Evolut Inst Econ Rev (2015) 12:349–373
DOI 10.1007/s40844-015-0025-y

ARTICLE

Worldwide aviation network vulnerability analysis: a complex network approach

Q. H. Anh Tran[1] · Akira Namatame[1]

Published online: 8 January 2016

Abstract Transportation networks play a crucial role nowadays, not only in positive terms that include human mobility and the exchange of goods, but also in negative terms that include the spread of diseases and other malignancies. Among these networks, the worldwide aviation network (WAN) has become the largest and one of the most important global transportation networks in modern society, supporting the traffic of billions of passengers traveling between thousands of airports on millions of flights every year. Since the WAN has become one of the indispensable infrastructures in our daily lives, understanding its structure and vulnerability is an essential issue. In this work, we apply complex network analyses to elucidate the hidden characteristics of the network. We first construct a global aviation network using datasets obtained from an open source project named OpenFlights. We then clarify the topological and spatial characteristics of the constructed network and provide a binary status model to investigate the dynamics of the network with emphasis on its vulnerability as a result of extreme events. Our results may contribute to the understanding of the response of the network to disturbances and provide insights on the construction of a robust network and the improvement of the network resilience.

Keywords WAN · Breakdown · Vulnerability · Resilience

JEL Classification C81 · R41

✉ Q. H. Anh Tran
newsv2004@gmail.com

Akira Namatame
nama@nda.ac.jp

[1] Department of Computer Science, National Defense Academy of Japan, Yokosuka, Kanagawa, Japan

1 Introduction

The worldwide aviation network (WAN) plays a crucial role in our modern society nowadays. Through this system, activities related to culture, society, and the current economy can be closely and easily connected with active exchanges between regions and countries all over the world. Because the WAN is essential for the smooth exchange of people and goods, thereby aims towards the stimulation of the economy, it is important to study and understand its characteristics and dynamics.

Over the last few decades, complex network analyses have become widespread, attracting many researchers. Network analyses can provide insight into the topological structure, performance, and dynamics of complex systems. Hence, complex network theory has been widely used to characterize many systems in nature and society, including social, technological, and biological systems, such as the Internet, the World-Wide Web, computer networks, electrical power grid networks, metabolic networks, and others. More recently, the advancement of complex network theory has generated interest in the area of aviation systems.

In complex network theory, an aviation system can be modeled as a graph, in which airports act as nodes linked by directed flights. Interestingly, many real networks, including aviation networks, share some topological similarities, such as small-world properties, whereby the average topological distance between nodes is small, and leads to the emergence of hubs–nodes that have large degree of centrality (Amaral et al. 2002; Guimera et al. 2005). Additionally, the degree distributions in these systems, $P(k)$, have been discovered to exhibit heavy tails that are well-approximated by power-law behaviors of the form $P(k) \propto k^{-\beta}$, in which the exponent β takes values within the range $1 \le \beta \le 3$ (Albert and Barabási 2002). Other studies of aviation networks focused on analyzing the overall network features, as well as on identifying the importance of individual airports (Guimera and Amaral 2004b; Bagler 2008; Wang et al. 2011).

Even the increased importance and critical role that aviation network infrastructures play in our daily lives, they can unexpectedly lead to large-scale disasters that trigger socio-economic problems with tremendous costs. For example, the eruption of the Icelandic volcano, Eyjafjallajökull, in 2010 led to the cancellation of at least 60 % of daily European flights, affecting millions of passengers that lasted for 5 days (Bye 2011; Verma et al. 2014). Thus, investigating the topological features of aviation networks is important. However, understanding how to identify the relationship between their structure and their vulnerability is even more important.

A vast number of studies have clarified the fact that certain topological properties of complex networks have a strong impact on their stability. This observation has triggered intense research efforts aimed towards the understanding of the organizing principles of these networks and the interplay between topology and network dynamics (Albert and Barabási 2002; Krapivsky and Redner 2002; Tran and Namatame 2014). Major studies have evaluated the reliability and vulnerability of networks, and the results showed that, locally emerging failure due to error, interference from environmental conditions, or attacks, can lead to global economic losses and social disruption (Albert et al. 2000; Holme et al. 2002; Chassin and

Posse 2005; May et al. 2008; Buldyrev et al. 2010). In the context of worldwide aviation networks, there has been considerable effort in understanding how the network responds to damage (Meza et al. 2013; Verma et al. 2014), and how network structure impacts its dynamics, e.g., how this infrastructure leads to the unintended consequence of facilitating the rapid spread of infectious diseases (Hufnagel et al. 2004; Colizza et al. 2006), and the subsequent and necessary intentional shutdown of airports to block the transmission of infection (Colizza et al. 2007; Epstein et al. 2007). There has also been some investigation on system vulnerability to real disasters (Bagrow et al. 2011; Meza et al. 2013), that showed how natural events differ from synthetically modeled systems, how they differ from one another, and which common features they share.

In this article, we present a complex network approach for measuring the function and estimating the vulnerability of the WAN, which is constructed using datasets provided by an open source project known as OpenFlights (2014). The first part of this work discusses the structure of the WAN with regard to its topological geometry and spatial structure. We present an overview of global- and local-level measures for understanding the structure of the network. In the second part, we address the following questions: which model is suited to assess the vulnerability of the network? Is the often used topological integrity proper to do this? In the effort to address the question, we introduce a binary status model to assess the structural vulnerability of the WAN under different airport breakdown scenarios that perturb the network, i.e., intentional breakdown driven by various centrality strategies, and random breakdown. We find that the network may lose a comparable efficiency even when only a single airport fails, whereas the topological integrity of the network is still extremely high. Using the proposed model, we reveal the "robust yet fragile" property of the network, that is to say, the fact that the network is resilient to random breakdown but vulnerable to intentional breakdown which focus on, not only central, but also salient airports. The efficiency of the entire network wherefore depends on the protection of these significant airports. The findings may help towards the better understanding of the essential features of the network, providing a quantitative assessment of the most vulnerable airports, the vulnerability of which is often masked by the network's complexity.

2 Network structure of the WAN

2.1 Constructing the WAN from data

The WAN is the global network of airports that are connected through direct flights, and supports billions of passengers per year. Using a dataset downloaded from the website openflights.org, we first built an aggregate network of airports, considering all flights among all destination airports throughout the world. Since 2014, the datasets provided by OpenFlights project have comprised three databases:

- Airport database: The database contains more than 8000 airports that span the globe, and provides information on airport ID, airport name, city and country where an airport locates, IATA/FAA, and ICAO code, longitude and latitude, and other information
- Airline database: Each entry in the database contains information on airline ID, airline name, alias of airlines, IATA/FAA, and ICAO code, airline call-sign, country, and active status, that is, if the airline is, or has until recently, been operational or defunct
- Route database: The database contains nearly 70,000 routes between airports. Each entry in the database contains information on source airport ID, destination airport ID, code-share, and equipment codes for plane types generally used on this flight

The dataset is freely accessible. However, the disadvantage of this dataset is that it offers only airport, airline, and route database, and does not have any other data available, such as the number of flights, number of passengers, or the airline schedule (timetable) data service compared to other alternative commercial sources of data.

Although a variety of networks can be understood in terms of their topological connectivity, a number of systems are better captured by weighted networks in which edges are described in terms of weights that quantify their strengths. Moreover, in the context of aviation networks, spatial constraints also play an important role, resulting in a complex interplay between topology, weight, and geography. In order to study the vulnerability of such a network, weight and spatial constraints must therefore be considered along with the topological quantities.

We model the WAN by a weighted network, in which nodes represent airports, and edges represent nonstop flights. Mathematically, we represent the network by a $N \times N$ weighted matrix, $\mathbf{W}$, where N is the number of airports. The elements in the ith row and jth column of matrix $\mathbf{W}$ are expressed in terms of w_{ij}, which quantify the coupling strength between airport i and airport j. Normally, depending on the context, w_{ij} might reflect the number of direct flights or the number of passengers of the corresponding flight between i and j during a period of time. However, due to the particularity of the provided dataset, we define w_{ij} as the number of routes between i and j. If $w_{ij} > 0$, i and j are connected and there exists a determined number of flight routes between them. If $w_{ij} = 0$, the two airports are not connected, meaning that no flight operated between them. A simple example of weight definition is shown in Fig. 1.

