Characterizing the Efficiency-Robustness Trade-offs in US Domestic Airline Networks

(Dated: December 23, 2009)

PACS numbers: Valid PACS appear here

I. INTRODUCTION

Airline networks form the backbone that carries the bulk of a country's transportation load. The US domestic airlines carried more than 53 million people in the first quarter of 2009, over 746,400 flights, amounting to a revenue of more than 46 billions [1]. In addition to catering to transportation needs, airline networks have long reaching socio-political impacts. They are critical infrastructure wherein operational inefficiencies can cost billions of dollars every year [2–4]. In recent times, airline networks also pose the threat of propagating contagious diseases such as SARS and swine influenza [5].

Airline networks have evolved over a period of time as a result of design decisions governed largely by economic considerations of airline companies, and constrained by geo-political relations between countries [6]. There are studies on the impact of the hub and spoke structure of US domestic airlines following deregulations in the 1970s. There are conflicting views on the economic advantages of hub and spoke structures as opposed to point to point structures [7–10]. Recently, there are suggestions that a co-existence of hub and spoke and point to point networks can lead to stable equilibria by means of game theoretic analyses [11].

The underlying structure or topology of an airline network not only impacts its day to day performance, but also its resilience in the phase of uncertainties and disasters. The structure also influences important decisions such as location of hubs, airline code sharing, mergers [12, 13], schedules and pricing [10, 14, 15]. Hence, a detailed understanding of the structure of airline networks is important.

In this work, we analyze the networks of seven major US domestic carriers from a complex networks perspective. We study their structural properties with respect to two parameters critical to network performance: *efficiency* and *robustness*. We characterize the trade-offs between efficiency and robustness in terms of multiple performance metrics using a graph theoretic framework. We observe that US domestic airline networks have similar structural properties that are possibly a result of similar design trade-offs.

There are several efforts in the last few years towards analyzing airline networks. Guimerà *et al.* [16] study the world-wide airport network (WAN) and report "anomalous" high betweenness centrality values for nodes that have relatively low degrees. They argue that power-law generating models such as *preferential attachment* (PA) cannot explain this observation. Instead, they propose that the anomalous centrality is a result of a network designed under several geo-political constraints leading to well-defined, geographically separated community structures [6]. Further, they classify cities based on their intra and inter-community "participation" into roles such as hubs, provincial hubs, connectors and peripheral nodes.

Barrat *et al.* [17] consider a weighted network model of WAN, and analyze the degree and betweenness distributions, clustering coefficient and the *mixing pattern* [18] of nodes. Here, weights on edges are expressed in terms of the number of available passenger seats between pairs of airports. They report that considering weights allows a richer characterization of a network. In a related work, Barthélemy *et al.* propose a model of network evolution based not only on the network topology, but also on the strengths of interaction between nodes represented by edge weights [19].

Colizza *et al.* [5] study the role of the air transportation network in the spreading pattern of diseases. They report a correlation between the heterogeneity in the structure of the air transportation network and the seemingly erratic, global scale sprading of diseases.

Domestic airline networks have also been studied in an effort to better understand the dynamics of transportation networks at a lower scale. Li and Cai [20] observe that the degree distribution of the airport network of china (ANC) follows a double Pareto distribution. Bagler [21] reports that the airport network of India (ANI) is a small world like ANC and WAN, however with a much smaller clustering coefficient. Further, ANI suggests a *disassortative mixing* with hubs having a lot more low-degree neighbours than high-degree neighbours. This is in contrast with WAN wherein *assortative mixing* can be observed [17]. Similar analysis on the Italian airport network shows a small-world network characterized by a fractal nature [22]. Sienkiewicz and Hołyst [23] report a statistical analysis of transportation networks of 22 Polish cities. Similar to other networks in this class, small-world behaviour with a hierarchical structure is observed.

In general, the following approaches can be found in literature concerning the study of airline networks: (1) analyzing properties such as degree and node betweenness distributions, clustering coefficient and mixing, such that they can be related to standard network generation models like preferential attachment and small-world networks, and (2) developing a theory or model to generate networks that show properties characteristic of airline networks. While these approaches have resulted in many important insights, there is still a need for an in depth understanding how different aspects of the structure of airline networks relate to their overall performance. Also, while network generation models explain certain properties, such as power law degree distributions or large clustering coefficients, there is a need to validate network generation models with respect to a wider set of properties [?].

In this work, we use multiple metrics for efficiency and robustness that characterize the different aspects of optimality (or performance) in networks. We observe that the most of the seven airline networks that we study separately show similar trade-offs. This work is part of a larger project in which we study several classes of real world complex networks including airline networks, food webs, supply chains and trade networks in search of structural motifs, within classes as well as across classes, that lead to optimal performance in disparate networks. We conjecture that these optimal motifs are potential underpinnings for designing new complex networks.

