Probability Distribution and Their Non-linear Relationship between Node Degree and Clustering Coefficient of Aviation Network of China Based on Complex Network

Cheng Xiangjun¹, Yu Aihui² and Chen Xumei¹
1. School of Traffic & Transportation, Beijing Jiaotong University, Beijing 100044, China
2. Civil Aviation Management Institute of China, Beijing 100102, China

Abstract: In order to reveal the complex network feature of aviation network of China, probability distribution of node degree and clustering coefficient of aviation network of China was researched according to statistics data of civil aviation of China. It was verified that node degree had power function probability distribution. Clustering coefficient of nodes with exponential function probability distribution was discovered. It was found that node degree and clustering coefficient had single peak nonlinear relationship. At the left side of the peak, there is no certain relationship between them. At the right side of the peak, clustering coefficient became smaller with the rise of node degree and there was negative exponential function relationship between them by regression analysis.

Key words: Aviation network of China, complex network feature, probability distribution, regression analysis, curve fitting.

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 of civil aviation airlines among 203 airports in 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. It was verified that node degree of aviation network of China has the power function probability distribution. It was discovered that the node clustering coefficient of aviation network of China has the exponential function probability distribution. By regression analysis and curve fitting, correlation between node degree and clustering coefficient was discussed. Non-linear relationship between node degree and clustering coefficient was found. The value of clustering coefficient was scattered in range [2,12] of node degree; it monotonously declined in range [13,142] of node degree. Through analyzing, there is no correlation between node degree and clustering coefficient in range [2,12] of node degree; in range [13,142] of node degree, the value of clustering coefficient decreases with the increasing of node degree and there is negative exponential relationship between them.

2. Node Degree and Its Probability Distribution of Aviation Network of China

The degree \( D_v \) of node \( V_v \) in complex network is the amount of edges connected to this node [3].
Airport was regarded as node of aviation network. Airline between airports was regarded as edge. Table 1 about node degree interval of aviation network in 2015 of China was sorted through statistics data [4].

Probability distribution diagram of node degree of aviation network of China was drawn in Fig.1 using median and probability in Table 1. Let the node degree 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 between logarithm of node degree and logarithm of probability in Fig. 2. The correlation coefficient $r$ of scattered points in Fig. 2 was calculated by Eq. (1) [5].

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

Here, $n = 8$. Using the data in Table 1, the value of correlation coefficient $r$ was calculated, $r = -0.986$. The critical value of $r_{1-\alpha}$ was 0.834 found in critical value table [5] at degree of freedom $f = n - 2 = 6$ and level of significant $\alpha$ of 1%. Since $|r| = 0.986 > 0.834 = r_{1-0.01}$, the scattered points in Fig. 2 have significant linear correlation. Least square method [5] was used as an approach in Eq. (2) to fit the line with points in Fig. 2.

$$
\begin{align*}
\hat{\beta}_0 &= \overline{v} - \hat{\beta}_1 \overline{u} = 4.151 \\
\hat{\beta}_1 &= \frac{L_{uv}}{L_{uu}} = -1.845
\end{align*}
$$

The linear equation:

$$
    v = 4.151 - 1.845u
$$

To take $t$ test [6] of Eq. (3), test hypothesis is: $H_0: \beta_1 = 0$. When the hypothesis is true, there is:

$$
    \hat{\beta}_1 \sim N(0, \frac{\sigma^2}{L_{uu}})
$$
Here, $\hat{\beta}_1$ fluctuates near zero; statistic $t$ is build.

$$t = \frac{\hat{\beta}_1}{\hat{\sigma}} = \frac{\hat{\beta}_1 \sqrt{L_{uu}}}{\hat{\sigma}}$$

(5)

where:

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

(6)

Statistic $t$ was calculated by data: $t = -11.37$.

