Structure and Resilience of Networks: Airlines in USA

Daniel R. Wuellner,†‡ Soumen Roy,‡†∗ and Raissa M. D’Souza‡†∗

1Graduate Group in Applied Mathematics, University of California, Davis, CA 95616
2Center for Computational Science and Engineering
3Department of Mechanical and Aeronautical Engineering, University of California, Davis, CA 95616

Airline networks are methodically designed, engineered systems with structures that can vary considerably amongst distinct carriers. We analyze the flight networks of the seven largest passenger carriers in the USA, characterizing their topological structures and resilience properties. Structurally, we find the degree distribution of several of the networks, including the aggregate over the seven carriers, are well described by simple exponential distributions. Functionally, we find that networks with “large” k-core structures possess extreme resilience to both random and targeted removal of either airports (nodes) or flight paths (edges), with no significant increase in estimated travel time. Similar results are obtained when the targeted removal of airports is by degree or by betweenness, albeit the effect of the latter causes a faster breakdown of each carrier’s network. We introduce a rewiring scheme that preserves total number of daily flights and gate requirements while enhancing k-core structures and resilience (i.e., k-core resilience), which should augment our understanding of building resilient networks in general. Finally, our findings suggest that point-to-point topologies have larger k-core structures, providing new insight into the long standing debate on the optimality of such layouts when compared to hub-and-spoke arrangements.

Modern civilization relies on the efficient design, operation and maintenance of a variety of, often interdependent, infrastructure networks, with aviation networks being an archetypical example. Air travel is a principal means of fast and effective transportation of people and goods over large distances across countries or continents, around the globe. It is critical to the functioning of countries and the world economy as a whole. The aggregate network of air travel worldwide built by considering all flights amongst all destinations throughout the globe (the World Airline Network) has been the subject of much recent study [1, 2, 3, 4, 5]. The focus has been on analysis of overall flow patterns and the consequences for the spread of global epidemics [4], as well as identifying the overall importance of individual airports [5]. An aggregate level analysis has also been carried out on the airline networks of a few individual countries which show several similarities to the WAN, namely “scale-free” and small-world characteristics [6, 7].

Our interest is not in overall flow, but in design and operation of critical infrastructure. The aggregate view of air travel is built up from a collection of co-existing airline networks, operated independently by distinct entities. Each independent operator must build a well-connected and economically successful airline network which is resilient to random or systematic vagaries, ranging from acts of nature to terrorism. Furthermore, an individual airline has direct control only over their own network. We investigate what changes to network structure of an individual carrier can lead to improved efficiency and resilience.

Herein we analyze and contrast the network structures of the seven largest passenger airlines in the United States of America (USA). When broken down into networks of individual carriers, we show that many features differ from those of the aggregate view. Small-world attributes are exhibited by the networks of all the carriers, yet, rather than scale-free power law distributions, we find that the distribution in airport connectivity is better described by either a simple exponential decay or a cumulative log-normal distribution. More pronounced than distribution in connectivity, we find that Southwest Airlines (SW) stands apart from the other six carriers by its k-core structure (defined in detail below) and its extreme resilience to random or targeted deletion of nodes (airports) or edges (flight paths). Edge deletion corresponds to, for instance, weather preventing travel between two airports, while node deletion corresponds to temporary closure of an airport. SW has essentially built a core network, comprising more than half of its overall destinations, which is a dense mesh of interconnected high-degree (i.e., “hub”) airports. We explore the interplay between placing hubs in the periphery versus the core of a network and introduce a general network rewiring process which keeps constant the demand on each node and the amount of flow between nodes, that enhances the k-core structure and increases resilience of a network.

One fundamental consideration when building a new airline network, or expanding an existing one, is whether to prefer “point to point” (PP) or “hub and spoke” (HS) connectivity. Of course, aside from functional considerations, there are also financial reasons why one type may be preferred over the other. In the PP scenario, a passenger can travel on a direct non-stop flight to a range of destinations at shorter distances, but to travel considerable lengths has to transit and take multiple flights. In the HS scenario, in contrast, a passenger can travel non-stop only to a few central hubs, and from there transit to
their final destination (almost always requiring two-hops unless the hub is their ultimate destination). Rigorous analysis shows asymptotic optimality of HS models for spatial transportation networks with transfer costs [8]. Analytic arguments, backed by numerical simulations indicate that HS architectures are optimal for travelers wishing to minimize the number of connecting flights required instead of overall distance travelled [9]. Inspired in part by studies on airport networks, a general model of weighted networks via an optimization principle was proposed in which a clear spatial hierarchical organization, with local hubs distributing traffic in smaller regions, emerges as a result of the optimization [10]. Thus there seems to be a growing consensus in the literature regarding HS structures arising out of optimization of resources. However, real-world structures also need to be resilient and robust apart from having an optimal distribution of resources. In this work, we show that PP structures can be much more resilient than HS structures.

