

## Exploring the network structure and nodal centrality of China's air transport network: A complex network approach

Jiaoe Wang<sup>a</sup>, Huihui Mo<sup>a,b,c,\*</sup>, Fahui Wang<sup>d</sup>, Fengjun Jin<sup>a</sup>

<sup>a</sup> Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

<sup>b</sup> Graduate University of Chinese Academy of Sciences, Beijing 100039, China

<sup>c</sup> China Communications and Transportation Association, Beijing 100053, China

<sup>d</sup> Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA

### ARTICLE INFO

#### Keywords:

Air transport network  
Complex network  
Centrality  
China

### ABSTRACT

This paper uses a complex network approach to examine the network structure and nodal centrality of individual cities in the air transport network of China (ATNC). Measures for overall network structure include degree distribution, average path length and clustering coefficient. Centrality metrics for individual cities are degree, closeness and betweenness, representing a node's location advantage as being directly connected to others, being accessible to others, and being the intermediary between others, respectively. Results indicate that the ATNC has a cumulative degree distribution captured by an exponential function, and displays some *small-world* (SW) network properties with an average path length of 2.23 and a clustering coefficient of 0.69. All three centrality indices are highly correlated with socio-economic indicators of cities such as air passenger volume, population, and gross regional domestic product (GRDP). This confirms that centrality captures a crucial aspect of location advantage in the ATNC and has important implications in shaping the spatial pattern of economic activities. Most small and low-degree airports are directly connected to the largest cities with the best centrality and bypass their regional centers, and therefore sub-networks in the ATNC are less developed except for Kunming in the south-west and Urumchi in the northwest because of their strategic locations for geographic and political reasons. The ANTC is relatively young, and not as efficient and well-developed as that of the US.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

Network structure has been of great interest to transport geographers since the "Quantitative Revolution" in the 1950s (Haggett and Chorley, 1969). More recently, the advancement of *complex network theory* has generated an increasing body of literature on applications in transport systems. Network analysis in geography was rejuvenated with the influx of new concepts and methods. Unlike most studies on complex network analysis by physical scientists, studies by geographers (including this study) have interests beyond testing the statistical properties of the networks and classifying whether a network meets the criteria of *small-world* (SW) or *scale-free* (SF). Rather the spatial structure of the transport network, its distinctive roles of nodes and connection with actual traffic flows are the primary interests of geographic inquiry.

A *small-world network* generally has a small average path length and a large clustering coefficient (Watts and Strogatz,

\* Corresponding author at: Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. Tel.: +86 10 64889086; fax: +86 10 64889302.

E-mail address: [huihuimo@163.com](mailto:huihuimo@163.com) (H. Mo).

1998). In the case of incremental growth in a complex network, new nodes are more likely to connect with well-linked existing nodes (Barabási and Albert, 1999). Consequently, hubs tend to reinforce themselves, and lead to a *scale-free network*. A *scale-free network* is a network whose degree distribution conforms to a power-law (Newman, 2003). There have been a number of studies applying these concepts to analyze air transport networks. Earlier applications were primarily to illustrate and test theories or models (e.g., Amaral et al., 2000; Chi et al., 2003; Guimerá and Amaral, 2004). More recent work explored the implication of network analysis results. For example, Guimerá et al. (2005) observed that the worldwide airport network is a *scale-free small-world network* and the most-connected cities are typically not the most-central, implying the anomaly of centrality values. Complex network measures are also used for evaluating air transport networks of particular countries and airlines, as for Italy (Guida and Maria, 2007), India (Bagler, 2008), and the Lufthansa airline (Reggiani et al., 2009). These studies examined overall network structures and indices for individual nodes such as average path length, nodal clustering, nodal degree and degree distributions. One common oversight is the lack of discussion of geographic, political, and socio-economic factors that strongly influence the configuration and evolution of air transport networks. Focusing network

analysis on China provides the opportunity to explore such aspects.

China has the world's largest national population of 1.33 billion in 2008 (National Bureau of Statistics of China, 2009). Since the beginning of economic reforms and an open-door policy in 1978, China has experienced an average annual growth rate of 14.6 percent in per capita gross domestic product (GDP). These two aspects together contributed to it becoming the world's fastest growing aviation market (Granitsas, 2002). In the past three decades (1978–2008), China's air passenger volume rose by 15.8% yearly, faster than any other transport modes in the country (i.e., rail, road, and water). By 2008, China had 158 commercial airports in operation with an annual air passenger volume of 192 million and annual air passenger movement rose to 288.3 billion person-km, surpassed only by the United States. The air transport network of China (ATNC) has considerably expanded over time and gradually evolved towards a complex network (Li and Cai, 2004). However, the ATNC is relatively young as its commercial air passenger market did not reach a significant scale until the 1980s (Jin et al., 2004). The case study of China, with comparisons to some well-studied and mature air networks such as the US, will shed new light on the development of air networks in emerging economies.

Despite the rising role of air transportation in China, very little has been reported on the evolution of the ATNC, particularly in mainstream international journals. Most studies on the ATNC focused on the spatial pattern of air passenger flows and airport hubs. Among others, Jin et al. (2004) examined the geographic patterns of air passenger transport in China from 1980 to 1998; Zhou and Li (2005) analyzed the relationship between China's airport distribution and tourist development; and Wang and Jin (2007) found that the ATNC is mainly composed by city-pair connections with some primitive features of a hub-and-spoke system. Most recently, Ma and Timberlake (2008) used longitudinal air passenger flow data to analyze the leading cities in China at both the national and global levels during 1990–2005. Other studies (Le, 1997; Zhang, 1998; Liu, 2000; Zhang and Chen, 2003; Zhang and Round, 2008; Yang et al., 2008; Zhang and Round, 2009; Shaw et al., 2009) emphasized the regulation and management of China's airline industry and air transport system. Li and Cai (2004), the only study on ATNC using the complex network approach, identified its basic topological properties. However, like most studies by physical scientists, they did not provide in-depth analysis of geographic and socio-economic factors shaping the network structure.

Complex network theory offers a new set of analytical methods for spatial economic analysis to provide this new insight (Reggiani and Nijkamp, 2007). This study uses network analysis indices to examine the overall network structure and centrality of individual cities in the ATNC, and then analyzes their spatial patterns with relationship to economic and geographic factors.

## 2. The study area and data processing

Data for this study are obtained from the Civil Aviation Administration of China or CAAC (2009). The study area includes all cities with operating airports in mainland China (excluding Hong Kong, Macao and Taiwan) from October 28, 2007 to March 29, 2008. Air routes are the linkages in the network, operated by the following carriers: Air China Airlines, China Eastern Airlines, China Southern Airlines, Hainan Airlines, Shanghai Airlines, Shandong Airlines, Sichuan Airlines, Shenzhen Airlines, and Xiamen Airlines Ltd. etc.

