Case Study

Global air transport complex network: multi-scale analysis

Weisi Guo[1,5] · Bogdan Toader[2] · Roxana Feier[2,5] · Guillem Mosquera[3,5] · Fabian Ying[2] · Se-Wook Oh[2] · Matthew Price-Williams[4] · Armin Krupp[2]

Abstract
Almost half of the world's population is carried by airlines each year, and understanding this mode of transport is important from economic and scientific perspectives. In recent years, the increasing availability of data has led to complex network and agent interaction models which attempt to gain better understanding of the air transport network and develop forecasts. In this case study paper, we review existing research on two key approaches, namely: (1) a top-down multi-scale network science approach, and (2) a bottom-up entropy-maximization interaction network approach. Using simple socioeconomic indicators, we were able to construct a very accurate interaction model that can predict traffic volume, and the model can forward estimate the impact of population growth or fuel cost. Using network science approaches, we were able to identify community structures and relate them to economic outputs. We also saw how hubs evolved over time to become more influential. Looking into the future, using random graph theory, it seems that reduced flight cost will lead to increased hub influence. The disseminated knowledge in this case study paper will provide both academics and industry practitioners with steps forward to co-explore the interesting research landscape.

Keywords Air transport network · Complex network · Spatial interaction

1 Introduction

Air transport networks are complex networks that span across multiple distance scales (from a few km to 10,000 km) and multiplex together over 5000 airline operators and has strong inter-dependencies with socioeconomic drivers. The air transport network carry 3.5 bn passengers per year and generate over 30 m jobs globally. The analysis of air transport networks to better understand its network properties goes back for over 10 years [1–4]. Both global and regional studies have explored their complex network structure across different network scales [5–7] with multi-layer analysis [6, 8]. The analysis predominantly focus on robustness from attacks or failures [9, 10], efficiency [4], and structural evolution [7]. The air transportation network is also responsible for the propagation of knowledge and culture [11], infectious diseases [12–16], and understanding the network properties allows scientists to better estimate the intangible benefits and risks of global transportation. A detailed review of existing literature will be given in each relevant section of analysis in this paper.

1.1 Case study outline

This paper summarizes an intense collaboration project between Airbus (industrial practitioners) and academics that bring in new complexity methodologies to add new knowledge value. The goal is to review and explore the network science and interaction modelling methods that can be used to gain fundamental understanding into the complexity of air transport networks.

Two fundamental approaches are tackled in this review:

✉ Weisi Guo, wguo@turing.ac.uk | [1]School of Engineering, University of Warwick, Coventry CV4 7AL, UK. [2]Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK. [3]Mathematical Institute and the Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK. [4]Department of Mathematics, Imperial College, London SW7 2AZ, UK. [5]The Alan Turing Institute, London NW1 2DB, UK.

SN Applied Sciences (2019) 1:680 | https://doi.org/10.1007/s42452-019-0702-2

Received: 21 April 2019 / Accepted: 3 June 2019 / Published online: 8 June 2019

- Bottom-up entropy-maximization interaction model, which considers consumer choice;
- Top-down network science analysis, which seeks to uncover common statistical patterns and infer latent knowledge.

The former gives a complex and detailed understanding of how spatial networks (i.e., flights) form from spatial processes (i.e., airports) and what the weight of each edge (i.e., passenger volume) is with respect to cost (impedes flow) and benefit (attracts flow) functions that relate to consumer behaviour. The latter approach gives a statistical understanding into the fundamental network properties and how they evolve over time, enabling the application of generalized network scaling laws that can be used to predict the future structure of the network. Both the bottom-up and the top-down approach is of fundamental interest to network science and industry.

1.2 Data availability and network construction

Several air transport network data sources are available from academic and commercial databases. One of the most widely used commercial databases is the purchased OAG data. This case study paper will use a single month's sample in the year 2015, as well as open air transport data obtained from the US Bureau of Transportation Statistics to demonstrate results. The spatial resolution of the data includes 9000 global airports, each geo-tagged with coordinates, and the temporal resolution of the data are every civilian flight (dis-including cargo flights). Compared to open data, the purchased data from OAG offers a more comprehensive list of flights as well as passenger volume and flight class distribution (e.g. between first, business, and economy).

In order to construct a network from the data, airports are represented by nodes and flights are represented by weighted links. The vast majority of work uses regular scheduled flights and the seat number of each flight is used as a weight for the link. True passenger numbers (load) are commercially sensitive and cannot be obtained on a global scale. Each node, if connected to another, is usually a bi-directed connection with equal weighting (i.e., most flights transverse back and forth). When multiple flights exist between two airports, the total weight is the sum of the seats available. An example of the network is shown in Fig. 1.

1.3 Key industrial problems and interest

Industrial practitioners range from aircraft manufacturers to airline operators. Of fundamental interest to both parties is the future of airline routes, both in terms of their spatial patterns (including multi-hop routes) and their demand intensity (including temporal fluctuations). Understanding these patterns allows aircraft manufacturers, such as Airbus, to design future aircraft, which may take up to 20 years and are required to operate for another 30 years. Generally speaking, the problems posed by industrial practitioners can be broken down into the following:

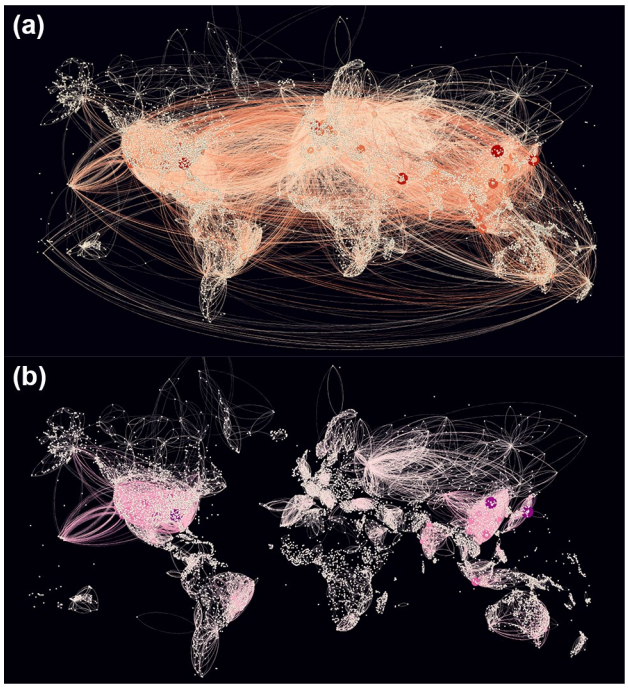

Fig. 1 Complex network of city nodes (airports) with directed and weighted air transport links. Node size reflects weighted degree and link line width indicates number of seats per month. **a** Global network comprises of 9033 nodes and 101042 links. **b** A number of domestic sub-graphs which comprises of 9032 nodes and 53496 links

- How do we predict the passenger flow capacity of existing routes?
- What are the vulnerable points in the network that can help prioritize redundancy and security [16]?
- How can we categorize air transport networks for different airlines to define their business model?
- How can socioeconomic data help to understand the future of the network?

Several resolutions are of interest, such as: airline business model (i.e., legacy, budget, regional, international), operational model (i.e., point-to-point, hub–spoke), geographic region (i.e., developed country, holiday destinations), time-span (i.e., post-disaster, post-merger), and flight range (i.e., long-haul).

1.4 Organisation

In Sect. 2, we give a literature review of bottom-up approaches such as spatial interaction models that have been applied to different transport scenarios. Focus will be on both pair-wise models such as the gravity law and the radiation model, as well as the Boltzmann-Lotka-Volterra (BLV) competitive interaction model [17]. A small-scale test case of its application to the air transport network will be given.