The majority of the larger airlines operating in the USA at present predominantly follow the HS pattern. This was not the case prior to 1978, when the USA Federal Government regulated air traffic, with special attention paid to ensure lower traffic (and hence lower-profit) routes were not ignored [11]. Such regulations effectively enforced PP architectures. Once the Federal Government deregulated the airline industry in 1978, most airlines gradually shifted to their current HS pattern, apparently finding such a HS architecture more desirable. A significant exception was Southwest Airlines (SW), which continued to build a PF system. As of the end of 2007, SW is the largest airline (by number of domestic passengers and domestic departures) not only in the United States, but also in the entire world [12]. Its sheer size together with the extremely consistent economic success of SW [13] provide strong evidence for the efficacy of PP networks. Ryanair and Easyjet are two examples of successful PP carriers in Europe [14]. Innovative management policies have played an important part in the success of SW and are studied extensively in business literature (for instance, Ref. [15]). Here our focus is on network infrastructure with a view to efficiently design or restructure individual airline networks so they are well-connected, robust and resilient to disturbances.

Our findings are of theoretical interest yet should also be relevant to entities engaged in designing or altering large-scale airline networks. For instance, expanding airline networks in developing nations need to consider the tradeoffs between PP and HS architectures. Airline carriers wanting to shrink an airline (i.e., eliminate flights with minimal impact), for instance as rising fuel prices require enhanced operating efficiency, need systematic approaches for identifying the appropriate manner. Finally, individual carriers need metrics to assess the quality of network infrastructure resulting from a merger with another carrier.

CONSTRUCTING THE AIRLINE NETWORKS

All certificated U.S. air carriers are required to file monthly reports with the U.S. Department of Transportation, Bureau of Transportation Statistics, detailing information on every flight segment flown during that month. This information is maintained in a public database, the “Air Carrier Summary: T-100 Domestic Segment (U.S. Carriers)” [16]. From this database we download information on every “scheduled passenger service” class flight segment flown by each of the seven largest U.S. passenger carriers for the entire 2007 calendar year. To isolate the structure of passenger carriers, we neglect the small fraction of flights by these carriers which are designated by the “cargo” (only class) or “non-scheduled passenger service” (charter) class. Yet, in order to compare the structure of a passenger carrier to a cargo-only air carrier, we also download all flights flown during the 2007 calendar year by two cargo-only carriers (Federal Express and United Parcel Service).

The seven largest US passenger airlines (by number of passengers flown) are in order, Southwest (SW), American Airlines (AA), Delta (DL), United Airlines (UA), Northwest (NW), US Airways (US), and Continental (CO). These seven carriers account for 61.6% of all passengers enplaned in 2007. We construct two distinct views of the network for each carrier, one which captures the connectivity (i.e., which airports are connected via direct flights), the other captures both connectivity and the total traffic flow between airports. For carrier $c$ we denote the first view by $G^c(N^c, E^c)$, and the latter $W^c(N^c, E^c)$. The vertices, $N^c$, are the same in both views and are the set of all airports listed as an origin or destination airport for carrier $c$ which are also included in that carrier’s list of official domestic destinations as stated on June 2008. This additional data “scrubbing” step eliminates airports used only for diverted aircraft which have substantially fewer numbers of flights than the official airports and otherwise introduce noise. We first construct the complete flight history for 2007, $W^c(N^c, E^c)$. A directed edge is added from each origin airport to its destination airport, with edge weight equal to the total number of flight segments from that origin to that destination flown by carrier $c$ in 2007. The unweighted (binary) version of this graph is $G^c(N^c, E^c)$, and is the equivalent of the “route map” for that carrier. It is the collection of airports serviced, with an edge between two airports if there is a direct connection between them. The $G^c(N^c, E^c)$ view focuses only on the connectivity of the network, while the $W^c(N^c, E^c)$ includes also the actual traffic along those connections.

We consider both node degree and strength. The out-degree of node $i$, $q_i^{out}$, is the number of edges originating at that airport in $G^c(N^c, E^c)$ (number of distinct destinations that can be reached directly from $i$). The in-degree, $q_i^{in}$ is the number of edges terminating at $i$ (number of distinct incoming origins). We find $q_i^{in} \sim q_i^{out}$ (airports are almost always connected in both directions)
We also consider the “strength”, $s_i$ of the $i$'th node, defined as in Ref. [3]. The in-strength (out-strength) of an airport is the total number of flights landing (departing) there, for that specific carrier, in 2007. Formally, the in-strength (out-strength) is the sum over all edge weights in $W^\text{in}(N^c, E^c)$ for edges terminating (originating) at that node. We find $s^\text{in}_i \sim s^\text{out}_i$, so for the remainder we treat all edges as undirected and set the undirected edge weights in $W^c$ to be the maximum edge weight in either direction.