II. PRELIMINARIES

A. Data

We consider seven major US domestic carriers: American Airlines, Southwest Airlines, Delta Airlines, United Airlines, Continental Airlines, US Airways and Northwest Airlines. Our data comes from the official flight schedules effective from May 1, 2008, available on the respective websites of the carriers. Table I has information about the sizes of each airline network.

Every domestic airport served by a carrier is denoted by a node in the carrier's network. We do not consider international destinations. We also do not consider flights operated by subsidiaries. An edge in a carrier's network indicates the presence of at least one direct flight belonging to the carrier between a pair of nodes. We observe that all the networks are symmetric. That is, an edge exists in both directions between every pair of nodes in all the networks. Thus, we replace a pair of directed edges between nodes with a single bidirected or undirected edge. In the rest of the paper we consider all networks to be undirected.

Both unweighted and weighted networks are considered. Here, weights on edges are simply the distances between airports computed from latitude and longitude information using the haversine formula [24].

B. Network Analysis Framework

Our framework is inspired by Venkatasubramanian *et al.*'s [25] work on the emergence of network topologies under varying constraints. They propose a formalism based on the Darwinian evolutionary theory: networks evolve their structures so as to optimize their short term operational performance (efficiency) as well as long term survival chances (robustness). There is also a cost constraint that networks cannot breach. A selection pressure variable decides the trade-off between short term and long term survival goals. Efficiency is measured in terms of the average path length (*apl*) in the network. Robustness is measured in terms of the size of the largest component in the networks upon a node deletion. Cost is a function of the number of edges in the network. The "star" topology emerges as most efficient and least robust, when the selection pressure is completely on efficiency. The "circle" topology emerges as the most robust and least efficient, when the selection pressure is completely on robustness. Hub and spoke structures of different types emerge for intermediate selection pressures. Thus, the structure of a network is the result of evolutionary adaptation under selection pressure, which governs the trade-offs between three critical system parameters, efficiency, robustness and cost.

Performance of a network has multiple facets, such as: low communication cost, congestion-free flow of traffic and presence of alternate paths in case of component failures. Hence, we extend the above formalism to propose measures of efficiency and robustness that model different aspects of performance.

Venkatsubramanian *et al.*'s work is pertinent, especially in the context of airline networks for another reason. Airline networks are generally known to have a hub and spoke structure. Since the above work yields different types of hubs and spokes structures for different selection pressures, a comparison between the evolutionarily optimal topologies and airline network topologies can lead to important insights regarding design trade-offs. We consider this question in a later section.

III. COST ANALYSIS

Cost is measured in terms of the number of edges in an unweighted network. The minimum number of edges ($\check{e}$) required to have a connected undirected graph is $n-1$. We do not associate any cost to a minimally connected graph. Any "extra" edge has an associated cost. All extra edges cost the same. An undirected clique has the highest cost, with $\hat{e} = \frac{n(n-1)}{2}$ (and $\hat{e} = n(n-1)$, for directed) number of edges. Thus, the *density* ($d^u w$) of a topology is defined as a ratio of the number of extra edges in a topology to the number of extra edges in the clique with the same number of nodes.

$$d^{uw} = \frac{e - \check{e}}{\hat{e} - \check{e}}$$

We also measure the *redundancy* (r^{uw}) of a network as the ratio of the number of edges, e, in the network to that in a minimally connected network (spanning tree), $\check{e}$.

$$r^{uw} = \frac{e}{\check{e}}$$

In case of weighted graphs, cost is measured in terms of the total weight over edges, E. A minimum spanning tree is the minimally connected network, with an assigned cost, $\check{E}$ 0. A complete graph has the highest cost,

