Evolution-Based Modelling of Complex Airport Networks

Emil A. Gegov*. Alexander E. Gegov**
Mark A. Atherton*. Fernand R. Gobet******

**School of Engineering and Design, Brunel University, Uxbridge, Middlesex UB8 3PH, UK (Tel: 0044-18952-65921; e-mail: emil.gegov@brunel.ac.uk)*
*** School of Computing, University of Portsmouth, Lion Terrace, Portsmouth PO1 3HE, UK (e-mail: alexander.gegov@port.ac.uk)*
**** School of Engineering and Design, Brunel University, Uxbridge, Middlesex UB8 3PH, UK (e-mail: mark.atherton@brunel.ac.uk)*
***** School of Social Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK (e-mail: fernand.gobet@brunel.ac.uk)*

Abstract: This paper presents a time-series model of the United States Airport Network as a directed, weighted network, with the weight representing the total number of passengers flying from an origin to a destination airport, in a two month time period. Six independent networks are built for a given year, in order to capture the seasonal variation of passengers. To explore the evolution of the network over the past two decades, three specific years are investigated: 1990, 2000, and 2010. The results highlight the growth of the network in terms of airports and connections, and suggest a scale-free, small-world topology. In addition, the ranked passenger distribution appears to follow a logarithmic trend, implying high heterogeneity in passengers on different connections.

Keywords: network modelling; United States Airport Network; topological evolution; ranked passenger distribution.

1. INTRODUCTION

In recent years, the availability of huge data sets has enabled researchers across many disciplines to model and to understand exceedingly complex systems, by using network modelling and analysis. For example, biological networks, such as metabolic (Morine et al. 2009), and gene co-expression (Carter et al. 2004); technological networks, such as the Internet (Alderson, Willinger 2005), and the power grid (Carreras et al. 2002); and social networks, such as friendship (Girvan, Newman 2002), and co-authorship (Barthelemy et al. 2005), have been widely studied, and interesting patterns have emerged. This research has shown that network modelling provides a powerful abstraction of networked complex systems in the real-world, that is able to strip away the detail of individual systems, while retaining the core information, such as network structure (topology), and dynamics (the function of the network). Hence, it is possible to model the evolution of complex systems at a high level, and to identify common properties, as well as trends, over time. This leads to a better understanding of complex systems, with potential benefits to many areas, such as medicine, technology, and the social sciences, to name a few.

1.1 Airport Networks

The air transportation network of a country or region is a critical component of its infrastructure, with huge impacts on its economy, the transportation of people, cargo, and mail, as well as the potential for propagating negative effects, such as globally spreading diseases (Guimera et al. 2005). Therefore, researchers from multiple disciplines have recently shown a lot of interest in this field, and with an abundance of available data, have made attempts to model and to analyse airport networks. This provides an understanding of how these networks operate; the critical airport nodes that connect otherwise distant locations; whether there are any naturally occurring community structures; and how the networks evolve over time.

1.2 Evolution

Depending on several key factors, such as geographical area, population, economic growth, tourism, and trade, the national airport network of a country may grow and change its topology considerably over time, driven mainly by the airlines, seeking to increase their short-term profits. This means that an airport network is constantly developing, or more precisely, evolving in response to the growing demands of the people using the network either directly as passengers, or indirectly as consumers of transported goods. Globalisation, and the introduction of more long-distance direct connections between far-apart regions does however present a serious threat to public health, as a small outbreak of a disease in a remote region may quickly turn into a global epidemic.

1.3 Related Work

Researchers working on airport networks have typically focused on the modelling of a national airport network, such as the Airport Network of China (Li, Cai 2004), and the Airport Network of India (Bagler 2008); or the World Airport Network (Guimera et al. 2005), which is the global network of all airports. However, most studies so far have either investigated the evolution of the network over a not significantly long time period (Amaral et al. 2000, Barrat et al. 2004, Xu, Harriss 2008), or have not modelled in detail by ignoring link directionality and link weights (Bounova 2009).

1.4 Contribution

This paper presents a more detailed model of the evolution of a complex airport network over a significant time period. The aim is to explore the development of the network, in order to expose growth patterns, and changes in structure as well as passenger demand. In addition, the ranked weight distribution of the links (Gegov et al. 2011), instead of the commonly used (cumulative) probability distribution of link weights, is used as a measure of the volume of passengers travelling between all connected airports. It was chosen because it contains information about the absolute numbers of passengers flying between airports, and every connection, or link, is explicitly present in the distribution.