The full network obtained from the dataset of OpenFlights contains $N = 8107$ airports and $E = 34{,}091$ edges. Since traffic is strongly symmetric for all airports, we carry out our analysis on the undirected WAN network so that the weights are symmetric ($w_{ij} = w_{ji}$) for all edges. After refining the raw data, converting the network to an undirected one, and abstracting the largest connected component from the entire network, we obtain the remaining undirected, weighted and strongly connected WAN with $N = 3010$, and $E = 17{,}450$ as shown in Fig. 2.

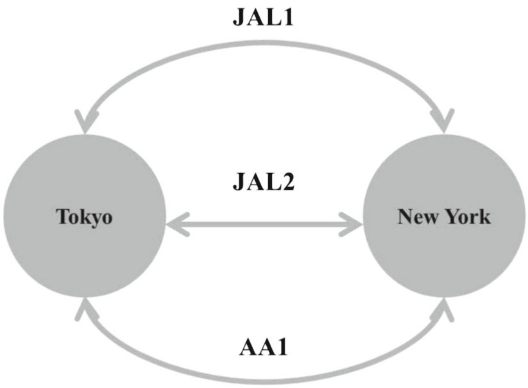


Fig. 1 An example of weight definition. Suppose that two carriers, namely, Japan Airlines (JAL) and American Airlines (AA), operate flights between Tokyo and New York. The JAL operates two flight types JAL1, JAL2, whereas the AA operates only AA1. The weight of the edge connecting these two airports will then be equal to the total of flight types ($w_{\mathrm{Tokyo,NewYork}} = 3$)

Fig. 2 Spatial structure of the WAN. The number of nodes $N = 3010$, and the number of edges $E = 17{,}450$. *Darker lines* indicate connections that have larger weights. As seen, connections that have large weights are operated within continents as well as inter-continentally. Most are connections between European–European, European–North American, North American–Asian, and Asian–Asian airports

2.2 Topological characteristics of the WAN

In the study of complex networks, one common characteristic is community structure. A network is said to have community structure if it can be easily divided into groups of nodes, such that the connection inside each group is dense, whereas the connection between groups is much sparser. Synthetic model-based generated networks, such as random graphs (Erdös and Rényi 1959), or scale-free graphs (Barabási et al. 2000), may not have any meaningful community structure. However, community structures are quite common in real networks. For example, communities are formed by common location, occupation, or interest in social networks, or by research topic in citation networks. Identifying communities within a network is important since it can provide insights into its topology and functionality.

Several methods of community detection have been developed. One of the most widely used methods is modularity maximization proposed by Newman (2004). Modularity is a quantity that measures the quality of a particular division of a network into communities. The modularity maximization method detects communities by searching over all possible divisions of a network. Since exhaustive search over all possible divisions is computationally intractable, practical approximate optimization algorithms are provided, such as greedy algorithms, and simulated annealing, with different methods offer different balances between speed and accuracy (Guimera and Amaral 2004a; Danon et al. 2005). A popular modularity maximization approach is the Louvain method by Blondel et al. (2008), which iteratively optimizes local communities until global modularity can no longer be improved, given perturbations to the current community state. Especially, the method is proved to be efficient in detecting communities in large networks. Hence, we apply the method to identify community structure of the WAN. Figure 3 shows the top 10 largest communities of the network based on size. As showed in the figure, it is not surprising that airports tend to form communities with others that are nearby due to political and geographical conditions. The largest community is the one that covers almost all airports in America. Interestingly, Heathrow airport in the United Kingdom is the only one airport of Europe that belongs to this community. It implies that Heathrow strongly connects to airports in America more than connecting to others in the same continent. The second one is the East Asia–Southeast Asia–Australia community, and the next is the community of the European airports. Additionally, Alaska of the United States, and Canada are detached from the rest of the America to become independent communities, ranking in the sixth and eighth by size, respectively.

Another key issue on how to characterize properties of complex networks is the identification of the most important nodes in the systems. Centrality is a concept that can deal with the issue. However, centrality is not a unique feature. In fact, it can be

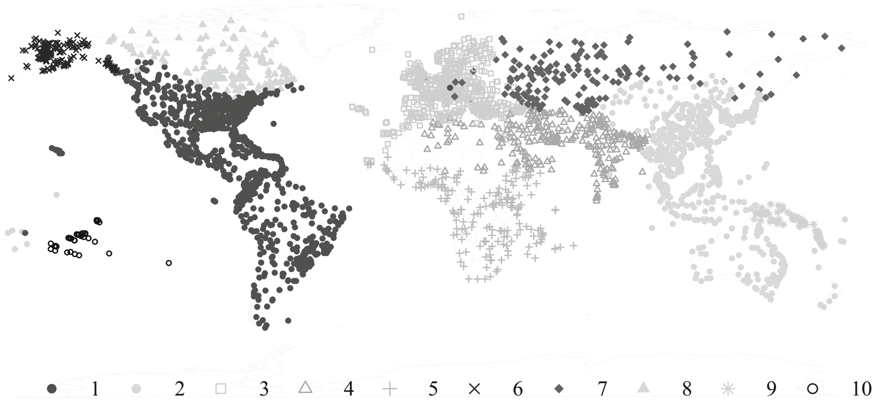


Fig. 3 Top 10 largest communities (in a total of 24 communities) of the WAN. The *legend* shows the sizes of communities represented by the number of nodes within communities in descending rank from the largest (number 1) to the smallest one (number 10)

quantified by various measures. Among them, degree centrality is a local and most intuitive quantity and provides insights on how to quantify the importance of a node. As mentioned, we represent the WAN by a weighted matrix $\mathbf{W}$ in which the elements in the ith row and jth column of matrix $\mathbf{W}$ are expressed as w_{ij}, which quantify the coupling strength between airport i and j. The strength of a node i can be defined as $s_i = \sum_j w_{ij}$, where j is the node that directly connects to i. If we consider that $w_{ij} = a_{ij} = 1$ for all $w_{ij} > 0$, meaning that we ignore the meaning of weights of edges in the network and assign all connected edges a unit value, the predefined strength s_i of node i becomes its degree centrality, k_i, that simply indicates how many connections node i has, $k_i = \sum_j a_{ij}$. In other words, the degree of node i is the unweighted version of the strength of node i. The simplified WAN corresponds to an average degree equal to 11.6, while the maximal one is equal to 240, exhibiting a strong degree heterogeneity. Moreover, the WAN is topologically a small-world: the average shortest path length, measured as the average number of edges separating any two nodes in the network, is equal to four.

An important issue is to concern how centrality ranking relates to the geographical information available for the network. Figure 4 shows the degree distribution map of the network. As observed, high-degree airports locate mostly in Europe, North America and East–Southeast Asia. In the top highest degree airports as shown in Fig. 5, Charles de Gaulle (CDG) is the most connected airport, that connects to 240 other airports all over the world. Other hub-like airports include Ataturk International Airport (IST), Flughafen Frankfurt am Main (FRA), Amsterdam Airport Schiphol (AMS), Beijing Capital International Airport (PEK), and others. The figure on the right of Fig. 5 shows the dominance of American, Asian, and especially European airports, in terms of degree centrality. None of the airports in other continents appear in this ranking list.

