

中国航空复杂网络的结构特征与应用分析

陈航宇 李慧嘉

(中央财经大学管理科学与工程学院 北京 100081)

摘要 随着航空运输的经济价值和社会价值的不断提高,航空网络作为航空运输实现的载体,对其网络结构进行研究与分析有着重要的意义。文中以中国主要航空公司的航班数据为基础,运用复杂网络理论,分析中国航空网的网络特性,证实了中国航空网络是具有无标度特性的小世界网络。通过对 2015 年中国航空复杂网络的基本统计特征进行分析发现,其平均路径长度下降,节点平均度增加,而聚类系数逐渐趋于稳定。之后,对中国航空复杂网络的节点指标、边指标及加权指标的相互影响进行统计分析,研究了不同指标变化对网络结构的影响及其现实意义。在研究中国航空网络连接偏好和结构特性的关联性分析中发现,其度度相关、度权相关及介数相关均呈现异配性。最后对研究结果进行应用分析和展望。

关键词 航空网络,复杂网络,统计特征,实证分析,幂律特性

中图分类号 TP393 **文献标识码** A

Analysis of Characteristics and Applications of Chinese Aviation Complex Network Structure

CHEN Hang-yu LI Hui-jia

(School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China)

Abstract With the continuous improvement of the economic and social value of air transport, as the carrier of air transport, the research and analysis of air transport network structure is of great significance. Based on the flight data of major airlines in China, this paper used complex network theory to analyze the network characteristics of China's aviation network, and proved that China's aviation network is a small-world network with scale-free characteristics. By analyzing the basic statistical characteristics of China Aviation Complex Network in 2015, we found that the average path length decreases, the average degree of nodes increases, and the clustering coefficient tends to be stable. After that, the paper analyzed the interaction of node index, edge index and weighted index of China Aviation Complex Network, and studied the influence of different index changes on network structure and its practical significance. In addition, the paper found that the degree-degree correlation, degree-weight correlation and betweenness-betweenness correlation, which reflect the connection preference and structural characteristics of China Airline Network, are negative. Finally, the application analysis and prospect of the research results were carried out.

Keywords Chinese airline network, Complex network, Statistical characteristics, Empirical analysis, Power law distribution

1 引言

随着航空运输的经济价值和社会价值的不断提高,航空网络作为航空运输实现的载体,对其网络结构的实证研究与分析有着重要的意义,能够帮助我们优化网络性能并更加有效地管理网络,同时可以为航空网络演化的相关研究挖掘问题和提供判断依据。

现在,虽然已经有不少研究使用复杂网络统计指标的度分布、平均路径长度、聚类系数、介数、点权和边权等实证分析了中美及世界航空网络,但在选取影响航空网络结构的因素以及分析相关性的实证目的上有许多不同。如 Barrat 等^[1]选择以乘客数量作为边的权重;Cai 小组对一个星期中某一天的航班数进行了标准化,再将其定义为边权^[2-3];为了研究航空网络连接机制,Barrat 等^[1]建立了点权统计指标,并以距离和吞吐量为权重,实证分析了与航空网络演化相关的动态

权重和空间约束的影响;王姣娥^[4]则将城市空间体系加入到研究中,与航空网络结构相关联,以探究它们之间的关系为目的进行实证研究;曾小舟等^[5]引入了机场吞吐量以及航距两个权重因素,通过对 2010 年中国航空网络的实证分析,重点研究了我国国内航空网络结构形成的特性以及网络结构形成中各个指标的影响。

本文先对中国航空网的复杂网络基本统计特征进行了描述,之后引入了边权和点权,实证分析了各个变量之间的相互影响作用,重点突出了研究的现实意义,并论证了中国航空网络的连接偏好和结构特性。相比于已有的中国航空网络实证分析,本次研究采用最新的数据,考虑的因素更多样,相关性分析更全面。最后,进行实证结果的应用分析和展望。