In Sect. 3, we give a review of top-down network science analysis on the air transport network. At the macroscopic level, we focus on degree distribution and centrality correlation measures to detect certain airport properties, as well as small-world network structures and implications on network resilience to failures. At the mesoscopic level, we will focus on how community detection, core–periphery profiling, and other methods can be used to identify network motifs such as hub–spoke structure to help industry understand the network better and design future aircrafts. Relationship with socioeconomic parameters will also be reviewed and analysed.

In Sect. 4, we review work on random graph models and how generic distance and hop-distance cost functions can be used to change the network structure (i.e., from random geometric graphs to random graphs). We use these cost functions to hypothesize on how the network structure can evolve and what it means for the business model of aircraft designers.

In the last section, we summarize the bottom-up and top-down approaches and how future researchers can move forward in this area to better understand the science of air transport networks.

2 Bottom-up approach: spatial interaction models

2.1 Pairwise models

Pairwise models are free from any global constraints (i.e., finite network commuter capacity bounded by total population), and as such have low computational complexity.

2.1.1 Gravity law

One method to measure flow is the widely used *gravity law* to infer the volume of flow between any two given cities [18]. The gravity law has been employed in various forms for over a century [19, 20], but as with many such laws, its theoretical underpinning comes in many forms (see below). Gravity laws generally describe the attractive force between two entities and has been used to describe to flow of a wide variety of goods (e.g. vehicles, goods, disease, and human beings) [21–24] and information (e.g. telephone calls and social media messages) [25–27] between cities and countries. The law consists of three main parameters: the weights of the two nodes (i.e., population P) and the rate of decay dependent on their Euclidean separation distance d. Continuing with the flow model used previously, the number of trips from location i to location j:

$$F_{ij} \propto P_i^{\alpha} P_j^{\beta} f(d_{ij}). \tag{1}$$

where $[\alpha, \beta]$ are parameter exponents and the function of distance $f(d)$ can take on many forms depending on the context of application. In the most classical gravity law case, the form of $f(d)$ is generally d^{-2}.

A thorough review of gravity laws and complex networks can be found in [18]. Almost all research will agree that population determines the flow of goods or people [28, 29]. The discrepancy between different models lies in what form the gravity law takes, especially for the distance function $f(d)$, and the parameters that weight the population, i.e., α and β in Eq. 1. Two studies in particular stand out as examples of global trade or good exchange [12, 22]. For air travel, one of the largest flow studies examined worldwide commuter traffic [22]. It was found that the travel pattern conformed to the following gravity law with a distance function $f(d_{ij}) = \exp(-d_{ij}/\kappa)$. For below 300 km, the nodes were asymmetrically weighted (i.e., directed links): $[\alpha = 0.46, \beta = 0.64, \kappa = 82]$. This is perhaps accounted for by travelling between home and work. For over 300 km, this study and many others like it found that flow is nearly symmetric (i.e., undirected) and the parameters are: $[\alpha = 0.35, \beta = 0.37]$. Other similar air traffic studies indicated that for long distance the values for population weighting are: $\alpha = \beta = 0.5$ [12]. One challenge with the gravity model is its sensitivity to parameterization and the following models overcome this.

2.1.2 Radiation model

Inspired by the gravity model, the radiation model [24] has recently been proposed to overcome all the aforementioned limitations. Using mobile data from commuters (traveling from home to work), in [24] the authors show that the flux is independent of key parameters in the job market, namely: (1) benefits of the job, (2) the number of jobs available at the location, and (3) the number of people N_c. Hence, unlike the gravity model, the radiation model is *parameter-free*. The average flux $\langle F_{ij} \rangle$ is predicted by:

$$\langle F_{ij} \rangle = F_i \frac{P_i P_j}{(P_i + s_{ij})(P_i + P_j + s_{ij})}, \tag{2}$$

where $F_i = P_i(N_c/N)$ denotes the total number of commuters transferring from i to j, and N is the total number of people in the country. The parameter s_{ij} denotes the population within a circle of radius r_{ij} that is centred around the location i (catchment area).

In general, these pairwise models only consider the attributes of nodes i and j, and do not bound the overall system with energy constraints that would otherwise capture some kind of competitive decision making process. Perhaps, the local studies (i.e., within cities or countries) do not need to consider a high degree of competition, but these models cannot be and have not been generalized to larger networks.

2.2 Multi-point entropy maximization models

Pairwise models suffer from the lack of competition between nodes [30, 31]. As such, they tend to work for non-competitive interactions and cannot accurately describe the competitiveness nature of the global air transport industry. Multi-point models consider all possible flows simultaneously and attempt to discover the most likely combination.

2.2.1 Boltzmann–Lotka–Volterra (BLV) formulation

We now review the BLV model [17], which has been applied to a wide range of competitive scenarios, such as financial spending patterns in shopping centres. The BLV model has the potential to predict the flow between different nodes of the network, given data related to the cost and the benefit of having flights between the airports. As such, it can test hypotheses related to the impact of changing costs and passenger benefits. Given a fixed number of spatial points (i.e., airports), there are a finite number of route configurations. Entropy in a spatial configuration context can be defined as the likelihood of forming certain combination of links. This is the foundation to the BLV model. The formulation is pinned on the maximizing the number of micro-states in the network (a term from statistical physics), which gives the most likely flow pattern [17]:

$$W(F_{i,j}) = \frac{F!}{\prod_{i,j} F_{i,j}!}, \tag{3}$$

where the weighted flow between two arbitrary nodes is $F_{i,j}$, and F is the total flow in the system. By taking logs and using Stirling's approximation, the above is equivalent to maximizing the Shannon entropy S in the system:

$$S = -\sum_{i,j} F_{i,j} \log(F_{i,j}). \tag{4}$$

At this point, the generic spatial interaction model needs to define clear benefit and cost functions. Constraining the link weights based on cost functions $c_{i,j}$ (i.e., distance and fuel cost), benefit functions b_j (i.e., attractiveness of destination city Z_j), and fundamental limits (i.e., total capacity of airports X_i), the most likely passenger flow $F_{i,j}$ can be found with Lagrange multipliers (α, β, γ). The general form of the predicted flow is given as [17]:

$$F_{i,j} = X_i \frac{\exp(\alpha b_j - \beta c_{i,j})}{\sum_k \exp(\alpha b_j - \beta c_k)}, \tag{5}$$

where the Lagrange multipliers are optimisation parameters that weigh each benefit and constraint. The benefit and constraint functions are given below for particular regional case studies.

2.2.2 Case study: Australia domestic network

Due to the vast computation required to consider a global or even a large regional air transport network, we consider an isolated and small domestic network such as Australia. In particular, we select the 5 largest airports: Sydney, Melbourne, Brisbane, Adelaide, and Perth. Given known data on the flow between these airports and the associated city data, we need to assume right cost-benefit functions and corresponding Lagrange multipliers. Reasonable assumptions based on existing literature can be developed for the cost-benefit functions. We assume that the decision of having flights between two airports only depends on the population P and separation distance d of the nearest cities.

Benefit function Preliminary results show that the number F_j of passengers flying to a city j has positive correlation with the population P_j of the city j. From that, we construct two ways to calculate b_j which are in line with gravity law equations [24]:

1. Model 1: $b_j = \log P_j$,
2. Model 2: $b_{i,j} = \log(P_i P_j)$

We use both benefit models to predict $F_{i,j}$, then we compare the predicted result and real data. We also compare the dynamics of $F_{i,j}$ with the increase and decrease of populations P_i of each city i by time.

Cost function Distance as a cost usually appears as a gravity law or exponential form, which is used in transportation cost functions [19, 20, 22, 24].

1. $F_{i,j} \propto d_{i,j}^{-a}$, where $a = 1, 2$ is similar to gravity or radiation law land-based travel models.
2. $F_{i,j} \propto \exp(-d_{i,j})$, where this exponential form is similar to Levy flight movement of birds and low friction systems.