In addition to the network structures of individual carriers, the aggregate airline network of the USA is of interest. We study three different aggregate views: Agg7, which is the aggregate over the seven largest passenger carriers; AggPass, which is the aggregate over all “scheduled passenger service” class flights flown during 2007 by all carriers (not just the seven largest); finally, AggAll is the aggregate over every single flight segment flown during 2007, regardless of service class or carrier. Formally, to construct the distinct aggregate views we take the union over all nodes and edges for the set of carriers involved: $G^{\text{Agg}} = \bigcup_i G^c_i$ where $G^c_i = \bigcup_j N^c_j$ and $E^c_i = \bigcup_j E^c_j$ and $W^{\text{Agg}}(N^{\text{Agg}}, E^{\text{Agg}})$, where $E^{\text{Agg}}$ is the sum over all the corresponding edge weights. Finally, in light of a recent merger between two of the carriers we study (NW and DL), [17] we construct the merged networks $G^{\text{NW+DL}}$ and $W^{\text{NW+DL}}$.

**INDIVIDUAL VERSUS AGGREGATE AND USEFUL DISCRIMINATORS**

We first compare and contrast the network structures of the distinct airlines. Results are summarized in Table I with the passenger airlines listed first, in order of increasing number of distinct airports serviced ($N$). The number of distinct direct connections between airports for that carrier is $E$ (the total number of edges in the unweighted, binary view $G^u(N^c, E^c)$). Included in the table also are the results for the three different aggregate views (Agg7, AggPass, and AggAll), the two cargo carriers Federal Express (FX) and United Parcel Service (UPS), and the tentative “NW+DL” network. The average airport degree, denoted $(q)$, is simply $\langle q \rangle = 2E/N$. The average shortest path length over all source-destination pairs in the network is denoted $\langle l \rangle$. This is the average number of flight segments required to fly from any airport in the network to any other. The average value of betweenness centrality [18] is denoted $\langle b \rangle$. The average clustering coefficient [19] is denoted $\langle C \rangle$.

For comparison, we generate a corresponding Erdős-Rényi (ER) random graph for each carrier, using that carrier’s $N$ and $E$ values. Results are in Table I. Note the values of $\langle l \rangle$ and $\langle b \rangle$ for the actual carriers agree extremely well with the values for the corresponding ER realizations, strongly suggesting that density alone determines these two properties. However, all remaining properties show significant differences between the real networks and ER equivalents. Furthermore, all carriers have $\langle l \rangle \ll ln N$ and values of $\langle C \rangle > \langle C_{\text{ER}} \rangle$, thus can be considered “small-world” networks.

To quantify the extent to which a network follows the “hub and spoke” (HS) pattern, the degree assortativity coefficient [20], $r$, seems to be a natural choice from a network theory perspective. $r > 0$ indicates a tendency of high-degree nodes to connect to other high-degree nodes. $r < 0$ indicates a tendency of high-degree nodes to connect to low-degree nodes. Intuitively, a larger negative (dissassortative) value of $r$ should indicate that the network follows the HS paradigm more closely. Previous studies have found the airport networks of China and India and the airline networks of European carriers to be strongly dissassortative (Refs. [6, 7, 21] respectively), while in contrast the WAN shows assortative behavior [3].

As can be seen in Table I, the two cargo carriers have dissassortative structures. The value of $r$ for SW is about half the magnitude of the other passenger carriers, supporting the widely held understanding that SW has a predominantly PP structure while the other carriers considered here have a more HS topology. Yet the value of $r$ for FX is substantially smaller in magnitude than that for SW, seemingly indicating a lack of HS structure though the topology of FX exhibits strong HS structure. In this context, we turn to a measure used in the transportation literature [22] to quantify the extent of HS structure, the Gini coefficient [23]. The degree Gini coefficient, $G(q)$, is defined for a network of size $N$ as,

$$G(q) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} |q_i - q_j|}{N^2 \langle q \rangle},$$  

where $\langle q \rangle = 2E/N$. It essentially measures the magnitude of the difference in node degree between all pairs of nodes in a network normalized by average node degree. Similarly, the strength Gini coefficient $G(s)$ is defined as $G(s) = (\sum_{i=1}^{N} \sum_{j=1}^{N} |s_i - s_j|)/(N^2 \langle s \rangle)$, with $\langle s \rangle = (\sum_{i=1}^{N} s_i)/N$. As seen in Table I the Gini coefficient metric correctly indicates that FX has a strong HS structure, whereas the value of $r$ misleadingly indicates a strong PP structure. Likewise, for AggPass and AggAll (which have strong HS topologies) the values of $r$ mistakenly indicate strong PP structures, while the values of $G(q)$ correctly indicate strong HS structures. In contrast to $r$ for all the networks analyzed herein, $G(q)$ and $G(s)$ consistently and unambiguously differentiate between the HS and PP structures. (Note, due to computational constraints we could not calculate $G(s)$ for AggAll and AggPass.) The Gini coefficient has been widely used in economics [23], ecology [24] etc. and, as found herein, network studies may benefit from inclusion of this metric. Interestingly, SW has a value approximately half the magnitude of the other carriers for both $r$ and $G(q)$.