2 样本描述与基本统计特征

本文研究的中国航空网络,是以全国各地主要机场为节

本文受国家自然科学基金项目(71871233,71401194),北京市自然科学基金(9182015)资助。

陈航宇(1998-),男,硕士生,主要研究方向为复杂网络和大数据分析等;李慧嘉(1985-),男,博士,副教授,主要研究方向为数据挖掘、复杂性科学、人工智能等,E-mail:Hjli@amss.ac.cn。

点,以机场间的往返航线为有向边,以机场间往返航线各自每日的航班数为边权,形成的有向加权网络(即两个节点之间有正反两条边,部分节点之间可能只有一条单向边)。其中,航空网络的节点、边、边权的统计以我国几家主要航空公司提供的 2015 年航班时刻表为准。本文中的中国航空网络一共包含了 152 个机场节点和 2556 条有向边。

网络节点的度、平均路径长度、聚类系数、介数和核数是描述复杂网络拓扑结构最基本的统计指标。

2.1 度和度分布

作为网络节点的机场,其度是指与该节点直接连接的节点个数,即与该节点机场通航的机场个数(只有单向通航的机场也计数)。节点的度可记为:

$$k_j = \sum_i a_{ij} \quad (1)$$

若节点 i 与节点 j 之间有边连接,则 a_{ij} 为 1,反之则为 0。通常,节点度值的大小反映了该机场的通达性和在整个航空网络中的重要性。度分布 $p(k)$ 反映了网络节点中度值为 k 的节点所占的比例。本文所统计的样本节点平均度为 16.93,即平均每个机场约与其他 17 个机场有直接的航线连接。图 1 描述了我国航空网络机场节点度值的概率分布,反映出中国航空网络节点度值整体上呈幂律分布^[6]:度值小于 20 的机场节点占机场节点总数的 72.4%;度值大于 100 的机场节点仅占机场节点总数的 1.3%。以上数据说明只有少数机场节点具有较大的度值。

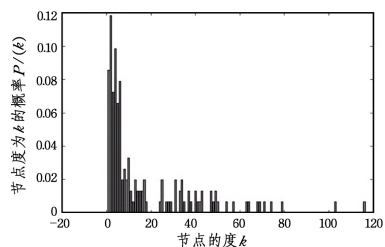


图 1 2015 年中国国内航空网络的概率分布

2.2 平均路径长度

相互连接的两个节点 i 和 j 之间边数最少的路径所包含的边数即为这两个节点间的距离 d_{ij} 。取尽网络中任意两节点的组合,网络的平均路径长度定义为所有组合之间距离的平均值:

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

其中, N 为网络节点数。由于本文所建立的航空网络为有向图模型,节点 i 到 j 的距离 d_{ij} 与节点 j 到 i 的距离 d_{ji} 不同,故本文中网络的平均路径长度 L 的计算式为:

$$L = \frac{1}{N(N-1)} \sum d_{ij} \quad (3)$$

中国航空网络的平均路径长度为 2.0458,意味着任意两个机场之间平均只需要转机 1 次就可以互相到达。

2.3 聚类系数

反映网络节点聚类情况的聚类系数 C_i 是指所有与节点 i 相连的节点之间实际相连的边数占这些点可能的最大连边数目的比例:

$$C_i = \frac{2E_i}{k_i(k_i-1)} \quad (4)$$

其中, k_i 为与节点 i 连接的其他节点的个数; E_i 为这 k_i 个节点间相互连接的边数(两点之间的双向边和单向边都按 1 条边计数)。整个网络的聚类系数 C 定义为网络中所有节点聚

类系数的平均值:

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (5)$$

其中, C 取 1 就说明网络中所有的节点都相连。本文中中国航空网络样本测得的聚类系数为 0.7059,说明了网络的集聚性较强。

具有 152 个节点和 2556 条有向边的中国航空复杂网络,其平均路径长度为 2.0458,聚类系数为 0.7059,与小世界网络特性相符^[7-9]。通过与刘宏鲲、周涛^[9]以及曾小舟等^[5]在分析 2005 年、2010 年中国航空网络数据时得出的结果相比较(如表 1 所列)可知:中国航空网络平均路径长度缩短的同时,机场节点的平均度值在上升,反映了网络的可达性和连通性在增强;聚类系数先下降后渐渐趋于稳定,下降阶段表明了网络中枢层级结构在提高,而渐渐趋于一个常数则说明中国航空网络并非随机形成,该常数聚类系数是中国航空网络的一个特征。这些数据证明,中国航空网络的整体结构在趋好,一个成熟、完善的航空网络体系正在逐步形成。

