

# Article Airline Network Structure: Motifs and Airports' Role in Cliques

Huijuan Yang <sup>1,2,\*</sup>, Meilong Le <sup>1</sup> and Di Wang <sup>3</sup>

- <sup>1</sup> College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 213300, China; mlle@nuaa.edu.cn
- <sup>2</sup> College of Airport Engineering, Civil Aviation Flight University of China, Guanghan 618300, China
- <sup>3</sup> Admissions Office, Tianfu New Area General Aviation Profession Academy, Chengdu 620500, China; wangdi0816@outlook.com
- \* Correspondence: hyang@cafuc.edu.cn

**Abstract:** The air transport system can be considered to be a complex network with airports as vertices and direct flights as edges. Research in this area contributes to the optimisation of the airline network and the sustainable development in transportation. This study chose Air China as an example to discover the dynamics of the airline network topologically. Serving as a critical agent of social and economic connections between cities, the airline network structure evolves over time. However, Air China maintains its multicentric and hierarchical structure and forms a mature point-to-point network with codeshare partners. This research also extracts key players at the airport level and investigates the topological structure of highly connected cliques. The results show that the combination of airports in the cliques may be affected by the airline capacity, traffic rights and interline cooperation. Meanwhile, smaller airports appear more often in cliques than hub airports, which can be interpreted and justified with slot limits at mega-airports. The weighted clique percolation method provides new insights to detecting overlapping communities, which can be characterized by geographical constraints. The shared vertices in the combined codeshare network indicate the possible hub shifting in the constantly changing aviation sector.

Keywords: airline network; motif; clique; clique percolation community detection

# 1. Introduction

The high demand for air transportation creates an enormous pressure on the infrastructure and the environment. Therefore, understanding the dimensions of the airline network becomes crucial for the long-term development towards sustainability [1]. With the tremendous growth of the complex network theory and its application, the air transport system has gradually formed as a complex network with airports as vertices and direct flights as edges [2]. As one of the most investigated networks, the intricate nature of the airline network has been widely analysed topologically at the macroscopic level. Although degree distribution, centrality correlation and small-world network structures are commonly used for the network property analysis, academics found that the abovementioned macroscopic indicators can hardly reflect the character of the individual nodes or modules in the airline network [3–5]. Subsequently, this research aims to propose a dynamic approach to discover the airline network from the mesoscopic level with a special focus on motifs and cliques.

Motifs were first introduced by Milo et al. [6]. A network motif is a small connected subgraph with a well-defined structure, which occurs significantly more frequently than it does in an ensemble of appropriately chosen random graphs. By appearing at higher frequencies, network motifs may have specific functions in information processing [7]. Hence, they are commonly considered to be the basic building blocks of complex networks [8]. In contrast, a subgraph which occurs less often than it does in a randomised network is defined as an anti-motif [1,9].

From the mesoscopic perspective, motifs focus on the microcosmic organisation structure and define network classes with topological interaction patterns [6]. They represent



Citation: Yang, H.; Le, M.; Wang, D. Airline Network Structure: Motifs and Airports' Role in Cliques. *Sustainability* 2021, *13*, 9573. https:// doi.org/10.3390/su13179573

Academic Editor: Lynnette Dray

Received: 22 July 2021 Accepted: 22 August 2021 Published: 25 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



the basic structures that control and modulate the behaviours of the complex network [10]. Since different sets of motifs comprise distinct types of networks, the existing literature replaced network motifs with representative glyphs and introduced them to the air transport industry. For example, Dunne and Shneiderman proposed three types of motifs for simplification, visualisation and interpretation, namely, fan, D-connector and D-clique motifs [11]. Similarly, Clarke and Clarke considered cliques (a collection of vertices for which all possible edges are included) and hub-and-spoke graphs (one vertex to which many other vertices are joined) as two important motifs in the aviation industry [12].

Despite all the possible combinations of vertices, three- and four-node motifs are extensively discussed in the air transport system. For instance, Du et al. defined the network motif as the local relationship pattern between any three airports [13]. Serving as a critical agent of social and economic connections between cities, the airline network is more sophisticated than a group of subgraphs with only three nodes. Bounova [14], Agasse-Duval and Lawford [9] further explored both three- and four-node undirected subgraphs for Southwest Airlines and provided contradictory results regarding the significance of motifs and hub-and-spoke graphs. Jin et al. identified the motifs and anti-motifs for 37 passenger airlines in China, illustrating the importance of adjusting the number of proper network motifs from the topological perspective [1]. Nevertheless, those academics did not capture and interpret the networks as the coexistence of structural subgraphs. How smaller subgraphs can be helpful in discovering and revealing the critical structures, no key roles in the complex network have been captured from the motifs [15]. How they influence the highly interconnected parts in the system is rarely discussed.

As one of the basic concepts in the mathematical area of graph theory, a clique represents a complete subgraph which requires every pair of distinct nodes to be connected with a unique edge in a simple undirected graph or a pair of unique edges in each direction [9]. Particularly, the critical well-connected vertices can be identified by extracting the cliques in the network. Moreover, the coexistence of structural cliques can be analysed and interpreted using community detection methods.

