Characterization of Delay Propagation in the US Air-Transportation Network

Pablo Fleurquin, José J. Ramasco, and Víctor M. Eguíluz

Abstract

Complex networks provide a suitable framework to characterize air traffic. Previous works described the world air-transport network as a graph where direct flights are edges and commercial airports are vertices. In this work, we focus instead on the properties of flight delays in the US air-transportation network. We analyze flight performance data in 2010 and study the topological structure of the network as well as the aircraft rotation. The properties of flight delays, including the distribution of total delays, the dependence on the day of the week and the hour-by-hour evolution within each day, are characterized with special attention to flights accumulating delays longer than 12 hours. We find that the distributions are robust to changes in takeoff or landing operations, different moments of the year, or even different airports in the contiguous states. However, airports in remote areas (Hawaii, Alaska, Puerto Rico) can show peculiar distributions biased toward long delays. Additionally, we show that long-delayed flights have an important dependence on the destination airport.

Keywords

Delay propagation, complex networks, airport and airline performance, ATM

Introduction

The generation, propagation, and eventual amplification of flight delays involve a large number of interacting mechanisms. Such mechanisms

Pablo Fleurquin
Corresponding author
Instituto de Física Interdisciplinaria y Sistemas
Complejos IFISC (CSIC-UIB) Campus UIB, E07122
Palma de Mallorca, Spain, and Innaxis Foundation
& Research Institute, Madrid, Spain
Email: pfleurquin@ifisc.uib-csic.es

José J. Ramasco
Instituto de Física Interdisciplinaria y Sistemas
Complejos IFISC (CSIC-UIB) Campus UIB, E07122
Palma de Mallorca, Spain

Víctor M. Eguíluz
Instituto de Física Interdisciplinaria y Sistemas
Complejos IFISC (CSIC-UIB) Campus UIB, E07122
Palma de Mallorca, Spain

can be classified as internal or external to the air-traffic system. The basic internal mechanisms include aircraft rotations (the different flight legs that comprise an aircraft itinerary), airport operations, passengers' connections, and crew rotation. In addition, external factors, such as weather perturbations or security threats, disturb the system performance and contribute to a high level of system-wide congestion. The intricacy of the interactions among all these elements calls for an analysis of flight delays under the scope of Complex Systems theory. Complexity is concerned with the emergence of collective behavior from the microscopic interaction of the system elements. Several tools have been developed to tackle complexity. Here we use Complex Networks theory and take a system-wide perspective to broaden the understanding of delay propagation. A network is a mathematical abstraction that represents systems of interacting entities as vertices (nodes) connected by edges (links) (see, for instance, Bocaletti et al. 2006, Newman 2010, or Barrat, Barthélemy, and Vespignani 2012 for recent reviews). Given the natural networked structure of the air-traffic system, we analyze the air-transport network formed by nodes representing airports and edges direct flights between them. The nature of such networks is highly dynamical since a different instance exists at every moment in time.

In this work we are interested in characterizing delays and how they may be transferred and amplified by subsequent operations, the so-called reactionary delays. Naturally reactionary delays spread across the network, so an understanding of the topological features of the air-transportation network, the properties of aircraft rotations, and the statistical features of flight delays is of great significance for subsequent modeling efforts (Fleurquin, Ramasco, and Eguíluz 2013a).

The remainder of the article is organized as follows. The next section provides a background review of the literature on complex networks, focusing on air transportation. Then, the used database is described, followed by the presentation of results on the characterization of the US air-transportation network, flight trajectories, and flight delays. Finally, we summarize our findings and point to further research questions.

Background

The use of network analysis to characterize complex systems has become widespread in the last two decades. The potential of graphs for describing social systems was pointed out almost a century ago (see Freeman 2004 for a review). However, the generalization of these concepts and tools had to wait much longer until the seminal works by Watts and Strogatz (1998), and Barabási and Albert (1999). Ever since, complex

networks have been applied in a growing range of disciplines such as technology (Huberman et al. 1998), biology (Jeong et al. 2001), and economy (Mantegna and Stanley 2007).

