Network Restructuring Models for Improved ATS Forecasts

Daniel A. DeLaurentis, Tatsuya Kotegawa, Aaron Sengstacken
School of Aeronautics and Astronautics
Purdue University
West Lafayette, IN 47907 USA
ddelaure@purdue.edu, tat@purdue.edu, asengsta@purdue.edu

***Abstract*— Current air traffic forecast methods employed by the FAA function under the assumption that the flight route network will not change, that is, no new flight routes will be added and no existing flight routes will be removed. However, in reality the competitive nature of the airline industry is such that new routes are routinely added between cities possessing significant passenger demand while other city-pairs are removed. This paper investigates models for forecasting network reconfiguration that exploit knowledge of network structure in the Air Transportation System (ATS), with the goal of improving overall forecast that drives policy and infrastructure enhancement decision-making.**

Keywords-forecast; network theory; air traffic

I. INTRODUCTION

In order to synthesize long term plans for new technology, infrastructure improvements, policy enhancements, and regulations for the Air Transportation System (ATS), an understanding of air traffic dynamics is needed (i.e., determining how, when and where would air traffic arise or shift in the future). To meet this need, the FAA Air Traffic Organization (ATO) Office of Performance Analysis and Strategy (PAS) produces air traffic forecasts to project future demand, identify operational shortfalls, determine workforce requirements, and estimate the benefits of future investments. In the current forecast algorithm, the projected schedules are based upon the assumption that the future route network structure will be the same as the current network structure. That is, no new direct service routes are added between cities and, thus, the existing airline hub airports will continue to operate as hub airports.

However, the flight service route network structure is likely to change over time. The competitive nature of the airline industry is such that new direct routes are routinely added between cities with significant passenger demand and routes are also removed when demand dwindles. In addition, the location and number of airline hubs are not fixed; within the past several years, two major hubs have been eliminated (St. Louis and Pittsburgh), one airline hub opened and subsequently closed (Washington Dulles International Airport), and several other hubs were substantially restructured. Looking further, scenarios are now taking shape in which environmentally-inspired imperatives may significantly modify the feasible sets of operations and network reconfiguration states. Overall, in order to enhance the ATS forecast precision, a better understanding of restructuring dynamics is required. *Motivated by this goal, research described in this paper is focused on investigating several models for forecasting the mechanism of network restructuring, in particular the aspect of new flight service route formation.* Families of parameters that describe the network topology are used as predictor variables in these models.

The remainder of the paper is organized as follows. After an introduction to network theory and some examples of its use in previous efforts for analyzing the ATS (Section II), Section III describes the data source and assumptions for all analysis. Detailed explanation of the three forecast algorithms developed up to date, along with key implications will follow in Section IV. Section V summarizes the interim results from these forecast algorithms.

II. NETWORK THEORY

A. Background

Multiple networks subsist in the overall ATS; the primary ones are summarized in Table I. The transport network topology was analyzed in the present study in which airports (nodes) are interconnected by flight routes (links). Modern Network Theory (also known as Network Science)[1,2] has produced powerful results from multiple domains (e.g. physics, information, social science, biology) in recent years concerning how real world networks evolve. Some researchers have begun to explore application for analyzing air transportation networks. Guimera et al analyzed the worldwide air transportation network topology and computed measures which characterized the relative importance of cities/airports.[3] Bonnefoy and Hansman[4] used a plot of the weighted degree distribution for light jet operations to understand the capability of airports to attract the use of Very Light Jets (VLJs). A significant body of work exists in the

ISBN: 978-0-615-20720-9

TABLE I. MULTIPLE, INTERACTING NETWORKS IN THE ATS

Network	Node (N) & Link(L)	Time Scale
Demand	N : *Homes/Business* L : *Demand for Trips*	Months/Years
Mobility	N : *Origin/Destination* L : *Actual PAX trips*	Days/Weeks
Transport	N: *Airports* L: *Flight Routes*	Days/Weeks
Operator	N: *Aircraft / Crew* L: *Mission*	Hours
Infrastructure	N: *Waypoints and Airports* L: *Flight Routes*	Months