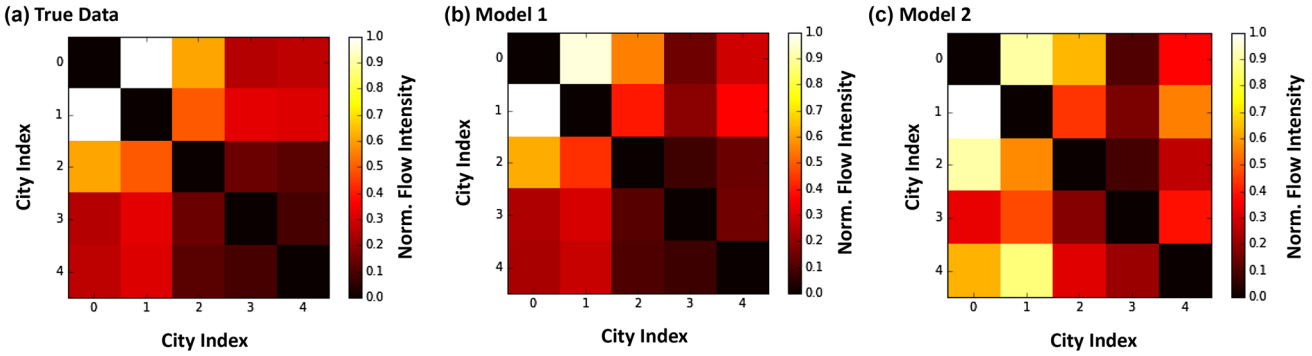

Fig. 2 Using entropy-maximization BLV model to predict passenger flow volume. Passenger flow **a** obtained from real data, **b** predicted by assuming $b_j = \log P_j$, and **c** predicted by assuming $b_{i,j} = \log(P_i P_j)$

In our study, we assume cost is linearly dependent on the distance: $c_{i,j} \propto d_{i,j}$, such that the flow is proportional to the exponential form of the distance $F_{i,j} \propto \exp(-d_{i,j})$ [see Eq. (5)]. In order to find the Lagrange multipliers α, β and γ such that the outputs $F_{i,j}$ of our model fit the flights data, we minimize the norm of the residual relative to the true flow data $F_{i,j_{\text{true}}}$:

$$f(\alpha, \beta, \gamma, z_s) = \|F_{i,j_{\text{true}}} - F_{i,j}\|_2 + \lambda \|\text{diag}(F_{i,j_{\text{true}}}) - \text{diag}(F_{i,j})\|_2, \quad (6)$$

where the second term enforces that the diagonal of the output is zero for fixed $\lambda > 0$. Since this is a global optimisation problem with a *non-convex* objective function, one cannot achieve perfect convergence, but the output of our calibrated model gives good relative number of flights between different airports. This means that we need to adjust the output of our model by multiplying the results by $C = \frac{M_{\text{data}}}{M_{\text{model}}}$, where M_{data} is the true maximum number of flights and M_{model} is the predicted maximum number of flights between any two of the five cities in our dataset.

Results Figure 2a shows the passenger flow F where each element F_{ij} is the flow of passenger from airport i to airport j. Each row and column corresponds to five different airports that are ordered from Sydney (SYD), Melbourne (MEL), Brisbane (BNE), Perth (PER), and Adelaide (ADL). Because each passenger flow is normalised, $F_{i,j}$ ranges from $0 \leq F_{i,j} \leq 1$. Figure 2b is the predicted passenger flow based on the benefit function as model 1: the benefit for a passenger to fly to airport j is $b_j = \log P_j$. In Fig. 2c, we use model 2 (i.e., $b_i = \log P_i P_j$) to calculate benefit function, and obtain the result. As one can observe, model 1 shows better agreement with the Australian air transportation data than model 2, with model 1 yielding an aggregated normalised flow intensity difference of 0.7 compared to those of 2.6 for model 2.

2.2.3 Future scope for research

The BLV model [17] has the advantage of finding the entropy-maximization solution to a competitive network flow problem, including the temporal dynamics. However, the non-convex nature of the BLV model means that unless there is native intuition on the benefit and cost functions (e.g. based on established studies), then discovering the correct function form and the parameters is costly. Nonetheless, the BLV model has been applied successfully to complex challenges in urban retail, mobility [30], and policing.

During this brief analysis of how the BLV model can be used to predict future passenger flows (flights), the benefit function depends only on the population of the cities where the airports are located (destination or both), and we modified the input of the model manually (the population of one city) in order to predict the future flows. If the benefit function would reflect the actual capacity of the airport, like we suggest above, then we can have a more natural evolution of the model: instead of modifying the input Z_j to the benefit function, we can let it evolve by the following rule: $\Delta Z_j = \epsilon(D_j - Z_j)Z_j$, where $D_j = \sum_i F_{i,j}$ is the total flow to each airport as predicted by our model. The sign of ΔZ_j depends on whether $D_j > Z_j$ (in which case the capacity of the airport should grow) or $D_j < Z_j$ (in which case the capacity of the airport should decline). At each time step, we update Z_j by adding ΔZ_j and then we re-calculate $F_{i,j}$ for each edge using the new benefit Z. For instance, this may understand the population and economic dynamics of BRIC countries and understand the contributing factors to flight demand. An even more sophisticated approach would take into account both the airport capacity and population size, and other socioeconomic data in addition, like GDP of the country/city.

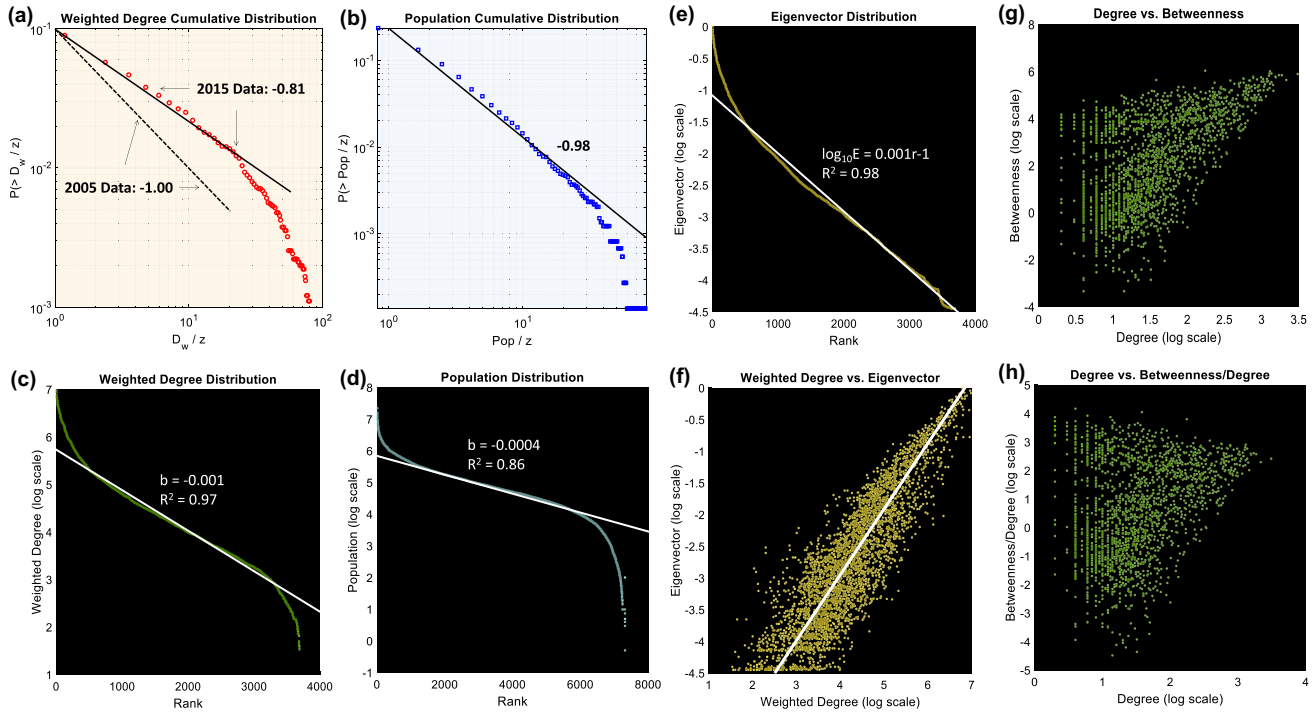

Fig. 3 Complex network properties of the global air transport network. **a**, **b** The normalised cumulative distribution of the weighted degree and population. **c**–**e** The rank distribution of the weighted degree and population and eigenvector centrality. **f**–**h** The correlated centrality values for each airport