Figures 6 and 7 show the strength distribution map and the top highest strength airports of the network. Similar to the degree distribution shown in Fig. 4, high-

Fig. 4 Degree distribution map of the WAN. Airports with high-degree centrality are indicated by the increased *sizes of the circles*

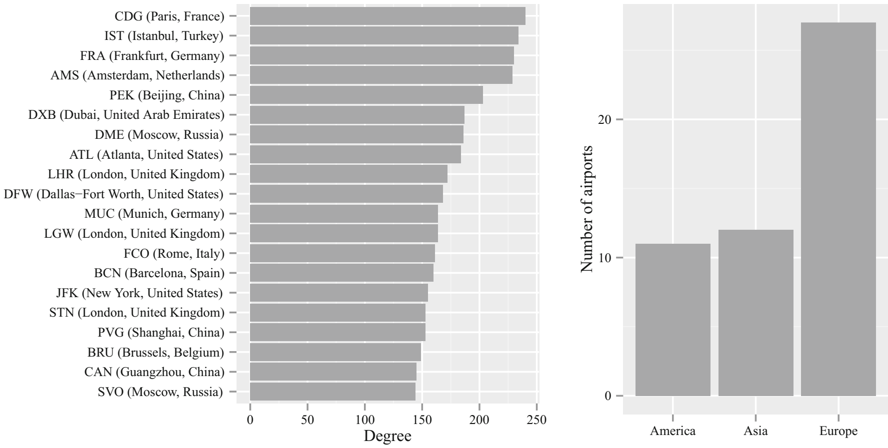


Fig. 5 Highest degree airports (top 20) in the network (*left*) and the cumulative number of highest degree airports (in the top 50) of each continent (*right*)

Fig. 6 Strength distribution map of the WAN. Airports with higher strengths are indicated by increased *circle sizes*

strength airports locate mostly in Europe, North America and East–Southeast Asia. The highest strength airport is Beijing Capital International Airport (PEK), followed by Charles de Gaulle (CDG), Pudong International Airport (PVG), Flughafen Frankfurt am Main (FRA), Ataturk International Airport (IST), and others. The figure on the right of Fig. 7 shows the dominance of American, European, and Asian airports in terms of strength centrality. Particularly, Asia has a number of major airports in this ranking, and none of the airports of other continents appear in this ranking.

As seen in Figs. 5 and 7, top highest degree airports also appear in the ranking of the top highest strength airports, indicating the positive relation between strength and degree. Figure 8 further shows the strong correlation between these two

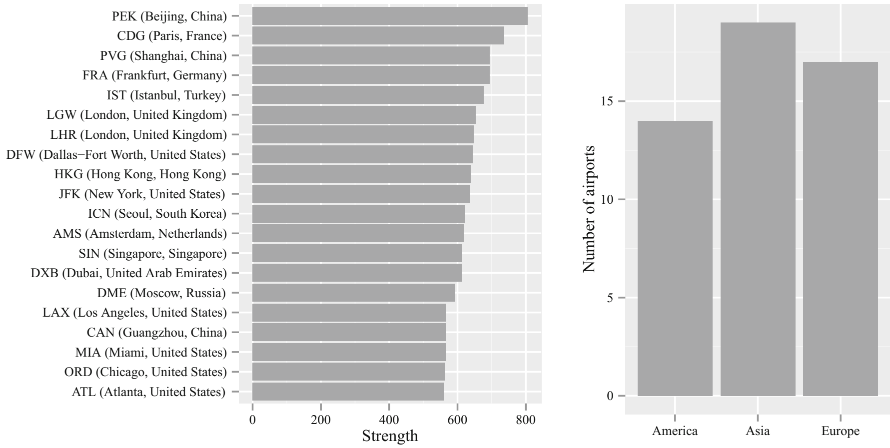
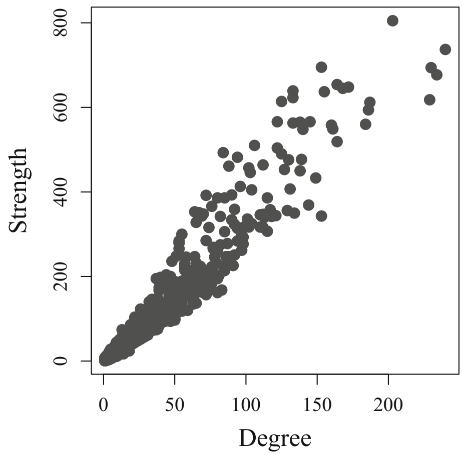


Fig. 7 Highest strength airports (top 20) in the network (*left*) and the cumulative number of highest strength airports (in the top 50) of each continent (*right*)

Fig. 8 Correlation between strength and degree



centralities. The correlation coefficient between these two centralities is equal to 0.97.

However, local measures, such as degree and strength, do not take into account global effects, such as for example the existence of crucial nodes, which may have a small degree or strength, but act as bridges between different communities of the network. In this context, a widely used quantity to investigate node centrality is the so-called node betweenness centrality (Freeman 1977). The betweenness centrality of a node is the number of shortest paths between any arbitrary pairs of nodes that passes through that given node. According to the definition, important nodes are therefore part of shortest paths than less important ones.

In weighted networks, weight of unequal edges render some specific paths more favorable than others in connecting two nodes. It thus seems natural to generalize the notion of node betweenness centrality through a weighted node betweenness centrality in which the shortest paths are replaced by weighted ones. In the context of aviation networks, such as the WAN, this notion is reasonable considering two airports to be close if heavy traffic exists between them. In other words, it is natural to assume that the proximity, p_{ij}, between two connected airports i and j is inversely dependent on the weight of the edge connecting them, i.e., $p_{ij} = 1/w_{ij}$. For any two airports i and j, the shortest weighted path between i and j is the one for which the total proximity of the edges forming the path between i and j is minimum. The weighted node betweenness centrality of an airport i is then defined as the total number of shortest weighted paths that pass through the airport. The weighted node betweenness centrality represents a trade-off between finding bridges connecting different parts of a network, taking into account weights, which demonstrate the fact that some edges may carry more traffic than others.

According to the definition of the weighted node betweenness, we exhibit the betweenness distribution of the WAN. As shown in Fig. 9, only few airports have large betweenness, becoming significant gateways so that most flights have to travel through. On the other hand, major airports do not contribute to transferring traffic at all, i.e., the weighted node betweenness is equal to zero. Hence, we can separate airports in the WAN into two groups: a group of gateways, and a group of peripheral airports.

Figures 10 and 11 show the weighted node betweenness distribution map and the top highest betweenness airports of the network. The most central airport is Heathrow (LHR), followed by the John F. Kennedy International Airport (JFK), Chicago O'Hare International Airport (ORD), Los Angeles World Airport (LAX), Changi Airport (SIN), and others. The figure on the right of Fig. 11 shows the dominance of European, Asian, and American airports, in terms of weighted node betweenness centrality. Different from the results as shown in Figs. 5 and 7, where

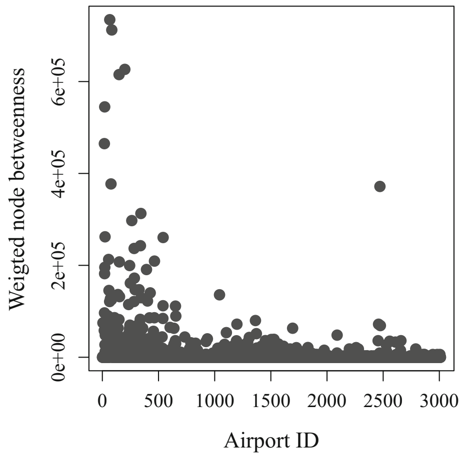


Fig. 9 Weighted node betweenness distribution of the WAN

Fig. 10 Weighted node betweenness distribution map of the WAN. Airports with high-weighted node betweenness are indicated by increased *circle sizes*. Airports with same characters belong to same community (see Fig. 3 for reference). The airports with the highest weighted node betweenness are major airports in the American community (interestingly, the highest weighted node betweenness airport is Hearthrow (LHR) in the United Kingdom, also belongs to this community)