TABLE II. DEFINITIONS FOR SELECTED NETWORK MEASURES

Parameter	Symbol	Description
Node	*N/A*	Airport
Node Degree	k_i	Number of flight routes existing at node *i*
Node Weight	w_i	Amount of operations associated with node *i*
Link Weight	r_{ij}	Amount of operations between node *i* and *j*
Clustering Coefficient	C_i	Measure of local cohesiveness for a node. Higher C_i implies that it is more likely an alternate connection path exists when a existing link fails
Eigenvector Centrality	x_i	A centrality measure of a node determined by its own and neighbors' degree. In the transport network, the importance of one airport is determined not only by its own number of routes supported, but also the number of routes and traffic level of airports with which it directly connects (an airport with high eigenvector centrality is likely to be very busy itself and also connected to other busy airports)
Population*	pop_i	Population within a 50 mile radius of node *i*

related domain of operations research on the design of optimal networks for particular instances and applications (e.g. schedule for an airline). However, these approaches generally do not pursue insight into the underlying structure of networks, the role this structure plays in future designs, nor the interplay between networks from multiple domains. Examination of the ATS using network theory at the national level and assessment of associated analysis models and techniques as a framework to provide both insight into ATS structure and a useful systems analysis has been a topic for our work[5]. The forecast of service route restructuring presented in this paper is one example application. Table II summarizes key network theory parameters that will be discussed and utilized for the remainder of this paper. More details can be found in [6].

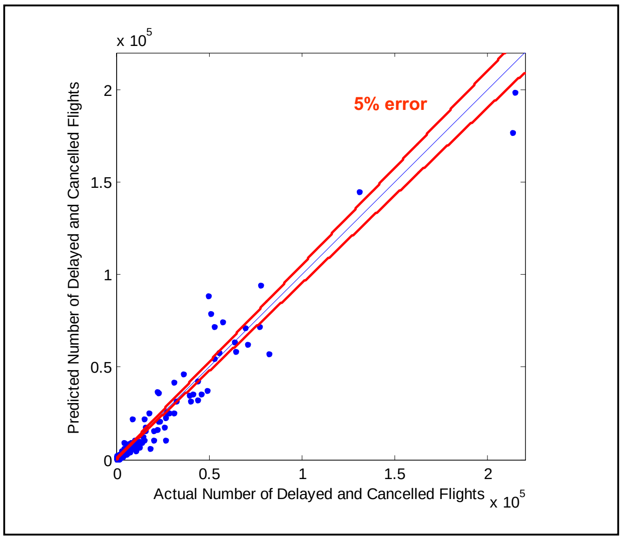


Figure 1. Prediction validation for 2004 delayed operations regression model (each data point is an airport)

The manner in which some of these parameters translate into real world performance and operations metrics is also topic of ongoing research[4]. One example of such a mapping is depicted in Eq. (1). This expression is a multivariate regression model for predicting the number of delayed operations for an airport using its degree, clustering coefficient, eigenvector centrality, degree weight and surrounding population as predictor variables.

$$\sqrt{Delayed\ Ops\ per\ Year} = 0.01928 + 0.147k_i + 0.02606C_i + 0.56722x_i + 0.20758w_i + 0.07462pop_i. \quad (1)$$

All variables are normalized using the corresponding maximum value, and the model produces a good coefficient of determination (R^2 = 0.95). The graph shown in Figure 1 displays the comparison between the actual and predicted number of delayed operations (for airports that registered at least one delay) for the 2004 ATS. A 5% error interval is also included. Eigenvector centrality and degree compose the majority of the regression model (significantly high F-values and Type II Sum of Squares compared to the other variables in the model). The number of expected delays can be forecasted using Eq. (1), but in order to reduce traffic congestion more attention should be placed on airports with not only high degree but also higher eigenvector centrality since these two

ISBN: 978-0-615-20720-9

variables are anticipated to be the main source of operation delays. The primary implication for utilizing network theory as an ATS analysis tool, then, is that these measures can be efficient indicators of network operational performance. Also, focusing on the high-level characteristic of the ATS network generates deeper understanding on the nature of the ATS without being overwhelmed by its complexity.

III. DATA SOURCE AND ASSUMPTIONS

The primary research conducted under this study follows a similar approach to the delay regression model presented in the previous section. The objective is to determine if network theory parameters can be utilized to identify unconnected city-pairs that are most likely to connect in the future. The data used for this study was obtained from Air Carrier Statistics database family maintained by the U.S. Bureau of Transportation Statistics[7]. In particular, the Form 41 T- 100 Domestic segment (All US Carriers) database was used to construct the network studied. The BTS monitors 2627 total airports; however, the ATS network analyzed in this study was restricted to airports that had at least one cumulative commercial flight since 1990. This criterion reduces the network size to 887 nodes (airports). Several different measures are available for use in defining a link, such as the number of passengers, available seats, flights scheduled or actually performed. Since the transport network was explored in this research, a link was constituted by performed passenger flights per year between airports. Each flight route was required to have a minimum of 24 annual flights to be defined as a link in order to filter out any spontaneous, irregular flights that may bring 'noise' to the network analysis. To further simplify the analysis, the ATS network was assumed to be undirected and the number of arrival and departure operations were simply added together to compute the w_i and r_{ij}.

