

ELSA

Empirically grounded agent based models
for the future ATM scenario

SESAR INNOVATION DAYS

Toulouse, 30/11/2011

Salvatore Micciché
University of Palermo, dept. of Physics

ELSA Project

Toward a complex network approach to ATM delays analysis



UNIVERSITÀ
DEGLI STUDI
DI PALERMO



SCUOLA
NORMALE
SUPERIORE
PISA

Joint work with

DEEP BLUE: Simone Pozzi (Project Coordinator)
Valentina Beato



SNS: Fabrizio Lillo



UNIPA: Rosario N. Mantegna
Salvatore Miccichè



Marc Bourgois: project officer

FOREWORD

*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**.*

TOOLS

DATA

RESULTS

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

NETWORKS IN ATM



Complexity tools: networks

NETWORKS IN ATM: metrics

The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles

R. Guimerà*, S. Mossa†, A. Turttschi‡, and L. A. N. Amaral**§

*Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208; †European Synchrotron Radiation Facility, B.P. 220, F-38043 Grenoble, Cedex, France; and ‡Avanade Deutschland GmbH, Campus Kronberg 7, D-61476, Kronberg, Germany

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved April 5, 2005 (received for review October 27, 2004)

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 multicomponent 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.

complex networks | betweenness centrality | critical infrastructures

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 cells 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 locales, villages, towns, and cities with at least one airport and establish links between pairs of locales 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 that are scale-free. In contrast to the prediction of scale-free

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

Betweenness - This a measure of how central is an airport in the network.

The real airport network behaves quite differently from a random network

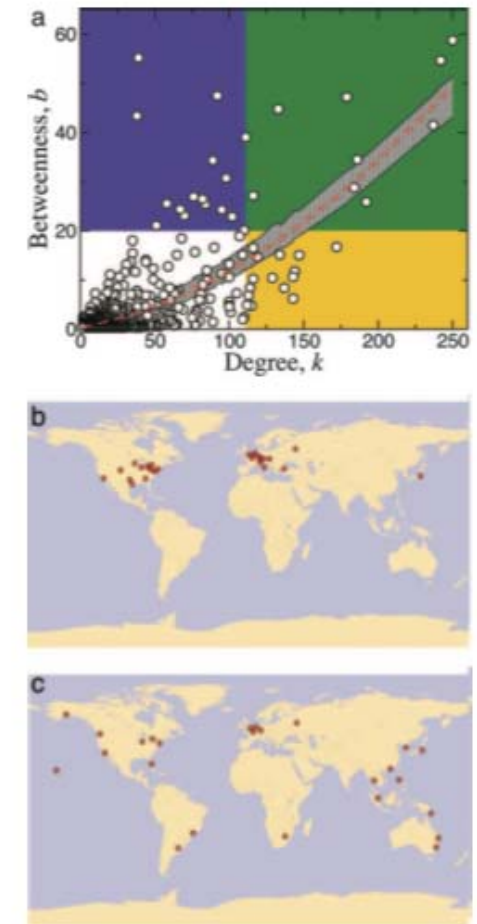


Fig. 2. Most-connected versus most-central cities in the worldwide air transportation network. (a) Betweenness as a function of the degree for the cities in the worldwide air transportation network (circles). For the randomized network, the betweenness is well described as a quadratic function of the degree (dashed line) with 95% of all data falling inside the gray region. In contrast to the strong correlation between degree and betweenness found for 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.

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.

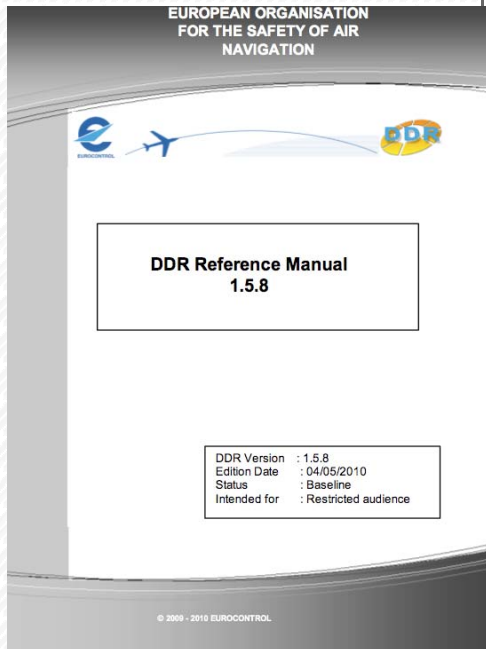
ISSUES:

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?

