Supplementary Information for
Systemic delay propagation in the US airport network

Pablo Fleurquin⋆,†, José J. Ramasco⋆ & Víctor M. Eguiluz⋆

⋆ Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB),
Palma de Mallorca, Spain.

† Innaxis Foundation & Research Institute, José Ortega y Gasset 20, Madrid, Spain.

In this document, we provide further information about the database, the model and the results discussed in the main text. We begin by describing the database used to construct the network, the flight's schedule and connectivity in Section 1. Then, in Section 2, we explain in detail the algorithm used to simulate the flight dynamics and the delay propagation across the network. And, finally, in Section 3, we consider further results of the model not included in the main text and study the stability of the model predictions when its parameters are changed. In particular, in Sec. 3.1 we verify the accuracy of the model predictions together with a sensitivity analysis to a variation in α. The sensitivity to changes in β is considered in Sec. 3.2. In section 3.3, we study the variability of the simulation results due to the model's internal stochasticity and further results of the model simulations for different magnitudes and days are shown in Sec. 3.4.

1   Database

1.1   Description

The database was obtained using the information available at the Bureau of Transport Statistics [1]. In particular, we used the Airline On-Time Performance Data, which is built with flight data provided by air carriers that exceed one percent of the annual national revenue for domestic scheduled service. The database comprehends $6,450,129$ scheduled domestic flights operated by 18 carriers connecting 305 different commercial airports. Considering all flights in 2010, not only those that report On-Time Performance Data, the number of scheduled domestic flights totalizes $8,687,800$ [2] and so our dataset includes information for 74% of the total. It is worth noting that this schedule is based on real events, which is not necessarily the original schedule in hands of the companies at the beginning of the day. If a flight gets canceled, diverted or even in milder situations the managers of an airline can introduce changes in the original schedule that we cannot trace back. However, given that these flights represent, respectively, the 0.20% and 1.75% of all flights in the database, one can expect these changes not to be of large magnitude.

Among the available data fields, we consider the following as the most relevant for our work: *Tail number, airline ID, airports of origin and destination, date of the flight, scheduled departure and arrival*

times, real departure and arrival times, and whether the flight was canceled or diverted. The tail number is a code that identifies the aircraft and that allows us to follow it along the daily plane rotation. Arrival and departure times (real or scheduled) considered refer to when the flight actually reaches or departs from the gate, we are not taking taxing or take-off and landing times as departure (arrival) times. We will exclude of the coming analysis diverted and canceled flights since they difficult the characterization of the delay propagation and are a small fraction of the total operations.
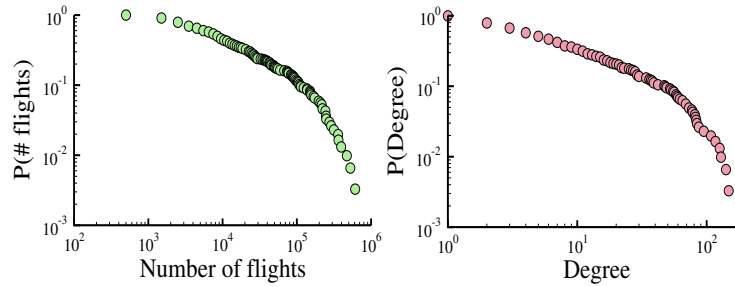

Figure S1: Cumulative distribution of the number of flights and number of different connections (degree) for the airports in 2010.

A network between airports can be built based on the data: Airports are the vertices and edges represent direct flights from one airport to another. Note that this is a directed graph and that depends on the time-scale of data aggregation. On an annual basis, the resulting US air-transportation network comprises 305 commercial airports and $2,318$ connections. Figure S1 depicts the complementary cumulative distribution of the number of flights and different connections for all the airports of the network in 2010. Both distributions confirm the presence of high heterogeneities in the airport network. The most active airports for 2010 are represented in Table S1, showing that the maximal degree corresponds to Atlanta International Airport (ATL) with 159 different connections. For the analysis of the clusters of congested airports, we will considered networks aggregated only during one day.

The average delay per delayed flights (those with positive delays) for 2010 is 29 minutes. This same value is used to define when an airport is considered congested in the main text but it can also be used to define days of operational problems or not. Those are respectively days whose average delay per delayed flight is over or below 29 minutes, respectively. Table S2 shows the ranking of the 20 best and worst days of the year according to their average delay for flights with positive delay.

The United States spans through several time zones. In order to unify criteria and simplify the analysis, we transform all the local times to the East Coast local time. Olson or tz database [3] is used to ensure an accurately timezones conversion from the respective local times in the database to the East Coast local time (EST in winter and EDT in summer time). Plotting the departure probability as a function of the scheduled departure in Figure S2, we can distinguish a zone that goes from 00am to 4am with almost no operations. We set thus the start of a new day in our analysis and simulations at 4am East Coast local time, that is, at 3am Central, 2am Mountain and 1am Pacific. In this way, the starting of the new day coincides with the low activity phase of air

Airport code	# edges	# flights
ATL	159	809, 869
ORD	147	608, 981
DFW	140	524, 206
DTW	128	314, 369
DEN	125	470, 592
MSP	116	246, 245
IAH	107	362, 562
SLC	94	246, 245
MEM	86	152, 730
MCO	83	241, 851

Table S1: Major airports according to their degree (number of different destinations).
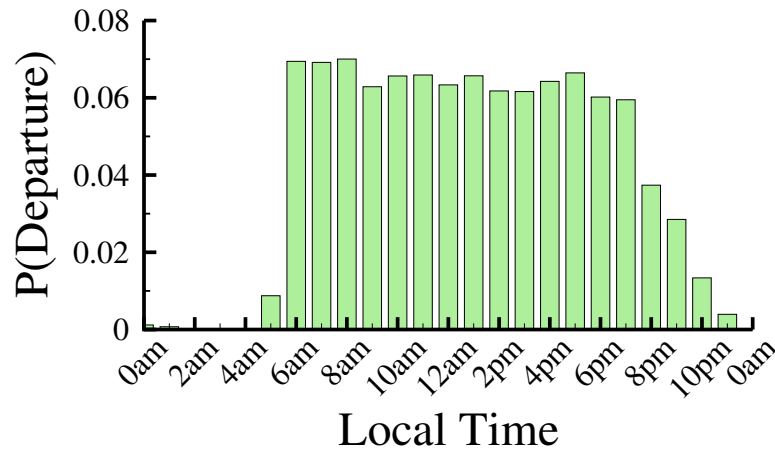
operations in most of the country.