The application of network theory to air transportation has a much shorter history, for which the first results were published in 2004 and 2005. The world air-transportation network is described as a graph formed with the passenger commercial airports as vertices and the direct flights between airports as edges (Barrat, Pastor-Satorras, and Vespignani 2004; Guimera, Sales-Pardo, and Amaral 2005), with a weight corresponding to the number of seats available in the connection. The main source of this database is the International Air Transport Association (IATA; http://www .iata.org), while some other studies have presented data from the US Bureau of Transport Statistics (BTS) or from OAG (http://www.oag.com). The initial work by Barrat, Pastor-Satorras, and Vespignani (2004) focused on the correlations between network topology and fluxes of passengers in finding a nonlinear relation between them: $w_{ij} = \left(k_i k_j \right)$ where w_{ij} is the number of seats available in the connection between airports i and j; k_i is the number of connections with other airports of airport I; and θ is a parameter whose value was estimated to be approximately ½. A second study by Guimera, Sales-Pardo, and Amaral (2005) included a network description and analyzed the degree (number of connections per node) and node strength (sum over the weights of the connections of a node) distributions, degree-degree correlations, density of triangles, and so on. The world air-transportation network was analyzed later with graph clustering techniques (Sales-Pardo et al. 2007) to classify airports according to their connectivity patterns. The seasonal evolution of the connectivity patterns in the US airports networks have also been investigated in Gautreau, Barrat, and Barthelemy (2009) and Pan and Saramaki (2011). The authors characterize along the year how the network connectivity varies, with more routes available in summer, as well as how the passenger fluxes modify. Recently, information on human mobility through the air-transportation network has also been used to model and forecast the propagation pathways of infectious diseases transmitted by contact such as influenza (Balcan, Colizza, et al. 2009; Balcan, Hu, et al. 2009).

Within the air-traffic management (ATM) community, even if reactionary delays can have a great impact on air-traffic performance (EUROCONTROL 2008–2011; ICCSAI Fact Books 2007–2011; Joint Economic Committee 2008), the research effort to understand delay propagation has been scarce so far, and mostly limited to a descriptive work (Ahmadbeygi et al. 2008; Beatty et al. 1999; Schaefer et al. 2001). A good review of previous

work on delay propagation can be found in Belobaba, Odoni, and Barnhart 2009 and Jetzki 2009. Some research efforts have begun to apply network theory (Bonnefoy and Hansman 2007; Wuellner, Roy, and D'Souza 2010) in combination with stochastic modeling (Janić 2005; Rosenberger et al. 2002) to the modeling of delay propagation (Bonnefoy and Hansman 2005; EPISODE 3 2009; Pyrgiotis, Malone, and Odoni 2013).

Data and Method

Data was obtained from the Airline On-Time Performance Data available at the US Bureau of Transportation Statistics webpage (http://www.bts.gov). This database provides information such as schedule and actual departure and arrival times, departure and arrival delays, origin and destination airports, taxi-in and taxi-out times, airline ID, tail number and flight date. Air carriers that exceed 1 percent of the total domestic scheduled-service passenger revenue report on-time data and the causes of delay.

We restricted our analysis to domestic flights conducted in the year 2010. Despite the fact that these data are two years old, no major changes concerning on-time performance has occurred since then. For the year 2010, 18 air carriers filed on-time performance data that represents a total of 6,450,129 flights from 305 airports. From this database 1.75 percent of the flights were canceled and 0.2 percent diverted. All scheduled domestic flights for the year 2010 (not only those from On-Time Performance Data) total 8,687,800 (Bureau of Transportation Statistics 2011); therefore, the data used represent 74 percent of all scheduled flights in 2010.

