WILEY | Hindawi

Research Article

Community Detection in Airline Networks: An Empirical Analysis of American vs. Southwest Airlines

Weiwei Wu [iD],[1] **Haoyu Zhang** [iD],[1,2] **Shengrun Zhang** [iD],[1] **and Frank Witlox** [iD][1,2]

[1]*College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, 210016 Nanjing, Jiangsu, China*
[2]*Department of Geography, Ghent University, Krijgslaan 281/S8, B9000 Gent, Belgium*

Correspondence should be addressed to Haoyu Zhang; 1123851665@qq.com

In this paper, we develop a route-traffic-based method for detecting community structures in airline networks. Our model is both an application and an extension of the Clauset-Newman-Moore (CNM) modularity maximization algorithm, in that we apply the CNM algorithm to large airline networks, and take both route distance and passenger volumes into account. Therefore, the relationships between airports are defined not only based on the topological structure of the network but also by a traffic-driven indicator. To illustrate our model, two case studies are presented: American Airlines and Southwest Airlines. Results show that the model is effective in exploring the characteristics of the network connections, including the detection of the most influential nodes and communities on the formation of different network structures. This information is important from an airline operation pattern perspective to identify the vulnerability of networks.

1. Introduction

Our world consists of many complex geographical networks, ranging from computer to social networks, to infrastructure or transport-related networks. Many researchers have attempted to unravel the properties associated with these complex networks [1]. In terms of transport, for instance, city streets [2], public transport networks [3, 4] and aviation networks [5, 6] have been investigated. With the discovery of small-world [7] and scale-free [8] properties in many natural and artificial networks, methods and techniques have emerged to improve the understanding of these complex networks. More recently, the study of community structures within networks have gained renewed attention. As defined by Girvan et al. [9] and Chen et al. [1], p. 890 "community structure refers to vertices that are gathered into several groups in which there is a higher density of edges within groups than among groups." A schematic illustration of such a graph with communities is shown in Figure 1.

Communities, also often referred to as clusters or modules, are groups of vertices that share common properties and/or play similar roles within a network [10]. In other words, apart from the topological structure, the vertex properties made

explicit in the communities are also important. This is because being able to detect these communities can help us to understand and utilize these networks more effectively. Moreover, it will allow us to discover hidden relations between vertices [11].

For this paper, we are interested in detecting community structures in airport and airline networks. Much of the existing work in this area is relevant to the development of our modelling approach. Gegov et al. [12], for instance, investigated community structure in a US airport network by considering both the topological properties and the volume of people traveling. Comparing the network structure with migration patterns, the identified relationships showed a clear overlap between US domestic air travel and migration. Guimerà et al. [13] identified communities in the worldwide air transport network, and demonstrated the multi-community structure of this worldwide network. Their analysis showed that the community structure cannot be explained exclusively based on geographical restraints, but that geo-political concerns should also be taken into account. Postorino and Versaci [14] proposed a fuzzy-based procedure to cluster airports using the geometric distance among airports as an intrinsic fuzzy variable. The result showed that this airport selection or
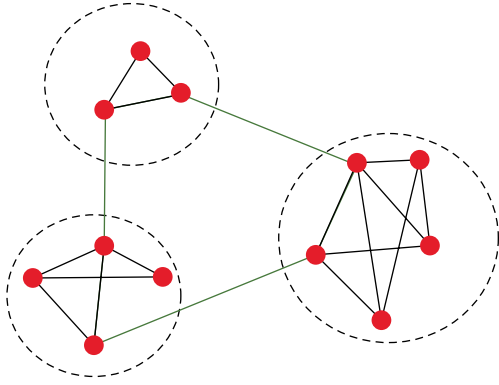
FIGURE 1: A toy network graph with three communities enclosed by the dashed circles (adapted from Fortunato and Castellano [32]).

classification is more appropriate compared to the classification provided within policy recommendations. Finally, Cong et al. [15] developed a spectral clustering algorithm to categorize airports by analyzing fluctuations in distance correlation. The results showed that there is one category of airports that controls the critical state of the network, and six airports in this group were found to be the most important airports in the Chinese air transport network.

However, few of the previous studies have considered the reality of networks, and communities were often only algorithmically defined [16]. In most networks, only distance is taken into account, and community detection is merely the end product of the algorithm. Without a clear definition of the network, the results can lead to conclusions of fuzzy communities and so-called unstable nodes [17]. Anomalies may arise especially in an air transport network, such as the mismatching of low degree and high betweenness of an individual node ([18, 19]). Furthermore, the most connected cities are not necessarily the most central [13]. These anomalies can result in the miscalculation of some marginal vertices, which lead to inaccurate community detection and ignore the effect of these vertices on both internal and external communities.

Essentially, air transport networks are not only spatially constrained when the vertices are both vastly topologically coupled and spatially clustered, but can also be property correlated between vertices. Barrat et al. [20] pointed to the clustering coefficient concept. The empirical research on the worldwide air transport network have shown the impact of edge-weight on network structure. That is, even with the same topology, the related internal structures and hierarchies of networks would be different due to different weights. Moreover, the detection of important nodes would be impacted within different weighted networks. Techniques used in identifying node importance are distinguished from network metric-based selection and robustness-related node importance perspective, and can be applied to a weighted network.

Sun et al. [21] considered passengers and distance as network weight to investigate the robustness of the worldwide air transportation network. Based on the network metric, 12 different 'attacking strategies' were used to analyze airport importance by measuring the unaffected passengers with rerouting. Using different metrics (e.g., node degree, weighted

betweenness centrality, size of giant component) not only allowed for the robustness of network to be evaluated, but also allowed for the comparison of the importance of node connectivity [22, 23], Motivated by Sun et al. [21], we hypothesized that not only route distance would affect the airport connection in terms of the network spatial aspect, but also that route traffic would impact on the capability of a network in terms of passenger rerouting. Therefore, both traffic and distance were considered as network weight in this paper. We aim to investigate the importance of airports in their capability to guarantee more passengers within the shortest route.

