

基于传递熵的机场间延误传播关系网络

王娟¹ 郝艳华²

¹(江西机电职业技术学院车辆工程学院 江西 南昌 330013)

²(咸阳师范学院物理与电子工程学院 陕西 咸阳 712000)

摘要 为了刻画机场间延误传播的真实动力学过程并理解延误传播机制,建立基于传递熵的机场间延误传播关系网络并进行实例分析。由航班数据处理得到各个机场的延误时间序列,通过传递熵因果分析方法构建所有机场间的延误传播关系网络,使用复杂网络理论对该网络进行分析。实例分析结果表明,在机场间的日常延误传播交互中,小型机场尤其是4C级机场总是在向外传播延误;对于航空公司而言,较小型航空公司的航空系统更容易发生延误传播。这些分析结果可以为制定延误传播缓解措施提供理论支撑。

关键词 航空运输 延误传播关系 传递熵 航班延误 因果分析 复杂网络

中图分类号 TP391 V335 文献标志码 A DOI:10.3969/j.issn.1000-386x.2024.09.052

DELAY PROPAGATION RELATION NETWORK AMONG AIRPORTS BASED ON TRANSFER ENTROPY

Wang Juan¹ Xi Yanhua²

¹(Department of Vehicle Engineering, Jiangxi Vocational College of Mechanical & Electrical Technology, Nanchang 330013, Jiangxi, China)

²(School of Physics and Electronics Engineering, Xianyang Normal University, Xianyang 712000, Shaanxi, China)

Abstract To describe the real dynamic process of inter airport delay propagation and understand the mechanism of delay propagation, a delay propagation relation network based on transfer entropy is established and analyzed with cases in this paper. The delay time series of each airport was obtained by flight data processing. The transfer entropy causal analysis method was used to construct the delay propagation network among all airports. The complex network theory was used to analyze the network. The results of the case study show that in the daily communication interaction of delay, small airports, especially 4C airports, are always transmitting delay to the outside. For airlines, the systems of smaller airlines are more prone to occur delay propagation. These results provide theoretical support for the formulation of delay propagation mitigation measures.

Keywords Air transportation Delay propagation relation Transfer entropy Flight delay Causal analysis Complex network

0 引言

随着当前民航运输业的快速发展,机场业务量迅速增长,航班延误的问题日益凸显^[1-4]。当上游航班出现延误时,与之共用飞行器或机组人员等资源的下游航班也会以极大概率发生延误,如果没有合理的方法可以控制此类延误传播,那么延误影响范围将会持

续扩大,这不仅会增加航空公司和机场的运营成本,甚至会危害到乘客个人、飞机、机场的安全。因此,对航班延误传播的研究是十分必要的。

到目前为止,航班延误传播领域已经有了较多研究,比如,Beatty等^[5]通过对一个航空公司的航班状态表的分析提出延迟乘子的概念来量化延误传播,代晓旭等^[6]基于传染病模型建立空中交通拥挤传播模型揭示空中交通拥挤传播规律,这些工作主要关注部分航

班、单个机场或航空公司的局部动力学。当前的民航运输系统是一个复杂系统,从复杂网络角度对该系统进行研究可以更全面地掌握其特性。例如,武喜萍等^[7]基于空中交通流量网络参考负荷容量级联失效模型和病毒传播模型构建了空中交通延误传播模型,李航等^[8]基于民用航空网络分析了灾害在网络中的扩散特点,这些工作中构建的网络均是以机场为节点,以真实航线为机场节点间的链接,这样的物理网络可以描述机场间的显式交互,但不能表示机场间延误的真实传播过程。目前仅有少数人尝试从时间序列因果分析角度来解决这一问题,例如,Zanin 等^[9]以及 Du 等^[10]基于格兰杰因果检验方法通过航班延误时间序列发现延误传播动力学表示,构建了机场间延误传播关系网络。但由于当前航空运营场景的复杂性,航班延误时间序列已经具有非线性特点,而格兰杰因果检验这一既定模型只能处理线性时间序列数据^[11]。为了实现非线性时间序列的准确因果分析,Montalto 等^[12]提出了传递熵这一概念,并已广泛应用在医疗和金融等领域。

