Characterization of Traffic and Structure in the U.S. Airport Network

Vineet Mehta, Feanil Patel, Yan Glina, Matthew Schmidt, Ben Miller and Nadya Bliss

MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02421

Abstract—In this paper we seek to characterize traffic in the U.S. air transportation system, and to subsequently develop improved models of traffic demand. We model the air traffic within the U.S. national airspace system as dynamic weighted network. We employ techniques advanced by work in complex networks over the past several years in characterizing the structure and dynamics of the U.S. airport network. We show that the airport network is more dynamic over successive days than has been previously reported. The network has some properties that appear stationary over time, while others exhibit a high degree of variation. We characterize the network and its dynamics using structural measures such as degree distributions and clustering coefficients. We employ spectral analysis to show that dominant eigenvectors of the network are nearly stationary with time. We use this observation to suggest how low dimensional models of traffic demand in the airport network can be fashioned.

I. INTRODUCTION

In this paper our aim is to characterize traffic in US national air space in order to gain insight that would subsequently lead to developing detailed demand models. We model the US air transportation system as a network to assist with the characterization of traffic demand. The U.S. airport network has been the subject of previous study, although not with the aim of developing traffic demand models. Xu and Harriss [1] have studied the structure of the U.S. intercity passenger air transportation network. The cumulative degree distribution of the network was shown to have a powerlaw tail. The vertex strengths (with number of passengers, fare and distance as incident edge weights) were shown to grow as a power of the vertex degree. As a result airports with higher degree were found to handle more traffic than those with smaller degree. The clustering property of the network was also examined through clustering coefficients. The clustering coefficient as a function of degree was found to have a power-law distribution, with lower clustering coefficients for higher degree vertices. This behavior is characteristic of a hierarchical network structure.

There have been similar studies of the airport network structure for other national networks and the world wide network. The evolution of the Chinese airport network over successive years has been analyzed by Zhang, et. al. [2]. The traffic growth in this network was found to correlate well with the growth in Chinese GDP. The Chinese airport network was shown to have a degree tail distribution which exhibits a two-regime power-law. The relationship between the clustering coefficient and degree showed vertices with smaller degree to have larger clustering coefficient. The vertex traffic strength was shown to grow as a power of vertex degree. The topological properties on the Chinese airport network were seen to remain nearly fixed over 2002 to 2009, even though airports were added and removed from the network and the network saw an exponential growth in passenger and cargo traffic. Bagler [3] has studied the weighted airport network of India. da Rocha [4] examined the structure and evolution of the Brazilian airport network over consecutive years and found similar degree, strength and clustering characteristics as other examinations of national airport networks. The worldwide air transportation network was also shown to exhibit some characteristics similar to other national airport networks [5].

We examine the temporal dynamics of the network by considering the behavior over a number of successive days. We find the network to be highly dynamic over successive days, and exhibit an unexpected level of spatial and temporal complexity. In the next section (Section II) we present the approach we have employed in constructing the daily airport networks. The resulting airport networks and their spatial structure is shown in Section III. The structural characteristics, as well as the dynamic behavior of the US airport network is discussed in Section IV. An analysis of the spectral characteristics of the airport network using the adjacency matrix is presented in Section V.

II. NETWORK CONSTRUCTION

In order to construct the US airport pair network, we employ the Aircraft Situation Display to Industry (ASDI) [6] data feed provided by the Federal Aviation Administration (FAA). The generation of graphs for the airport network relies on the use of *Flight Management Information (RT)* messages. The *RT* messages contain information about a flight's identity, origin, destination, departure/arrival times, and filed route. We construct each graph by either adding new airport pairs (edges) or updating the flight count on

This work was sponsored by the Assistant Secretary of Defense (Research and Engineering) under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

an existing edge for all flights reported by *RT* messages occurring within a 24-hour period. These graphs include flights with origin or destination outside the continental US territory. The graphs are then truncated to retain only airports that reside inside the bounding spherical quadrilateral specified by (latitude, longitude)-coordinates: $(25^\circ, -125^\circ)$ and $(50^\circ, -65^\circ)$.