Based on route traffic and distance, we propose an improvement of the Clauset-Newman-Moore (CNM) modularity maximization algorithm [1, 24] to detect communities in spatially constrained networks. A core community is not only more compact in space, but the nodes in the same community also have stronger flight correlations than those in a different community. The aim is to identify the regional features of network connections and to detect the most impacted nodes and communities in flight interaction, and how they affect the entire network.

This paper is organized as follows. In Section 2, we introduce the route-traffic-based community detection method, modified to account for route traffic and distance when considering two connected airports. Section 3 illustrates the application of our model using two case studies (American Airlines and Southwest Airlines). In Section 4, core communities and core airports of different airline networks are analyzed and we discuss the results in relation to different operation patterns in identifying the vulnerability of entire networks. Section 5 provides the conclusion and discusses directions for further research.

2. Methodology

The Clauset-Newman-Moore (CNM) modularity maximization algorithm formed the basis of our model for airport networks. We rely in part on the work by Chen et al. [1], and we modify their algorithm by defining the weights of the edges in the network as the function of the route-traffic correlation coefficient between two directly connected airports. The coefficient corresponds to the number of flights between the connected airports, such that the higher the number of flights, the greater the route-traffic correlation coefficient.

2.1. Modularity and CNM Algorithm. The CNM algorithm [24] is a community detection algorithm based on an agglomerative hierarchical clustering method, where groups of vertices are successively joined to form larger communities such that modularity gradually increases after communities are merged. The higher the value of the modularity, the better clustering of the network. Therefore, the basic concept of the CNM algorithm is the concept of modularity, and the approach is devised to maximize the modularity of the network [25].

The modularity, as pointed out by Wang et al. [26], is based on the idea that a random graph is not expected to have a cluster structure. Therefore, the possible existence of clusters is revealed by the comparison between the actual density of the edges and

the density that one would expect to have if the vertices of the graph were attached regardless of the community structure. A sound partition of a network should result in a considerably greater number of edges within communities than expected. The mathematical expression of modularity is [1], p. 894:

$$Q = \frac{1}{2m} \sum_{C \in P} \sum_{v,u \in C} \left(A_{vu} - \frac{k_v k_u}{2m} \right), \tag{1}$$

where C is a community, P is the community set of the network, and v and u are nodes in the community C. A_{vu} is an element of the adjacency matrix of the network. If there is an edge between v and u, then $A_{vu} = 1$; otherwise, $A_{vu} = 0$. In addition, let m be the total number of edges in the network, and k_v be the degree of node v. The higher the value of Q, the better the community structure. Therefore, Q is defined as a stop criterion for the community detection algorithm. To simplify the above expression, two additional variables are introduced:

$$e_{ij} = \frac{1}{2m} \sum_{v \in i, u \in j} A_{vu}, \tag{2}$$

$$a_i = \sum_j e_{ij}, \tag{3}$$

where e_{ij} is the fraction of edges that join vertices in community i to vertices in community j, and a_i is the fraction of edges that are attached to vertices in community i. Thus, in line with [1], p. 895, the function of Q can be transformed into:

$$Q = \sum_i \left(e_{ii} - a_i^2 \right). \tag{4}$$

Note that at the beginning, as Chen et al. (2015, 895) states "the CNM algorithm regards every vertex as a community and then merges them step-by-step." Furthermore, merging communities between where there are edges connected leads to an increase of modularity, which results in ΔQ. The pair of communities that results into the maximum ΔQ is then selected to form a new community. When ΔQ is negative, the process stops; and the network community structure is identified.

2.2. Route-Traffic Modularity and Modified Algorithm. As stated earlier, we also alter the CNM algorithm to distinguish a route-traffic-based community. As it is a spatially constrained network, route distance is taken into consideration. In addition, because the traffic in different airports will affect the interaction of connecting airports, we suggest a route-traffic modularity, which adds a weight of traffic correlation coefficient to the edges when calculating modularity. Using the correlation coefficient in probability theory, the correlation of route traffic between two airports can be presented as ($< \cdot >$ means taking the average):

$$r_{vu} = \frac{< f_{vu} f_{uv} > - < f_{vu} >< f_{uv} >}{\sqrt{\left(< f_{vu}^2 > - < f_{vu} >^2 \right) \left(< f_{uv}^2 > - < f_{uv} >^2 \right)}}. \tag{5}$$

Following the impact of route traffic and distance decay, the weight of the edges is defined as:

$$w_{vu} = \frac{r_{vu}}{d_{vu}^2}. \tag{6}$$

Where r_{vu} is the route-traffic correlation coefficient, where f_{vu} denotes the number of flight departures from airport v and arrivals at airport u, and d_{vu} is the normalized distance between airport v and u. Different values of the power of d_{vu} may lead to different community structures, therefore, we choose (based on the gravity model) to square d_{vu} when considering the spatial limitation [1]. According to Equation (2), we can calculate the element of the weighted matrix of $e^{traffic}$:

$$e^{traffic} = \frac{\sum_{v \in i, u \in j} w_{vu}}{2W} = \frac{\sum_{v \in i, u \in j} r_{vu}/d_{vu}^2}{\sum_{v,u} r_{vu}/d_{vu}^2}, \tag{7}$$

where W is the sum of weighted edges in the network. Let $a_i^{traffic}$ be the sum of elements in each row of the matrix,

$$a_i^{traffic} = \sum_j e_{ij}^{traffic}. \tag{8}$$

the route-traffic modularity can be defined as:

$$Q^{traffic} = \sum_i \left[e_{ii}^{traffic} - \left(a_i^{traffic} \right)^2 \right]$$
$$= \sum_i \left[\frac{\sum_{v,u \in i} w_{vu}}{2W} - \left(\frac{\sum_{v \in i} w_{vu}}{2W} \right)^2 \right]. \tag{9}$$

Similar to the CNM algorithm, the route-traffic-based community detection algorithm begins with each airport as a single community. It computes the increase of route-traffic modularity $\Delta Q^{traffic}$ by merging each pair of communities, and the pair that generates the maximum $\Delta Q^{traffic}$ is then merged. In this way, route-traffic communities can be clustered while considering both geographic distance and route traffic.