Results

Characterizing the United States Air-Transportation Network

The resulting air-transportation network is composed of 305 nodes denoting airports, and 2,318 edges accounting for direct connections between them (fig. 1). Airports are sized according to the logarithm of their average delay per flight. Even though the network is not completely bidirectional, for instance, there can be flights from A to B but not from B to A, most connections bear flights in the two directions. For example, we find that if we build daily networks with the flight information, 98 percent of the overall connections are bidirectional. Furthermore, the lowest percentage of bidirectional links measured in a daily network is 92 percent. Small airports are responsible for these minor anomalies. To simplify the analysis we symmetrized the network.
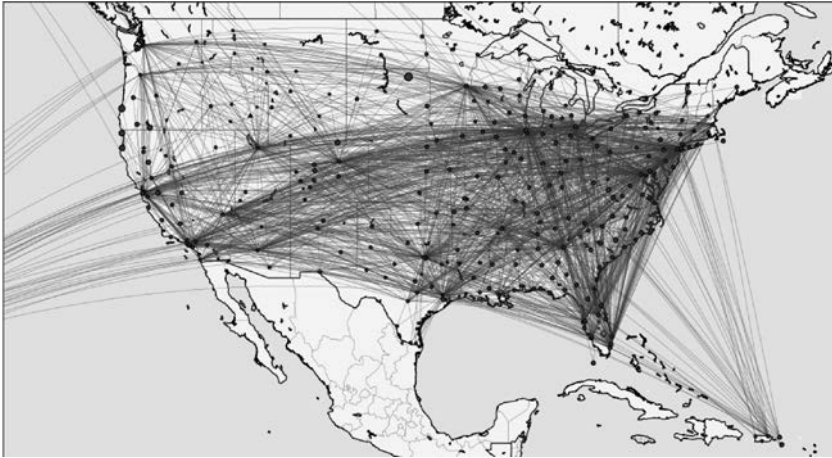
Figure 1 US Air Transport Network in 2010

As in the previous works, we define the degree of an airport as the number of different connections (airports of origin or destination of flights connecting with it). We can then calculate a degree distribution, taking into account the degrees or the number of flights of the airports across the network, and integrate it to obtain a cumulative distribution $F_X(x)$. For each value of x, the corresponding cumulative distribution tells us which fraction of airports with degree (number of flights) is lower than or equal to x. In figure 2, we show the complementary cumulative distribution of the degree and of the number of flights $1-F_X(x)$. Both distributions are wide and demonstrate the heterogeneities present in the network. Some few airports are large hubs with a large number of connections and flights, while most others have low traffic. These topological characteristics are well known for this network, but still are relevant for the dynamics of delay propagation.

Table 1 shows the ranking of the top-10 airports based on the number of different destinations (degree) and displays also the number of flights. The largest hub in the network is Atlanta International Airport (ATL) with 159 direct connections and the average degree of the whole network of 15.2.

Flight Trajectories

An important ingredient to characterize the propagation of reactionary delays is the rotation of the aircrafts. The database contains the tail number of the planes, which allows us to track their movements throughout the day. In figure 3, we show the percentage of aircrafts taking a certain number of
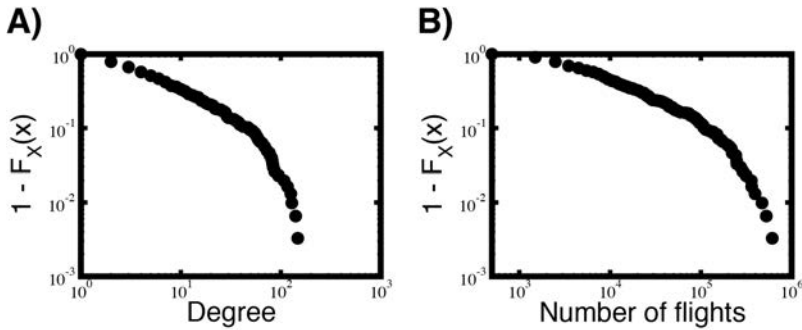
Figure 2 Complementary Cumulative Distribution of the Degree

Table 1/Top-10 Airports Ranked according to the Number of Connections to Other Airports

Airport	Number of Flights	Degree
ATL: Hartsfield-Jackson Atlanta International Airport	809,869	159
ORD: Chicago O'Hare International Airport	608,981	147
DFW: Dallas/Fort Worth Interntional Airport	524,206	140
DTW: Detroit Metropolitan Wayne County Airport	314,369	128
DEN: Denver International Airport	470,592	125
MSP: Minneapolis-Saint Paul International Airport	246,416	116
IAH: George Bush Intercontinental Airport	362,562	107
SLC: Salt Lake City International Airport	246,245	94
MEM: Memphis International Airport	152,730	86
MCO: Orlando International Airport	241,851	83

leaps per day. It can be seen that 80 percent of trajectories are composed of a number of leaps between two and seven. Very few planes do longer rotations due to the constraint of daily time periods and the duration of the flights.