The outcome of the aforementioned processing are daily graphs G_n , where n is the daily index. Each graph G_n is specified by a set of edges E_n and a set of vertices V_n . The set V_n is composed of vertices $\{v_i\}_n$. Each vertex v_i denotes an airport with geographic position $\vartheta = (\phi, \theta)$, where ϕ and θ specify the latitude and longitude respectively. The set E_n is composed of vertex two-tuples $\{(v_i, v_j)_n | v_i, v_j \in V_n\}$. For the analysis in this paper we consider $\{G_n\}$ to be undirected weighted graphs. The weight of each edge $e_{k,n} = (v_i, v_j)_n$ is denoted by $w_{k,n}$, with k as the edge index.

Here we define three parameters that are employed in subsequent sections for quantifying graph characteristics: vertex degree, vertex strength, and local clustering coefficient. The degree κ_i is defined as the number of edges incident upon the vertex v_i . The strength s_i of vertex v_i is defined as the sum of edge weights for all edges incident upon the vertex. In order to define the clustering coefficient C_i , we first define the vertex neighborhood $N_i = \{v_j : e_k = (v_i, v_j) \in E\}$. For a vertex with κ_i neighbors it is possible to define a maximum of $\kappa_i(\kappa_i-1)/2$ edges amongst the neighbors. The local clustering coefficient C_i measures the number of actual edges within the neighborhood as a fraction of the maximum number possible.

$$C_i = \frac{2|\{e_k\}|}{\kappa_i(\kappa_i - 1)}, \ e_k = (v_i, v_j) \in E \land \ v_j \in N_i$$
(1)

III. U.S. AIRPORT NETWORK

The US airport network and its dynamic character is depicted in Figure 2. The graphs shown in Figure 2(a-c) capture the aggregate air traffic between city pairs on a particular date. The graph in Figure 2(d) is the *union-graph* \bar{G}_{\cup} . This graph is computed from a set of N daily weighted graphs $\{G_n\}$, with edge sets $\{E_n\}$ and vertex sets $\{V_n\}$. The edge set \bar{E}_{\cup} and vertex set \bar{V}_{\cup} of \bar{G}_{\cup} are given as:

$$\bar{E}_{\cup} = \bigcup_{n=0}^{N-1} E_n, \quad \bar{V}_{\cup} = \bigcup_{n=0}^{N-1} V_n$$
 (2)

The weight of the k-th edge, \bar{w}_k , in the set \bar{E}_{\cup} is computed as:

$$\bar{w}_k = \frac{1}{N} \sum_{n=0}^{N-1} \hat{w}_{k,n}$$
(3)

The quantity $\hat{w}_{k,n}$ is equal to the weight $w_{k,n}$ for $k \in E_n$, and zero for $k \notin E_n$. The strength \bar{s}_i of vertex $v_i \in \bar{V}_{\cup}$ is the sum of weights $\bar{w}_{l(i)}$, where l(i) is the index over edges in the incidence set L(i) of the vertex v_i .

The graphs shown in Figure 2(a-c) depict the volume of traffic carried on links between city pairs. The straight line edges are meant to depict linkages between city pairs, and should not be mistaken as travel paths between the cities. The edges are assigned colors such that *blues* denote lower traffic volume, and *reds* denote higher traffic volume. The edge colors make it easier to visually discern that daily graphs are dominated by lower volume edges, and have relatively fewer higher volume edges. This is verified by examining the distribution of edge weights in Figure 1.



The variations in spatial edge density and coloring of edges in these plots illustrate the changes in traffic across the national airspace over consecutive days. This is visually most apparent in comparing graph Figure 2(c), which shows the traffic for a Saturday, with graphs for the other days. The size and color of the vertices depict their strength, which is the number of flights that were served by the airport during that day. The major airports such as Chicago O'Hare International Airport (KORD), Hartsfield-Jackson Atlanta Airport (KATL), Denver International Airport (KDEN), and Dallas/Fort Worth International (KDFW) stand out as the busiest airports by traffic volume.