表 1 中国航空网络 2005 年、2010 年与 2015 年基本结构特征变化比较

	节点数	平均度	平均路径长度	聚类系数
本文(2015 年)	152	16.93	2.06	0.71
曾小舟等(2010 年)	161	14.72	2.14	0.70
刘宏鲲等(2005 年)	121	11.39	2.26	0.75

2.4 介数

网络中包含节点 m 的最短路径的条数定义为节点 m 的介数。在本文的航空网络图中,若节点 i 到节点 j 以及节点 j 到节点 i 的最短路径同时经过节点 m ,则节点 m 的介数计数为 2;若节点 i 与节点 j 之间只有单向最短路径通过节点 m ,则节点 m 的介数计数为 1。机场节点介数的大小反映了机场在整个航空网络中的重要性和影响力,介数越大的节点在网络中的中枢性越强。本文中中国航空网络样本节点的平均介数为 566.57,其中介数为 0 的占 28.9%,介数小于 10 的占 46.1%,介数最大的前 3 个节点介数分别为 14 182,10 840,6 030。由此可见中国航空网络节点介数的分布呈现明显的无标度特征,少数节点的影响力巨大,起到了交通枢纽的作用。

度值表明了机场节点的规模和通达性,介数则表明了机场节点的影响力和中枢性。图 2 所示为基于网络数据所建立的节点介数 g 与节点度值 k 的分布关系。当节点度值小于 25 时,分布关系表现出较大的波动性;当度值较大时,两者之间呈现幂律关系,幂律指数为 3.0406。整体来看,两者之间呈现正相关趋势,说明度值越大,介数往往也越大。

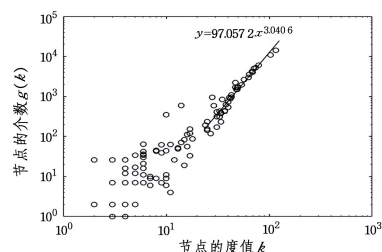


图 2 中国航空网络节点介数与度的分布

2.5 核数

一个图的 k -核是指反复去掉度小于或等于 k 的节点后所

剩余的子图。存在于 k -核的节点,在 $(k+1)$ -核中被除去,则 k 就是该节点的核数。节点核数中的最大值称为网络的核数。节点的核数可以表明节点在核中的深度。本文中中国航空网络节点的平均核数为 9.96,最大值为 25,即网络图的核数为 25。这说明中国航空网络中大部分节点所处的深度较大,不会因为其他节点受到破坏而轻易脱离出网络,整个网络具有较深的层次。本文建立了中国航空网络节点核数与度值之间的关系,如图 3 所示。当度值较小时,两者之间呈现幂律关系,幂律指数分别为 1.4537 和 1.0232;当节点度值大于 25 时,其核数稳定在某个值左右。

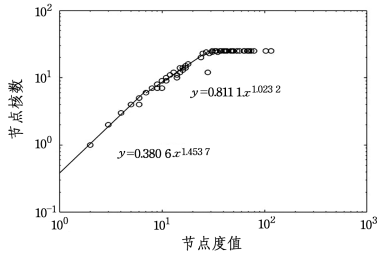


图 3 中国航空网络节点的核数与节点的度值 k 之间的分布

3 度值、介数对网络节点加权重影响的统计分析

3.1 节点加权重与度的统计分布

机场每日客流量总和能够更精确地衡量机场规模。本文中以机场每日航班数量代替机场每日客流量,通过加权重这个指标来描述,加权重定义为:

$$S_i = \sum_{j \in \Gamma_i} w_{ij} + \sum_{j \in \Gamma_i} w_{ji} \quad (6)$$

其中, Γ_i 是节点 i 的临近集, w_{ij} 为节点 i 到节点 j 的航班数量, w_{ji} 为节点 j 到节点 i 的航班数量。加权重反映了该机场的客流量大小。显然,节点的加权重依赖于该点的度值 k 。根据中国航空网络的数据,建立 $S_i(k) \sim k^\beta$ 的关系,如图 4 所示。