DATA: what we received



M1/M3 and ALL_FT files

DATA: CLEANING

```
20100701_m1_2.so6 (modified) -
File Edit Search Preferences Shell Macro Windows
/usr2/micciche/FILES-ATM/DATI/FLIGHT_2/20100701_m1_2.so6 3952 bytes

LPPG *5PG LPPG LEMD E120 002000 002505 0 81 0 CALLSIGN 100701 100701 2940.5833 152.8667 2942.3 128.8 137400625 1 15.8746 0
*5PG OL LPPG LEMD E120 002505 003305 81 171 0 CALLSIGN 100701 100701 2942.3 128.8 2923.8 143.1833 137400625 2 20.7776 0
OL MONOT LPPG LEMD E120 003305 004215 171 190 0 CALLSIGN 100701 100701 2923.8 143.1833 2885 147.4333 137400625 3 38.9024 0
MONOT GIPEX LPPG LEMD E120 004215 004705 190 190 2 CALLSIGN 100701 100701 2885 147.4333 2865 147.4333 137400625 4 20.6595 0
GIPEX NEV LPPG LEMD E120 004705 005605 190 190 2 CALLSIGN 100701 100701 2865 147.4333 2862 147.4333 137400625 5 3.7347 0
NEV LEMIN LPPG LEMD E120 005605 010510 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 6 40.5662 0
LEMIN RISUN LPPG LEMD E120 010510 011000 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 7 47.4656 0
RISUN LARON LPPG LEMD E120 011000 011425 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 8 43.7451 0
LARON GUERE LPPG LEMD E120 011425 011535 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 9 24.6534 0
GUERE LMG LPPG LEMD E120 011535 012715 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 10 37.4627 0
LMG NORON LPPG LEMD E120 012715 013305 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 11 22.4297 0
NORON VELIN LPPG LEMD E120 013305 014150 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 12 36.723 0
VELIN SAV LPPG LEMD E120 014150 014705 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 13 21.5004 0
SAV ENSAC LPPG LEMD E120 014705 015545 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 14 13.3058 0
ENSAC RILON LPPG LEMD E120 015545 020040 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 15 21.7309 0
RILON MAGEC LPPG LEMD E120 020040 020350 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 16 12.0577 0
MAGEC BTZ LPPG LEMD E120 020350 020855 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 17 4.7289 0
BTZ URUNA LPPG LEMD E120 020855 021150 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 18 17.9659 0
URUNA SSN LPPG LEMD E120 021150 021250 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 19 17.122 0
SSN DITOP LPPG LEMD E120 021250 021710 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 20 20.504 0
DITOP PPN LPPG LEMD E120 021710 022105 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 21 17.122 0
PPN NOLSA LPPG LEMD E120 022105 022555 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 22 11.5305 0
NOLSA ALEPO LPPG LEMD E120 022555 022715 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 23 5.6086 0
ALEPO GARVU LPPG LEMD E120 022715 022955 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 24 11.5305 0
GARVU HERMI LPPG LEMD E120 022955 024050 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 25 46.5436 0
HERMI BAN LPPG LEMD E120 024050 024315 190 190 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 26 9.9964 0
BAN PINAR LPPG LEMD E120 024315 024805 190 189 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 27 20.647 0
PINAR *7BAN LPPG LEMD E120 024805 025020 189 167 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 28 9.398 0
*7BAN BUDOM LPPG LEMD E120 025020 025420 167 130 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 29 15.4788 0
BUDOM ASBIN LPPG LEMD E120 025420 030205 130 70 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 30 29.1498 0
ASBIN LEMD LPPG LEMD E120 030205 030830 70 0 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 31 21.8808 0
```

31 segments

```
20100701_m3_2.so6 (modified) - /usr2/micciche/FILES-ATM/DATI/FLIGHT_2/
File Edit Search Preferences Shell Macro Windows
/usr2/micciche/FILES-ATM/DATI/FLIGHT_2/20100701_m3_2.so6 3842 bytes

LPPG *5PG LPPG LEMD E120 000600 001410 0 81 0 CALLSIGN 100701 100701 2940.5833 152.8667 2942.3 128.8 137400625 1 15.8746 0
*5PG OL LPPG LEMD E120 001410 001545 81 140 0 CALLSIGN 100701 100701 2942.3 128.8 2923.8 143.1833 137400625 2 20.7776 0
OL MONOT LPPG LEMD E120 001545 002325 140 190 0 CALLSIGN 100701 100701 2923.8 143.1833 2885 147.4333 137400625 3 38.9024 0
MONOT GIPEX LPPG LEMD E120 002325 002755 190 204 0 CALLSIGN 100701 100701 2885 147.4333 2865 147.4333 137400625 4 20.6595 0
GIPEX KUSEK LPPG LEMD E120 002755 002855 204 211 0 CALLSIGN 100701 100701 2865 147.4333 2862 147.4333 137400625 5 3.7347 0
KUSEK *1002R LPPG LEMD E120 002855 003945 211 264 0 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 6 40.5662 0
*1002R GUSON LPPG LEMD E120 003945 004945 264 270 0 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 7 47.4656 0
GUSON LMG LPPG LEMD E120 004945 005905 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 8 43.7451 0
LMG NORON LPPG LEMD E120 005905 010435 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 9 24.6534 0
NORON VELIN LPPG LEMD E120 010435 011250 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 10 37.4627 0
VELIN SAV LPPG LEMD E120 011250 011750 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 11 22.4297 0
SAV ENSAC LPPG LEMD E120 011750 012605 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 12 36.723 0
ENSAC RILON LPPG LEMD E120 012605 013100 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 13 21.5004 0
RILON MAGEC LPPG LEMD E120 013100 013425 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 14 13.3058 0
MAGEC BTZ LPPG LEMD E120 013425 013950 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 15 21.7309 0
BTZ URUNA LPPG LEMD E120 013950 014045 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 16 12.0577 0
URUNA SSN LPPG LEMD E120 014045 014055 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 17 4.7289 0
SSN DITOP LPPG LEMD E120 014055 014215 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 18 17.9659 0
DITOP PPN LPPG LEMD E120 014215 014630 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 19 17.122 0
PPN NOLSA LPPG LEMD E120 014630 015140 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 20 20.504 0
NOLSA ALEPO LPPG LEMD E120 015140 015300 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 21 5.6086 0
ALEPO GARVU LPPG LEMD E120 015300 015530 270 270 2 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 22 11.5305 0
GARVU HERMI LPPG LEMD E120 015530 020535 270 254 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 23 46.5436 0
HERMI BAN LPPG LEMD E120 020535 020725 254 228 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 24 9.9964 0
BAN PINAR LPPG LEMD E120 020725 021145 228 184 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 25 20.647 0
PINAR *7BAN LPPG LEMD E120 021145 021250 184 172 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 26 9.398 0
*7BAN BUDOM LPPG LEMD E120 021250 021520 172 130 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 27 15.4788 0
BUDOM ASBIN LPPG LEMD E120 021520 022135 130 64 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 28 29.1498 0
ASBIN *1002S LPPG LEMD E120 022135 022615 64 32 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 29 16.7506 0
*1002S LEMD LPPG LEMD E120 022615 022750 32 0 1 CALLSIGN 100701 100701 2862 147.4333 2862 147.4333 137400625 30 5.172 0
```

30 segments

Also, segments may change name, !!!