Most cities in the data have a single airport. For some large cities with multiple airports, the data are combined with one entry for each city. In other words, each node in this study represents a city instead of an airport. For example, Shanghai includes data from both the Pudong and Hongqiao airports. Meanwhile, some small

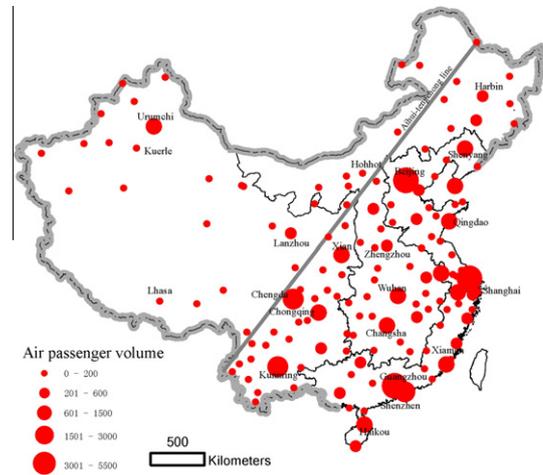


Fig. 1. Air passenger volumes in cities in China 2007 (in 10,000).

Table 1  
Top 15 cities by air passenger volume in China 2007.

Rank	Name	No. of air routes	Population <sup>a</sup> (million)	GRDP <sup>a</sup> (billion Yuan)	Air passenger volume (million)
1	Beijing	92	11.42	920.76	54.37
2	Shanghai	82	13.09	1206.61	51.55
3	Guangzhou	79	7.73	657.05	30.96
4	Shenzhen	65	2.12	680.16	20.62
5	Chengdu	58	5.03	209.19	18.57
6	Kunming	58	2.33	104.87	15.73
7	Hangzhou	43	4.20	325.79	11.73
8	Xi'an	57	5.49	132.95	11.37
9	Chongqing	56	15.26	294.10	10.36
10	Xiamen	42	1.67	138.79	8.68
11	Wuhan	42	5.10	270.90	8.36
12	Changsha	44	2.19	129.90	8.07
13	Nanjing	39	5.34	301.57	8.04
14	Qingdao	44	2.76	213.85	7.87
15	Dalian	39	2.93	225.88	7.28

<sup>a</sup> In municipal districts. Source: China City Statistical Yearbook, 2008; Statistical Data on Civil Aviation of China, 2008.

airports without regular flights are excluded. Both direct and stop-over air routes are considered and combined in the data set. The latter are divided into two parts: from the departure city to the stopover city, and then from the stopover city to the destination city. Duplicated air routes are removed and only one route connects each city-pair. With these adjustments the final network is constituted by 144 cities and 1018 unique air routes. Fig. 1 shows the spatial distribution of the 144 cities and their air passenger volumes.

Table 1 summarizes the GRDP, population, air route, and air passenger volume of the top 15 cities in 2007. All these 15 cities are located in eastern China, i.e., east of the famous “Aihui-teng-chong Line” (see Fig. 1 for location), an imaginary “geo-demographic demarcation line” in China (Hu, 1935). According to the 2000 census, 90.8% of population lived east of this line which accounts for just 43% of national area (Yue et al., 2003). The concentrations of population and economy in eastern China are a major factor in shaping the spatial pattern of air transportation in China.

## 3. Methods

In this paper, the ATNC is abstracted as a connected network  $G = (V, E)$  by  $V$  and  $E$ , where  $V = \{v_i; i = 1, 2, \dots, n\}$ ,  $n = |V|$  is the

number of nodes (vertices), and  $E = \{e_i; i = 1, 2, \dots, m\}$ ,  $m = |E|$  is the number of edges (links). To represent the network, a connectivity (adjacency) matrix  $A_{n \times n}$ , is created such that an element  $a_{ij} = 1$  when a flight link exists between city-pair  $i$  and  $j$  and  $a_{ij} = 0$  otherwise. Two nodes are defined to be *neighbors* if there is a link between them. This section introduces the indices for assessing the overall network structure, the measures for centrality of individual nodes, and the techniques for evaluating the relationship between some indices.

### 3.1. Network structure measures

Several basic indices are used to measure the configuration of a network with a set of edges and nodes.

#### 3.1.1. Degree distribution

Degree is the number of edges that a node shares with others (Barabási and Albert, 1999). For a network with  $n$  nodes, if  $n_k$  of them have degree  $k$ , the degree distribution  $p(k)$  is defined as the fraction of these  $k$ -degree nodes, i.e.,  $n_k/n$ .  $P(k)$  represents the cumulative degree distribution, i.e., the fraction of nodes with degrees greater or equal  $k$ , written as:

$$P(k) = \sum_{k'=k}^{\infty} p(k') \quad (1)$$

The average degree of a network, denoted as  $\langle k \rangle$ , is the average number of neighbors (i.e., directly connected nodes) a node has in the network.

#### 3.1.2. Average path length

Average path length ( $L$ ) is defined as the average number of edges along the shortest paths for all possible node-pairs in the network (Watts and Strogatz, 1998), written as:

$$L = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i>j} d_{ij} \quad (2)$$

where  $d_{ij}$  is the number of edges for the shortest path from  $i$  to  $j$ , and the diameter of a network is defined as the maximum value of all  $d_{ij}$ .

#### 3.1.3. Clustering coefficient

The clustering coefficient ( $C_i$ ) of a node  $i$  is the portion of actual edges ( $E_i$ ) between the nodes ( $k_i$ ) within its *neighborhood* divided by the maximal possible edges ( $k_i(k_i - 1)/2$ ) between them (Watts and Strogatz, 1998), written as:

$$C_i = \frac{E_i}{k_i(k_i - 1)/2} \quad (3)$$

Note that the neighborhood of node  $i$  includes all the nodes directly connected to it but excludes the node  $i$  itself. A larger  $C_i$  value means that the node has a more compact system of connections with its neighbors. In a fully-connected network,  $C_i$  of all nodes equals 1.  $C_i$  of nodes with  $k_i = 1$  equals 0.

The clustering coefficient of the whole network  $C$  is the average of all individual  $C_i$ 's, presented as:

$$C = \frac{1}{n} \sum_{v_i \in V} C_i \quad (4)$$

The larger the value of  $C$  is, the more likely nodes are to reach one another within a short topological distance (i.e., connections or transfers).

According to the aforementioned three indices, the characteristics of regular network, random network, *small-world* network, and *scale-free* network are summarized in Table 2. A *regular network* is a connected graph in which each vertex is connected by the same

**Table 2**  
Characteristics of various networks.

Network	Average path length, $L$	Clustering coefficient, $C$	Degree distribution, $P(k)$
Regular network	Long	Large	Point to point
Random network	Short	Small	Binomial or poisson
<i>Small-world</i> network	Short	Large	Exponential or power-law
<i>Scale-free</i> network	Short	Large	Power-law
Real network	Short	Large	Similar power-law

way exactly as its neighboring vertices. A *random network* is obtained by starting with a set of  $n$  vertices and adding edges between them at random. A *small-world network* is a network that between the regular and random network and has a small average path length and a high clustering coefficient. In a *small-world* network, most nodes are not neighbors of one another, but most of them can be reached by a small number of edges. Many real world networks such as the Internet connectivity and gene networks are represented by *small-world* networks. A *scale-free network* is a network whose degree distribution follows a power-law, at least asymptotically. *Scale-free* networks are noteworthy because many empirically observed networks appear to be *scale-free*, including protein networks, citation networks, and some social networks (Albert and Barabási, 2002). Also, a *small-world* network can be generated from a regular network by re-wiring the cut edges with probability, and a *scale-free* network can be generated by the preferential attachment algorithm.