3 Top-down approach: complex network models of air transport

The complexity of the air transport network has led many to apply network science to better understand its properties at macroscopic (network properties), and mesoscopic (community properties) levels. Existing work is abundant with snap-shot analysis of network structure (i.e., degree profile, modularity, closeness). However, longitudinal analysis is rare, because the data is expensive to obtain. This section will review both existing research and conduct longitudinal case studies on sub-regions of the air transport network.

3.1 Macroscopic network properties

3.1.1 Previous studies

For macroscopic studies, degree rank, degree distribution and betweenness distribution are the most well studied [1, 32]. Previous studies found that both the degree (unweighted) and the betweenness (unweighted) have a complementary cumulative distribution that obeys a truncated power-law. The normalised gradient (slope) is found to be approximately − 1.0 for degree and − 0.9 for betweenness [1, 32]. Previous studies have also shown that the degree, betweenness, and closeness rank distributions of the Chinese air transport network was found to obey an exponential distribution [33]. Furthermore, the centrality measures are positively correlated with passenger numbers, which indicates that airports that are important from a network perspective also experience the most number of passengers. Another interesting aspect of complex networks is the small-world property, which also applies to airline networks (clustering coefficient is an order of magnitude higher than the random graph equivalent). Furthermore, it was found that average shortest path d grows with $\log(S)$, where S is the number of nodes in the network [1].

3.1.2 Case study: global air transport network in 2015

Centrality distributions Figure 3 shows the complex network of airport (nodes) connected by directed and weighted air transport links. Node size reflects weighted degree and link line-width indicates number of seats per month (aggregated over the flights). (a) global network over one example month comprises of 9033 nodes and 101042 links; and (b) a number of domestic sub-graphs (national), which comprises of 9032 nodes and 53496 links.

Figure 3a, b show the normalised cumulative distribution of the weighted degree and population. The results

confirm established knowledge that the normalised weighted degree (normalised with respect to mean z) exhibits a power-law form:

$$P(> D_w/z) \propto (D_w/z)^{-a} \quad (7)$$

which has been previously confirmed back in 2005 [1, 32]. The gradient (slope) a is found to be − 0.81 for our 2015 data (compared to − 1.00 for 2005 [1]), indicating that there is a diffusion of transportation flow towards a larger number of highly connected hubs. Similarly, a power-law exists in the cumulative distribution of the normalised cities' population Pop/z), which has been well established at both the global and domestic (national) levels.

Centrality correlations Looking further, of particular interest in the context of airline networks is the degree and betweenness correlation. A high correlation indicates the hub–spoke (HS) model, whereby highly connected airports (degree) also act as shortest-path (betweenness) for multi-hop routes (see Fig. 3g). In particular, the variance is small for hubs, giving confidence to the conclusion. Figure 3h looks at the correlation between degree and betweenness per link (betweenness/degree). The results show that the lower-bound of the scatter plot increases the betweenness/degree as degree increases. This shows that hubs not only have a lot of shortest paths and connections, but the number of shortest paths per link is also higher than non-hub airports. Other results also reinforce the notion that hubs can be detected by degree profiling and are important. For example, Fig. 3f shows that degree is highly correlated with eigenvector centrality, indicating that airports with a high number of connections are also airports with important connections.

In Fig. 4, we select the top 50 hubs and show a strong correlation between degree and betweenness centrality (data from 2016). We track the correlation from 1988 to 2018, showing that the correlation falls towards the late 90s, but dramatically increases from late 90s to today (correlation increase from 0.47 to 0.85), which corresponds to the significant fall in air travel costs to consumers. In Sect. 4, we give a more theoretical foundation on what factors drive the HS model, and theorize that the cost of flight changes have led to an increase in HS model.

Relation to population rank In Fig. 3's c, d show the rank distribution of the weighted degree and population. In particular, we note that the data generally obeys an exponential rank distribution

$$D_w \propto \exp(-br), \quad (8)$$

where r is the rank and b is given in Fig. 3d and e. Whilst the coefficient of determination (R-squared) values show that the exponential distribution can explain 97% and 86% of the variations, there exists a *King* and *Pauper* effects which cannot otherwise be explained by any other known

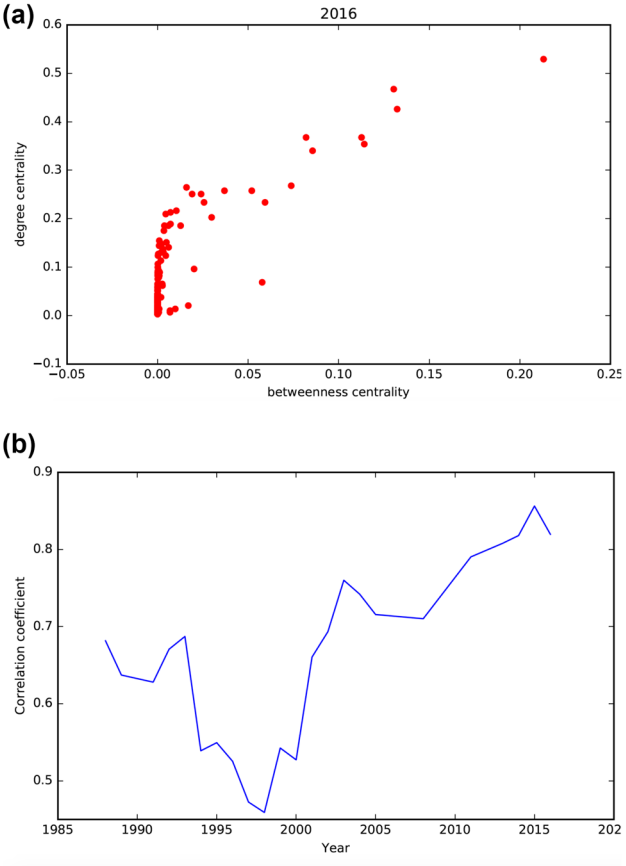

Fig. 4 Hub–spoke model using degree-betweenness correlation: **a** hubs tend to have high degree and betweenness correlation (data from 2016), and **b** correlation has evolved to be stronger after 2000

statistical distributions. The first few ranked cities have an order of magnitude higher (King effect) air transport degree and population. The tail ranked cities have an order of magnitude lower (Pauper effect) air transport degree and population. This is not observable on the cumulative distribution plots, and is evident in both the global graph and within each sub-graph at the domestic level (see results in Fig. 5). More interestingly, most of the King airports relate to the core of the network and we will demonstrate that the air transport network has a core–periphery structure.

