Analysis of the Chinese provincial air transportation network

Wen-Bo Du a,b,c, Bo-Yuan Liang a,b,c, Chen Hong a,d,*, Oriol Lordan e

a School of Electronic and Information Engineering, Beihang University, Beijing 100191, PR China
b Beijing Key Laboratory for Network-based Cooperative Air Traffic Management, Beijing 100191, PR China
c Beijing Laboratory for General Aviation Technology, Beijing 100191, PR China
d College of Information Technology, Beijing Union University, Beijing 100101, PR China
e Universitat Politècnica de Catalunya-BarcelonaTech, C/Colom no. 11, Terrassa 08222, Spain

HIGHLIGHTS

- We examine the Chinese provincial air transportation network via the complex network theory.
- We investigate the variation of flight flow within 24 h.
- We found an obvious tide phenomenon in the dynamics of the Chinese provincial ATN for high output level of tertiary industry.

ARTICLE INFO

Article history:
Received 5 June 2016
Received in revised form 10 August 2016
Available online xxxx

Keywords:
Chinese air transportation network
Structural property
Flight flow
Complex network

ABSTRACT

The air transportation system is of a great impact on the economy and globalization of a country. In this paper, we analyze the Chinese air transportation network (ATN) from a provincial perspective via the complex network framework, where all airports located in one province are abstracted as a single node and flights between two provinces are denoted by a link. The results show that the network exhibits small-world property, homogeneous structure and disassortative mixing. The variation of the flight flow within 24 h is investigated and an obvious tide phenomenon is found in the dynamics of Chinese provincial ATN for high output level of tertiary industry. Our work will offer a novel approach for understanding the characteristic of the Chinese air transportation network.

© 2016 Published by Elsevier B.V.

1. Introduction

In the past twenty years, the theory and application of complex networks have attracted a great deal of attention from different scientific communities [1–3], such as network modelling [4–6], cascading failures [7–10], evolutionary games [11–14], traffic dynamics [15–17] and optimization [18–23]. As a crucial infrastructure with an enormous impact on local, national, and international economies, air transportation network (ATN) has been extensively studied via complex network framework in recent years [24–27]. Guimerà et al. [28] studied the worldwide ATN and found that the network is a scale-free small-world network. They suggested that the community structure of the worldwide ATN cannot be explained solely based on geographical constraints and that geopolitical considerations must be taken into account. Barrat et al. [29] examined the weighted worldwide ATN and found that high-degree airports have a progressive tendency to form interconnected groups with high-traffic links. Li and Cai [30] examined the airport network of China. The result reveals that the Chinese airport
network shows small-world property, and the cumulative degree distribution of Chinese airport network (both directed and undirected) obeys a two-regime power law.

In most previous studies, air transportation networks are primarily investigated with three different scales: worldwide [31,32], continental [33,34] and national [35–37]. In the case of national scales, some major countries, such as US [38–42], China [43–46], India [47] and Brazil [48], are extensively investigated. The results of these studies show that national ATNs can display different features, such as two-regime power-laws degree distribution, rich-club phenomenon and small-world property and so on. Commonly, a node denotes an individual airport or aggregated airports within the same city. It is well known that air transportation network is sensitive to the economical distribution and geopolitical constraints [28], and provinces can effectively reflect these properties in China. Consequently, it is natural to investigate Chinese ATN in the province level. Meanwhile, provinces of China have exhibited heterogeneous economical developing levels and formed diverse regional airline networks. Hence it is meaningful to identify the roles of provinces in Chinese ATN and analyze how the air traffic interactions between provinces are organized. In this paper, we investigate the features of the Chinese provincial ATN and the variation of flight flow within one day. The results show that the network displays small-world property, disassortative mixing level and is heavily affected by the output of tertiary industry.

The paper is organized as follows. In Section 2 we demonstrate the Chinese provincial ATN and the related network properties in detail. In Section 3, simulation results and correspondent theoretical analysis are provided. Finally, the work is summarized in Section 4.