The same callsign appears with two different IFPS IDs

DATA: CLEANING

```
EFPS;EPRE;AEP/24;AEP;R320;20110601100000;BB95112653;20110601100000;FPL;FP  
FGSL;GMME;CALLSIGN;;F900;;BB95112653;20110601100000;FPL;FP  
EDDE;EDDT;DIH400;DIH;A321;20110601104500;A180660780;20110601104500;FPL;FP
```

```
EFPA;EPWA;FIN/43;FIN;A190;20110601142900;AA82671149;20110601142000;FPL;FP  
FGSL;GMME;CALLSIGN;;F900;20110601100000;BB95113198;20110601100000;FPL;FP  
EDD1;EDDT;MAYR604;MAY;A320;20110601070700;BB05100000;20110601071000;FPL;FP
```

This makes problematic a comparison with M1/M3 files because the IFPS code does not appear in the M1/M3 files

```
SME_EFES EIME EIES MD02 120342 120333 3 0 1 1331204 110601 110601 2433.4 370.0000 2433.5107 371.0007 147558177 13 1.5000 0  
FGSL_TMS FGSL GMME F900 111000 141314 0 400 0 CALLSIGN 110601 110601 225.3167 522.5167 1368.4667 326.8 147558178 1 1158.7421 0  
TMS_TAM FGSL GMME F900 141314 141319 400 400 2 CALLSIGN 110601 110601 1368.4667 326.8 1368.4 326.7333 147558178 2 .0907 0  
TAM_SIHA FGSL GMME F900 141319 150647 400 430 0 CALLSIGN 110601 110601 1368.4 326.7333 1710 210.7333 147558178 3 357.2242 0  
SIHA_MNA FGSL GMME F900 150647 152556 430 430 2 CALLSIGN 110601 110601 1710 210.7333 1833.5167 171.7 147558178 4 128.1 0  
MNA_MEZAB FGSL GMME F900 152556 153946 430 430 2 CALLSIGN 110601 110601 1833.5167 171.7 1894.8167 90 147558178 5 93.0295 0  
MEZAB_SEFRA FGSL GMME F900 153946 155900 430 430 2 CALLSIGN 110601 110601 1894.8167 90 1980 -24.7 147558178 6 129.0602 0  
SEFRA_ARIAM FGSL GMME F900 155900 161104 430 430 2 CALLSIGN 110601 110601 1980 -24.7 2033.3667 -97.6167 147558178 7 80.93 0  
ARIAM_FES FGSL GMME F900 161104 163626 430 312 1 CALLSIGN 110601 110601 2033.3667 -97.6167 2035.65 -300.8167 147558178 8 168.6495 0  
FES_VARAS FGSL GMME F900 163626 164711 312 114 1 CALLSIGN 110601 110601 2035.65 -300.8167 2041.5333 -380.9 147558178 9 66.67 0  
VARAS_SBI FGSL GMME F900 164711 165157 114 40 1 CALLSIGN 110601 110601 2041.5333 -380.9 2050.5833 -399.85 147558178 10 18.1143 0  
SBI_GMME FGSL GMME F900 165157 165427 40 0 1 CALLSIGN 110601 110601 2050.5833 -399.85 2043.0833 -405.1 147558178 11 8.6685 0  
EDDE;EDDT;DIH400;DIH;A321;20110601104500;A180660780;20110601104500;FPL;FP
```

flight ID

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?

Confronto_m0_m1_m3_prova.xls							
	A	B	C	D	E	F	G
1	DAY	ID of the FLIGHT	AIRPORT START	AIRPORT END	IOBT from ALL_FT	start time first segment from M1 file	start time first segment from M3 file
2	20110601	CALLSIGN 1	EDLW	EBAW	130000	130600	130600
3	20110601	CALLSIGN 2	EVRA	EETN	134500	135000	135500
4	20110601	CALLSIGN 3	LIPX	EBBR	130000	131000	141800
5	20110601	CALLSIGN 4	FGSL	GMME	110000	111000	111000
6	20110601	CALLSIGN 5	LIPQ	LIPQ	93000	93500	92900
7	20110601	CALLSIGN 6	LLNV	LLNV	100000	101000	101000
8	20110601	CALLSIGN 7	LLNV	LLNV	41500	42500	42500
9	20110601	CALLSIGN 8	LSZS	EGGW	120000	121000	122500
10	20110601	CALLSIGN 9	LGIR	HSSS	83000	83500	83400
11	20110601	CALLSIGN 10	UKBB	UKFB	120000	121500	121100
12	20110601	CALLSIGN 11	UKBB	UKFB	120000	121500	121100

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)