### 3.2. Centrality measures

Centrality measures the relative importance of a node within a network. In this paper, three indices—degree centrality, closeness centrality, and betweenness centrality—are used to capture a node's importance as being directly connected to others, being accessible to others, and being the intermediary between others.

#### 3.2.1. Degree centrality

As previously discussed, degree centrality is the number of edges that a node shares with others, and thus symbolizes the importance of the node in a network (Freeman, 1977, 1979). Degree centrality of a node  $i$  reflects its connectivity in the network and is defined as:

$$C_D(i) = \sum_{j=1}^n a_{ij} \quad (5)$$

where element  $a_{ij} = 1$  when a direct link exists between nodes  $i$  and  $j$  and  $a_{ij} = 0$  otherwise. If the network is directed, then we usually define two separate measures of degree centrality, namely in-degree and out-degree. In-degree is a count of the number of edges directed to the node, and out-degree is the number of edges that the node directs to others. Since the in-degree is nearly perfectly correlated with out-degree in the ATNC, the network is considered symmetric (Li and Cai, 2004), and thus regarded as an undirected graph in this study.

#### 3.2.2. Closeness centrality

Closeness centrality measures the extent to which a node is close to all other nodes along the shortest path and reflects its accessibility in a given network. The closeness of node  $i$  is written as:

$$C_C(i) = \frac{n-1}{\sum_{v_j \in V, i \neq j} d_{ij}} \quad (6)$$

In other words, a node's closeness is the inverse of the average shortest distance from that node  $i$  to all other nodes in a given network (Sabadussi, 1966). The larger the  $i$  value, the more convenient it is to reach other nodes.

3.2.3. Betweenness centrality

Betweenness centrality measures the extent to which a particular node lies between other nodes in a network, as first described by Anthonisse (1971) and Freeman (1977). A node tends to be more powerful if it is on the shortest paths connecting many node-pairs, as it may be in a position to broker or mediate connections between these pairs. The betweenness of a node  $i$  is defined as the ratio of all shortest paths passing through it and reflects its transitivity. Thus,

$$C_B(i) = \sum_{k \neq i \neq j \in N} \sigma_{kj}(i) / \sigma_{kj} \quad (7)$$

where  $\sigma_{kj}$  is the sum of all shortest paths between nodes  $v_k$  and  $v_j$ , and  $\sigma_{kj}(i)$  is the number of shortest paths that pass through  $v_i$ . Nodes that occur on many shortest paths between other nodes have higher betweenness than those that do not.

3.3. Correlation measures

3.3.1. Degree correlation

Degree correlation demonstrates the extent of a node's degree related to the average degree of its neighbors. This index reflects the node's connection preference. If high-degree nodes tend to link with each other, this tendency is referred to as assortativity. Otherwise, high-degree and low-degree nodes tend to connect with each other is referred to as disassortativity (Newman, 2003). Considering the node  $v_i$  with degree  $k$  and its  $k_i$  neighbors (each  $v_j \in N_i$ ), the average degree of  $N_i$  is defined as:

$$K(i) = \frac{1}{k_i} \sum_{v_j \in N_i} k_j \quad (8)$$

The average degree of all  $k$ -degree nodes  $N_i$  (neighbors of all nodes with  $k$ -degree) is defined as:

$$\overline{K}(k) = \frac{1}{N(k)} \sum_{v_i \in V, k_i=k} K(i) \quad (9)$$

where  $N(k)$  equals the number of  $k$ -degree nodes. Degree correlation refers to the relationship between  $k$  and  $\overline{K}(k)$ .

3.3.2. Clustering-degree correlation

Clustering-degree correlation demonstrates the extent of a node's clustering coefficient related to its degree. The average clustering coefficient  $C(k)$  of all  $k$ -degree nodes  $N_i$  is given by:

$$C(k) = \frac{1}{N(k)} \sum_{v_i \in V, k_i=k} C_i \quad (10)$$

4. Topological analysis of the ATNC

4.1. Cumulative degree distribution conforming to an exponential function

The ATNC's cumulative degree distribution follows an exponential function as  $P(k) = 0.705e^{-0.047k}$  ( $R^2 = 0.977$ ), shown in Fig. 2. That is to say, a few busy cities at the top dominate the system with a large number of air routes, and the number of routes to each city declines quickly and levels off towards small cities, most of which have only 1–3 air routes. For example, the top 20% cities account for a majority (65%) of all air routes, and the bottom one-

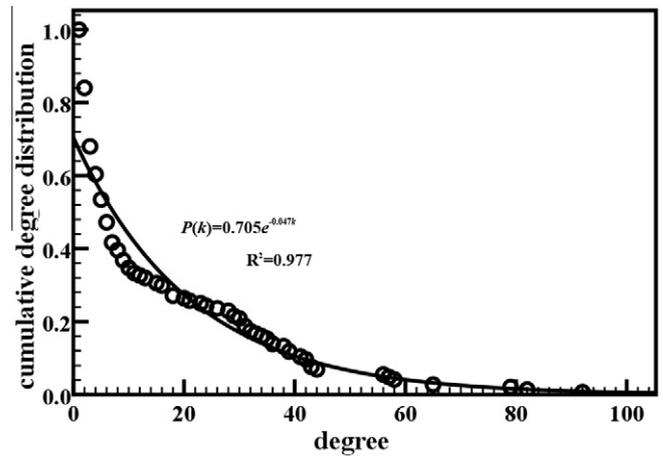


Fig. 2. Cumulative degree distribution in the air transport network of China.

third cities (46 of the 144 cities in the network) are only connected by one or two air routes. The distribution of air passenger volume in the ATNC has even a steeper slope with the top 20% cities accounting for 87.7% of all passenger volume in 2007. Therefore, the air passenger volumes are even more concentrated in a few large cities than are the air routes in China. The cumulative distributions of degree and air passenger volume approximately follow the Pareto principle, also known as the 80–20 rule: about 80% of the effects come from 20% of the causes. The average degree is 14.14 in the ATNC.

Such a distribution pattern is found in most self-organized complex systems in nature, technology, and society. In other words, the ATNC exhibits the statistical property of a self-organized system following some “organic” order in its evolution over time. This is different from a scale-free network, whose degree distribution pattern is better characterized by a power function. In general, the degree distribution captured by an exponential function (as in the ATNC) exhibits a steeper decline, i.e., more dominance of large airports, than that by a power function. This is more common in a developing country than in a developed one. Table 3 compares the air transport network structure of China to other countries and the world as reported in the literature. All share some properties of a small-world network. The ATNC has an average degree value ( $\langle k \rangle$ ) similar to those of the air transport networks of India, Italy and the world, but much smaller than the average degree value in the US. Note that the US air transport network is much larger than that of the ATNC with about twice as many nodes and more than six times the number of edges. There are much fewer edges per node in the ATNC ( $1018/144 = 7.07$ ) than in the US ( $6566/272 = 24.14$ ).