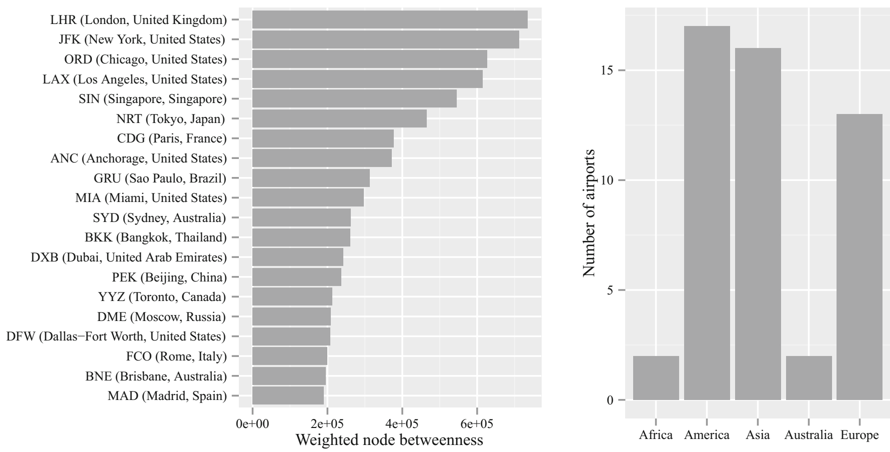


Fig. 11 Highest weighted node betweenness airports (top 20) in the network (*left*) and the cumulative number of highest weighted node betweenness airports (in the top 50) of each continent (*right*)

European airports dominate, America has a number of major airports in this ranking list.

Figures 5, 7 and 11 show the geographical distribution of the most central airports ranked according to different centrality measures, highlighting the different properties and biases of different centrality measures. A significant level of positive correlation between degree and strength is observed in Fig. 8, i.e., airports that have a large degree typically also have large strength. However, Fig. 12 shows weak correlations between weighted node betweenness and degree as well as weighted

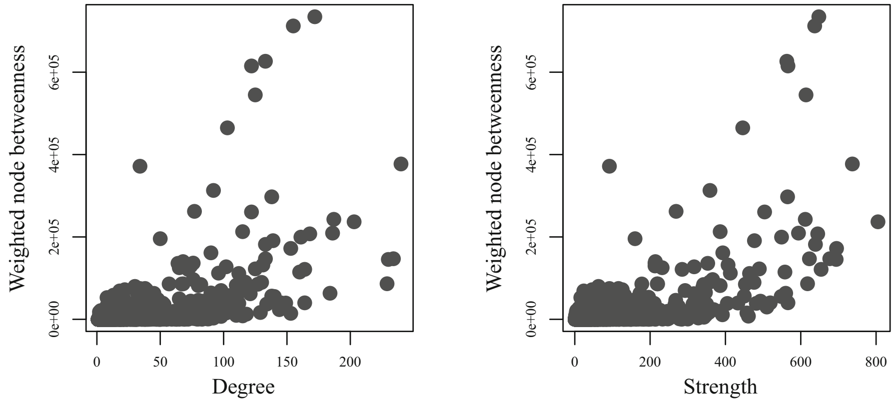


Fig. 12 Correlation between weighted node betweenness and degree (*left*), weighted node betweenness and strength (*right*) for each airport in the network

node betweenness and strength, i.e., the correlation coefficients are equal to 0.58 and 0.64, respectively. We easily observe that the most connected airports do not necessarily have the largest weighted node betweenness centrality and vice versa [these results are similar to the published results of Guimera and Amaral (2004b), Barrat et al. (2005)]. Additionally, some airports which are very central according to a given centrality become peripheral according to other criteria. For example, Amsterdam Airport Schiphol (AMS) is the fourth highest connected airport in terms of degree centrality, but ranks only 43rd in terms of weighted node betweenness. On the other hand, Ted Stevens Anchorage International Airport (ANC) is the eighth highest weighted node betweenness airport but does not even appear in the ranking list of the top airports that have high degree and strength.

Topological measures that do not take notice of weights, such as degree centrality may miss the economical dimension of the network since weights reflect traffic and economic realities. Strength centrality is the characteristic that surmounts the drawbacks of degree centrality by paying attention to weights. Nevertheless, both degree and strength consider only local effects of complex systems. On the other hand, weighted node betweenness centrality is the measure that pinpoints the most important nodes in each geographical zone by taking into account both weights and global effects. In particular, the weighted node betweenness appears as a balanced measure which combines traffic importance with topological centrality. Hence, investigating several centrality measures of a system is essential to characterize its properties.

In a large variety of real networks, centrality measures are broadly distributed reflecting the strong structural heterogeneity of the networks. We observe the similar behavior of the WAN for the degree, strength, and weighted node betweenness, as shown in Fig. 13.

Similar to the definition of weighted node betweenness centrality, we define weighted edge betweenness centrality of an edge as the total number of the shortest

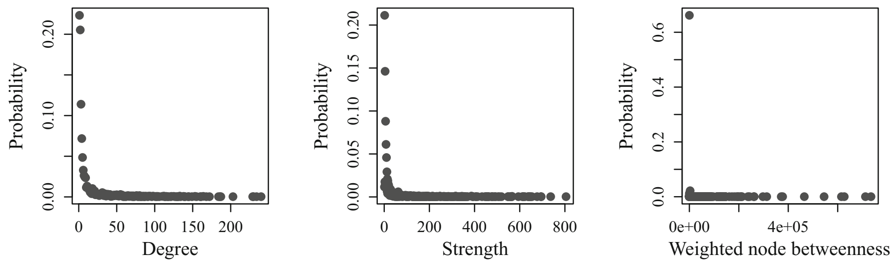


Fig. 13 Probability distribution function of the degree (*left*), strength (*center*) and weighted node betweenness (*right*) of the WAN

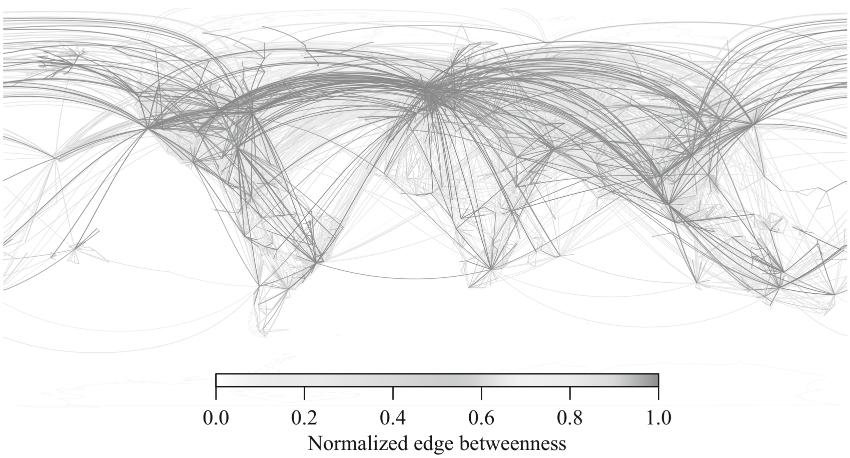
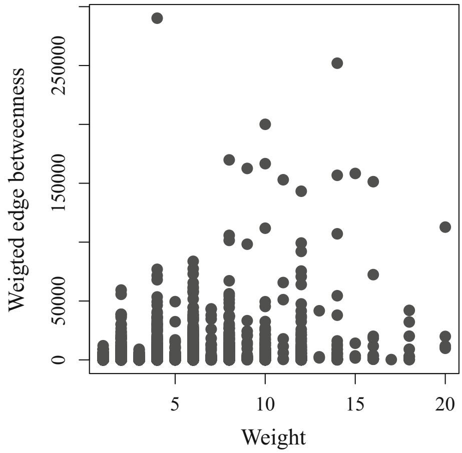


Fig. 14 Weighted edge betweenness distribution map of the WAN. *Darker lines* represent edges that have higher weighted edge betweenness. The *legend* represents the normalized weighted edge betweenness where the maximum value is supposed to be unity

weighted paths that pass through that edge. Edges which are frequently used as parts of shortest weighted paths, may therefore have high weighted edge betweenness. Figure 14 shows the weighted edge betweenness distribution map of the WAN. As observed, Hearthrow (LHR) becomes the largest gateway of the network since most traffic is routed through it.