Within a day, some of the aircraft trajectories are *closed walks*, that is, a sequence of airports starting and ending at the same airport, but most of the aircraft trajectories do not close at the end of the day. In figure 4 we show the percentage of closed walks per day during 2010. We can conclude that these trajectories are a small percentage with respect to the total number of aircraft rotations. This finding does not mean that the trajectories will not close taking into account the longer periods of time (weeks, months, or years).
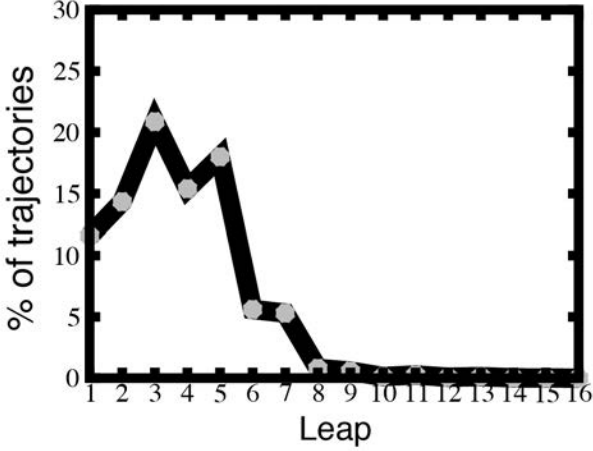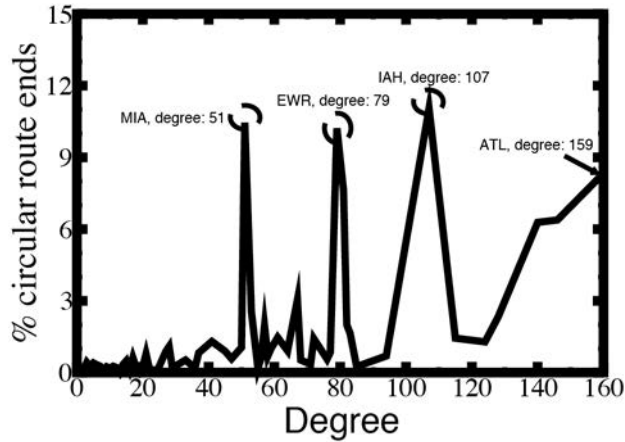
Figure 3 Percentage of Daily Aircraft Trajectories with Given Number of Leaps
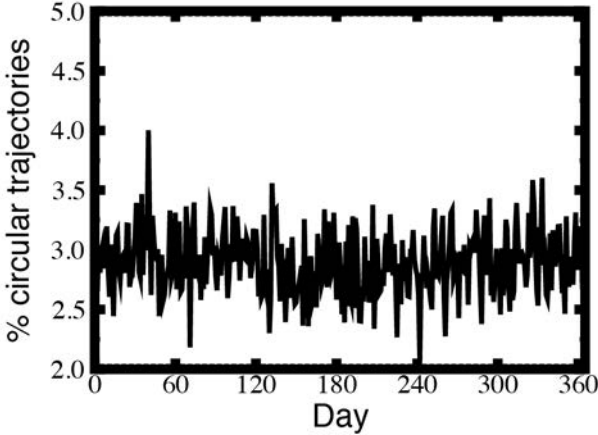
Figure 4 Percentage of Closed Walks

Regarding the previous result, another way of classifying the airports (besides connectivity) is by the fraction of closed walks that starts in each airport. These airports are not necessarily the ones with highest degree (see fig. 5). Assuming that the airline hubs (airlines' centers of operations) are those airports with a larger percentage of closed rotations, we can conclude that the network hubs (nodes with highest degree) do not always coincide with the airlines hubs.