The union-graph \bar{G}_{\cup} shown in Figure 2(d) is computed from a set of 29 daily graphs over the period 26 April 2011 to 24 May 2011. These graphs show expected as well as some unexpected features. We had anticipated that a set of major airports would continue to represent the vertices with dominant strength over consecutive days. However, we did not expect the amount of variation found in the sub-dominant vertices and edges in the set of daily graphs examined. A remarkable contrast is found in comparing the average number of edges in the daily graph $|\bar{E}_n| =$ $\frac{1}{N} \sum |E_n|$ to the number of edges in the expected graph, $|\bar{E}|$. The quantity $|\bar{E}_n|$ is approximately 10*K*, while $|\bar{E}_{\cup}|$ is approximately 80*K*. This surprisingly large difference suggests a level of dynamicism in the daily air traffic over the national airspace that we have not seen reported elsewhere.

IV. STRUCTURAL CHARACTERISTICS

In this section we employ basic structural metrics such as vertex degree, edge weights, and vertex clustering for characterizing both static and dynamic aspects of the daily airport network graphs. The tail distributions of degree κ for G_n over selected days are shown in Figure 3. The degree tail distributions $P_n(\kappa > K)$ for the consecutive days are found to be similar. The mean distribution is found to exhibit a power law decay, $P(\kappa > K) \sim \kappa^{-\alpha}$, with $\alpha \approx 1.17$.

The variation in the clustering coefficient with vertex degree κ is given in Figure 4. The clustering coefficient is found to exhibit approximately a uniform distribution with degree κ . This behavior is in contrast with results reported in other investigations of the airport network. We believe this difference can be explained by examining the clustering coefficient for the *intersection-graph* G_{\cap} . We define the *intersection-graph* \bar{G}_{\cap} as the intersection of as set of graphs $\{G_n\}$. The *intersection-graph* has an edge set $\bar{E}_{\cap} = \cap E_n$, a vertex set $\bar{V}_{\cap} = \cap V_n$, and weights $\{\bar{w}_k\}$ that are averaged over the weight sets $\{w_{k,n}\}$. The variation in the clustering coefficient with vertex degree k for the *intersection-graph* \bar{G}_{\cap} is given in Figure 5. The near monotonic decay of the clustering coefficient of the \bar{G}_{\cap} is consistent with behavior reported by others.

The variation of average vertex strength with vertex degree is given in Figure 6. The growth trend in strength with degree is similar to vertex strength behavior reported in previous studies. This behavior indicates that airports with larger degree tend to also on average carry more traffic on their city pair links.

The temporal dynamics of the US airport network in the aggregate can be seen by examining the daily variation in flight count, as shown in Figure 7. We find that flight counts exhibit a weekly cyclic trend. It is interesting that this periodic trend is also seen in the edge (Figure 8) and vertex counts over successive days. These results indicate the growth in traffic volume during the week is coupled with an increase in the diversity of cities that host flights.

The results in Figure 8 suggest that there are city pairs for which flights occur with periodicity less than a day, and perhaps as infrequently as once in a week, or longer. In fact there are flights that occured only once over the set of 29 days examined. In order to quantify the number of new city pairs visited over successive days, we examine the one lag difference of our graph time series $\{G_n\}$. In particular we consider the complement of the intersection between successive day graphs G_n and G_{n-1} .

$$g_{n,1} = (G_n \cap G_{n-1})^C$$
(4)

The edge count of the complement graphs $\{g_{n,1}\}$ for successive lags in the daily time series is given in Figure 9. This figure shows that the geospatial traffic pattern varies

substantially from day-to-day. Although the range of variation in the edge count over a week is approximately 3K, the disparity in city pairs visited between successive days can be nearly as high as 13K. As a result more than 1/3of the flights between two successive days are to a different set of cities. We have also examined the edge counts for complement graphs $\{g_{n,m}\}$, with m = 7, in an attempt to remove the apparent seven day sesonality as evident in Figure 8. However, the edge counts for $\{g_{n,7}\}$ were also found to be highly non-stationary over successive days.