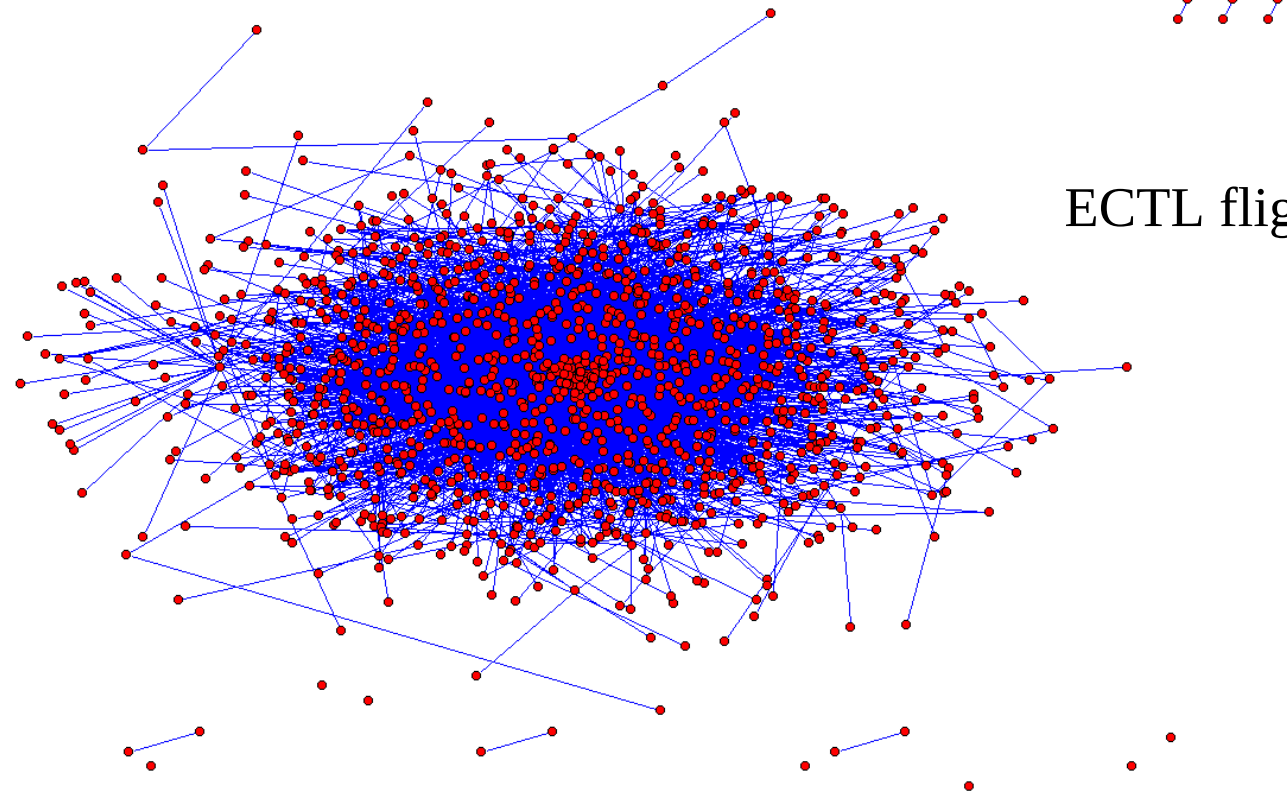
SOME PRELIMINARY RESULTS:
Network of airports

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

ROUTE network

nodes - airports

$\langle k \rangle$ average degree - number of connections with other airports

$\langle s \rangle$ average strength - number of flights in each airport

$\langle l \rangle$ average path length - minimum number of flights that connects two airports

	day	nodes	flights	links	$\langle s \rangle$	$\langle k \rangle$	$\max(s)$	$\max(k)$	largest cluster	$\langle \ell \rangle_s$	$\langle \ell \rangle_k$
W	June 1	1375	31224	8498	45.4	12.4	1452	221	1356	3.34	3.29
Th	June 2	1296	27660	7757	42.6	12.0	1426	226	1264	3.33	3.27
F	June 3	1291	28280	8039	43.8	12.5	1384	219	1256	3.31	3.23
Sa	June 4	1153	25289	7642	43.8	13.3	1343	245	1126	3.17	3.15
Su	June 5	1188	28175	8104	47.4	13.6	1451	232	1167	3.14	3.12
M	June 6	1342	30435	8243	45.3	12.3	1494	221	1327	3.35	3.29
T	June 7	1348	30397	8131	45.1	12.1	1494	220	1313	3.34	3.28

TABLE I

FULL NETWORK. ℓ IS THE PATH LENGTH. FOR A DETAILED EXPLANATION, SEE TEXT.