3. Applying the Model to Two Case Studies

We test our model using two case studies. The case studies were selected so that we focus on the same geographical area (i.e., the US). While both the selected airlines have a strong market reach within that area and are in competition with each other, they operate with different strategy patterns. American Airlines (AAL) is a full service carrier with scheduled domestic flights for 150 cities. After merging with the US Airways Group in 2013, AAL became one of the largest airlines in the world. It has been shown that the network structure created by the merger of these two carriers has impacted on hub structures, accessibility and physical coverage of the network [27]. By contrast, Southwest Airlines (SWA) is the largest low-cost carrier in the world and it has the most scheduled domestic flights in United States. In 2017, the traffic volume of AAL was 144 million and the traffic volume of SWA reached 156 million, which occupied 17.9 and 18.3% of the air traffic in the US respectively. Although both airlines are the largest airlines of America, their operation patterns—which determines how they attract passengers and how their network operates—are vastly different.
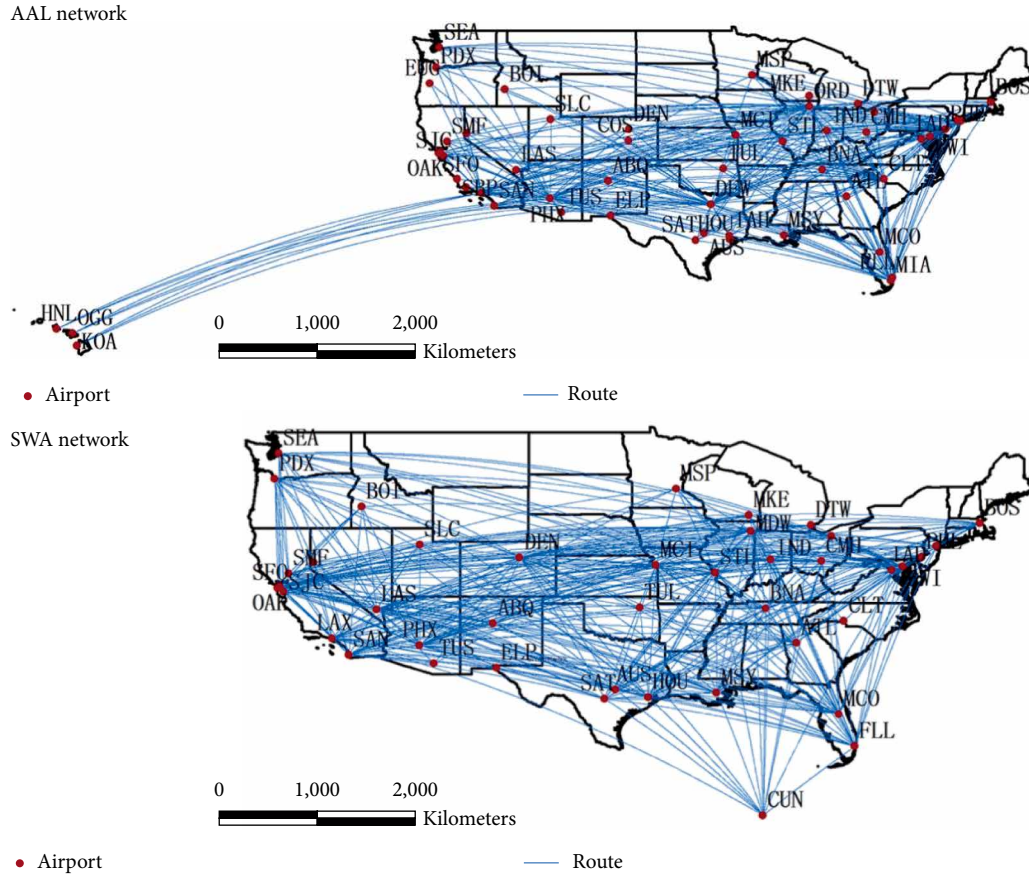
FIGURE 2: AAL Network and SWA Network.

Data from https://www.oag.com and https://www.flight-stats.com on American Airlines (AAL) and Southwest Airlines (SWA) were collected from May to June 2017, and the networks of these two airlines are shown in Figure 2 based on their scheduling flights. Using the community detection method discussed above, the community detection maps of these two airlines were produced (Figure 3). Different communities are divided by different colors, with core airports denoted by larger circles and names.

Figure 3 shows that the community structures of the two airlines are completely different. The American Airlines network is divided into 5 main communities with 7 core airports, while the Southwest Airlines network is divided into 6 main communities and 8 core airports (core airports are shown by larger dots). In the AAL network, the main airports include LAX, PHX and MIA. In the SWA network, the main airports include SAN and STL.

From the correlation coefficient distribution function $f(r_{vu})$ of both networks (Figure 4), the route correlation coefficient of AAL has a greater range (ranging from 4.06 to 5.52, with an average value of 0.73) compared to the route correlation coefficient of the SWA network (ranging from −0.83 to 0.98, with an average value of 0.08).

From Figure 4, the AAL network showed the strongest features of connectivity to preferential nodes and is considered to be the closest to the hub-and-spoke network. In fact, it presented a sparse-strong style (i.e., a few edges among the communities with large weights), which means that traffic is concentrated between hub airports and there are fewer routes between communities. Therefore, we can assume a 'tendency' towards a hub hierarchy or hub-and-spoke configuration in the AAL network. This is also based on the appearance of nodes such as PHX, LAX, and MIA, which are structured as hubs in the framework of AAL activities.

The traffic in the SWA network showed that flights are scheduled in a more balanced manner over the entire network and presented in a dense-weak style (i.e., many edges among the communities with small weights). This means that more routes are connected for both inside and outside communities, but with traffic distributed sparsely. This characteristic leads to a city-to-city network configuration. This is not surprising given SWA's low cost operation.