In complex networks, the community is one form of mesoscale structures, which are usually densely connected internally but sparsely connected to the outside [3,16]. Although community detection plays a key role in complex network analysis, design and optimisation, the traditional algorithm reveals the underlying community structure by removing edges based on their betweenness [17,18]. Further, methods like this sort each vertex in one community and fail to detect the overlap in the communities. Rather than being clique-driven, those methods tend to be node-driven, targeting the low-order structures in the network [19]. Moreover, by focusing on the topological matter, the complex network is usually constructed as unweighted and undirected. Additional information is naturally neglected, such as flight schedule, aircraft type and the operator [13]. While flights are not equivalent, the dynamics of weights along the routes should be taken into account in proportion by either flight frequency or passenger number [2]. Last but not least, the airline network can be refined as a multilayer network operated by different carriers [20]. The existing studies of the airline industry mainly focus on the single layer operated by one selected airline, which leaves the structure of the integrated multilayer network unclear [21].

To explore the configuration of the airline network and resolve the limitations mentioned above, this research paid attention to the mesoscopic level and proposed a bottom-up dynamic approach. First of all, this study introduced a motif detection technique and investigated how small and tight components build up to solidarity and connect large networks. Then, the cliques and crucial vertices were extracted from the motifs in order to capture the high-order connectivity patterns. Lastly, a weighted clique percolation method was adopted to examine the dynamic spatial distribution of cliques [22]. To verify the effectiveness of the proposed method, this research chose the scheduled network of Air China for the case study to provide new insights and understanding of the air transportation system. This paper is structured as follows: Section 2 discusses the methodology and the dataset adopted in this research; Section 3 identifies and examines the motifs, cliques and influential nodes in the community; Section 4 discusses the findings and concludes this paper by emphasising the new insights.

### 2. Methods and Dataset

## 2.1. Methods

This research proposes a dynamic network approach to discover the complex network at the mesoscopic level. A series of detecting techniques was adopted to identify the motifs, cliques, high-order communities, as well as influential nodes in the system.

#### 2.1.1. Motif Detection

FANMOD serves as the primary analysis tool in motif detection using the Rand-ESU algorithm [23]. Frequency, p-value and Z-score are three statistical indicators widely used in motif evaluation. The frequency of a given subgraph S with n nodes is defined as follows:

$$Frequency = \frac{n(S)}{N} \tag{1}$$

where n(S) denotes the appearance time for the given subgraph and N denotes the total time that all subgraphs with n nodes appear in the given network.

The *p*-value is a probability ranging from 0 to 1. A bigger *p*-value shows a higher chance of observing more appearance time of a given motif in a random network than it does in the given network. Therefore, when the *p*-value drops to a certain threshold, the subgraph can be considered to be the "network motif".

The Z-score is a statistical significance index, which compares the local network structure with a randomised network. The Z-score for motif *m* is defined as follows:

$$Z_m = \frac{(N_m - Avg_m)}{std_m} \tag{2}$$

where  $N_m$  is the time that motif *m* is observed in the network and  $Avg_m$  and  $std_m$  are the mean and standard deviation of the time that motif *m* is observed in a random network, respectively. In this sense, the motif in a large network tends to obtain a higher Z-score. Rather than the absolute significance, the normalised Z-score (SP) is crucial in emphasising the relative significance of subgraphs when comparing networks of different sizes [8].

## 2.1.2. Cliques and the Weighted Clique Percolation Method

The clique percolation method was introduced by Palla et al. [24]. A clique community is defined as a set of adjacent *k*-cliques, which means two cliques sharing k - 1 nodes. In other words, a *k*-clique has k(k - 1)/2 connected edges, while two adjacent *k*-cliques denote two *k*-cliques sharing k - 1 nodes. Indeed, two three-cliques are adjacent if they share exactly two nodes, which is equivalent to an edge. Likewise, two adjacent four-cliques share three nodes, which is equivalent to a three-node clique.

Farkas et al. applied an extension of the original algorithm to search for modules in weighted networks [22]. The weight of a subgraph is calculated using the geometric mean of its link weights and defined as the subgraph intensity. Therefore, the intensity of a k-clique (C) is written as follows:

$$IN(C) = \left(\prod_{i < j; i, j \in C} w_{ij}\right)^{2/k(k-1)}$$
(3)

where k(k-1)/2 and  $w_{ij}$  denote the number of edges and the weight between nodes *i* and *j*, respectively.

A weighted clique community is defined as a maximal set of *k*-cliques with intensities higher than an optimal *in*. In this sense, modules can be reached via a series of adjacent *k*-clique connections. More importantly, it allows overlaps between communities. Thus,

finding the optimal *in* for each *k* becomes the key for the clique percolation algorithm. If the *in* is too big, the program excludes all the *k*-cliques. On the other hand, a small *in* includes all the *k*-cliques, which makes it difficult to detect any community. Ideally, the size distribution of the communities follows the power law. When the number of communities is small, Farkas et al. proposed optimising *in* based on the variance of communities, which is defined as follows:

$$\chi = \sum_{\alpha \in \alpha} \frac{nc_{\alpha}^{2}}{\left(\sum_{\beta} nc_{\beta}\right)^{2}}$$
(4)

where  $nc_{\alpha}$  denotes a group of communities excluding the largest one, while  $nc_{\beta}$  denotes a group of communities excluding  $nc_{\alpha}$  and the largest one [22]. Therefore, the maximal variance ( $\chi$ ) is associated with the optimal *in* for each respective *k*.