Figure 15 shows the lack of correlation between the weighted edge betweenness and the edge weight of each connection in the WAN, indicating that connections which have a large weight do not necessarily have a large weighted edge betweenness, and vice versa. Particularly, the largest betweenness edge is the one that connects Chicago O'Hare International Airport (ORD) and Ted Stevens Anchorage International Airport (ANC) although only a few types of flights are operated between them, i.e., $w_{\mathrm{ORD,ANC}} = 4$.

Fig. 15 Correlation between weight and weighted edge betweenness of each connection in the WAN



3 Vulnerability analysis of the WAN

With the development of air transportation systems, it becomes more and more important to grasp the dynamics and vulnerability of the networks. Particularly, the WAN generalizes several questions concerning the vulnerability of weighted networks. In this section, we propose a model to gain insight into the response behavior of the WAN to catastrophic events.

3.1 Model

Recently, a great deal of attention has been devoted to the analysis of the resilience of both artificial networks and real networks. The analysis of network resilience has been largely investigated for unweighted networks (Albert et al. 2000; Callaway et al. 2000; Holme et al. 2002). In this context, a widely used indicator for assessing network resilience is the topological integrity, which is often measured by the size of the largest connected component of the network. This maps network resilience to a percolation model, where the aim is to determine the number of nodes that must be eliminated through a specific elimination strategy to reach a threshold at which the network disintegrates into small components. Considering topological integrity as network resilience, heterogeneous networks with a scale-free degree distribution reveal their robustness to situations in which nodes are eliminated randomly. On the other hand, the targeted attacks on nodes following their degree rank is extremely damaging, leading to the total fragmentation of the network by eliminating only a small fraction of nodes (Albert et al. 2000). Moreover, the elimination of nodes that have large node betweenness centrality typically leads to an even faster fragmentation (Holme et al. 2002). However, prior research showed that the theoretical percolation approach that focuses on the topological integrity is not suitable for the study of aviation systems (Meza et al. 2013).

Another indicator to assess network resilience is the systematic changes in the distribution form of network standard centrality measures, such as degree, strength,

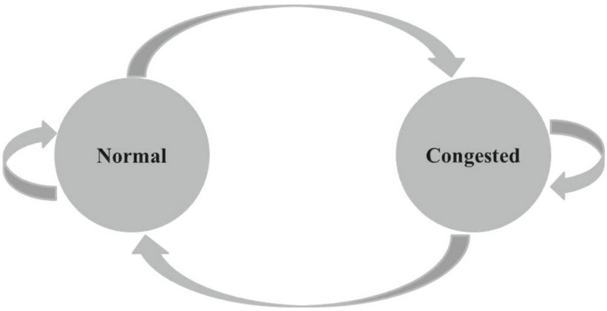


Fig. 16 Simple model that describes the two states of airports in the network. Normally, airports are in "normal" state. If a breakdown occurs at one or more airports because of an external cause, the most efficient paths between a certain number of remaining airports will change and the load will be redistributed. Airports that are not overloaded, remain in "normal" state, while airports with loads that exceed their capacities immediately become congested. In subsequent time steps, congested airports are less used by passengers than usual, and they may remain in a "congested" state at first but recover gradually from the "congested" to the "normal" state. The system recursively loops in this manner until the network converges to a stable state, where all airports can finally handle their loads

and betweenness. However, prior research also found that the functional form of degree, strength, and betweenness distributions are surprisingly stable to disruptions in the case of the WAN (Meza et al. 2013). Therefore, changes in the referred standard centrality distributions are also unsuitable for assessing the impact of disasters on the network.

We know that many real networks are not only specified by their topology, but also by the dynamical properties of processes taking place on them, such as the traffic among components of the system. The functionality of the network can be temporarily damaged in terms of traffic, even if the physical structure is still globally well connected. Therefore, considering network resilience by taking traffic into account may be more reasonable than topological factors such as integrity or changes in centralities distributions. According to this fact, we provide a binary status model, namely normal or congested model, in order to successfully capture the dynamics and vulnerability of the WAN.

In most real networks, the breakdown of one or more nodes may be sufficient to cause the entire system to collapse. This is based on the fact that the breakdown of a single node, so that the node is assumed to be eliminated from the network, not only has direct consequences on the network performance, but can also cause overloads on other nodes, thus generating a cascading effect. To take into account this phenomenon, we model the dynamics of the WAN using plausible assumptions about the load and overload of airports (Latora and Marchiori 2001; Motter and Lai 2002; Kinney et al. 2005).

As mentioned in Sect, 2.2, the proximity, p_{ij}, of direct connection between airport i and airport j is defined as its reciprocal coupling weight, $p_{ij} = 1/w_{ij}$. Thus, the proximity captures the intuitive notion that strongly (or weakly) coupled airports are close to (or distant from) each other. We define the length of a path that starts at i_0 and ends at i_k, passing through intermediate airports $i_1, i_2, \ldots, i_{k-1}$ as

$l_{i_0 \to i_k} = \sum_{n=0}^{k-1} p_{i_n i_{n+1}} = \sum_{n=0}^{k-1} 1/w_{i_n i_{n+1}}$. Given the set $S(i, j)$ of paths that connect airports i and j, we define the distance from i to j, $d_{i \to j}$, as the shortest path length from i to j, i.e., $d_{i \to j} = \min_{S(i,j)} l_{i \to j}$. Since the network is symmetric, we have $l_{i \to j} = l_{j \to i} = l_{ij}$, and therefore $d_{i \to j} = d_{j \to i} = d_{ij}$. The efficiency of a path between i and j is then defined as $\epsilon_{ij} = 1/d_{ij}$. According to the definition, the path that has the shortest path length, becomes the most efficient one among all paths that connect i and j.

We consider a model in which we assume that passengers travel with equal probability from any departing airport to any destination airport by following the most efficient paths in the network. The assumption is plausible since it agrees with real aviation systems. We define the load, $L_i(t)$, of airport i as the total number of most efficient paths between any two airports that pass through i at time step t, i.e., the weighted node betweenness of airport i,

$$L_i(t) = \text{Load (weighted node betweenness) of airport i at time step t} \tag{1}$$

We assign to each airport i a capacity C_i, which is simply assumed to be proportional to the initial load $L_i(0)$,

$$C_i = \alpha L_i(0) \tag{2}$$

where α is the tolerance parameter, indicating the maximum load that an airport can handle. Here, the tolerance parameter α also accounts for the budget of network construction or resource allocation.

If a breakdown occurs at one or more airports because of an external cause, such that the airports cannot work at all, and eliminated from the network, the most efficient paths between a certain number of remaining airports will change, leading to load redistribution within the network. In other words, the breakdown of one or more airports naturally leads the passengers to choose another routes to their desired destination, causing a global change of flow within the network. This might lead to a situation in which a certain number of airports are forced to carry more load than their capacity, thus becoming congested (overloaded), which in turn would result in a degradation of their performance. Such performance degradation subsequently modifies the most efficient paths, redistribute the load on the network, and cause new airports to become congested. If the influence caused by the initial breakdown is small, congestion will remain local. On the other hand, if the influence caused by the initial breakdown is sufficiently large, congestion will spread throughout the system, causing an avalanche effect and remarkable degradation of system performance. Congested airports can recover their normality if their load decreases below their capacity as a result of load redistribution. Eventually, the degradation stabilizes when all remaining airports can handle their respective loads (Fig. 16).

When the network is in a normal state, all existing connections between any pairs of airports are equal to their initial weights, meaning that all connections are working perfectly. However, an overload at some airports caused by the occurrence of the initial breakdown leads to congested airports that cannot operate as usual, leading to the degradation of performance (weights) of all direct connections to the congested airports. The performance (weight) of the connection between airports i and j, evolves over time as follows,

$$w_{ij}(t+1) = \begin{cases} w_{ij}(0)\dfrac{C_i}{L_i(t)}, & \text{if} \quad \mathrm{L_i(t) > C_i} \\ w_{ij}(0), & \text{otherwise} \end{cases} \tag{3}$$

where j is the first neighbor of i. $w_{ij}(0)$ is the initial weight (performance) of the connection between i and j. $w_{ij}(t+1)$ is the weight (performance) of the connection between i and j at the next time step.

