SPACE-INDEPENDENT COMMUNITY STRUCTURE DETECTION IN UNITED STATES AIR TRANSPORTATION

Emil Gegov*. Alexander Gegov**. M. Nadia Postorino***
Mark Atherton*. Fernand Gobet****

*School of Engineering and Design, Brunel University, Uxbridge UB8 3PH, UK
(e-mail: emil.gegov@brunel.ac.uk, mark.atherton@brunel.ac.uk)
**School of Computing, University of Portsmouth, Portsmouth PO1 3HE
(e-mail: alexander.gegov@port.ac.uk)
*** Department of Informatics, Mathematics, Electronics and Transport, University of Reggio Calabria,
Reggio Calabria 89122, Italy (e-mail: npostorino@unirc.it)
****School of Social Sciences, Brunel University, Uxbridge UB8 3PH, UK
(e-mail: fernand.gobet@brunel.ac.uk)

Abstract: This article presents an evolution-based model for the US airport network. 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, reveals a better picture of the communities within.

Keywords: air transportation; community structure; United States Airport Network.

1. INTRODUCTION

Transportation networks are a good example of spatial networks. In this case, network topology is entangled with spatial aspects such as the location of nodes and the length of links. Such networks are also characterised 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, 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 number of transported passengers. 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 (Barrat et al. 2004).

In recent years, the analysis of complex transport networks has received considerable attention, mainly in terms of commuting networks (De Montis et al. 2007, Patuelli et al. 2007). Airport networks have also been studied to characterise their level of degree correlations and clustering (community structure), their evolution in time, and their potential scale-free properties (Guimerà et al. 2005, Gegov et al. 2011). Community structure is when nodes can be

grouped into sets of nodes such that each set is more densely connected inside than with the rest of the network. Community structure is later explained in more detail. The emergence of community structures in airport networks has implications for network efficiency and its socio-economic characteristics. In terms of network efficiency, 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 specific individual traveller's needs, and the location and distribution of relevant activities.

2. US DOMESTIC AIRPORT NETWORK

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 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, a time coordination of flights at the hub - according to a "flight wave" concept (Burghouwt, De Wit 2005), and the integration of air services at the hub (e.g. baggage transfer). This kind of air service increases both the number of served destinations and the load factor (fraction of filled seats) for each aircraft. In fact, point-to-point services that guarantee the same number of destinations from each origin could have too low demand levels to assure

profitability to the airline. The main disadvantage 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. 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 (Bounova 2009). Consequently, the USAN transformed from a hub-and-spoke to a small-world architecture, with high clustering and low characteristic path length (average number of air trips one needs to take to get from A to B, considering all pairs of airports).

3. METHOD

This study investigates 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. In this way it is possible to study, over time, both dynamics *on* the network in terms of traffic fluctuations and dynamics *of* the network in terms of topology fluctuations. Fig. 1 shows a map of the US regions and states, including the locations of the main airports in terms of passenger flows.
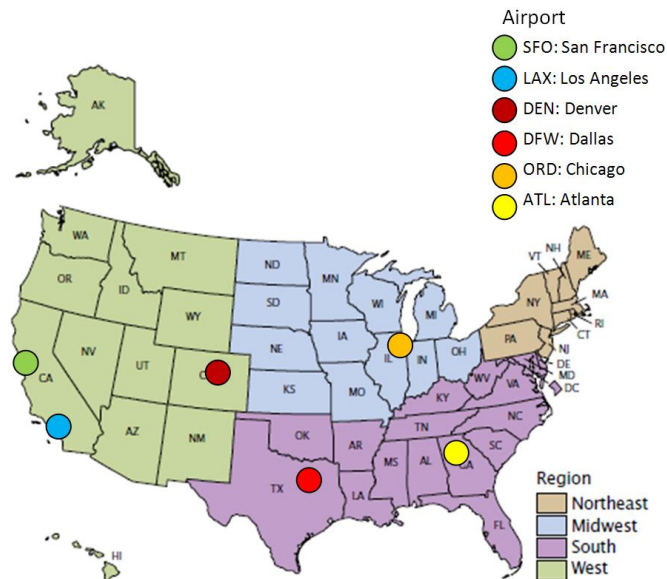

Fig. 1. US macro-regions and major airports in 2010. Adapted from (Mackun et al. 2011).

4. COMMUNITY STRUCTURE

Community structure is a prominent feature in many biological, social and technological complex systems (Meunier et al. 2010, Blondel et al. 2008). It 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.