2. METHODOLOGY

Evolution-based modelling of any complex network can be defined as a process that takes as input some specific network data, and returns a complete network model of these data. In other words, all local interactions between pairs of nodes for some time period are mapped onto a global network model, representing the structure and dynamics of the real complex network, for the period under study. In this way, it is possible to determine how the network is evolving over time, in terms of its topology and interactions.

2.1 Selecting Data

First, it is necessary to decide which specific interactions in the network are of particular interest. For example, in an airport network, these can be the number of passengers flying between airports, the number of aircraft flying between airports, or quite possibly, any other metric describing the link between a pair of airports. Then, a long enough time interval is chosen, such that there are available data to be modelled, and the scale of the observed evolution is maximised. The chosen interval is partitioned into equal time slices, depending on the required level of granularity. In the case where a long interval and high granularity result in an unfeasible number of time slices, a sample of those can be selected for the actual modelling.

2.2 Network Modelling

A network is essentially a set of nodes and links, so if the data are in the form of node pairs (in most cases it is), it is easy to build a network directly from the data: for each pair, insert a directed link from the source node to the target node, labelling the link with the given weight (representing strength of interaction). Hence, a snapshot of the evolving network is generated for each time slice of the data.

3. UNITED STATES AIRPORT NETWORK

Here, a case study of a continuously developing air transportation network that is vital for the mobility of millions of passengers per day is presented. The United States Airport Network (USAN) was chosen for several reasons. Firstly, it is large and growing, so it is clearly a good candidate for studying network evolution. Secondly, there is a lack of detailed models that trace the network for more than a few years. Thirdly, there is a large quantity of available data, dating back to 1990, when the network looked very different to what it is today. Essentially, this is an application of the evolution-based modelling methodology from the previous section, with an additional part describing the network analysis.

3.1 Data Sets

The number of passengers flying from an origin to a destination airport was chosen as the variable for this study, because it is the common choice in the literature, and it is perhaps the most influential factor in the expansion and organisation of the network. The longest possible time period – from 1990 to 2010 – was selected, based on the availability of data for this period. To investigate seasonal variation within a given year and to build more precise models of the network, time slices of length two months offer a good balance, so a year is divided in six equal parts. To reduce the huge amount of modelling (120 networks), without losing too much information, only three years are modelled in this study: 1990, 2000, and 2010. These years capture the oldest, the intermediate, and the newest, open source states of the network. Unfortunately, the data for the end of 2010 are still unreleased at the time of writing, so the network snapshot for November – December is not included in this study. All the data is obtained from the Bureau of Transportation Statistics (http://www.bts.gov/), and is publicly available.

3.2 Network Model

The model consists of seventeen network snapshots: six for 1990, six for 2000, and five for 2010. Each network is directed and weighted, and includes a number of isolated nodes and self-loops. The directed links reflect the difference in passengers flying from A to B and vice versa. The link weight represents the total passengers travelling from A to B in a time slice (January – February, March – April, etc.). Isolated nodes denote airports that handled aeroplane departures and/or arrivals, but no actual passengers. Self-loops occur when an aeroplane takes off and lands at the same airport for some reason, such as an emergency. Over the past twenty years, the USAN experiences dramatic growth: airports triple from about 350 to over 1,100, and direct connections double from 5,000 to 10,000.

3.3 Network Analysis

Graph theory offers numerous statistical parameters that usually measure some structural property of the underlying network, so for the purposes of this study, the most prominent parameters are selected for analysis. Since they are quite general, they are often used across many disciplines that exploit the potential of network modelling. In this paper, we investigate six individual parameters (defined further below): number of nodes (*N*); number of links (*E*); size of Giant Connected Component (*GCC*); average degree ($<k>$); characteristic path length (*L*); and clustering coefficient (*C*). In addition, we compute three functions: the in-degree distribution $P(k_{\text{in}})$; the out-degree distribution $P(k_{\text{out}})$; and the ranked weight distribution $W(r)$. Note that $W(r)$ is an indicator of network dynamics, as opposed to network structure.