We carried out a detailed analysis of betweenness centrality [18] in the manner of Ref. [5], for all the passenger airlines. For a few airlines, we do find examples of airports where the betweenness of the nodes is relatively
higher in comparison to their degree (e.g., IAH for CO, PHX for US, STL for AA and LAX for DL). However, this mismatch is not as strongly disproportionate as that of say the Anchorage airport in Ref. [5]. Hence, we would like to classify our observation as “weak anomalous centrality”.

We analyze the distribution of node degree and node strength, with \( p(q) \) the observed probability of a carrier having a node of degree \( q \) and \( p(s) \) the observed probability of having a node with strength \( s \). These raw probability distributions for our networks are extremely noisy, thus we construct the complementary cumulative distributions \( P(x) = \sum_{x \geq x} p(x) \). Figure 1 shows the cumulative degree distributions, while Fig. 2 shows the cumulative strength distributions.

For each carrier we analyze how well the empirically observed degree distribution can be fit by a theoretical distribution, considering the following forms for the theoretical cumulative distribution function: 1) power law (pl), 2) exponential (exp), 3) stretched exponential (se), 4) power law with exponential decay (pled), 5) cumulative log-normal (chn) distribution. We use the nonlinear least squares fitting routine of the R Statistical Computing platform to solve for the parameters values for each candidate distribution which provide the best fit to the data. Finally, we calculate the residual sum of squares between these best fit candidate distributions and the empirical data. In almost all cases, one of the candidate distributions clearly minimizes this difference. See Table III in Appendix A. Though of course there exist more rigorous methods for extracting the best fit power law exponent to a data set, the airline networks analyzed herein are too far from power law distributions to warrant the overhead associated with such techniques.

The theoretical distribution which best describes the aggregate over the seven passenger carriers (Agg7) is a simple exponential distribution. The aggregate over all passenger carriers (AggPass) is best described by a cumulative log-normal distribution. The aggregate over all flights flown in 2007 is best fit by a power law with exponential tail. Furthermore, as can be seen in Fig. 1 for all three distinct aggregate views, the tail decays more sharply than exponential. The SW network, similar to AggPass, is best described by the cumulative log-normal distribution. The other six individual carriers studied all have networks with degree distributions that are well described by simple exponential distributions. However,

### Table I: Basic network properties of the carriers. \( N \) and \( E \) denote the number of nodes and edges respectively, and \( \langle q \rangle \) the mean node degree. \( \langle l \rangle \), \( \langle b \rangle \) and \( \langle C \rangle \) denote the mean of the geodesic, betweenness, and clustering coefficient distributions. \( r, G(q), G(s) \) denote assortativity and degree and strength Gini coefficients. \( \alpha(q) \) is the skewness of the degree distribution.

| Carrier | \( N \) | \( E \) | \( \langle q \rangle \) | \( \langle l \rangle \) | \( \langle b \rangle \) | \( \langle C \rangle \) | \( r \) | \( G(q) \) | \( G(s) \) | \( \alpha(q) \) |
|---------|------|------|-----------------|-------------|-------------|-----------------|-------|----------|----------|---------|
| SW      | 64   | 892  | 27.88           | 1.542       | 0.0091      | 0.731            | -0.177| 0.254    | 0.490    | 226.277 |
| US      | 96   | 556  | 11.58           | 1.990       | 0.0108      | 0.672            | -0.367| 0.521    | 0.779    | 1053.811|
| CO      | 117  | 736  | 12.58           | 1.935       | 0.0083      | 0.628            | -0.330| 0.512    | 0.823    | 1742.802|
| UA      | 121  | 737  | 12.18           | 1.983       | 0.0084      | 0.640            | -0.320| 0.498    | 0.806    | 1839.590|
| AA      | 121  | 1163 | 19.22           | 1.889       | 0.0076      | 0.646            | -0.280| 0.461    | 0.794    | 1542.013|
| NW      | 132  | 753  | 11.41           | 2.023       | 0.0080      | 0.624            | -0.269| 0.493    | 0.760    | 2130.143|
| DL      | 133  | 906  | 13.62           | 1.943       | 0.0073      | 0.586            | -0.272| 0.499    | 0.790    | 2168.703|
| NW+DL   | 163  | 1529 | 18.76           | 1.985       | 0.0062      | 0.617            | -0.256| 0.497    | 0.767    | 2682.666|
| UPS     | 107  | 606  | 11.33           | 1.929       | 0.0090      | 0.620            | -0.249| 0.427    | 0.628    | 1618.748|
| FX      | 334  | 1355 | 8.11            | 3.060       | 0.0062      | 0.579            | -0.047| 0.548    | 0.696    | 1457.096|
| AggPass | 817  | 9688 | 23.72           | 3.181       | 0.0027      | 0.639            | 0.185 | 0.630    | -       | 8758.680|
| AggAll  | 1258 | 17437| 27.72           | 3.005       | 0.0016      | 0.557            | 0.097 | 0.677    | -       | 17484.512|

### Table II: Properties of Erdős-Rényi (ER) random graphs with \( N \) and \( E \) corresponding to the carriers in Table I.