Figure S2: Probability of flight departure as a function of the scheduled departure hour.

We also notice that for daily networks 98% of the edges are bidirectional on average, i.e., if there is a flight from A to B there is always a flight from B to A. Taking this into account, we symmetrized the network to simplify the cluster analysis.

Problematic days		Satisfactory days	
DATE	Average delay (mins.)	DATE	Average delay (mins.)
Oct, 27	54.3	Apr, 19	16.9
Mar, 12	53.0	Oct, 09	17.2
Dec, 12	51.9	Nov, 11	17.3
Jan, 24	49.8	Apr, 14	17.6
Feb, 24	49.1	Oct, 08	18.0
May, 31	46.8	Set, 11	18.4
May, 21	45.5	Apr, 15	18.4
May, 14	44.6	Oct, 13	18.5
Jun, 23	44.6	Apr, 17	18.5
Jul, 13	44.3	Nov, 10	18.8
Jun, 24	42.8	Nov, 09	18.9
Jul, 12	42.7	Mar, 06	19.1
Jan, 21	41.5	Oct, 12	19.2
Jul, 29	41.4	Mar, 17	19.3
Jun, 15	41.2	Feb, 28	19.5
Jun, 27	40.5	Oct, 16	19.5
Mar, 20	40.5	Apr, 13	19.5
Mar, 11	39.9	Nov, 26	19.5
Aug, 22	39.7	Set, 09	19.6
Jan, 25	39.5	Set, 20	19.7

Table S2: Ranking of the 20 worst/best days of the year 2010 according to their daily average delay for flights with positive delay.

1.2   Annual fraction of connecting passengers for each US commercial airport

Another key input for modeling the delay propagation over the network is the connection between flights. The previous database has no information regarding flight connectivity, neither for the crews nor the passengers. In order to at least approximate the heterogeneity of the airports in this sense, we used the T100 Domestic Market (US carriers) and the DB1B Ticket information downloaded from the BTS page [1]. These documents allow us to obtain an approximation of the annual fraction of connecting passengers for each airport. The information of T100 corresponds to the total number of passengers who have a flight departing from an airport regardless of their real point of origin ($Passengers_{T100}$). On the other hand, the database DB1B contains a 10% sample of the number of passengers whose itinerary originated in each given airport ($Passengers_{DB1B}$). So for each airport we can get an approximation of the annual fraction of connecting passengers as:

$$\frac{Passengers_{T100} - 10.Passengers_{DB1B}}{Passengers_{T100}} \tag{S1}$$

Although our model is based on flight not passenger connectivity, we assume that these ratios are related, which is always better than assuming arbitrary values, with α controlling the intensity

of such relation.

Airport code	Fraction of connecting passengers
ATL	0.81
ORD	0.72
DFW	0.75
DTW	0.69
DEN	0.71
MSP	0.68
IAH	0.75
SLC	0.73
MEM	0.81
MCO	0.69

Table S3: Fraction of connecting passengers for the top ten airports in degree.

2   Model description

To simulate the delay propagation, we developed an agent-based model that combines within the same framework queuing and a schedule based approach dynamics.

2.1   Overview

As stated in the main text, one of the purposes of this model is to understand how delays propagate and magnify considering internal operational factors and schedule. As it will be explained further below, "extrinsic" or primary delay is given at the initial steps of the simulation to the first flight of the day for some aircraft rotations, and then let this perturbation evolve multiplying or diminishing the delay according to the particular structure of the system. Concretely, the model dynamics will be based on three subprocesses which are: (i) aircraft rotation, (ii) flight connectivity and (iii) airport congestion. The last two are independent from each other, and can be turned on/off to explore the relevance of each subprocess in the delay propagation dynamics. Aircraft rotation, on the other hand, is intrinsic to the schedule and so we do not switch it off.

   We use one-minute intervals as the basic time step unit in the model and proceed in each simulation until the schedule of a selected day is completed (all flights had completed their itinerary). In most cases, this means slightly more than $1,440$ minutes. This time interval allows the simulation to execute actions at a realistic concurrent time-scale and is the finest level available in the data. As mentioned in the previous section, 4am East Coast local time is set as the starting point for airport operations and to begin the aircraft rotational sequences. By this selection, we ensure that most aircraft rotational sequences are sorted correctly and it is the natural choice considering the daylight time flow in the United States. Also, as mentioned before, to arrange the schedule in a real sequential order we converted time data from Local operation time to Eastern local time.

2.2 Hierarchy of Objects

2.2.1 Aircraft (tail-number)

The airplane is the primary fundamental agent of the simulation. The number of airplanes that participate in the simulation varies with the day considered, but it is around $4,000$. Each aircraft is unique and comes identified by their tail number. This code allows us to reconstruct the rotational sequence of the plane during the day. This sequence can be subdivided in individual flight legs or point-to-point flights.

