

AN EXPLORATION OF EVOLUTION PATHS FOR AIR TRANSPORTATION NETWORKS

Jung-Ho Lewe*, Kisun Song*, Dimitri N. Mavris* *Georgia Institute of Technology, Atlanta, GA, U.S.A.

Keywords: Air Transportation History, Evolutionary Algorithm, Connection Enplanement, Hub Airport

Abstract

Embracing the idea that the U.S. air transportation network (ATN) has been shaped by endogenous and exogenous changes over time, an ATN evolution path is devised to collectively represent continuous and consecutive events. The path is made up of discrete points chronologically populated in a two-dimensional evolution space where spatial expansion and volumetric growth are characterized. By applying different assumptions for these two factors, different shapes of evolution path are generated and investigated: linear, convex, and concave. The convex path was determined to fit the reference data the best. A detailed investigation into the convex path was performed as a case study. The result demonstrates how the evolution unfolds its dynamics on 100 airports that handle more than 95% of the nation's demand volume.

1. Introduction

The US Air Transportation Network (ATN) is a very complex combination of heterogeneous systems consisting of airspace and airport environments for operations of aircraft. Various factors such as economic conditions. technologies, and deregulation have transformed the ATN. Despite the ATN's overwhelming complexity and unpredictability, these broad and unique characteristics have naturally been motivating research topics for network scientists and researchers. One of the main areas is understand attempting to the ATN by investigating historical trends and changes. Some investigators analyzed part of structure aspects, using various network-associated metrics to retrieve meaningful knowledge out of the complexity. [1, 2, 3] Others investigated

historical changes associated with the properties of the ATN to understand its behavior. [4, 5, 6] Most of these studies aimed to figure out the fundamental dynamics of the ATN.

For this purpose, an air transportation network evolution model has been developed and elaborated to capture the dynamics of the ATN based on the network evolution. [7, 8, 9] This model employs the evolution concept by defining evolutionary environments and considering the ATN as pseudo ecosystems that evolve under these environments. What the model really focuses on is mimicking the airlines' adaptations to changing environments. Therefore, the performance of the model relies on how the evolutionary environments are being fed.

The historical evolution process has been deployed with the complex variations of demands and airports over time. To implement this process, two key driving forces were identified and abstracted: volumetric progression and spatial expansion. This model starts from defining an evolution path in a two-dimensional space called evolution space with spatial expansion as the horizontal axis and volumetric progression as the vertical one, respectively. An evolution path comprises a series of points that conceptually describes what path the ATN has been through during its evolution over time. Conceivably, all else being equal, different path shapes result in different networks. This thinking led to a series of questions that the authors will attempt to address. To what extent will different paths affect the network outcomes? What is the implication of these differences? If the modeling objective is to replicate the current ATN, which path is better than the others? (through the intermediate snapshots as well as in the final state)

In order to answer these questions, this paper explores various evolution paths and models the ATN by changing the configuration of an evolution path. Since the target network is the current US ATN, the performance of an evolution path is evaluated by the similarity between the outcome network and the reference obtained from historic data.

2. Methodology

2.1 Evolution Path and Environment Setup

An evolution path is made up of evolution points that define how the environment changes in terms of space and volume. Three basic shapes are considered: linear, concave (curved upward) and convex (curved downward) as illustrated in Fig. 1. The linear path is the simplest path with an assumption that the volume and the space are deployed at the same pace. The concave shape is based on the assumption that spatial expansion drives the evolution of the ATN, meaning that a sizable amount of demand volume exists in the early stage of evolution and a rapid spatial expansion in the later stage shapes the final ATN. On the other hand, the convex path is driven by volumetric progression. It is based on the assumption that many pivotal airports exist in the early stage of evolution and a significant demand volume follows later. It is important to distinguish between the demand and the enplanement. The intrinsic demand volume (τ) goes into the network and the 'networked' demand volume becomes the enplanement (E).



