ELSA
Empirically grounded agent based models for the future ATM scenario

SESAR INNOVATION DAYS
Tolouse, 30/11/2011

Salvatore Miccichè
University of Palermo, dept. of Physics

ELSA Project
Toward a complex network approach to ATM delays analysis
Joint work with

DEEP BLUE:
Simone Pozzi (Project Coordinator)
Valentina Beato

SNS:
Fabrizio Lillo

UNIPA:
Rosario N. Mantegna
Salvatore Miccichè

Marc Bourgois: project officer
This paper describes a project that is part of SESAR Workpackage E, which is addressing long-term and innovative research. The project was started on May 2011 so this description is limited to an outline of the project objectives augmented by some early findings.
ELSA Objective

» Analyse, describe and model the dynamics of the ATM system
  » in the current scenario
  » in the future SESAR scenario(s)

Three main steps:

» **WP1** - characterisation of statistical regularities in the current scenario
  » *analysis of ATM data with Complex Systems techniques*

» **WP2** - simulation of the emergent properties of the trajectory-based SESAR scenario
  » *development of an Agent Based Model*

» **WP3** - inform the design of a tool to monitor, predict and intervene on the ATM system
  » *design of a prototype of a decision support tool*
**Complexity tools: networks**

**Network** - it is a graph with nodes connected by links
The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles

R. Guimerà*, S. Mossa*, A. Turtschi†, and L. A. N. Amaral

*Department of Chemical and Biological Engineering, Northeastern University, Boston, MA 02115; †European Synchrotron Radiation Facility, B.P. 220, F-38043 Grenoble, Cedex, France; ‡European Bioinformatics Institute, Campus Hinxton 7, D-41476, Knokke, Germany

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved April 1, 2009; reviewed for publication October 21, 2009

We analyze the global structure of the worldwide air transportation network, a critical infrastructure with an enormous impact on local, national, and international economies. We find that the worldwide air transportation network is a scale-free small-world network. In contrast to the prediction of scale-free network models, however, we find that the most connected cities are not necessarily the most central, resulting in anomalous values of the centrality. We demonstrate that these anomalies arise because of the multi-community structure of the network. We identify the communities in the air transportation network and show that the community structure cannot be explained solely based on geographical constraints and that geopolitical considerations have to be taken into account. We identify each city’s global role based on its pattern of intercommunity and intracommunity connections, which enables us to obtain scale-specific representations of the network.

Much research has been conducted on the definition of models and algorithms that enable one to solve problems of optimal network design (9, 10). However, a worldwide, “system” level analysis of the structure of the air transportation network is still lacking. However, just as one cannot fully understand the complex dynamics of ecosystems by looking at simple food chains (11) or the complex behavior in cities by studying isolated biochemical pathways (12, 13), one cannot fully understand the dynamics of the air transportation system without a “holistic” perspective. Modern “network analysis” (14–18) provides an ideal framework within which to pursue such a study.

We analyze here the worldwide air transportation network. We build a network of 3,883 nodes, villages, towns, and cities with at least one airport and establish links between pairs of nodes that are connected by nonstop passenger flights. We find that the worldwide air transportation network is a small-world network (19) for which (i) the number of nonstop connections from a given city and (ii) the number of shortest paths going through a given city have distributions with extreme tails.

The real airport network behaves quite differently from a random network.

Fig. 2. Most-connected versus most-central cities in the worldwide air transportation network: (a) Betweenness as a function of the degree for the cities in the worldwide air transportation network circles. For the randomized network, the betweenness is well described as a quadratic function of the degree (dashed line) with 95% of all data falling inside the gray region. In contrast, the strong correlation between degree and betweenness found for the randomized networks, the air transportation network comprises many cities that are highly connected but have small betweenness and, conversely, many cities with small degree and large betweenness. We define a blue region containing the 25 most central cities in the world and a yellow region containing the 25 most connected cities. Surprisingly, we find there are only a few cities with large betweenness and degree (green region, which is the intersection of the blue and yellow regions). (b) The 25 most connected cities in the world. (c) The 25 most central cities in the world.