In other words, when airport i becomes congested at time step t, i.e., $L_i(t) > C_i$, it is assumed that the performance of the connections from i to all of its first neighbors at the next time step decreases linearly with the ratio of overload, $L_i(t)/C_i$. Otherwise, all connections from i to its first neighbors work perfectly as normal, i.e., $w_{ij}(t+1) = w_{ij}(0)$.

The performance of the WAN (network efficiency) at the time step t is defined as the average efficiency of all paths in the network,

$$E(t) = \frac{1}{{}_N C_2}\sum_{i \neq j} \epsilon_{ij}(t) = \frac{2}{N(N-1)}\sum_{i \neq j}\frac{1}{d_{ij}(t)} \tag{4}$$

where N is the number of airports, $\epsilon_{ij}(t)$ is the efficiency of the path that connects airport i and j, $d_{ij}(t)$ is the distance between i and j, and $E(t)$ is the network efficiency at the time step t.

In order to quantify how well the network operates before and after the breakdown of an airport i, we define the damage, D, as the normalized network efficiency loss,

$$D = \frac{E(0) - E_i}{E(0)} = 1 - \frac{E_i}{E(0)} \tag{5}$$

where $E(0)$ is the efficiency of the network before breakdown, i.e., all airports are in normal state, and E_i is the efficiency of the network after the breakdown of airport i.

Using this model, we can follow the dynamical response of the system subjected to breakdowns. In particular, we can capture how the breakdown in one location can propagate and have consequences over the entire network.

3.2 Results

In order to mitigate the damage caused by initial breakdowns, preventing overload on remaining airports is significant. Intuitively, the most effective method to do this is to increase the tolerance parameter, α, which is defined in Eq. 2, as much as possible. If the tolerance parameter α is sufficient large, the network might not result in a cascading effect since airports have sufficient capacities to handle their increased loads, and network efficiency would remain unaffected by the breakdown of some airports. However, the increase of α is often limited by cost, making it difficult to construct a network with a very large α. In fact, infrastructures that

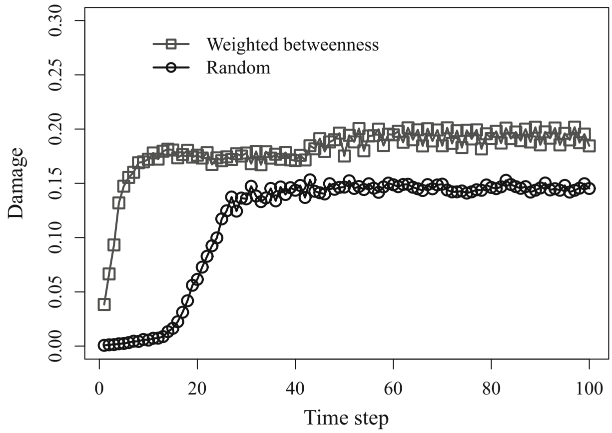


Fig. 17 Evolution of the damage over time in the case of $\alpha = 1.05$. The *lines with squares* and *circles* represent the evolution of the damage caused by the elimination of a randomly chosen airport, or highest weighted node betweenness airport, respectively

support the flow of energy, information, or people at all scales often operates near maximum capacity (Vespignani 2009). According to this fact, to assess the resilience of the WAN, we consider that α takes values in the range of $\alpha \in [1, 2]$. $\alpha = 1$ represents the extreme case where capacities of all airports are assumed to be equal to their initial loads. On the other hand, the capacities of all airports are assumed to be double their initial loads in the case of $\alpha = 2$.

We simulate the initial breakdown by eliminating one or more airports randomly, subjected to random breakdown–random airports are chosen to be eliminated, or intentionally, subjected to intentional breakdown–central airports are chosen to be eliminated, and monitor the progression of overloading dynamics on remaining airports.

Figure 17 shows the evolution of the damage caused by the breakdown of a random airport (circle) and an airport with the highest weighted node betweenness (square), over time in the case of $\alpha = 1.05$. As observed, the damage fluctuates but finally converges to a stable state. We determine the median of the series of damage as a single representative damage of the WAN.

Figure 18 shows the relationship between the tolerance parameter α and the damage for a random breakdown scenario. As observed, increasing α is an efficient way to reduce the damage of the random breakdown scenario. The damage is on average almost completely absorbed when $\alpha \geq 1.1$. The elicited result in the right figure of Fig. 18 additionally shows the effectiveness of increasing the tolerance parameter to random breakdown, i.e., the larger the value of α, the lower the probability that a large damage will occur.

For intentional breakdowns, we propose a series of topological and weight-depending centrality measures that can be used to identify the most important nodes of the network, such as degree, strength and betweenness. We first consider a single breakdown of the high-centrality airport, i.e., Heathrow (LHR) for the highest

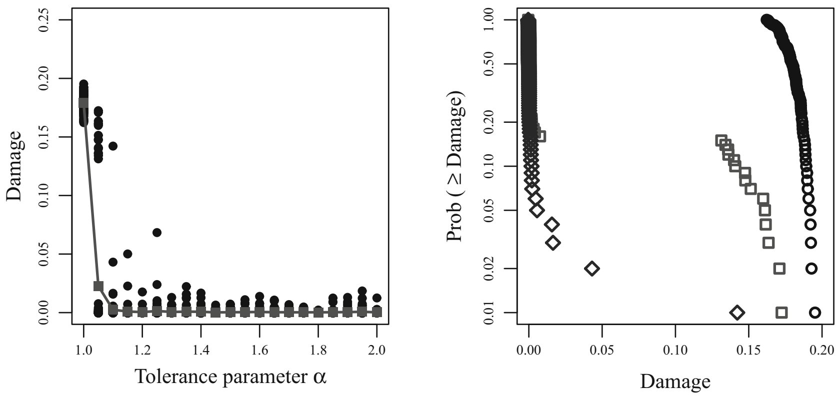


Fig. 18 (*Left*) relationship between the tolerance parameter α and the damage for a random breakdown scenario. Each *dot* represents the damage to the network after eliminating a single airport randomly. We conducted 1000 simulations for each value of the tolerance parameter α to obtain the data. The *solid squares* represent average values for each tolerance parameter. (*Right*) cumulative damage distribution after random breakdown of a single airport for different tolerance parameter schemes: $\alpha = 1$ (*circles*), $\alpha = 1.05$ (*squares*), and $\alpha = 1.1$ (*diamonds*)

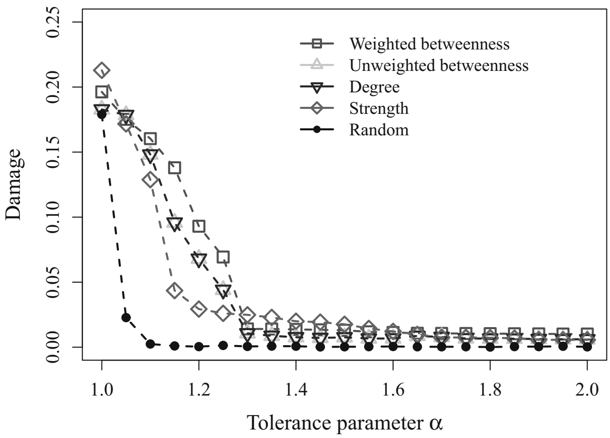


Fig. 19 Comparison of the damage as a result of intentional breakdown and random breakdown. Airport with highest weighted node betweenness (*squares*), unweighted node betweenness (*upward triangles*), degree (*downward triangles*), and strength (*diamonds*) is chosen for the case of intentional breakdowns, while randomly (*dots*) for the case of random breakdown. Note that the result of random breakdown is obtained from the *left* figure of Fig. 18

weighted node betweenness airport, Charles De Gaulle (CDG) for the highest unweighted node betweenness, and also the highest degree airport, and Beijing Capital International Airport (PEK) for the highest strength airport.