4.2. Average path length above 2

The average path length is an indicator of the convenience of travelling in a given network. Table 4 summarizes the minimum number of flights for travelling for all city-pairs. Over 98% of city-pairs are reachable by changing two flights or less. About 70% of city-pairs are accessible by direct flights or changing one flight. Only about 10% of city-pairs are connected by direct flights. Based on Eq. (2), the ATNC's average path length is 2.23. In other words, it takes over one-flight change on average to connect all city-pairs. This number is slightly larger than a random network ( $L_r \approx 1.88$ ) of the same size. The ATNC's diameter is 5 (at least four flight changes), which exists in connections between the isolated Qjemo and six other cities (i.e., Heihe, Baoshan, Dehong, Fuyang, Liping, and Linyi). Qjemo in southern Xinjiang has only one direct

**Table 3**  
Characteristics of the air transport networks of China and other countries/regions.

Author	Country	No. nodes ( <i>n</i> )	No. edges ( <i>m</i> )	Average degree ( $\langle k \rangle$ )	Average path length ( <i>L</i> )	Clustering coefficient ( <i>C</i> )	Network structure
Bagler (2008)	India	79	455	11.52	2.26	0.66	SW
Guimerá et al. (2005)	World	3883	27,051	13.93	4.4	0.62	SF SW
Guida and Maria (2007)	Italy	50	310	12.40	1.98~2.14	0.07~0.1	SF SW Fractal
Xu and Harriss (2008)	US	272	6566	48.28 <sup>a</sup>	1.84~1.93	0.73~0.78	SW
In this paper	China	144	1018	14.14	2.23	0.69	SW

<sup>a</sup> Calculated by the authors based on the info reported in the paper.

**Table 4**  
Distribution of air routes by number of connection flights.

Shortest path	No. of paths	Percentage of air routes (%)	Cumulative percentage of air routes (%)	No. of flights needed to be changed
1	2036	9.89	9.89	0
2	12,200	59.25	69.14	1
3	5976	29.02	98.16	2
4	362	1.76	99.92	3
5	18	0.09	100.00	4

air route to Kuerle, and then connects through Urumchi (the capital of Xinjiang) in order to reach other cities in the network.

The average path length of 2.23 in the ATNC is very similar to that of India's air transport system (2.26) and slightly above that of Italy's (1.98–2.14), but larger than that of the US (ranging from 1.84 to 1.93) (Table 3). A relatively larger average path length in the ATNC implies that China, like in other emerging countries, remains at a less-developed stage in air transportation with much room to improve in efficiency of connections.

#### 4.3. Clustering coefficient close to 0.7

The ATNC's clustering coefficient is 0.69, much larger than for a random network ( $C_r \approx 0.098$ ) of the same size. A larger clustering coefficient confirms the high-degree of concentration identified earlier and also implies a high probability for travelling with fewer transfers in the network. The clustering coefficient of 0.69 in the air transport network of China is slightly above that of India (0.66) but lower than that of the US (0.73–0.78) (Table 3). With comparison to the US, the ATNC has a smaller average degree value (i.e., cities on average have fewer direct links to other cities), a larger average path length and a smaller clustering coefficient (i.e., potentially more transfers are needed to connect any two cities in the network). All indicate that the air transport infrastructure in China is less mature or efficient than that of the US. It might also imply that the effect of the hub-and-spoke system of the US has not been completely felt yet within the ATNC.

## 5. Node centrality assessment

### 5.1. Statistical distributions

#### 5.1.1. Exponential decline in centrality values

The distribution of all three centrality indices generally conforms to an exponential function (Table 5 and Fig. 3), with  $R^2$  above 0.93. That is to say, the centrality value declines exponentially with the node's ranking. The top 20% cities (i.e., 29 most-central nodes) account for 65.1% of air routes, 93.1% of transfers, and 84.0% air passenger volumes. From Table 5, closeness has the flattest slope (−0.004), betweenness has the steepest slope (−0.137), and the slope for degree is between those two (−0.032). The steep curve of betweenness indicates that a few hub cities account for most of the transfer capacity.

**Table 5**  
Best fitting models for centrality indices (*y*) vs. rankings (*x*).

	Function	<i>a</i>	<i>b</i>	$R^2$	<i>n</i>
Degree, $C_D$	$y = ae^{bx}$	60.284	−0.032	0.976	144
Closeness, $C_C$	$y = ae^{bx}$	0.605	−0.004	0.936	144
Betweenness, $C_B$	$y = ae^{bx}$	0.206	−0.137	0.965	67 <sup>a</sup>

<sup>a</sup> Excluding nodes with betweenness = 0.

Table 6 reports the top 20 cities by degree, closeness, and betweenness. Beijing and Shanghai are ranked at the top by all three indices. Based on both degree and closeness, the next three cities are: Guangzhou, Shenzhen, and Chengdu. Kunming and Urumchi make to the top 5 by betweenness in place of Shenzhen and Chengdu. However, Kunming is ranked the sixth by degree and the eighth by closeness, and Urumchi is ranked only the 28th and 27th by degree and closeness, respectively. Kunming is the gateway to the southwestern region and Urumchi is a regional hub in the northwestern region. Both enjoy considerably large numbers of connection flights and thus have relatively high values of betweenness. Similarly, some cities in peripheral regions such as Hohhot, Guiyang, Xining, Lanzhou, and Kuerle also appear in the top 20 cities by betweenness, but do not make to the top 20 lists by degree or closeness. These cities are regional connector hubs in the network.

The 23 cities with the lowest degree ( $C_D = 1$ , only one air route to other cities) are mostly located in less developed regions near the borders such as those in Yunnan and Xinjiang. In terms of closeness, Qiemo in Xinjiang has the lowest value, followed by Ge'ermu in Qinghai, Fuyang in Anhui, Liping in Guizhou, and Akesu, Aletai, and Tacheng in Xinjiang. All these cities only have one air route to their regional hubs, such as Kuerle, Xining, Hefei, Guiyang, and Urumchi, and have no connections to national hubs. Therefore, it usually takes multiple connection flights for them to connect with other nodes in the network. Seventy seven cities have a betweenness value of zero, indicating that no shortest paths between other city-pairs pass through them. These cities are the peripheral 77 nodes in the network.