When the network is too small, only a few communities can be expected, which makes it hard to establish a stable  $\chi$  estimation. Further, a specific amount of communities may be detected by chance. Under this circumstance, entropy becomes another option [25]. It is based on the Shannon Information and the entropy of the community partition. More precisely, the most surprising community partition is defined as the lowest probability of knowing to which community a randomly selected node belongs. For instance, the surprisingness equals zero if only one community is detected in a network. This can be explained by the high probability that a randomly selected node belongs to this community because there is only one. Moreover, high chances are that a randomly selected node belongs to a larger community when the communities are not equal in size. Therefore, the surprisingness is only higher if the communities tend to be more equalised in size.

The distribution of entropy values for each k can be obtained from the permutation test. The test creates permutations for the network and extracts the highest entropy for each k before calculating the confidence interval of the entropy. By comparing the entropy with the upper bound of the confidence interval, the optimal *in* for respective k can be spotted and interpreted. As a result, the entropy value which exceeds the confidence interval can be considered more surprising than it would be expected by chance. The entropy can be defined as

$$entropy = -\sum_{i=1}^{NC} p_c * \log_2 p_c$$
(5)

where *NC* denotes the number of communities and  $p_c$  denotes the probability of being in community *c*.

### 2.2. Dataset

Air China was chosen for the case study to enable knowledge discovery and pattern detection of the airline network. A weekly scheduled nonstop flight dataset (from 1 August 2019 to 7 August 2019) was obtained from OAG, including the origin, destination, operating and codeshare carriers of each flight.

Codeshare agreements dramatically influence airline network configuration and shape the market dynamics worldwide [3]. Consequently, the codeshare network operated by 19 Star Alliance partners and 14 carriers outside the alliance was investigated in this study. From the multilayer perspective, each airline should correspond to a different layer. However, the topological properties of the airline network result from the multilayer character rather than the single-layer one [26]. Hence, the codeshare network is tested as an integrated graph to investigate the partners' contributions and the way they affect Air China's network.

During the week, Air China connected 195 airports with 9970 scheduled flights, among which 976 were unique edges. For the combined network, the number of flights was 25,379, among which 2247 were unique. In total, 333 airports were connected with the help of partnerships.

To analyse the topological structure of the airline network, the dataset was treated as an unweighted and undirected network in motif detection and clique extraction. However, flights were weighted by the weekly frequencies for the clique percolation community detection. With continuous growth in demand for air travel, the concept of multiple airport regions (MARs) has emerged since the 1990s [27]. However, this study was primarily focused at the airport level in order to identify the key players in multiple airport regions. Therefore, each airport represented a vertex while each direct flight connecting an airport pair denoted an edge.

## 3. The Network Structure of Air China

The aviation industry in China has been growing at an impressive rate, resulting in complex dynamics in the network. The hub-and-spoke configuration is proposed as one of the most efficient structures and is commonly used by most major airlines worldwide [2]. The Chinese authorities have been trying to establish a US-style hub-and-spoke network to enhance the maturity in the passenger aviation sector [28]. However, the short loops resembling "braids" in Figure 1 reveal the opposite. Figure 1 plots the networks of Air China and its codeshare partners. The edges in the graph are directed and weighted by daily frequencies. The darker colour and wider arrows demonstrate higher frequency and better connectivity. As the flag carrier, Air China primarily serves the domestic market via Beijing Super Hub and Chengdu Shuangliu International Hub. However, it is hard to identify the hub-and-spoke configuration from the complicated cluster since most of the cities are highly connected. Meanwhile, the spatial organisation of the codeshare network forms one sparse cluster in Canada and three relatively dense clusters in Australia, Germany and the United States. Although those countries are geographically far away from China, this shows the potential of hub airports abroad in concentrating flows. More specifically, the intercontinental flights comprise the trunk lines fed by domestic routes. In this sense, the hub-and-spoke structure seems to be more precise with codeshare agreements.



Figure 1. Network structure of Air China and the combined codeshare network.

### 3.1. Motif

To maintain consistency with the previous study, this research captured the three-/four-node motifs for Air China and its partners. Table 1 illustrates all the detected subgraphs in the networks, including motifs and anti-motifs. They are listed by the order of the absolute values of Z-sores.