3.2 Mesoscopic network properties

The global network can be de-constructed into different sub-graphs. For example, each airline can form a subgraph [34], or the links on each continent can be detected through community structure analysis (modularity) [1]. In [8], a multi-layer network is constructed that comprises of major international airlines and low-cost budget airlines in Europe. It was found that the degree distribution of each

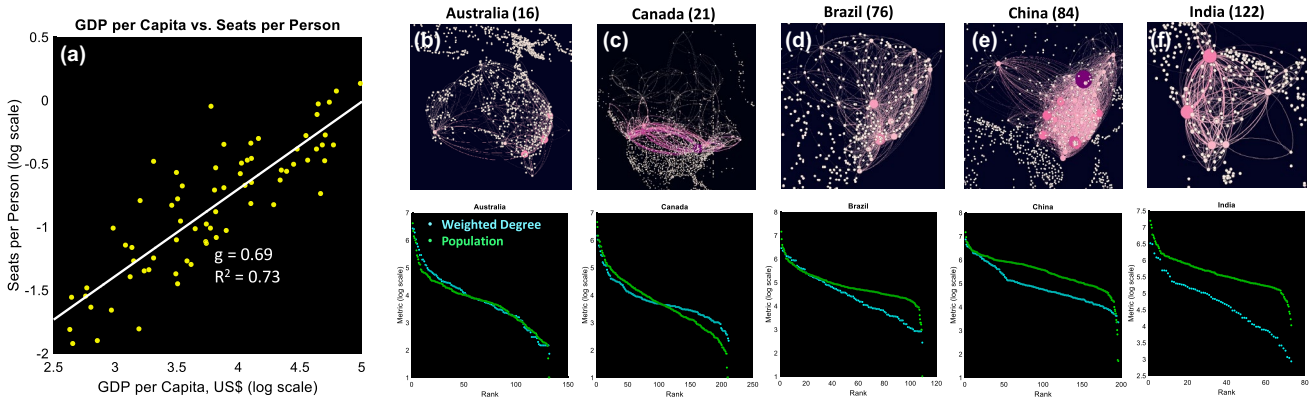

Fig. 5 Domestic air transport network sub-graph centrality distributions and relation to personal wealth. **a** The relationship between the nation's cities' population and airport degree distributions with the national GDP per capita. **b**–**f** The domestic sub-graphs for each country and their population and airport degree distributions. Each nation's GDP per capita rank is given in the brackets

sub-graph did not necessarily conform to the power-law distribution observed at the continental or global scale [1]. In general, it was found that major international formed connections that contained distribution tails which were orders of magnitude higher than the power-law, and budget airlines formed connections that had a degree tail distribution which was poorly connected, indicating a Pauper effect. The robustness [35] of the air transport network subject to random removal was tested in [4, 9, 10, 32], and it was found that the existing network structure has been designed for efficiency and is not resilient against failures or attacks.

3.2.1 Domestic network centrality and relation to wealth

The global air transport network includes both international and domestic flights, and the latter can be regarded as a set of sub-graphs. Figure 3d and e demonstrated that the rank distribution of both the city's population and airport weighted degree fit an exponential distribution. We discover that despite the variety of domestic sub-graph patterns for different countries (see Fig. 5b–f), the same exponential distributed degree rank also exists in each sub-graph alongside the similar exponentially distributed population rank. A key observation is that each country's difference between the sub-graphs' population and airport degree rank distributions is correlated with the GDP per capita of the country. We measure the difference by the ratio of the average area under the graphs, which can be interpreted as the average number of flight seats per person (data is for per month). Fig. 3a shows that the ratio is positively correlated with the GDP per capita i (2015 world bank) via a power-law relationship $\frac{\mathbb{E}[D_w]}{\mathbb{E}[\text{Pop}]} \propto i^g$, where g is found to be 0.69 and can explain for approximately 73% of the variations in each domestic sub-graph's population and degree distribution differences. On a statistical level, the relationship is intuitive in the sense that individual wealth determines the frequency of domestic flights and reasonably well understood [36]. However, what is less well understood until our discovery is the close relationship between the degree and the population rank distributions and the universality of the distribution for every nation. The higher resolution understanding of the distribution means that should new cities be constructed or there is a change in the demographics of one region, researchers can potentially use the relationship found to estimate the resulting adjustments needed in the degree distribution and use it as a proxy for network rewiring (i.e., plan new flight paths).

3.2.2 Core–periphery structure

An intuitive understanding of a network core often refers to a subset of nodes that are densely connected among themselves, whilst the periphery is loosely connected to the core [37]. There already exists different algorithms to detect a core structure based on certain purposes, therefore, it is important to choose the appropriate one. The core profiling method [38] used here considers the degree of nodes in core and the link density within the core. First, nodes are ranked based on decreasing order of degree. For each node, the number of links k_r^+ that connected with nodes having a higher degree than the selected node was recorded. After the k_r^+ sequence is generated, the boundary of the core is able to obtain by detecting the peak of the sequence, after which k_r^+ decreases steadily. A demonstration for the 500 airports is shown in Fig. 7.

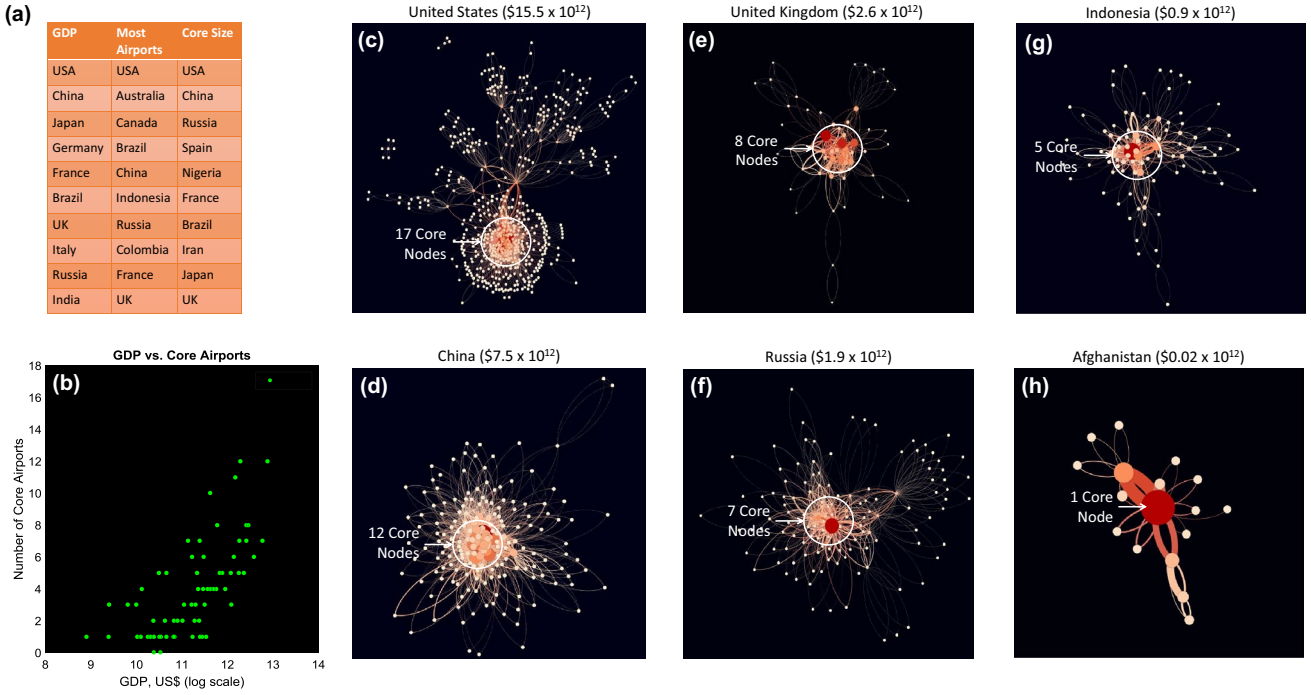

Fig. 6 Core–periphery structure of domestic air transport networks. **a** Table of top 10 ranked countries. **b** Relation between core size and GDP, and **c**–**h** six example core–periphery structures for different countries