本文从整个民航运输系统的角度出发,使用传递熵因果分析方法从非线性到港延误时间序列中挖掘机场间延误传播的真实动力学,即机场间延误传播关系。基于所有机场间延误传播关系构建延误传播关系网络,通过复杂网络理论及其相关指标揭示延误传播关系网络的性质,从而发现延误传播场景下存在的实际问题。这有助于理解机场间的延误传播机制,为制定延误传播缓解措施提供理论支撑,对航班延误的改善有指导意义。

1 基于传递熵的延误传播关系网络

1.1 数据描述

本文使用的数据集由中国民航信息网络股份有限公司提供,包含了2016年12月1日至31日所有离港和到港机场均在国内的航班运行状态数据,涉及30余万个航班,200余个机场,30余家航空公司。对于每个航班,可用的信息包括,航空公司 IATA 代码、航班号、飞行器注册号、计划离港时间、计划到港时间、实际离港时间、实际到港时间、离港机场 IATA 代码、到港机场 IATA 代码。

1.2 数据预处理

本文将每个航班的到港延误时长按照机场进行聚合,且以小时为单位时间窗口,而本文的研究对象是航班粒度的机场间延误传播,所以到港延误时长需要针

对航班量进行平均,从而得到平均到港延误时间序列。具体计算方式如下:

$$D^i(d, h) = \frac{\sum T_{\text{act}} - T_{\text{sched}}}{N^i(d, h)} \quad 1 \leq d \leq 31, 1 \leq h \leq 24 \quad (1)$$

式中:所有变量均是基于第 i 个机场第 d 天所有计划到港时间在第 h 小时内的所有航班的数据计算得到; $D^i(d, h)$ 表示所有航班的到港延误时长的平均; T_{act} 表示某个航班的实际到港时间; T_{sched} 表示某个航班的计划到港时间; $N^i(d, h)$ 表示航班总量。那么第 i 个机场到港延误时间序列集合可以表示为:

$$S = \{D^i(1, 1), D^i(1, 2), \dots, D^i(2, 1), D^i(2, 2), \dots, D^i(d, h)\} \quad (2)$$

此时间序列的长度为 $T = d \times h$ 。

Lyapunov 指数是研究混沌的一个重要参数。最大 Lyapunov 指数大于 0,就可判定该系统为混沌的^[13]。依据上述方法计算所有机场 2016 年 12 月整月的到港延误时间序列,接着计算每个序列的最大 Lyapunov 指数,其中的最大值约为 0.16,这表明当前到港延误时间序列是混沌的,即具有非线性特点。

1.3 网络构建算法

本文提供一种基于传递熵的机场间延误传播关系网络构建算法。传递熵是一种非参数统计量,用于测量两个随机过程之间的有向信息传递量。在本文的场景下,有向信息的传递即是延误在机场间的传播,而平均到港延误时间序列对应一个随机过程,因此传递熵可以用于检测两个机场间的延误传播关系。具体如下:

假设 X_t 和 $Y_t (t \in T)$ 分别表示机场 A 和机场 B 的平均到港延误时间序列,令:

$$\Omega = \{X_{T-1}, X_{T-2}, \dots, X_{T-p}, Y_{T-1}, Y_{T-2}, \dots, Y_{T-p}\} \quad (3)$$

式中: p 表示滞后阶数。初始化嵌入向量的集合 $V = \emptyset$ 。计算出一个 $W_1 \in \Omega$ 满足 W_1 与 X_t 的互信息为最大,即:

$$W_1 = \operatorname{argmax}_{W \in \Omega} I(X_t; W) \quad (4)$$

将满足条件的 W_1 嵌入到集合中,即 $V_1 = \{W_1\}$ 。之后继续进行第 k 步嵌入 ($1 < k \leq \dim(\Omega)$),选择一个 W_k 满足:

$$W_k = \operatorname{argmax}_{W \in \Omega \setminus V_{k-1}} I(X_t; W, V_{k-1}) \quad (5)$$