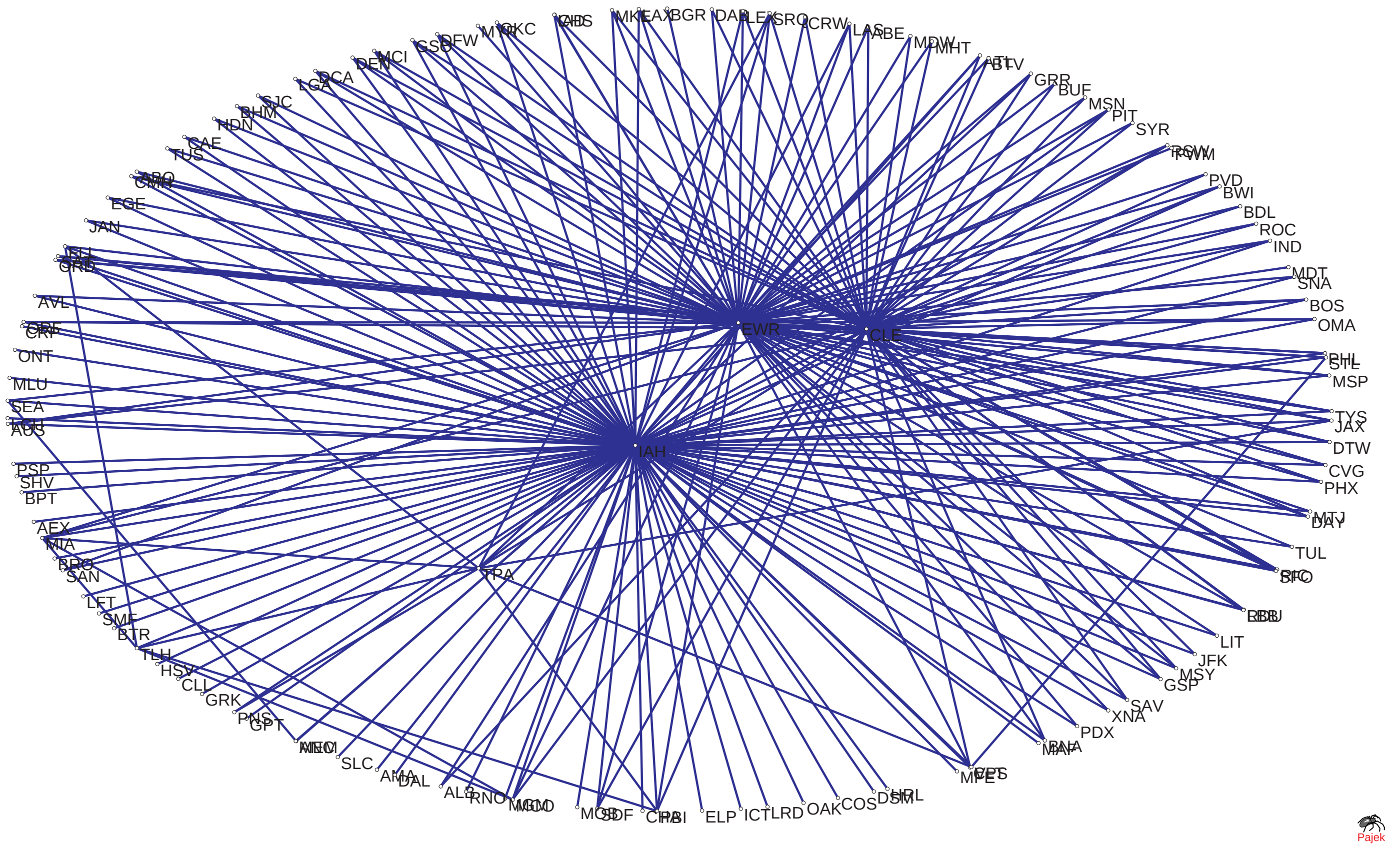

FIG. 1: Topology of Continental Airlines

$\hat{E}$, which is found by computing the weights between all pairs of nodes. Thus, we have equivalent definitions of weighted density, d^w, and weighted redundancy, r^w.

$$d^w = \frac{E - \check{E}}{\hat{E} - \check{E}}$$

$$r^w = \frac{E}{\check{E}}$$

Table I shows the cost measures of the seven airline networks. We can observe that six of the seven (except for Southwest Airlines), have a unweighted density, d^{uw}, around 0.03 and a weighted density, d^w around 0.04. The low values densities are due to the hub and spoke structure of these networks, where in a small number of central hubs serve a large number of destinations. The cost of these six networks is close to that of a spanning tree. On the other hand, Southwest Airlines is very dense. Figures 1 and 2 show topologies of Continental Airlines and Southwest Airlines respectively.

IV. EFFICIENCY ANALYSIS

Efficiency measures the *operational performance* of an airline network. There are a number of ways in which efficiency can be defined. We start with an unweighted network. The *diameter*, which is the longest of all pairs shortest paths in a network, models the upper bound on the number of stopovers or flight changes in a network. The *average path length* models the number of stopovers on average.