2.2.2 Point-to-point flight

This is the basic schedule unit. It is the minimum package of information used as an input to relocate an aircraft from an origin to a destination airport, meeting the planned schedule. During their itinerary an aircraft can be in one of two flight phases: block-to-block or turn-around phase. The former is the time elapsed from the airport origin gate to the airport destination gate. The latter is defined as the time the aircraft remains parked at the airport gate (Figure S3).
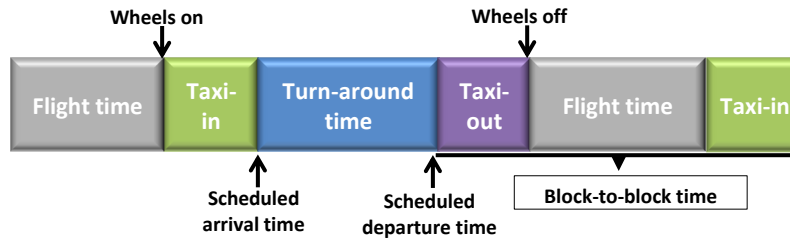

Figure S3: Turn-around and block-to-block time/phase definition.

Flights are characterized by a tail-number, origin airport, destination airport, schedule departure time ($T_{sch.d}$) and schedule arrival time ($T_{sch.a}$). Block-to-block time (T_b) between two airports is calculated as:

$$T_b = T_{sch.a}^j - T_{sch.d}^i, \tag{S2}$$

where j corresponds to destination airport and i to the origin one. Another issue worth noting in our model is that, in the block-to-block phase we do not allow for delay absorption or reduction. This could only be achieved in the turn-around phase by means of the difference between the actual arrival time of the previous flight leg and the scheduled departure time of the next flight leg.

2.2.3 Air carrier (airline id)

Air carriers are the second level unit in the model. Each aircraft has an airline associated via the airline code id. Only aircrafts having the same airline id are allowed to interact during the process of flight connectivity (see 2.3.2 for further details).

2.2.4 Airport

The airport is an intermediate-level entity located in space coordinates, where interactions among aircrafts take place. This interaction occurs indirectly through the schedule, flight connections or airport queues (see 2.3.3 for further details). Each airport is different from the others because of their planned capacity and the local aggregation of the schedule. Airports play the role of nodes in the transport network.

2.2.5 Clusters of congested airports

This is a high-level entity that represents interactions between airports. The clusters are formed by airports whose average (departure) delay per flight is higher or equal to 29 minutes and are linked by a direct connection (see 2.6 for further details). In most cases, we are interested in the largest cluster of the full day (or by hour in some cases). The size of a cluster is measured according to the number of airports that belong to it.

Figure S4 shows a representation of two clusters (Cluster A and B) constituted by airports whose average departure delay per flight in a certain time period is equal or larger than 29 minutes (red dots). Apart from this condition airports within these clusters are linked by a direct connection. In this case, cluster A correspond to the largest cluster in a certain time period according to the number of airports that form this cluster.
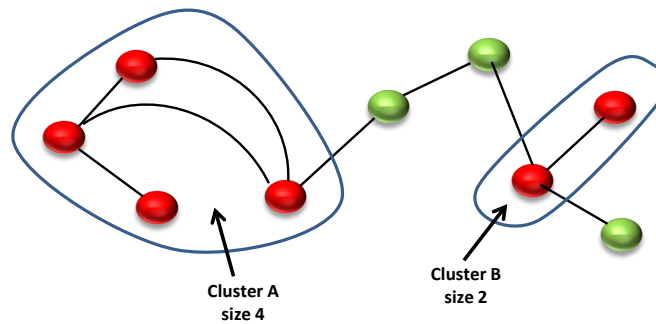

Figure S4: Red dots correspond to airports whose average departure delay per flight in a certain time period is > than 29 minutes. Green dots correspond to airports whose average departure delay per flight in a certain time period is leq than 29 minutes.

2.3 Subprocesses

2.3.1 Aircraft rotation

During a day, each aircraft has an itinerary to accomplish that in the vast majority of cases consists of two or more flight legs. Naturally, to complete a flight leg, the previous ones have to be fulfilled, e.g., it is not possible to depart from San Francisco to Honolulu if the airplane has not completed the previous leg from Atlanta to San Francisco. Besides this evident situation, if an aircraft arrives late (inbound delay) and the delay cannot be absorbed by the turn-around time it will depart late in the next flight leg (Figure S5). Usually, a buffer time is included in the turn-around phase to absorb this type of delay but this is already incorporated in the schedule obtained from the data.

Another feature of this subprocess, is that in the turn-around phase each aircraft, when arrived, has to comply with a minimum service time T_s, in the simulations set as 30 minutes. This service time includes operations such as refueling, passenger unboarding/boarding, luggage handling, safety inspection, etc.
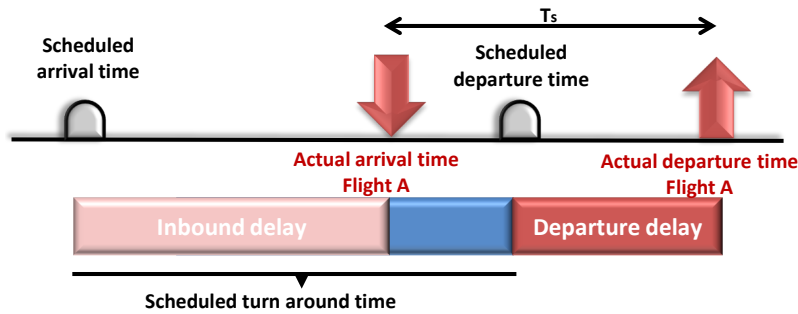

Figure S5: Aircraft rotation description.

2.3.2 Flight connectivity

In addition to rotational reactionary delay, the need to wait for load, connecting passengers and/or crew from another delayed airplane from the same fleet (airline id) may cause, as well, reactionary delay.