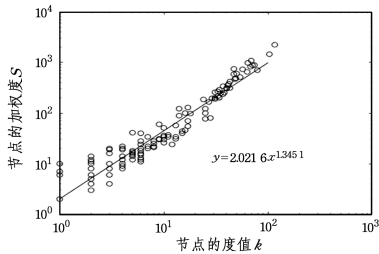


图 4 中国航空网络节点加权重对度的分布

图 4 反映出,中国航空网络度值较小的机场加权重分布差异较大,随着度值的增加,这种差异逐渐减小,度值大于 35 之后,相同度值的节点之间加权重的差别较小。机场加权重与机场规模(度值 k) 呈现正幂律关系,幂律指数 $\beta^k = 1.3451$,这说明机场的度值越大,其客流量往往也越大,且这种正相关性随着度值增加越来越明显(曲线随度值的增加变陡峭)。

3.2 节点加权重与介数的统计分布

节点加权重与介数的统计分布特性,能够揭示机场客流量与机场重要性和中枢性的关系。图 5 为基于本文中网络数据所建立的机场节点加权重与节点介数 g 的分布曲线 $S_i(g) \sim g^\beta$ 。从中可知,机场节点加权重与介数呈正相关性,系数 $\beta = 0.6133$,但没有度值对节点客流量影响那么敏感($\beta^k = 1.3451$)。

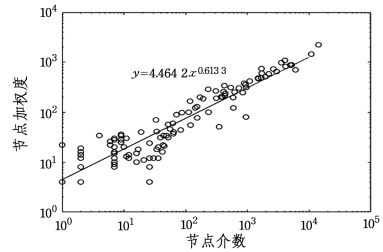


图 5 中国航空网络节点加权重与介数的统计关系

4 度值和介数对网络节点间边权和边介数影响的统计分析

4.1 节点间边权与度值的统计分布

节点间的边权反映了边所代表的航线客流量大小,节点间边权与度值的分布关系可以反映机场规模和通达性对机场所处航线客流量的影响。本文将节点 i 与节点 j 之间双向边的边权加总后作为节点 i 与 j 之间的边权,将节点 i 与 j 的度值加总后作为边权所对应的度值,建立了节点间边权与度值的分布关系,如图 6 所示。图中显示,当节点间边权较小时,边权与度值没有明显的相关关系;当边权较大时,节点间边权与度值呈现正相关性。由此可知,当航线的客流量较小时,航线客流量基本不受其所连接的机场规模大小的影响,当航线客流量超过一定大小后,航线所连接的机场规模越大,航线客流量就越大。

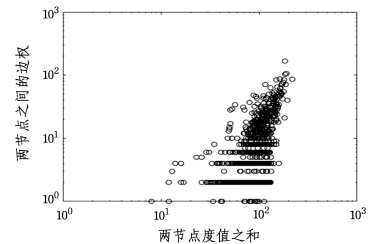


图 6 中国航空网络节点间边权与度值的统计关系

4.2 节点间边介数与度值的统计分布

将网络中包含两节点间边的最短路径的条数定义为两节点间的边介数。边介数反映了航线的中枢性和影响力。节点间边介数与度值的分布关系可以反映机场规模和通达性对机场所处航线中枢性和影响力的影响。本文将节点 i 与节点 j 之间双向边的边介数加总作为节点 i 与 j 之间的边介数,将节点 i 与 j 的度值加总作为边介数所对应的度值,建立了节点间边介数与度值的分布关系,如图 7 所示。

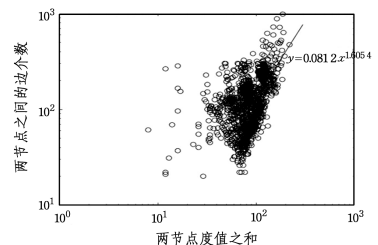


图 7 中国航空网络节点间边介数与度值的统计关系

从图 7 可以看出当两节点度值之和小于 70 时,节点间的边介数波动较大,边介数与度值没有明显的相关关系;当两节点度值大于 70 时,节点间的边介数与度值近似于幂律分布,幂律指数为 1.6054。以上数据说明:当机场节点度值较小时,其所在航线的中枢性和影响力几乎不受机场规模的影响;而

