A DYNAMICAL MODEL FOR THE AIR TRANSPORTATION NETWORK

Massimiliano Zanin, Lucas Lacasa, Miguel Cea, Samuel Crístobal
Innaxis Foundation & Research Institute
Velázquez 157, 28002 Madrid, Spain
Email: mzanin@innaxis.org

KEYWORDS

LT: Discrete Event Modelling and Simulation in Logistics, Transport and Supply Chains

ABSTRACT

Although in the literature several Air Transportation Networks have been modeled through the mathematical framework of Complex Networks, less attention has been devoted to their evolution: specifically, topology is studied as static, excluding any schedule information. Simulating the growth of an air network is not an easy task, due to the great amount and heterogeneity of agents interacting (aircrafts, controllers and more). In this work we use a recently developed tool, called Scheduled Network, to define an algorithm to simulate such a growth; the basic assumption is that the cost for passengers should be minimized, which is approximated with the time needed to go from one airport to another one. Some results are presented, and the role and importance of hubs (that is, central airports where great part of the flights are concentrated) is discussed.

INTRODUCTION

The Air Transportation Network has been rapidly evolving in the last decades, mainly due to the many technological changes introduced: more efficient aircraft have modified any previous business perspective, giving birth for instance to low-cost airlines. The increment in the number of operations, and the consequent need for optimal investment strategy, is a significant challenge for regulation and control authorities: for example, Eurocontrol is forecasting a growth of up to 220% in the number of flights inside Europe in the time window $2005 - 2030$ (STATFOR, 2008).

Intuitively, the study of the structure of this network and the forecast of its evolution are extremely valuable tools for policy makers: but indeed are not easy tasks, due to the great number of agents interacting and to the complex interactions between them. Inside the Complex Science paradigm, we have chosen an instrument which simplifies that task, and we have adapted it to this context: that is, Complex Network theory.

Complex Networks are mathematical objects with a simple definition (Boccaletti et al., 2006; Newman, 2003): nodes (which represent any kind of physical or virtual entities) connected through links (once again, any kind of relations) following a given topology. Exploiting this generality, many studies of the structure of real and virtual systems (Costa et al., 2008), and properties of such topologies (Costa et al., 2007) have been proposed.

Although powerful, this network framework has an important drawback: the structure is considered as static, and does not include any time evolution. Let us show an example of how this can lead to some problematic consequences. In a previous work, Guimerà (Guimerà et al., 2005) constructed the network of worldwide flights by connecting two cities if there were a direct flight connecting them; no information regarding flights duration or allowed concatenations is therefore taken into account. For instance suppose that a customer has a flight at 12:00 from Stockholm to Paris, and another flight the same day from Paris to Toulouse at 13:00; clearly we cannot claim that Stockholm and Toulouse are connected in that time window by a path of length 2, as the passenger would be flying at 13:00 and would miss the second plane: he/she will have to wait until the next flight, maybe the following day.

To overcome this class of problems, an extension of Complex Network which includes the time scheduling has been developed by the authors (Zanin et al., 2009). This Scheduled Networks approach allows dynamically activating links according to an external source of information: as a consequence, it is possible to directly study its topology and extract useful information, as mean rotation times and system efficiency.

In this work, we have applied Scheduled Networks to several virtual air networks with the aim of forecasting their evolution. Initial conditions are defined as airport positions and flux of passengers between them: after that, links (that is, flights) are sequentially added to minimize an objective function which represents the cost for customers to travel across the network. The remainder of the work is organized as follows: first, an introduction to the mathematics of Scheduled Networks is presented; after that, several virtual networks are constructed, and some conclusions about the behavior of such systems is discussed.

SCHEDULED NETWORKS

The first step to construct an algorithm for simulating the growth of air traffic networks is to define an extension of Complex Networks which includes a time schedule. So, we start with the definition of a static directed network,

Proceedings 23rd European Conference on Modelling and Simulation ©ECMS Javier Otamendi, Andrzej Bargiela, José Luis Montes, Luis Miguel Doncel Pedrera (Editors)
ISBN: 978-0-9553018-8-9 / ISBN: 978-0-9553018-9-6 (CD)

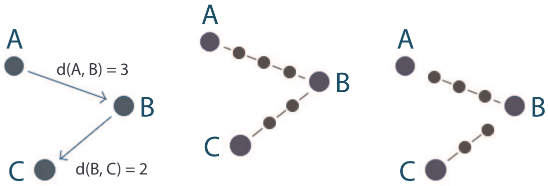

Figure 1: (Left) Representation of a simple directed network, which has been expanded (Center) to include a representation of time (i.e. the time needed to travel along a link). At the Right, both connections of the network are deactivated, by deleting the first link of each path.

where n *nodes* are connected through *links*; those connections are represented in an *adjacency matrix* $A_{n \times n}$, where the element a_{ij} has a value of 1 if there exists a link between nodes i and j, and zero otherwise (Bollobas, 2002).