	Air China Only									Combined with Codeshare Partners						
	Motif ID	Motif	Frequency (Original)	Mean Frequency (Random)	Standard Deviation (Random)	Z-Score	SP	<i>p-</i> Value	Motif ID	Motif	Frequency (Original)	Mean Frequency (Random)	Standard Deviation (Random)	Z-Score	SP	<i>p</i> -Value
Three- node	3a	•	96.89%	99.53%	0.0031	-8.4094	-0.7071	0.999	3a	•	92.94%	99.91%	0.0026	-26.5020	-0.7071	1.000
	3b	$\triangle$	3.11%	0.47%	0.0031	8.4094	0.7071	0.001	3b	$\triangle$	7.06%	0.09%	0.0026	26.5020	0.7071	$\leq 0.001$
Four- node	4a		0.05%	0.00%	0.00%	10.468	0.4974	0.001	4a		0.27%	0.00%	0.00%	57.8600	0.7361	≤0.001
	4b		4.05%	8.69%	0.0051	-9.1556	-0.435	0.999	4d		2.11%	0.02%	0.0006	32.9330	0.4190	$\leq 0.001$
	4c	$\geq$	6.15%	1.11%	0.0057	8.8714	0.4215	0.002	4b		12.84%	24.38%	0.0041	-28.3870	-0.3611	1.000
	4d		0.79%	0.07%	0.0009	8.2666	0.3928	0.001	4c	$\geq$	13.21%	0.25%	0.0056	23.1380	0.2944	≤0.001
	4e		88.76%	89.08%	0.0004	-7.5756	-0.36	0.997	4f		0.60%	2.49%	0.0013	-14.7010	-0.187	1.000
	4f		0.20%	1.05%	0.0013	-6.7136	-0.319	0.997	4e	•	70.97%	72.87%	0.0014	-13.7490	-0.1749	0.999

Table 1. Motif detection results for Air China and its partners.

Two types of three-node motifs were identified in the network. The 3a motif represents two point-to-point flights connected by a hub airport. However, the *p*-value and the negative Z-score indicate that the 3a motif is not significant. Consequently, it is defined as the anti-motif. The only significant three-node motif is the 3b motif, which demonstrates a complete graph called a clique. Although the frequency of the 3b motif in Air China's network is relatively low (3.11%), the partnership more than doubles the result to 7.06%. That means the partnership increases the proportion of the 3b motif's appearance time in the total time that all the three-node subgraphs appear in the network.

Six types of four-node motifs appear in Air China's network, among which three are anti-motifs (4b, 4e and 4f). Particularly, the 4e motif fits the hub-and-spoke motif definition from Clarke and Clarke, where many other vertices are joined to one vertex [12]. However, the hub-and-spoke glyph in this study was diagnosed as insignificant. Like the 3b motif, the 4a motif was captured as the clique in four-node subgraphs. From 4a, 4d to 4c, the number of edges decreases gradually. Social network analysis refers to the unbalances in 4d and 4c as conflicts, which could spread throughout the network [29]. In aviation, those conflicts indicate a lack of direct flights between two airports, where the demand for air travel is not enough to justify the connection. The lower efficiency of the unbalanced motifs suggests the clique to be a sign of network maturity [1]. Remarkably, the frequency, Z-score and SP show that the partnership further enhances the network with significant improvements in mature motifs (4a) and less mature ones (4d and 4c). It is also noticeable that the cliques represent complete point-to-point subgraphs in the airline network, which is the opposite of the hub-and-spoke ones.

The motif detection results are in good agreement with those of Jin et al. [1], indicating that the significant basic glyphs of Air China's network remained the same from 2015 to 2019. Shreds of evidence reflect that Air China maintained its multicentric and hierarchical structure for the time being. Meanwhile, the relatively lower frequencies of motifs in this research can be explained by airline network expansion. New destinations are usually less connected compared with mature markets. Moreover, this research includes regional and international destinations, which are located sparsely and remotely from Air China's home market. Although Agasse-Duval and Lawford claimed that the number of subgraphs generally increases with the size of the network [9], the topological structures are not the priority for airline operation and capacity allocation. Subsequently, the appearance frequency of a given motif drops with the growth of the network.

#### 3.2. Clique

As discussed in Section 3.1, a complete airline subgraph not only reveals the critical players, but also sheds light on the connectivity and maturity of the overall airline network. The clique detection process confirmed that several regions in the network are highly connected (see Figure 2). Indeed, 567 three-node cliques were found among 104 destinations in Air China's network. In particular, 91 cliques were comprised of airports located outside mainland China, more than half of which are located in Asia, including Japan (21), Taiwan (15), Thailand (12) and Hong Kong (11). The partnerships add 1,814 three-node cliques, among which 1,366 are exclusively comprised of airports located in mainland China. Since foreign airlines are seldom authorised with the fifth freedom, those flights can only be operated by Chinese partners, namely Juneyao Airlines (HO), Shandong Airlines (SC) and Shenzhen Airlines (ZH).

The partnerships also introduced 448 three-node cliques with regional and international destinations. It is noticeable that 73 of them were completely comprised of international airports. The spatial distribution of the triangular subgraphs shows that Europe (33) and the United States (31) are the most popular regions (see Figure 2a). With the help of Air Macau (NX) and Shenzhen Airlines (ZH), 110 domestic city pairs are fully connected to Macau and form cliques. Likewise, 29 cliques are identified, with Hong Kong as one of its vertices, by aggregating the networks operated by Cathay Pacific (CX), Dragonair (KA), Shandong Airlines (SC) and Shenzhen Airlines (ZH). Although Uni Airways (B7) and EVA Air (BR) operate 86 flights to/from Taiwan every week, they do not contribute to the existing cliques. It implies that their destinations have already been covered by Air China's own network.