2. The model

2.1. The Chinese provincial ATN

The data used in this paper are provided by the Air Traffic Management Bureau (ATMB) of China. We consider the data within one week, from June 1, 2015 to June 7, 2015. In the Chinese provincial air transportation network, all airports located in a province are denoted as a single node. It is noteworthy that municipalities, such as Beijing, Shanghai, Tianjin and Chongqing, are also considered as a single node. Table 1 displays the number of airports \( N_a \) in each province with decreasing order. One can see that Xinjiang has the maximum number of airports \( N_a = 17 \) while the value of \( N_a \) corresponding to Tianjin, Hong Kong and Macao is only one. We establish link between node pairs if there are direct flights between them. Fig. 1 shows the structure of the weighted Chinese provincial ATN, which contains \( N = 33 \) nodes and \( M = 462 \) links, and the size of nodes and the width of links are proportional to the number of flights. The average degree of the network is 28, the average shortest path length of the network is 1.13 and the clustering coefficient is 0.883, indicating that Chinese ATN is a densely connected small-world network. The diameter of the network is 2, and over 85% province pairs are directly connected. Table 2 displays a comparison between the basic network parameters of the Chinese provincial ATN and the Chinese ATN [37] whose nodes are airports and links are direct flights. This shows that the Chinese provincial ATN is smaller and denser than the Chinese ATN.

2.2. Network properties

To examine the features of the Chinese provincial ATN, we investigate some properties commonly used in complex network literatures [1,2], including the cumulative degree distribution, degree correlation and the node strength.
Table 1
The number of airports $N_a$ in each province.

<table>
<thead>
<tr>
<th>Province</th>
<th>$N_a$</th>
<th>Province</th>
<th>$N_a$</th>
<th>Province</th>
<th>$N_a$</th>
<th>Province</th>
<th>$N_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xinjiang</td>
<td>17</td>
<td>Gansu</td>
<td>9</td>
<td>Hunan</td>
<td>6</td>
<td>Shanghai</td>
<td>3</td>
</tr>
<tr>
<td>Sichuan</td>
<td>16</td>
<td>Liaooning</td>
<td>8</td>
<td>Shanxi</td>
<td>6</td>
<td>Henan</td>
<td>3</td>
</tr>
<tr>
<td>Inner Mongolia</td>
<td>16</td>
<td>Zhejiang</td>
<td>7</td>
<td>Hebei</td>
<td>5</td>
<td>Ningxia</td>
<td>3</td>
</tr>
<tr>
<td>Guangdong</td>
<td>13</td>
<td>Guangxi</td>
<td>7</td>
<td>Jilin</td>
<td>5</td>
<td>Hainan</td>
<td>2</td>
</tr>
<tr>
<td>Heilongjiang</td>
<td>12</td>
<td>Shanxi</td>
<td>6</td>
<td>Fujian</td>
<td>5</td>
<td>Beijing</td>
<td>2</td>
</tr>
<tr>
<td>Yunnan</td>
<td>12</td>
<td>Anhui</td>
<td>6</td>
<td>Tibet</td>
<td>5</td>
<td>Tianjin</td>
<td>1</td>
</tr>
<tr>
<td>Guizhou</td>
<td>10</td>
<td>Jiangxi</td>
<td>6</td>
<td>Qinghai</td>
<td>4</td>
<td>Hong Kong</td>
<td>1</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>9</td>
<td>Hubei</td>
<td>6</td>
<td>Chongqing</td>
<td>3</td>
<td>Macao</td>
<td>1</td>
</tr>
<tr>
<td>Shandong</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2
A comparison between basic network parameters of the Chinese provincial ATN and the Chinese ATN.

<table>
<thead>
<tr>
<th></th>
<th>Chinese provincial ATN</th>
<th>Chinese ATN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>33</td>
<td>203</td>
</tr>
<tr>
<td>Edges</td>
<td>462</td>
<td>1877</td>
</tr>
<tr>
<td>Average degree</td>
<td>28</td>
<td>18.48</td>
</tr>
<tr>
<td>Average shortest path length</td>
<td>1.13</td>
<td>2.19</td>
</tr>
<tr>
<td>Diameter</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.883</td>
<td>0.73</td>
</tr>
</tbody>
</table>

- The cumulative degree distribution $p(k)$:
  The cumulative degree distribution captures the probability of finding a node with connections not less than $k$. As a key property of complex networks, the cumulative degree distribution has been widely used to characterize the structural and dynamical properties of many real systems [5,35,36]. Here, the cumulative degree distribution is used to characterize the structural heterogeneity of the Chinese provincial ATN.