This structure is now transformed to include the time needed to travel a connection, as shown in Fig. 1. To the *primary nodes*, which correspond to the original nodes of the graph, some *secondary nodes* are added, that are used to delay movements according to the scheduling information. Those secondary nodes are virtual, as they do not exist in the real system under study, and represent the time length of each link. The new adjacency matrix is now the combination of a *differential* adjacency matrix dA, which is constant and represents one time step, and an *Activation Matrix*, which holds the scheduling information. The global adjacency matrix, and therefore dA, is a square matrix of $n \times n$ elements, where n is the sum of primary and secondary nodes ($n = n_p + n_s$). To simplify the notation, S is constructed so that the primary nodes are represented first, followed by secondary ones. This new adjacency matrix is divided into four parts, as in Eq. 1.

$$A = \left[\begin{array}{c|c} P & R \\ \hline Act & T \end{array} \right] \tag{1}$$

Within Graph theory, the adjacency matrix fully characterizes the set of allowed movements that a given agent can perform inside a network. In a similar way, each one of the four sub-networks has a specific function:

- Persistence matrix (P): is a $n_p \times n_p$ identity matrix, which allows agents in a primary node to stay there indefinitely.

- Activation Matrix (Act): this matrix represents the schedule of the network, and is dynamically updated at each time step. When a link is activated, agents are allowed to move from a primary node to the first secondary node of that link.

- Reception Matrix (R): it moves an agent from the last secondary node of a link to the primary node which is the end of the link.

- Transfer Matrix (T): this matrix moves agents in a secondary node to the next node of the same link, so that they can travel until the destination airport; in other words, it simulates the passage of time.

A deeper explanation of how to construct the new Adjacency Matrix can be found in (Zanin et al., 2009), while an example of the resulting structure of the network of Fig. 1 can be found in the following matrix - note that the link from B to C is activated:

$$A = \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \tag{2}$$

It is easy to chain together several adjacency matrices, to represent a whole time window. The static adjacency matrix is the same in each time step, and just the Activation Matrix is updated following the given schedule. Mathematically, it can be expressed as:

$$\begin{aligned} A = A(t) \cdots A(t+\delta-1) A(t+\delta) = \\ = \prod_{p=t}^{t+1} [dA + Act(p)] \end{aligned} \tag{3}$$

In the original work (Zanin et al., 2009) this framework has been used to calculate several important metrics of the European Air Transportation Network, like efficiency and rotation time. Mathematics apart, now we want to simulate the growth of a virtual air network: and this tool helps us in creating a more realistic simulation by including the time dimension inside the system.

MODEL DEFINITION

Our model includes 9 virtual airports, placed in a regular lattice. Each airport is defined by two parameters: (i) the number of passengers that want to travel from that given airport, and (ii) a fitness, i.e. a value representing the relative attractiveness of that node for travelers. When both values are included in two vectors $\overline{P}_i$ and $\overline{F}_i$, with $1 \leq i \leq 9$, the effective number of passengers going from node j to k is:

$$p_{j,k} = P_j \frac{F_k}{\sum\limits_{1 \leq i \leq 9, i \neq j} F_i} \tag{4}$$

In other words, the passengers that are ready to depart from one airport are distributed to the other 8 according to their attractiveness. As time is included in the system, it is necessary to define at which time of the day

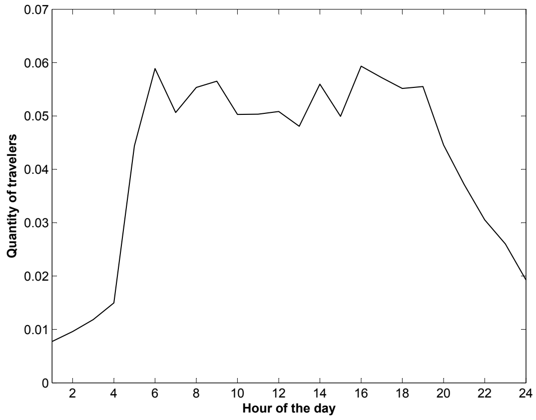

Figure 2: Proportion of passengers traveling through the European Air Network as a function of the time of the day.

those customers want to travel; with that aim, we have extracted from real European traffic data an estimation of the mean number of passengers as a function of the hour of the day. Results are shown in Fig. 2, and have been included in all simulations.