Figure 5 Percentage of Daily Aircraft Trajectories Ending at an Airport as a Function of the Airport Degree

Flight Delay Characterization

We have described the topology of the network and the rotation of the flights. The next step is to focus on the real data regarding flight delays. We plot in figure 6 the complementary cumulative distribution of departure and arrival delays for all flights of 2010, $1–F_x(x)$ First, we notice that just like the degree and flight distribution, the delay distribution is broad with a slight hump at values of the delay around and larger than 700 minutes. Second, we find that there is no significant difference for both types of delays (arrival and departure delays), the day of the week or the season of the year (fig. 7). The cumulative distribution for different airports (fig. 8) shows a broad variety of behaviors. Displayed in the figure are a remote airport from the mainland like Honolulu International Airport (HNL), and two continental hubs, namely Dallas/Fort Worth International Airport (DFW) and Denver International Airport (DEN). We can see that, unlike HNL, DFW and DEN still show a slight hump in the distribution. On the other hand, Honolulu displays a broader distribution. This is probably due to the longer duration of the flights with destination or origin in HNL that allows for an easier absorption of short delays. The delays in the islands can be, therefore, much larger than those in the continent, and as a consequence the distribution becomes more skewed.

In order to understand the nature of the hump in the delay distributions, we extract the flights with departure delay above 12 hours and
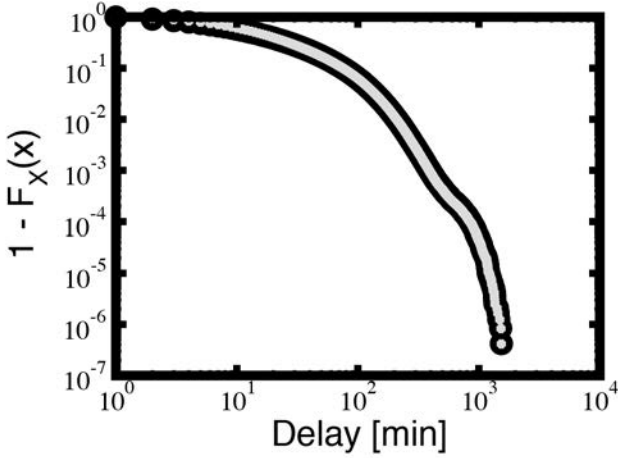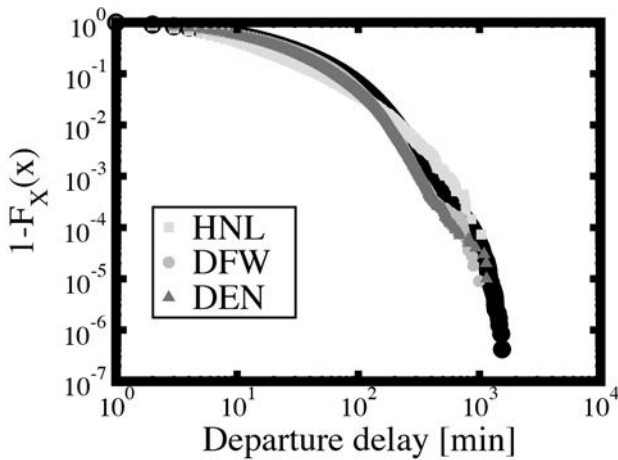
Figure 6 Complementary Cumulative Distribution Function of Departure and Arrival Delays in 2010
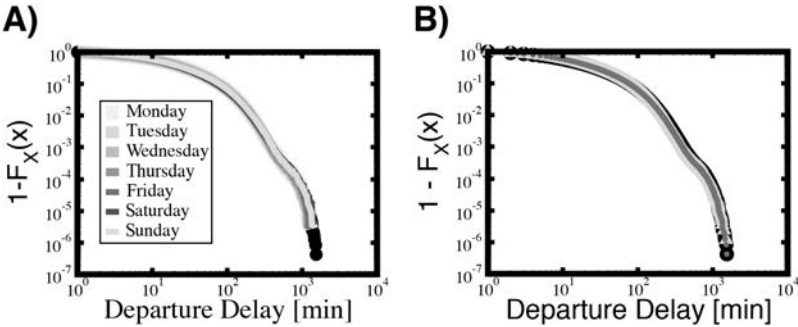