- The degree correlation $k_{nn}(k)$:
  For each node $i$, the average degree of its nearest neighbors is
  \[ k_{nn}(k_i) = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j, \]
  where $k_i$ is the degree of node $i$, $A_{ij}$ is the adjacency matrix of the network and $N$ is the size of the network. The degree correlation $k_{nn}(k)$ is the average degree of the nearest neighbors of nodes with degree $k$ [49].

  \[ k_{nn}(k) = \sum_{k'} k' P(k'|k), \]
  where $P(k'|k)$ is the conditional probability which following a link of a $k$-degree node we get to a $k'$-degree node. Here, we adopt $k_{nn}(k)$ to reflect the relationship between the degrees of nodes in Chinese provincial ATN.

- The assortativity coefficient $r$:
  To investigate the degree correlation in more detail, we adopt a widely used network property: the assortativity coefficient, which quantitatively measures the assortative mixing level of networks [50]
  \[ r = \frac{M^{-1} \sum_i j_i k_i - \left[ M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[ M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}, \]
  where $j_i$, $k_i$ are the degrees of the nodes at the ends of the $i$th link, with $i = 1, \ldots, M$ and $M$ is the total number of links in the network. Obviously, $-1 \leq r \leq 1$. $r > 0$ represents assortative networks, and $r < 0$ denotes disassortative networks. In this paper, we use the network assortativity coefficient to characterize the assortative mixing level of the Chinese provincial ATN.

- The node strength $S_i$:
  The strength $S_i$ of node $i$ is usually considered as [29]
  \[ S_i = \sum_{j \in I_i} w_{ij}, \]
  where $I_i$ represents the neighbor set of node $i$, and $w_{ij}$ indicates the weight of the link between $i$ and $j$. The strength of a node integrates the information both with its connectivity and the importance of link weights. Here, the weight of a link is defined by its number of flights within a week, where only domestic flights are taken into account, i.e., flights whose origin or destination is outside China are disregarded.
Fig. 2. (a) The cumulative degree distribution $p(k)$. (b) The cumulative distribution of node strength $p(S)$. (c) The degree correlation $k_{nn}(k)$ as a function of degree $k$. (c) The strength correlation $S_{nn}(S)$ as a function of node strength $S$, where strength correlation $S_{nn}(S)$ is the average strength of neighbors of nodes whose strength is $S$.

3. Simulation results

Firstly, we investigate the structural heterogeneity of the Chinese provincial ATN. Fig. 2(a) shows the cumulative degree distribution of the network. One can see that the value of $p(k)$ is stable when $k$ is low while decreases very quickly when $k$ becomes large, which illustrates that the Chinese provincial ATN is of a homogeneous topological structure. Fig. 2(b) displays the cumulative distribution of node strength $p(S)$. One can see that the distribution follows an exponential function, reflecting a heterogeneous distribution of the flight flow between provinces.

To address the degree correlation of the Chinese provincial ATN, we investigate the dependence of $k_{nn}(k)$ on $k$ (Fig. 2(c)). One can see that the value of $k_{nn}(k)$ decreases as the increment of $k$, which means that the higher is the degree of a node, the lower is the average degree of its nearest neighbors, i.e., high-degree nodes tend to connect to nodes with low-degree, reflecting a disassortative mixing level of the Chinese provincial ATN. To measure the disassortative mixing level of the network quantitatively, we calculate the assortativity coefficient $r$ of the network and the value of $r$ is $-0.12$, which further demonstrates that the Chinese provincial ATN is a disassortative network. Fig. 2(d) plots the relationship between the strength correlation $S_{nn}(S)$ and the node strength $S$, where strength correlation $S_{nn}(S)$ is the average strength of neighbors of nodes whose strength is $S$. This shows that $S_{nn}(S)$ exhibits a negative correlation with the node strength $S$, which means that high-strength nodes prefer to link to low-strength nodes in the Chinese provincial ATN.