In recent years, research on complex networks has proposed many community detection methods (Lancichinetti, Fortunato 2009) that aim to discover the most sensible partition of a network into communities. Most of them work on the principle of modularity (Newman, Girvan 2004) optimisation, aiming to maximise the modularity benefit function describing the quality of a network partition into communities. 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 in (1) (Expert et al. 2011). In essence, modularity measures how sharply the modules are defined.

Q = (fraction of links within communities)

$\quad$ − (expected fraction of such links) $\qquad$ (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 network properties and organisation. 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, Girvan 2004) preserves the strength (total weight on all adjacent links) of nodes in the random networks:

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

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. Therefore, 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 (Ball et al. 2011, Estrada, Hatano 2009). To address this, (Expert et al. 2011) 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:

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

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, similarly to gravity, as distance between objects increases.

Expert's null model is ideal for uncovering space-independent community structure, and hence, it is the one applied to the USAN network model. The necessary inputs are the adjacency matrix (encoding topology and passenger flows), the distance matrix (containing the Euclidean distance between all pairs of airports), the importance vector (holding the passenger flow at each airport), and the bin size, which is used to bin the data from the distance matrix. For application to our model, several bin sizes were tested and, after comparing their effect on the deterrence function of the algorithm, a bin size of one was chosen due to distance being expressed in terms of degrees of arc length, where one degree is approximately 60 miles.

5. RESULTS

Expert's spatial community detection method, described earlier, is applied to each of the 18 USANs (each one representing the topology and passengers for a bi-monthly period). The output is a vector, assigning each airport to a specific community of airports, in which all members have particularly strong interactions in terms of passenger flows between them, given their physical separation. Figs. 2-19 represent the USAN at various stages over time, where each airport is denoted by a circle, the size of which 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. 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.

5.1. Year 1990

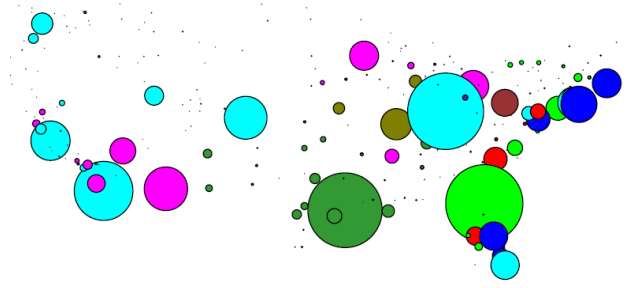Figs. 2-7 depict bi-monthly snapshots of the USAN for the year 1990.
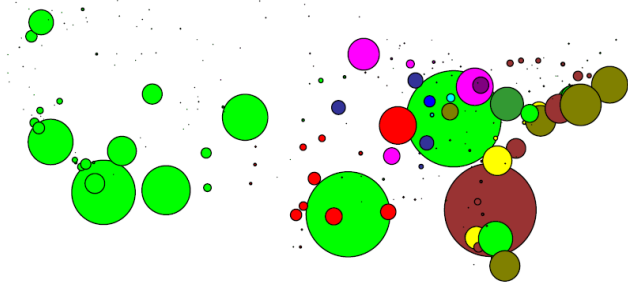