| ER     | \( \langle l \rangle \) | \( \langle b \rangle \) | \( \langle C \rangle \) | \( r \) | \( G(q) \) |
|--------|-----------------|-------------|-----------------|-------|----------|
| SW     | 1.533           | 0.0090      | 0.446           | -0.065| 0.086    |
| US     | 2.070           | 0.0116      | 0.118           | -0.034| 0.145    |
| CO     | 2.101           | 0.0097      | 0.106           | -0.025| 0.143    |
| UA     | 2.151           | 0.0098      | 0.095           | -0.040| 0.156    |
| AA     | 1.864           | 0.0074      | 0.158           | -0.008| 0.109    |
| NW     | 2.242           | 0.0097      | 0.091           | -0.000| 0.154    |
| DL     | 2.103           | 0.0085      | 0.102           | 0.012  | 0.151    |
| NW+DL  | 1.977           | 0.0061      | 0.116           | 0.003  | 0.116    |
| UPS    | 2.134           | 0.0110      | 0.101           | -0.051| 0.148    |
| FX     | 3.002           | 0.0061      | 0.035           | 0.017  | 0.190    |
| Agg7   | 1.810           | 0.0042      | 0.181           | 0.012  | 0.090    |
| AggPass| 2.457           | 0.0018      | 0.029           | -0.006| 0.113    |
| AggAll | 2.51            | 0.0012      | 0.022           | -0.018| 0.108    |
FIG. 1: Cumulative degree distribution, $P(q)$, for each carrier shown on a log-log scale with the linear-linear view inset. Solid lines are the best fit theoretical distribution: cln for SW, pled for AA and CO, exp for DL, NW, and UA; pled for US and FX; cln for UPS; exp for Agg7; cln for AggPass; pled for AggAll. Note, the tail on each of the three different aggregate views decays more quickly the exponential.
K-CORE STRUCTURES

The SW network is distinguished from the networks of the other carriers by the metrics of Table I, yet the difference in topology is even more pronounced when the k-core structures of the distinct carriers are compared. The k-core of the network is a subgraph constructed by iteratively pruning all vertices with degree less than k. For instance, starting from an original network, we remove all nodes with degree q < k and their corresponding edges, then successively remove all nodes (along with their edges) which are now of degree q < k in the pruned network, and continue iterating until all remaining nodes have q ≥ k. The remaining subgraph is the k-core. We also consider the k-shell, which consists of all nodes which are present in the k-core but not in the (k+1)-core. Likewise, the “coreness” of node i, denoted c_i, is defined as the largest value of k for which the node is a member of the k-core.

The k-core decomposition is a computationally inexpensive way of revealing additional details about the structural role of nodes beyond their degrees and has lately been the focus of several studies in network theory. It has been used to predict protein functions from protein-protein interaction networks and amino acid sequences and to identify the inherent layered structure of the protein interaction network. More recently, the method of k-shell decomposition has been used to arrive at a model of internet topology. The additional structural information revealed by k-core decomposition has been used to generate random graphs with a specified “k-core fingerprint” and simulate the AS network of the internet.

Figure 2 (a) shows the k-core structure of all the carriers studied herein. Here F(k) is the fraction of all nodes with coreness greater than or equal to k. Note that for SW all nodes i have c_i ≥ 7, and the fraction of nodes with large coreness decays much more slowly than for the other carriers. Two key differences are prominent when comparing the k-core structure of SW to the other carriers: the value of k_{max} and the occupancy of the k_{max} shell. For k_{max}, in spite of having the smallest number of nodes N, SW achieves the highest k-core, with value k_{max} = 20, and normalized value k_{max}/N = 0.312. The next largest is American Airlines (AA) with k_{max} = 17 with normalized value k_{max}/N = 0.140.) With respect to occupancy, that of the largest shell in SW is especially remarkable, with 53% of all airports belonging to the k_{max}-core. In contrast, for AA, 26% belong to the k_{max}-core.

Figure 2 (b) compares the k-core structure of SW and AA to the k-core structure of similarly sized power-law (PL) and Erdős-Rényi (ER) random graphs (with N = 64 for SW and N = 121 for AA). The PL curve is the average over an ensemble of 50 independent realizations, each generated by applying the configuration model to a degree sequence selected from a power law distribution with exponent 1.8 (consistent with the power law fit to the world-wide air-transportation network from ). The ER curve is the average over 50 independent realizations generated with the respective N and number of edges.
RESILIENCE

Of great interest is the individual passenger carrier’s resilience to random edge deletion and targeted and random node deletion. Edge deletion corresponds to, for instance, disturbances such as weather preventing travel between a pair of airports (i.e., deletion of a flight path). Node deletion corresponds to the closure of an airport. There is extensive literature investigating various real and simulated networks’ resilience to both random and targeted node and edge removal. Previous work found that simulated random power-law networks are robust to random node deletion but vulnerable to targeted attack 34. Different targeted attack strategies have been investigated on various real and simulated networks using a variety of metrics, notably average inverse geodesic distance (also called ‘network efficiency’) and the relative size of the largest connected component 35. The robustness of graphs with various kinds of degree distributions have also been studied recently, e.g. in Refs. 37, 38 and references therein.