As observed in Fig. 19, eliminating the most central airports results in significant damage to the entire network as compared to eliminating a single random airport. We only need to set $\alpha = 1.1$ to completely eliminate damage in the case of random

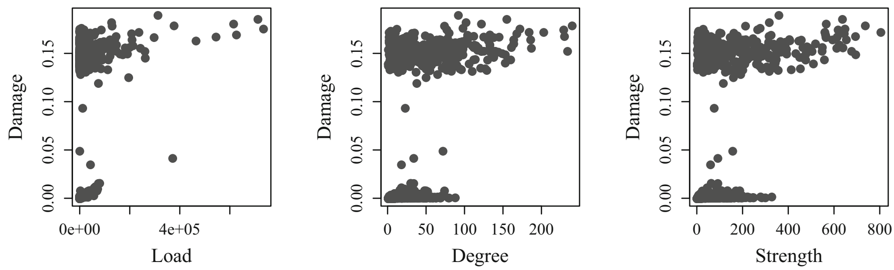


Fig. 20 Correlation between load and damage (*left*), degree and damage (*center*), strength and damage (*right*), for each airport in the case of $\alpha = 1.05$. We observe no correlation between centrality measures and damage, i.e., when α is excessively small, most airports bring significant damage to the system, indicating that not only central airports but peripheral airports also are excessively damaging

breakdown. However, a higher α value is required for intentional breakdown scenarios. The breakdown of these central airports can result in a loss of network efficiency that is as high as 20 %. In the WAN, central airports play a crucial role because they connect to other major airports to guarantee the connectivity of the network. They are usually selected as pathways to reduce the distance between each pair of airports, and their absence thus increases the efficiency of the network. The breakdown of one of them, e.g., Heathrow (LHR), leads to major changes in network connectivity and traffic flow, causing a massive degradation of network efficiency. Hence, the existence of central airports conversely becomes a weak point of the network that could be targeted by intentional attacks. In the WAN, Heathrow (LHR) might be the most important airport according to a number of criteria. It has the highest weighted node betweenness centrality. The removal of Heathrow (LHR) alone is enough to cause the highest system performance degradation compared to an attack targeting another central airport of the network. However, we can think of a core of central airports that play an important global role rather than a single, most important airport, with Heathrow (LHR) being the most prominent entity of this core.

We further investigate the importance of the individual airports in the network by answering the question "How does the breakdown of an airport directly affect the entire system?". Since most infrastructures often operate near their maximum capacity, we fix the tolerance parameter with a series of small values, $\alpha = 1.05$, $\alpha = 1.1$, $\alpha = 1.2$, and calculate the damage of each airport, when it breaks, to the network. Figures 20, 21, 22 show the relationship between the load, degree, and strength, of each airport and its damage. As a result, high load, high degree as well as high strength airports have the strongest impact on the network performance for all the cases of the investigated tolerance parameter α.

The top damaging airports for all cases of the tolerance parameter α, as listed in Table 1, and their corresponding normilized centrality measures are shown in Fig. 23. The centrality measures of each airport are obtained by dividing them by the maximum values of corresponding centrality measures for normalization.

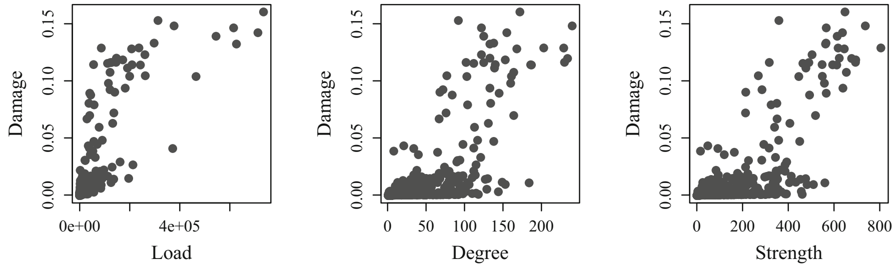


Fig. 21 Correlation between load and damage (*left*), degree and damage (*center*), strength and damage (*right*), for each airport in the case of $\alpha = 1.1$. We observe a weak correlation between centrality measures and damage, i.e., central airports are damaging

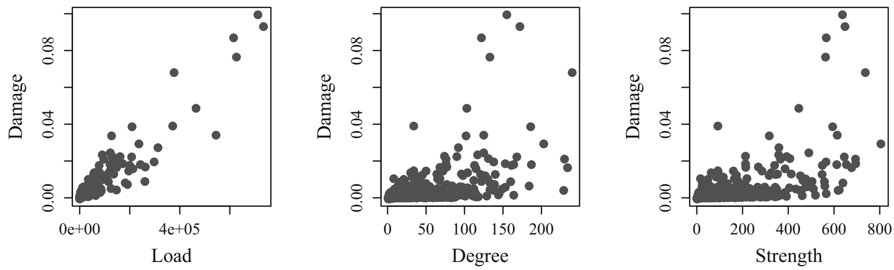


Fig. 22 Correlation between load and damage (*left*), degree and damage (*center*), strength and damage (*right*), for each airport in the case of $\alpha = 1.2$. We clearly observe a strong correlation between load and damage, but a weak correlation between degree (strength) and damage

Table 1 List of top damaging airports for different values of the tolerance parameter α

Rank	$\alpha = 1.05$	$\alpha = 1.1$	$\alpha = 1.2$
1	GRU (Sao Paolo)	LHR (London)	JFK (New York)
2	JFK (New York)	GRU (Sao Paolo)	LHR (London)
3	OSL (Oslo)	CDG (Paris)	LAX (Los Angeles)
4	LAX (Los Angeles)	LAX (Los Angeles)	ORD (Chicago)
5	CDG (Paris)	JFK (New York)	CDG (Paris)
6	BOG (Bogota)	SIN (Singapore)	NRT (Tokyo)
7	GVA (Geneva)	MIA (Miami)	ANC (Anchorage)
8	LHR (London)	ORD (Chicago)	DME (Moscow)
9	GSP (Greenville)	PEK (Beijing)	SIN (Singapore)
10	LPA (Gran Canaria)	AMS (Amsterdam)	OSL (Oslo)

We can recognize the appearance of very crucial airports in the figure, such as Heathrow (LHR), Los Angeles International Airport (LAX), Charles de Gaulle (CDG), John F. Kennedy International Airport (JFK), Chicago O'Hare International

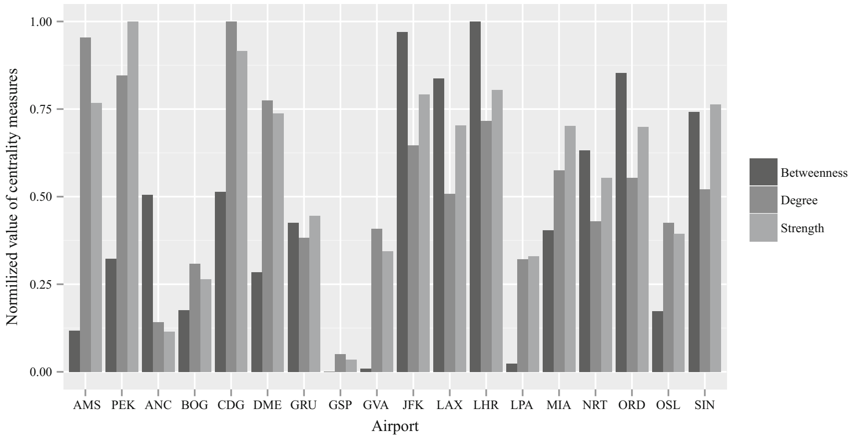


Fig. 23 The top damaging airports listed in Table 1 and their corresponding normilized centrality measures