(a) Airports in 3-node cliques

(b) Airports in 4-node cliques



**Figure 2.** (a) Critical airports detected in 3-node cliques; and (b) Critical airports detected in 4-node cliques. Notes: the blue points denote destinations connected by Air China; the red points denote destinations connected by codeshare partners.

The network of Air China forms 361 four-node cliques, connecting 57 airports. More specifically, 46 out of the 361 cliques include regional and international routes, such as flights to Japan (16), Taiwan (10), Hong Kong (6), Thailand (5), Germany (4), South Korea (4) and France (1). Additionally, those 46 four-node cliques are comprised of three domestic airports and one regional/international airport. This is probably because of the limitation of airline capacity and traffic rights. As the flag carrier, Air China primarily serves the domestic market, which leaves a small amount of capacity to routes with the fifth freedom. Further, since the cabotage (seventh, eighth and ninth freedoms) rarely applies in the real world, Air China can hardly schedule flights between two points in one or two foreign countries. Hence, four-node cliques can only be formed with three domestic airports and one regional/international one.

The combined network expands the total number of four-node cliques to 3437, approximately five times the original number. Cliques involving regional/international airports increase dramatically to 684. Particularly, 589 of them are fully comprised of airports located in Asia. Macau is included in almost half the number of those all-Asian-airport cliques (262), outranking Japan (149), South Korea (61), Thailand (54), Hong Kong (52), Taiwan (10) and Singapore (1). This can be explained by the slot limitation at busy airports. Specifically, hub airports running at or close to their capacity are less likely to connect to regional airports in the developing area. Subsequently, airports like Hong Kong usually prefer allocating more slots to metropolitan cities such as Beijing and Shanghai to maintain their market shares with relatively high frequencies. On the contrary, less busy airports like Macau tend to cover more destinations with lower frequencies.

Geographically, the 11 cliques entirely comprised of international airports are located in North America and Europe, operated by Star Alliance members United Airlines (UA) and Lufthansa (LH). On the other hand, most cliques in those regions are comprised of three or two local airports and one or two airports in China. The forms of those cliques distinguish themselves from the abovementioned all-international-airport ones by establishing the interline cooperation between Air China and Star Alliance members.

#### 3.3. Clique Percolation Community Detection

The clique percolation method proposes an algorithm to detect the interaction patterns of cliques. For weighted networks, the algorithm only considers the detected *k*-cliques further when their intensity exceeds a specified threshold *in*. A big threshold may rule out all the communities while a small one may include all the cliques, leading to the same community partition as for the unweighted model. In this sense, the threshold *in* becomes vital in high-order community detection. Initially, the maximum edge weight is tested as the upper limits for *in* as it was recommended by Farkas et al. [22]. The last parameter to be set is the steps. Theoretically, smaller steps are preferred since small changes in steps could lead to rather different results. However, when steps are too small, the computation time increases considerably. Considering the upper limits for *in*, 0.1 was selected as the step, which should be appropriate to find a broad community size distribution. Subsequently, 112 and 119 were set for the network of Air China and the combined codeshare network in steps of 0.1.

Only one community was identified for the *k*-clique in Air China's network. Hence, the entropy was used to optimise *in* for the respective *k*. The largest entropy for k = 3 and k = 4 was 0.999981 and 0.871684. The permutation test shows that only the confidence interval for k = 4 exceeded the upper bound (see Table 2). In this sense, the entropy values for k = 4 can be considered more surprising than would already be expected by chance alone. The largest entropy denotes the most surprising community partition, which captures a low probability of knowing the community of a randomly selected vertex. Therefore, 4 is acceptable as the optimal *k*. In 4-node cliques, 57 airports are grouped as one community, while other airports are isolated. Although the airline network may seem to be complex and sophisticated, the topological structure of high-order cliques to be a small network which cannot be further divided into more communities.

 k
 95% CI (Lower)
 95% CI (Upper)

 3
 1.73391192
 1.87706607

 4
 0.00102004
 0.01341099

Table 2. The permutation test results for Air China's network.

For the combined codeshare network, at least three communities were found for three-node cliques. Subsequently, the optimal *in* (27.2) was identified at the point of the maximal variance ( $\chi = 5.62811791$ ). Eventually, three communities were identified among 58 airports. The 275 isolated nodes in the network included 135 nodes identified in three-cliques and 140 nodes outside the cliques. Initially, the clique percolation algorithm was designed to measure the overlapped vertices in the network. However, no shared node was found in three-clique communities, which implies no key airports interconnecting the coexistence of structural subgraphs on three continents.

