

## **COMMUNITY STRUCTURE DETECTION IN THE EVOLUTION OF THE UNITED STATES AIRPORT NETWORK**

**EMIL GEGOV**

*School of Engineering and Design, Brunel University, Kingston Lane, Uxbridge UB8 3PH, UK*

*emil.gegov@brunel.ac.uk*

**M. NADIA POSTORINO**

*Department of Informatics, Mathematics, Electronics and Transport, University of Reggio Calabria, Reggio*

*Calabria 89122, Italy*

*npostorino@unirc.it*

**MARK ATHERTON**

*School of Engineering and Design, Brunel University, Kingston Lane, Uxbridge UB8 3PH, UK*

*mark.atherton@brunel.ac.uk*

**FERNAND GOBET**

*School of Social Sciences, Brunel University, Kingston Lane, Uxbridge UB8 3PH, UK*

*fernand.gobet@brunel.ac.uk*

This paper investigates community structure in the US Airport Network as it evolved from 1990 to 2010 by looking at six bi-monthly intervals in 1990, 2000 and 2010, using data obtained from the Bureau of Transportation Statistics of the US Department of Transport. The data contained monthly records of origin-destination pairs of domestic airports and the number of passengers carried. The topological properties and the volume of people travelling are both studied in detail, revealing high heterogeneity in space and time. A recently-developed community structure detection method, accounting for the spatial nature of these networks, is applied and reveals a picture of the communities within. The patterns of communities plotted for each bi-monthly interval reveal some interesting seasonal variations of passenger flows and airport clusters that do not occupy a single US region. The long-term evolution of the network between those years is explored and found to have consistently improved its stability. The more recent structure of the network (2010) is compared with migration patterns among the four US macro-regions (West,

Midwest, Northeast and South) in order to identify possible relationships and the results highlight a clear overlap between US domestic air travel and migration.

*Keywords:* Air transportation; community structure; United States Airport Network; migration.

## 1. Introduction

A common approach to dealing with complex systems is to make use of network theory to simulate symmetric or, more generally, asymmetric relations among discrete objects. Transportation [19], the Internet [1], mobile phone [21], power [15], social, and neural networks are just some examples of complex systems where network theory has been applied. In fact, complex systems are often modelled as complex networks, i.e. graphs with non-trivial topological characteristics, because this provides a powerful abstraction that can eliminate the unnecessary complexity of the system while maintaining the key properties and interactions. Complex networks usually have nodes and links arranged in space, but topology alone does not contain all the vital information and the spatial constraints could, and typically do, affect the structure and properties of these spatial networks (see [6] for an overview).

Transportation networks are a good example of spatial networks. Their network topology is not only characterised by spatial aspects such as the location of nodes and the length of links but also by the association of a “transport cost” to the link length; implying that longer links are typically balanced by some benefit, such as connecting to a high-degree node, or a node in an attractive location. Transportation networks typify the specific nature of spatial networks particularly with regard to issues such as congestion, fast-growing urban sprawl and disease propagation. Network structure and dynamics play a key role in most, if not all, of these challenges. Transportation networks can be planar (not to be confused with *planar graphs*), as in road and rail networks, or non-planar, as in airport networks. In addition, transportation networks are usually weighted, where the link weight describes the intensity of some form of interaction, e.g. the amount of traffic. Air transportation networks are an important example of spatial networks. Nodes identify airports and links represent the existence of a direct air service among them. Weights on links may represent the number of passengers flying on that connection, and the distribution of weights is an initial indication of the existence of possible strong heterogeneities [5].

The existence of links among airports depends on factors related to both airline strategies and passenger demand. Airlines decide to operate at a given airport on the basis of a significant demand, allowing them to reach satisfactory load factors. Location and socio-economic characteristics of the airport catchment area are the key factors generating air traffic demand. The airport choice made by both airlines and travellers depends on factors that can be ultimately reduced to time and monetary costs. For example, reduced airport charges may help airlines to offer lower air fares to potential travellers, and hence to induce more flights. The airport network is an example of a heterogeneous network where the hubs have high connectivity, high weight (in terms of traffic) and long-distance links [5].

In recent years, the analysis of complex transport networks has received considerable attention, mainly in terms of commuting networks [19, 35, 37]. Airport networks have also been studied to characterise their level of degree correlations and clustering, their evolution in time, and their potential scale-free properties [3, 22, 24, 40]. Concerning airport networks, it is of interest to look at their *community structure*, which is a prominent feature in many biological [33], social [7] and technological [7] complex systems. Community structure is defined as the presence of highly *intra*-connected *modules* of nodes that are loosely *inter*-connected to the rest of the network. In other words, nodes are organised in clusters and most links are inside those clusters. The reason for this phenomenon is that nodes that share functional similarity and/or dependency tend to interact more and therefore they should be more connected. There are two main advantages of this community architecture: the first is efficiency, as most interactions are within modules which are internally well-connected, thereby reducing the path length (the number of links that separate a pair of nodes); and the second is robustness, as entire modules may fail autonomously, without severely affecting the operation of other modules, and hence, the function of the entire network. Therefore, the emergence of community structures in airport networks has implications for their efficiency and robustness, as well as their socio-economic characteristics. In terms of network robustness, network failure due to external factors such as bad weather conditions, volcanic eruptions, and political or security issues, may have significant impact on the air traffic depending on the criticality of the involved nodes and the extent of their influence. In terms of socio-economic characteristics, the emergence of community structure depends on the location and distribution of relevant activities. Concentration of activities in a given area generally means concentration of short trips in

that area, and this is a typical commuting pattern. For medium-long distance trips, the main contributing factor is mass migration rather than commuting, and air transportation plays an important role in facilitating easier migration of workers. Within larger countries, such as the United States, a new kind of “commuting by air” can be identified, as people working in different parts of the country during the week return home at weekends. The changes in the availability, frequency and cost of air travel facilitate trips for migrants located far from traditional gateways (large airports with hub functions (hub-and-spoke) and inter-continental links) [12].

Over the past few decades air travel in the US has changed considerably. Apart from the obvious increase in the number of airports, connections and passengers, the structure (topology) of the US Airport Network (USAN) has transformed, thereby affecting all aspects of air travel. Up to the 1970s the USAN had mainly a hub-and-spoke architecture: flights coming from many origins (spokes) converge to the airport (hub) from which new flights start toward other destinations (spokes). The hub-and-spoke architecture is characterised by a high spatial network concentration, and a time co-ordination of flights at the hub according to a “flight wave” concept [11]. The ideal wave is the set of arriving and departing flights such that for each arriving flight there is a departing one allowing travellers to get an easy transfer to the final destination, and the integration of air services at the hub (e.g. baggage transfer). The main disadvantage of the hub-and-spoke architecture for passengers is that they would have to change flights at the hub, taking more time to reach their final destination. Furthermore, passengers travelling between other destinations may experience poor service, including infrequent flights and many changes [25]. As a result, a number of low-cost airlines emerged in the 1980s, providing point-to-point direct services between poorly connected destinations. One example is JetBlue, which is still considered very successful even when compared against larger airlines, such as American Airlines and United Airlines [8]. Consequently, the resulting USAN topology is a combination of both hub-and-spoke and point-to-point architectures.