4. Community Analysis

Given that the position of an airport in the carriers' hub hierarchies may influence network structures and other market variables such as average airfares [28], core community/airport and their influence should be further explored [29]. In the following sections, we discuss a number of issues related to community analysis. We examine the core communities, the core nodes (airports) and the peripheral nodes (airports).
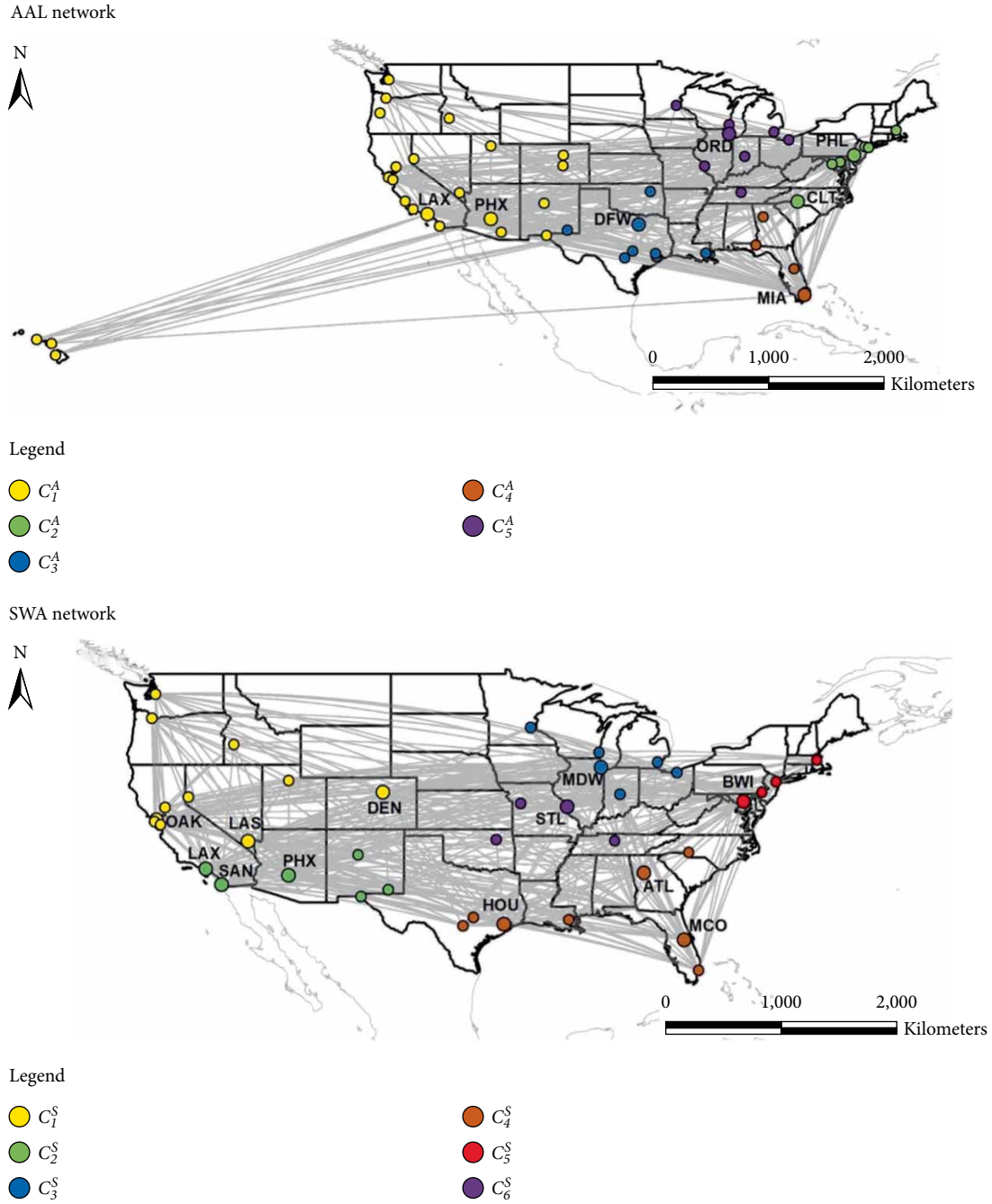
AAL network

SWA network

FIGURE 3: Community detection of AAL and SWA airline networks.

4.1. Core Community. The link between city size and air service is complex; and the way airlines serve cities of different sizes is worth exploring [30]. The correlation coefficient between airports impacts on how the various nodes relate and link to one another, and ultimately determine the configuration of the entire network structure. In this paper, w_{uv} between two airports is used as an edge weight for the core airport correlation coefficient $r(u)$ and the community correlation coefficient $r(C_i^A)$ is calculated based on a network metric, i.e. weighted betweenness centrality, which indicates a node's ability to stand between other nodes, and therefore, to control

the flows among them. Betweenness centrality of node u is calculated as follows:

$$B(u) = \sum_{s \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}}. \tag{10}$$