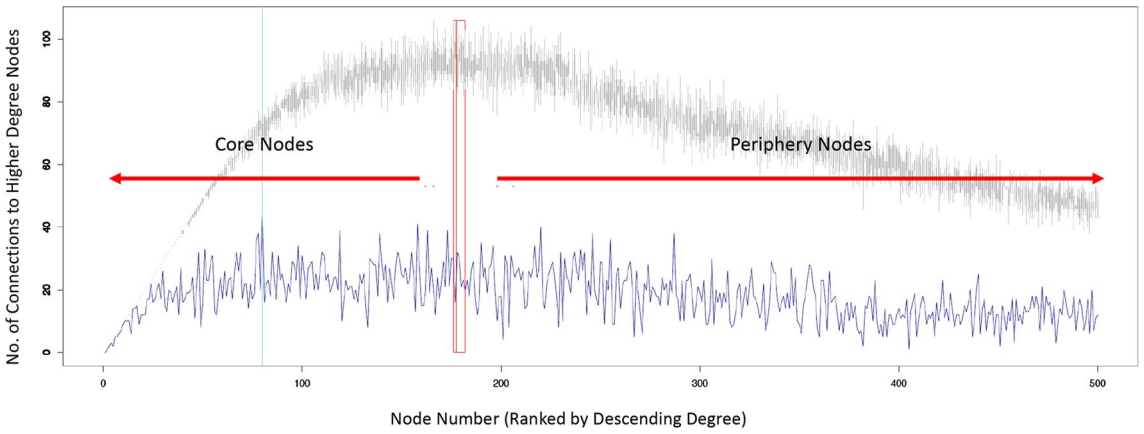

Fig. 7 Core classification: for 500 airports. X-axis indicates the decreasing degree rank of node, Y-axis is the number of connections it has with a higher ranked node (k_r^+), and the red line shows the cut-off between core and periphery classification

At a domestic sub-graph level, a core–periphery structure also exists. Figure 6a shows the top 10 countries in terms of the GDP, most airports, core size, and relative core size. The relationship between each domestic sub-graph's core size and the nation's GDP is shown in Fig. 6b. Figure 6c–h show the core–periphery structure for 6 example nations in descending order GDP and corresponding descending order of core size.

The global air transport network contains a core with approximately 80 nodes (less than 1%), whilst the remaining 9000 are peripheral nodes. The relatively small core size demonstrates the economic efficiency of the network, as well as its low robustness to random and targeted failures. This has been established previously in [32], but not done so with the understanding of core–periphery structure properties. We compared the

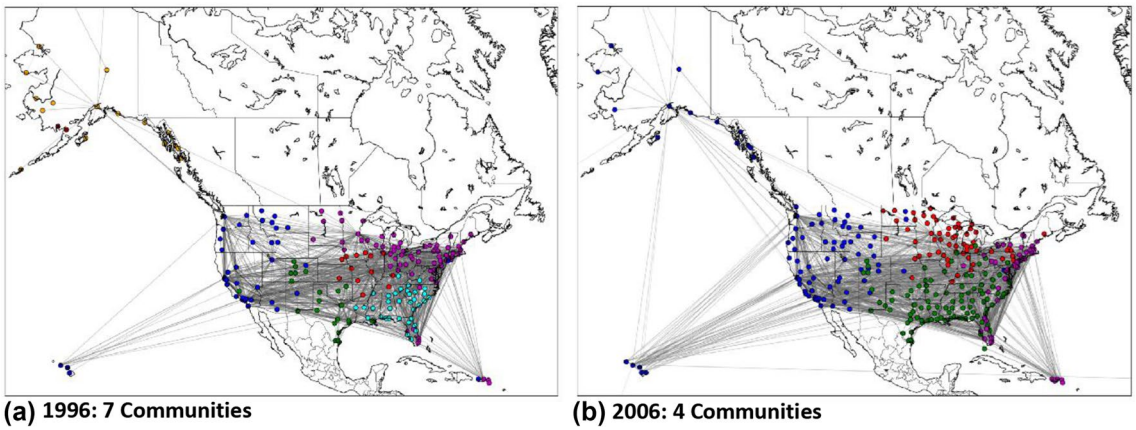

Fig. 8 Community structure of flights network in the US in the month of January in three different years: **a** 2016 (4 communities), and **b** 1996 (7 communities). Data from the Bureau of transportation statistics

Table 1 Summary of the main methods for detecting communities in complex networks

Detection method	Community indicator
Spectral clustering	Eigenspace closseness
Modularity optimization	Higher link density
Statistical inference	Higher link likelihood
Spin–spin interactions	Low energy domains
Coupled oscillators	Phase synchronization
Markov processes	Random walk confinement

current air transport network to a *random network* in which the nodes are the same, but the links were rewired randomly. Therefore, the number of nodes and links, as well as the degree distribution were maintained in the random network [39]. By comparing the relative core size and the core link density between the real network and the random networks, we found that the air transport networks form more cohesive cores, which results in higher stability and topological robustness in the face of perturbations (e.g. attacks or failures [40]).

3.2.3 Evolving communities

Communities are a form of mesoscale structure in networks. Roughly speaking, they are defined as groups of nodes that are densely connected internally and sparsely connected to other groups in the network. There are a number of ways to detect and define community structures from the underlying data of air transport. However, given the ill-defined nature of network communities, selecting a suitable detection method is still discretionary to the researcher's needs and intuition, both in terms of computing complexity and data characteristics. Here, in Table 1, we present the main general classes of community detection methods currently in use across the literature, referring to their strengths and weaknesses (Fig. 7).

In this particular case, we used the Louvain method (a form of modularity maximization particularly suited for large networks) [41]. Crucially, we don't need to specify the number of communities in the network, that is detected automatically by the algorithm. In the Fig. 8, we show identified communities for the US domestic flights network in the month of January for three different years: 2016, 2006, and 1996. Edges in the network are weighted by the number of flights in the respective time period. We notice that detected communities decrease in number over time, and they align with US geographical areas. For example, in 2016, there are 4 communities consisting of: the East Coast and Puerto Rico (purple); the Midwest (red); the South-East (green); and the Western States, including Alaska and Hawaii (blue).

Some structural changes occur over time, especially in the south-east United States. One reason for this could be the consolidation of regional airlines such as JetBlue, which offers many flights along the East Coast and relatively few flights to other regions. Community structure is useful for market segmentation based on route density. Furthermore, by looking at how communities evolve over time, we may be able to pick up changes in the state of the market in a particular region. For our US case study, more work is necessary to understand how communities change over time and what are the factors that drive those changes.

3.2.4 Route changes and classification

Another method for detecting substructures is route classification. An airline's network evolves constantly, with routes being added and discontinued from year to year (see example below for United Airlines). One question is whether we can characterize these routes based on features such as: distance between origin and destination, degree (or weighted degree) of origin and destination (or difference between them), and socioeconomic indicators of the areas serviced. Ideally, this would give an indication of what kind of routes an airline is adding or removing from its network.

As a proof-of-concept, we analysed the 10% of the passenger data in the US for the second quarter of the years 1993–2015. This allows us to estimate the actual travel to high accuracy and we can infer results about the weighted domestic flight network. While the total air travel has increased (see Fig. 9a), there is a clear shift towards longer flights (Fig. 9b, c, note the order of the curves). At the same time, the total number of different routes has decreased, pointing towards an evolution of a hub and spoke structure. Future research in this promising area can focus on developing proprietary unsupervised learning methods for classification, with particular attention to churn and the relationship between operator type and the flight route.

4 Future of air transport networks

We assume that cities are randomly and uniformly distributed. The critical assumption is that we assume that the number of routes is constant and that we make no assumptions on which routes should or shouldn't exist or what the range of a route should be. That means the model is a pure theoretical spatial graph, aimed at only analyzing its fundamental properties as a function of distance cost.