#### 5.1.2. Consistency between degree and closeness but less with betweenness

Table 6 shows that the rankings by degree and closeness are generally consistent. The same 18 cities appear in the top 20 lists for both indices. Dalian and Harbin are among the top 20 list by degree but not by closeness; and conversely Guilin and Sanya make to the top 20 list by closeness but not by degree. As pointed previously, the rankings of cities by betweenness can be significantly different from those by degree and closeness. Some cities are highly connected but play a relatively insignificant role for transferability such as Nanjing, Hangzhou, Wenzhou, Linyi, and Ningbo. On the other side, some cities are less connected but serve as important transfer hubs such as Kuerle, Diqing, Kelamayi, and Urumchi. The latter are usually located in the peripheral areas, and play an important role as connector hubs for sub-regions as

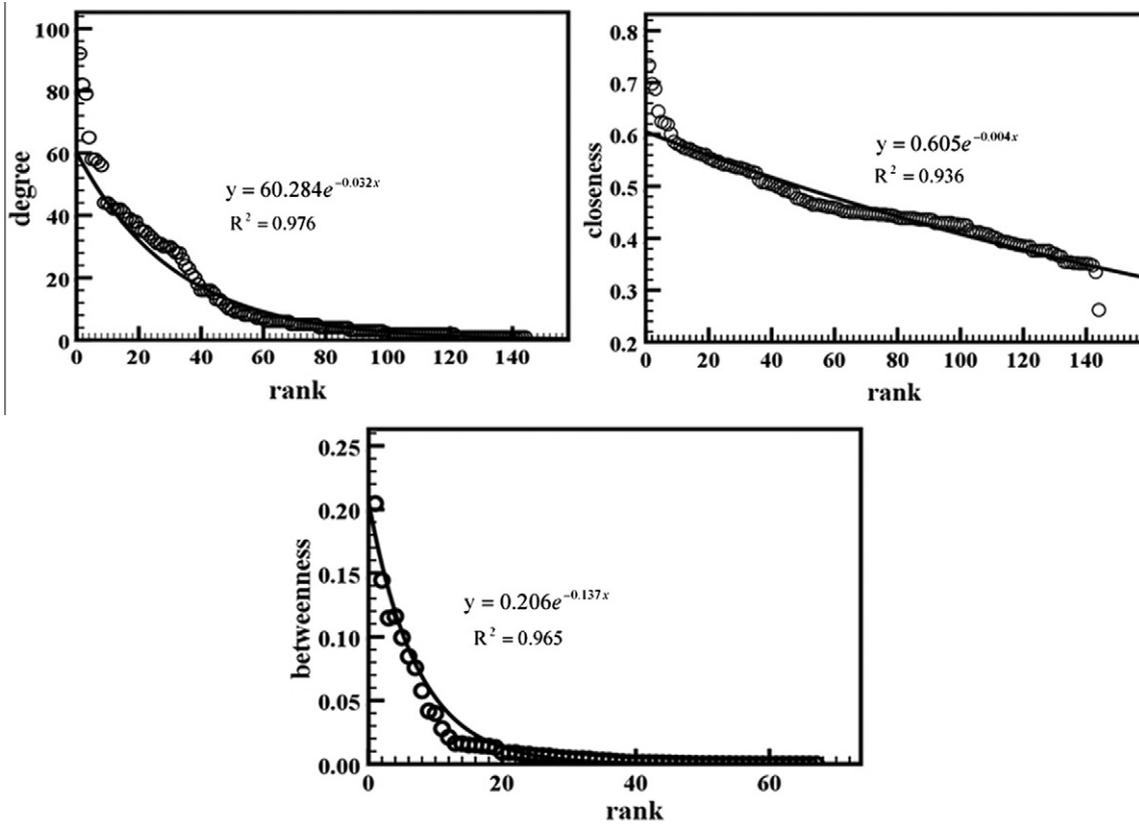


Fig. 3. Statistical distributions of degree, closeness and betweenness.

Table 6  
Top 20 cities by degree, closeness, and betweenness.

Rank	Degree, $C_D$	Closeness, $C_C$	Betweenness, $C_B$	Rank	Degree, $C_D$	Closeness, $C_C$	Betweenness, $C_B$
1	Beijing	Beijing	Beijing	11	Hangzhou	Zhengzhou	Hohhot
2	Shanghai	Shanghai	Shanghai	12	Zhengzhou	Wuhan	Guiyang
3	Guangzhou	Guangzhou	Kunming	13	Wuhan	Ji'nan	Haerbin
4	Shenzhen	Shenzhen	Guangzhou	14	Xiamen	Taiyuan	Wuhan
5	Chengdu	Chengdu	Urumchi	15	Haikou	Qingdao	Hefei
6	Kunming	Xi'an	Xi'an	16	Dalian	Hangzhou	Xining
7	Xi'an	Chongqing	Chengdu	17	Nanjing	Guilin	Lanzhou
8	Chongqing	Kunming	Shenzhen	18	Ji'nan	Haikou	Kuerle
9	Qingdao	Changsha	Chongqing	19	Taiyuan	Sanya	Guilin
10	Changsha	Xiamen	Changsha	20	Harbin	Nanjing	Zhengzhou

shown above. The correlation coefficient is as high as 0.924 between degree and closeness, but much less (0.644) between the closeness and betweenness.

5.2. Spatial patterns

The 144 cities with commercial airports are scattered almost evenly across the three major regions, with 49 in the eastern region, 42 in the central region and 53 in the western region (Table 7; see Fig. 1 for division of the three regions). As noted however, population and economy are not so evenly distributed. Hence, the spatial inequality of their centrality measures is evident (Fig. 4). Generally, cities in the east have better centrality than those in the west. Over half of the top 20 cities with the best centrality values are in the eastern region, about one fourth in the central region and less than one fourth in the western region. The average degree of cities in the eastern region is 22.71, much higher than the central region (10.83) and the western region (8.83). In terms of degree and closeness, the most-central cities are mainly

Table 7  
Centralities and air passenger volumes by region.

Region	Number of cities	Mean			Air passenger volume (million)
		Degree, $C_D$	Closeness, $C_C$	Betweenness, $C_B$	
Eastern	49	22.71	0.50	0.012	259.1
Central	42	10.83	0.46	0.003	45.42
Western	53	8.83	0.43	0.009	83.01

clustered around the greater Beijing area, the Yangtze River Delta and the Pearl River Delta. Provincial capitals such as Xi'an, Chengdu, Kunming, Zhengzhou, Wuhan, and Changsha also have high centrality values. The cities in peripheral areas as discussed earlier, such as in Xinjiang, Qinghai, Gansu, and Tibet, have low centrality values in terms of degree and closeness, excluding Urumchi and Lanzhou. In terms of spatial distribution, betweenness has the highest inequality, the degree next, and the closeness the least.

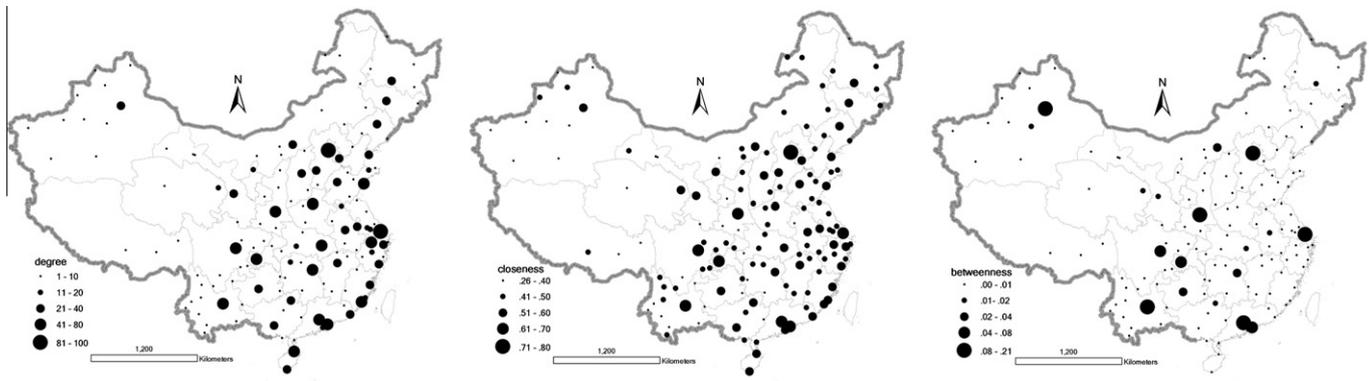


Fig. 4. Spatial distributions of degree, closeness and betweenness.