Fig. 2. JAN-FEB 1990
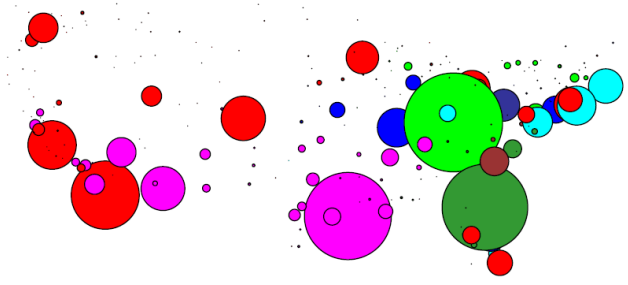

Fig. 3. MAR-APR 1990

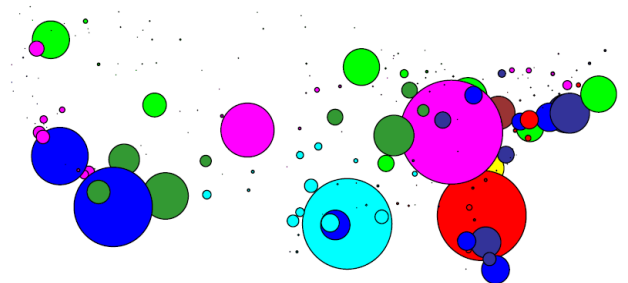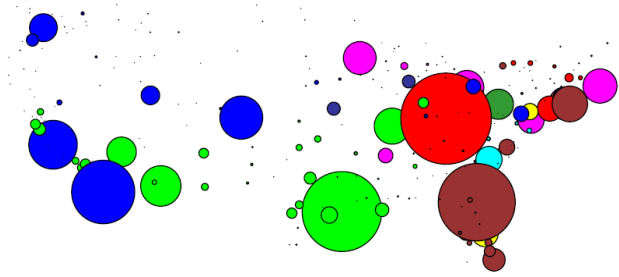

Fig. 4. MAY-JUN 1990

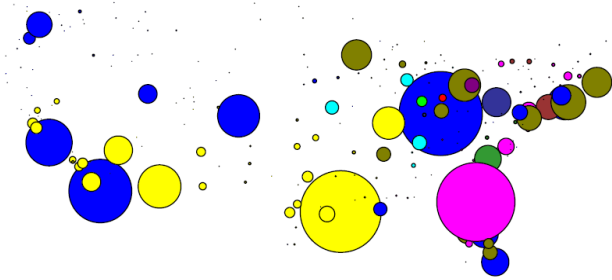Fig. 5. JUL-AUG 1990

Fig. 6. SEP-OCT 1990

Fig. 7. NOV-DEC 1990

In Jan-Feb (Fig. 2) there is a well-defined cyan community of west-coast airports, such as Los Angeles (LA) and San Francisco, together with Chicago, suggesting high passenger mobility between those locations. In Fig. 3 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. 4 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. 5) 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. 6) sees an overall decline in air travel, matching the end of the tourist season, and two large communities in blue and green. In Fig. 7 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 the blue regions.

5.2. Year 2000

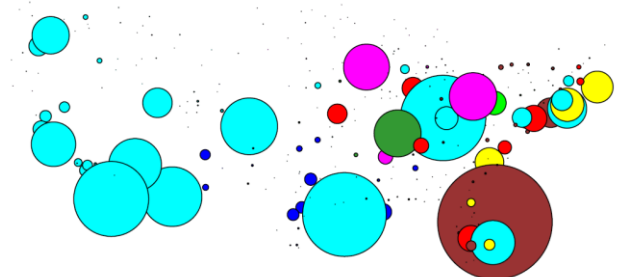Figs. 8-13 depict bi-monthly snapshots of the USAN for the year 2000.
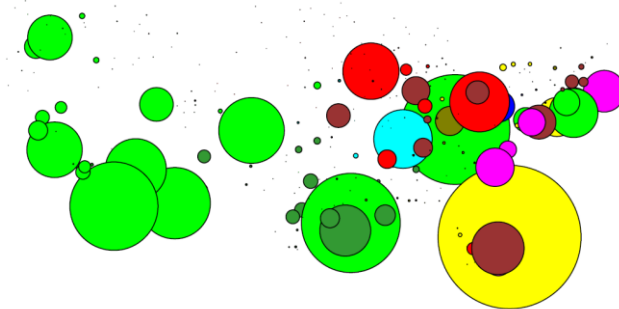

Fig. 8. JAN-FEB 2000
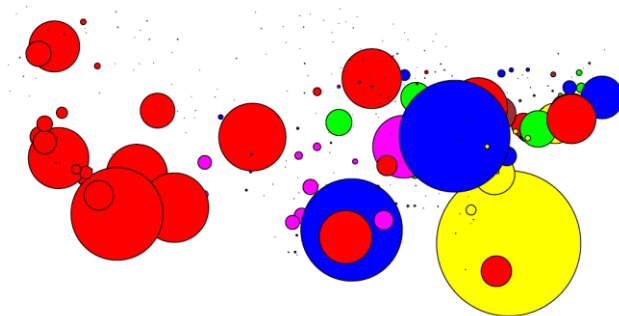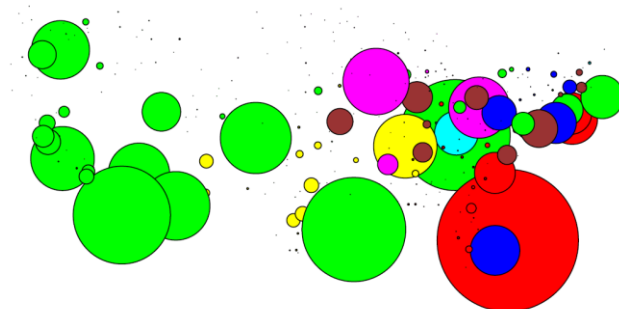