Migration can be thought of as population redistribution within a country or between countries. It is often linked to an asymmetric distribution of employment and affluence: people are attracted to areas with better job markets, services and quality of life. These aspects relate to the concept of “city competitiveness”, in other words, attractive cities (or regions) are efficient, accessible and offer economic opportunities to both investors and workers [9, 14, 16, 17]. In terms of accessibility, attractive areas have efficient

transport systems mainly in terms of external connections linking those areas to other parts of a large territory. In this context, air services can play an important role because they provide fast links among distant locations, even though there may be alternative forms of transportation. In large countries, such as the US, the domestic airport network is a key factor in facilitating domestic migration, i.e. the movement of people within the United States. Particularly, migrants are defined as people moving among states (inter-state migration). Incoming migration (in-migration) is defined as movements into an area during a given period, while outgoing migration (out-migration) is defined as movements out of an area during the same period.

The aim of this paper is to identify a coherent community structure in the evolving USAN, and to use it to investigate the relationship between air travel and migration within the US. Section 2 presents the research methodology; section 3 summarises the main results; section 4 provides a detailed discussion; and section 5 concludes this paper.

## **2. Research Methodology**

This section presents the network model and the current state-of-the-art in community structure detection. In addition to a description of the USAN network model, there are details about the origin of the data being modelled, the mechanisms of community structure, and a justification for choosing a particular method for detecting spatially-independent communities.

### **2.1. Network Model**

To investigate the evolution of the USAN from 1990 to 2010 the network is modelled in a discrete time-series consisting of three stages: 1990, 2000 and 2010. Each of those is further split into six bi-monthly intervals, in order to capture finer temporal detail and to explore seasonal variations in the network. Hence, the network model consists of 18 network snapshots depicting topology and traffic for a two-month time-slice. Each network is defined by a set of nodes (the airports) and a set of links (the direct flight connections), representing topology. In addition, the links are weighted by the total number of passengers (both inbound and outbound) that flew on that connection within the specified time-slice. Data were obtained from the Bureau of Transportation Statistics (BTS) [10], of the US Department of Transport. More specifically, data contained monthly records of origin-destination pairs of domestic airports and the number of passengers carried. Using network modelling both dynamics *on* the network in terms of

traffic fluctuations and dynamics of the network in terms of topology fluctuations are studied. The more recent structure of the network (reference year: 2010) will be compared with migration patterns among the four US macro-regions (West, Midwest, Northeast and South), in order to identify possible relationships. Fig. 1 shows a map of the US regions and states, including the locations of the main airports in terms of 2010 passenger flows.

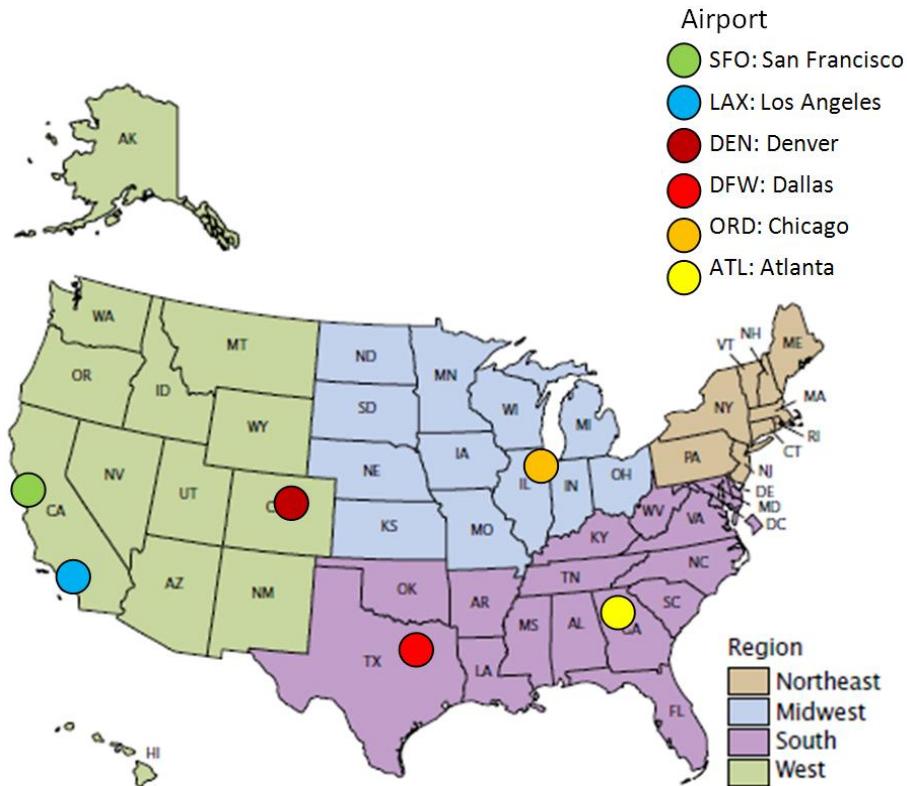


Fig. 1. US macro-regions and major airports in 2010. Adapted from [31].

## 2.2. Community Structure

In recent years, research on complex networks has proposed many community detection methods [28] that aim to discover the most sensible partition of a network into communities. Most of them work on the principle of modularity [34] optimisation, aiming to maximise the modularity benefit function describing the quality of a network partition. The more links that fall within a community compared to an ensemble of

benchmark random networks with the same community structure, then the more bias there is for links to connect to nodes belonging to the same community, and therefore the higher the modularity  $Q$  (Eq. 1). In essence, modularity measures how sharply the modules are defined.

$$Q = (\text{fraction of links within communities}) - (\text{expected fraction of such links}) \quad (1)$$

The expected fraction of links within communities is calculated from an ensemble of random networks that resemble the network under scrutiny in terms of its strength (total weight on all adjacent links) distribution. In addition, it is necessary to quantify the average level of interaction between a pair of nodes, and this is achieved by defining a null model matrix  $P_{ij}$  that describes the expected weight of a link between nodes  $i$  and  $j$ , over the ensemble. The standard choice for  $P_{ij}$ , defined by Newman and Girvan (NG) [34], preserves the strength of nodes in the random networks:

$$P_{ij}^{\text{NG}} = k_i k_j / 2m \quad (2)$$

where  $k_i$  is the strength of node  $i$  and  $m$  is the total weight in the network. A limitation of this null model, and of community detection methods that use it, is that only network topology and traffic are considered, but this is insufficient for networks embedded in space, such as the USAN. The reason for this is that most spatial networks (excluding the Internet for example) are very biased towards short-range connections due to the cost involved in long-range interactions in physical space. In terms of topology, an airport network is not a typical spatial network, as long-range connections are common. However, in terms of traffic, the higher financial and temporal costs involved in long-range travel play an important role for passengers, thereby affecting the flow on the network. Hence, standard community detection methods (typically based on the NG null model) will discover communities of nodes that are spatially close, as opposed to communities that have particularly strong internal interactions [29, 13, 4, 2, 20]. To address this, Expert et al. [21] proposed an alternative null model for  $P_{ij}$  that takes into account the effect of space by favouring communities of nodes  $i$  and  $j$  that are more connected than expected, given the physical distance  $d_{ij}$  between them:

$$P_{ij}^{\text{Spa}} = N_i N_j f(d_{ij}) \quad (3)$$

where  $N_i$  is the importance (typically the strength) of node  $i$  and  $f(d_{ij})$  is the function that incorporates the effect of space. This so-called deterrence function describes the expected level of interaction between nodes  $i$  and  $j$  that are separated by some distance  $d_{ij}$ . In other words, the function defines how interaction decays, analogous to gravity, as distance between objects increases.