For example, if the distance penalty for a flight reduces, how will it affect the network properties? To this end, we construct a 2-D random geometric graphs (RGG) with a Poisson Point Process (random uniform), whereby the probability of connect is weighted by, such that $Q_{i,j} = Kd^{-\alpha}$, where K is a normalizing factor (i.e., ticket cost). We attempt to construct RGG with a fixed number of nodes and links for a fair comparison of centrality metrics. As such, the expected number of links $E = \sum_{i,j} Q_{i,j}$, yielding $K = \frac{E}{\sum_{i,j} d_{i,j}^{-\alpha}}$. Therefore, the probability of a link forming is:

$$Q_{i,j} = E \frac{d_{i,j}^{-\alpha}}{\sum_{k,l} d_{k,l}^{-\alpha}}. \tag{9}$$

The resulting graph tends to vanish for large values of α (i.e., ticket price K is too large to compensate), so E is only maintained for certain α values (from 0 to 3).

In Fig. 10, we show 8 values of α uniformly distributed from 0 to 3 (represented by different colours in the scatter plot). For a high value of α (i.e., 2–3), the spatial graph shows weak to no correlation between degree and betweenness. This indicates that it is better to travel point-to-point or not travel by air, and as such well connected (high degree) are not prominent transfer hubs (high betweenness). For a low-medium value of α (i.e., 0–2), the non-spatial graph shows a strong correlation between degree and betweenness. This indicates that the hub airports are also the best airports for minimum hop transfers. As such, one conclusion that we can draw is as follows. Traditionally, the cost of flying was high and point-to-point (PP) transportation was prevalent. As the cost reduced (especially since 2000s), the structure of the network is statistically more likely to move to a hub–spoke (HS) network, because large-hubs can afford efficient take-off and landing and logistics. We can see this trend in the data given in Fig 5, where there is a dramatic increase in the HS model since 2000 (correlation increase from 0.47 to 0.85).

This has a profound effect on the design of future aircrafts, as the PP model would prefer small to medium sized aircrafts (e.g. Boeing 777/787 and Airbus A330), whereas a HS model would perhaps prefer high-capacity jumbo-jets (e.g. Boeing 747 or Airbus A380).

5 Conclusions

Almost half of the world's population is carried by airlines each year, and understanding this mode of transport is important from economic and scientific perspectives. In this case study paper, we reviewed both bottom-up (max. entropy agent model) and top-down (network science) approaches to better understand the fundamental science behind air transport networks. A summary of key key findings is given in Fig. 11.

In Sect. 2.2, using simple socioeconomic indicators, we were able to construct a very accurate entropy-maximization interaction model that can predict traffic volume for Australia. Using the population and distance functions, the spatial interaction model can forward estimate the impact of population growth. In Sect. 3.2, using historical data, we were able to identify how hubs evolved over time to become more influential. In Sect. 4, looking into the future, using random graph theory, it seems that reduced flight cost will lead to increased hub influence.

Future research will integrate the flow dynamic data into the complex network analysis, which can be done

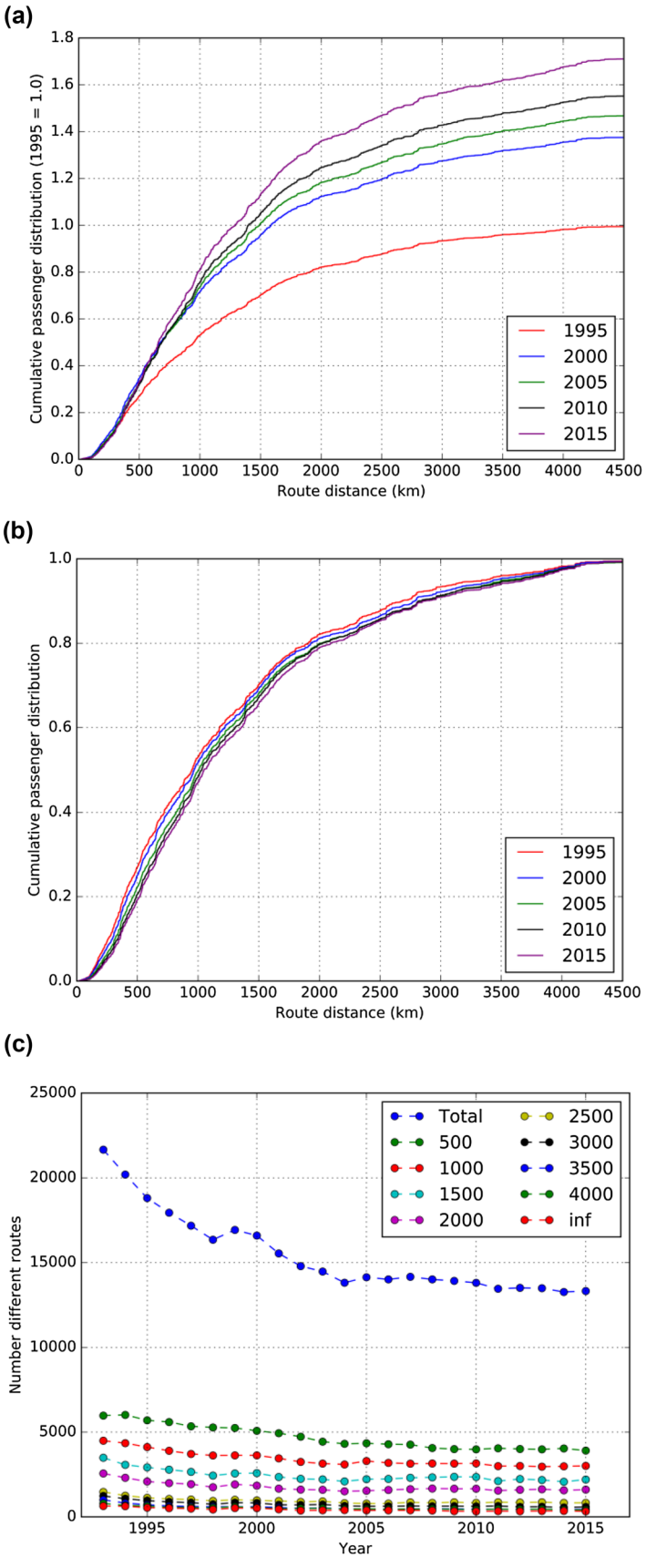

Fig. 9 Route classification: **a**, **b** Cumulative distribution of route distance for different years. **c** Number of routes for different distance classes

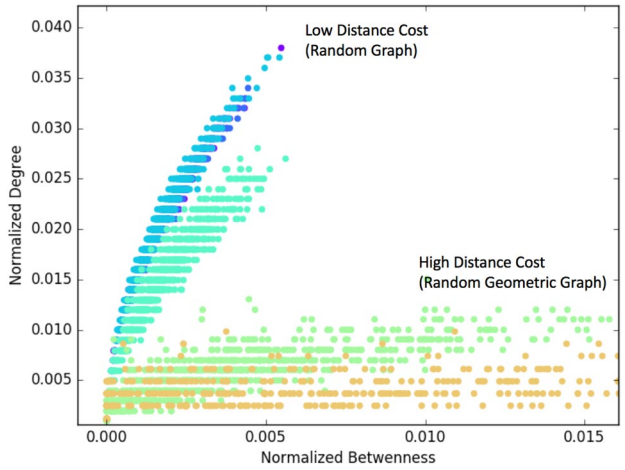

Fig. 10 Relationship between airport degree and betweenness for random graphs: distance dependent cost function drives the formation of random graphs with spatial to non-spatial characteristics, resulting in different levels of hub–spoke prominence

Key finding	Difference over literature	Section and figures
Competitive spatial interaction can predict passenger flow volume	Accounts for competition dynamics compared to gravity and radiation models	Section 2.2 and Fig. 2.
Hub structure becoming more dominant since 2000	Novel definition of hub using centrality correlation of degree and betweenness	Section 3.2 and Fig. 5.
Large economies have high core size in domestic networks, which is less economically efficient, but more robust	Core periphery structure improves over existing numerical analysis	Section 3.2 and Fig. 6.
Fuel cost affects the future network structure, with low fuel cost preferring point-to-point and high cost preferring hub-spoke transport	Random spatial graph analysis using a general distance cost metric	Section 4 and Fig. 10.