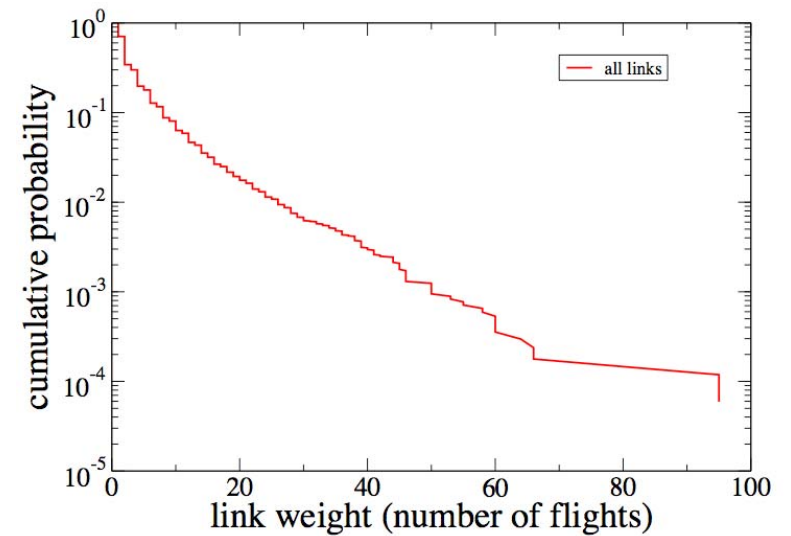
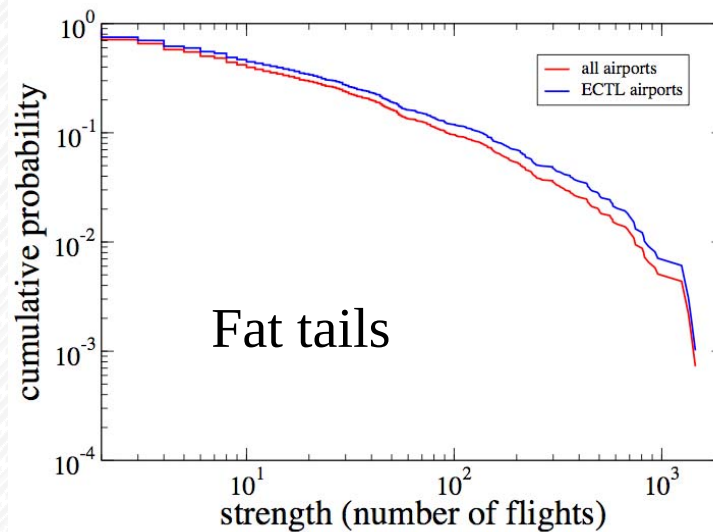
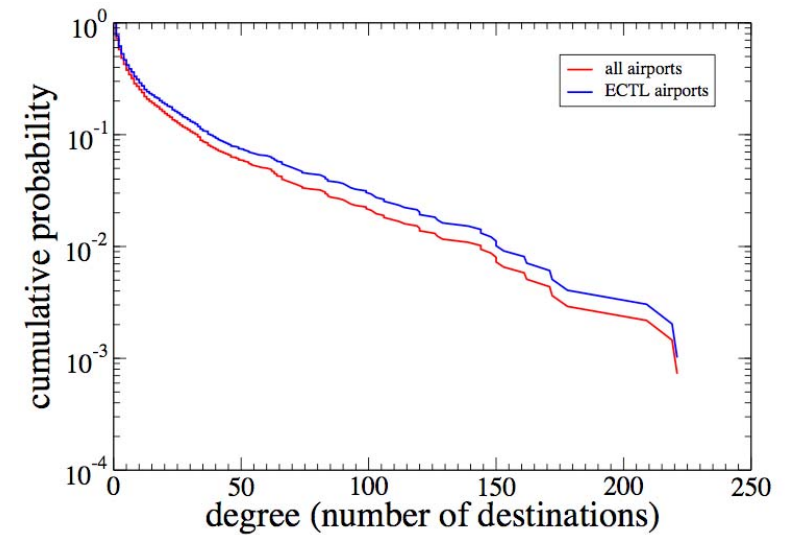
highly connected system

flight
network

SOME PRELIMINARY RESULTS:
Network of Airports
ROUTE NETWORK

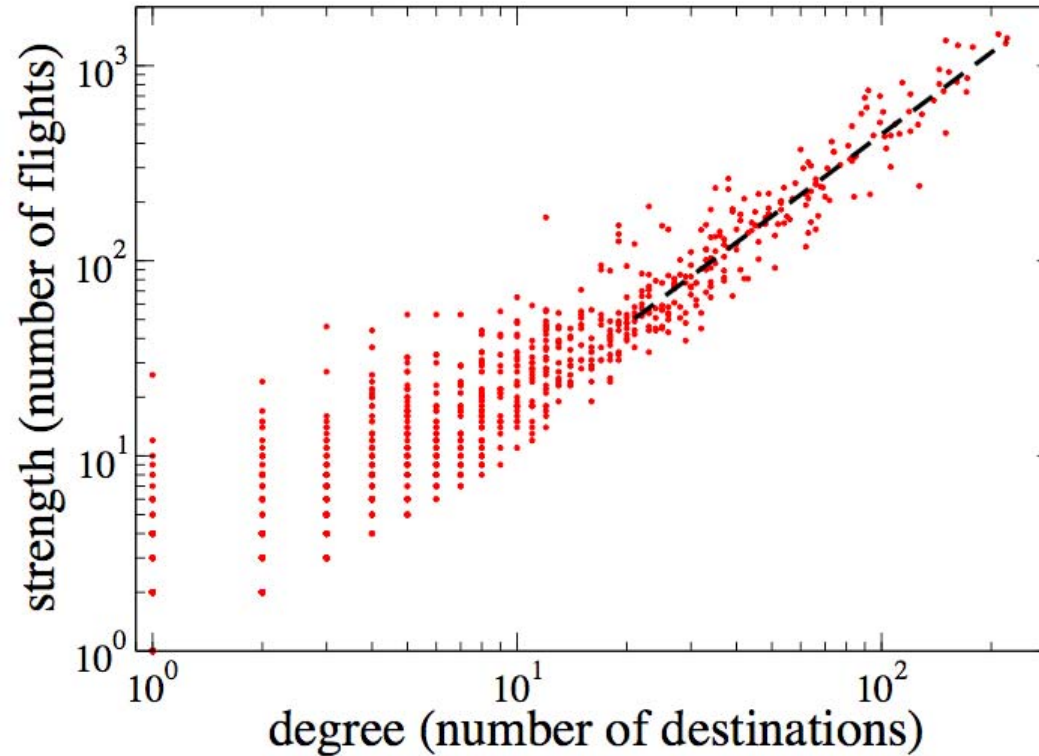
ROUTE network:
some metrics
distributions