To quantify the performance of the networks under the various deletion processes, we use two topological measures: the size of the largest connected component (denoted $S$) and a relative global travel cost metric (introduced below and denoted $T$) which includes scaled contributions accounting for both spatial (geographic) distance and geodesic distance (hop-count).

The travel cost metric is defined as follows. Note, to include information on geographic distance, we first augment $G^*(N^c, E^c)$ by including on each edge the geographic length of the edge. Then for every possible source-destination pair $(i, j)$, we apply Dijkstra’s algorithm 39 (as implemented in NetworkX 40) to determine the path with the shortest geographic distance connecting $i$ and $j$. In the event that there is an edge directly between $i$ and $j$, the shortest path is simply $(i, j)$. If the path consists of a sequence of edges (denote these $(i, i_1), (i_1, i_2), \ldots, (i_m, j)$), we calculate the total path length $d_{ij}$ by adding the length of the edges:

$$d_{ij} = l_{i_1i} + l_{i_1i_2} + \ldots + l_{i_mi}.$$

Next we convert the geographic path length to a ‘flight time’ by dividing by a characteristic velocity ($v = 500$ miles/hour). For each of the $m$ intermediate nodes in the path we add a fixed ‘transfer cost’ of $\theta = 1.0$ hour to account for layover time:

$$t_{ij} = \frac{d_{ij}}{v} + m\theta.$$

Finally, we can define the travel cost for the whole network or for just a subset of nodes in the network $M \subseteq N^c$ as the sum over all of the included path costs:

$$T(M) = \frac{1}{2} \sum_{i \in M} \sum_{j \in M} t_{ij}.$$

Note, the travel cost over the entire network is $T(N)$.

With these two metrics ($S$ and $T$) in mind, we analyze the resilience of the carrier networks to random edge deletion and to random and targeted node deletion. We first consider random edge deletion as increasing numbers of edges are deleted. For each number of deleted edges, we generate an ensemble of 50 randomly selected sets of edges to delete. Figure 3 (a) shows the results for $S$ (the relative size of the largest connected component) as more edges are removed. Remarkably, SW has
FIG. 3: Two metrics illustrating the resilience of the passenger carriers: (a) Relative size of the largest connected component \(S\) of each passenger carrier’s network as a function of the proportion of edges removed by random failure \(r\) averaged over 50 realizations. (b) Travel cost metric \(\frac{T(M)}{T_0(M)}\) \((M\) is the set of nodes in the largest connected component\), evaluated on the largest connected component of the passenger carrier’s networks as a function of \(r\) averaged over 50 realizations. Representative error bars (with height one standard error above and below the mean) are shown in both (a) and (b) on SW and US.

FIG. 4: \(S\) of each passenger carrier’s network as a function of the proportion of nodes removed by targeted attack \(t\). The node (and its edges) with the largest existing degree and (inset) betweenness centrality are iteratively removed at each step. Targeting by betweenness rather than degree causes more rapid breakdown of each carrier’s network. The dashed diagonal line depicts the maximal size of \(S\) under this process for any network (i.e. the size of \(S\) for a complete graph).

nearly 98% of its nodes in largest connected component even after the deletion of 80% of its edges (and remains at 100% connected for every realization in the ensemble until 30.8% of the edges are removed). In contrast, all of the other carriers have realizations that start losing full connectivity after the deletion of fewer than 2% of edges.

Once some nodes are disconnected, there is no shortest path to any of these disconnected nodes so the travel cost over the whole network is formally infinite. Consequently, when calculating the travel cost we consider only the nodes in the largest connected component of the randomly damaged graph. We calculate the travel cost between all source-destination pairs in this subset in the original graph, \(T_0(M)\), and in the damaged graph, \(T(M)\). Limiting the cost metric to the largest connected component gives a measure of the efficiency of the underlying structure independent of the disconnected nodes. Finally, we normalize the damaged travel cost through the largest connected component by the travel cost for this subset in the original network to obtain the relative travel cost of the damaged network \(\frac{T(M)}{T_0(M)}\). In this manner, we eliminate network size effects by comparing the performance of the damaged network only with the corresponding original network.

As shown in Fig. 3 (a), for all of the carriers, the majority of the network remains connected even after removing a high percentage of the edges. Yet the HS networks are fragile in the sense that a small set of nodes can be completely disconnected from the network even in the low regime of edge deletion which reasonably models weather disturbances (< 20%). This result is consistent with the prevalence of low-degree nodes occupying the lower \(k\)-shells in the HS networks.