Figure 7 Complementary Cumulative Distribution Function of Departure Delays in 2010

compare them with all the flights of 2010. Plotting the departure delay as a function of the scheduled departure time we can distinguish how flights with delay greater than 12 hours are more abundant than the baseline at the beginning and at the end of the day (see fig. 9a). The opposite behavior can be observed for flights with departure delay below 12 hours, which show an almost flat delay distribution. Regarding this point, we plotted the delay distribution for flights with different scheduled departure times in figure 9b. The hump becomes more evident in the distribution of flights departing between 00am and 5am, and between 1pm and 11:59pm (local times) indicating a relatively higher abundance of long

Figure 8 Complementary Cumulative Distribution of the Departure Delays in 2010 (Black Circles), and Single Airports
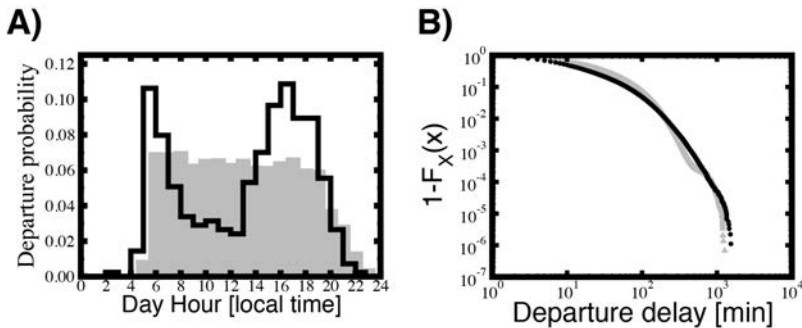

Figure 9 Fraction of Departures as a Function of the Scheduled Departure Hour

delayed flights. Note that even so, the fraction of delay flights is small compared with the total.

Another feature of long-delayed flights is their strong dependence on the destination airport. In table 2, we compare the data for long-delayed flights with two sets of randomly selected flights: one among all flights (delayed or not) and the other only with delayed flights. From the data 51 airports (16%) are the destinations of 414 delayed flights. If the 414 flights are randomly chosen, the number of destination airport increases to 120 (more than double the results from the real data) regardless of how the flights were chosen. This means that a bias exists toward a smaller set of destination airports. Note that the same phenomenon is not observed for the departure airports that are in the same range both in the data and in

Table 2/Flights with Departure Delay Higher than or Equal to 12 Hours

	Flights	Origin	Destination	Days	Tail	AR_ID
With problem	414	118	51	226	346	14
Total	6,341,340	305	305	365	5,081	18
Percentage	0.01%	38.00%	16.00%	62.00%	7.00%	77.00%
Randomly Chosen 414 Flights						
With problem	414	114	120	248	392	18
Total	6,341,340	305	305	365	5,081	18
Percentage	0.01%	38.00%	39.00%	67.00%	8.00%	100.00%
Randomly Chosen 414 Flights Delayed						
With problem	414	112	120	246	383	18
Total	6,341,340	305	305	365	5,081	18
Percentage	0.01%	36.00%	39.00%	67.00%	7.00%	100.00%
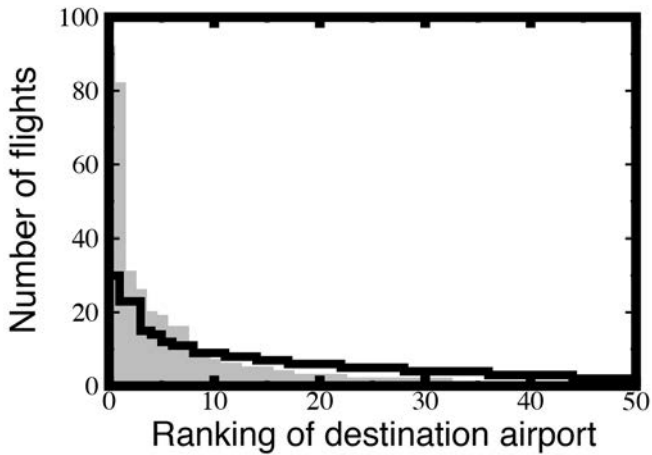
Figure 10 Ranking of the Number of Flights Delayed 12 hours or Longer for the 51 Destination Airport from the Data (Blue Bars) and the Randomly Selected Airports (Red Line)

the randomly selected flights. Other variables such as days, tail-number, or air carriers remain the same. In figure 10, we plotted the number of flights with long delays versus the ranking of destination airport with respect to the number of long-delayed flights. The data correspond to the gray bars, while the randomly selected set of flights is the black curve. In the data,

the first eight airports are destination of 75 percent of the long-delayed flights, while in the randomly selected set the first eight airports total only 52 percent.