For each flight at a particular airport, connections from that airport are randomly chosen as follows. Firstly, we take a ΔT window prior to the scheduled departure time of the flight. Secondly, we distinguish possible connections of the same airline from other flights, that have a scheduled arrival time within the ΔT window (Flights B and D in the example of Figure S6). Finally, from these possible connections we select those with probability $\alpha *$ flight connectivity factor. The flight connectivity factor was defined in 1.2 and α is an effective parameter of control that allows to modify the strength of this effect in the model. For instance, $\alpha = 0$ means that there is no connection between flights with different tail number, while $\alpha = 1$ makes the fraction of connecting flights of the same airline equal to the fraction of connecting passengers in the given airport. In
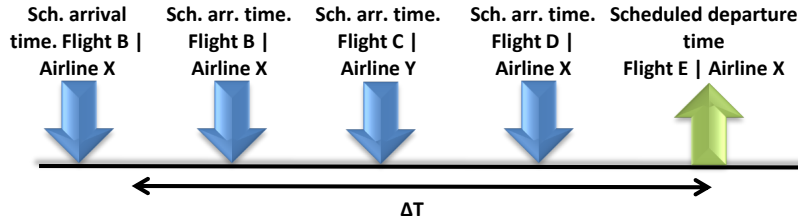
Figure S6: Possible connections within flights of the same airline.

the simulations, α is varied according to the case under study and ΔT is always taken to be 180 minutes (3 hours).

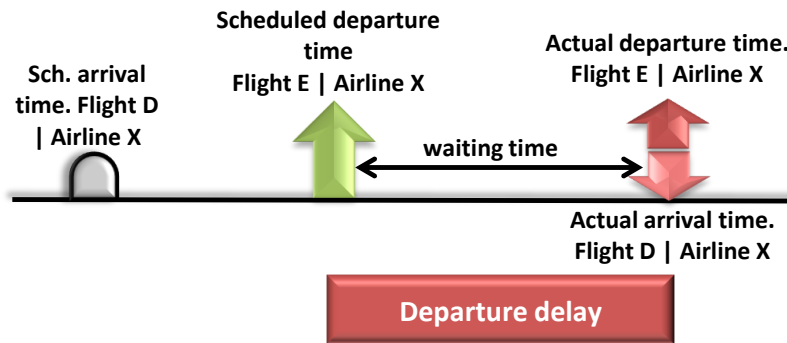Let us suppose that from the previous example Flight D was randomly selected.

Figure S7: Flight connectivity description.

By this subprocess an airplane is able to fly if and if only their connections have already arrived to the airport, if not it has to wait until this condition is satisfied (Figure S7). It is important to note that flight connectivity is the only source of stochasticity in the model due to a lack of knowledge about the real flight connections within the schedule.

2.3.3 Airport congestion

Because airports are entities with a finite capacity, the possibility of their congestion has to be introduced in the model. Interactions between aircrafts other than the ones defined by the schedule (flight connectivity and aircraft rotation) are in this way taken into account. This occurs indirectly through an airport's queue. That is to say that delays from airplanes of different airlines can delay others because they congest the airport. The delay spreading does not surge so easily as in the previous cases, it requires a cumulative effect of several delayed aircrafts to perturb the airport efficiency and once this condition is meet the delay spread to other aircrafts and affect other airlines.

We assume a "First in-First Served" queuing protocol that is the most widely used queue operation and simple to introduce in the model. In the simulations each airport will have a capacity that varies throughout the day according to the Scheduled Airport Arrival Rate ($SAAR$). This means that for every airport we measure the scheduled flights that arrive per hour and this is the nominal capacity for each hour of the day (Figure S8). Due to reactionary delays aircrafts may not arrive as planned and the Real Airport Arrival Rate ($RAAR$) will vary. Whenever $RAAR > SAAR$, a queue begins to form with the arriving aircrafts. Naturally, airplanes that are not in queue are being served and this service time lasts T_s (see 2.3.1 for further details). It should be noticed that once an aircraft starts to be served this process cannot be interrupted no matter how $SAAR$ varies. We define another effective control parameter β in order to modify the nominal capacity of the
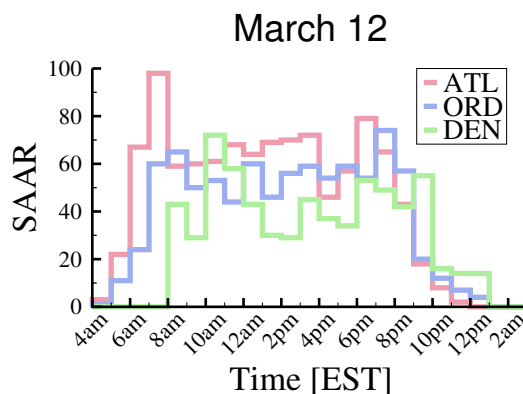

Figure S8: Example of SAAR for three major airports: Atlanta International Airport (ATL), O'Hare International Airport (ORD) and Denver International Airport (DEN).

airports. This parameter multiplies the $SAAR$ and in the simulations presented here affects all the airports in the same way. For instance, if we want to introduce a buffer capacity of 20%, β is set to 1.2.

2.4 Initial conditions

Initial condition refers to the situation of the first flight of an aircraft sequence, meaning when, where and the departure delay of this flight. As mentioned in the main text, variations on this situation can have a great impact on the delay propagation. In other words, the dynamics of delays over the network is highly sensitive to the initial conditions.
We characterized initial conditions by the average delay per flight for the first flights of all the aircraft sequences and by the fraction of airplanes that their first flight was delayed. Comparing the ranking of the 20 worst and best days of 2010 (Figure S9) we can observe that it is most likely that if a day started with unfavorable initial conditions it will likely produce large congested clusters.