得到 W_k 后,将 W_k 嵌入到集合 V 中,使得 $V = \{V_{k-1}, W_k\}$ 。给出 $I(X_t; W, V_{k-1})$ 的经验分布并通过非参数统计方法检验 W_k ,若 W_k 不满足显著性条件(例如大于 $I(X_t; W, V_{k-1})$ 的 0.95 分位点),则嵌入过程停止,嵌入向量集合 $V = V_k$ 。在嵌入向量集合 V 中减去来自 Y_t 的历史向量 Y' 的贡献, $V' = V \setminus Y'$,计算传递熵:

$$T_{Y \rightarrow X} \equiv H(X_t | V') - H(X_t | V) \quad (6)$$

若传递熵大于零,则 Y_t 对 X_t 具有因果关系,即机场 B 将延误传播至机场 A ;若传递熵等于零,则机场 B 与机场 A 之间没有延误传播关系。

依据上述的延误传播关系发现方法,可以输入 n 个机场对应的延误时间序列从而得到 $n \times n$ 大小的邻接矩阵 A ,其中 $a_{ij} = 1$ 表示机场 i 将延误传播至机场 j ,即存在一条从机场 i 到机场 j 的有向链接, $a_{ij} = 0$ 表示机场 i 未将延误传播至机场 j ,即不存在一条从机场 i 到机场 j 的有向链接。以机场为网络节点,依据邻接矩阵判断节点间是否存在链接,从而得到一个延误传播关系网络。

1.4 网络分析指标

图 1 展示了两个机场的延误时间序列,例如机场 A 每次产生延误时,一定时间内机场 B 也会产生延误,此时表明机场间可能存在延误的传播。我们使用传递熵度量两个机场之间的延误传播关系,之后基于机场间成对的延误关系构建全局的延误传播关系网络。

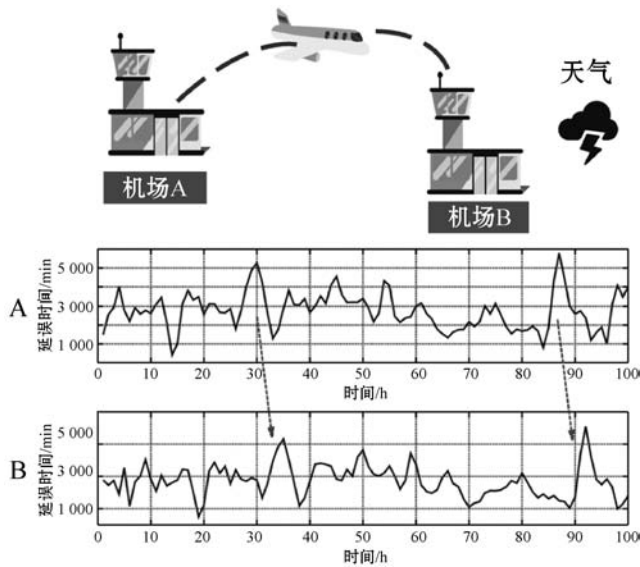


图 1 机场延误时间序列及延误传播关系

复杂网络理论及其相关指标是分析现实世界中复杂系统的重要手段,它可以从多个角度进行复杂系统的分析。当前的民航运输系统包含机场众多,延误传播场景复杂,因此本文使用复杂网络这一工具分析民航运输系统中的延误传播关系。网络分析指标如表 1 所示。