The best unweighted diameter, D^{uw}_{min}, is 1, which is the diameter of a clique or a complete graph, wherein every node has a direct connection to every other node. A topology is most efficient if the diameter is 1. In other words, the best *lower bound* on efficiency, $\hat{\eta}^{uw}_{max}$, occurs when the diameter is 1.We measure the *lower bound on efficiency* of a network with diameter, D, against that of a clique, which is 1:

$$\hat{\eta}^{uw} = \frac{1}{D}$$

Similarly, a clique has the best unweighted apl, L^{uw}_{min}, of 1. The best *average case* efficiency in an unweighted network, $\bar{\eta}^{uw}_{max}$, occurs when the apl is 1. We measure the

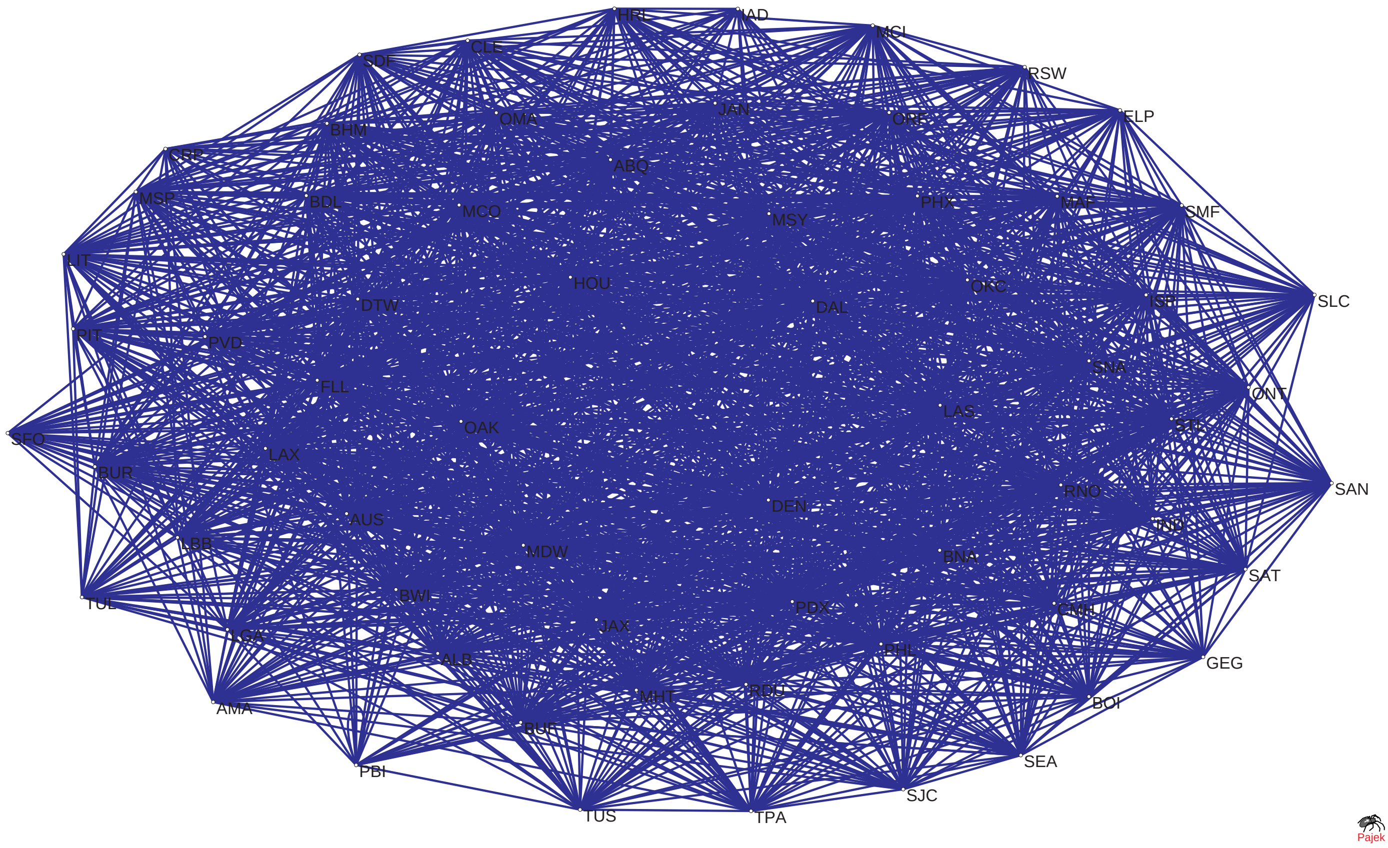

FIG. 2: Topology of Southwest Airlines

absolute average case efficiency of a network with apl, L, against that of a clique, which is 1:

$$\bar{\eta}^{uw} = \frac{1}{L}$$

In case of a weighted network, shortest paths are computed by considering weights or distances on edges. Therefore, the diameter is the greatest distance between any pair of nodes in the network. The average path length is the average distance between pairs of nodes. Owing to the triangle inequality, a direct link between two nodes is always shorter than an indirect path. Thus, similar to the case of unweighted networks, the best weighted diameter and apl occur in a complete graph. We use relations similar to the above to compute the values for efficiencies in weighted networks.