Fig. 3(a) shows the relationship between the node strength $S$ and the inner flow of provinces $F_{inner}$, where $F_{inner}$ is the sum of the flights inside the province. Guangdong is of the highly developed economy and possesses two first-tier cities (Guangzhou and Shenzhen), and thus the $S$ value of Guangdong is the largest. Most provinces have small number of inner flights except for Yunnan, Xinjiang and Inner Mongolia. Yunnan owns many famous tourist destinations, such as Kunming, Shangri-La and Lijiang, and the land transportation between these tourist destinations is relatively underdeveloped. Consequently, Yunnan is of the largest number of inner flights though its area is not very large. Next we plot the relation between the rate of inner flow $R_{inner}$ and the area of provinces (Fig. 3(b)), where $R_{inner}$ is the rate between $F_{inner}$ and the sum of $F_{inner}$ for all provinces. For top 4 provinces in area, $R_{inner}$ of Xinjiang and Inner Mongolia is remarkably higher than that of Qinghai and Tibet. This can be attributed to the regional economic difference: the number of airports for Qinghai and Tibet is 4 and 5 respectively, and they always fall into the last two in the GDP ranking in the past decades.

It is known that the tertiary industry is of an important impact on the development of air transportation [46,51]. Fig. 4 shows the relationship between the node strength $S$ and the output of tertiary industry of provinces. One can see that the value of $S$ increases as the increment of the output of tertiary industry. According to the number of provinces, we
Fig. 3. (a) The node strength $S$ as a function of the inner flow of provinces $F_{\text{inner}}$, where $F_{\text{inner}}$ is the sum of the number of flights inside the province. (b) The relationship between the rate of inner flow $R_{\text{inner}}$ and the area of provinces, where $R_{\text{inner}}$ is the rate between $F_{\text{inner}}$ and the sum of $F_{\text{inner}}$ for all provinces.

Fig. 4. The node strength $S$ as a function of the output of tertiary industry of provinces. Here, provinces are divided into three levels: high level (HL), middle level (ML) and low level (LL), where each level corresponds to 11 provinces.

To further examine the effect of tertiary industry on air transportation, we consider the variation of the flight flow within 24 h. Fig. 5(a) depicts the variation of the total number of flights $F_{\text{total}}(t) = F_d(t) + F_a(t)$, where $F_d(t)$ is the number of departure flights and $F_a(t)$ represents the number of arrival flights. This shows that the three output levels of tertiary industry exhibit similar circumstances: the flight flow is high in the daytime while it is low in the late-night. Besides, $F_{\text{total}}(t)$ of HL is always larger than those of ML and LL.

Fig. 5(b) shows the variation of the flight difference $D_f(t)$, where $D_f(t) = F_d(t) - F_a(t)$. One can see that the value of $D_f(t)$ under HL shows an obvious tide phenomenon, which exhibits a deep valley in the morning and a remarkable peak in the late-night. In detail, from 5:00 to 7:00, it is the morning peak of departure, but only a few flights arrive; from 7:00 to 9:00, the arrival flights increase while the departure flights decreases, resulting in the increment of $D_f(t)$; from 9:00 to 21:00, there is a balance between arrival and departure flights; from 21:00 to 23:00, it is the evening peak of arrival and thus $D_f(t)$ monotonously increases; from 23:00 to 5:00, the arrival flights decrease to almost zero. For ML and LL, the value of $D_f(t)$ also displays a week tide phenomenon. Fig. 5(c) depicts the variation of the cumulative differential flow $C_f(t)$, where $C_f(t)$ is the cumulative value of $D_f(t)$. One can see that the value of $C_f(t)$ will always reach to zero at the end of the time period whatever the output level is, indicating a balance between the departure flight flow and the arrival flight flow within one day. We further discuss the variation of the number of departure flights $F_d(t)$ and the number of arrival flights $F_a(t)$ for the three levels (Fig. 5(d)). For LL, the differences between departing flow and arriving flow are quite small in the whole day. For ML, there is a weak tide phenomenon. In the case of HL, the number of departure flights is remarkably larger than that of arrival flights in the morning, and an opposite effect is observed in the late-night. To better demonstrate the dynamics of the Chinese provincial ATN during 24 h, we plot the percentage of departure/arrival flights under three time periods: 7:00–7:59, 15:00–15:59 and 23:00–23:59 (Fig. 5(e), (f) and (g)), where yellow (blue) color denotes the percentage of departure (arrival) flights. This shows that the percentage of departure decreases from 7:00 to 23:59, yet an opposite phenomenon is observed for the percentage of arrival flights.