Fig. 9. MAR-APR 2000

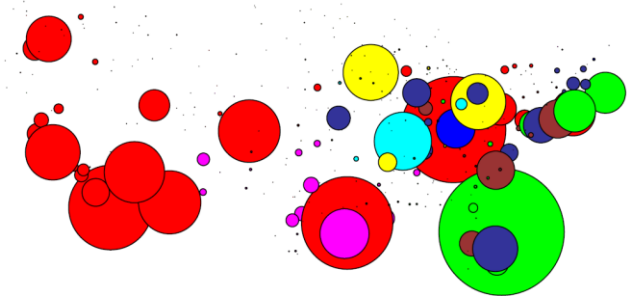Fig. 10. MAY-JUN 2000

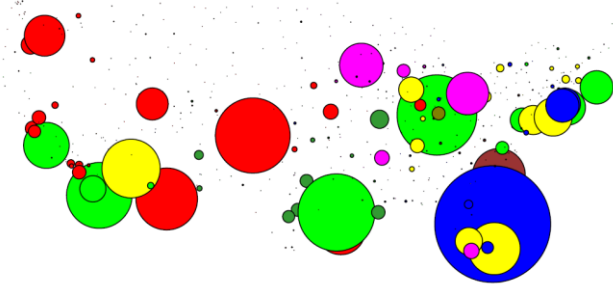Fig. 11. JUL-AUG 2000

Fig. 12. SEP-OCT 2000
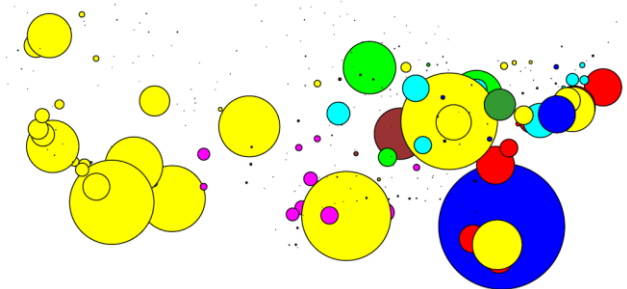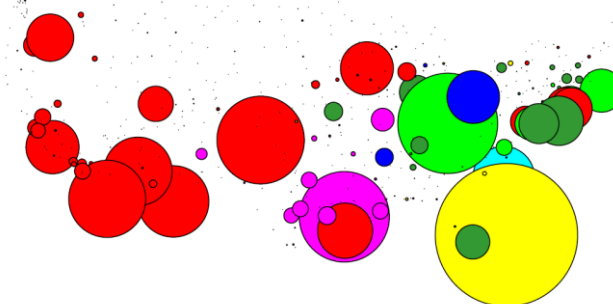

Fig. 14. JAN-FEB 2010

Fig. 13. NOV-DEC 2000

Fig. 15. MAR-APR 2010

Jan-Feb in Fig. 8 displays a prevailing cyan community of most major airports dominating the west and a large part of the rest of the US. In Fig. 9, 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. 10 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 northeast. 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. 11) 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. 12 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. 13 is similar to the previous network for Sep-Oct.

5.3. Year 2010

Figs. 14-19 depict bi-monthly snapshots of the USAN for the year 2010.
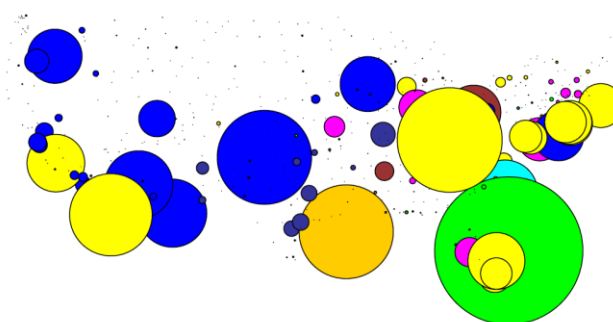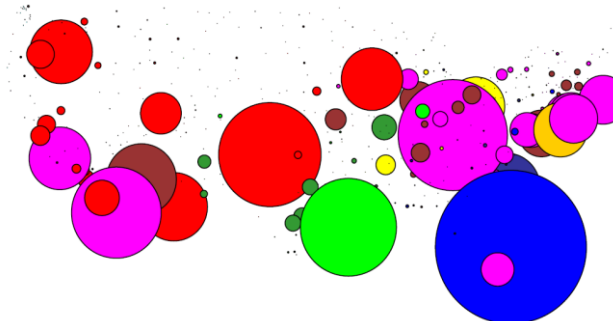