当机场节点度值较大时,机场的规模越大、通达性越好,其所在航线的中枢性和影响力越高。

4.3 节点间边介数与介数的统计分布

节点间边介数与介数的分布关系可以反映机场中枢性对机场所处航线中枢性和影响力的影响。图8反映了中国航空网络节点间边介数与介数的统计关系,两者之间呈现幂律关系,幂律指数分别为0.3747和0.4531,说明航线中枢性和影响力与其连接的机场的中枢性正相关,且随着其所连接机场介数的增大,这种正相关性更加明显;同时,当航线的边介数超过一定数值后,机场节点的度值对航线边介数的影响大于节点介数的影响(边介数与度值的统计分布幂律指数为1.6054),说明机场的规模和通达性对航线中枢性的影响大于机场本身中枢性的影响,即机场连接的其他机场增多,相比于机场的网络中枢性增加,更能使机场所处航线的影响力和中枢性提高。

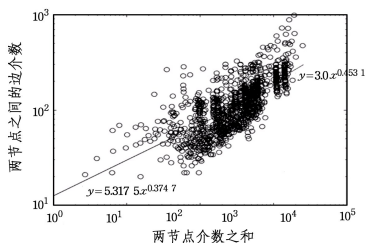


图8 中国航空网络节点间边介数与介数的统计关系

5 关联性分析

本节通过实证分析研究中国航空网络的连接偏好和结构特性,以考查中国航空网络的度度相关性、度权相关性、介数相关性以及簇度相关性。

5.1 度度相关性

因节点度值不同而造成的节点连接时选择倾向于不同的现象为度度相关性。节点的邻点平均度记为:

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j \quad (7)$$

其中, k 为节点度值,节点 j 与节点 i 相连。将网络中节点度都为 k 的所有节点 N_k 的邻节点的平均度 $k_{nn,i}$,取平均,得到度为 k 的节点的邻点平均度:

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i, k_i=k} k_{nn,i} \quad (8)$$

如果 $k_{nn}(k)$ 随 k 增加,即 $k_{nn}(k)-k$ 曲线斜率大于零,则表示度大的节点倾向于连接其他度大的节点,称为度正相关(或同配性);反之,如果 $k_{nn}(k)$ 随 k 递减,表示度大的节点倾向于连接其他度小的节点,称为度负相关(或异配性);如果 $k_{nn}(k)$ 不随 k 变化,是常数,则称节点的度是不相关的^[11]。图9所示为中国航空网络节点邻点平均度 k_{nn} 与度值 k 的分布关系,整体来看其呈现明显的异配性,说明度大的节点倾向于连接其他度小的节点。

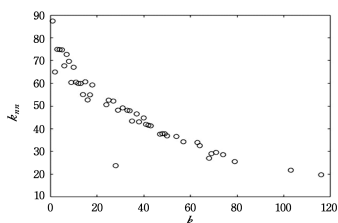


图9 中国航空网络节点邻点平均度 k_{nn} 与度值 k 的分布关系

5.2 度权相关性

影响节点连接选择偏好的因素不只有度值一个,节点的客流量同样影响节点之间的吸引力。本文通过把加权度引入度度相关分析模型的方式来研究这个因素。定义含权的节点 i 的所有邻节点 j 的平均加权度为:

$$k_{nn,i}^w = \frac{1}{k_i} \sum_{j \in V(i)} S_j \quad (9)$$

图10描述了邻节点平均加权度随节点度值 k 以及节点加权度值 S 的变化分布情况。

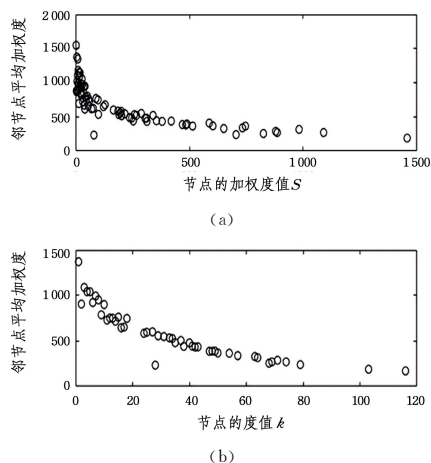


图10 邻节点平均加权度随节点度值 k 以及节点加权度值 S 的变化分布曲线

观察图10可知,整体上看,邻节点平均加权度与节点度值 k 以及节点加权度值 S 呈现明显的负相关性,说明度大的节点偏好连接加权度小的节点,加权度大的节点同样也偏好连接其他加权度小的节点;同时,随着节点度值和加权度值的升高,这种负相关性逐渐减弱,其中节点加权度值与邻节点平均加权度的负相关性减弱得尤为明显,说明机场客流量较小时其客流量对节点之间连接的影响效果显著。

5.3 介数相关性

介数相关性描述的是节点根据介数相互选择的偏好。一个介数为 g 的节点 i ,与 j 节点相连,则节点 i 的邻点平均介数记为:

$$g_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} g_j \quad (10)$$

将网络中节点介数都为 g 的所有节点 N_g 的邻节点的平均介数 $g_{nn,i}$ 进行平均,得到“介数为 g 的节点的邻点平均介数”:

$$g_{nn}(g) = \frac{1}{N_g} \sum_{i, g_i=g} g_{nn,i} \quad (11)$$

中国航空网络节点的邻点平均介数 g_{nn} 与介数 g 的分布关系如图11所示。整体来看,其呈现明显的负相关性,说明介数大的节点偏好连接其他介数小的节点。

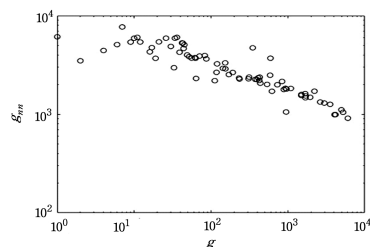


图11 中国航空网络节点的邻点平均介数 g_{nn} 与介数 g 的分布关系

5.4 簇度相关性

节点的邻节点相互连接的集聚程度与节点度值的相关性称为簇度相关性。本文分析了中国航空网络度与平均聚类系数之间的关系,如图12所示。图中节点度值小于25时,节点的平均聚类系数大致稳定在0.8~0.9之间;在度值较大时聚类系数出现了大幅度下降,这说明少数度值较大的机场成为了中国航空网络的中心,网络结构的“轴辐式”特性逐渐形成^[12]。

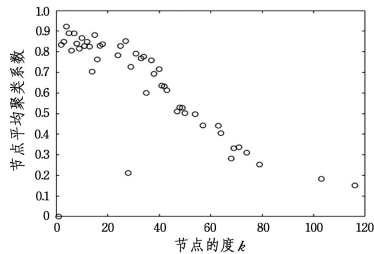


图12 中国航空网络聚类系数与度之间的分布关系

6 应用分析

6.1 针对性防护

中国航空网络的无标度特性使得随机故障不易对网络造成重大影响,但蓄意攻击却能直接对网络造成重大破坏^[13]。因此,筛选出网络中度值和介数等指标较大的节点^[14],加强对对应机场的防护措施,既可以减少意外事故在这些关键机场发生的概率,防止整个航空网络陷入瘫痪状态,又可以针对性打击恐怖组织的蓄意攻击,保护重要机场免遭恶意破坏。

6.2 演化预测

通过对2005年、2010年和2015年中国航空网络的实证分析发现,中国航空网络整体结构在趋好,网络正处于发展和完善的时期。研究航空网络的复杂特性、关联性质,并基于网络特性构建中国航空网络演化模型,仿真航空网络演化发展,探究网络成型的驱动力^[15-16],以为航空网络的安全运营、相关部门的科学管理规划以及航空运输企业制定运输方案提供参考。