(01/06/2011)



ROUTE network:

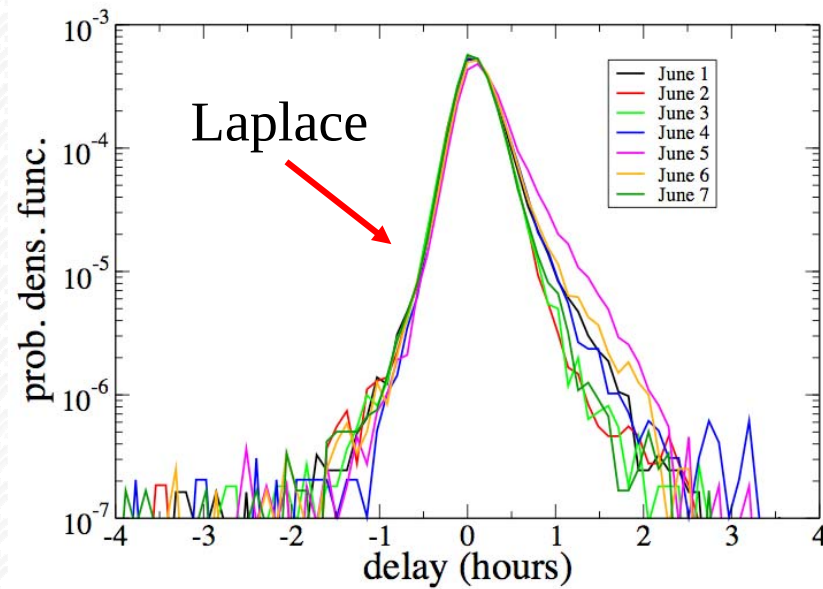
Fit: 1.39



SOME PRELIMINARY RESULTS:
Network of Airports
ROUTE NETWORK

If the number of destinations becomes twice, than the number of flights increases by a factor $2^{1.39}=2.45$

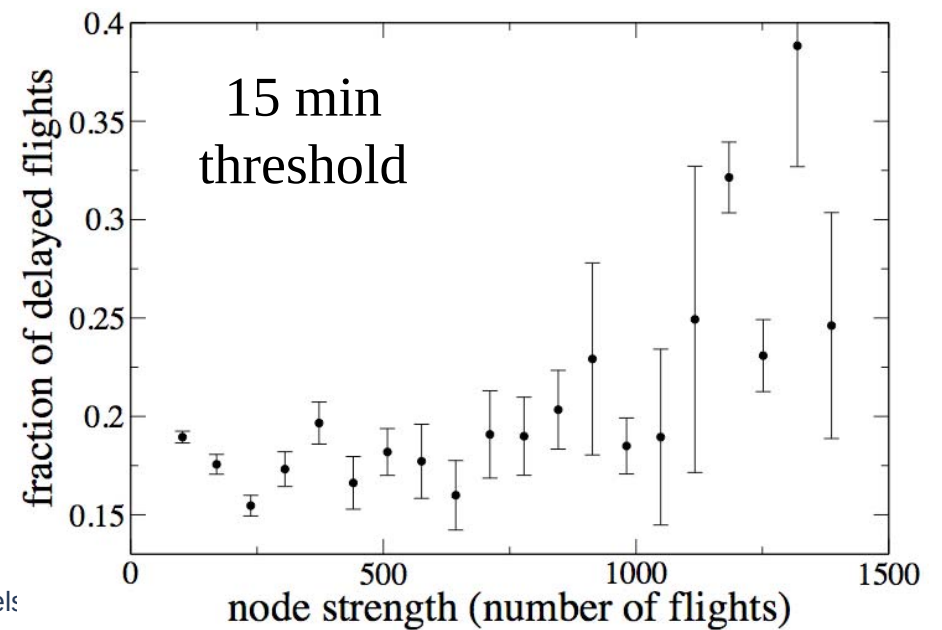
SOME PRELIMINARY RESULTS:
delay



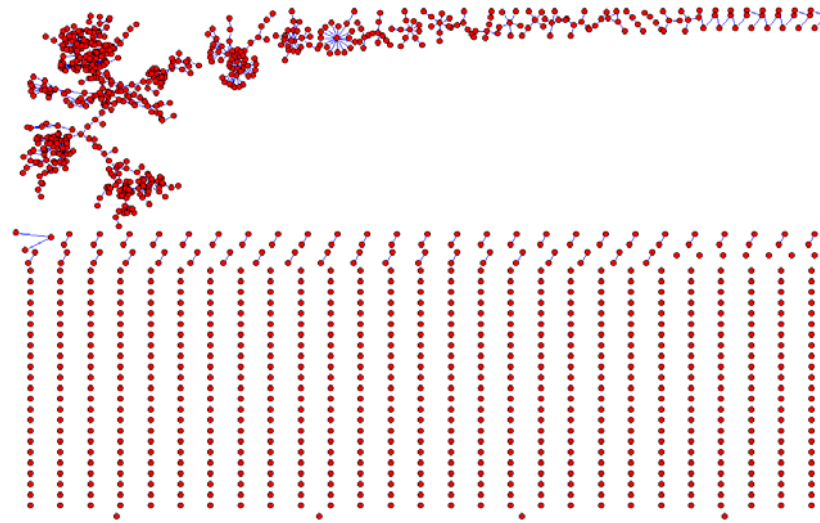
Negative
delays

7 days are pooled together

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



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

FUTURE WORK 1: statistically validated networks - FDR

OPEN ACCESS Freely available online



Statistically Validated Networks in Bipartite Complex Systems

Michele Tumminello^{1,2}, Salvatore Micciché³, Fabrizio Lillo^{2,3,4}, Jyrki Piilo⁵, Rosario N. Mantegna^{2*}