Where $\sigma_{st}(u)$ is the number of the shortest paths between node s to node t that pass through node u, and σ_{st} is the overall number of paths between nodes s and t [21]. The airport correlation coefficient $r(u)$ and the community correlation coefficient $r(C_i^A)$ can be presented respectively as:
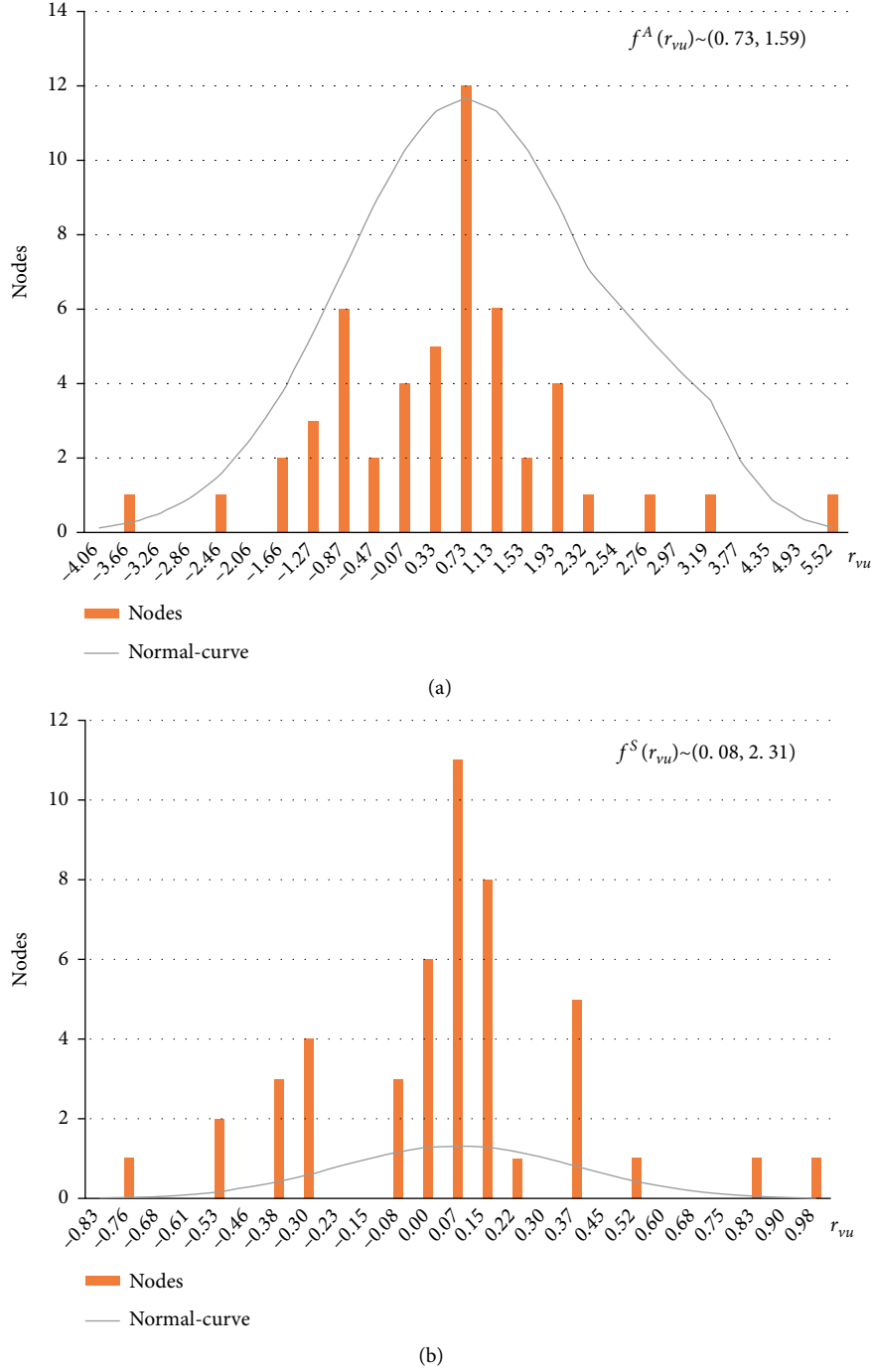
(a)

(b)

Figure 4: Correlation coefficient distribution of AAL and SWA. (a) Correlation coefficient distribution of AAL network. (b) Correlation coefficient distribution of SWA network.

$$\overline{r(u)} = \sum_{v} \frac{<B(u)B(v)> - <B(v)><B(u)>}{\sqrt{\left(<B(u)^2> - <B(v)>^2\right)\left(<B(v)^2> - <B(u)>^2\right)}}, \quad (11)$$

$$\overline{r\left(C_i^A\right)} = \sum_{u \in C_i^A} \overline{r(u)}. \quad (12)$$

Therefore, it is worth examining the average correlation coefficient of each community in more detail and to draw some conclusions from this Tables 1 and 2.

In Table 1, size refers to the number of airports in a community and degree refers to the degree of a community, which is the average degree of all nodes in this community. The internal and external edge respectively presents the quantity of edges that is inside and outside a community. The average community correlation coefficient $\overline{r\left(C_i^A\right)}$ represents the centrality and influence of a community. In Table 1, the core communities of AAL have larger $\overline{r\left(C_i^A\right)}$ values and fewer core airports. This means that the community is dominated by a small number of hub airports that may lead to larger impacts on other communities. In particular, the largest average

TABLE 1: Community analysis of the AAL network.

AAL	Key airports	Core Community	$\overline{r\left(C_i^A\right)}$	Size	Degree	Internal edges	external edges
C_1^A	LAX, PHX	√	0.84	24	492	178	136
C_2^A	CLT,PHL	√	0.72	7	147	21	105
C_3^A	DFW		0.63	7	121	19	83
C_4^A	MIA		0.54	4	49	6	37
C_5^A	ORD	√	0.79	10	209	44	121

TABLE 2: Community analysis of the SWA network.

SWA	Key airports	Core community	$\overline{r\left(C_i^S\right)}$	Size	Degree	Internal edges	external edges
C_1^S	DEN, LAS, OAK	√	0.75	11	366	50	266
C_2^S	LAX, PHX, SAN	√	0.69	6	148	15	118
C_3^S	MDW		0.61	5	145	10	125
C_4^S	ATL, HOU, MCO	√	0.77	9	280	36	208
C_5^S	BWI		0.51	5	174	10	154
C_6^S	STL		0.43	3	86	3	80

TABLE 3: Change in community correlations when top 10 airports are 'attacked.'

	AAL network				SWA network		
Top	Airport	$\overline{r(u)}$	$\left\|\overline{r\left(C_i^A\right)} - \overline{r\left(C_i^A\right)}'\right\|$	Top	Airport	$\overline{r(u)}$	$\left\|\overline{r\left(C_i^S\right)} - \overline{r\left(C_i^S\right)}'\right\|$
1	DFW	0.25	0.34	1	LAX	0.37	0.23
2	CLT	0.17	0.04	2	BWI	0.26	0.30
3	ORD	0.11	0.29	3	MDW	0.22	0.27
4	PHL	0.10	0.06	4	DEN	0.20	0.11
5	LAX	0.07	0.09	5	LAS	0.1	0.11
6	MIA	0.21	0.22	6	HOU	0.1	0.14
7	PHX	0.04	0.10	7	PHX	0.09	0.23
8	BOS	0.02	0.18	8	ATL	0.09	0.14
9	JFK	0.02	0.27	9	OAK	0.09	0.17
10	LAS	0.02	0.22	10	MCO	0.09	0.15

coefficient of community C_1^A (0.84) reveals a strong correlation between airports in this community. This demonstrates strong connection between airports in this community. In other words, internal flights of community C_1^A are more easily transmitted, and therefore it is the most frequently contacted community.