Table I shows a summary of the efficiency analysis for the seven domestic carriers. Except for Southwest airlines, all other airline networks have an unweighted diameter, D^{uw} of 3. This is indicative of a hub and spoke arrangement, with a small "core" of hubs that form a complete graph. Such a hub and spoke arrangement achieves a small diameter with a small number of edges, $e \approx n-1$, for large n, nearly equal to that in a spanning tree. In fact, we can observe that all airline networks except Southwest have a small number of large hubs (figure **??**). They have an apl which is very close to that of a star network. A star network has an apl ≈ 2, when n is large. This is due to the fact that the networks have a large number of "extra" edges in addition to the $n-1$ required to build a minimally connected network, which help in reducing the *apl*.

Maximizing the symmetry in the distribution of distances between pairs of nodes can also serve as a useful measure of efficiency. Therefore, per node *eccentricity* (longest of all shortest paths from a node) and *closeness centrality* values can also be used to define efficiency.

1. Robustness

Robustness measures the resilience of a network in the face of node and edge failures. Robustness is often defined in terms of the *skew* in the *importance* of nodes and edges. As such, centrality measures viz; degree, node betweenness and edge betweenness distributions can be used. When there is a skew in the centrality measures, a small number of nodes and/or edges are more important than the others. Thus, their failure affects the network's performance much more than failures in the rest of the network. On the other hand, a symmetric central-

TABLE I: Summary of efficiency results for the seven domestic airline networks.

Carrier	**American**	**Continental**	**Delta**	**Northwest**	**Southwest**	**United**	**US**
# Nodes (n)	82	125	94	105	66	88	79
# Edges (e)	169	266	210	207	1618	205	207
Cost and Efficiency Measures for Unweighted Airline Networks							
Density (d^{uw})	0.03	0.02	0.03	0.02	0.75	0.03	0.04
Redundancy (r^{uw})	2.090	2.14	2.23	1.99	24.89	2.36	2.65
Diameter (D^{uw})	3	3	3	3	2	3	3
APL (L^{uw})	1.99	2.06	2.03	2.13	1.24	2.06	2.10
PL Dist. (1)	0.05	0.03	0.05	0.04	0.75	0.06	0.07
(2)	0.91	0.88	0.87	0.79	0.25	0.83	0.77
(3)	0.04	0.09	0.08	0.17	0.00	0.11	0.16
Cost and Efficiency Measures for Weighted Airline Networks							
Density (d^{w})	0.04	0.02	0.04	0.02	0.76	0.04	0.04
Redundancy (r^{w})	14.09	15.10	16.64	11.30	196.46	16.58	15.73
Diameter (D^{w}) (km)	11075	8182	10288	10613	4368	10236	9476
APL (L^{w}) (km)	3264	2377	3048	2740	1969	3089	3304

ity distribution ensures robustness to random as well as targetted node/edge failures.

One of the definitions of robustness we use is skew in degree centrality. We define this as the difference in the maximum degree in the graph ($\hat{p}$) and the mean degree of the nodes ($\bar{p}$). For a connected graph of n nodes, the worst skew occurs for the star topology. The central node has a degree of $n-1$ and all the nodes surrounding it have a degree of 1. Therefore, the worst skew is $\frac{(n-1)(n-2)}{n}$. The best skew is 0, when all the nodes have the same degree. This occurs when the topologies are *regular* graph topologies as in a circular topology or a clique. This holds for both directed and undirected graphs. Thus,

$$\rho = 1 - \frac{n(\hat{p} - \bar{p})}{(n-1)(n-2)}$$

Another way to measure robustness is in terms of *connectivity* (λ). Connectivity is the minimum number of nodes or edges whose removal renders the network disconnected. In case of an undirected graph, the tree topologies have the worst connectivity of 1, and the circle has the worst connectivity of 1 in directed graphs. For both cases, the clique has the best connectivity, $n-1$. Thus, robustness, when defined in terms of connectivity is:

$$\rho = \frac{\lambda - 1}{n - 2}$$

V. FUTURE WORK

VI. CONCLUSION

The conclusion goes here.

Acknowledgment

The authors would like to thank...

[1] Source: Bureau of Transportation Statistics, T-100 Market

[2] K. J. Button. *An Assessment of the Capacity and Congestion Levels at European Airports*. Journal of Air Transport Management, 1999.