6.3 大数据分析

随着数据收集和处理技术的快速发展,现实世界网络的类型和维度呈现出爆炸式增长,如何处理这种百万级甚至亿级数据是当下面临的一个重要难题。近十年间,我国航空旅客运输量年均增速接近15%;航空货邮运输量年均增长率约为10%;具有定期航班的城市和机场均增加了1.5倍;开通的定期航班航线条数也增长了1.5倍。未来对中国航空网络的研究势必将结合当下的大数据技术,使得模型更加有效^[17-19]。

结束语 本文对中国航空复杂网络结构进行了实证分析,实证结果符合无标度的小世界网络特性。航空网络的连接度和通达性不断被优化,网络集聚性趋于稳定,中国航空网络整体结构在趋好。机场的度值和机场的介数与核数表现出正相关性。机场客流量随机场通达性和中枢性的提高而增

大,其中通达性的影响力度更大。航线客流量和枢纽性与航线两头机场的通达性和中枢性表现出一定的正相关性,这种表现随着规模的增大越来越明显。度相关和介数相关均呈现异配性,表明通达性和中枢性高的大规模机场倾向于与其他通达性和中枢性低的小机场连接。

参考文献

- [1] BARRAT A, BARTHELEMY M, VESPIGNANI A. The effects of spatial constraints on the evolution of weighted complex networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(5): P05003.
- [2] LI W, CAI X. Statistical analysis of airport network of China [J]. *Physical Review E*, 2004, 69(4pt2), 046106.
- [3] WANG R, CAI X. Hierarchical structure, disassortativity and information measures of the US flight network [J]. *Chinese Physics Letters*, 2005, 22(10): 2715-2718.
- [4] 王姣娥, 莫辉辉, 金凤君. 中国航空网络空间结构的复杂性[J]. *地理学报*, 2009, 64(8): 899-910.
- [5] 曾小舟, 唐笑笑, 江可申. 基于复杂网络理论的中国航空网络结构实证研究[J]. *交通运输系统工程与信息*, 2011, 11(6): 175-181.
- [6] 崔博. 中国民用航空网络的结构分析与空间集聚问题研究[D]. 南京: 南京航空航天大学, 2014.
- [7] WATTS D J, STROGATZ S H. Collective dynamics of "small-world" networks[J]. *Nature*, 1998, 393: 440-442.
- [8] NEWMAN M E J. Models of the small world: a review[J]. *Journal of Statistical Physics*, 2000, 101: 819-840.
- [9] STROGATZ S H. Exploring complex networks [J]. *Nature*, 2001, 410(6825): 268-276.
- [10] 刘宏巍, 周涛. 中国城市航空网络的实证研究与分析[J]. *物理学报*, 2007, 56(1): 106-112.
- [11] NEWMAN M E J. Assortative mixing in networks[J]. *Physical Review Letters*, 2002, 89, 208701.
- [12] 王姣娥, 莫辉辉. 中国航空网络演化过程的复杂性研究[J]. *交通运输系统工程与信息*, 2014(1): 71-80.
- [13] 王俊超, 殷志远, 冯光柳. 复杂网络特性及可靠性分析-以中国航空网为例[J]. *微型电脑应用*, 2013(11): 13-16.
- [14] 闫玲玲, 陈增强, 张青. 基于度和聚类系数的中国航空网络重要性节点分析[J]. *智能系统学报*, 2016(5): 586-593.
- [15] ROCHA L E C. Dynamics of Air Transport Networks: A Review from a Complex Systems Perspective[J]. *Chinese Journal of Aeronautics*, 2017, 30(2): 469-478.
- [16] 李想. 中国航空网络效率的空间演化及其影响因素研究[D]. 上海: 华东师范大学, 2018.
- [17] 冯慧芳, 柏凤山, 徐有基. 基于轨迹大数据的城市交通感知和路网关键节点识别[J]. *交通运输系统工程与信息*, 2018(3): 42-47.
- [18] 邵峰晶, 孙仁诚, 隋毅, 等. 基于复杂网络的大数据分析[J]. *计算机仿真*, 2018(5): 1-8.
- [19] 窦梅, 孙仁诚, 葛兆富, 等. 公共交通档案大数据有效分析与利用研究[J]. *档案学研究*, 2018(4): 59-61.