Expert's null model has been shown to uncover space-independent community structure in the Belgian mobile network [21], where the NG null model fails to do so. Hence, to identify coherent community structure that reflects particularly high-traffic connections while also considering spatial effects, Expert's null model is applied to the USAN network model. The inputs are the adjacency matrix  $A_{ij}$  (encoding topology and passenger flows), the distance matrix  $D_{ij}$  (containing the Euclidean distance between all pairs of airports), the importance vector  $N_i$  (holding the passenger flow at each airport), and the bin size, which is used to bin the data from the distance matrix. Due to distance being expressed in terms of degrees of arc length, where one degree is approximately 60 miles, the largest distance in the distance matrix is 149. Therefore, it is necessary to select a bin size such that the bins are sufficiently populated, without losing too much spatial resolution. The bin populations and the deterrence function were checked for bin sizes 0.1, 1, 2 and 3, and 1 was chosen as it provided balanced bin populations and a smooth deterrence function. The output of Expert's null model is the modularity matrix  $Q_{ij} = (A_{ij} - P_{ij}^{Spa})$ , which is then fed into a community detection algorithm [27] (Expert et al. used the same algorithm in [21]) that searches for a network partition, maximising modularity.

### 3. Results

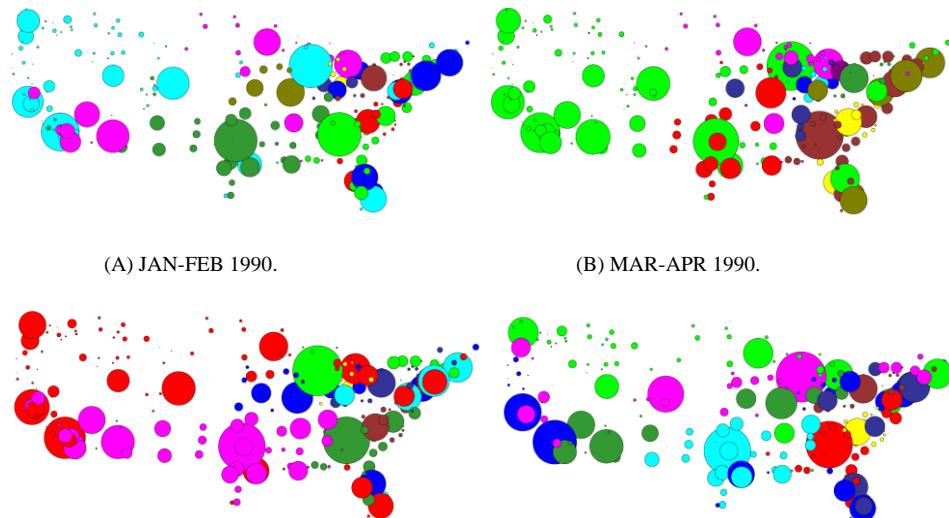
Expert's model is used in each of the 18 USANs (representing topology and passengers for a bi-monthly period). The output of the community detection algorithm is a vector, assigning each airport to a specific community in which all members have particularly strong interactions in terms of passenger flows between them, given their physical separation. Due to the potentially large number of nearly-optimal partitions [23], a non-deterministic implementation of the algorithm is applied twice to each network snapshot, in order to discover better partitions and to check their stability (similar partitions for the same snapshot). This is achieved using Normalised Variation of Information (NVI) [32], which measures the distance between two partitions in the range 0-1 (0 if they are

identical, approaching 1 if they are very different). The average NVI values across the six snapshots for the years 1990, 2000 and 2010 are 0.40, 0.34 and 0.26, respectively. These values indicate that the community detection is considerably stable.

Figs. 2-4 represent the USAN at various stages over time, where each airport is denoted by a circle with a surface area that is directly proportional to the passenger flow (inbound and outbound passengers), and the colour represents the community. Airport connections and airport-to-airport flows are not shown for clarity, and colour is not consistent across the networks as it is only used to differentiate between different communities in a single network (the software used does not allow the user to consistently assign colours to communities). In other words, the figures below depict the size of airports by passengers handled, and the groups of identically coloured airports that have particularly strong passenger flows between them. Alaska, Hawaii and the Mariana Islands are not shown here but they represent a very small fraction of the network. The airport in the bottom right is for the Virgin Islands. In the following analysis of results, the term “hub” is used to describe an airport that handles a high volume of passengers, and the terms “community” and “cluster” are used interchangeably.

### 3.1. Year 1990

Fig. 2 depicts bi-monthly snapshots of the USAN for the year 1990.



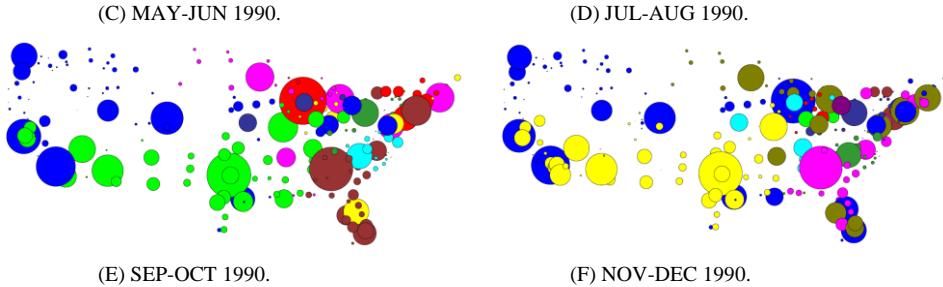


Fig. 2. Community structure in the USAN in the year 1990. Colour denotes community and node surface area is proportional to passenger flow at airport.

In Jan-Feb (Fig. 2(A)) there is a well-defined cyan community of west-coast airports, such as Los Angeles (LA) and San Francisco, together with Chicago, indicating high passenger mobility between those locations. In Fig. 2(B) the network for Mar-Apr implies a particularly large community (light-green) of the main US airports. This means that there were particularly active interactions between all the light-green locations during this time, in contrast to the previous image for Jan-Feb. May-Jun in Fig. 2(C) displays a geographically clustered set of communities in the east, together with the largest community in red which spans almost the entire US. In other words, the geographically clustered communities represent the regions where passengers mainly flew locally, and the red community refers to long-distance passengers. Jul-Aug (Fig. 2(D)) shows a very inter-mixed network, with significant long-distance travel suggested by the spatial spanning of the communities. However, the cyan Dallas cluster is an exception, as it covers only Dallas and small nearby airports. Sep-Oct (Fig. 2(E)) sees an overall decline in air travel flagged by the noticeable reduction in general size of circles, matching the end of the tourist season, and two large communities in blue and green. In Fig. 2(F) Nov-Dec has no major change in traffic patterns apart from the fact that Chicago (a key US hub) is taken over by the spanning blue community, implying that it was used extensively for air travel, particularly among these blue regions.

### 3.2. Year 2000

Fig. 3 depicts bi-monthly snapshots of the USAN for the year 2000.