In the USAN model, *N* is the total number of US airports; *E* is the total number of one-way domestic connections; *GCC* is the number of airports in the largest connected subnetwork; $<k>$ is the average number of domestic connections per airport; *L* is the average number of flights that need to be taken to get from A to B; and *C* is the expected proportion of airport neighbours (all connected to an airport) that are connected themselves. The latter two of those are calculated for an undirected network due to computational complexity, but most connections are bidirectional anyway, so the results should be fairly accurate. $P(k_{\text{in}})$ and $P(k_{\text{out}})$ are the probability distributions of a randomly chosen airport having k_{in} incoming and k_{out} outgoing connections, respectively. By extracting the first two data points (0 and 1 connection) and taking them as separate parameters *p* and *q*, the degree distributions are well-approximated by a power-law fitting function of the form $P(k) = ak^n$, where *a* is the scaling factor, *k* is in/out-degree, and *n* is the exponent. $W(r)$ is the rank-ordered passenger distribution on all network connections. For systematic analysis across all networks, $W(r)$ is normalised to be in the range (0, 1]. This function is well-approximated by a logarithmic fit of the form $W(r) = bLn(r) + c$, where *b* is the scaling factor, *Ln* is the natural logarithm, *r* is the rank, and *c* is the coefficient. Hence, the functions are described by their parameters: p_{in}, q_{in}, a_{in}, and n_{in} of $P(k_{\text{in}})$; p_{out}, q_{out}, a_{out}, and n_{out} of $P(k_{\text{out}})$; and, *b* and *c* of $W(r)$. To sum up, the networks are analysed in terms of six individual parameters (denoted by capital letters), and ten function parameters (denoted by lower case letters).

4. RESULTS

The single parameters are calculated using Network Workbench; the degree distributions are fitted using the EzyFit toolbox for Matlab; and the ranked weight distributions are fitted in SPSS. For each parameter and for each of the three years (1990, 2000, and 2010), the mean parameter value and the Standard Error of the Mean (SEM) of all six network snapshots were calculated. The SEM indicates the amount of bimonthly variation. Figs. 1-16 illustrate the trend of each parameter average over the twenty-year period, and the vertical error bars (where visible, due to higher variance) indicate the SEM. Figs. 1-6 present the six individual network parameters in green. Figs. 7-14 show the eight degree distribution parameters in blue for in-degree and orange for out-degree. Figs. 15 and 16 report the ranked weight distribution parameters, *b* and *c*, in red. The results are discussed in the next section.

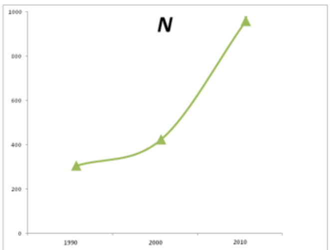

Fig. 1. Airports.

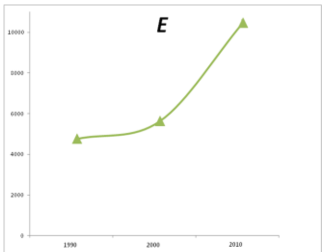

Fig. 2. Connections.

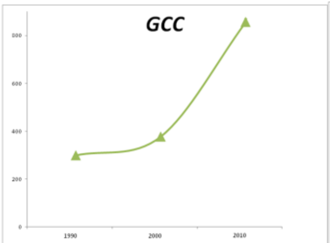

Fig. 3. Connected airports.

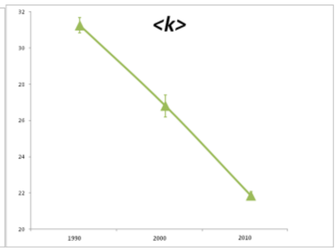

Fig. 4. Average connections.

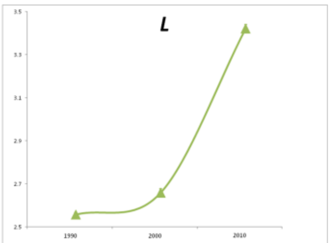

Fig. 5. Average hops.

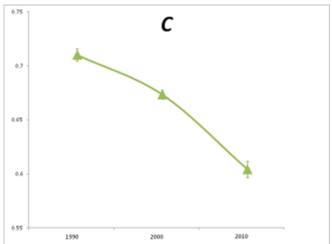

Fig. 6. Clustering.

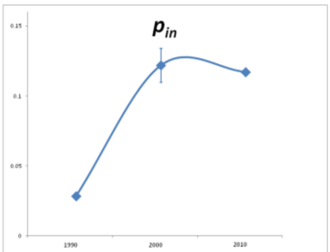

Fig. 7. P(0 connections in).

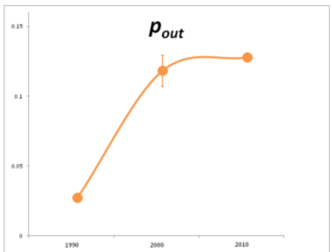

Fig. 8. P(0 connections out).

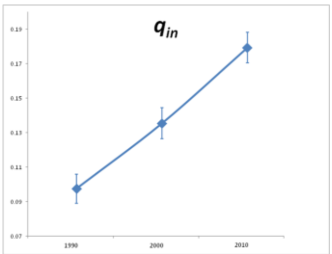

Fig. 9. P(1 connection in).