表 1 网络分析指标

网络分析指标	计算式	含义
节点数	N	延误传播关系网络中的机场数目
链接数	M	延误传播关系网络中的链接数目

续表 1

网络分析指标	计算式	含义
入度	$k_i^{\text{in}} = \sum_{j=1}^N a_{ji}$	影响机场 i 的机场数目
出度	$k_i^{\text{out}} = \sum_{j=1}^N a_{ij}$	机场 i 影响的机场数目
度	$k_i = k_i^{\text{in}} + k_i^{\text{out}}$	与机场 i 有延误传播关系的机场数目
平均度	M/N	与某个机场有延误传播关系的平均机场数目
链接密度	$L = \frac{M}{N(N-1)}$	具有延误传播关系的机场对占所有机场对的比例
同配系数	$S = \frac{1}{M} \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij}$	反映度值相近的顶点是否倾向于互相连接
聚集系数	$\bar{C} = \frac{1}{N} \sum_{i=1}^N \frac{T(i)}{k_i(k_i-1) - 2k_i^{\leftrightarrow}}$	反映网络的聚集程度, $T(i)$ 表示通过机场 i 的有向三角形的数目, k_i^{\leftrightarrow} 表示机场 i 的双向链接数
模块度	$Q = \frac{1}{M} \sum_j \left(a_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{M} \right) \delta_{ij}$	反映网络划分为多个社团的强度,若机场 i 和机场 j 在同一社团,则 δ 为 1,否则为 0
效率	$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$	反映延误在机场之间延误传播的容易程度
网络基序	$Z(G') = \frac{F_C(G') - \mu_R(G')}{\sigma_R(G')}$	反映任意三个机场之间的关系模式, G' 表示某个网络基序, $F_C(G')$ 表示原网络 G 中 G' 的频率, $\mu_R(G')$ 和 $\sigma_R(G')$ 分别表示 G' 在多个随机网络的频率平均值和标准差

2 实验分析

2.1 延误传播关系网络全局特点

为了分析全国机场间的延误传播关系,本文依据 1.2 节对 2016 年 12 月整月的航班运行状态数据进行处理得到所有机场的平均到港延误时间序列,作为 1.3 节算法的输入从而得到一个全国机场间的延误传播关系网络。图 2 中包含 205 个机场节点,555 条代表延误传播关系的有向边,其中圆形机场节点的大小与航班量成正比,节点越大表示该机场航班量越大,而菱

形机场节点的大小与节点度成正比,节点越大表示该机场度越大。

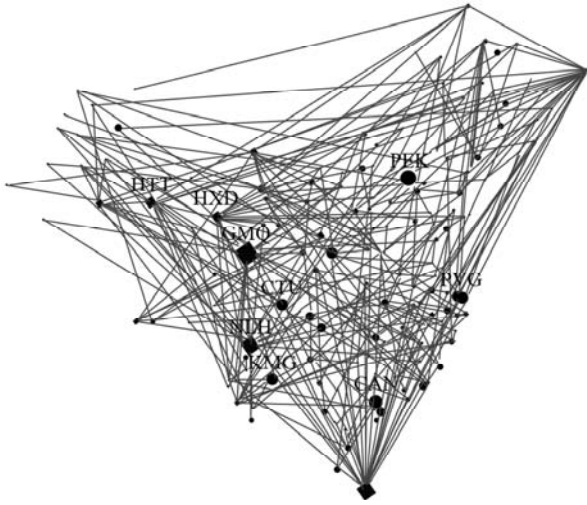
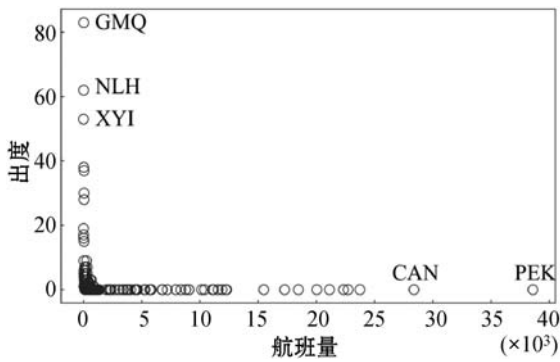
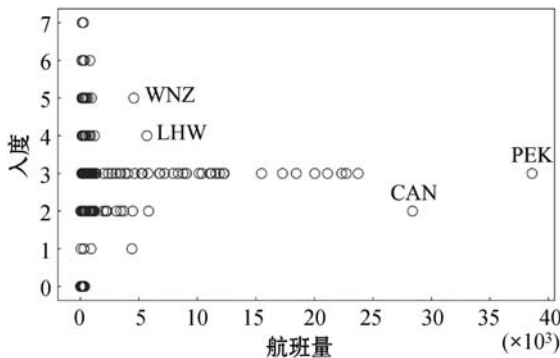