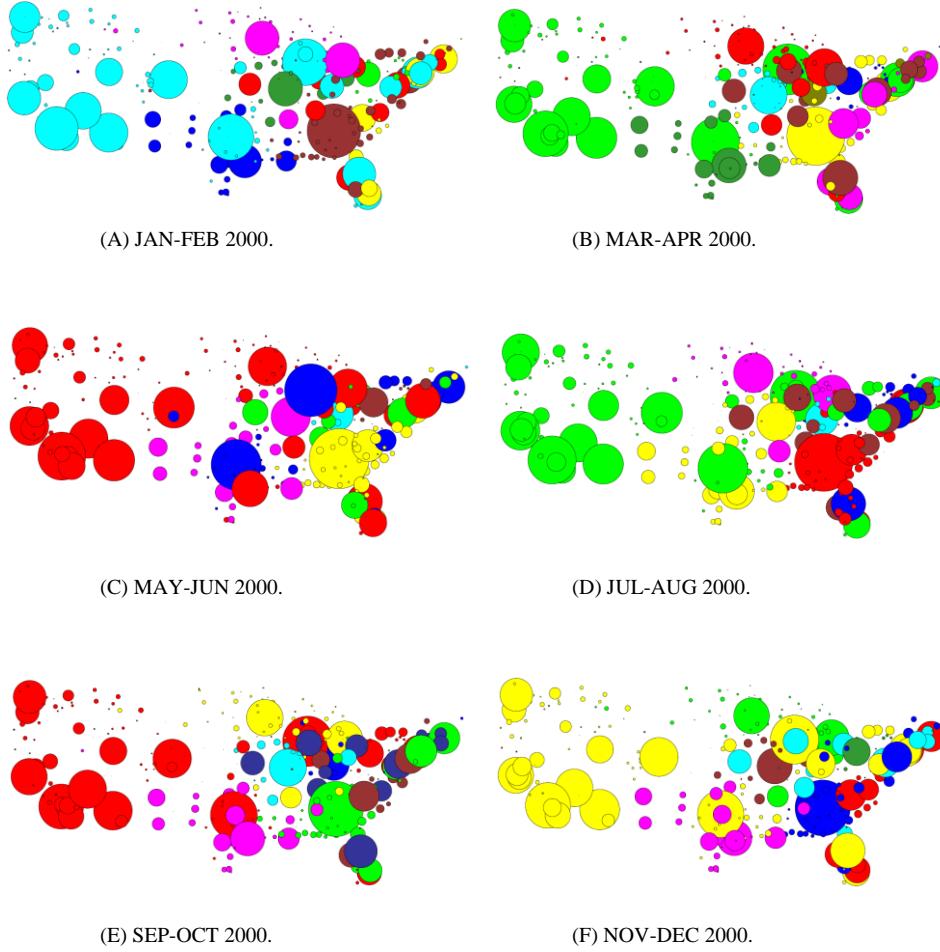


Fig. 3. Community structure in the USAN in the year 2000. Colour denotes community and node surface area is proportional to passenger flow at airport.

Jan-Feb in Fig. 3(A) displays a prevailing cyan community of most major airports dominating the west and a large part of the rest of the US. In Fig. 3(B), Mar-Apr displays a very similar pattern but the number of passengers has increased, which is reflected by the larger circles. In particular, yellow Atlanta (ATL) is clearly the leading US airport in terms of passengers handled during this period. May-Jun in Fig. 3(C) suggests that Dallas and Chicago have separated from the largest community in the previous image, forming their own community (in blue) with a few more airports in the north-east. Again, Atlanta is nearly the only member of its yellow cluster, but its size implies that it

plays the role of the main hub in the US, connecting many of the other regions. This will be explored in more detail in the discussion section. Jul-Aug (Fig. 3(D)) appears similar to the networks for Jan-Apr, with a main green cluster covering most of the US and Atlanta still on its own. In Fig. 3(E) Sep-Oct the number of passengers has predictably decreased. The east appears to be mixed while the west, Dallas and Chicago are all part of the same red cluster. Nov-Dec in Fig. 3(F) is similar to the previous network for Sep-Oct.

### 3.3. Year 2010

Fig. 4 depicts bi-monthly snapshots of the USAN for the year 2010.

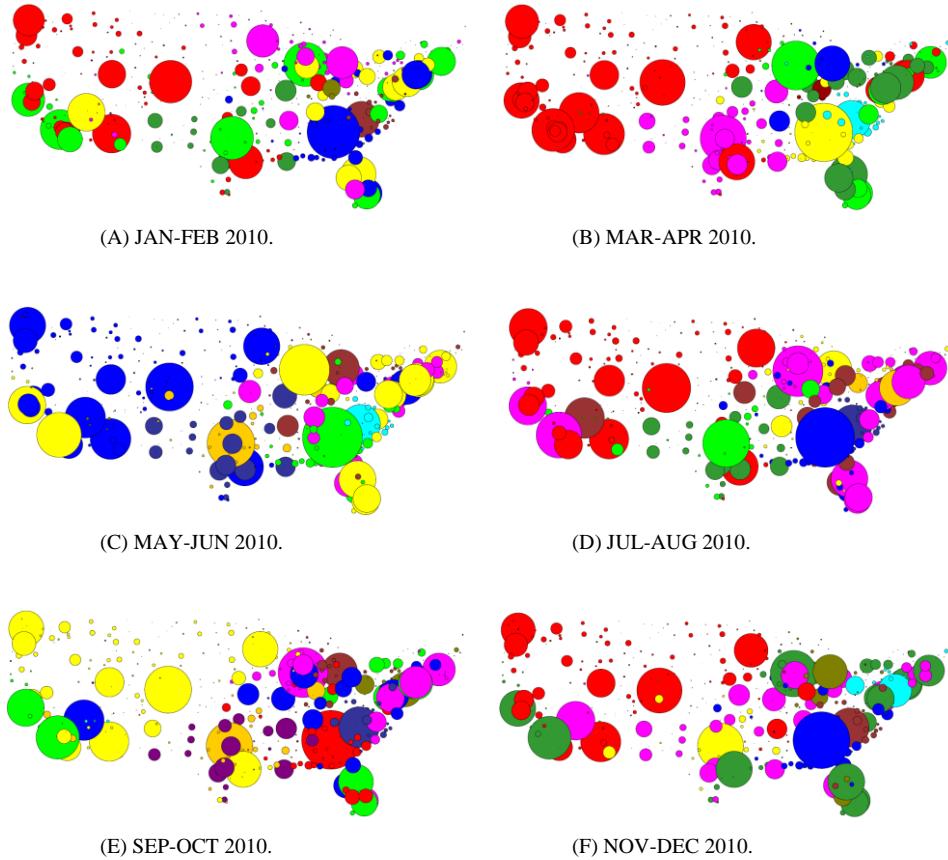


Fig. 4. Community structure in the USAN in the year 2010. Colour denotes community and node surface area is proportional to passenger flow at airport.