¹ Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, ² Dipartimento di Fisica, Università di Palermo, Palermo, Italy, ³ Santa Fe Institute, Santa Fe, New Mexico, United States of America, ⁴ Scuola Normale Superiore di Pisa, Pisa, Italy, ⁵ Department of Physics and Astronomy, Turku Centre for Quantum Physics, University of Turku, Turku, Finland

Abstract

Many complex systems present an intrinsic bipartite structure where elements of one set link to elements of the second set. In these complex systems, such as the system of actors and movies, elements of one set are qualitatively different than elements of the other set. The properties of these complex systems are typically investigated by constructing and analyzing a projected network on one of the two sets (for example the actor network or the movie network). Complex systems are often very heterogeneous in the number of relationships that the elements of one set establish with the elements of the other set, and this heterogeneity makes it very difficult to discriminate links of the projected network that are just reflecting system's heterogeneity from links relevant to unveil the properties of the system. Here we introduce an unsupervised method to statistically validate each link of a projected network against a null hypothesis that takes into account system heterogeneity. We apply the method to a biological, an economic and a social complex system. The method we propose is able to detect network structures which are very informative about the organization and specialization of the investigated systems, and identifies those relationships between elements of the projected network that cannot be explained simply by system heterogeneity. We also show that our method applies to bipartite systems in which different relationships might have different qualitative nature, generating statistically validated networks in which such difference is preserved.

Citation: Tumminello M, Micciché S, Lillo F, Piilo J, Mantegna RN (2011) Statistically Validated Networks in Bipartite Complex Systems. PLoS ONE 6(3): e17994. doi:10.1371/journal.pone.0017994

Editor: Eshel Ben-Jacob, Tel Aviv University, Israel

Received: December 22, 2010; **Accepted:** February 17, 2011; **Published:** March 31, 2011

Copyright: © 2011 Tumminello et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

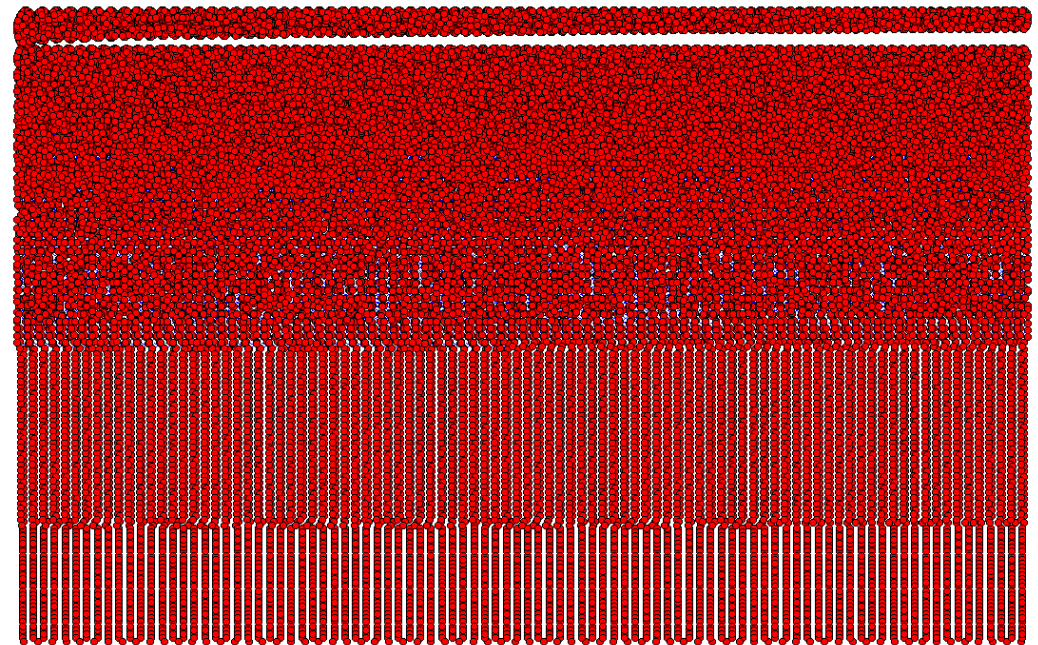
Funding: J.P. acknowledges financial support by the Magnus Ehrnrooth Foundation and the Vilho, Yrjö, and Kalle Väisälä Foundation, F.I., S.M. and R.N.M. acknowledge partial support from the Complex World Network funded by EUROCONTROL under the SESAR Work Package 6 framework, Contract Ref. 10-220210-C3. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: r.mantegna@gmail.com

The network includes
2_fdr: 28395/31224 nodes, 148509 links
2_fdr: 5891 connected components
2_fdr: 95 nodes in largest

FDR-network of flights



Bipartite system of segments and flights

FUTURE WORK 2: bipartite
system of segments and flights

Segment 1	Flight 1
Segment 1	Flight 2
Segment 1	Flight 3
Segment 1	Flight 4
Segment 1	Flight 5
Segment 2	Flight 2
Segment 2	Flight 3
Segment 2	Flight 6
....	

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.