The source of the ATS network evolution can be broken down into four basic categories—flight route addition due to network expansion or reconfiguration, and flight route removal due to network contraction or reconfiguration. Reconfiguration refers to the 're-wiring' of links within a defined set of nodes; no new nodes are added and no pre-existing nodes are removed to create or destroy a link. Network contraction and expansion are opposite to reconfiguration, with links either being created by connecting to newly developed nodes or removed by detaching nodes from an existing set. Figure 2 illustrates the morphing of the ATS network categorized in these four evolution sets—it can be seen that the vast majority of the flight routes removed and created are a result of network reconfiguration. Thus, all forecast model and results described in the following section will only examine the mechanism of flight route construction due to reconfiguration. Investigation of the mechanism for the other three evolution categories (flight route removal due to network contraction / reconfiguration and flight route addition due to expansion) can be done relatively easily by supplying historical data sets to the algorithm that corresponds to the evolution category of interest.

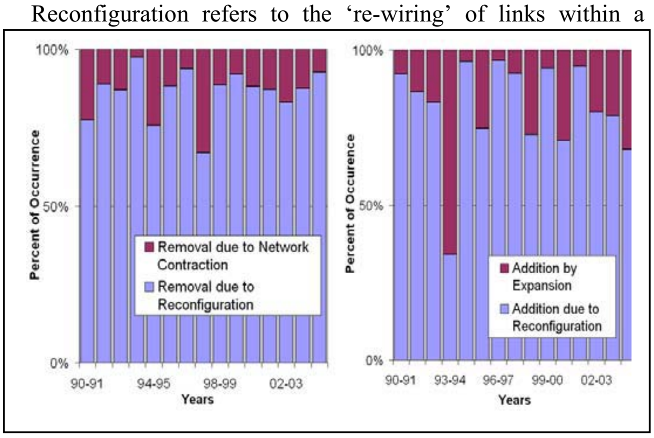


Figure 2. Variation in source of ATS network topology evolution

IV. ROUTE CONSTRUCTION FORECAST ALGORITHMS

Three prototype forecast algorithms were created, compared and contrasted a) the logistic regression model, b) fitness function model and c) the artificial neural network approach. In this paper, the logistic regression model is discussed in detail. A brief summary for each approach is listed below.

Logistic regression is a statistical method to train a probability curve for event occurrence based on historical data input. The event for which the occurrence probability is calculated will be the construction of a new flight route between unconnected city-pairs and the inputs will be the parametric characteristics of the flight route. The iteratively-reweighted least squares (IRLS) method was utilized as the algorithm to fit the regression model with historical data.

Fitness function model is a network growth logic which operates under principles of the scale free network model where nodes with higher importance, or fitness value, are granted a higher probability to construct a new link. The initial composition of the function that computes nodal fitness projects growth that favors highly connected nodes (a hub-and-spoke type growth) that is typical in the ATS today. However, the fitness function can be modified to investigate the efficacy of various types of network growth mechanisms corresponding to a mix of different business models.

The **Artificial Neural Network (ANN)** is composed of a set of interconnected neurons that mimic human brain activity in attempting to develop optimal input-output mappings for prediction. Though some underlying fundamentals are similar to logistic regression, the ANN usually has higher precision. One drawback is that the relationship between input and output remains to be a 'black box'—it cannot be expressed in terms of explicit equations as is typical in conventional statistical models. Also, due to the higher computational requirements of the ANN algorithm, the network to be analyzed via ANN must be kept relatively small.

ISBN: 978-0-615-20720-9

A. Logistic Regression Model

A new flight route in the network context represents a new pathway between unconnected node pairs. The characteristic of new routes can be described by observing the traits of the airport pairs that create the route. The traits of the airport pair can be captured from two perspectives: a) by examining the list of parameters for each of the airports and b) by examining the relative difference of parameters between the airports. A record of network parameters (referred to as [*parameter*] *list*) for each airport involved in a new route indicates the type of airports that are most likely to be involved in a new connection. On the other hand, a record of parameter difference, or *deviation*, between the airport pairs produces a pair-wise measure that may be better able to characterize new connection formation. In particular, the type of connection can be categorized into either homogenous (connection between 'large-large' airports or 'small-small' airports) or heterogeneous (connection between 'small-large' airports).