To check the $t$ distribution table [5], at significant level $\alpha$ of 0.01 and degree of freedom $f = n - 2 = 6$, the value of $t_{\alpha=0.01}$ in table is $3.143$.

So, $|t| = 11.37 > 3.143 = t_{\alpha=0.01}$, null hypothesis $H_0$ is refused. The linear correlation of Eq. (3) is significant. The fitting curve Eq. (7) of probability distribution of node degree is derived from Eq. (3).

$$y = 63.5x^{-1.845}$$

(7)

Fitting points of curve (Eq. (7)) with sample points in Fig. 4 show a good fitting effect. It illustrates that the probability distribution of node degree of aviation network of China is power function curve.

### 3. Node Clustering Coefficient and Its Probability Distribution of Aviation Network of China

The definition of clustering coefficient $C_j$ of node $v_j$ in complex network [3] is in Eq. (8).

$$C_j = \frac{E_j}{C^2_{k_j}}$$

(8)

In the formula:

$k_j$—the amount of neighbor nodes of node $v_j$, there is an edge between the node and its neighbor node;

$E_j$—the existed edges among $k_j$ neighbor nodes of node $v_j$;

$C^2_{k_j}$—all the possible edges among $k_j$ neighbor nodes of node $v_j$.

According to Eq. (8), the node clustering coefficient of 203 airports of China in year 2015 was investigated. Since 12 airports among them only have one airline of each airport, there is only one neighbor node for these 12 nodes. It makes the possible edges among neighbor nodes be zero and let the denominator be zero to make the formula be invalid. So these 12 airport nodes have no node clustering coefficient and the rest 191 airport node can be calculated node clustering coefficient.

The adjacent matrix $A = (a_{ij})_{203 \times 203}$ was got from statistic data, where:
The clustering coefficient of each node was calculated by adjacent matrix $A$. Program code of language C to calculate node clustering coefficient is followed.

```c
unsigned short i, j, k, N, M;
unsigned short NeiNum;
unsigned short Neibour[203];
float NeiLink;
float C[203];
N = 203;
for (i = 1; i <= N; i++)
{
    NeiNum = 0;
    for (j = 1; j <= N; j++)
    {
        if (A[i][j] == 1)
        {
            NeiNum = NeiNum + 1;
            Neibour[NeiNum] = j;
        }
    }
    NeiLink = 0;
    M = NeiNum;
    for (j = 1; j <= M; j++)
    {
        for (k = 1; k <= M; k++)
        {
            if (A[Neibour[j]][Neibour[k]] == 1)
            NeiLink = NeiLink + 1;
        }
    }
    if (M == 1) C[i] = 0;
    else C[i] = NeiLink / (M * (M - 1));
}
```

Probability distribution diagram of node clustering coefficient of aviation network of China was drawn in Fig. 5 using median and probability in Table 2. Let the node clustering coefficient be $x$ and probability be $y$. Let $v = \ln y$. Take $x$ as abscissa and $v$ as ordinate to draw associated scatter diagram between node clustering coefficient and logarithm of probability in Fig. 6.

![Fig. 5 Probability distribution diagram of node clustering coefficient of aviation network of China.](image)

![Fig. 6 Relationship diagram between node clustering coefficient and logarithm of probability.](image)

**Table 2  Frequency and probability of node clustering coefficient of aviation network of China.**

| Node clustering coefficient interval | [0.0, 0.1) | [0.1, 0.2) | [0.2, 0.3) | [0.3, 0.4) | [0.4, 0.5) | [0.5, 0.6) | [0.6, 0.7) | [0.7, 0.8) | [0.8, 0.9) | [0.9, 1] |
|-------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|----------|
| Median                             | 0.05       | 0.15       | 0.25       | 0.35       | 0.45       | 0.55       | 0.65       | 0.75       | 0.86       | 0.95     |
| Frequency                          | 1          | 2          | 7          | 10         | 9          | 12         | 20         | 15         | 26         | 89       |
| Probability                        | 0.0052     | 0.0105     | 0.0367     | 0.0524     | 0.0471     | 0.0628     | 0.1047     | 0.0785     | 0.1361     | 0.466    |
The correlation coefficient $r$ of scattered points in Fig. 6 was calculated by Eq. (10).