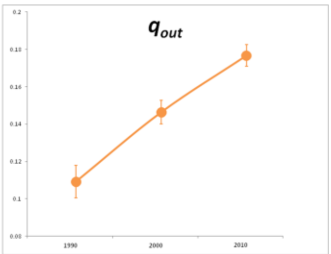

Fig. 10. P(1 connection out).

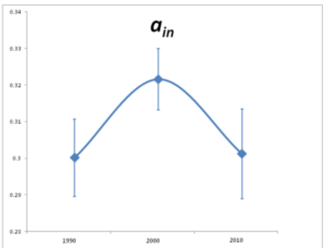

Fig. 11. Scaling factor a_{in}.

Fig. 12. Scaling factor a_{out}.

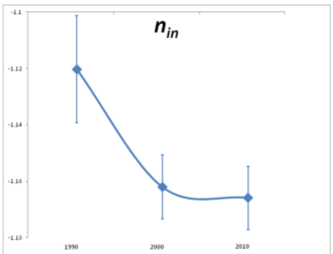

Fig. 13. Exponent n_{in}.

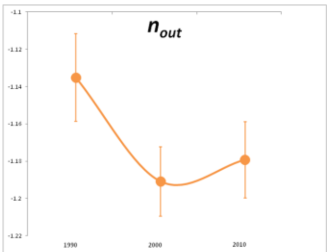

Fig. 14. Exponent n_{out}.

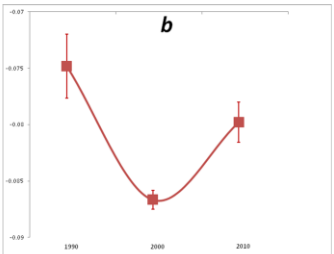

Fig. 15. Scaling factor b.

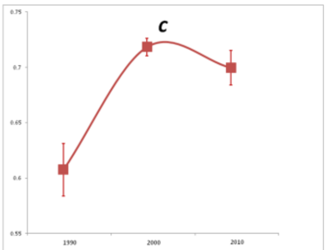

Fig. 16. Coefficient c.

5. DISCUSSION

The obtained results are discussed in three parts. Section 5.1 addresses the individual parameters, which are based on the global structure of the entire network. Section 5.2 covers the degree distribution parameters, which describe the structure of the USAN in terms of the airports' number of incoming and outgoing connections from/to other airports. Section 5.3 focuses on the weight distribution parameters, which highlight the high heterogeneity in the number of passengers on different connections.

5.1 Individual Network Parameters

Figs. 1-3 show the growth of the network in terms of airports, connections, and connected airports. Clearly, the expansion is much larger from 2000 to 2010, indicating a non-linear growth process. This observed behaviour is not unusual, as any transportation network is constantly affected by economic decisions, supply and demand, and many other factors. What is rather unusual is the fact that the average number of airport connections, Fig 4, displays a linear decline in time, due to the faster increase in number of airports compared to the number of airport connections. This means that many (probably small size) airports were introduced but they were not interconnected that well, unless already established airports lost some connections. Because of this rapid growth, the average hop length (Fig. 5) between any two airports in the US jumped from 2.5 to 3.5, within the past ten years. However, this does not imply that the average journey would need more changes; to the contrary, the network was optimised over time to reduce the changes of the average passenger by interconnecting airports with higher passenger demands, and disconnecting those less profitable. This is evident from the recent boom in low-cost airlines, providing many point-to-point flights between poorly connected destinations. Based on these facts, it is natural to assume that the clustering in the network increases, but Fig. 6 contradicts this; again, this must be due to the huge number of new airports. All these parameters have confirmed the immense development of the USAN, particularly in the first decade of the 21st century, and the next section explains this phenomenon in a little more detail.

5.2 Degree Distribution Parameters

Figs. 7 and 8 show the probability of an airport having zero incoming and outgoing connections, respectively. In other words, this parameter measures the proportion of very remote airports that only have some arrivals, or departures, per month. Clearly, the fraction rises from 1990 to 2000, indicating a significant increase in such poorly connected airports, but more interesting is the 2000 to 2010 period, which experienced no major change. Figs. 9 and 10 present the fraction of airports with just one incoming and outgoing connection, respectively. Again, these trends quantify the presence of minor airports, which increases linearly over the two decades. Figs. 11 and 12 report the fitting functions' estimates for the parameters from the previous two figures. Basically, they confirm that the fits are not able to approximate (especially for the year 2000) the first two data points that were extracted as p and q, since they do not obey the power-law relationship that the rest of the data does. The key parameter in a power-law is the exponent, as it controls the skew of the distribution. Therefore, between 1990 and 2000, Figs. 13 and 14 suggest an increasing exponent in absolute terms, since the scale of the figures is negative. This implies stronger preferential attachment, which means that already highly connected airports obtained more connections, while poorly connected airports received few new, or even lost existing, connections. The fact that the change between 2000 and 2010 is small, suggests that although there was a lack of point-to-point flights in the 90s, it may have been resolved in the 00s.