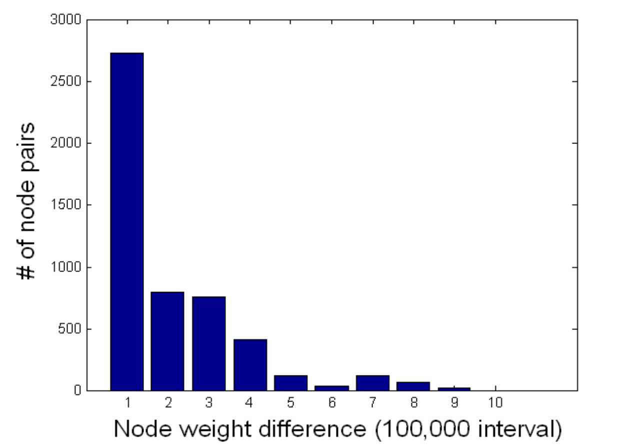


Figure 3. Node weight list distribution for new flight routes established in the ATS network between 1990 and 2005

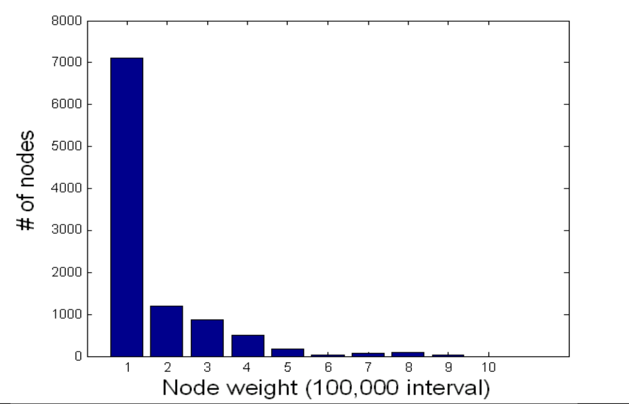


Figure 4. Node weight deviation divergence distribution for new flight routes established in the ATS network between 1990 and 2005

By combining insights from the parameter list and parameter deviation traits of the airport pairs that construct a new connection, the patterns that facilitate new flight routes can be extracted. The histograms in Figure 3 and 4 illustrate the distribution of node weight list and deviation for airports that formed new flight routes in the ATS network between 1990 and 2005. Figure 3 shows that most of the nodes involved in flight route constructions had relatively low traffic (between 1 and 100,000 annual operations), and Figure 4 shows that the difference (deviation) in the traffic of nodes involved in new links was mostly homogenous. The implication is that most new flight routes are established between airports that have lower traffic. A similar exercise was carried out for the remainder of the network parameters listed in Table II.

Parametric data are fed into the logistic regression model via design matrix X which ultimately gives node pairs that follow such trends higher likelihood of connection. Design matrix X is structured as shown in Eq. (2) for which all network theory variables in Table II are included, along with the distance information between node i and j. The second column of X, rij , signifies the occurrence of a new flight route construction for node i and j between observation years. If a new route is established between i and j a '1' is placed in r_{ij} and if not a '0' is placed.

$$X = \begin{bmatrix} 1 & r_{ij} & k_i & k_j & abs(k_i - k_j) & w_i & \cdots \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (2)$$

Based on the design matrix input, the regression model computes the variable parameter estimates using the standard iteratively-reweighted least squares (IRLS) algorithm and feeds the estimates into Eq. (3) which computes the probability of an unconnected node pair ij will construct a new flight route.

$$P_{connect,\,ij} = \frac{1}{1 + e^{-\hat{x}X_2,ij}} \quad (3)$$

X_2 in Eq. (2) is a matrix that contains the network parameter and parameter deviation information structured identically to X, except X_2 only includes data for unconnected node pairs. The design matrix X contains information for all connected and non-connected node pairs for probability curve training purposes. After Eq. (2) has been computed, $P_{connect,ij}$ is compared to a random number (*rand*) between 0 and 1 and the algorithm predicts a new flight route construction between node i and j if $P_{connect,ij} > rand$.

B. Accuracy Measures

Three accuracy measures are employed to assess the forecast precision.

$$Accuracy\ 1 = \frac{number\ of\ correctly\ predicted\ routes}{total\ number\ of\ predicted\ routes} \quad (3)$$