The cost metric gives a more subtle portrait of the effect of removing edges from the carrier network, Fig. 3 (b). For small fraction of edges deleted all carriers have similar performance. Yet once more than 15% of edges are deleted SW exhibits the lowest ensemble-averaged travel cost through the largest component. Intuitively, a well-connected (high density) PP structure permits the carrier to use the majority of its nodes as viable transfer points between most source-destination pairs, so even if an edge deletion process eliminates the original shortest path between a pair there is likely to remain other nearly-shortest paths. We expect HS networks which route the
REDESIGNING NETWORKS: K-CORE ENHANCING, COST PRESERVING, REWIRING

Of great interest is understanding how to increase the resilience of an existing network. We show here one rewiring process which can increase binary edge density, \(k\)-core (both in terms of the value of \(k_{\text{max}}\) and the occupancy of larger \(k\)-shells), and consequently resilience to node and edge deletion without increasing flight or airport requirements. Given a set of four nodes connected as shown in figure 5(a), the missing connection to form a 4-clique can be created by routing a small number of flights along the existing edge connecting nodes 1 and 3. To preserve the gate requirements, a flight originally between nodes 3 and 4 is rerouted between 2 and 4. In this manner, the total number of flights above all edges and the gate requirements (the in-strength and out-strength of each node) remain constant. The addition of the edge connecting 1 and 3 raises the coreness of at least one of these two nodes.

To test this rewiring scheme, we extracted the ‘Daily 2-flight minimum’ weighted subnetwork for each carrier \(c\), formed by rescaling all edge weights \(s_{ij} \rightarrow \left\lfloor \frac{2s_{ij}}{k_{\text{max}}} \right\rfloor\) and removing all edges with new weight less than 2 (each edge must begin with at least two daily flights in order to be considered for rewiring by our scheme).

Fig. 6 shows the results for UA (chosen since it has a particularly shallow \(k\)-core structure) after adding 10, 20, 30, and 50 percent additional edges according to the flight and gate-preserving scheme. The initial network has 62 nodes and 113 edges. The rewiring scheme increases the original binary edge density by 9.7% upon adding 10% extra edges and by 18.6% upon adding 20% extra edges. More so, it moves lower \(k\)-core nodes into higher cores. Fig. 6(a) shows the resulting enhanced \(k\)-core structure. Fig. 6(b) shows the increased resilience with respect to \(S\), the fraction of nodes in the giant component. The rewired networks are also more resilient to random edge removal (not shown here).

While the specific many-variable optimization problems solved by the carriers may preclude such simple rewirings, this example suffices to show the existence of strength-preserving transformations which increase binary edge density and consequently network resilience to node and edge failure. It is difficult to conceive of a degree preserving rewiring which enhances the \(k\)-core structure of networks as the degree sequence essentially determines the \(k\)-core structure of a network [41, 42].

CONCLUSION

We have analyzed the network structures of the major passenger airlines in USA with an aim of furthering understanding of how to better design and operate critical infrastructure. We find the degree distributions of several of the networks, including the aggregate views, are well described by simple exponential or cumulative log-normal distributions. We notice that the Gini coefficient can supplement the structural knowledge gained from degree assortativity in networks. We establish connections between \(k\)-core structures of networks and how airlines can be well-connected and resilient to disturbances as diverse as random and targeted attacks or failures, by both degree and betweenness. Similar results hold whether the disturbance involves airports or flight paths. Furthermore, using a travel time heuristic introduced herein, we find that connectivity can be maintained without much increase in travel time, despite significant disruption.

Of the seven largest USA passenger air carriers, especially remarkable is Southwest Airlines, where more than half of all nodes belong to the \(k_{\text{max}}\)-core leading to extreme resilience for that network. We present a simple rewiring scheme which can help existing airlines approach such \(k\)-core structures, without increase of flight or air-
FIG. 6: Applying the rewiring of Fig. 5 to the daily 2-flight minimum UA network increases $k$-core structures and resilience to degree-targeted node deletion (removed exactly as in Fig. 4). (a) Cumulative $k$-core distribution, $F(k)$, of the increasingly rewired UA network, along the SW daily 2-flight minimum for reference. (b) $S$ as a function of $t$ for the networks in (a).

craft requirements.

Our findings on $k$-core and resilience suggest that hierarchical networks could be especially susceptible to targeted attacks or failures, given the rare population of the highest $k$-cores of such networks. The future design and operation of critical infrastructure may benefit from analyzing the tradeoffs of core versus peripheral placement of hub nodes. Hubs located in the core of a network substantially increase efficient connectivity yet are critical targets as without them, the network loses connectivity. Hubs in the periphery (low $k$-cores) offer smaller benefits with respect to efficient connections, yet if they are disabled the connectivity of the core of the network remains largely unaffected.

It is worth noting that the effect of targeted attack by betweenness, rather than by degree, is significantly more pronounced on each carrier’s network. This complements previous studies on the importance of betweenness in the World Airline Network [5] and suggests that betweenness is an important criterion for consideration in critical infrastructure networks.

Our findings may be applicable to the design and operation a range of infrastructure networks, for instance other transportation networks, power-grids, and telecommunication networks. In addition, including analysis of resilience properties of networks would augment current studies on the optimal distribution of resources or facilities in a given geographical area. The analysis herein is a first step. Real weather would correspond to correlated edge deletion, not random. Moreover we neglect scheduling and restrict ourselves to the domestic routes of some international carriers.