Degree - number of destinations that can be reached from an airport

Betweenness - This a measure of how central is an airport in the network.
DATA: what we received

DDR data: M1/M3 and ALL_FT files

What are the different file types?

SO6
Stores SAAM 4D trajectories and limited flight information. Contains points coordinates. M1 means last filled flight plan. M3 means last filled flight plan updated with radar data. Mainly used in SAAM for queries, analysis, display, animation ...
The first point of every trajectory is the beginning of the runway.

EXP2
stores flights information, but no trajectories. Used for SAAM as a traffic demand for the assignment.

ALL_FT
Stores flight information and trajectories. Used by NEVAC and CFMU simex (at least for the historic ones). Do not contain point’s coordinates, so to be used in a given environment.

ISSUES:
What is the last filled flight plan?
What is the threshold that triggers a correction in the M3 file?
What exactly are the times reported in the M1/M3 files?
M1/M3 and ALL FT files

30 segments

Also, segments may change name, .... !!!
The same callsign appears with two different IFPS IDs. This makes problematic a comparison with M1/M3 files because the IFPS code does not appear in the M1/M3 files.
DATA: CLEANING

347_ENV_20110601.ALL_FT contains:
37338 records, i.e. different IFPS IDs
33704 records relative to day 01/06/2011
29401 distinct callsigns relative to day 01/06/2011

20110601_m1.so6 contains
31150 different flight IDs
28631 distinct callsigns

20110601_m3.so6 contains
31138 different flight IDs
28621 distinct callsigns

The overall intersection is 28613

The same for landing?
The same for take-off times?
SOME PRELIMINARY RESULTS:
WHAT DATA DID WE CONSIDER?

M1 and M3 files

We put aside the information on segments at this stage.

We were interested in two aspects:
1. The *network* structure of the system *airport*-flights
2. A network characterization of *flight delays*

1. We set a *link* between two airports if there is *at least a flight connecting them*.

2. We define: *flight delay* = *(arrival time in M3)*
   minus
   *(arrival time in M1)*
FLIGHT network: weighted and directed
The weight of each link is the number of flights. The number of flights from A to B is in general different from the number of flights from B to A, although these numbers are very close.

ROUTE network: weighted and undirected
The weight of each link is the total number of flights.

ECTL flights only
SOME PRELIMINARY RESULTS:
Network of Airports
ROUTE NETWORK

nodes - airports
<k> average degree - number of connections with other airports
<s> average strength - number of flights in each airport
<l> average path length - minimum number of flights that connects two airports

<table>
<thead>
<tr>
<th>day</th>
<th>nodes</th>
<th>flights</th>
<th>links</th>
<th>&lt; s &gt;</th>
<th>&lt; k &gt;</th>
<th>max(s)</th>
<th>max(k)</th>
<th>largest cluster</th>
<th>&lt; l &gt; s</th>
<th>&lt; l &gt; k</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>June 1</td>
<td>1375</td>
<td>31224</td>
<td>8498</td>
<td>45.4</td>
<td>12.4</td>
<td>1452</td>
<td>221</td>
<td>3.34</td>
<td>3.29</td>
</tr>
<tr>
<td>Th</td>
<td>June 2</td>
<td>1296</td>
<td>27660</td>
<td>7757</td>
<td>42.6</td>
<td>12.0</td>
<td>1426</td>
<td>226</td>
<td>3.33</td>
<td>3.27</td>
</tr>
<tr>
<td>F</td>
<td>June 3</td>
<td>1291</td>
<td>28280</td>
<td>8039</td>
<td>43.8</td>
<td>12.5</td>
<td>1384</td>
<td>219</td>
<td>3.31</td>
<td>3.23</td>
</tr>
<tr>
<td>Sa</td>
<td>June 4</td>
<td>1153</td>
<td>25289</td>
<td>7642</td>
<td>43.8</td>
<td>13.3</td>
<td>1343</td>
<td>245</td>
<td>3.17</td>
<td>3.15</td>
</tr>
<tr>
<td>Su</td>
<td>June 5</td>
<td>1188</td>
<td>28175</td>
<td>8104</td>
<td>47.4</td>
<td>13.6</td>
<td>1451</td>
<td>232</td>
<td>3.14</td>
<td>3.12</td>
</tr>
<tr>
<td>M</td>
<td>June 6</td>
<td>1342</td>
<td>30435</td>
<td>8243</td>
<td>45.3</td>
<td>12.3</td>
<td>1494</td>
<td>221</td>
<td>3.35</td>
<td>3.29</td>
</tr>
<tr>
<td>T</td>
<td>June 7</td>
<td>1348</td>
<td>30397</td>
<td>8131</td>
<td>45.1</td>
<td>12.1</td>
<td>1494</td>
<td>220</td>
<td>3.34</td>
<td>3.28</td>
</tr>
</tbody>
</table>