When k = 4, only two communities were found. Similarly, entropy and the permutation test were chosen as the primary indicators. Since the entropy value (1.0021173) exceeded the upper bound (0.007377024), 4 was acceptable as the optimal *k*. Figure 3b shows the spatial distribution of the communities. Generally, three airports in Canada are separated from the large community, while the Beijing Capital International Airport (PEK) and the Shanghai Pudong International Airport (PVG) are shared between the two four-clique communities as influential nodes. Topologically, a hub is to a node with a high degree [30]. Although Air China's physical hubs are in Beijing and Chengdu, the result illustrates the worldwide market power of Beijing and Shanghai as international hubs, which connect passengers to the entire network. In other words, the overlapped airports act as the boundary spanners between communities while others obtain connections within the group. Additionally, the community detection algorithm checked the adjacency of subgraphs by demonstrating whether two four-cliques share three of their vertices. Since all the vertices were adjacent to each other, the shared three nodes formed a complete subgraph (a three-clique). Therefore, the detected community was defined as the maximal group of four-cliques that can be reached through a series of adjacent four-cliques. Hence, the blue dots in Figure 3b represent overall better connectivity between airports within the community. Three Canadian airports were left behind due to the lack of shared nodes.

(a) 3-clique communities



(b) 4-clique communities



**Figure 3.** (a) 3-clique communities detected for the combined codeshare network; and (b) 4-clique communities detected for the combined codeshare network. Notes: yellow, blue and red nodes in (a) denote three communities; blue and yellow nodes in (b) denote two communities while red dots denote the shared nodes (PEK and PVG).

Scholars used to detect low-order communities based on edge betweenness. They agreed that communities in the airline business cannot be explained exclusively by geographical considerations [4,31]. Nonetheless, the spatially isolated three three-clique communities are geographically separated from each other (see Figure 3a). Likewise, the four-clique community further points out that the isolation does not necessarily mean separation by continents or countries (see Figure 3b). Basically, the geographical location of the partners' network results in the geographical separation of clique communities. Therefore, this result captures not only the concentration of airline capacity, but also the highly intensive subnetworks in the codeshare network, which may provide some new insights in community detection in the air transport industry.

# **4. Findings and Discussion, Contribution, Limitations and Future Work of the Study** *4.1. Findings and Discussion*

With continuous growth in demand for air travel, the airline network has become complex and widespread. This research investigates the airline network topologically and contributes to the sustainable development of air transport.

Hub oligarchy seems to be the most efficient organisation for transportation system [4]. Thus, the hub-and-spoke subgraph used to be recognised as a mature modular structure in the airline network. Previous literature found the dynamics of the airline network toward a higher spatial concentration [2]. For instance, Button [32], Goetz and Sutton [33] noticed that the topology of the airline network evolved from point-to-point to hub-and-spoke in Europe and America, respectively. Gradual centralisation has also been found in a prototypical low-cost carrier network [14]. Nevertheless, clear evidence from motif detection illustrates that the hub-and-spoke structure is not topologically significant in Air China's network. In contrast, the point-to-point structure represented by complete subgraphs sheds light on the connectivity and maturity of the overall network. The results are consistent with the work of Jin et al. [1], confirming that Air China maintains its multicentric and hierarchical structure for the time being. The difference in airline network configurations is probably due to the centralisation of developed urban areas in China. This also raises the question of whether a hub-and-spoke network fits the strategic plan of all full-service carriers regardless of the geographical, economic or political issues.

The critical destinations in the airline network were extracted by clique detection, revealing the majority of Air China's operation at the domestic level. Pieces of evidence demonstrated the partners' contribution to Air China's network either by offering higher frequencies on existing routes or exploring new destinations which had not yet been covered by Air China. Remarkably, the abovementioned contributions were also limited by traffic rights, geographical location and socioeconomic situation of the airports. For example, the domestic routes can only be operated by Air China's Chinese partners. In this sense, a marketing carrier could benefit from an expanded network using codeshare agreements while complying with local regulations. Moreover, the slot limitation at busy airports leads to the result that less busy airports cover more destinations with lower frequencies and appear more often in cliques. At the continental level, Europe and the United States are the most popular regions considering the spatial distribution of the cliques. The cliques in those areas distinguish themselves by establishing the interline cooperation between Air China and Star Alliance members.

Furthermore, the vertices are not in general equivalence in the networks as they may seem to be in the mathematical formulation of graphs [12]. Community interactions are consequential in capturing the opportunities and constraints and predicting the evolution of the network as a whole. To understand how an individual vertex is embedded in communities of a complex network, communities are identified using a weighted clique percolation method. The method captured not only the concentration of airline capacity, but also the highly intensive subnetworks in the codeshare network. Besides, the results provide new insights by identifying spatially isolated communities in the air transport industry, which has not been observed with traditional community detection techniques. Topologically, the structure of cliques in Air China's network has improved the overall connectivity between airports within the community. The patterns of intercommunity and intracommunity connections were obtained by discovering the overlapped communities.

Two shared influential nodes were found in the four-clique communities. Particularly, the Shanghai Pudong International Airport outranks Air China's international hub in Chengdu Shuangliu and is one of the shared nodes interconnecting the coexistence of structural subgraphs. Multiple hub connections always come with additional cost–benefit considerations. The global market power of Shanghai proves that the network could grow out of a possible geopolitically core region and connect to the rest of the world [34]. Although some academics claim that the criticality is rather stable over time [35], change happens constantly in the aviation sector. Air China may encounter more hub shifting like this after the Beijing Daxing International Airport (PKX) and the Chengdu Tianfu International Airport (TFU) opened to the public in September 2019 and June 2021, respectively. Consequently, subsequent studies are necessary for the adaptation of Air China's strategic management.