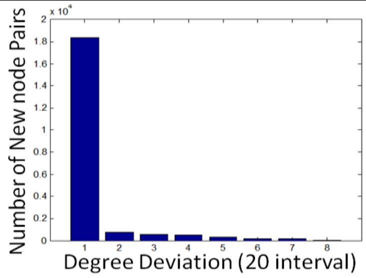
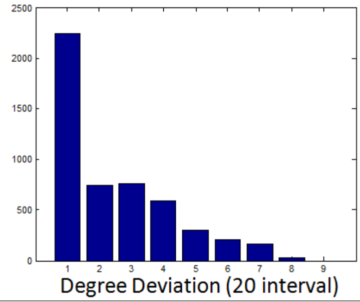
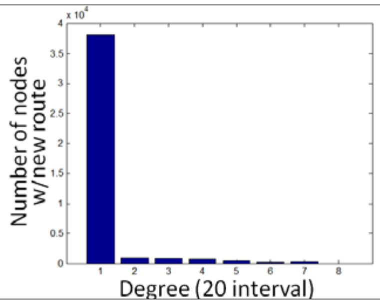
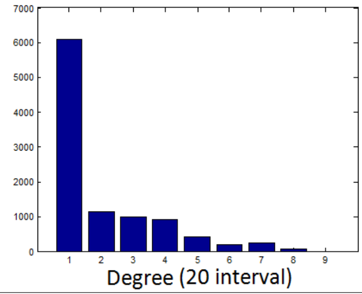
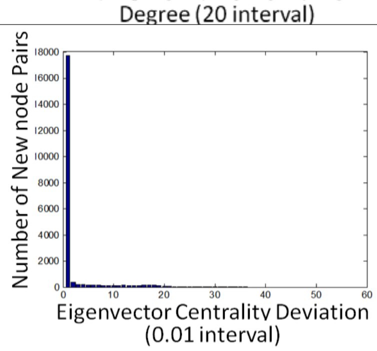
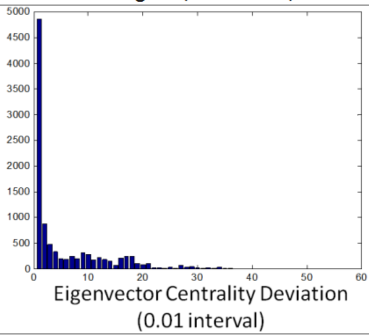
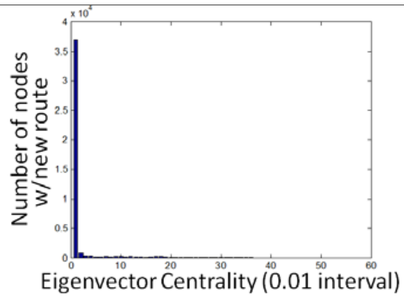
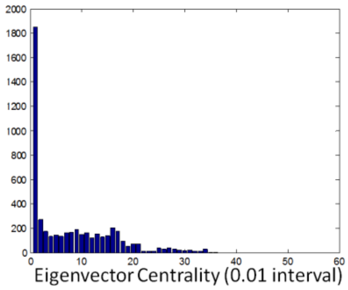
 ISBN: 978-0-615-20720-9

$$Accuracy\ 2 = \frac{number\ of\ correctly\ predicted\ routes}{number\ of\ actual\ new\ routes} \quad (4)$$

Accuracy 1 shown in Eq. (3) was used to check how many new routes the forecast algorithm was predicting in order to obtain the correct new route. If the algorithm is predicting thousands of new routes to acquire only few correct new routes, accuracy 1 will be very low. On the other hand, accuracy 2, shown in Eq. (4) simply describes how many of the predicted new routes were correct, with respect to the number of actual new routes.

Accuracy 3 is a special type of accuracy measure which the coherence in distribution of characteristic trends for new links between the data and forecast model is examined. This is done by comparing the node parameter list and divergence histogram curve from the data and forecast algorithm, such as those seen in Figure 3 and Figure 4. The goal of employing accuracy 3 is to make sure that the forecast methods are predicting the future ATS network in the 'right direction'; a formal equation to describe accuracy 3 currently does not exist.

TABLE III. SUMMARY OF LOGISTIC REGRESSION RESULTS FOR ACCURACY 3

	Logistic Regression Results Distribution (1990-2005 Cumulative)	***Historical Distribution(1990-2005 Cumulative)***
Degree Deviation		
Degree Log		
Eigenvector Cetrality Deviation		
Weight Log		

ISBN: 978-0-615-20720-9

C. Logistic Regression Results and Discussion

Results for an iteration of the logistic regression model are shown on Table IV; output for each year is an average over 10 runs. Inputs to the model consist of all the network theory parameters and deviation values for variables listed on Table II as well as the distance between airport pairs. Forecasts are done on a year-over-year basis; that is, parameters from only the previous year are utilized for the forecast.

The 'Correctly Predicted' column indicates the total number of correctly predicted routes for that year while the 'Total Predicted' column indicates the total number of predicted routes from the forecast algorithm. The logistic regression model has relatively high Accuracy 2 but low Accuracy 1 across all years, indicating that the algorithm can correctly forecast a significant number of new routes but does so by forecasting many additional routes in the process.