$$ r = \frac{L_{vv}}{\sqrt{L_{xx}L_{vv}}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad (10) $$

Here, $n = 10$. Using the data in Table 2, the value of correlation coefficient $r$ was calculated, $r = 0.934$. The critical value of $r_{0.01}$ was 0.765 found in critical value table [5] at degree of freedom $f = n - 2 = 8$ and level of significant $\alpha$ of 1%.

Since $|r| = 0.934 > 0.765 = r_{0.01}$, the scattered points in Fig. 6 have significant linear correlation. Least square method was used as an approach in Eq. (11) to fit the line with points in Fig. 6.

$$ \begin{align*}
\hat{\beta}_0 &= \bar{v} - \hat{\beta}_1 \bar{x} = -4.887 \\
\hat{\beta}_1 &= \frac{L_{xy}}{L_{xx}} = 3.869
\end{align*} \quad (11) $$

The linear equation:

$$ v = -4.887 + 3.869x \quad (12) $$

The points of fitting line (Eq. (12)) were drawn with the sample points in one diagram of Fig. 7.

To take $t$ test [7] of Eq. (12), test hypothesis is: $H_0: \beta_1 = 0$.

When the hypothesis is true, there is:

$$ \hat{\beta}_1 \sim N(0, \frac{\sigma^2}{L_{xx}}) \quad (13) $$

Here, $\hat{\beta}_1$ fluctuates near zero, statistic $t$ was build.

$$ t = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{L_{xx}}} = \frac{\hat{\beta}_1 \sqrt{L_{xx}}}{\hat{\sigma}} \quad (14) $$

where:

$$ \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{v}_i)^2 \quad (15) $$

Statistic $t$ was calculated by data: $t = 7.37$.

To check the $t$ distribution table [5], at significant level $\alpha$ of 0.01 and degree of freedom $f = n - 2 = 8$ the value of $t_{0.01}$ in table is 2.896.

So, $|t| = 7.37 > 2.896 = t_{0.01}$, null hypothesis $H_0$ is refused. The linear correlation of Eq. (12) is significant.

The fitting curve Eq. (16) of probability distribution of node clustering coefficient was derived from Eq. (12).
Fitting points of curve (Eq. (16)) with sample points in Fig. 8 show a good fitting effect. It illustrates that the probability distribution of node clustering coefficient of aviation network of China is exponential function curve.

4. Non-linear Relationship between Node Degree and Clustering Coefficient of Aviation Network of China

The 191 pairs of points consisting of node degree and clustering coefficient were drawn in Fig. 9. It illustrated non-linear relationship between node degree and clustering coefficient. The range of node degree was [2,142] and the range of node clustering coefficient was [0,1]. The value of node clustering coefficient declines monotonously from 13 to 142 of node degree.

Since the relationship between node degree and clustering coefficient was different that could be seen in Fig. 9 in range [2,12] and range [13, 142] of node degree, the 115 points in range [2,12] of node degree were drawn in Fig. 10 to study first.

Let the node degree be \( x \) as abscissa and node clustering coefficient be \( y \) as ordinate in Fig. 10.

The correlation coefficient \( r \) of scattered points in Fig. 10 was calculated by Eq. (17).

\[
 r = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{17}
\]

Here, \( n = 115 \). The value of correlation coefficient \( r \) was calculated, \( r = -0.15 \). The critical value of \( R_{p=0.05, f=113} \) was 0.183 found in critical value table at degree of freedom \( f = n - 2 = 113 \) and level of significant \( \alpha \) of 5%. Since \( |r| = 0.15 < 0.183 = R_{p=0.05, f=113} \), the scattered points in Fig. 10 have no linear correlation.