FUTURE WORK 3: spreading of delays over network

The role of the airline transportation network in the prediction and predictability of global epidemics

Vittoria Colizza*, Alain Barrat[†], Marc Barthélemy^{*,†}, and Alessandro Vespignani[‡]

^{*}School of Informatics and Center for Biosciences, Indiana University, Bloomington, IN 47401; and [†]Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8623, Université Paris-Sud, Batiment 210, F-91405 Orsay, France

Communicated by Giorgio Parisi, University of Rome, Rome, Italy, December 8, 2005 (received for review June 18, 2005)

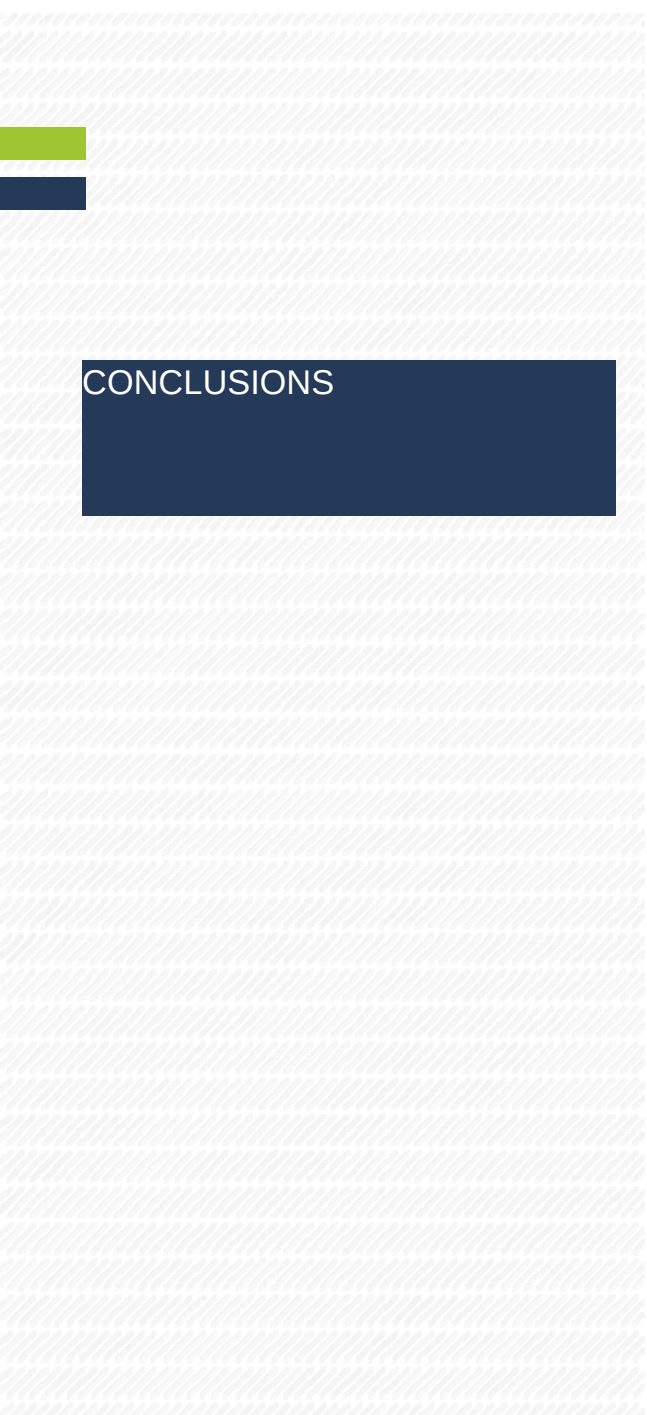
The systematic study of large-scale networks has unveiled the ubiquitous presence of connectivity patterns characterized by large-scale heterogeneities and unbounded statistical fluctuations. These features affect dramatically the behavior of the diffusion processes occurring on networks, determining the ensuing statistical properties of their evolution pattern and dynamics. In this article, we present a stochastic computational framework for the forecast of global epidemics that considers the complete worldwide air travel infrastructure complemented with census population data. We address two basic issues in global epidemic modeling: (i) we study the role of the large scale properties of the airline transportation network in determining the global diffusion pattern of emerging diseases; and (ii) we evaluate the reliability of forecasts and outbreak scenarios with respect to the intrinsic stochasticity of disease transmission and traffic flows. To address these issues we define a set of quantitative measures able to characterize the level of heterogeneity and predictability of the epidemic pattern. These measures may be used for the analysis of containment policies and epidemic risk assessment.

model including the full International Air Transport Association (www.iata.org) database, aiming at a detailed study of the interplay among the network structure and the stochastic features of the infection dynamics in defining the global spreading of epidemics. In particular, whereas previous studies have generally been focused in the *a posteriori* analysis of real case studies of global epidemics, the large-scale modeling presented here allows us to address more basic theoretical issues such as the statistical properties of the epidemic pattern and the effect on it of the complex architecture of the underlying transportation network. Finally, such a detailed level of description allows for the quantitative assessment of the reliability of the obtained forecast with respect to the stochastic nature of the disease transmission and travel flows, the outbreak initial conditions, and the network structure.

Results and Discussion

The Air-Transportation-Network Heterogeneity. The International Air Transport Association database contains the world list of airport pairs connected by direct flights and the number of available seats on any given connection for the year 2002. The

complex systems | epidemiology | networks



A few issues about data are still open

CONCLUSIONS

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.

 **THANKS FOR YOUR ATTENTION**

elsa@dblue.it

salvatore.micciche@unipa.it



UNIVERSITÀ
DEGLI STUDI
DI PALERMO



SCUOLA
NORMALE
SUPERIORE
PISA

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