Most flight transfers are concentrated in several major cities. That can be seen in the highest betweenness value registered for Beijing (0.20), which quickly drops to 0.04 in Changsha (10th ranked), and to 0.01 in Guilin (19th ranked). In fact more than half of the cities (77) have the betweenness value of zero. The polarization of betweenness explains the very significant economic roles played by large cities such as Beijing and Shanghai, although the effect is less for other large cities such as Shenzhen and Wuhan.

## 6. Correlation analysis

### 6.1. Disassortativity in degree correlation

Based on Eqs. (8) and (9), the ATNC's degree correlation equals  $-0.429$ , and they follow a linear function as  $y = -0.429x + 54.64$  ( $R^2 = 0.886$ ), as shown in Fig. 5. A node's degree  $k$  is negatively correlated with the average degree of its next neighbors  $K(i)$ , particularly among moderate- and high-degree cities (e.g., when  $k \geq 8$ , with  $y = -0.414x + 53.88$ ,  $R^2 = 0.947$ ). The higher degree a node has, the lower the average degree of its neighbors ( $K(i)$ ) is. For example, Beijing has the highest degree of 92 and the lowest  $K(i)$  value of 19.4; Shanghai has the second highest degree of 82 and also a low value of  $K(i)$  (21.7); and Guangzhou with the third highest degree of 79 has a low  $K(i)$  value (22.6). Among low-degree cities ( $k < 8$ ), the pattern is less clear. For example, 4-degree cities have the maximum  $\bar{K}(k)$  value (average of  $K(i)$  for all 4-degree cities) of 62.2, followed by 5-, 6- and 2-degree cities. Among individual cities, the highest value  $K(i)$  equals 87 in three 2-degree cities such as Dongying, Jiamusi, and Qiqihaer. All these three cities are connected to Beijing and Shanghai with the highest degrees in

the network. In Yunnan and Xinjiang, low-degree cities are directly linked with their capitals (Kunming and Urumchi) that anchor their respective sub-networks in the regions.

In the worldwide air transport network, when  $k \leq 10$ , the network is assortative ( $k$  and  $\bar{K}(k)$  are positively correlated); when  $k \geq 10$ , the variance of  $\bar{K}(k)$  is very small (Barrat et al., 2004). For the air transport network of the US, when  $k < 30$ , the network is assortative; when  $k \geq 30$ , it becomes disassortative (Barrat et al., 2005). In the ATNC, when  $k < 8$ , the node degree has little relationship with that of its neighbors; when  $k \geq 8$ , the network is disassortative. The ATNC is much smaller than the worldwide or the US air transport network, and thus the cutoff number between the distinctive assortativity patterns differ from those (10 for the world, 30 for the US and eight for China). It is not necessarily assortative for low-degree cities as both the worldwide and the US networks are, but is clearly disassortative for moderate- and high-degree cities like the US (i.e., lower  $\bar{K}(k)$  corresponds to higher degree  $k$ ). In the ATNC, many low-degree cities are prefecture-level central cities, and their main air routes are linked to provincial capitals that generally enjoy better centrality. Major hubs reinforce themselves and prevent the formation of sub-hubs that are close to them. The phenomena may be termed as the "shadow effect" described originally by Taaffe (1959). This helps explain the general pattern of disassortativity of the ATNC, particularly for high-degree cities.

### 6.2. Nonlinear clustering-degree relation

Fig. 6 shows the relationship between clustering coefficients and degrees, resembling an inverted-V shape. The trend can be

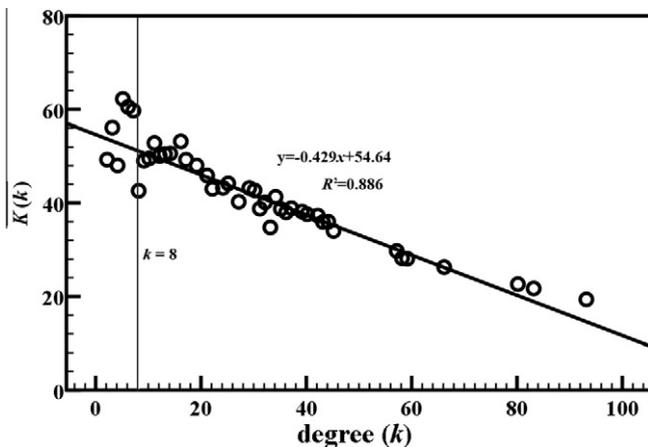


Fig. 5. Degree correlation.