[3] J. A. Brueckner. *Internalization of airport congestion: A network analysis*. International Journal of Industrial Organization, 2005.

[4] S. A. Morrison and C. Winston. *Another look at airport congestion pricing*. American Economic Review, 2007.

[5] V. Colizza, A. Barrat, M. Barthélemy and A. Vespignani. *The role of the airline transportation network in the prediction and predictability of global epidemics*. Proc. Natl. Acad. Sci. USA, Vol. 103, pp. 20152020.

[6] R. Guimerà, S. Mossa, A. Turtschi and L. A. N. Amaral. *The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles*. Proceedings of the National Academy of Sciences, 2005.

[7] J. A. Brueckner and Y. Zhang. *A Model of Scheduling in Airline Networks: How a Hub-and-Spoke System Affects Flight Frequency, Fares and Welfare*. Journal of Transport Economics and Policy, Vol. 35, No. 2 (May, 2001), pp. 195-222.

[8] G. Burghouwt and J. Hakfoort. *The evolution of the European aviation network, 19901998*. Journal of Air Transport Management, 2001.

[9] K. Button. *Debunking some common myths about airport hubs*. Journal of Air Transport Management, 2002.

[10] S. Berry, M. Carnall and P. T. Spiller. Airline hubs: costs, markups and the implications of customer heterogeneity.