At this point, it is necessary to define a target function for the system, which should be minimized (or maximized) when creating the network of connections. In real life, airlines have a clear mission: maximize the company value. Benefits are a complex combinations of incomes (flight tickets, airplanes renting, . . .) and costs (airplanes depreciation, fuel, and so on); moreover, some companies (specifically, freighter airlines) have to dynamically rearrange connections to respond to customers' needs. With the aim of simplifying this complex space of variables, we have defined our function to be minimized as the cost for passengers to travel from a node to another, that is, the time required:

$$z = \sum_{i \in \mathrm{P}} T(i) \tag{5}$$

with P the group of all passengers, and $T(i)$ the time needed by passenger i expressed in hours. This is a realistic approach, as airlines can charge passengers according to the usefulness of the flight (or of the group of connected flights), which is in inverse proportion to the duration. We also further simplify the system by only considering passenger transportation (that is, no freighter airlines), which networks are more static.

With all the above definitions, it is clear that the best solution would be a fully-connected graph; to impose some economical restriction to this problem, we limit the number of allowed flights for each graph (or the number of available airplanes), and we check the evolution of the system when this quantity is gradually increased.

The growing algorithm is therefore defined as follows. At the beginning airports are not connected; after that, at each time step, a new single flight is added to the system. To determine which connection has to be chosen,

(40 links)

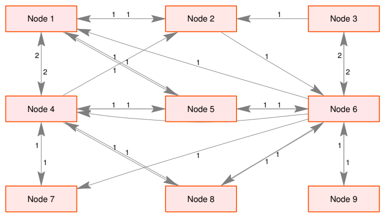

Figure 3: Growth of the aeronautical network without any hub, and linear costs. In this case, the resulting topology depends only on the statistical fluctuations of the number of passengers in each airport.

all possible flights (that is, each possible pair of airports, and each possible hour of departure) are checked, and the one which minimizes the global cost is added - what is mathematically called a *greedy algorithm*. Of course, this algorithm wants to emulate a real situation where an airline has enough business volume and decides to add a new flight.

We have simulated two main situations:

1. A regular lattice, where airports have similar number of passengers and fitness: only the geographical position is therefore defining the final network topology.

2. To the previous situation, a hub is added. In the aeronautical context, a hub is a central airport where most of the flights converge, and where maintenance is carried out. In our system, we have introduced such hub incrementing by 100% the number of passengers and the fitness of the central airport (Node 5 in all Figures).

SIMULATIONS' RESULTS

The result of growing the virtual aeronautical network described in the previous Section is shown in Fig. 3 and Fig. 4. To check the differences introduced by a hub, in Fig. 3 is shown the result for the first configuration (that is, all nodes have similar numbers of passenger and fitness); similarly, in Fig. 4 are shown four steps in the evolution of the second configuration, with a central hub: the evolution of the mean cost for this case is represented in Fig. 5.

Some considerations should be emphasized. First, in the case of Fig. 4 the emerging structure, at least for low number of connections, is an hub-and-spoke. This topology is usually defined as the best way to construct an aeronautical network, because with a limited number of flights all airports can be connected: of course, passengers have to take two different flights, one from the node of origin to the hub, and the second from the hub to the destination. Indeed this is the solution found by the growing algorithm; nevertheless when we further add connections, such star structure is lost, and many flights appear to link peripheral airports. In other words, what we have created in this simulation is a complex network

(18 links)

(30 links)

(40 links)

(60 links)

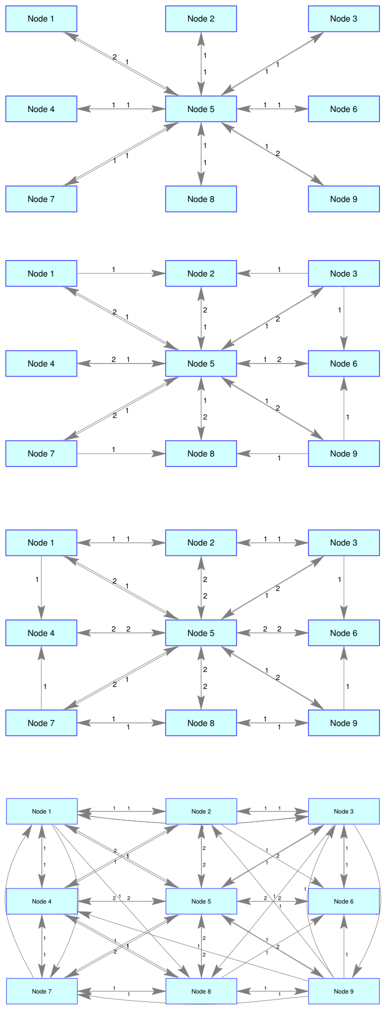

Figure 4: Growth of the aeronautical network with a central hub, and linear costs. Results are shown for 18, 30, 40 and 60 links.