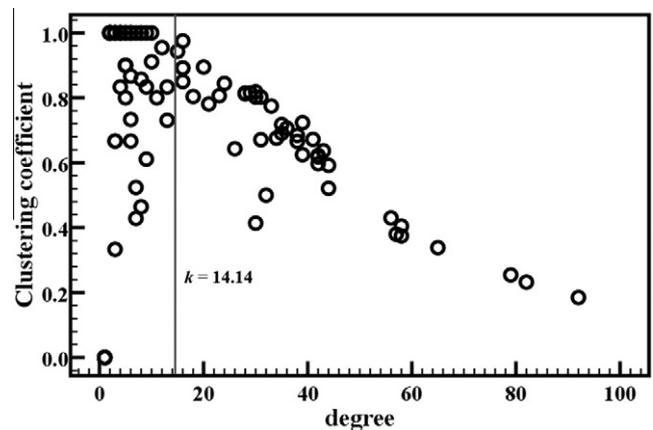
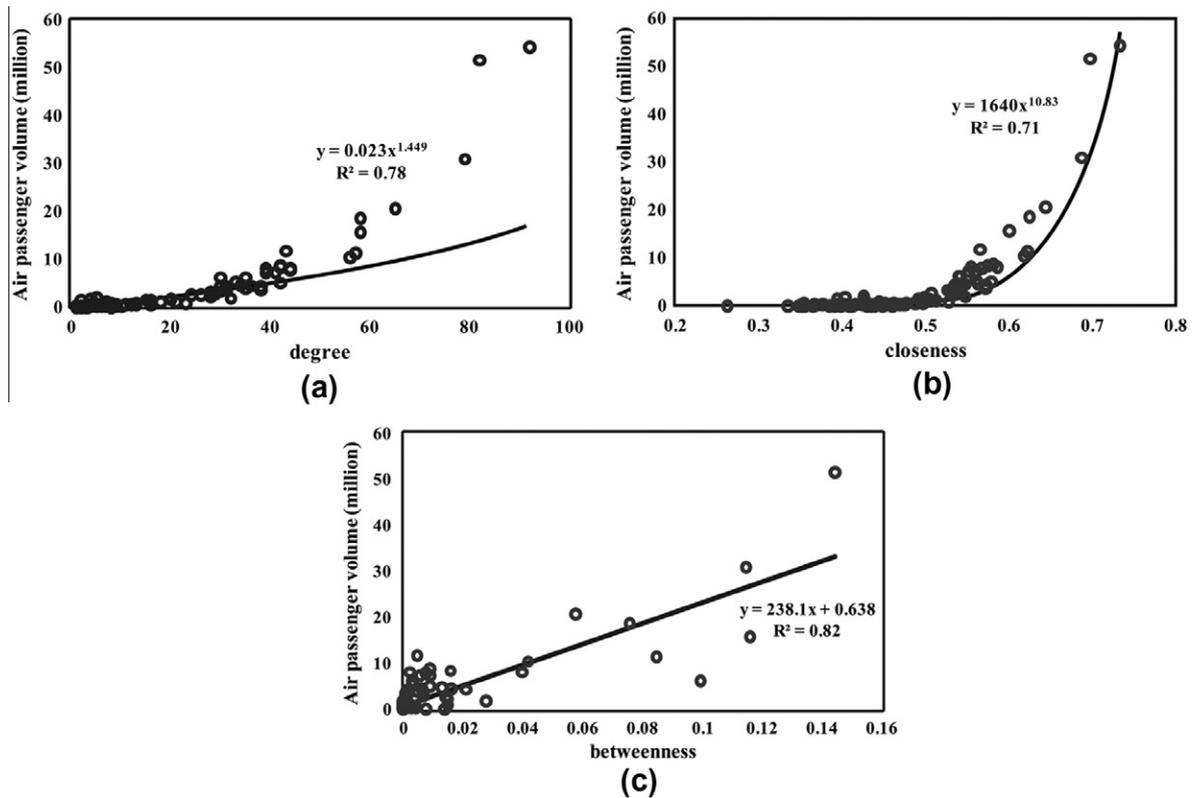


Fig. 6. Correlation between degree and clustering coefficient.

**Table 8**  
Relationship between centralities and the air passenger volume, population, and GRDP.

Correlation coefficient	Degree	Closeness	Betweenness	Air passenger volume	Population	GRDP
Degree	1	0.924 <sup>(a)</sup>	0.766 <sup>(a)</sup>	0.837 <sup>(a)</sup>	0.771 <sup>(a)</sup>	0.793 <sup>(a)</sup>
Closeness		1	0.644 <sup>(a)</sup>	0.708 <sup>(a)</sup>	0.698 <sup>(a)</sup>	0.709 <sup>(a)</sup>
Betweenness			1	0.905 <sup>(a)</sup>	0.750 <sup>(a)</sup>	0.770 <sup>(a)</sup>

<sup>a</sup> Pearson correlation is significant at the 0.001 level (2-tailed) with 144 samples.



**Fig. 7.** Relationship of air passenger volume vs degree, closeness, and betweenness.

captured by two parts: (1) when the degree is below the network average (14.14), the clustering coefficient and degree indices are positively related with a correlation coefficient of 0.41; and (2) when the degree is over the network average, they are strongly negatively related with a correlation coefficient of  $-0.92$ . In the ATNC, 54 cities, i.e., 37.5% of all cities, have the highest clustering coefficient of 1.0, and have low degrees ranging from 2 to 10. In other words, neighbors of (i.e., cities directly connected to) these 54 cities are fully connected among themselves through direct flights. Who are their neighbors? Among the 54 cities, 36 (66.7%) have a direct link to Beijing, 25 (46.3%) to Shanghai, and 23 (42.6%) to Guangzhou. This illustrates the case that low-degree cities tend to directly connect with well-linked cities, which usually have direct connections among each other (i.e., inter-hub air routes). On the other side, after a city passes a threshold of degrees (14 in our case), higher degree (and usually larger) cities tend to be surrounded by lower-degree (and smaller) cities, which are less well-connected among themselves. Therefore, higher degree cities are associated with lower clustering coefficients in this group of cities.

**6.3. Association of centralities with air passenger volume, population and GRDP**

This sub-section examines the relationships between the three centrality indices and the air passenger volume, population, and

gross regional domestic product (GRDP). Table 8 reports the correlation coefficients between them, and shows that centrality indices are all highly correlated with economic indicators of cities such as their air passenger volume, population, and GRDP. As an example, Fig. 7a–c shows the best-fitting trendlines between air passenger volume versus degree, closeness, and betweenness after we account for a variable's scale effect by measuring each variable in its original values or logarithms. Note that the best fitting function is a power function for air passenger volume versus degree (or closeness), but a linear function for air passenger volume versus betweenness. In other words, the air passenger volume in a city increases geometrically with its degree and closeness, and linearly with its betweenness.

**7. Conclusions**

This paper has used complex network theory to examine the overall structure of China's air transport network and the centrality of individual cities. Major findings are summarized as follows.

- (1) The air transport network of China (ATNC) has *small-world* characteristics (like the air networks of the world and other countries such as the US, India and Italy), but is not a *scale-free network* (as is found in the worldwide and Italian air networks). Its degree distribution is best captured by an exponential function, indicating more dominance of large

airports particularly the “big three” (Beijing, Shanghai and Guangzhou) than is found in other statistical patterns such as a power function. These results confirm the links between the air network and the underlying settlement geography of the country.

- (2) The average degree measures the average number of direct links between two cities, the average path length reveals the depth of air transport system, and the clustering coefficient reflects the intensity of interconnectivity of the system. When compared to the US air network, the ATNC has a smaller average degree value, a larger average path length and a smaller clustering coefficient. Thus it seems that the air transport network of China is less mature than that of the US, but results confirm that it is consistent with that found in other emerging economies.
- (3) These outcomes are especially prominent in the measures of centrality indices for individual cities. Among these three indices, degree and closeness are generally consistent with each other but not necessarily with betweenness. All three measures are highly correlated with socio-economic indicators of cities such as air passenger volume, population, and GRDP. This confirms that the overall centrality of the cities in a network, which reflects in turn the spatial pattern of economic activities – captures the crucial aspect of location advantage that moulds an air transport network. The role in the network of cities with high air passenger volumes, such as Nanjing, Hangzhou, Wenzhou, Shenzhen and Ningbo are explainable by their economic power and tourism attraction, while for others such as Urumchi, Kunming, Hohhot, and Guiyang, outcomes are attributable to geographic locations as regional hubs facilitating connection flights. For Beijing, Shanghai, and Guangzhou, their high air passenger volumes are explained by both of the aforementioned factors. Hence the character of the network reflects not only the underlying development of the country but also its distinctive settlement pattern with many large cities in its heavily populated eastern half.
- (4) Like the US, the ATNC is largely disassortative (i.e., higher degree cities surrounded by lower-degree neighbors with direct links, and vice versa), particularly for moderate- and high-degree (i.e.,  $k \geq 8$ ) cities. Such a “shadow effect” in China is partially attributable to cities of lower administrative levels with direct links mainly with cities of higher levels (typically the “big three” and/or provincial capitals). That outcome reflects the influence exerted by Beijing, Shanghai and Guangzhou in particular upon the parts of the country that surround them.
- (5) In general however, unlike the hub-and-spoke system in the US, sub-networks in the ATNC are less developed. The few exceptions are “regional connector hubs” such as Kunming in the southwest and Urumchi in the northwest because of their strategic locations for geographic and political reasons.