Finally, we study the correlation between the number of international flights $N_{\text{inter}}$ and the number of domestic flights of provinces $N_{\text{dom}}$ (Fig. 6), where $N_{\text{dom}}$ is the sum of the node strength $S$ and the inner flow of provinces $F_{\text{inner}}$. One can see
Fig. 5. (a) The variation of the total number of flights $F_{\text{total}}(t) = F_d(t) + F_a(t)$ under different output levels of tertiary industry, where $F_d(t)$ is the number of departure flights and $F_a(t)$ denotes the number of arrival flights. (b) The variation of the differential flow $D(t) = F_d(t) - F_a(t)$. (c) The variation of the cumulative differential flow $C_d(t)$, where $C_d(t)$ is the cumulative value of $D(t)$. (d) The variation of $F_d(t)$ and $F_a(t)$ under three different levels. The percentage of departure/arrival flights under three time periods: (e) 7:00–7:59, (f) 15:00–15:59 and (g) 23:00–23:59, where yellow (blue) color denotes the percentage of departure (arrival) flights and the size of nodes is proportional to the number of flights.

Table 3

The number and percentage of flights for major provinces, where $N_{\text{dom}}$ is the number of domestic flights, $P_{\text{dom}} = N_{\text{dom}} \times 100\% / N_{\text{all,dom}}$ (where $N_{\text{all,dom}}$ is total number of domestic flights in China), $N_{\text{inter}}$ is the number of international flights and $P_{\text{inter}} = N_{\text{inter}} \times 100\% / N_{\text{all,inter}}$ (where $N_{\text{all,inter}}$ is total number of international flights in China).

<table>
<thead>
<tr>
<th>Province</th>
<th>$N_{\text{dom}}$</th>
<th>$P_{\text{dom}}$</th>
<th>$N_{\text{inter}}$</th>
<th>$P_{\text{inter}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>11 325</td>
<td>14.59%</td>
<td>2333</td>
<td>10.31%</td>
</tr>
<tr>
<td>Shanghai</td>
<td>11 645</td>
<td>15.00%</td>
<td>4332</td>
<td>19.15%</td>
</tr>
<tr>
<td>Guangdong</td>
<td>15 485</td>
<td>19.95%</td>
<td>2371</td>
<td>10.48%</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>2 444</td>
<td>3.15%</td>
<td>6183</td>
<td>27.34%</td>
</tr>
</tbody>
</table>

that most provinces are of small number of international flights except Shanghai, Beijing, Guangdong and Hong Kong. Since Shanghai, Beijing, Guangzhou (located in Guangdong province) are international aviation hubs of China, it is reasonable that the value of $N_{\text{inter}}$ of them is large. Especially, as a famous international metropolis, Hong Kong has the largest number of international flights although its number of domestic flights is small. To further explore the importance of these major provinces, we display their information in detail (Table 3). We can see that Guangdong province possesses almost one fifth of the total domestic flights, while Hong Kong owns more than one fourth of the total international flights.

Please cite this article in press as: W.-B. Du, et al., Analysis of the Chinese provincial air transportation network, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.067
Fig. 6. The number of international flights $N_{\text{inter}}$ as a function of the number of domestic flights of provinces $N_{\text{dom}}$, where $N_{\text{dom}}$ is the sum of the node strength $S$ and the inner flow of provinces $F_{\text{inner}}$.

4. Conclusion

To summarize, we have analyzed the features of the Chinese provincial air transportation network (ATN) with complex network mechanism, where all airports located in each province are represented as a single node. The results demonstrate that the network exhibits small-worldness (i.e., short characteristic path length and high clustering coefficient). The structure of the network is homogeneous while the distribution of flight flow on the network is heterogeneous. By investigating the relationship between $k_{\text{out}}(k)$ and $k$, we found that the network displays disassortative mixing level. We also explored the variation of the flight flow within one day under different output levels of tertiary industry. It is found that, under high output level, a strong tide phenomenon exists in the traffic dynamics, where the number of departure flights is remarkably larger than that of arrival flights in the morning, whereas the opposite effect is exhibited in the late-night.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant Nos. 91538204, 61425014, 61521091), National Key Research and Development Program of China (Grant No. 2016YFB1200100), and National Key Technology R&D Program of China (Grant No. 2015BAG15B01).

References