However, the correlation coefficient of the core community of the SWA network is much smaller than that of the AAL network. Its highest value is 0.75 for community C_1^S and the lowest value is 0.43 for community C_6^S, which shows a weaker connection between airports inside the community. The largest community in the AAL network is C_1^A which includes 24 airports, while the largest community of SWA is C_1^S which has only 11 airports but is dominated by three key airports.

It can be seen that the core community of SWA has more core airports and less coefficient values. This is because flights are distributed uniformly to several main airports and more

flights are connected outside communities. Therefore, the core community of SWA has less impact on other communities of these airports compared to the core community of AAL.

4.2. Core Airport. The correlation coefficient of a single node can be understood as a measure of centrality and influence, where the highest values are usually matched to the best connected node. This is because the best connected airports have a greater impact on the network, as they can "control" a significant number of flights. Furthermore, geographical distance between nodes focuses on their 'ease-of-access' to other nodes. The core airport is determined by the ranking of the airport correlation coefficient $\overline{r(u)}$, which considers both the correlation impact inside and outside the community. In Table 3, a list of ranked airports in the AAL network and the SWA network is provided to show the important connectivity of airports inside and outside the community. It also shows the absolute change in value of community correlations (i.e.

TABLE 4: Key airports in different communities.

	AAL				SWA			
	$\overline{d_{inside}^A}$	$\overline{d_{outside}^A}$	$\overline{r_{inside}^A}$	$\overline{r_{outside}^A}$	$\overline{d_{inside}^S}$	$\overline{d_{outside}^S}$	$\overline{r_{inside}^S}$	$\overline{r_{outside}^S}$
PHX	23	19	3.49	2.81	5	26	0.21	0.19
LAX	23	15	3.12	2.82	5	28	0.18	0.16

$\overline{r(C_i^A)}$ and $\overline{r(C_i^S)}$) when these different airports are "attacked" (removed from the network) at each time. $\overline{r(C_i^A)}'$ and $\overline{r(C_i^S)}'$ denotes the community correlations after removal.

The top 10 airports are the key airports in their own communities. For both AAL and SWA networks, the removal of a core airport has less impact on the core community, especially those with more than two core airports. This is because of the homogeneity of flight scheduling on these routes and that most of the flights can be rerouted. Therefore, core communities often have better robustness. For non-core communities, their dominant airports such as DFW in C_3^A and BWI in C_5^S, have a greater impact on the overall connectivity of communities, especially when they are the only leading airport in the community. In addition, the average community correlation decline in AAL network is 0.131, while it is 0.185 in SWA network. This indicates that the network structure of AAL is robust against airport failures. The smaller decrease in correlation value for each airport also indicates that the removal of the key airports might not easily reduce the connectivity of other airports both inside or outside the community. Table 4 shows the results for two specific airports (PHX and LAX) in the AAL network used to investigate the impact of key airports in different communities.

In Table 4, $\overline{d_{inside}}$ refers to the number of airports that are connected with PHX/LAX airport in the same community and $\overline{d_{outside}}$ refers to the number of airports that are connected outside. The difference in core airports between the two networks is clear. In the AAL network, PHX dominates 42 airports with an average correlation coefficient of 3.49, with 23 airports inside the community and 19 outside. The correlation coefficient between PHX and the 23 connected airports within the community is 0.68 more than the correlation coefficient outside the community. This shows the dominate position of PHX airport in community C_1^A.

In the SWA network, PHX only dominates 5 airports inside the community but dominates 26 airports outside the community. However, the difference between the average inside and the average outside community correlation coefficient is only 0.02. This means that there is little difference between airports inside or outside the community. In other words, PHX is of the same importance in its own community as well as in the entire network.

As LAX is also a key airport in the same community as PHX in both airline networks, results were similar to that of PHX. However, the lower value of $\overline{r_{inside}}$, shows that LAX has a less prominent position compared to PHX.

The results indicate that the core airports in the AAL network have a stronger impact compared to airports in the same community. By contrast, the core airports in the SWA network have a stronger impact compared to airports outside of the community. The different relationships among communities have an important effect on the impact of core airports when the modularity values are equal.

4.3. Airport Dominance and Control. The detection method is also useful for analyzing airports that have an "ambiguous" geographical position (e.g., Kansas City International Airport, MCI). In previous studies, airport communities were divided by distance. However, distance is only one kind of airport clustering. In addition, nearby airports are considered to be in the same community, and therefore airports that are equal in distance to core airports can be easily confused. For example, the route distance of MCI-ORD and MCI-DFW are approximately equal; thus it is difficult to determine whether it is dominated by ORD or DFW when only distance is considered. However, results show that the route-traffic-based model can eliminate this confusion.

In the AAL network, MCI is dominate by ORD in community C_5^A. This is because flights between route MCI-ORD are more frequent, and ORD has a larger impact on traffic in MCI.

From Figure 5, it can be noted that as traffic in ORD and DFW increase from 5% to 40%, the difference between the two curves of r_{ij} becomes more noticeable. This indicates the control effect of ORD. Furthermore, apart from the core airports, other airports (such as STL and TUL) from the surrounding community have minimal impact on MCI.

In the SWA network, MCI is dominated by STL in community C_6^S, which is closer than the core airport MDW in community C_3^S. STL is one of the hub airports of the SWA airline and MDW is a secondary airport in the city of Chicago. However, SWA airline is easily influenced by MDW due to its frequent traffic flow. Therefore, airports have a stronger connection in the same community in the AAL network, while they are more likely to be influenced by other core airports in surrounding communities within the SWA network structure.