The simulations can be initialized by two different ways depending on the case under study: from data or random initial conditions.
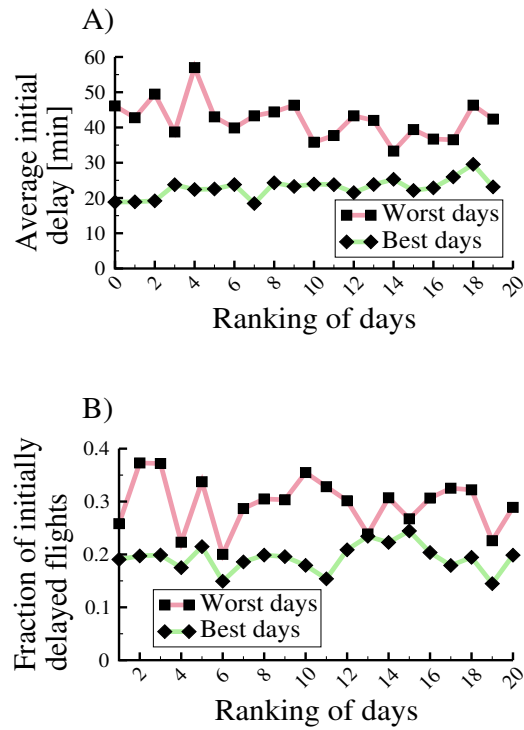
Figure S9: Initial conditions of the 20 worst days (red) and of the 20 best days (green) of 2010.

2.4.1 From the data

Initializing the model "from the data" means to replicate exactly the situation of the first flights of all the aircrafts sequences for a particular day.

2.4.2 Random initial conditions

When random initial conditions are set, initial delays are reshuffled among all possible aircrafts, so when and where may vary. Two inputs are needed: initial delay and fraction of flights initially delayed. For instance,

Initial delay: 20 minutes

Percentage of airplanes initially delayed: 10%

Suppose that the number of aircrafts for one day simulation is $4,000$. In this example, 400 aircrafts will have their first flight departing with an initial delay of 20 minutes.

2.5   Decision Tree

Model flowchart summary including all the subprocesses. Flowchart objects in green and red will be explained separately.

- *Generate class objects*:
  Once the data is loaded for a particular day into the data class object, the remaining class objects are created using this data structure. These objects are:

    - Airport list: Indexation of all the airports that operated that day.
    - SAAR matrix: Includes the hourly capacity (schedule airport arrival rate) for every airport in the list.
    - Airport Flight Connectivity Factor List.
    - Adjacency list: Contains the network structure for that day.
    - Tail number: Indexation of all the aircrafts that operated that day.
    - Schedule: For each flight the schedule object contains the information described in section 2.2.2, initial delay (see 2.4), flight index, flight status (on land "L", flying "F" and in service or in queue "S"), inbound index (previous flight leg index) and connections (see 2.3.2). All flights are initialized with flight status "L".
    - Tail number situation: For each aircraft contains the origin airport, the destination airport, the scheduled and actual block-to-block time and the departure delay (initial, inbound, queue and due to connections).
    - Airports tail number queue: For each airport contains the aircrafts ordered as First in - First served.
    - Airports flight queue: The same as the previous one but indexed with flight number.

- *Update class objects & t +1 min*:
  Objects as Schedule and Airport tail number and flight queues are synchronous updated for each time step.

- *Aircraft arrival*:
  The flight status is changed from "F" to "S" and the airport's tail number and flight queues are updated.

- *Aircraft can depart?*:
  The aircraft can depart if the service time (30 minutes) is complete and the are no flight connections to wait for. Initial flight legs of an itinerary are considered as already served.

- *Aircraft departure*:
  Tail number situation and origin airport queues are updated. The actual block-to-block time is reset. Flight status is changed from "L" to "F".

- *Previous flight legs are not completed?*:
  Check if the inbound index is among the flight connections and the flight status is "L".

- *Aircraft is in service?*:
  Inspect if the flight status for the aircraft is "S" and the service time is different from zero or the aircraft position at the airport queue is less than airport capacity.

- *Aircraft is in airport's queue?*:
  Check if the flight status is "S" and the service time is zero.

- *Connection/s not landed?*:
  Verifies if the number of connections in the schedule for the flight is zero and the flight status is "L".

2.6 Clustering

1. Create a cluster list with all airports labeled as -1 (unexplored).

2. Create an empty list (active list) to include the airports to inspect while traversing the adjacency list (network).

3. While unexplored airports continue to exist in the cluster list:

   - For each airport in the cluster list:
     - Check if the airport is unexplored and the average delay per flight for the airport is greater than 29 minutes.
     - If it is so, label the airport with its index and insert the airport index in the active list.
     - Else, label the airport as -2 (not delayed).
     - While the active list continue to have airports to explore:
       - For each airport in the active list:
         * Explore its neighbors in the adjacency list.
         * Check if they are labeled as unexplored and their average delay per flight is greater than 29 minutes.
         * If it is so, label them with the same index as before and insert the airport index in the active list.
         * Else, label the airport as "not delayed".
       - Remove from the active list the airports that their neighbors had been explored.

2.7   Overview of the model parameters

Parameter	October 27	March 12	December 12	July 13	October 9	April 19
T_s [min]	30					
ΔT [min]	180					
α	0.263	0.190	0.265	0.075	0.020	0.020
β	1.0					
Initial Condition	"From the data"					

Table S4: Overview of default values of the model's parameters. The values of α correspond to the best fit for the day.
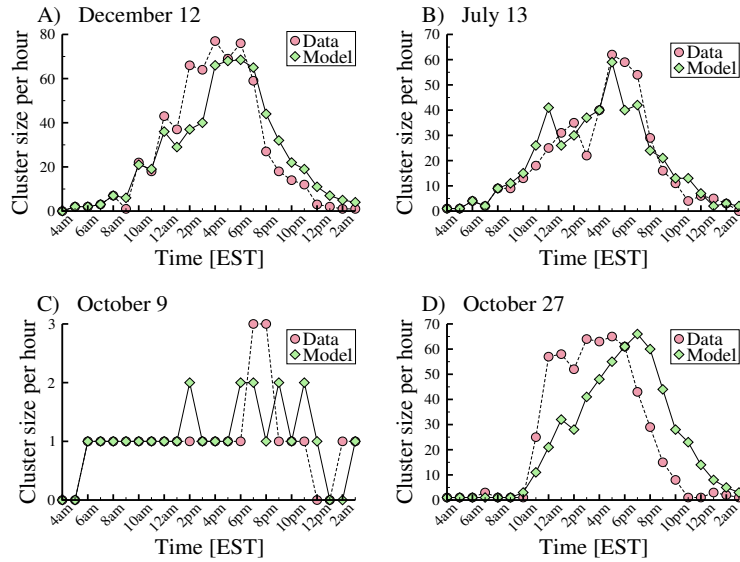
3   Model simulations