图2 全国机场间的延误传播关系网络

可以看出,相比于度较大的机场节点,经过航班量较大的机场节点的链接并不多。进一步考察机场航班量与机场节点度之间的关系,图3(a)表示机场节点出度与机场航班量之间的关系,可以明显看出两者是有相反关系的,航班量较大的节点反而出度很小,这与在其他航空运输网络中观察到的情况是一致的,例如文献[9]和文献[14]。图3(b)表示机场节点入度与机场航班量之间的关系,可以看到入度与航班量没有明显关系,但大多数航班量较大的机场节点入度仍较小。



(a) 出度与航班量的关系



(b) 入度与航班量的关系

图3 机场出入度与机场航班量的关系

基于以上分析可得,在机场间的日常延误传播交

互中,航班量较小的机场频繁出现,且总是在向外传播延误。出现这种现象的原因可能在于,随着机场航班量的增加,单个航班的延误对机场的影响会慢慢被“稀释”,航班量较大的机场也有着更好的缓解延误的举措和设施,而航班量较小的机场缺少这样的条件,所以在航班运营过程中,航班量较小的机场可能会导致加重延误的传播。

表2列出了此延误传播关系网络的网络指标值,以及由此网络构造的1000个随机网络的指标平均值。构造随机网络的目的是与通过传递熵构建的延误传播网络进行对比,进而了解延误传播的特点,因为随机网络代表了一种平均的网络状态。由平均度约为2.707317可知,每个机场平均与3个机场有延误传播关系,而由链接密度可知,此网络较为稀疏。此网络的平均聚集系数、模块度和效率均高于随机网络的平均值,这表明此网络有较强的聚集趋势、相对明显的社区结构并且延误更容易传播。此网络的同配系数低于随机网络的平均值,这表明该网络的异配性更高,即度值较大的节点倾向于连接度值较小的节点,而由上述可知节点度数与航班量存在相反关系,因此航班量较大的机场与航班量较小的机场之间的延误传播更为明显。

表2 延误传播关系网络指标值

拓扑指标	此延误传播关系网络指标值	1000个随机网络指标平均值
平均度	2.707317	2.707317
链接密度	0.013271	0.013271
同配系数	-0.079135	-0.003809
聚集系数	0.020543	0.011369
模块度	0.626412	0.425472
效率	0.390365	0.315730

2.2 延误传播关系网络局部特点

2.1节对延误传播关系网络进行了整体分析,而本节我们需要关注网络的局部区域。中国民航局为了评估一个机场的定期航班运输服务能力,划定了机场飞行区等级,由高到低分别为4F、4E、4D、4C、3C级机场,其中4C级机场约占全国机场总量的53%。将机场飞行区等级放入延误传播关系网络中,借助网络基序来分析延误传播关系网络中的局部关系模式。本文对2016年12月的每一天均构建延误传播关系网络。图4上方展示的是31个网络中的每一种网络基序的显著性水平,图4下方展示的是网络基序的具体组成,共13种。