Effective community and airport detection is important for understanding the different patterns in airline network configurations in the consideration of geographical, air transport-political and economic factors such as supply, traffic demand and costs. Based on the results, two possible reasons can be identified for different community structure and hub airports in the two airline networks. First reason is the different network configurations. AAL airline is a full-service carrier, which offers a variety of services and network linkages; while SWA is a typical low-cost carrier, which offers a limited number of services in specific segments of the network such as regional airports at low prices and mainly dominate on a point-to-point basis. More short- and medium-range flight route is provided by SWA, which increases the connectivity of airports on a smaller scale. Second reason is the different operation patterns. The historical operation pattern and company strategies have placed limitations on its networks structure expansion, not only in terms of complementarity and competition, but also in terms of the target passengers that they attract. The higher frequency and flexibility in flight scheduling of SWA enable branch routes to become more important and have a greater impact on the robustness of the entire network.
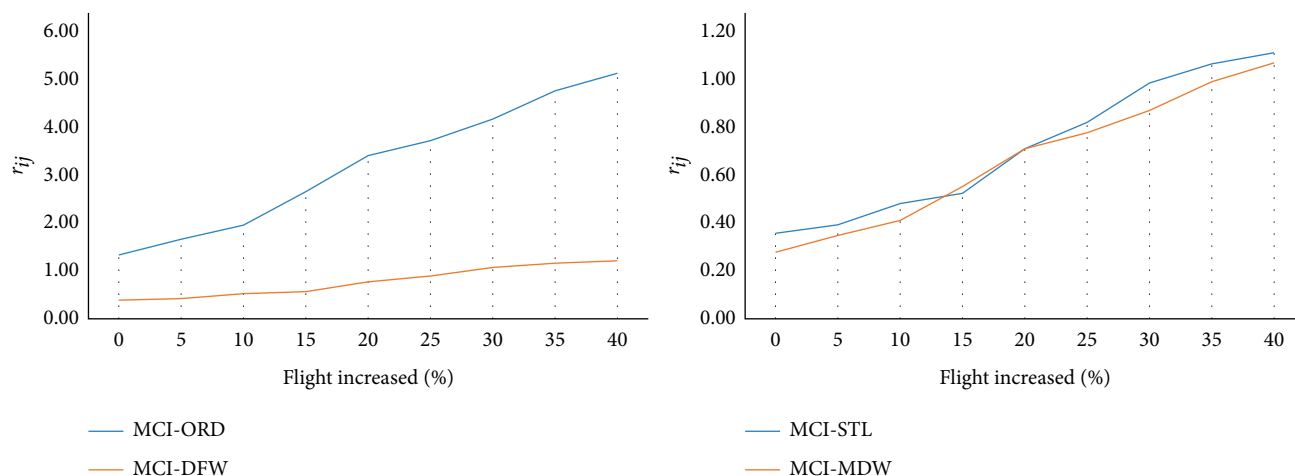
FIGURE 5: Sensibility analysis of the MCI correlation coefficient.

5. Conclusion

The purpose of this paper is to propose a route-traffic-based method of detecting communities in airline networks that have different operation patterns to identify their community structures that are both highly connected and spatially clustered.

Our model is an improved version of the Clauset-Newman-Moore (CNM) modularity maximization algorithm, where both route distance and passenger volumes are taken into account. This extension is useful because the relations among airports are defined not only based on their topological network structure, but also by traffic between connecting airports. The method was tested using the American Airlines and Southwest Airlines networks. Results show that the model is effective in analyzing the interaction and clustering properties for different airline patterns without diverging from realistic networks or mismatching airports. The three main findings are as follows:

(1) The route-traffic-based method of detecting communities, which considers the actual flight operation (airline passenger flows between routes) of airlines, is more accurate compared to network community detection using only the physical distance between airports. The core airports of the communities are consistent with the actual hub airports of the airlines. Moreover, there is a clearer and more explicit classification especially for those more geographically "ambiguous" (fuzzy) airports (i.e., airports that fall in the catchment area of different dominant airports), which is helpful to further study the relationship between airports.

(2) The differences in operating patterns of airlines lead to different network topological and community structures. This difference is mainly due to flight scheduling factors such as flight focus on trunk route or branch route, rather than objective factors such as route distance. For the overall structure of the network, the traditional full service airlines have fewer communities with a greater number of airports, sparse correlation/bridge edges and higher weights on the trunk route. Low cost airlines tend to have more communities with fewer airports, dense correlation/bridge edges and dispersion route-weight [31].

(3) For the connected airports inside and outside of communities: First, links between core airports (regardless of whether it is inside or outside of the community) are stronger in the network of full service airlines, and the impact of these core airports is greater for both internal and external nodes that are organized as communities. The connection between other nodes in different communities is much weaker, and the influence between core airports is much stronger than that of other airports in the same community. In other words, the connection between core airports is strong, and non-key airports are more likely to be affected by key airports outside of the community. Second, the connections between internal nodes in the same community of a low-cost airline network are closely linked. The airports are mostly "controlled" by the core airports in the same community, while the key airports are more affected by other airports both inside and outside of the community. In other words, the relationship between the key airports is weaker, and the non-key airports are mainly dominated by the key airports of its own community. That is, the key airports are more prominent in the community.

In this study, characteristics of the network connections were investigated. Nodes and communities with the most impact on the development of different network structures were identified, and how these affect the entire network was examined. Furthermore, several drivers on network structuring, including correlation coefficient and degree were analyzed to compare the different relationships between community structures and airline operating patterns.

Two issues have been identified in this study that can be addressed in future research. First, more airline networks can

be analyzed to further validate the application of our model. Different airlines have different networks and operation patterns. The effectiveness of the route-traffic-based community detection method can be further tested and refined using different airline networks. Second, more community detection variables can be considered. Although we explored route traffic as the main factor that determine the core airports and their dominance, future research can also consider other factors such as time and flight delay in the model.

Data Availability

The numeric data used to support the findings of this study are available from the corresponding author (Haoyu Zhang, Email: 1123851665@qq.com) upon request. The data sheets are saved in EXCEL which is collected from the historical data on website: https://zh.flightaware.com.