In an attempt to further characterize the dynamics of the graph time series $\{G_n\}$ we examine the autocorrelation function over the edge weights. In order to compute the autocorrelation we treat \overline{G}_{\cup} as a *root-graph*, for the set of graphs $\{G_n\}$. We then define the set of graphs $\{\widehat{G}_n\}$ that have edge and vertex sets identical to \overline{G}_{\cup} . The edge weights of G_n are mapped onto \widehat{G}_n for shared edges, and set to zero for all other edges. The estimate of the autocorrelation function for the weights of each edge is given as:

$$R_{\tilde{w}_k,\tilde{w}_k}(n) = \frac{1}{N-1} \sum_{m=0}^{N-1} \tilde{w}_{k,m} \tilde{w}_{k,m+n}$$
(5)

The time series $\tilde{w}_{k,n}$ is obtained by converting $w_{k,n}$ to be zero-mean and unit variance. The averaged autocorrelation function over all edges is plotted in Figure 10. The range bars that bracket the averaged autocorrelation function indicate the maximum and minimum values of the autocorrelation function amongst all the edges. The boxes at each lag give the standard deviation in the autocorrelation values over all edges. The average and standard deviation of autocorrelation shows the temporal behavior of flights on the majority of edges in US airport network to be weakly correlated. The maximum values of autocorrelation for each lag show the presence of highly correlated components. These highly correlated components tend to also carry a high volume of traffic. These results suggest that standard time series techniques that employ time series differencing and autocorrelation to construct forecasting models, are unlikely to yield low dimensional models that would capture aggregate, as well as fine scale features of traffic demand. In the next section we examine spectral techniques that may hold greater promise in development of such low dimensional models.

V. SPECTRAL CHARACTERISTICS

In this section we discuss the spectral characteristics of the set of adjacency matrices $\{\hat{A}_n\}$ associated with the normalized graphs $\{\hat{G}_n\}$. The eigenvalue magnitude spectrum $\{|\hat{\lambda}_i|\}$ for a set of consecutive days is given in Figure 11. The structure of the eigenvalue magnitude spectrum is similar over the days examined. The eigenvalue spectrum follows a power-law decay, $\hat{\lambda}_i \sim i^{-\beta}$, with $\beta \approx 1.04$.

The aggregate temporal behavior of the eigenvectors is examined by observing their normalized autocorrelation



(c) Saturday, 04/30/2011 (d) Union-Graph \overline{G}_{\cup} Figure 2. (a-c) G_n : Daily Geospatial graphs of the US airport network; (d) \overline{G} : Expected value graph



Figure 3. Degree distribution for daily graphs

function $R_{\hat{x},\hat{x}}(n)$, which is defined as:

$$R_{\hat{\boldsymbol{x}}_{j},\hat{\boldsymbol{x}}_{j}}(n) = \frac{1}{N-1} \sum_{m=0}^{N-1} \hat{\boldsymbol{x}}_{j,m}^{T} \hat{\boldsymbol{x}}_{j,m+n}$$
(6)

where $\hat{\pmb{x}}_{j,n}$ is the eigenvector corresponding to the eigen-



Figure 4. Clustering coefficient for daily graph

value $\hat{\lambda}_{j,n}$, for the graph \hat{G}_n . The normalized autocorrelation function of the leading daily eigenvectors is given in Figure 12. The normalized autocorrelation function for the leading eigenvectors shows a temporal character that is quasi-stationary. The apparent decay in the autocorrelation



Figure 5. Clustering coefficient for the intersection graph



Figure 6. Variation in the vertex strength with degree

function with lags is due to finite sample size effects. The quasi-stationary behavior of the autocorrelation function suggests that it may be possible to model the leading order dynamics of the airport graphs by employing the spectral decomposition of the adjacency matrix.

$$\hat{A}_{n} = \sum_{j=0}^{|\hat{V}_{n}|-1} \hat{\lambda}_{j,n} \hat{Q}_{j,n}$$
(7)