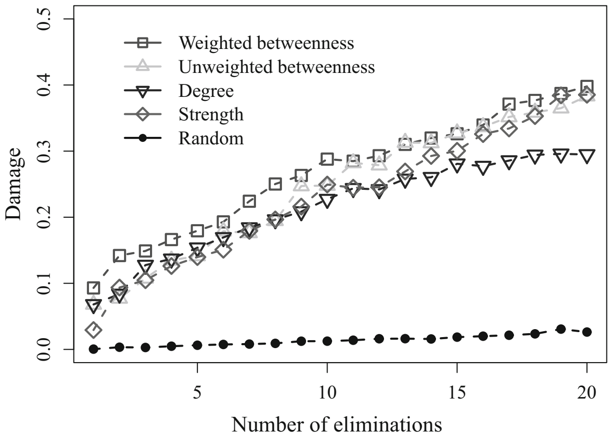


Fig. 24 Comparison of the damage as a result of multiple intentional breakdown and random breakdown, for the case of $\alpha = 1.2$. A certain number of airports with highest weighted node betweenness (*squares*), unweighted node betweenness (*upward triangles*), degree (*downward triangles*), and strength (*diamonds*) are chosen for the case of intentional breakdowns, while randomly (*dots*) for the case of random breakdown. Naturally, as the number of eliminations increase, the damage also increases linearly for intentional breakdown scenarios, but remains almost unchanged for random breakdown

Airport (ORD), Changi Airport (SIN), Miami International Airport (MIA), and Narita International Airport (NRT), which have large degree, strength, as well as betweenness centrality. The breakdown of these airports naturally elicits the highest damage to the WAN. Besides, Amsterdam Airport Schiphol (AMS) and Beijing Capital International Airport (PEK) apprear in the list since they both have large degree and strength, although their betweenness values are low. On the other hand, Ted Stevens Anchorage International Airport (ANC), which has only high betweenness, but low degree and strength, is also listed as one of the most damaging airports. We also observe some other airports emerging in the list, such as Oslo Airport (OSL) in Norway, Guarulhos Airport (GRU) in Brazil, El Dorado International Airport (BOG) in Colombia, Geneva Airport (GVA) in Switzerland, Las Palmas de Gran Canaria Airport (LPA) in Spain, and the most surprising one is the Greenville International Airport (GSP) in the United States, which has a very low degree, strength, and betweenness.

We then study the behavior of damage measures in the presence of multiple random, and different intentional breakdown scenarios. In this case, a number of airports are eliminated concurrently according to their initial ranking calculated based on the integral network (Fig. 24).

4 Conclusion

Complex systems in nature and society often contain an enormous amount of information. Extracting meaningful information from such systems is not a trivial task because information is often masked by their complexity. Recently, complex network theory has become one of the most efficient frameworks for this issue.

Using network analyses, we investigated the WAN. The data was obtained from the open source project named OpenFlights. We first clarified the topological characteristics of the network. We found that the network exhibits a heterogeneous distribution of degree, strength, and betweenness, covering many orders of magnitude. The WAN is a small-world, with an obvious community structure. Our results revealed that the network has been developed with Heathrow airport at its the center, connecting other central airports, in the effort to form the core of the network.

We then investigated the dynamics of the network, with emphasis on the understanding of the resilience of the network to extreme events. Since the topological integrity, which has been widely used to assess the vulnerability of complex networks is not suitable for the study of aviation networks, such as the WAN, we introduced a binary status model, which took weights and traffic into account as its key features. We first characterized relevant topological and weighted centrality measures and then used these quantities as selection criteria for the elimination of airports. We showed that centrality driven breakdowns are capable to lead to a massive loss of network efficiency, even at very low levels of damage in the connectivity pattern. That is to say, central airports have a strong impact on network resilience, and the breakdown of one of them results in serious consequences to the entire system. This provides a clear identification of the airports which need the highest level of protection against such attacks. Our analysis provided a closer look on how the worldwide air transportation system has been constructed and how it responds to catastrophic events. The results may contribute to further developments in air transportation systems studies in general.

References

Albert R, Barabási A (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47

Albert R, Jeong H, Barabási A (2000) Error and attack tolerance of complex networks. Nature 406:378–382

Amaral L, Scala A, Barthélémy M, Stanley H (2002) Classes of small-world networks. Proc Natl Acad Sci 97(21):11149–11152

Bagler G (2008) Analysis of the airport network of india as a complex weighted network. Phys A 387(12):2972–2980

Bagrow J, Wang D, Barábasi A (2011) Collective response of human populations to large-scale emergencies. PLoS One 6:e17680

Barabási A, Albert R, Jeong H (2000) Scale-free characteristics of random networks: the topology of the world-wide web. Phys A 281:69–77

Barrat A, Barthélémy M, Vespignani A (2005) The effects of spatial constraints on the evolution of weighted complex networks. J Stat Mech 2005:P05003

Blondel V, Guillaume J, Lambiotte R, Lefebvre R (2008) Fast unfolding of community hierarchies in large networks. J Stat Mech 2008:P10008

Buldyrev S, Parshani R, Paul G, Stanley H, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. Nature 464:1025–1028

Bye B (2011) Volcanic eruptions: science and risk management. http://www.science20.com/planetbye/volcanic_eruptions_science_and_risk_management-79456. Accessed 11 Aug 2015

Callaway D, Newman M, Strogatz S, Watts D (2000) Network robustness and fragility: Percolation on random graphs. Phys Rev Lett 85:5468

Chassin D, Posse C (2005) Evaluating north american electric grid reliability using the barabási-albert network model. Phys A 355(2):667–677

Colizza V, Barrat A, Barthélémy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci 103:2015–2020

Colizza V, Barrat A, Barthélémy M, Valleron A, Vespignani A (2007) Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. PLoS Med 4:e13

Danon L, Duch J, Díaz-Guilera A, Arenas A (2005) Comparing community structure identification. J Stat Mech 09:P09008

Epstein J, Goedecke D, Yu F, Morris R, Wagener D et al (2007) Controlling pandemic: the value of international air travel restrictions. PLoS One 2:e401

Erdös P, Rényi A (1959) On random graphs. Publ Math 6:290–297

Freeman L (1977) A set of measures of centrality based upon betweenness. Sociometry 40:35–41

Guimera R, Amaral L (2004a) Functional cartography of complex metabolic networks. Nature 433(7028):895–900

Guimera R, Amaral L (2004b) Modeling the world-wide airport network. Eur Phys J B 38:381–385

Guimera R, Mossa S, Turtschi A, Amaral L (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc Natl Acad Sci 102(22):7794–7799

Holme P, Kim B, Yoon C, Han S (2002) Attack vulnerability of complex networks. Phys Rev E 65:056109

Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci 101:15124–15129

Kinney R, Crucitti P, Albert R, Latora V (2005) Modeling cascading failures in the north american power grid. Eur Phys J B 46:101–107

Krapivsky P, Redner S (2002) A statistical physics perspective on web growth. Comput Netw 39:261–276

Latora V, Marchiori M (2001) Efficient behavior of small-world networks. Phys Rev Lett 87(19):198701–198704

May R, Levin S, Sugihara G (2008) Complex systems: ecology for bankers. Nature 451:893–895

Meza O, Grady D, Thiemann D, Bagrow J, Brockmann D (2013) Eyjafjallajökull and 9/11: the impact of large-scale disasters on worldwide mobility. PloS One 8:e69829

Motter A, Lai Y (2002) Cascade-based attacks on complex networks. Phys Rev E 66:065102(R)

Newman M (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69:066133

OpenFlights (2014) http://openflights.org

Tran Q, Namatame A (2014) Design robust networks against overload-based cascading failures. Int J Comput Sci Artif Intell 4(2):35–44

Verma T, Araújo N, Herrmann H (2014) Revealing the structure of the world airline network. Sci Rep 4:5638

Vespignani A (2009) Predicting the behavior of techno-social systems. Science 325:425–428

Wang J, Mo H, Wang F, Jin F (2011) Exploring the network structure and nodal centrality of china's air transport network: a complex network approach. J Transp Geogr 19(4):712–721