Figure S10: Evolution of the largest cluster size for A) December 12 (α = 0.265), B) July 13 (α = 0.075), C) October 9 (α = 0.002) and D) October 27 (α = 0.263)

3.1   Model validation and sensitivity to α

Figure S10 displays results for other days different from the ones presented in the text. Results for December 12 (Figure S10 A), July 13 (Figure S10 B) and October 9 (Figure S10 C) confirm
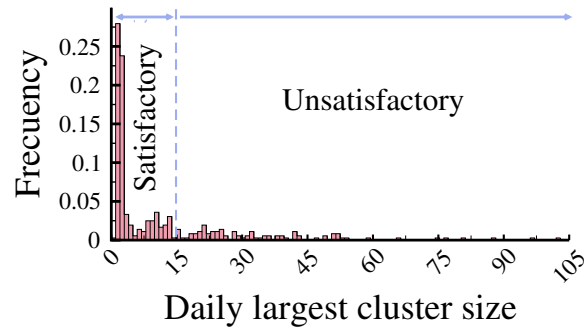
Figure S11: Frequency of the largest cluster size for all days of 2010.

that the model is in good agreement with the data when α is fitted for each day. In the case of October 27 (Figure S10 D), the size of the cluster evolved much faster than the model prediction, although the size could be predicted. Analyzing the possible explanation to this difference, we found that severe weather conditions occurred that day across an important part of the country [4] affecting flights in airports such as Hartsfield-Jackson (Atlanta), John F. Kennedy (New York), La Guardia (New York), St. Paul (Minneapolis), O'Hare (Chicago), Philadelphia and Newark. External perturbations were not explicitly introduced in the model so we cannot expect to be able to reproduce well these days delay dynamics.

In the previous sections we have defined days/airports with problems as those whose average delay per delayed flight was over 29 minutes. Another way, related to the previous one, of classifying the days is by means of the largest cluster size of the day. To do so, we set a cluster size that corresponds to 15 airports so that if the largest cluster size in a day is higher than this threshold
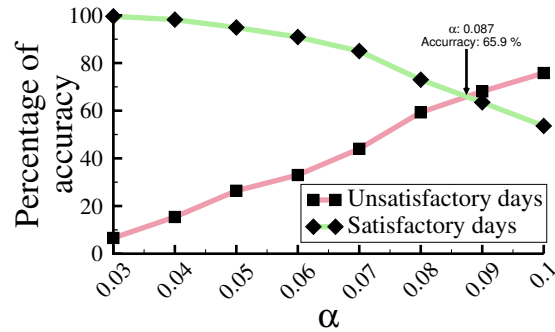

Figure S12: Exploring the model forecast accuracy by varying the α parameter. All days of 2010 are taken into account.

the day is labeled as problematic or unsatisfactory. On the other hand, if it is less than 15 airports the day is labeled as satisfactory. This threshold was selected because in the distribution of largest cluster size there exists a small depression at this value (Figure S11). This particular value for the threshold is arbitrary. Still, we have repeated the analysis with some other thresholds and checked that the main conclusions are maintained.

Unsatisfactory days		Satisfactory days	
DATE	Accurate Prediction	DATE	Accurate Prediction
Oct, 27	No	Apr, 19	Yes
Mar, 12	Yes	Oct, 09	Yes
Dec, 12	Yes	Nov, 11	Yes
Jan, 24	No	Apr, 14	Yes
Feb, 24	Yes	Oct, 08	No
May, 31	No	Set, 11	Yes
May, 21	Yes	Apr, 15	Yes
May, 14	No	Oct, 13	Yes
Jun, 23	Yes	Apr, 17	Yes
Jul, 13	Yes	Nov, 10	Yes
Jun, 24	No	Nov, 09	Yes
Jul, 12	Yes	Mar, 06	Yes
Jan, 21	Yes	Oct, 12	Yes
Jul, 29	Yes	Mar, 17	No
Jun, 15	Yes	Feb, 28	Yes
Jun, 27	No	Oct, 16	Yes
Mar, 20	Yes	Apr, 13	Yes
Mar, 11	Yes	Nov, 26	Yes
Aug, 22	Yes	Set, 09	No
Jan, 25	Yes	Set, 20	Yes

Table S5: Ranking for the top 20 days by the average delay for flights with positive delay. Model accuracy according to the classification of each day in satisfactory or unsatisfactory. The model is able to predict unsatisfactory days with an accuracy of 70% and satisfactory ones with an 85%.