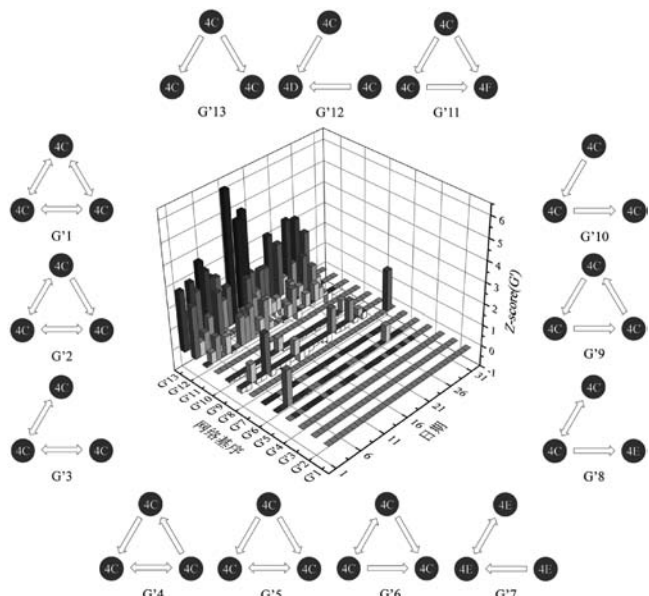


图 4 31 个延误传播关系网络中网络基序的显著性水平

每种网络基序的日变化均可以归因于民航运输系统延误情况的不断变化,但可以明显看到,在大部分时间中 G'11、G'12 和 G'13 的显著性水平均较高,这意味着 4C 级机场可将延误传播至多个机场,并且不只是同级别机场间的延误传播甚至包含 4F 和 4D 级较大型的机场,而机场飞行区等级又与航班量成正相关,因此这种关系模式与 2.1 节中提到的航班量和出度之间的相反关系相吻合,即在机场间的日常延误传播交互中,大型机场总是在接收延误从而缓解延误传播,小型机场尤其是 4C 级机场总是在向外传播延误从而导致或加重延误传播。G'5、G'7 和 G'9 均是在不同的某些天显著性水平较高,可能是由于这些天在不同区域发生了大面积延误事件,不同的机场发生延误,造成不同的网络基序的显著性水平增大,其中包括 4C 级机场之间的延误双向传播、4E 级机场之间的延误传播、4C 级机场之间的延误级联传播。对于 G'8,可以明显看到某些天的显著性水平较高,而某些天的显著性水平是小于零的,表明延误情况的变化可显著影响该网络基序的显著性水平。

2.3 航空公司延误传播关系网络特点

民航运输系统由多家航空公司参与运营,本文前面进行的分析均未区分航空公司,而本节将会按照航空公司对全国延误传播关系网络进行分析。具体而言,本文通过航班量将数据集中涉及的航空公司进行排序,取出航班量前十的航空公司,如图 5 所示,其中 CZ、MU 和 CA 是国内的三大航空公司,其他航空公司的规模相对较小。以这十家航空公司为例,基于各个航空公司整月的航班运行状态数据得到各个航空公司

延误传播关系网络,对比分析这十个航空公司的延误传播情况之间的差异。

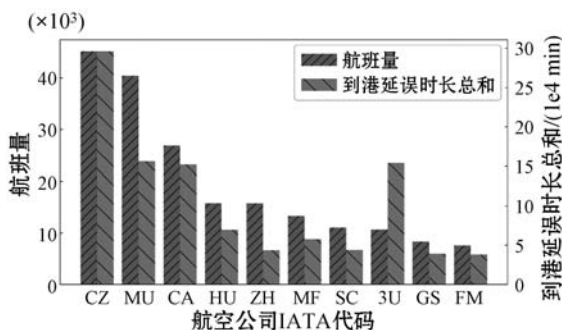


图 5 航班量前十的航空公司及其对应的到港延误时长总和

理论上,一个航空公司航班越多,发生延误传播的概率就越大,而图 5 展示的 3U 航空公司虽然航班量较小,但其到港延误时长总和却与 CA 航空公司相差无几。图 6 中展示的各个航空公司的延误传播关系网络中的节点数和链接数出现了类似的现象,3U、GS、FM 三个航班量最小的航空公司对应的延误传播关系网络的节点数和链接数却不是最小的,其中仍是 3U 航空公司涉及的机场数较多,即延误传播范围比其他较小型航空公司要大。