#### 4.2. Contribution

Research on the airline network is of great importance in understanding and optimising its structure and sustainable development. To fill the gap in the existing literature, this research focused on the way that smaller subgraphs compound the larger structures, capturing and interpreting the airline network as the coexistence of structural subgraphs.

This paper first expanded the research scope to regional and international destinations. Then, it introduced mature techniques from complex network theories and presented a dynamic bottom-up roadmap to uncover the hidden cluster configuration in the airline industry. Indeed, this study confirms the results from the previous literature regarding Air China's multicentric and hierarchical point-to-point network structure. It also raises the question of whether a hub-and-spoke network fits all full-service carriers. More importantly, this research examines the key roles in the airline network at the airport level. The results show that the combination of airports in the cliques may be affected by airline capacity, traffic rights and interline cooperation. Meanwhile, smaller airports appear more often in cliques than hub airports, which can be interpreted and justified with slot constraints at mega-airports.

Rather than being node-driven, the community detection method in this study is clique-driven, targeting the high-order structures and the overlap in the communities. Weights along the routes and layers of the network were taken into account to reflect the inequivalence in flights and the contributions of codeshare partners. Although the airline network may seem to be complex and sophisticated, the topological structure of high-order cliques tends to be a small network. The algorithm provides geographically separated communities, which have not been obtained with traditional techniques. However, the separation does not necessarily mean geographical isolation by countries or continents. Basically, the geographical location of the partners' network results in the geographical separation of clique communities. In other words, the geographical separation reflects the partners' contribution and market power in a certain area. Although most techniques prefer considering high-degree nodes as critical [35], this paper considered the shared nodes in the overlapping area as influential. In this sense, two shared nodes in Air China's codeshare network prove their global market power in connecting the domestic market to the rest of the world. It is also noticeable that one of the shared nodes is not the original hub airport of Air China. While the hub shifting phenomenon reveals the contradiction between physical and topological networks, it raises another question of whether this phenomenon is widespread in the air transport industry.

#### 4.3. Limitations and Future Work of the Study

This research explores the spatial distribution of the community structure and simplifies the complex network in reality. However, only one airline and its 33 codeshare partners were examined in this study with a relatively small sample size. Further research is required with a regional or worldwide dataset to measure the robustness of this method and promote sustainable development in the air transport industry. During the analysis, this paper also raised two questions, first of all, whether a hub-and-spoke network fits all full-service carriers; second, whether the topological hub shifting phenomenon is widespread in the air transport industry. Consequently, subsequent studies are necessary for the adaptation of airlines' strategic management, especially in multiple-airport regions.

**Author Contributions:** Writing—original draft preparation, H.Y.; Writing—review and editing, H.Y., M.L. and D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Civil Aviation Administration of China (FDQT0006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available at www.oag.com (accessed on 20 January 2020).

**Acknowledgments:** The authors are grateful to the anonymous reviewers and the editor for their valuable suggestions, which improved the manuscript considerably.

**Conflicts of Interest:** The authors declare that there is no conflict of interest regarding the publication of this paper.