Fig. 4(A) Jan-Feb has two large clusters in red and green covering the west and a large part of the US, respectively. Atlanta (blue) is still the largest hub but passenger demand is low due to the low season. Mar-Apr in Fig. 4(B) shows an increase in passengers and a clearly dominating red community in the west. The south is covered by the pink Dallas cluster, and yellow Atlanta and light-green Chicago are the first and second largest hubs, respectively. May-Jun in Fig. 4(C) is different in two respects. Firstly, Chicago has formed a yellow cluster covering the south-west and the east, and secondly, orange Dallas has separated from the south cluster, so it has become more of a long-distance travel airport than in the previous two months. Atlanta is still the largest airport by far, providing the connections for many more passengers than any other airport in the US. Jul-Aug (Fig. 4(D)) is very similar to May-Jun. This means that there is a particularly high volume of travellers among the east coast, the west coast and Chicago. Sep-Oct (Fig. 4(E)) has a good mix of many clusters, suggesting that during these months there has been more long-distance travel within the US. The green, yellow and blue communities are particularly well spread out, highlighting the extent of long-range travel. Nov-Dec (Fig. 4(F)) is similar to the previous two months but now the Chicago and LA clusters have merged again (see May-Jun and Jul-Aug), forming one of the two largest clusters (red and green).

#### **4. Discussion**

First of all, it is important to highlight the fact that some communities have airports that are very far apart, suggesting that spatial community detection discovers more meaningful communities that are not occupying a single region on the map. The seasonal variation within each of the three years, and the long-term evolution of the network between those years, are explored in the following two sub-sections.

##### ***4.1. Seasonal Variation***

The seasonal variation in passenger flows within each year is investigated qualitatively by visually examining the obtained community structure, and quantitatively, using Normalised Mutual Information (NMI) [18].

In terms of qualitative analysis, there appear to be significant changes in the community structure of the USAN in 1990. In other words, there were considerable seasonal variations in the volume of passengers on network connections. Specifically,

Jan-Feb had a very mixed structure, Mar-Apr had a large (green) super-cluster, and the rest of the year was mixed again, with some similarities between May-Jun and Sep-Oct. In the last two months of the year, Chicago joined the blue LA cluster, forming a similar structure to Jan-Feb, which indicates the presence of an annual cycle of passenger demand. Throughout 2000 (apart from May-Jun), the community structure remained fairly stable, implying low seasonal variation. In particular, the network had a large super-cluster covering most of the US, and Atlanta was the super-hub. May-Jun, however, was different as Dallas and Chicago were in a separate cluster of their own, so there was a particularly strong passenger flow between them and other smaller airports in the north-east, during these months. In 2010, similarly to 1990, there were notable fluctuations in the community structure of the network. Jan-Feb was mixed, Mar-Apr had a dominant red cluster, and in the rest of the year there were two dominant clusters (Denver and Chicago). LA and San Francisco formed their own community in green in Sep-Oct.

Quantitative analysis of network snapshots involves NMI, which measures the similarity between two network partitions (in this case two consecutive snapshots), returning 1 if they are identical and 0 if they are completely independent. It is typically used to quantify the stability of community structure over time, but it is also used in tests of community detection algorithms [28]. In order to calculate NMI, it was necessary to filter airports that do not appear in all snapshots for a given year. These few, small airports are rarely used and their traffic is very low, so their effect on the network is insignificant.

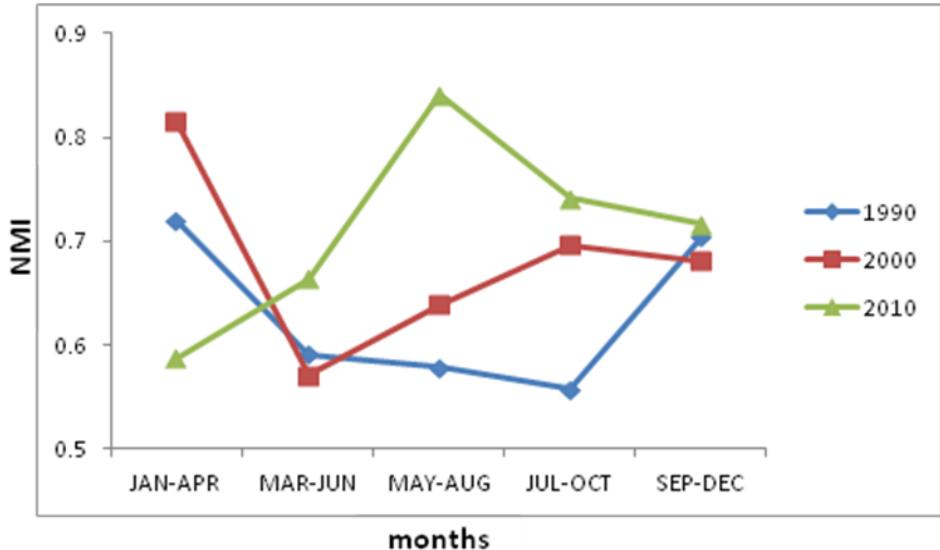


Fig. 5. Normalised Mutual Information (NMI) of consecutive network snapshots.

Fig. 5 presents NMI over time. For example, JAN-APR refers to the stability of the community structure in the period January to April, using the NMI of the partitions for Jan-Feb and Mar-Apr. The connecting lines do not indicate continuity, but are there to facilitate interpretation of the graph. Fig. 5 suggests that in general, the community structure is fairly stable over the course of a year as the NMI is always above 0.5. In addition, annual stability has increased over the three years investigated as the average NMIs for 1990, 2000 and 2010 are 0.63, 0.68 and 0.71, respectively. Specifically, for May-Aug and Jul-Oct the network has shown consistent improvement in stability over its evolution, whereas for Jan-Apr it has become more unstable. The intervals Mar-Jun and Sep-Dec are virtually unchanged over the two decades. In particular, Jan-Apr 2000 and May-Aug 2010 were highly stable ( $NMI > 0.8$ ), while 1990 was a relatively unstable year. The existence of an annual cycle is confirmed and quantified by calculating the NMI of the pair Jan-Feb and Nov-Dec, which is 0.69, 0.79 and 0.52 for the years 1990, 2000 and 2010, respectively. In other words, in terms of community structure, Jan-Feb resembles Nov-Dec (not so much in 2010), indicating the presence of an annual cycle of passenger demand.

#### 4.2. Evolution

This section describes the evolution of the USAN from 1990 to 2010, by focusing on three key issues: volume of air travel, bi-monthly snapshots, and the main hub Atlanta. In addition, the migration levels between, and within, the four US macro-regions are discussed.

#### *4.2.1. Volume of Air Travel*

The quantity of domestic air traffic can be described by the total number of passengers carried across the USAN. Since the surface area of the airport nodes in Figs. 2-4 is proportional to the number of passengers, it is easy to determine the volume of air travel by observing the size of the nodes.

The volume of air travel grew significantly from 1990 to 2000, with a particularly strong concentration of travellers via Atlanta. The first decade of the 21<sup>st</sup> century, however, did not see a significant increase in air travel, which, to a certain extent, may have been caused by key events, such as the September 11 terrorist attacks in 2001, and the start of the global economic recession in 2008. It is interesting that although most airports did not grow much from 2000 to 2010, there are some, such as Denver, that did experience a steady growth in terms of passengers. The specific changes in passenger distribution among airports are highlighted by the changes in the size of circles in Figs. 2-4.

#### *4.2.2. Bi-monthly Snapshots*

In addition to the analysis of seasonal variation, it is also necessary to study long-term evolution, by focusing on individual bi-monthly snapshots and observing the changes in the network from 1990 to 2000, and from 2000 to 2010. Therefore, each of the six bi-monthly periods is analysed separately in order to illustrate the precise changes in passenger flows and community structure for the specified period, that have occurred in each of the two decades.