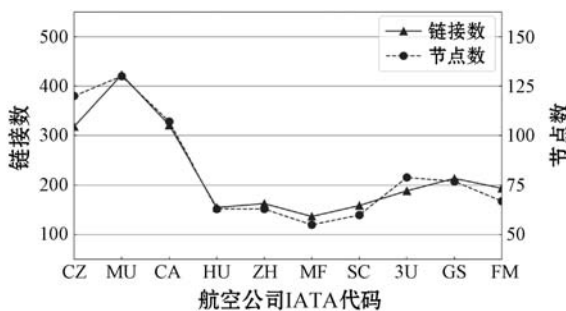


图 6 各个航空公司的延误传播关系网络中的节点数和链接数

图 7 展示的是各个航空公司的延误传播关系网络的同配系数和效率,结合图 6 可以明显看到三大航空公司与其余较小型航空公司的区别。由于三大航空公司通航城市和航班量均较多,使得可能的延误传播范围也较大,所以对应网络的链接数和节点数均会远远大于其余的航空公司。而在延误传播的效率方面,三大航空公司又低于其余的航空公司,尤其以 MF、SC 两个航空公司效率最高,这意味着较小型航空公司的航空系统更容易发生延误传播,而对于三大航空公司来说,可能因为网络规模较大或者拥有更好的缓解延误传播的资源使得延误传播较为困难。从图 7 的同配系数可以看出,较小型航空公司拥有较小的同配系数,依然是以 MF、SC 两个航空公司最小,这意味着较小型航空公司中度值相差较大节点倾向于相互连接,网络呈现星状,这同样使得延误在较小型航空公司的航空系统内快速传播。

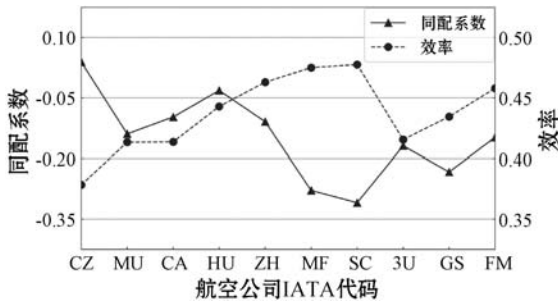


图7 各个航空公司的延误传播关系网络的同配系数和效率

3 结语

本文提出一种基于传递熵的机场间延误传播关系网络构建方法,该方法可以从当前非线性到港延误时间序列中挖掘到机场间延误传播的真实动力学表示。接着利用复杂网络理论及其相关指标从整个国内民航运输系统的角度分析了机场间延误传播机制,得到的结果可为当前航班延误的改善提供参考。分析发现小型机场尤其是4C级机场,总是导致或加重延误传播,由此机场管理者可以着重考虑改善4C级机场的基础设施以及设定更好的延误缓解策略。其次发现较小型航空公司的航空系统延误传播效率高,这就需要航空公司优化当前的航班排班和资源分配。

参考文献

- [1] 何月娇. 基于状态机的民航飞行电报数据并发处理模型[J]. 计算机应用与软件, 2018, 35(6): 114-116, 162.
- [2] 徐涛, 刘泽君, 卢敏. 基于RBM-BPNN的民航潜在高价值旅客预测[J]. 计算机应用与软件, 2019, 36(9): 58-63.
- [3] 徐涛, 孙媛媛, 卢敏. 基于灰色神经网络的航线客流量预测[J]. 计算机应用与软件, 2020, 37(1): 31-36.
- [4] 王红, 李浩飞, 邸帅. 民航突发事件实体识别方法研究[J]. 计算机应用与软件, 2020, 37(3): 166-172.
- [5] Beatty R, Hsu R, Berry L, et al. Preliminary evaluation of flight delay propagation through an airline schedule[J]. Air Traffic Control Quarterly, 1999, 7(4): 259-270.
- [6] 代晓旭, 胡明华, 田文, 等. 利用传染病模型研究空中交通拥挤传播规律[J]. 交通运输系统工程与信息, 2015, 15(6): 121-126.
- [7] 武喜萍, 杨红雨, 韩松臣. 基于复杂网络的空中交通特征与延误传播分析[J]. 航空学报, 2017, 38(S1): 113-119.
- [8] 李航, 胡小兵. 一种改进的民用航空网络空间脆弱性模型[J]. 交通运输系统工程与信息, 2018, 18(4): 202-208.
- [9] Zanin M, Belkoura S, Zhu Y. Network analysis of Chinese air transport delay propagation[J]. Chinese Journal of Aeronautics, 2017, 30(2): 491-499.
- [10] Du W, Zhang M, Zhang Y, et al. Delay causality network