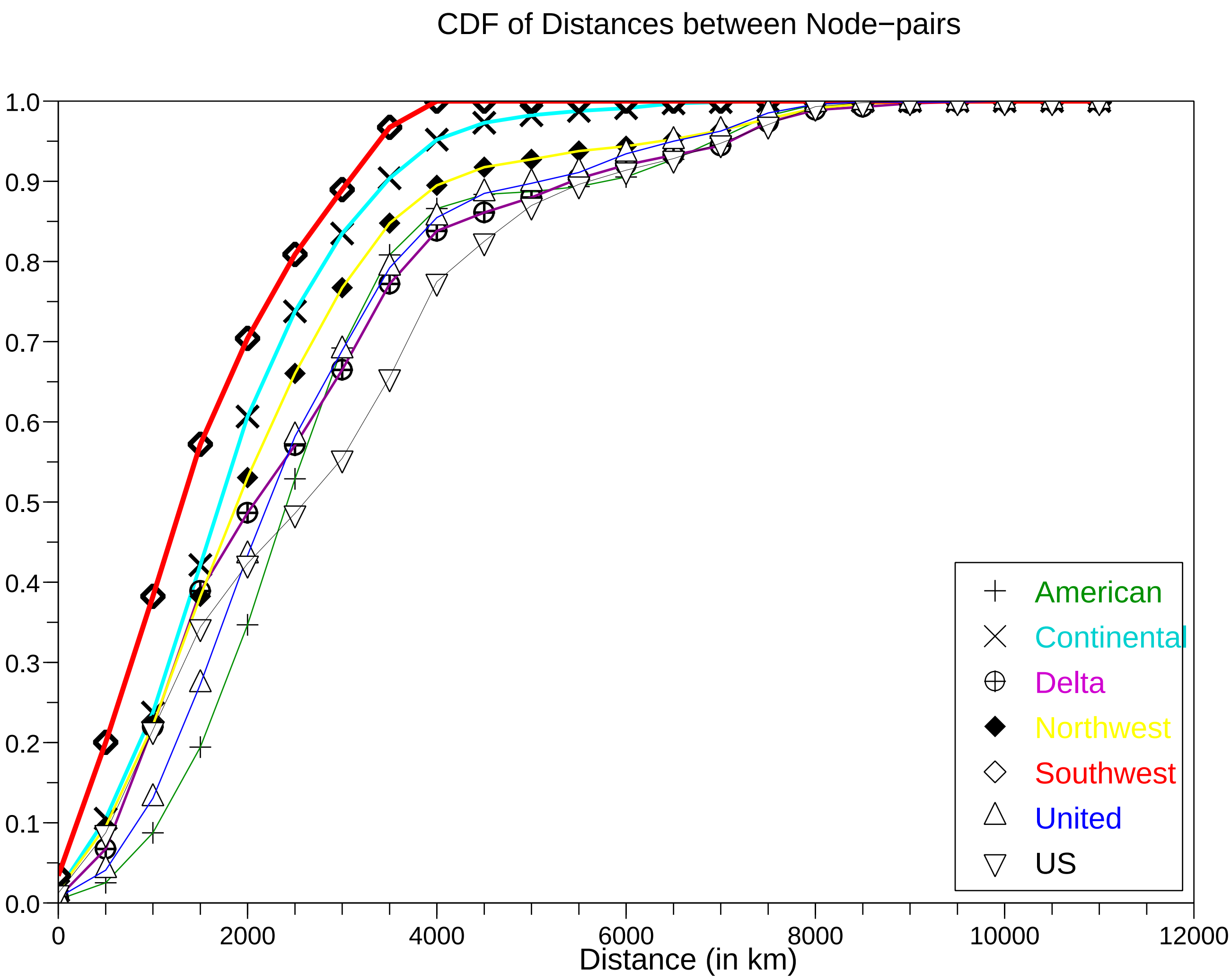

FIG. 3: Path Length Distribution between Node Pairs

TABLE II: Summary of robustness results for the seven domestic airline networks.

Carrier	**American**	**Continental**	**Delta**	**Northwest**	**Southwest**	**United**	**US**
Node Degree based Robustness Measures							
Max Degree ($\hat{p}$)	79	106	87	84	65		
Min Degree ($\check{p}$)	1	1	1	1	1	20	1
Mean Degree ($\bar{p}$)	4.12	4.26	4.47	3.94	49.03		
Degree Skew (p_{skew})	0.22	0.19	0.20	0.19	0.005		
Node Betweenness based Robustness Measures							
Max Betweenness ($\hat{nb}$)	0.89	0.79	0.85	0.50	0.10		
Betweenness Skew (nb_{skew})	0.88	0.78	0.84	0.49	0.08		