Fig. 11 Summary of key results and advance over literature

either explicitly through differential equation models [42] or using passenger flow data as a proxy [43].

Acknowledgements W.G. is supported by Resilient Ecosystems project under the EPSRC Grant EP/R041725/1. B.T., R.F., F.Y., and S.O. are supported by the Oxford Centre for Doctoral Training in Industrially Focused Mathematical Modelling under EPSRC Grant EP/L015803/1. G.M. is supported by the Warwick Centre for Doctoral Training in Mathematics for Real-World Systems under EPSRC Grant EP/L015374/1. All authors would like to thank the Data Centric Engineering Program under Lloyds Register Foundation, and the Alan Turing Institute under the EPSRC Grant EP/N510129/1 for funding the data study, and Airbus for providing guidance.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest.

References

1. Guimera R, Amaral L (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc Natl Acad Sci: PNAS 102:7794–7799
2. Zanin M, Lillo F (2013) Modelling the air transport with complex networks: a short review. Eur Phys J Spec Top 215:5–21
3. Verma T, Araujo N, Herrmann H (2014) Revealing the structure of the world airline network. Sci Rep 4:5638
4. Zhou Y, Wang J, Huang G (2019) Efficiency and robustness of weighted air transport networks. Transp Res Part E Logist Transp Rev 122:14–26
5. Xu Z, Harriss R (2008) Exploring the structure of the U.S. intercity passenger air transportation network: a weighted complex network approach. GeoJournal 73:87
6. Du W, Zhou X, Lordan O, Wang Z, Zhao C, Zhu Y (2016) Analysis of the Chinese airline network as multi-layer networks. Transp Res E Logist Transp Rev 89:108–116
7. Li W, Cai X (2015) Temporal evolution analysis of the European air transportation system: air navigation route network and airport network. Transp B Transp Dyn 3:153–168
8. Cardillo A, Zanin M, Gomez-Gardenes J, Romance M, del Amo A, Boccaletti S (2013) Modeling the multi-layer nature of the European air transport network: resilience and passengers re-scheduling under random failures. Eur Phys J Spec Top 215:23–33
9. Lordan O, Sallan J, Simo P, Gonzalez-Prieto D (2015) Robustness of airline alliance route networks. Commun Nonlinear Sci Numer Simul 22:587–595
10. Wandelt S, Sun X, Cao X (2015) Computationally efficient attack design for robustness analysis of air transportation networks. Transp A Transp Sci 11:939–966
11. Guo W, Vecchio M, Pogrebna G (2017) Global network centrality of university rankings. R Soc Open Sci 4:171172
12. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) Prediction and predictability of global epidemics: the role of the airline transportation network. Proc Natl Acad Sci: PNAS 103:2015–2020
13. Colizza V, Barrat A, Barthelemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci 103:2015–2020
14. Balcan D, Goncalves B, Hu H, Ramasco J, Colizza V, Vespignani A (2010) Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. J Comput Sci 1:132–145
15. Nicolaides C, Cueto-Felgueroso L, Gonzalez M, Juanes R (2012) A metric of influential spreading during contagion dynamics through the air transportation network. PLoS ONE 7:e40961
16. Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. Science 342:1337–1342
17. Wilson A (2008) Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. J R Soc Interface 5:865–871
18. Barthélemy M (2011) Spatial networks. Phys Rep 499(1–3):1–101
19. Anderson J (1979) Theoretical foundation for the gravity equation. Am Econ Rev 69:106–116

20. Bergstrand J (1985) The gravity equation in international trade: some microeconomic foundations and empirical evidence. Rev Econ Stat 67:474–481
21. Poyhonen P (1963) A tentative model for the volume of trade between countries. Weltwirtschaftliches Archiv 90:93–100
22. Balcan D, Colizza V, Goncalves B, Hu H, Ramasco J, Vespignani A (2009) Multiscale mobility networks and the large spreading of infectious diseases. Proc Natl Acad Sci: PNAS 106:21484–21489
23. Kaluza P, Kolzsch A, Gastner MT, Blasius B (2010) The complex network of global cargo ship movements. J R Soc Interface 13:1093–1103
24. Simini F, Gonzalez M, Maritan A, Barabasi A (2012) A universal model for mobility and migration patterns. Nature 484:96–100
25. Lambiotte R, Blondel V, de Kerchove C, Huens E, Prieur C, Smoreda Z, dooren PV (2008) Geographical dispersal of mobile communication networks. Physica A 387:5317–5325
26. Krings G, Calabrese F, Ratti C, Blondel V (2009) A gravity model for inter-city telephone communication networks. J Stat Mech Theory Exp. https://doi.org/10.1088/1742-5468/2009/07/L07003
27. Liben-Nowell D, Nowak J, Kumar R, Raghavan P, Tomkins A (2005) Geographic routing in social networks. Proc Natl Acad Sci: PNAS 102:11623–11628
28. Gonzalez M, Hidalgo C, Barabasi A (2008) Understanding individual human mobility patterns. Nature 453:779
29. Jung W, Wang F, Stanley H (2008) Gravity model in the Korean highway. Europhys Lett: EPL 81:48005
30. Piovani D, Arcaute E, Uchoa G, Wilson A, Batty M (2018) Measuring accessibility using gravity and radiation models. R Soc Open Sci 5:171668
31. Orsino A, Guo W, Araniti G (2018) 5G multiscale mobility : a look at current and upcoming models in the next technology era. IEEE Veh Technol Mag 13(1):120–129
32. Lordan O, Sallan J, Simo P, Gonzalez-Prieto D (2014) Robustness of the air transport network. Transp Res Part E 68:155–163
33. Wang J, Mo H, Wang F, Jin F (2011) Exploring the network structure and nodal centrality of China's air transport network: a complex network approach. J Transp Geogr 19:712–721
34. Li W, Cai X (2004) Statistical analysis of airport network of China. Phys Rev E 69:046106
35. Louzada VHP, Araujo NAM, Verma T, Daolio F, Herrmann HJ, Tomassini M (2015) Critical cooperation range to improve spatial network robustness. PLoS ONE. https://doi.org/10.1371/journal.pone.0118635
36. CAPA (2014) Air travel rises with a country's wealth. Law of nature, or can government policy make a difference? CAPA, Technical report
37. Verma T, Russmann F, Araujo N, Nagler J, Herrmann H (2016) Emergence of core–peripheries in networks. Nat Commun 7:10441
38. Ma A, Mondragon R (2015) Rich-cores in networks. PLoS ONE 10:e0119678
39. Maslov S, Sneppen K, Zaliznyak A (2004) Detection of topological patterns in complex networks: correlation profile of the internet. Phys A Stat Mech Appl 333:529–540
40. Williams M, Musolesi M (2017) Spatio-temporal networks: reachability, centrality and robustness. R Soc Open Sci 3:160196
41. Blondel V, Guillaume J, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008:P10008
42. Gao J, Barzel B, Barabasi A (2016) Universal resilience patterns in complex networks. Nature 530:307
43. Pagani A, Mosquera G, Alturki A, Johnson S, Jarvis S, Wilson A, Guo W, Varga L (2019) Resilience or robustness: identifying topological vulnerabilities in rail networks. R Soc Open Sci 6:181301