Accuracy 3 outcomes for both list and deviation distributions for degree and eigenvector centrality are shown in Table III. The first important finding was that the distributions produced by the logistic regression differ from those observed from historical data. In particular, the logistic regression allocates too much preference for connection to airports and airport pairs with small valued network parameters. This result, however, does not mask a second important finding from these results: across both parameter list and deviation distributions [for nodes with new connections], new routes were being established primarily between 'small' airports, whether defining small by degree or centrality significance. Appropriately, the logistic regression model distributes higher probability to establish connections between these small airports…it just distributes too much importance to these. Owing to the fact that the current ATS network is dominated by hub-and-spoke style architectures[8], there exist many small, spoke airports and very few large, hub airports. Since there are more small nodes in the network, many small-to-small airport pairs arise as candidates for flight route construction. Abundant small-to-small airport connection candidates coupled with the forecast model favoring small-to-small airport connections from historical trends results in significant number of over-predictions for small-to small airport connections. This conclusion is the message conveyed from simultaneous consideration of Accuracy 1 and 3 metrics.

TABLE IV. SUMMARY OF LOGISTIC REGRESSION RESULTS

Year	Correctly Predicted	Total Predictions	Accuracy1	Accuracy 2
1990	N/A	N/A	N/A	N/A
1991	83.7	1217.6	0.0700	0.4314
1992	88.7	1427.6	0.0624	0.4264
1993	68.8	1065.1	0.0658	0.4145
1994	80.0	1365.8	0.0594	0.4645
1995	146.7	1395.8	0.1077	0.3976
1996	82.6	1409.8	0.0596	0.4325
1997	78.2	1489.0	0.0535	0.3476
1998	62.7	1177.6	0.0552	0.3968
1999	71.3	1488.4	0.0488	0.4006
2000	113.6	1980.2	0.0580	0.3381
2001	52.0	1286.3	0.0410	0.3824
2002	388.8	2328.9	0.1676	0.2878
2003	128.2	1123.5	0.1159	0.2728
2004	120.5	1043.1	0.1157	0.2953
2005	104.4	1088.1	0.0970	0.2806
Average	111.3	1392.5	0.0785	0.3713

D. Brief Introduction and Analysis of the Artificial Neural Network

The Artificial Neural Network (ANN) is composed of a set of interconnected neurons that mimic human brain activity. Through supervised back-propagation training techniques, an ANN is able to achieve desired input-output mapping by adjusting the weights associated with each neuronal connection in the network. While the basic concept underlying ANN is similar to that of the logistic regression, the ANN usually has higher accuracy due to its higher degrees of freedom. However, the relationship between input and output for a trained ANN remains difficult to describe, unlike the logistic regression model. Also, the size of the network that can be analyzed with an ANN was restricted to a smaller size (~250 nodes) than for the logistic regression model due to the computational intensity of the ANN training algorithm.

The ANN approach proceeded via a feed-forward, fully-connected network algorithm[9]. After training the ANN with historical data, it was used to predict connections between two airport nodes. To capture the ATS network dynamics, the airport metrics for the previous three years were used at the input neuron layer resulting in an input layer of 63 neurons, a hidden layer consisting of 126 neurons, and a single output neuron. The input neurons represent two airport nodes, the hidden layer neurons used a *tan-sig* activation function, and the output neuron used a *log-sig* activation function. The single output neuron indicated the connectivity between the two airport nodes—1 for connected, 0 unconnected. The training data consisted of 50% of the historical data, while 25% was used for testing and 25% for validation. Once again, it should be noted that, for research reported here, the ANN was used to forecast only a subset of the ATS, mainly due to current computational limitations. In particular, historical data from the American Airlines (composed of routes operated by American Airlines, American Eagle and Executive Airline) and Southwest Airlines Transport Networks were employed to evaluate the accuracy of the ANN algorithm.

The trained ANN had extremely high accuracy rates in predicting new flight routes, with a minimum value of 70% for both Southwest and American Airlines Transport Networks. The ATS network used for the ANN forecast algorithm was abbreviated to 224 nodes (recalling that the logistic regression model considers 887 nodes). The 224 nodes included in the ANN training were the most active nodes in the ATS, excluding smaller, inactive airport nodes. Results for the two airline network forecasts along with translation to Accuracy 1 and 2 are shown below in Table V and VI.

ISBN: 978-0-615-20720-9

TABLE V. TRAINED ANN RESULTS (SOUTHWEST AIRLINES NETWORK)

		Historical Data	
		Connect	Disconnect
Network Simulation	Connect	2850	1104
	Disconnect	738	380706
Accuracy 1 = 72.08%			
Accuracy 2 = 79.43%			

TABLE VI. TRAINED ANN RESULTS (AMERICAN AIRLINES NETWORK)

		Historical Data	
		Connect	Disconnect
Network Simulation	Connect	7291	2788
	Disconnect	2962	372357
Accuracy 1 = 72.33%			
Accuracy 2 = 71.11%			