[1] Amaral LAN, Scala A, Barthélemy M, Stanley HE (2000) Proc. Natl. Acad. Sci. USA 97, 11149-11152.
[2] Guimerà R, Amaral LAN (2004) European Physical Journal B, 38, 381-385.
[3] Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) Proc. Natl. Acad. Sci. USA 101, 3747-3752.
[4] Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) Proc. Natl. Acad. Sci. USA 103, 2015-2020.
[5] Guimerà R, Mossa S, Turchetti A, Amaral LAN (2005) Proc. Natl. Acad. Sci. USA 102, 7794-7799.
[6] Li W, Cai X (2004) Phys. Rev. E 69, 046106.
[7] Bagler G (2008) Physica A 387, 2972-2980.
[8] Aklous D (2008) Math. Proc. Camb. Phil. Soc. 145, 471.
[9] Gastner MT, Newman MEJ (2006) Phys. Rev. E 74, 016117.
[10] Barthélemy M, Flammini A (2006) J. Stat. Mech. L07002.
[11] Siddiqi A (2003) Deregulation and Its Consequences, http://www.centennialofflight.gov/essay/CommercialAviation/Dereg/ Tran8.htm
[12] World Air Transport Statistics (WATS) for 200, 52’nd Edition, IATA, http://www.iata.org/ps/publications/wats-passenger-carried.htm
[13] Southwest turns a profit for 69’th straight quarter, July 25, 2008 http://www.nytimes.com/2008/07/25/business/25air.html
[14] Easyjet and Ryanair flying high on the Southwest model: Charting the ups and downs of low-cost carriers (2006) , Strategic Direction, 22, 18-21.
[15] Gittell JH (2005), The Southwest Airlines Way. McGraw-Hill, New York.
[16] http://www.transtats.bts.gov/Fields.asp?Table_ID=259
[17] http://www.delta.com/help/faqs/merger/index.jsp
[18] Freeman LC (1977) Sociometry 40, 35-41.
[19] Watts DJ, Strogatz SH (1998) Nature 393, 440-442.
[20] Newman MEJ (2002) Phys. Rev. Lett. 89, 208701.
[21] Han D-D, Qiana J-H, Liu J-G (2009) Physica A, 388,
APPENDIX A: DATA FITTING

We analyze the statistical distribution which best fits the empirical data of the complementary cumulative degree distribution, $P(q) = \sum_{i \geq q} p(i)$, for each individual airline carrier’s network, choosing amongst: 1) power law, 2) exponential, 3) stretched exponential, 4) power law with exponential decay, 5) cumulative log-normal distribution. We use the nonlinear least squares fitting routine of the R Statistical Computing platform [25] to solve for the parameters values for each candidate distribution which provide the best fit to the data. Finally, we calculate the residual sum of squares between these best fit candidate distributions and the empirical data. The explicit values are given in Table III. The smallest value of the residual sum of squares over all the distributions is highlighted in red. If multiple candidate functions do equally well up to the first two significant figures, both are highlighted in red. In all these cases it is the EXP and PLED which perform equally well (and the “power law” portion of the PLED has exponent quite close to zero).

TABLE III: Residual sum of squares of the best fit functional form to the empirical airline data, choosing amongst power law (PL), exponential (EXP), stretched exponential (SE), power law with exponential decay (PLED) and the cumulative of a log-normal (CLN).

| Source | PL | EXP | SE | PLED | CLN |
|--------|----|-----|----|------|-----|
| AggAll | 0.7723 | 0.4222 | 7.516 | 0.03428 | 0.04669 |
| AggPass | 0.8448 | 0.3540 | 7.226 | 0.06245 | 0.05677 |
| Agg7 | 2.169 | 0.1066 | 15.32 | 0.1047 | 0.3098 |
| SW | 3.102 | 0.5565 | 17.94 | 0.1569 | 0.03924 |
| AA | 1.599 | 0.04187 | 7.875 | 0.03506 | 0.0466 |
| CO | 0.6395 | 0.02408 | 3.788 | 0.01919 | 0.03244 |
| DL | 0.8928 | 0.01375 | 4.431 | 0.01361 | 0.02473 |
| NW | 0.632 | 0.4055 | 3.558 | 0.04939 | 0.05967 |
| UA | 0.6775 | 0.03523 | 3.752 | 0.03515 | 0.03662 |
| US | 0.5079 | 0.04065 | 3.305 | 0.03742 | 0.05414 |
| FX | 0.2517 | 0.05657 | 1.568 | 0.02053 | 0.0382 |
| UPS | 0.8909 | 0.06618 | 3.941 | 0.04466 | 0.02802 |
FIG. 7: Cumulative strength distribution, $P(s)$, for each carrier shown on a log-log scale. These distributions span several decades of range. In contrast, the corresponding cumulative degree distributions (shown in Fig. 1) terminate orders of magnitude earlier. SW has no nodes with $s < 2000$, hence its plot begins with that value.