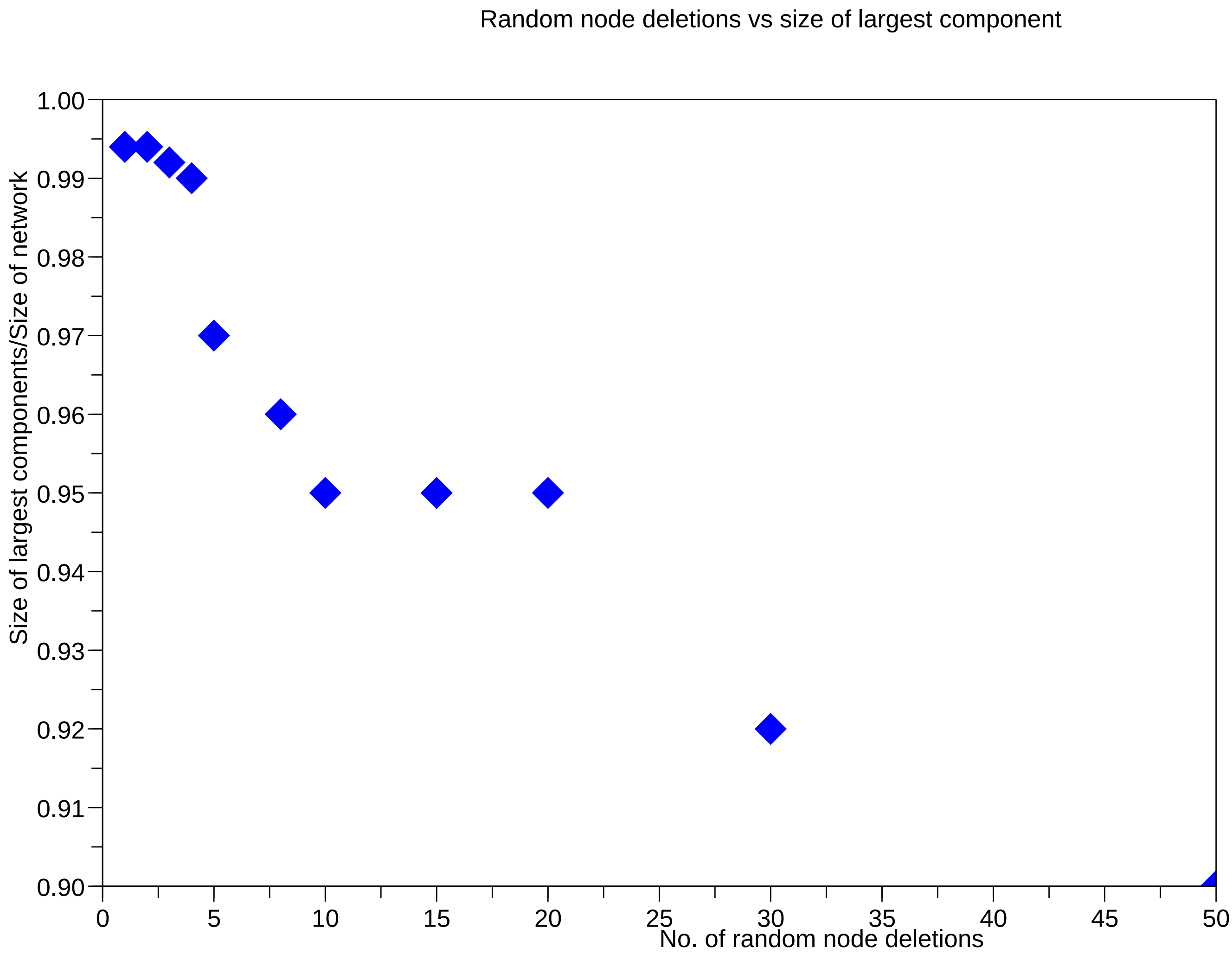

FIG. 4: Effect of random node deletions on functional robustness.

Competition Policy and Antitrust, 2006

[11] M. Alderighi, A. Cento, P. Nijkamp and P. Rietveld. *Network competition: the coexistence of hub-and-spoke and point-to-point systems*. Journal of Air Transport Management, 2005.

[12] N. Adler and K. Smilowitz. *Hub-and-spoke network alliances and mergers: Price-location competition in the airline industry*. Transportation Research Part B, 2007.

[13] S. Alumur and B. Y. Kara. *Network hub location problems: The state of the art*. European Journal of Operational Research, 2008.

[14] R. P. McAfee, V. Te Velde. *Dynamic pricing in the airline industry*. Economics and Information Systems, 2006.

[15] S. J. Forbes. *The effect of air traffic delays on airline prices*. International Journal of Industrial Organization, 2008.

[16] R. Guimera and L. A. N. Amaral. *Modeling the worldwide airport network*. The European Physical Journal B, 2004.

[17] A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani. *The architecture of complex weighted networks*. Proceedings of the National Academy of Sciences, 2004.

[18] M. E. J. Newman and M. Girvan. *Mixing patterns and community structure in networks*. Proceedings on a Workshop on Statistics on Networks, 2007.

[19] M. Barthélemy, A. Barrat, R. Pastor-Satorras and A. Vespignani. *Characterization and modeling of weighted networks*. Physica A: Statistical Mechanics and Its Applications, 2005.

[20] W. Li and X. Cai. *Statistical analysis of airport network of China*. Phys Rev E, 2004.

[21] G. Bagler. *Analysis of the Airport Network of India as a complex weighted network*. Physica A: Statistical Mechanics and its Applications, 2008.

[22] M. Guida and F. Maria. *Topology of the Italian airport network: A scale-free small-world network with a fractal*

structure?. Chaos, Solitons and Fractals, 2007.

[23] J. Sienkiewicz and J. A. Holyst. *Statistical analysis of 22 public transport networks in Poland*. Physical review. E, Statistical, nonlinear, and soft matter physics, 2005.

[24] http://en.wikipedia.org/wiki/Haversine$_f ormula$

[25] V. Venkatasubramanian, S. Katare, P. R. Patkar and F. Mu. *Spontaneous emergence of complex optimal networks through evolutionary adaptation*. Computers and Chemical Engineering, 2004.

[26] R. Albert and A.-L. Barabási. *Statistical mechanics of complex networks*. Review of Modern Physics, 74, pp. 47-97, 2002.

[27] A.-L. Barabási. *Scale-Free Networks*. Scientific American, 288, pp. 60-69, May 2003.

[28] S.H. Strogatz. *Exploring complex networks*. Nature, Vol. 410, Issue 6825, pp. 268-276, 2001.

[29] R. Albert, H. Jeong and A.-L. Barabási. *Diameter of the world-wide web*. Nature, Vol. 401, Issue 6749, pp. 130-131, 1999.

[30] D.J. Watts and S.H. Strogatz. *Collective dynamics of 'small-world' networks*. Nature, Vol. 393, Issue 6684, pp. 440-442,v 1998.

[31] J.M. Kleinberg. *Navigation in a small world*. Nature, Vol. 406, Issue 6798, pp. 845, 2000.

[32] M.E.J Newman. *Modularity and community structure in networks*. In proc. National Academy of Sciences, Vol. 103, No. 23, pp. 8577-8582, 2006.

[33] R. Guimera, M. Sales-Pardo and L.A.N. Amaral. A network-based method for target selection in metabolic networks. Bioinformatics 23, 1616-1622 (2007).

[34] Smart, AG, Amaral, LAN, Ottino, JM. Cascading failure and robustness in metabolic networks. Proc. Natl. Acad. Sci. U. S. A. 105, 13223-13228 (2008).