In overview, the research shows that the rapid development of the air transport network in China has produced a distinctive pattern. In recent years, new and small airports in China are inclined to supply direct links to the top hubs and so bypass the regional ones, resulting in underdeveloped regional centers. For example, Handan airport opened in 2007 only has direct flights to Shanghai, Guangzhou and Chongqing. Qinhuangdao airport opened in 2006 only has direct flights to Shanghai and Shenzhen. Jixi, Yichun and Daqing, all opened in 2009, only have direct flights to Beijing. This approach to air traffic planning means the hierarchical structure of the air transport system in China identified in the analysis may be re-enforced as it evolves.

## Acknowledgements

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (Grant No. 40635026 and No. 40928001) and the Postdoctoral Science Foundation of China (Grant No. 200902133). We appreciate the detailed and helpful comments by anonymous reviewers and Professor Kevin B. O'Connor that helped us prepare improved versions.

## References

- Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.
- Amaral, L.A.N., Scala, A., Barthélemy, M., et al., 2000. Clases of small-world networks. *PNAS* 97 (21), 11149–11152.
- Anthonisse, J.M., 1971. The rush in a directed graph. Technical Report BN9/71. Stichting Mahtematisch Centrum, Amsterdam.
- Bagler, G., 2008. Analysis of the airport network of India as a complex weighted network. *Physica A* 387, 2972–2980.
- Barabási, A., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
- Barrat, A., Barthélemy, M., Pastor-Satorras, R., et al., 2004. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* 101 (11), 3747–3752.
- Barrat, A., Barthélemy, M., Vespignani, A., 2005. The effects of spatial constraints on the evolution of weighted complex networks. *J. Statist. Mech.: Theory Exp.* 5, P05003.
- Chi, L.P., Wang, R., Su, H., et al., 2003. Structural properties of US freight network. *Chin. Phys. Lett.* 20 (8), 1393–1396.
- Civil Aviation Administration of China (CAAC), 2009. <[http://www.caac.gov.cn/C1/200807/t20080723\\_17530.html](http://www.caac.gov.cn/C1/200807/t20080723_17530.html)> (accessed 02.11.09.).
- Freeman, L.C., 1977. A set of measuring centrality based on betweenness. *Sociometry* 40, 35–41.
- Freeman, L.C., 1979. Centrality in social networks: conceptual clarification. *Social Networks* 1 (3), 215–239.
- Granitsas, A., 2002. Caught in a down draught. *Far Eastern Econom. Rev.* 164, 96–98.
- Guida, M., Maria, F., 2007. Topology of the Italian airport network: a scale-free small-world network with a fractal structure? *Chaos, Solitons Fractals* 31, 527–536.
- Guimerá, R., Amaral, L.A.N., 2004. Modeling the world-wide airport network. *Eur. Phys. J. B* 38, 381–385.
- Guimerá, R., Mossa, S., Turtschi, A., et al., 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *PANS* 102 (22), 7794–7799.
- Haggett, P., Chorley, R.J., 1969. *Network Analysis in Geography*. Edward Arnold, London.
- Hu, H.Y., 1935. The distribution of population in China. *J. Geogr. Soc. China* 2, 33–74 (in Chinese).
- Jin, F., Wang, F., Liu, Y., 2004. Geographic patterns of air passenger transport in China, 1980–1998, imprints of economic growth, regional inequality, and network development. *Prof. Geogr.* 56 (4), 471–487.
- Le, T., 1997. Reforming China's airline industry, from state-owned monopoly to market dynamism. *Transport. J.* 37 (2), 45–62.
- Li, W., Cai, X., 2004. Statistical analysis of airport network of China. *Phys. Rev. E* 69, 1–6.
- Liu, S.J., 2000. Study on reform of air transport in China. *Manage. World* 4, 66–77.
- Ma, X., Timberlake, M., 2008. Identifying China's leading world city: a network approach. *Geojournal* 71 (1), 19–35.
- National Bureau of Statistics of China. *China City Statistical Yearbook*, 2008. China Statistics Press, 2008, Beijing (in Chinese).
- National Bureau of Statistics of China. *China Statistical Yearbook*, 2009. China Statistics Press, 2009, Beijing (in Chinese).
- Newman, M.E.J., 2003. The structure and function of complex networks. *Siam Rev.* 45 (2), 167–256.
- Reggiani, A., Nijkamp, P., 2007. Transport networks and metropolitan development: new analytical departures. *Networks Spatial Econom.* 7, 297–300.
- Reggiani, A., Signoretti, S., Nijkamp, P., et al., 2009. Network measures in civil air transport: a case study of Lufthansa, networks, topology and dynamics. *Theory Appl. Econom. Social Syst.* 613, 257–282.
- Sabidussi, G., 1966. The centrality index of a graph. *Psychometrika* 31 (4), 581–603.
- Shaw, S.L., Lu, F., Chen, J., et al., 2009. China's airline consolidation and its effects on domestic airline networks and competition. *J. Transp. Geogr.* 17, 293–305.
- Statistical Data on Civil Aviation of China, 2008. China Civil Aviation Press, Beijing.
- Taaffe, E.J., 1959. Trends in airline passenger traffic: a geographic case study. *Ann. Assoc. Am. Geogr.* 49 (4), 393–408.
- Wang, J., Jin, F., 2007. China's air passenger transport: an analysis of recent trends. *Eurasian Geogr. Econom.* 48 (4), 469–480.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440–442.
- Xu, Z., Harriss, R., 2008. Exploring the structure of the US intercity passenger air transportation network: a weighted complex network approach. *Geojournal* 73, 87–102.

- Yang, X., Tok, S.K., Su, F., 2008. The privatization and commercialization of China's airports. *J. Air Transp. Manage.* 14, 243–251.
- Yue, T.X. et al., 2003. Numerical simulation of population distribution in China. *Populat. Environ.* 25, 141–163.
- Zhang, A., 1998. Industrial reform and air transport development in China. *J. Air Transp. Manage.* 4 (3), 155–164.
- Zhang, A., Chen, H., 2003. Evolution of China's air transport development and policy toward international liberalization. *Transport. J.* 42 (3), 31–49.
- Zhang, A., Round, D., 2008. China's airline deregulation since 1997 and the driving forces behind the 2002 airline consolidations. *J. Air Transp. Manage.* 14 (3), 130–142.
- Zhang, Y., Round, D., 2009. The effects of China's airline mergers on prices. *J. Air Transp. Manage.* doi:10.1016/j.jairtraman.2009.04.005.
- Zhou, P., Li, Y., 2005. An analysis of the relationship between China's airport distribution and tourist development. *Air Transp. Bus.* 120, 4–9.