##### *January-February*

In terms of community structure, 1990 has a mixed pattern of clusters apart from the south (Fig. 2(A)), 2000 has a large cyan super-cluster covering all of the US (Fig. 3(A)), and 2010 again has a mixed structure (Fig. 4(A)). This indicates that in 1990 and 2010 there were numerous popular connections that saw a large number of air passengers, but in 2000 the passengers were more evenly distributed among the possible connections,

resulting in a single super-community. In addition, Atlanta and some airports in the south and north-east had their own specific traffic patterns, as shown in Fig. 3(A).

#### *March-April*

Generally, the community structure for this period is stable, but from 2000 to 2010 there is a clear transition of two hubs – Chicago and Dallas – from the main cluster to their own local-scale clusters (Figs. 3(B) & 4(B)). In other words, these two airports became regional hubs in the first decade of the 21<sup>st</sup> century, at least for the months of March and April.

#### *May-June*

Community structure changes significantly for the period 1990-2010, highlighting the specific changes in passenger trends over the years. In particular, 1990 is composed of one large red cluster covering all but the south, one medium-sized pink cluster in the south, and several regional clusters (Fig. 2(C)). This structure indicates that the red airports are the national long-range hubs, the south is somewhat more isolated, and the rest of the airports provide more local services. On the other hand, in 2000 Chicago and Dallas belong to the same cluster, and Atlanta is by far the top airport in the US (Fig. 3(C)). In 2010, there are two main clusters – the Chicago cluster in yellow, and the Denver cluster in blue – that cover the US together with Atlanta and Dallas, acting as national super-hubs (Fig. 4(C)).

#### *July-August*

Community structure in July-August suggests that in 1990 passengers preferred specific long-range connections (Fig. 2(D)). Most clusters cover large areas of the US, so many people travelled all over the US, specifically among airports of the same colour. On the other hand, in 2000 passengers were more evenly distributed within the green cluster, and more intricately concentrated on certain routes only in the north-east (Fig. 3(D)); while in 2010 the picture is, again, completely different, with two large clusters in red and pink, and two key hubs – Atlanta and Dallas – in blue and green, respectively (Fig. 4(D)).

#### *September-October*

The network in 1990 (Fig. 2(E)) is mainly composed of the blue LA cluster and the green Dallas cluster, with Chicago and Atlanta as hubs, and the usual mix of clusters in

the densely populated north-east. In 2000, however, there is one red super-cluster, Atlanta is the main hub, and there is also a lot of activity in the Chicago region, as illustrated by the many colours that indicate the specific passenger trends in September-October (Fig. 3(E)). 2010 has a mix of multiple large clusters revealing new passenger flows (Fig. 4(E)). This is a sign of long-range travel among community members that are far apart.

#### *November-December*

In 1990 (Fig. 2(F)) the USAN is split into a large blue cluster and a yellow Dallas cluster in the south, but in 2000 (Fig. 3(F)) they have converged to a single yellow super-cluster, covering all but some regions in the north-east and the main hub Atlanta. In 2010 (Fig. 4(F)), the super-cluster has broken down, leaving Dallas as a national hub, and two red and green clusters spanning a large part of the US.

#### *4.2.3. Atlanta*

The role of ATL as a leading US airport depends on factors, such as air services and their locations, as well as investments into growth and development. In 1967, the city of Atlanta and the airlines began to work on a master plan for the future development of the airport. Many investments were made in the following years, leading to new passenger terminals, runways, and facilities both inside (such as the people mover system linking parts of the terminal), and outside (such as the Red/Gold rail line, operated by the Metropolitan Atlanta Rapid Transit Authority, linking the airport to the counties of Fulton and DeKalb, in addition to Atlanta itself).

ATL is also the primary base of many airlines, such as Delta Air Lines, who built one of the world's largest airline bases in 1930. Delta was an early adopter of the hub-and-spoke system, with Atlanta as its primary hub between the Midwest and Florida. This gave it an early competitive advantage, as Florida has been an attractive destination within the US for many decades. Although there is a decrease in the volume of migration in recent years, Florida and the South are still very popular destinations.

In 1990, Atlanta was one of the three leading US airports for domestic flights. By 2000 it became the top airport (Figs. 2-4). Atlanta is also the only significant member in its community for all three years. This implies that it is equally well connected to other airports, thereby possibly serving as a national hub. Since ATL handles so many passengers but there are no other major airports of the same colour, it follows that all

ATL connections have relatively similar traffic loads, with longer connections having less traffic due to the effect of spatial separation. Therefore, ATL has no strong preferential attachment to any other major airport. To verify that Atlanta is a national hub, it is also necessary to check its number of direct connections. Table 1 summarizes ATL's number of connections and the highest number of connections for the months Jan-Feb, in each of the years studied.

Table 1. Atlanta's connections.

	<b>1990</b>	<b>2000</b>	<b>2010</b>
Atlanta	101	142	167
Max	139	142	172

Clearly, Atlanta ranks very high in terms of connections, so it has a direct influence on a large part of the US territory. For example, in Jan-Feb 2010, ATL handled 10.7 million passengers (top in the US) on 167 connections, with an average of 64,000 passengers per connection, compared with the US highest figures of 172 and 73,000, respectively. In summary, ATL became the top US hub for domestic flights by the year 2000.

#### *4.2.4. Migration*

According to recent figures and US Census data [39], American people move many times during their adult lives, mainly in their twenties. Preferred destinations of domestic migration were Southern states, mainly Florida, possibly because they are considered attractive places to live and work. Although US domestic migration has fallen noticeably since the 1980s, it is still higher than that within most other developed countries and during the period 2000-2004 it continued to redistribute the country's population [36]. Nevertheless, the current slowdown in domestic migration due to the impact of the economic situation has changed the picture of movements within the US. In-migration towards states like Arizona, Florida and Nevada has slowed down, while Massachusetts, New York and California now have considerably less out-migration [26, 39]. In the years 2009 and 2010 mobility among states slowed nationwide and only a small percentage difference was observed during the two-year period (Table 2).

Table 2. In-migration, representing the number of people migrating to specific US states in 2009-2010 [39].