in air transport systems[J]. Transportation Research Part E: Logistics and Transportation Review, 2018, 118: 466-476.

- [11] Maziarz M. A review of the granger-causality fallacy[J]. The Journal of Philosophical Economics: Reflections on Economic and Social Issues, 2015, 8(2): 86-105.
- [12] Montalto A, Faes L, Marinazz D. MuTE: A MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy[J]. PloS one, 2014, 9(10): e109462.
- [13] 孙海云, 曹庆杰. 混沌时间序列建模及预测[J]. 系统工程理论与实践, 2001(5): 106-109, 113.
- [14] Andrewa C, Grahama T, Massimilianob Z. Towards superior air transport performance metrics-imperatives and methods[J]. Journal of Aerospace Operations, 2013, 2(1): 3-19.

(上接第347页)

- [3] 徐朝农, 吴建雄, 徐勇军. 时延有界的PD-NOMA物联网高可靠接入算法[J]. 通信学报, 2020, 41(9): 210-221.
- [4] 吕海斌. 基于边缘计算的多用户计算卸载方法研究[D]. 哈尔滨: 哈尔滨工程大学, 2019.
- [5] 危泽华, 曾玲玲. 基于Stackelberg博弈论的边缘计算卸载决策方法[J]. 数学的实践与认识, 2019, 49(11): 91-100.
- [6] 裘华东, 张燕, 涂莹. 移动云计算嵌套式两阶段博弈计算卸载算法[J]. 科学技术与工程, 2018, 18(32): 58-63.
- [7] 田贤忠, 姚超, 赵晨, 等. 一种面向5G网络的移动边缘计算卸载策略[J]. 计算机科学, 2020, 47(S2): 286-290.
- [8] 季子豪, 江凌云. 一种基于遗传算法的多站点协同计算卸载算法[J]. 计算机工程与科学, 2021, 43(3): 426-434.
- [9] 彭军, 王成龙, 蒋富, 等. 一种车载服务的快速深度Q学习网络边云迁移策略[J]. 电子与信息学报, 2020, 42(1): 58-64.
- [10] 李梓杨, 于炯, 王跃飞, 等. Flink环境下基于负载预测的弹性资源调度策略[J]. 通信学报, 2020, 41(10): 92-108.
- [11] 贾嘉, 慕德俊. 基于人工蜂群的云计算负载均衡算法[J]. 科学技术与工程, 2020, 20(16): 6532-6537.
- [12] Chen Z, Wang X D. Decentralized computation offloading for multi-user mobile edge computing: A deep reinforcement learning approach[J]. EURASIP Journal on Wireless Communications and Networking, 2020, 55(1): 123-144.
- [13] Hong Z C, Huang H W, Guo S, et al. QoS-aware cooperative computation offloading for robot swarms in cloud robotics[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 4027-4041.
- [14] Asheralieva A, Niyato D. Learning-based mobile edge computing resource management to support public blockchain networks[J]. IEEE Transactions on Mobile Computing, 2019, 55: 1156-1169.
- [15] Chen X F, Zhang H G, Wu C, et al. Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning[J]. IEEE Internet of Things Journal, 2019, 26(3): 4005-4018.