5.3 Weight Distribution Parameters

The ranked passenger distribution is the only characteristic of the dynamics on the network that is considered in this paper, and as such, cannot be taken as a complete description of the function of the network. Nevertheless, the results are interesting, and can be used as a basis for further analysis. Figs. 15 and 16 depict the two parameters of the logarithmic fit, and although further work is necessary to arrive at more precise conclusions, one thing is certain: the USAN exhibits considerable passenger variability over the course of a year. This is demonstrated by the error bars in the figures.

6. CONCLUSION

Evolution-based modelling of networks promises to be a useful tool for extracting detailed information about the complex interactions in networks that are typically getting larger, as demonstrated by this United States Airport Network case study. Specifically, it is necessary to build network models that satisfy three key conditions: modelling a significantly long time period; capturing fine temporal detail by high-resolution snapshots of the state of the system; and including multiple system features by using more complex network models. The approach described in this paper is simple and straightforward, and may be applied to the study of any transportation network, or more generally, to any evolving complex network. Further work on this case study will focus on finding community structure, additional measures for network dynamics, and forecasting future trends. The authors look forward to exciting new developments in the multi-disciplinary field of complex networks, which may apply to any domain-specific question.

ACKNOWLEDGEMENTS

The authors would like to thank the Brunel Research Support and Development Office for funding this project; Marc Barthelemy for recommending a reliable source of free data; Nadia Postorino for invaluable comments on the manuscript; and last but not least, Georgi Dimirovski for his kind invitation to COSY 2011.

REFERENCES

Alderson, D. & Willinger, W. 2005, "A contrasting look at self-organization in the Internet and next-generation communication networks", *IEEE Communications Magazine*, vol. 43, no. 7, pp. 94-100.

Amaral, L.A.N., Scala, A., Barthélémy, M. & Stanley, H.E. 2000, "Classes of small-world networks", *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11149-11152.

Bagler, G. 2008, "Analysis of the airport network of India as a complex weighted network", *Physica A-Statistical Mechanics and its Applications*, vol. 387, no 12, pp. 2972-2980.

Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. 2004, "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.

Barthelemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. 2005, "Characterization and modeling of weighted networks", *Physica A-Statistical Mechanics and its Applications*, vol. 346, no. 1-2, pp. 34-43.

Bounova, G. 2009, *Topological evolution of networks: case studies in the US airlines and language Wikipedias*, Massachusetts Institute of Technology.

Carreras, B.A., Lynch, V.E., Dobson, I. & Newman, D.E. 2002, "Critical points and transitions in an electric power transmission model for cascading failure blackouts", *Chaos*, vol. 12, no. 4, pp. 985-994.

Carter, S.L., Brechbühler, C.M., Griffin, M. & Bond, A.T. 2004, "Gene co-expression network topology provides a framework for molecular characterization of cellular state", *Bioinformatics*, vol. 20, no. 14, pp. 2242-2250.

Gegov, E., Gobet, F., Atherton, M., Freudenthal, D. & Pine, J. 2011, "Modelling Language Acquisition in Children using Network Theory", *European Perspectives on Cognitive Science*, New Bulgarian University Press, Sofia, 21/05/2011.

Girvan, M. & Newman, M.E.J. 2002, "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.

Guimera, R., Mossa, S., Turtschi, A. & Amaral, L.A.N. 2005, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7794-7799.

Li, W. & Cai, X. 2004, "Statistical analysis of airport network of China", *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 69, no. 4 Pt 2, pp. 046106.

Morine, M.J., Gu, H., Myers, R.A. & Bielawski, J.P. 2009, "Trade-offs between efficiency and robustness in bacterial metabolic networks are associated with niche breadth", *Journal of Molecular Evolution*, vol. 68, no. 5, pp. 506-515.

Xu, Z. & Harriss, R. 2008, "Exploring the structure of the U.S. intercity passenger air transportation network: a weighted complex network approach", *GeoJournal*, vol. 73, no. 2, pp. 87-102.