The results displayed in Tables V and VI are separated into four cells. The sum of rows in the table describes the forecast results by the ANN, and the sum of columns describes the actual status of the unconnected node pairs. For example, in the American Airlines results (Table VI), the ANN forecasted a total of (7,291+2,788) 10,079 new flight routes (city pairs). Out of this total number of predicted new routes, in actuality 7,291 formed connections as determined from the historical data while 2,788 were disconnected (i.e., 'false alarms'). Similarly, the ANN forecasted that (2,962+372,357) 375,319 node pairs would remain disconnected but in actuality 2,962 out of these 375,319 made a connection. The overall accuracy results of the ANN are impressive when compared to the logistic regression model; however, it is difficult to extract any insights on the ATS evolution mechanism itself, since the relationships inside the trained ANN do not relate directly to the meaning of the input data (it is just an optimal prediction configuration). It is noted here again that the network size was significantly reduced in the ANN case.

E. Brief Introduction and Analysis of the Fitness Function Method

The fitness function model is a network growth logic which operates under the fundamentals of scale-free network model[8]. In this type of growth mechanism and network model, nodes with higher importance, or fitness value, are granted a higher probability to participate in a new link. The procedure begins by reading in the network topology from the previous year. For each node in the network, a fitness value was calculated through a specific functional composition of several nodal metrics listed in Table II. The initial functional composition used in the research was simply a ratio of individual nodal parameter of airports and the network sum of that parameter. For example, if a particular node has k=10 and the total k for the entire network is 100, its fitness function will be 10/100 = 0.1. This type of fitness function projects growth that favors highly connected and important nodes (a hub-and-spoke type growth) that is typical in the ATS today.

However, the fitness function can be modified to allow various types of network growth mechanisms corresponding to a different mix of business models that might emerge in the future. This ability to tailor scenarios in an explicit manner dealing directly with service provider behavior is an attractive advantage of this approach. Subsequent to the fitness calculation, a pair-wise fitness was calculated for each node pair, and this was used to determine a probability of linking for all unconnected node pairs in the network. Links are added to the topology based on those pairs with high link probability (under some randomness).

Unlike the logistic regression model and the ANN approach, for which historical trends were directly projected to forecasting, the fitness function algorithm employs insights from growth models developed from the network science domain. Various combinations of network parameters (summarized in Table II) were investigated for the fitness function to determine which combination best suited the forecasting task. The fitness function that combines distance, degree, eigenvector centrality and nodal weight produced the forecast with highest accuracy. Results for an iteration of this fitness function model are shown on Table VII, noting once again that output for each year is an average over 10 runs. In comparison with the logistic regression model, the fitness function model produces poor results in the form of Accuracy 1 and 2. The problem of 'over-forecasting' was not resolved.

Surprisingly, however, the fitness function has improved Accuracy 3 results especially in the parameter list histograms (not displayed). The fitness function seems to develop the correct traits for choosing the nodes that develop new routes, but the specific prediction of 'which nodes' is relatively low perhaps due to the large pool of new connection candidates (there are approximately 4 million unconnected node pairs to choose from!). Even though Accuracy 1 and 2 for the fitness function approach may be lower than the ANN or logistic

TABLE VII. . SUMMARY OF LOGISTIC REGRESSION RESULTS

Year	Correctly Predicted	Total Predictions	Accuracy1	Accuracy 2
1990	N/A	N/A	N/A	N/A
1991	36.9	1217.6	0.0700	0.4314
1992	45.3	1427.6	0.0624	0.4264
1993	28.5	1065.1	0.0658	0.4145
1994	31.1	1365.8	0.0594	0.4645
1995	75.3	1395.8	0.1077	0.3976
1996	33.6	1409.8	0.0596	0.4325
1997	46.7	1489.0	0.0535	0.3476
1998	30.9	1177.6	0.0552	0.3968
1999	29.8	1488.4	0.0488	0.4006
2000	51.6	1980.2	0.0580	0.3381
2001	23.3	1286.3	0.0410	0.3824
2002	139.5	2328.9	0.1676	0.2875
2003	54.7	1123.5	0.1159	0.2728
2004	45.4	1043.1	0.1157	0.2953
2005	48.1	1088.1	0.0970	0.2806
Average	111.3	1392.5	0.0785	0.3713

ISBN: 978-0-615-20720-9

regression, the results appear sensible for an algorithm that does not depend on historical trends. This latter fact makes it difficult to judge the fitness function model performance comparative to the other two models described in this paper that purely utilizes historical trends as input. In addition, models that are independent of historical trends have an distinct advantage over models are that dependent—it is more likely that the forecast accuracy can be maintained even if the characteristic of the ATS is significantly shifted from the nature of past trends. The logistic regression and ANN model will be able to forecast the ATS at higher accuracy levels if and only if the ATS continues to evolve in the direction it has been evolving. However, if new policy, technology or operation methods that revolutionizes the ATS are introduced, the logistic and ANN will require new historical data to accumulate before further accurate forecast can be made. With algorithms like the fitness function model, change in ATS characteristics can be readily introduced by appropriately adjusting the fitness function calculation. Combinations of the models, therefore, also seem like a promising avenue for further research.