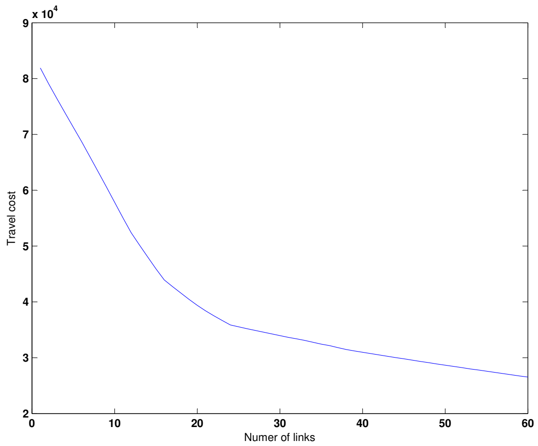

Figure 5: Evolution of the mean travel time of passengers in a given node as a function of the number of links; the configuration considered is with a central hub and linear costs.

of connections, where peripheral flights are usually operated by regional or low-cost airlines.

(40 links)

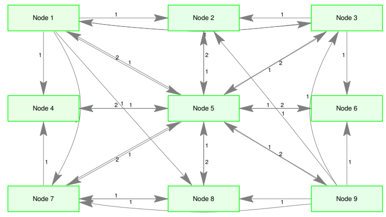

Figure 6: Growth of the aeronautical network with a hub in node 5 and quadratic costs.

The question which may arise is what is the main reason to create a hub-and-spoke structure when few flights are allowed: the presence of a *natural* hub (i.e. a node with a higher volume of passengers), or its central position? The answer can be deduced from Fig. 3: when no node is artificially promoted to hub by increasing its traffic, the central position is not enough for node 5 to become a central airport.

In Figs. 3 and 4 we have supposed that the cost for users is a linear function of the time required for the trip (as in Eq. 5); to ensure that results do not depend on the form of the target function, we have also calculated the network evolution for a quadratic cost, in the form:

$$z = \sum_{i \in \mathrm{P}} T(i)^2 \qquad (6)$$

With this new cost, results does not differ substantially and fewer links are needed to break the star topology (see Fig. 6). Of course, to create a more realistic simulation for real world air networks, it would be necessary to define a better economical approximation to the function which define the cost for users - or a better approximation of airlines benefits.

In order to check the robustness of obtained results, we have executed different simulations with noise in the flux of passengers. Obtained networks have similar characteristics; as may be expected, the regular structure is broken if the noise level is increased: see Fig. 7. Also, we have calculated the sensitivity of the solution of Fig. 4 with 30 links to noise applied to customers' fluxes; values of matrix P are now random variables with a normal distribution, centered in $p_{i,j}$, and whose standard deviation is the noise intensity. Results are shown in Fig. 8. It is worth noting the importance of such calculation when made in a real air transport network, as it would allow a quantification of the uncertainty embedded in the system.

CONCLUSIONS

From the above, some important informations can be inferred, which are of utmost importance for policy making and for the strategic planning of an airline.

First, the hub-and-spoke structure of connections, which is considered the standard configuration for an aeronautical network, is useful only as long as the number of flights is lower than a certain threshold: and this

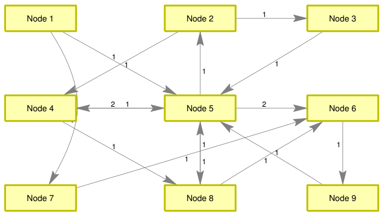

Figure 7: Growth of the aeronautical network with a hub in node 5, linear costs and 18 links, when a strong noise is added to the matrix P.