The significance of the destination airports could be related to Ground Delay Program (GDP) instituted by the US Federal Aviation Administration (FAA; see http://www.fly.faa.gov/Products/AIS_ORIGINAL/shortmessage .html). This program is implemented to control air-traffic volume to airports where the estimated demand is expected to surpass the airport arrival rate. When a GDP is issued, flights destined to the affected airport are not permitted to depart before their controlled departure time.

Conclusion

In summary, we have analyzed the characteristics of the US air-transportation network with a focus on flight delays. The air-transportation network is built by connecting pairs of airports if they have a direct flight. We studied the network topological properties such as the distribution of the number of flights or the number of connections per airport. These features show the broad heterogeneity of the air-transport network in accordance with previous works. In addition to the topology, we consider the properties of the aircraft rotation throughout the day and the characteristics of the delays. The aircraft rotation shows a complicated and highly heterogeneous profile. Some aircrafts itineraries are essentially round trips while others do not close in a simple periodic way. The heterogeneity of the rotation procedures can play an important role in the development and propagation of delays.

Regarding the delays, we show that the delay distributions show long decays both for arrival and departure delays, irrespective of the day of the week and season. Long tails are usually indicative of the complex nature of the mechanisms contributing to the propagation of delays. In this case, the system is not necessarily working under critical conditions but the combined action of several factors, such as connecting passengers or crew, a predetermined schedule, and the geographical distance of the airports, can contribute to reach a similar system state at a global level. Whether the air-transport network is a system at criticality is an open question that deserves further research. We study also the properties of the flights with a delay longer than 12 hours, showing a relative concentration of long-delayed flights early in the morning or late in the afternoon. The destination airport seems to be a key player for the surge of flights with long delay.

These results are relevant in order to better characterize flight delays from a statistical perspective. Subsequent efforts aimed at modeling delay spreading in the air-transport networks, such as the recent works by

Fleurquin, Ramasco, and Eguíluz (2013a, 2013b), should have into account the statistical patterns described here both in the model development and validation.

Notes

References

Ahmadbeygi, S., A. Cohn, Y. Guan, and P. Belobaba. 2008. "Analysis of the Potential for Delay Propagation in Passenger Aviation Flight Networks." *Journal of Air Transport Management* 14:221–36.

Balcan, D., V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani. 2009. "Multiscale Mobility Networks and the Spatial Spreading of Infectious Diseases." *Proceedings of the National Academy of Sciences USA* 106:21484–89.

Balcan, D., H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani. 2009. "Seasonal Transmission Potential and Activity Peaks of the New Influenza A(H1N1): A Monte Carlo Likelihood Analysis Based on Human Mobility." *BMC Medicine* 7:45–57.

Barabási, A.-L., and R. Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286:509–12.

Barrat, A., M. Barthélemy, and A. Vespignani. 2012. *Dynamical Processes on Complex Networks.* New York: Cambridge University Press.

Barrat, A., R. Pastor-Satorras, and A. Vespignani. 2004. "The Architecture of Complex Weighted Networks." *Proceedings of the National Academy of Sciences USA* 101:3747–52.

Beatty, R., R. Hsu, L. Berry, and J. Rome. 1999. "Preliminary Evaluation of Flight Delay Propagation through an Airline Schedule." *Air Traffic Control Quarterly* 7:259–70.

Belobaba, P., A. Odoni, and C. Barnhart. 2009. *The Global Airline Industry.* Chichester, UK: John Wiley and Sons.

Bocaletti, S., V. Latora, Y. Moreno, M. Chavez, and D.-U., Hwang. 2006. "Complex Networks: Structure and Dynamics." *Physics Reports* 424:175–308.