The introduction of the threshold allow us to define a binary variable associated to the performance of the network each day. Since the model requires a fit in α to reproduce the precise dynamics of the congested clusters, the aim of this exercise is to set a generic value of α and study how many of the satisfactory/unsatisfactory days are actually predicted. According to our definition, during 2010, 75% of the days get a satisfactory performance. In order to assess the model correspondence with reality, we have to take into account that satisfactory days outweigh unsatisfactory ones. Naturally, with a high α the model simulations predict unsatisfactory days with high accuracy but provide many false positives for satisfactory days. On the other hand, with a low α, most of days with small clusters are successfully predicted but not those with large congested clusters. Bearing this in mind, we defined the percentage of accuracy as a tradeoff between

March 12		
Airport Code	Precentage Realizations	Accurate Prediction
ATL	100.0	Yes
CWA	100.0	Yes
DFW	100.0	No
DLH	100.0	Yes
EAU	100.0	No
EYW	100.0	Yes
FLL	100.0	Yes
GGG	100.0	No
MGM	100.0	Yes
MIA	100.0	Yes
ORD	100.0	No
SJT	100.0	No
STT	100.0	Yes
TOL	100.0	No
BHM	99.8	Yes
CAK	99.5	Yes
CHA	98.6	Yes
FAY	98.4	Yes
MEM	98.3	Yes
HSV	97.9	No

Table S6: Top 20 ranking of airports that appear more frequently in the largest cluster for the model results compared to what actually occured on March 12.

the percentage of accuracy for satisfactory and unsatisfactory days. Figure S12 show the fraction of correct predictions both for satisfactory and unsatisfactory days. Both curves cross at a value of $\alpha = 0.087$ and at an accuracy rate of 65.9%. Obviously, this is a simplistic technique to measure performance. A more elaborate technique should include appropriate economic considerations to take into account that the cost related to false positives, claiming that a day is going to have a large congested cluster without actually occurring, and false negatives, not being able to predict a major collapse, are different. Even so, this simple method provides us with a quantitative framework to validate the model and to assess the importance of including further mechanisms in the simulation.

Another accuracy test was done to check if the model is able to predict not only the size but the airports that comprises the largest cluster of the day. We selected March 12 whose largest cluster is formed by 97 airports. The model is stochastic, so we run it for 1500 realizations. Comparing the data with the model results for the top 97 airports most frequently appearing in the largest cluster, the model accurately identify 57.8% of them. Table S6 displays the Top 20 airports which are more prevalent in the simulations showing if they appeared in the real data as part of the largest cluster for March 12. This is a first comparison, since the real cluster is coming from a single realization in

a particular day it cannot be taken as a definitive validation of the model. However, an accuracy of 57.8% with such a simple framework is already encouraging.

3.2 Analysis of the model sensitivity to changes in β

	April, 19			March, 12	
β	Average largest cluster size [airports]		β	Average largest cluster size [airports]	
1.0	1.0		1.0	92.0	
0.9	1.0		1.1	92.3	
0.8	1.0		1.2	89.9	
0.7	1.0		1.3	88.7	
0.6	1.3		1.4	86.4	
0.5	16.8		1.5	86.4	

Table S7: β variation for April 19 with an α fixed at 0.02: airports should work at half the scheduled capacity to transform this day into an unsatisfactory one. β variation for March 12 with an α fixed at 0.19: an increase of 50% in β decreases the size of the cluster by only a 7%.

As stated in the main text, the model is able to reproduce the clusters of congested airports by fitting α while fixing β to 1. In fact, as shown in Table S7 the model has a low sensitivity to a variation in the β coefficient.

For April 19 Table S7 shows that only by cutting the scheduled capacity by half, the day will start to have systemic problems according to the size of the largest cluster. In the case of March 12 the scheduled airport capacity is increased by 50% and the results indicates that this increment does not change the overall picture. Furthermore, Figure S13 B shows that increasing airports' capacity will not ease off the propagation of delays. The reason for this is that the main cause of delay spreading, flight connections within the schedule, is independent of the airport capacity. Conversely, by reducing β by at least 50% can worsen the situation (Figure S13 A). Such decrease on the airports' capacity can act as a trigger to new primary delays (different from the initial ones) that later on will spread in a cascading effect due to the flight connectivity. Although, a decrease on the scheduled capacity of 50% for every airport in the network is not likely to occur in practice, a much realistic situation could be an airport or group of airports operating undercapacity when severe weather conditions are met.

In any case, airport congestion could be the source for primary delays but it does not seem to be an important force behind their network-wide propagation.

3.3 Stochastic variability of the results

Because of the stochasticity included in the model each realization has a slightly different outcome. Figure S14 displays the variability between model realizations of the results for March 12 considering a confidence interval of 95%. Simulations in this case were done using initial conditions "from the data"; this means that the stochasticity is caused only by flight connectivity. No matter which set of flight connections are randomly selected, March 12 will continue to display a large cluster.
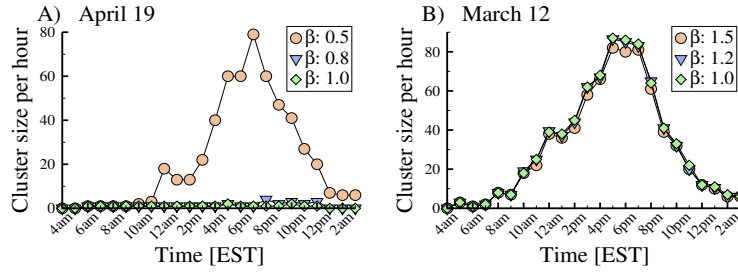
Figure S13: Dependence of the hourly largest cluster with aS variation in β. A) April 19 and B) March 12.
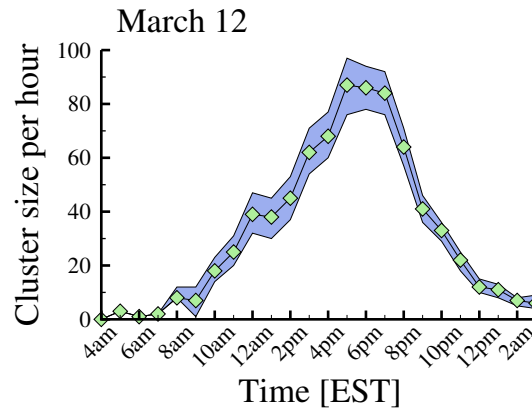

Figure S14: Exploring the variability of the model results.