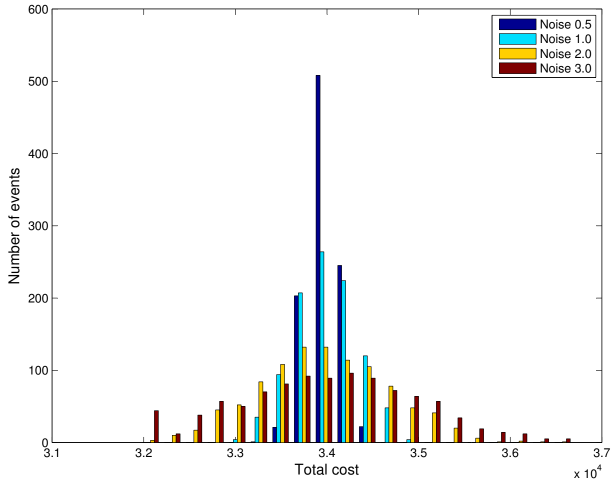

Figure 8: Distribution of the obtained total cost when different noise levels are added to the matrix P, with the network configuration of Fig. 4. This represents the sensitivity of the system to random events.

situation can be found only in new markets. The reality is much more complex: regional airlines cover connections between non-hub airports, and low-cost companies try to dodge big airports to avoid high costs. The only way to take into account all those aspects is to construct a complex framework which allows a simulated growth of the network: this tool can indeed help in understanding the future evolution of the real air transportation system.

Second consideration: hubs appear as a consequence of the higher volume of passengers of some airports, and are not necessarily related with their physical position. This explains some interesting examples, like the hub of Munich, which is in the southern border of Germany. At the same time, it is not feasible to design a hub for political interests: if the selection is not supported by economical considerations in term of efficiency for passengers, it would be difficult for that hub to develop.

In further studies based on this model, we want to better understand the economical and geographical factors that can affect the choice of the location of a hub, and its role in the diffusion process. Moreover, the next step will be to apply this framework to real air networks, with the aim to forecast their evolution, as well as other factors: sensitivity to noise or random events, or uncertainty of the whole system. In turn this would allow a better planning of infrastructure investments and air structure modifications.

ACKNOWLEDGMENTS

Authors wish to thank all Innaxis Foundation group, as well as four anonymous referees, for the helpful suggestions.

REFERENCES

STATFOR, the EUROCONTROL Statistics and Forecast Service (2008). *Flight Movements 2008 - 2030*.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.-U. (2006). Complex networks : Structure and dynamics. *Physics Reports*, 424:175.

Newman, M. E. J. (2006). The structure and function of complex networks. *SIAM Review*, 45:67–256.

Costa, L. da F., Oliveira, O. N., Travieso, G., Rodrigues, F. A., Boas, P. R. V., Antiqueira, L., Viana M. P. and Correa da Rocha, L. E. (2008). Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications. *arXiv:0711.3199v2 [physics.soc-ph]*.

Costa, L. da F., Rodrigues, F. A., Travieso, G. and Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56:167–242.

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. U.S.A.*, 102.

Zanin, M., Lacasa, L. and Cea, M. (2009). Dynamics in Scheduled Networks. *Chaos*, in press.

Bollobas, B. (2002). *Modern Graph Theory*. Springer.

AUTHOR BIOGRAPHIES

MASSIMILIANO ZANIN was born in Verona, Italy, in 1982. Mr. Zanin is currently studying Management Aeronautics that is taught in the Autónoma University of Madrid, Spain. His working experiences include software development in the field of industrial logistic, for several Spanish handling firms and as free-lance. Actual research activities are centered on the complex structure and non-linear behavior of several real world systems, like air traffic network, recommendation algorithms or chaotic cryptography applications. His email is `mzanin@innaxis.org`.

LUCAS LACASA was born in Madrid (Spain) in 1979. He received a BSc in Theoretical Physics by the Universidad Complutense de Madrid (Spain) in 2004, and both Msc and PhD in Physics of Complex Systems by the Universidad Politécnica de Madrid (Spain) in 2006 and 2009 respectively. He has worked in several aspects of complex systems analysis including critical phenomena or nonlinear time series analysis. His current research activities include the application of Complexity Science to the Air transport system.

MIGUEL CEA was born in Madrid (Spain) in 1978. He received a BSc, MSc and PhD in Mathematics by the Universidad Autónoma de Madrid (Spain) in 2004, 2006 and 2009 respectively. He has worked in several aspects of numerical modeling including optimal design and computational graph theory. His current research activities include the modeling and simulation of multicomponent systems, in particular the design of new algorithms to characterize the performance of the Air transport system.

SAMUEL CRÍSTOBAL was born in Malaga (Spain) in 1981. He received a BSc in Mathematics by the Universidad Complutense de Madrid (Spain) in 2008 and he is currently undertaking a Msc in Mathematic Research. He also graduated as a Telecommunication Engineer by the Universidad Politécnica de Madrid in 2005. He has worked in multidisciplinary aspects of mathematics and applications like Chaotic dynamical systems for CDMA modulations, methods of Commutative Algebra for computational optimization and study of Elliptic Curves and their applications in Cryptography. His current research activities are focused on developing a model of ATM network capable of obtain advanced key performance indicators for the future SESAR scenarios.