### References

- 1. Jin, Y.; Wei, Y.; Xiu, C.; Song, W.; Yang, K. Study on Structural Characteristics of China's Passenger Airline Network Based on Network Motifs Analysis. *Sustainability* **2019**, *11*, 2484. [CrossRef]
- Zanin, M.; Lillo, F. Modelling the air transport with complex networks: A short review. *Eur. Phys. J. Spec. Top.* 2013, 215, 5–21. [CrossRef]
- 3. Reggiani, A.; Nijkamp, P.; Cento, A. *Connectivity and Concentration in Airline Networks: A Complexity Analysis of Lufthansa's Network;* Tinbergen Institute: Amsterdam, The Netherlands, 2011.
- 4. Guo, W.; Toader, B.; Feier, R.; Mosquera, G.; Ying, F.; Oh, S.; Price-Williams, M.; Krupp, A. Global air transport complex network: Multi-scale analysis. *SN Appl. Sci.* **2019**, *1*, 680. [CrossRef]
- Guimera, R.; Sales-Pardo, M.; Amaral, L. Classes of complex networks defined by role-to-role connectivity profiles. *Nat. Phys.* 2007, 3, 63–69. [CrossRef] [PubMed]
- Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. Science 2002, 298, 824. [CrossRef] [PubMed]
- 7. Shen-Orr, S.; Milo, R.; Mangan, S.; Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genetics*. **2002**, *31*, 64–68. [CrossRef]
- Yan, L.; Wang, J.; Han, J.; Wang, Y. A Significance-Driven Framework for Characterizing and Finding Evolving Patterns of News Networks. In Proceedings of the 2012 International Conference on Artificial Intelligence and Computational Intelligence, Chengdu, China, 26–28 October 2012.
- 9. Agasse-Duval, M.; Lawford, S. Subgraphs and Motifs in a Dynamic Airline Network. Available online: https://arxiv.org/abs/18 07.02585 (accessed on 31 December 2019).
- 10. Benson, A.; Gleich, D.; Leskovec, J. Higher-order organization of complex networks. Science 2016, 353, 6295. [CrossRef] [PubMed]
- 11. Dunne, C.; Shneiderman, B. Motif Simplification: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs. In *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013;* Association for Computing Machinery: New York, NY, USA, 2013.
- 12. Clarke, B.S.; Clarke, J.L. *Predictive Statistics: Analysis and Inference beyond Models*, 1st ed.; Cambridge University Press: Cambridge, UK, 2018.
- 13. Du, W.; Zhou, X.; Lordan, O.; Wang, Z.; Zhao, C.; Zhu, Y. Analysis of the Chinese Airline Network as multi-layer networks. *Transp. Res. Part E Logist. Transp. Rev.* **2016**, *89*, 108–116. [CrossRef]
- 14. Bounova, G. Topological Evolution of Networks: Case Studies in the US Airlines and Language Wikipedias. Ph.D. Dissertation, Institute of Technology, Massachusetts, MA, USA, 2009.
- 15. Li, J.; Wen, X.; Wu, M.; Liu, F.; Li, S. Identification of key nodes and vital edges in aviation network based on minimum connected dominating set. *Physica A* 2019, *541*, 123340. [CrossRef]
- 16. Li, H.-J.; Bu, Z.; Wang, Z.; Cao, J. Dynamical clustering in electronic commerce systems via optimization and leadership expansion. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5327–5334. [CrossRef]
- 17. Newman, M.; Girvan, M. Finding and evaluating community structure in networks. Phys. Rev. E. 2004, 69, 026113. [CrossRef]
- 18. Li, H.-J.; Wang, L.; Zhang, Y.; Perc, M. Optimization of identifiability for efficient community detection. *New J. Phys.* 2020, 22, 063035. [CrossRef]

- 19. Guimerà, R.; Mossa, S.; Turtschi, A.; Amaral, L. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA* 2005, 102, 7794–7799. [CrossRef]
- 20. Li, H.-J.; Wang, L. Multi-scale asynchronous belief percolation model on multiplex networks. *New J. Phys.* 2019, 21, 015005. [CrossRef]
- 21. Sun, X.; Wandelt, S. Complementary strengths of airlines under network disruptions. Saf. Sci. 2018, 103, 76–87. [CrossRef]
- 22. Farkas, I.; Abel, D.; Palla, G.; Vicsek, T. Weighted network modules. New J. Phys. 2007, 9, 180. [CrossRef]
- 23. Wernicke, S. A faster algorithm for detecting network motifs. In Proceedings of the 5th Workshop on Algorithms in Bioinformatics (WABI '05). Lecture Notes in Bioinformatics, Mallorca, Spain, 3–6 October 2005.
- 24. Palla, G.; Derenyi, I.; Farkas, I.; Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 2005, 435, 814–818. [CrossRef] [PubMed]
- 25. Fortunato, S. Community detection in graphs. Phys. Rep. 2010, 486, 75–174. [CrossRef]
- Cardillo, A.; Gomez-Gardenes, J.; Zanin, M.; Romance, M.; Papo, D.; del Pozo, F.; Boccaletti, S. Emergence of network features from multiplexity. *Sci. Rep.* 2013, *3*, 1344. [CrossRef]
- 27. Sun, X.; Wandelt, S.; Hansen, M.; Li, A. Multiple airport regions based on inter-airport temporal distances. *Transp. Res. Part E: Logist. Transp. Rev.* **2017**, *101*, 84–98. [CrossRef]
- 28. Airline Business, Chinese Hub and Spoke. 2001. Available online: https://www.flightglobal.com/chinese-hub-and-spoke/35577. article (accessed on 6 January 2020).
- 29. Kouznetsov, A.; Tsvetovat, M. Social Network Analysis for Startups; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2011.
- 30. Sun, X.; Gollnick, V.; Wandelt, S. Robustness analysis metrics for worldwide airport network: A comprehensive study. *Chin. J. Aeronaut.* **2017**, *30*, 500–512. [CrossRef]
- 31. Verma, T.; Araújo, N.A.M.; Herrmann, H.J. Revealing the structure of the world airline network. Sci. Rep. 2014, 4, 5638. [CrossRef]
- 32. Button, K. Debunking some common myths about airport hubs. J. Air Transp. Manag. 2002, 8, 177–188. [CrossRef]
- 33. Goetz, A.; Sutton, C. The Geography of Deregulation in the U.S. Airline Industry. *Ann. Assoc. Am. Geogr.* **1997**, *87*, 238–263. [CrossRef]
- 34. O'Kelly, M. Global Airline Networks: Comparative Nodal Access Measures. Spat. Econ. Anal. 2016, 11, 253–275. [CrossRef]
- 35. Sun, X.; Wandelt, S.; Cao, X. On Node Criticality in Air Transportation Networks. Netw. Spat. Econ. 2017, 17, 737–761. [CrossRef]