Conflicts of Interest

All authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contributions

Haoyu Zhang wrote the manuscript; Haoyu Zhang, Weiwei Wu and Frank Witlox provided the idea; Shengrun Zhang and Haoyu Zhang provide the data and case study; and Haoyu Zhang, Weiwei Wu, Shengrun Zhang and Frank Witlox revised the manuscript.

Funding

Supplementary Materials

The uploaded excel refers to data we used in the paper. (*Supplementary Materials*)

References

[1] Y. Chen, J. Xu, and M. Xu, "Finding community structure in spatially constrained complex networks," *International Journal of Geographical Information Science*, vol. 29, no. 6, pp. 889–911, 2015.

[2] T. Hathway, "Successful community participation in local traffic proposals," *Journal of Advanced Transportation*, vol. 31, no. 2, pp. 201–213, 1997.

[3] X. Sun, S. Wandelt, N. Dziku, and F. Linke, "Air passenger flow communities between countries," in *Control Conference*, IEEE, pp. 9276–9281, Chengdu, China, 2016.

[4] A. Ceder, "Operational objective functions in designing public transport routes," *Journal of Advanced Transportation*, vol. 35, no. 2, pp. 125–144, 2010.

[5] J. Wang, H. Mo, F. Wang, and F. Jin, "Exploring the network structure and nodal centrality of China's air transport network: A complex network approach," *Journal of Transport Geography*, vol. 19, no. 4, pp. 712–721, 2011.

[6] T. Jia and B. Jiang, "Building and analyzing the US airport network based on en-route location information," *Physica A Statistical Mechanics & Its Applications*, vol. 391, no. 15, pp. 4031–4042, 2012.

[7] D. Watts and S. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, pp. 440–442, 1998.

[8] L. Adamic, B. Huberman, A. Barabási, R. Albert, H. Jeong, and G. Bianconi, "Power-law distribution of the World Wide Web," *Science*, vol. 287, no. 5461, p. 2115, 2000.

[9] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[10] R. Stimson, S. Baum, P. Mullins, and K. O'Connor, "A typology of community opportunity and vulnerability in metropolitan Australia," *Papers in Regional Science*, vol. 80, no. 1, pp. 45–66, 2001.

[11] Y. Sun, L. Mburu, and S. Wang, "Analysis of community properties and node properties to understand the structure of the bus transport network," *Physica A Statistical Mechanics & Its Applications*, vol. 450, no. 450, pp. 523–530, 2016.

[12] E. Gegov, M. N. Postorino, M. Atherton, and F. Gobet, "Community structure detection in the evolution of the United States airport network," *Advances in Complex Systems*, vol. 16, no. 16, p. 1350003, 2013.

[13] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, 2005.

[14] M. N. Postorino and M. A. Versaci, "Geometric fuzzy-based approach for airport clustering," *Advances in Fuzzy Systems*, vol. 2014, no. 2, pp. 1–12, 2014.

[15] W. Cong, M. Hu, B. Dong, Y. Wang, and C. Feng, "Empirical analysis of airport network and critical airports," *Chinese Journal of Aeronautics*, vol. 29, no. 2, pp. 512–519, 2016.

[16] J. Wang and Y. Jing, "Comparison of spatial structure and organization mode of inter-city networks from the perspective of railway and air passenger flow," *Acta Geographica Sinica*, vol. 72, no. 8, pp. 1508–1519, 2017.

[17] A. Reggiani, P. Nijkamp, and A. Cento, "Connectivity and concentration in airline networks: a complexity analysis of Lufthansa's network," *European Journal of Information Systems*, vol. 19, no. 4, pp. 449–461, 2010.

[18] B. X. Zhu, *Analysis of the structure and evolution of China's aviation network [Ph.D., thesis]*, Fudan University, China, 2009.

[19] J. Wang, H. Mo, and F. Jin, "Spatial structural characteristics of Chinese aviation network based on complex network theory," *Acta Geographica Sinica*, vol. 64, no. 8, pp. 899–910, 2009.

[20] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3747–3752, 2003.

[21] X. Q. Sun, V. Gollnick, and S. Wandelt, "Robustness analysis metrics for worldwide airport network: A comprehensive study," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 500–512, 2017a.

[22] X. Q. Sun, S. Wandelt, and F. Linke, "On the topology of air navigation route systems," *Proceedings of The Institution of Civil Engineers-Transport*, vol. 170, no. 1, pp. 46–59, 2017b.

[23] S. Wandelt, X. Q. Sun, D. Z. Feng, M. Zanin, and S. Havlin, "A comparative analysis of approaches to network-dismantling," *Scientific Reports*, vol. 8, no. 1, p. 13513, 2018.

[24] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 70, no. 2, p. 066111, 2004.

[25] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.

[26] X. Wang, G. Liu, J. Li, "A detecting community method in complex networks with fuzzy clustering," in *International Conference on Data Science and Advanced Analytics (DSAA*, IEEEShanghai, China, 2014.

[27] US Government Accountability Office, *Airline Mergers: Issues Raised by the Proposed Merger of American Airlines and US Airways*, Government Accountability Office Reports, Washington, D.C, USA, 2013.

[28] S. Zhang, B. Derudder, and F. Witlox, "The impact of hub hierarchy and market competition on airfare pricing in US hub-to-hub markets," *Journal of Air Transport Management*, vol. 32, no. 3, pp. 65–70, 2013.

[29] K. Gopalakrishnan, H. Balakrishnan, and R. Jordan, "Clusters and communities in air traffic delay networks," in *American Control Conference*, IEEE, pp. 3782–3788, Boston, MA, USA, 2016.

[30] K. O'Connor and K. Fuellhart, "Change in air services at second rank cities," *Journal of Air Transport Management*, vol. 28, pp. 26–30, 2013.

[31] Y. Min, X. Jin, Y. Ge, and J. Chang, "The role of community mixing styles in shaping epidemic behaviors in weighted networks," *PLoS One*, vol. 8, no. 2, p. e57100, 2013.

[32] S. Fortunato and C. Castellano, "Community detection in graphs," in *Encyclopedia of Complexity and Systems Science*, R. A. Meyers, Ed., vol. 1, Springer, Berlin, Germany, 2009, eprintarXiv:0712.2716.