aviation network and airline rate was drawn in Fig. 19. The correlation coefficient r of scattered points in Fig. 19 was calculated.

$$r = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{36}$$

Here, $n = 5$. Using the data in Table 9, the value of correlation coefficient r was calculated, $r = -0.964$. The critical value of $r_{\substack{\alpha=1\%\\ f=3}}$ was 0.959 found in critical value table [6] at degree of freedom $f = n-2 = 3$ and level of significant α of 1%. Since $|r| = 0.964 > 0.959 = r_{\substack{\alpha=0.01\\ f=3}}$, the scattered points in Fig. 19 have significant linear correlation. Least square method [6] was used as an approach in Eq. (30) to fit the line with points in Fig. 19.

Table 8 The relation between the average path length of aviation network and airline rate of China.

Year	1988	1994	2001	2008	2015
Average path length of aviation network	2.592	2.356	2.27	2.296	2.136
Airline rate	0.0742	0.0798	0.0871	0.0841	0.0938

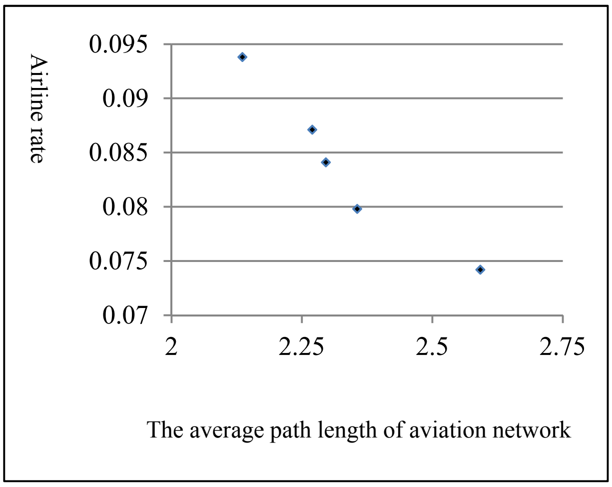

Fig. 19 The relationship scattered points diagram between average path length and airline rate.

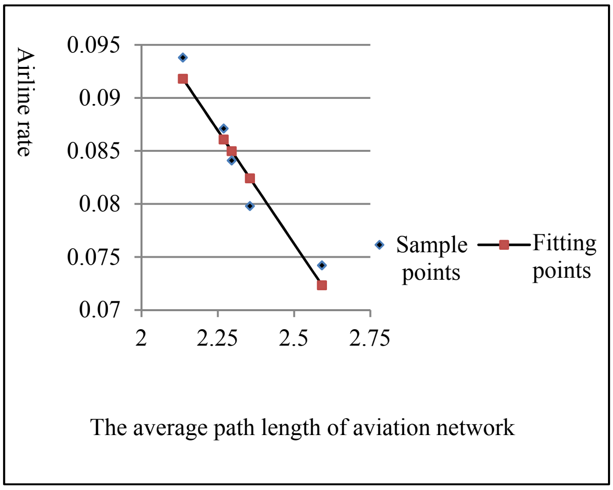

Fig. 20 The fitting effect diagram of linear relationship between average path length and airline rate.

$$\begin{cases} \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 0.183 \\ \hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}} = -0.0427 \end{cases} \tag{37}$$

The linear equation:

$$y = -0.0427x + 0.183 \tag{38}$$

The points of fitting line (Eq. (38)) were drawn with the sample points in one diagram of Fig. 20. The fitting effect was good in Fig. 20. To take t test [6] of Eq. (38), test hypotheses is: $H_0 : \beta_1 = 0$. When the hypotheses is true, there is:

$$\hat{\beta}_1 \sim N(0, \frac{\sigma^2}{L_{xx}}) \tag{39}$$

Here, $\hat{\beta}_1$ fluctuate near zero, statistic t is build.

$$t = \frac{\hat{\beta}_1}{\sqrt{\dfrac{\hat{\sigma}^2}{L_{xx}}}} = \frac{\hat{\beta}_1 \sqrt{L_{xx}}}{\hat{\sigma}} \tag{40}$$

where:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{41}$$

Statistic t was calculated by data in Table 9 and Eqs. (38), (40), (41): $t = -6.15$.

To check the t distribution table [6], at significant level α of 0.01 and degree of freedom $f = n - 2 = 3$ the value of $t_{\left|\substack{\alpha=0.01 \\ f=3}\right.}$ in table is 4.541.

So, $|t| = 6.15 > 4.541 = t_{\left|\substack{\alpha=0.01 \\ f=3}\right.}$, null hypotheses H_0 is refused. The linear correlation of Eq. (38) is significant.

The linear relationship between the average path length of aviation network and airline rate in Fig. 20 illustrated that the synchronous evolution of the average path length of aviation network and airline rate of China had linear trace.

5. Conclusion

On the basis of statistics data of China civil aviation in years 1988, 1994, 2001, 2008 and 2015, the probability distribution and its evolution trace of average path length of nodes of aviation network of China were researched. According to the theory of complex network, 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 average length of node path and its evolution trace of aviation network of China. Floyd algorithm to calculate the path length between any two nodes of network was applied and average length of node path of aviation network was obtained according to this algorithm. It was discovered that average length of node path to other nodes had normal distribution function in each year. At meantime, the location parameter and scale parameter of normal distribution function had linear evolution trace. Airline rate was an index to describe the density of airline. It was found that average length of node path of aviation network of China evolved synchronously with airline rate and they had linear relationship.

Reference

[1] Hossian, M. M., and Alam, S. 2017. "A Complex Network Approach towards Modeling and Analysis Australia Airport Network." *Journal of Airport Transport Management* 60: 1-9.

[2] 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.

[3] Division of Development and Planning of Civil Aviation Administration of China. 1990, 1996, 2003, 2010, 2017. *Civil Aviation Statistical Yearbook of China*. Beijing: Civil Aviation Press of China.

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

[5] Zhang, G. F., and Zhang, Y. 2017. *Probability and Statistics*. Beijing: Higher Education Press.

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