<b>State</b>	<b>Year 2010</b>	<b>Year 2009</b>	<b>Diff %</b>	<b>State</b>	<b>Year 2010</b>	<b>Year 2009</b>	<b>Diff %</b>
Alabama	108,951	124,658	-0.14	Montana	35,641	31,015	0.13
Alaska	36,345	40,474	-0.11	Nebraska	51,290	53,214	-0.04
Arizona	223,324	226,457	-0.01	Nevada	103,179	109,257	-0.06
Arkansas	79,214	85,857	-0.08	New Hampshire	39,423	37,940	0.04
California	445,972	460,161	-0.03	New Jersey	130,101	136,212	-0.05
Colorado	187,240	182,854	0.02	New Mexico	74,237	64,797	0.13
Connecticut	79,360	81,546	-0.03	New York	276,167	277,482	0.00
Delaware	31,713	35,085	-0.11	North Carolina	265,206	284,171	-0.07
District of Columbia	51,244	38,907	0.24	North Dakota	30,100	29,970	0.00
Florida	495,857	475,871	0.04	Ohio	174,773	171,894	0.02
Georgia	250,469	280,221	-0.12	Oklahoma	106,720	117,850	-0.10
Hawaii	53,581	53,270	0.01	Oregon	117,521	127,489	-0.08
Idaho	55,871	57,790	-0.03	Pennsylvania	241,855	232,316	0.04
Illinois	206,014	206,151	0.00	Rhode Island	32,335	32,108	0.01
Indiana	127,925	132,755	-0.04	South Carolina	152,710	33,616	0.78
Iowa	72,706	74,704	-0.03	South Dakota	25,777	145,873	-4.66
Kansas	95,127	102,695	-0.08	Tennessee	159,778	29,632	0.81
Kentucky	118,622	122,184	-0.03	Texas	490,738	168,174	0.66
Louisiana	98,291	90,957	0.07	Utah	78,163	511,166	-5.54
Maine	27,962	24,672	0.12	Vermont	22,529	90,375	-3.01
Maryland	165,096	174,958	-0.06	Virginia	260,813	19,390	0.93
Massachusetts	143,247	148,500	-0.04	Washington	191,784	271,600	-0.42
Michigan	117,581	118,054	0.00	West Virginia	39,791	192,654	-3.84
Minnesota	89,911	90,944	-0.01	Wisconsin	93,586	50,155	0.46
Mississippi	73,135	67,245	0.08	Wyoming	28,046	95,475	-2.40
Missouri	146,093	150,271	-0.03	Puerto Rico	31,732	30,889	0.03

Despite the current tendency to stagnancy, the role of the airport network in the context of US domestic migration is important. Since an airport network is continuously evolving depending on passenger demand, it is increasingly well-optimised for a number of functions, such as carrying more passengers, minimising flight changes for the average passenger, and making profit. As the USAN has evolved to attract passengers that are typically travelling to popular destinations, it is directly facilitating migration. Although most passengers fly for short-term business or leisure, there is evidence that a significant fraction of passengers are in fact migrating with a migration probability inversely proportional to the distance [30, 38].

According to Census data, Figs. 6-8 show the migration patterns for the years 1990, 2000 and 2010. Data refer to people that are moving to a given macro-region or within it. The scale is relative to the maximum value and therefore not consistent across the three years, but they are comparable, in order to identify any potential variations in migration patterns over the two decades. Migration within the macro-regions is higher than that among them (decay of interaction as distance increases), and migration within the South is the highest, suggesting strong dynamics among the member States. Furthermore, the South region attracts the most people from outside for all three years. This is in line with the fact that Atlanta airport (located in the South) has the highest passenger flow, as discussed above, but it does not necessarily follow that the entire flow is related to the South, as many of the passengers will change flights in Atlanta en route to other regions. Nevertheless, the migration patterns do have a clear overlap with the community structure discovered in the USAN.

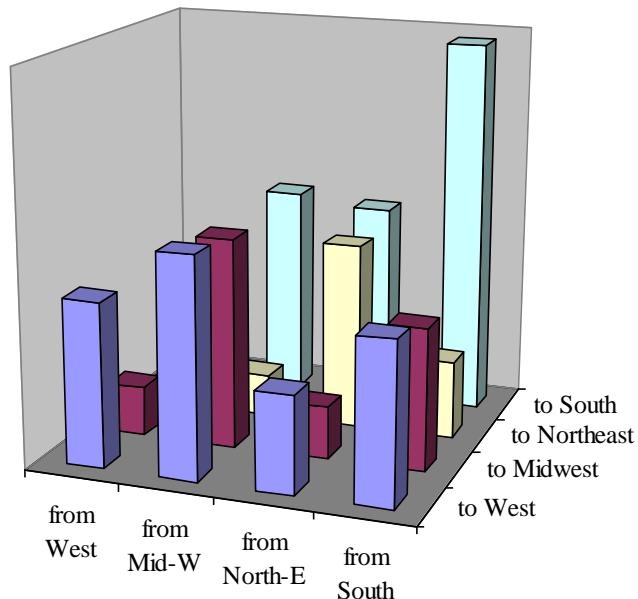


Fig. 6. 1990 migration patterns among the four macro-regions: West, Midwest, Northeast and South.

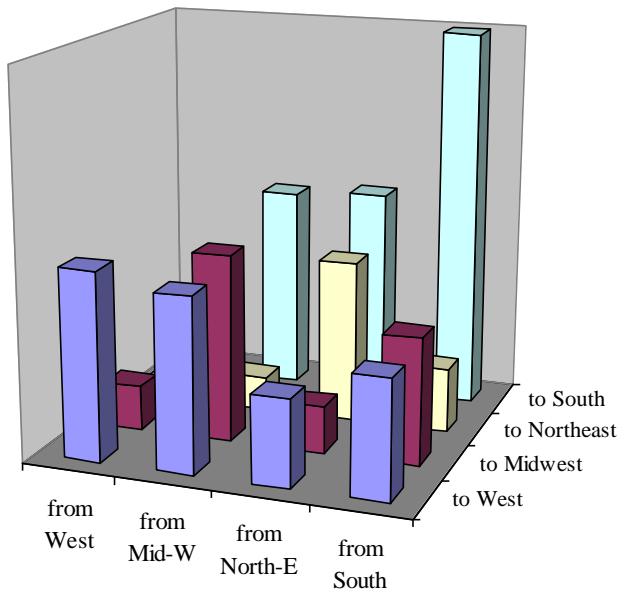


Fig. 7. 2000 migration patterns among the four macro-regions: West, Midwest, Northeast and South.

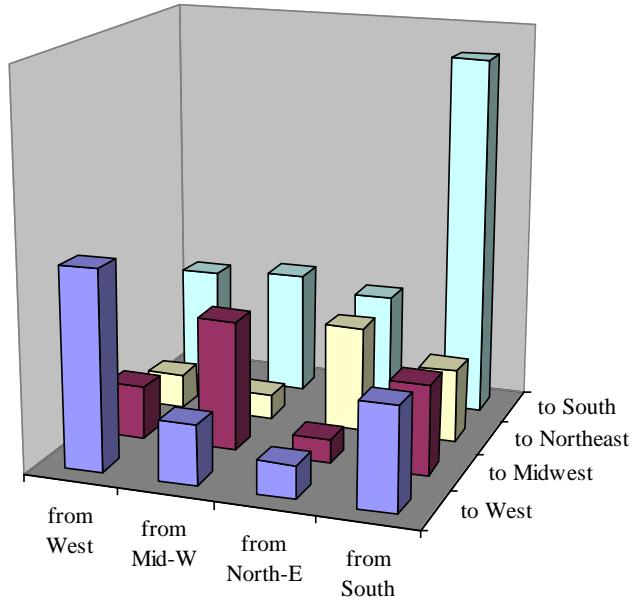


Fig. 8. 2010 migration patterns among the four macro-regions: West, Midwest, Northeast and South.