The matrix $\hat{Q}_{j,n}$ is the spectral projection matrix corre-



Figure 7. Daily Flight Count: 04/26/2011 to 05/24/2011



Figure 8. Daily Edge Count: 04/26/2011 to 05/24/2011



Figure 9. One Lag Difference Edge Count: 04/26/2011 to 05/19/2011

sponding to the eigenvalue $\hat{\lambda}_{j,n}$. Since $\hat{Q}_{j,n}$ is function of the eigenvectors, we expect to approximate it's behavior as stationary in modeling the leading order behavior of graphs \hat{G}_n . The application of this approximation allows \hat{A}_n to be modeled by the matrix \hat{M}_n

$$\hat{M}_n \approx \sum_{j=0}^{J-1} \hat{\lambda}_{j,n} \hat{Q}_{j,0} \tag{8}$$

where $\hat{Q}_{j,0}$ is the *j*-th spectral projection matrix of \hat{A}_0 ,



Figure 10. Averaged autocorrelation function



Figure 11. Eigenvalue Magnitude Spectra For 26 April to 29 April, 2011



Figure 12. Normalized autocorrelation function of the leading daily eigenvectors

 $\hat{\lambda}_{j,n}$ is the *j*-th eigenvalue of \hat{A}_n , and *J* is the rank of the approximation. The solid line in Figure 13 corresponds to \hat{M}_0 , and shows that much of the information in \hat{A}_0 can be approximated by a low-rank model. The dashed line in



Figure 13. Residual power for successively higher rank approximations

Figure 13 correspond to \hat{M}_2 . Despite using the projection matrix \hat{A}_0 to compute the model \hat{M}_2 , the model still contains much of the information of \hat{A}_2 . A similar result is found for other days' adjacency matrices $\hat{A}_{n\neq 0}$. This shows that the

eigenvectors corresponding to the largest eigenvalues of A_n are fairly static. With a method of modeling $\hat{\lambda}_{j,n}$ we would be able to create a low-rank approximation of \hat{A}_n without having to re-calculate the eigenvectors of \hat{A}_n . We expect this result to also aid in developing low dimensional predictive models of the daily national airspace wide traffic demand.

VI. CONCLUSIONS AND FUTURE WORK

In our examination of the U.S. airport network we have attempted to account for all flights that are tracked by the FAA's Traffic Flow Management System. We have found that these flights have a significant impact on the topological structure of the airport network. We have also found the temporal dynamics of the US airport network to be more complex than previously reported. The structure of the network, as characterized by its edges, is found to vary remarkably from day to day. The non-stationary components of the network were found to equalize the distribution of clustering coefficients. We have also performed a spectral analysis of the U.S. airport network. The eigenvalue spectrum of the network shows a power law decay. The structure of the eigenvalue spectrum is similar, even though the eigenvalues themselves exhibit oscillatory behavior over successive days. The leading order behavior of the eigenvectors is found to be quasi-stationary. The observation has been used to construct low rank models of the the airport network. This approach can be useful in developing national airspace wide predictive models for traffic demand.

VII. ACKNOWLEDGMENTS

The authors extend their gratitude for the support provided by the MIT Lincoln Laboratory Technology Office.

REFERENCES

- [1] Z. Xu and R. Harriss, "Exploring the structure of the U.S. intercity passenger air transportation network: a weighted complex network approach," *GeoJournal*, vol. 73, 2008.
- [2] J. Zhang, X. Cao, W. Du and K. Cai, "Evolution of Chinese airport network," *Physica A*, vol. 389, 2010.
- [3] G. Bagler, "Analysis of the airport network of India as a complex weighted network," *Physica A*, vol. 387, 2007.
- [4] L. da Rocha, "Structural evolution of the brazilian airport network," *Journal of Statistical Mechanics: Theory and Experiment*, 2009.
- [5] 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*, vol. 102, no. 22, 2005.
- [6] Federal Aviation Administration, "Traffic Flow Management System to Aircraft Situation Display to Industry Interface Control Document for the Traffic Flow Management Modernization Program," 2009.