V. Conclusion and Future Work

Current air traffic forecast methods employed at the FAA function under the assumption that the flight route network will not change, that is, no new flight routes will be added and no existing flight routes will be removed. In reality, the competitive nature of the airline industry and the potential need for new policies relating to the environment are such that new routes are routinely added between cities possessing significant passenger demand and other city-pairs are removed.

Research performed under this project and described in this paper explored means to understand network reconfiguration dynamics in the ATS. In particular, the aim was to expand the capabilities of the existing ATS forecast methods developed by the FAA, ultimately leading to improved decision-support in maintaining and enhancing the ATS. Employing network theory variables and concepts as a foundation to characterize the network of flight service routes in the ATS, three families of models were developed and tested: a) Logistic regression, b) a network topology based fitness function method, and c) an artificial neural network (ANN) algorithm. Results indicate that each has merit under differing accuracy metrics and each has methodological drawbacks. Advantages and disadvantages were documented. Overall, the logistic regression appears to capture more likely new city pairs, though in an inefficient manner as compared to the fitness function model. The ANN has superb prediction capabilities but was only tested on a sub-set of the network data due to computational and time constraints of this short duration study.

There still is much room for expansion in the current ATS forecast capabilities described in this paper. First, means are available to increase all accuracy measures for each forecast algorithm. Some proposed methods to meet this goal include the implementation of more accurate and precise data of the ATS (i.e. ETMS instead of BTS), parsing the ATS network into sub-networks such as specific aircraft class or service provider, and removing variables deemed insignificant to the model or causing high multicollinearity. Forecasting based on multiple previous years for the logistic regression and fitness function model may also increase accuracy. Eventually, multiple forecast methods may be merged to go beyond the limit of individual methods. Second, enhanced ability to implement future scenarios will greatly improve the value of this research. All forecast methods are essentially based under an assumption in which the future ATS will grow in the way it has in the past. However, this is not true. New types of airline services, emergence of innovative technologies as well as new regulations and policies will impact the future state of network configurations; each of these may also drastically change the fundamental principles of operation of the ATS. In order to anticipate the effect for some of these ground-breaking factors in the forecast algorithms, a better understanding and mapping of the ATS is required. Finally, combining the best of the algorithms for predicting new city pairs with the FAA's current forecast method (based largely on the FRATAR algorithm) constitutes the most immediate next step.

Acknowledgment

The authors acknowledge sponsorship of this research by the FAA Air Traffic Organization Office of Performance Analysis and Strategy. The contributions of Donald Fry, En-pei Han and Sricharan Ayyalasomayajula at Purdue University are also acknowledged and appreciated.

References

[1] Newman, M., "The structure and function of complex networks," *SIAM Review*, Vol. 45, 2003, pp. 167-256.

[2] Albert, R., Barabási, A.-L., "Statistical mechanics of complex networks," *Rev. Mod. Phys*. Vol. 74, pp. 47-97 2002.

[3] Guimera, R., Mossa, S., Turtschi, A., Amaral, L., "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proc. Nat. Acad. Sci.*, Vol. 102, 2005, pp. 7794-7799.

[4] Bonnefoy, P., Hansman, R., "Potential Impacts of Very Light Jets in the National Airspace System," *Journal of Aircraft*, Vol. 44, No. 4, 2007. pp. 1318-1326.

[5] DeLaurentis, D., Han, E-P., Kotegawa, T., "Network-Theoretic Approach for Analyzing Connectivity in Air Transportation Capacity Networks," accepted for publication in *AIAA Journal of Aircraft*..

[6] DeLaurentis, D., Kotegawa, T., "Establishment of a Network-based Simulation of Future Air Transportation Concepts," AIAA 2006-7719, Sep., 2006

[7] U.S. Bureau of Transportation Statistics (BTS), www.bts.gov; [accessed numerous times over study period]

[8] Conway S., "Scale-free Networks and Commercial Air Carrier Transportation in the United States," Proceedings of the 24th Congress of the International Council of the Aeronautical Sciences (ICAS), Yokohama, Japan, 2004.

[9] Tsoukalas, L. H. and R. E. Uhrig (1997). *Fuzzy and Neural Approaches in Engineering*,John Wiley and Sons, Inc.

 ISBN: 978-0-615-20720-9