Then the 76 points in range [13,142] of node degree were drawn in Fig. 11 to study the relationship between node degree and clustering coefficient.
Let the node degree be \( x \) and node clustering coefficient be \( y \). Let \( v = \ln y \). Let \( x \) be abscissa and \( v \) be ordinate to draw associated scatter diagram between node degree and logarithm of node clustering coefficient in Fig. 12.

The correlation coefficient \( r \) of scattered points in Fig. 12 was calculated by Eq. (18).

\[
r = \frac{L_{vx}}{\sqrt{L_{xx}L_{yy}}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (v_i - \bar{v})^2}}
\]

(18)

Here, \( n = 76 \). The value of correlation coefficient \( r \) was calculated, \( r = -0.96 \). The critical value of \( r_{0.01,74} \) was 0.296 found in critical value table at degree of freedom \( f = n - 2 = 74 \) and level of significant \( \alpha \) of 1\%. Since \( |r| = 0.96 < 0.296 = r_{0.01,74} \), the scattered points in Fig. 6 have significant linear correlation. Least square method was used as an approach in Eq. (19) to fit the line with points in Fig. 12.

\[
\begin{align*}
\hat{\beta}_0 &= \bar{v} - \hat{\beta}_1 \bar{x} = 0.0616 \\
\hat{\beta}_1 &= \frac{L_{vx}}{L_{xx}} = -0.0147
\end{align*}
\]

(19)

The linear equation:

\[ v = 0.0616 - 0.0147x \] (20)

The points of fitting line (Eq. (20)) were drawn with the sample points in one diagram of Fig. 13.

To take \( t \) test of Eq. (20), test hypothesis is:

\[ H_0 : \beta_1 = 0 \]

When the hypothesis is true, there is:

\[ x \sim N(0, \frac{\sigma^2}{L_{xx}}) \] (21)

Here, \( \hat{\beta}_1 \) fluctuates near zero, statistic \( t \) was build.

\[ t = \frac{\hat{\beta}_1 \sqrt{L_{xx}}}{\hat{\sigma}} \] (22)

where:

\[ \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (v_i - \hat{v}_i)^2 \] (23)

Statistic \( t \) was calculated by data: \( t = -28.98 \).

To check the \( t \) distribution table, at significant level \( \alpha \) of 0.01 and degree of freedom \( f = n - 2 = 74 \) the value of \( t_{0.01,74} \) in table is 2.383. So, \( t = 28.98 > 2.383 = t_{0.01,74} \), null hypothesis \( H_0 \) is refused. The linear correlation of
Eq. (20) is significant. The fitting curve Eq. (24) of relationship of node degree with node clustering coefficient is derived from Eq. (20).

\[ y = 1.0635e^{-0.0147x} \]  

Fitting points of curve (Eq. (24)) with sample points in Fig.14 show a good fitting effect. It illustrates that the node degree and node clustering coefficient of aviation network of China are negative exponential relationship at right side of curve peak.

5. Conclusion

Passenger data of civil aviation airlines among 203 airports in 2015 of China were disposed 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 edge. Node degree of aviation network of China was got by establishing adjacent matrix. After programming, node clustering coefficient of aviation network was calculated through computer program. It was verified that node degree of aviation network of China has the power function probability distribution. It was discovered that the clustering coefficient of aviation network of China has the exponential function probability distribution. By regression analysis and curve fitting, correlation between node degree and clustering coefficient was discussed. Non-linear relationship between node degree and clustering coefficient was found. The value of clustering coefficient was scatted in range [2,12] of node degree, it monotonously declined in range [13,142] of node degree. Through analyzing, there is no correlation between node degree and clustering coefficient in range [2,12] of node degree; in range [13,142] of node degree, the value of clustering coefficient decreases with the increasing of node degree and there is negative exponential relationship between them.

References

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

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

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