In Figure S14 we can differentiate a growing phase that goes from 4am to 5pm and a declining phase from 5pm onwards. As already said, merging is critical for the size evolution of the clusters. Because in the first hours of an unsatisfactory day there are several clusters, thus more possible combinations of merging events, the growing phase is characterized by a stronger variability than the declining phase. The latter, depicts a low variability and as Figure S15 A shows the number of clusters do not increase during this phase. All in all, this indicates that no atomization into smaller clusters is produce when the size diminishes. The cluster size dissolves continuously.

3.4   Further results on cluster and individual airport dynamics

Besides the evolution of the size of the largest cluster per hour, dynamics can be characterized by the evolution of the number of clusters during the day (see Figure S15). While in a satisfactory day (Figure S15 C and D) the number of clusters varies in each hour without a recognizable pattern, in an unsatisfactory day (Figure S15 A and B) the number of clusters increase in the first hours of
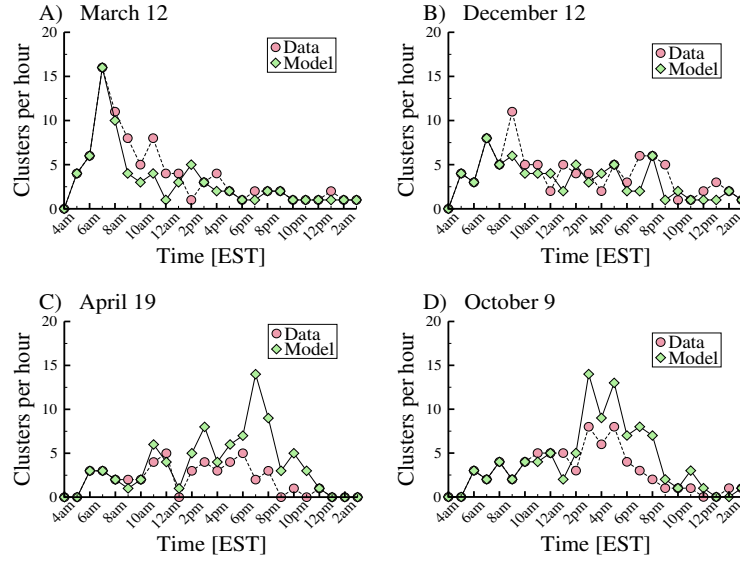
Figure S15: Evolution of the number of clusters. Comparison between data and model results for: A) March 12, B) December 12, C) April 19 and D) October 9.

the morning and from then on decay merging into fewer clusters, in most cases, in the afternoon (eastern time). This high-level interaction dynamics between clusters appears to be crucial in the evolution of an unsatisfactory day, where high-degree nodes play an important role to make this
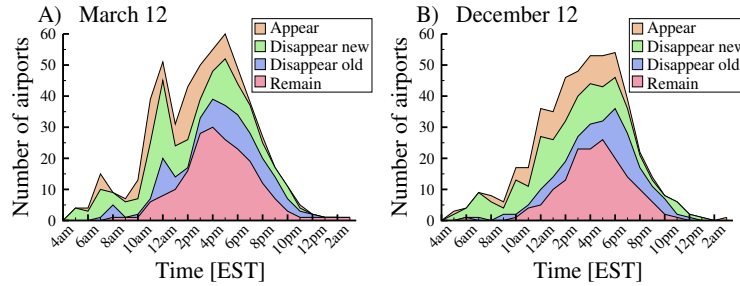

Figure S16: Results obtained from the data. Number of airports that belongs to the largest cluster of the day for A) March 12 and B) December 12. Red color indicates the number of "old" airports (that their average departure delay per flight has been > 29 minutes at least in the previous hour as well), while new airports that match this condition are shown in orange. The nodes which their average departure delay per flight will drop below 29 minutes in the next hour are shown in green (if they have been in problem for only one hour) and blue (if they have been in problem at least for two hours).

Airport code	days in largest cluster	Airport code	days with problems
ACV	100	OTH	167
CEC	80	CEC	138
SFO	54	ACV	136
OTH	52	LMT	111
MOD	49	MOD	90
EWR	45	CIC	86
CIC	45	MFR	70
LMT	44	BRW	62
MFR	43	CRW	60
CRW	41	MLB	60

Table S8: Top 10 raking of airports in number of days belonging to the largest congested cluster or in number of days with problems.

merging event come about.

However events involving individual nodes occur and varies dramatically from time to time. Figure S16 displays how nodes that belong to the largest cluster of the day vary their condition rapidly. One hour they are above the 29 minutes threshold and next they recover and vice versa. Most nodes switch from one state to the other very quickly, although some few nodes repeat their condition at least two time steps (red series).

In order to study the temporal persistence of airports in the largest congested cluster across the whole database, we display in Table S8 the list of the top 10 airports in days in the largest congested cluster and in days with problems. Although some airports appear in both lists, the order changes and both sets are not exactly equal. In these lists, there is a strong component of airports located in the West Coast. We think that this is due to the time difference between East and West Coasts. Flight operations initiate before in the East Coast and so the delays can propagate Westwards toward the end of the day. In the results in the main paper, we show that the largest congested clusters are not persistent, at least not in more than ~ 50% of the airports between different days. The airports in Table S8 are those most persistent in the largest congested cluster. It is interesting to notice that only two major hubs, Newark (EWR) and San Francisco (SFO), are present in the top 10 list.

References

[1] Bureau of Transport Statistics of the US Government, Web page: `http://www.bts.gov`

[2] BTS press release of March 22, 2011, available at
`http://www.bts.gov/press_releases/2011/bts017_11/html/bts017_11.html`

[3] pytz - World Timezone Definitions for Python, `http://pytz.sourceforge.net`

[4] News report available at
    `http://articles.cnn.com/2010-10-27/us/`
    `us.weather_1_tornado-damage-tornado-sightings-airport-delays?_s=PM%3AUS`