TABLE I
FULL network. ℓ is the path length. For a detailed explanation, see text.

highly connected system

ROUTE network
SOME PRELIMINARY RESULTS:
Network of Airports
ROUTE NETWORK

ROUTE network: some metrics distributions

(01/06/2011)
ROUTE network:
Fit: 1.39

If the number of destinations becomes twice, then the number of flights increases by a factor $2^{1.39} = 2.45$
SOME PRELIMINARY RESULTS:

Delay

There seem to be a tendency such that overloaded airports show larger delays.

Laplace

Negative delays

7 days are pooled together

15 min threshold

fraction of delayed flights vs. node strength (number of flights)
ROUTE FDR-network of airports

The network includes
1_fdr: 743/1375 nodes, 804 links
1_fdr: 74 connected components
1_fdr: 431 nodes in largest

FDR-network of flights

The network includes
2_fdr: 28395/31224 nodes, 148509 links
2_fdr: 5891 connected components
2_fdr: 95 nodes in largest
**Bipartite system of segments and flights**

<table>
<thead>
<tr>
<th>Segment 1</th>
<th>Flight 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 1</td>
<td>Flight 2</td>
</tr>
<tr>
<td>Segment 1</td>
<td>Flight 3</td>
</tr>
<tr>
<td>Segment 1</td>
<td>Flight 4</td>
</tr>
<tr>
<td>Segment 1</td>
<td>Flight 5</td>
</tr>
<tr>
<td>Segment 2</td>
<td>Flight 2</td>
</tr>
<tr>
<td>Segment 2</td>
<td>Flight 3</td>
</tr>
<tr>
<td>Segment 2</td>
<td>Flight 6</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
</tr>
</tbody>
</table>

Nodes can be either segments or flights.

Two segments are connected if they have at least a flight in common.

Two flights are connected if they have at least a segment in common.

We can study the segments topology and how it behaves with respect to delays. Are there segments that show peculiarities with respect to delays?
Once we have the networks (airports, segments, ...) what do we do?

**Propagation of delays over network**

Divide each day in 24 intervals of 1 hour each.

For each node (for example, airport, ..... and for each time interval one can consider the number of departure flights. One can compute how many are delayed (M1 vs M3 comparison) for more than 15 minutes. The same for arrivals.

Such analysis can give insights about the way delays spread over the network.
A few issues about data are still open

The network approach seems feasible. It can prove to be extremely useful when studying the ATM system.

Propagation of delays over networks at the level of airports and flight segments.
ELSA Objective

» Analyse, describe and model the dynamics of the ATM system
  » in the current scenario
  » in the SESAR scenario(s)

Three sub-objectives:
» characterisation of statistical regularities in the current scenario

» simulation of the emergent properties of the trajectory-based SESAR scenario

» inform the design of a tool to monitor, predict and intervene on the ATM system