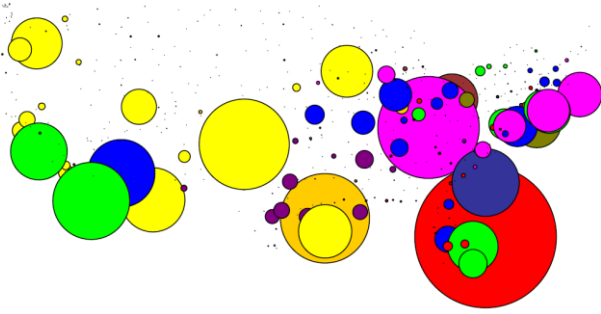Fig. 16. MAY-JUN 2010

Fig. 17. JUL-AUG 2010
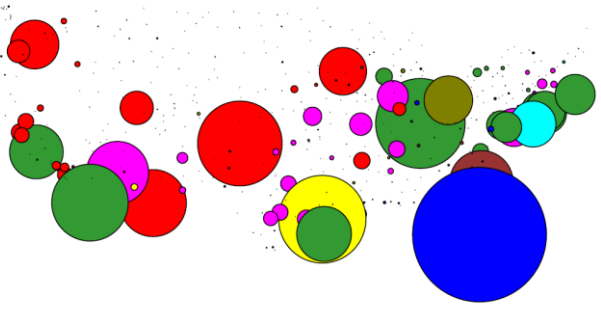
Fig. 18. SEP-OCT 2010

Fig. 19. NOV-DEC 2010

Fig. 14 Jan-Feb has two large clusters in red and green that cover the west and a big 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. 15 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. 16 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. Jul-Aug (Fig. 17) 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, possibly due to high business and leisure long-distance travel in the summer. Sep-Oct (Fig. 18) 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. 19) 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).

6. CONCLUSIONS

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, the spatial community detection method has identified a more realistic picture of the intricate structure within the network, which is invaluable for our understanding of this critical transportation system. Furthermore, the proposed network model may be used for urban and transport planning, and for forecasting future trends in the US Airport Network.

REFERENCES

Ball, B., Karrer, B. and Newman, M.E.J. (2011). Efficient and principled method for detecting communities in networks. *Physical Review E*, 84, 036103.

Barrat, A., Barthèlemy, M., Pastor-Satorras, R. and Vespignani, A. (2004). The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA,* 101, 3747-3752.

Blondel, V., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.*, 10, P10008.

Bounova, G. (2009). *Topological Evolution of Networks: Case Studies in the US Airlines and Language Wikipedias*, PhD Thesis. MIT Press, Massachusetts.

Burghouwt, G. and De Wit, J. (2005). Temporal configurations of European airline networks. *Journal of Air Transport Management*, 11, 185-198.

De Montis, A., Barthélemy, M., Chessa, A. and Vespignani, A. (2007). The structure of interurban traffic: a weighted network analysis. *Environment and Planning B: Planning and Design*, 34, 905 – 924.

Estrada, E. and Hatano, N. (2009). Communicability graph and community structures in complex networks. *Applied Mathematics and Computation*, 214, 500–511.

Expert, P., Evans, T., Blondel, V. and Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proc. Natl. Acad. Sci. USA*, 108, 7663–7668.

Gegov, E., Gegov, A., Atherton, M. and Gobet, F. (2011). Evolution-based modelling of complex airport networks. *Proceedings of COSY 2011: International Conference on Complex Systems*.

Guimerà, R., Mossa, S., Turtschi, A. and Amaral, L.A.N. (2005). The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA*, 102, 7794-7799.

Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical Review E*, 80, 056117.

Mackun, P., Wilson, S. and with Fischetti, T., Goworowska, J. (2011). *2010 Census Brief*.

Meunier, D., Lambiotte, R. and Bullmore, E. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience,* 4, 1-11.

Newman, M.E.J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69, 026113.

Patuelli, R., Reggiani, A., Gorman, S.P., Nijkamp P. and Bade, F.-J. (2007). Network analysis of commuting flows: a comparative static approach to German data. *Netw. Spat. Econ*, 7, 315–331.