Bonnefoy, P., and R. J. Hansman. 2005. "Emergence of Secondary Airports and Dynamics of Regional Airport System in the United States." Report ICAT 2005-02.

———. 2007. "Scalability and Evolutionary Dynamics of Air Transportation Networks in the United States." Proceeding of the 7th AIAA Aviation Technology and

Operations Conference. Belfast, Northern Ireland, September 18–20, 2007, http://arc.aiaa.org/doi/abs/10.2514/6.2007-7773.

EPISODE 3. 2009. "Report on Macro Modelling on Global Performances at Network-Wide Level 2009." Deliverable D3.3.5-02. http://www.episode3.aero/.

EUROCONTROL. 2009–2012. "EUROCONTROL Annual Reports 2008–2011." https://www.eurocontrol.int/publications.

Fleurquin, P., J. J. Ramasco, and V. M. Eguíluz. 2013a. "Systemic Delay Propagation in the US Airport Network." *Scientific Reports* 3:1159.

———. 2013b. "Data-driven Modeling of Systemic Delay Propagation under Severe Meteorological Conditions." Tenth USA/Europe Air Traffic Management R&D Seminar, Chicago, IL.

Freeman, L. C. 2004. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver: Empirical Press.

Gautreau, A., A. Barrat, and M. Barthelemy. 2009. "Microdynamics in Stationary Complex Networks." *Proceedings of the National Academy of Sciences USA* 106:8847–52.

Guimera, R., M. Sales-Pardo, and L. A. N. Amaral. 2005. "The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles." *Proceedings of the National Academy of Sciences USA* 102:7794–99.

Huberman, B. A., P. L. Pirolli, J. E. Pitkow, and R. M. Lukose. 1998. "Strong Regularities in World Wide Web Surfing." *Science* 280:95–97.

ICCSAI. 2007–2011. "Fact Books on Air Transport in Europe, 2007–2011." http://www.iccsai.eu/factbook.html.

Janić, M. 2005. "Modeling the Large Scale Disruptions of an Airline Network." *Journal of Transportation Engineering* 131:249–60.

Jeong, H., S. Mason, A. L. Barabasi, and Z. N. Oltvai. 2001. "Lethality and Centrality in Protein Networks." *Nature* 411:41–42.

Jetzki, M. 2009. "The Propagation of Air Transport Delays in Europe." PhD diss., Department of Airport and Air Transportation Research RWTH, Aachen University.

Joint Economic Committee (JEC), US Senate. 2008. "Your Flight Has Been Delayed Again: Flight Delays Cost Passengers, Airlines and the US Economy Billions." Report, May 22. http://www.jec.senate.gov/.

Mantegna, R., and E. H. Stanley. 2007. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press.

Newman, M. 2010. *Networks: An Introduction*. New York: Oxford University Press.

Pan, R. K., and J. Saramaki. 2011. "Path Lengths, Correlations, and Centrality in Temporal Networks." *Physical Review E* 84:016105.

Pyrgiotis, N., K. M. Malone, and A. Odoni. 2013. "Modeling Delay Propagation within an Airport Network." *Transportation Research C* 27:60–75.

Rosenberger, J., A. J. Schaefer, D. Goldsman, E. L. Johnson, A. J. Kleywegt, and G. L. Nemhauser. 2002. "A Stochastic Model of Airline Operations." *Transportation Science* 36:357–77.

Sales-Pardo, M., R. Guimera, A. A. Moreira, and L. A. N. Amaral. 2007. "Extracting the Hierarchical Organization of Complex Systems." *Proceedings of the National Academy of Sciences USA* 104:15224–29.

Schaefer, L., and D. Millner. 2001. "Flights Delay Propagation Analysis with the Detailed Policy Assessment Tool." *Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics* 2:1299–1303.

Watts, D. J., and S. H. Strogatz. 1998. "Collective Dynamics of 'Small-world' Networks." *Nature* 393:440–42.

Wuellner, D., S. Roy, and R. M. D'Souza. 2010. "Resilience and Rewiring of the Passenger Airline Networks in the United States." *Physical Review E* 82:056101.