## 5. Conclusion

The US Airport Network is a complex system that is continuously evolving to meet the growing demands for air travel. Investigating the community structure within has illuminated important hidden characteristics of the network's topology and dynamics. Specifically, the findings reveal high heterogeneity in both space and time. In other words, the network is non-uniform (in space) and non-linear (in time) in terms of its connections and traffic. In addition, spatial community detection has identified a more realistic picture of the intricate structure within the network, which is invaluable for understanding this critical transportation system. Furthermore, this network model may be used for forecasting future trends in the US Airport Network. For example, the identification of reliable communities can be the first step to study how external factors, such as natural disasters (e.g. tornados, which are common in large parts of the US), affect the function of the network. Moreover, the communities emerging from socio-economic interactions, as in the case of migration, reflect both the social influence radii and the activity system configuration (the distribution of activities in terms of location). Again, variations in the activity system will possibly modify such relationships and the resulting community structure. Finally, there is a clear relationship between domestic US air travel and migration. In particular, the identified community structures map well onto the migration patterns among the four macro-regions and within the region.

## References

- [1] Alderson, D. and Willinger, W., A contrasting look at self-organization in the Internet and next-generation communication networks, *IEEE Communications Magazine* (2005), 94-100.
- [2] Almendral, J.A., Leyva, I., Li, D., Sendiña-Nadal, I., Havlin, S. and Boccaletti, S., Dynamics of overlapping structures in modular networks, *Physical Review E* **82** (2010) 016115.
- [3] Amaral, L.A.N., Scala, A., Barthélémy, M. and Stanley, H.E., Classes of small-world networks, *Proc. Natl. Acad. Sci. USA* **97** (2000) 11149–11152.
- [4] Ball, B., Karrer, B. and Newman, M.E.J., Efficient and principled method for detecting communities in networks, *Physical Review E* **84** (2011) 036103.
- [5] Barrat, A., Barthélémy, M., Pastor-Satorras, R. and Vespignani, A., The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101** (2004) 3747-3752.
- [6] Barthélémy, M., Spatial networks, *Physics Reports* **499** (2011) 1-101.

- [7] Blondel, V., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E., Fast unfolding of communities in large networks, *J. Stat. Mech.* **10** (2008) P10008.
- [8] Bounova, G., *Topological Evolution of Networks: Case Studies in the US Airlines and Language Wikipedias*, PhD Thesis (Massachusetts: MIT Press, 2009).
- [9] Bulu, M., *City Competitiveness and Improving Urban Subsystems: Technologies and Applications* (US: IGI Global, 2011).
- [10] Bureau of Transportation Statistics, <http://www.bts.gov/>
- [11] Burghouwt, G. and De Wit, J., Temporal configurations of European airline networks, *Journal of Air Transport Management* **11** (2005) 185–198.
- [12] Button, K.J., Economic aspects of regional airport development. In Postorino, M.N. (Ed.), *Developments of Regional Airports* (Southampton, UK: WIT Press, 2010).
- [13] Calabrese, F., Dahlem, D., Gerber, A., Paul, D., Chen, X., Rowland, J., Rath, C. and Ratti, C., The connected states of America: quantifying social radii of influence, *Proceedings of the IEEE International Conference on Social Computing* (2011).
- [14] Camagni, R., On the concept of territorial competitiveness: sound or misleading?, *Urban Studies* **39** (2002) 2395–2411.
- [15] Carreras, B.A., Lynch, V.E., Dobson, I. and Newman, D.E., Critical points and transitions in an electric power transmission model for cascading failure blackouts, *Chaos* **12** (2002) 985–994.
- [16] Cervero, R., Efficient urbanisation: economic performance and the shape of the metropolis, *Urban Studies* **38** (2001) 1651–1671.
- [17] Chorianopoulos, I., Pagonis, T., Koukoulas, S. and Drymoniti, S., Planning, competitiveness and sprawl in the Mediterranean city: the case of Athens, *Cities* **27** (2010) 249–259.
- [18] Danon, L., Díaz-Guilera, A., Duch, J. and Arenas, A., Comparing community structure identification, *J. Stat. Mech.* (2005) P09008.
- [19] De Montis, A., Barthélémy, M., Chessa, A. and Vespignani, A., The structure of interurban traffic: a weighted network analysis, *Environment and Planning B: Planning and Design* **34** (2007) 905 – 924.
- [20] Estrada, E. and Hatano, N., Communicability graph and community structures in complex networks, *Applied Mathematics and Computation* **214** (2009) 500–511.
- [21] Expert, P., Evans, T., Blondel, V. and Lambiotte, R., Uncovering space-independent communities in spatial networks, *Proc. Natl. Acad. Sci. USA* **108** (2011) 7663–7668.

- [22] Gegov, E., Gegov, A., Atherton, M. and Gobet, F., Evolution-based modelling of complex airport networks, *Proceedings of COSY 2011: International Conference on Complex Systems* (2011).
- [23] Good, B.H., de Montjoye, Y.-A. and Clauset, A., Performance of modularity maximization in practical contexts, *Physical Review E* **81** (2010) 046106.
- [24] Guimerà, R., Mossa, S., Turtschi, A. and Amaral, L.A.N., The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles, *Proc. Natl. Acad. Sci. USA* **102** (2005) 7794-7799.
- [25] Hsu, C.-I. and Wen, Y.-H., Determining flight frequencies on an airline network with demand-supply interactions, *Transportation Research Part E: Logistics and Transportation Review* **39** (2003) 417–441.
- [26] Internal Revenue Service (IRS) migration data 2010, <http://www.irs.gov/taxstats/>
- [27] Jutla, I.S. and Mucha, P.J., A generalized Louvain method for community detection implemented in MATLAB, <http://netwiki.amath.unc.edu/GenLouvain> (2011).
- [28] Lancichinetti, A. and Fortunato, S., Community detection algorithms: a comparative analysis, *Physical Review E* **80** (2009) 056117.
- [29] Lancichinetti, A., Radicchi, F., Ramasco, J.J. and Fortunato, S., Finding statistically significant communities in networks, *PLOS ONE* **6** (2011) e18961.
- [30] Levy, M., Scale-free human migration and the geography of social networks, *Physica A* **389** (2010) 4913-4917.
- [31] Mackun, P., Wilson, S. and with Fischetti, T., Goworowska, J., *2010 Census Brief* (2011).
- [32] Meila, M., Comparing clusterings by the variation of information. In Schölkopf, B. and Warmuth, M. (Eds.), *Learning Theory and Kernel Machines, Lecture Notes in Computer Science* (New York: Springer, 2003).
- [33] Meunier, D., Lambiotte, R. and Bullmore, E., Modular and hierarchically modular organization of brain networks, *Frontiers in Neuroscience* **4** (2010) 1-11.
- [34] Newman, M.E.J. and Girvan, M., Finding and evaluating community structure in networks, *Physical Review E* **69** (2004) 026113.
- [35] Patuelli, R., Reggiani, A., Gorman, S.P., Nijkamp P. and Bade, F.-J., Network analysis of commuting flows: a comparative static approach to German data, *Netw. Spat. Econ.* **7** (2007) 315–331.
- [36] Perry, M.J., Domestic Net Migration in the United States: 2000 to 2004, *U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau* (2006).

- [37] Rouwendal, J., Search theory and commuting behaviour, *Growth Change* **35** (2004) 391–418.
- [38] Schwartz, A., Interpreting the effect of distance on migration, *J. Polit. Eco.* **81** (1973) 1153–1169.
- [39] US census data 2011, <http://www.census.gov/>
- [40] Wuellner, D.R., Roy, S. and D'Souza, R.M., Resilience and rewiring of the passenger airline networks in the United States, *Physical Review E* **82** (2010) 056101.