Spatial Expansion



2.2 Historical Review of Evolution Path

Since the inception phase of the commercial transport, many airports have been constructed and have started their services. Through the time period between the two World Wars and the political and military tensions, the majority of the airports in the US were constructed despite the limited need for commercial air transportation. Fig. 2(a) represents the debuts of the top US 100 major airports from 1920 to 2016 [10] depicted as 10-year intervals and Fig. 2(b) illustrates the enplanement history from 1950 to the present [11, 12, 13] with a simple extrapolation in the years before 1950 (three dots on the dotted line).

According to Fig. 2(a), most of the airports were constructed prior to 1970. Fig. 2(b) shows that the enplanement growth follows an S-shaped logistics curve. Various endogenous and exogenous factors affected the changes in the



Fig. 2. Historical Trends of the ATN Evolution (10-year intervals)

growth rate such as economy, population, the end of the cold war, the advancement of aircraft technology, airline deregulation, security issues, etc. Matching the timeline, these two trajectories can be combined. As a result, a convex shape is obtained as shown in Fig. 2(c). Therefore, a conjecture can be stated, namely that a convex path as shown in Fig. 1 would be the best fit as it better represents the historical data for the ATN. A corollary to this statement is that a concave path would be the worst performing shape as it is by far the most different from the convex curve in Fig. 2(c). Detailed discussion on results will be addressed in the subsequent section.

2.3 Preprocessing of Reference Data

The y-coordinate of the evolution space represents the fractional increment of demands. The demand (τ) is the source for the ATN to form and evolve. Its reference data is Airline Origin and Destination Survey (DB1B) [14], a 10% sample airline ticket data from reporting carriers, and the enplanement data from T-100D [15] was used for validation. It was not easy to track multiple-destination trips in the database, so single destination round trips were considered for the study. As tourists go back to their origin areas, the demand and enplanement matrices become symmetric. Through the interconnecting preprocess from the DB1B and T-100D, a symmetric demand input data called Symmetric DB1B (SDB1B) is created in a matrix form. An intrinsic element for the demand between airport *i* and airport *j* is represented by τ_{ii} and the corresponding symmetric enplanement reference data is also obtained. We use the SDB1B as the demand input and enplanements extracted from SDB1B were scaled up to the amount of T-100D for validation.

The x-coordinate of the evolution space represents the number of airports. We retrieved the information for spatial expansion from the National Plan of Integrated Airport Systems (NPIAS) [16] from 2011 to 2015, released by the Federal Aviation Administration (FAA). It says that 503 airports offer commercial services. Naturally, both the demand matrix and the enplanement matrix become very sparse in the cells of minor airports (travel demands are close to zero) such that it is only relevant to consider the major airports. We decided to consider the top 100 major airports as they account for almost 95% of total network volume (enplanement) for a year to simulate the strong Hub-and-Spoke (H&S) ATN evolution. Airports in close proximity are combined into one since they collectively serve the base demand of a metropolitan area. For example, most demands associated with the New York metropolitan area are handled by three major airports, JFK, LGA and EWR, even though each resides in a different administrative region. Likewise, several other metropolitan areas with multiple large airports are considered as shown in in Table 1. As a result, the 100 airports in the model actually contain 113 airports.

Metro Area	Airports (Code)	Code
New York City	EWR, JFK, LGA	@NY
Los Angeles	BUR, LAX, ONT, SNA	@LA
San Francisco	OAK, SFO, SJC	@SF
Chicago	MDW, ORD	@CH
Washington	BWI, DCA, IAD	@WA
Dallas	DAL, DFW	@DA
Houston	HOU, IAH	@HU
Miami	FLL, MIA	@MI

Table 1. Metroplex Airports

2.4 Enplanement Breakdown

In order to handle the composition of network volumes (often interchangeably called enplanements), a characterization scheme called the PACE breakdown was developed. In this explanation, enplanements at airport i (E_i) are decomposed into productions (P_i) , attractions (A_i) , and connections (C_i) at airport *i* such that $P_i + A_i + C_i = E_i$. P_i is the sum of passenger boarding departing from airport i as the initial trip origin and A_i indicates the sum of passenger boarding arriving at airport *i* as the final destination. C_i is the sum of enplanements using airport i as the connection airport (neither origin nor destination). Both P_i and A_i are intrinsic in nature such that they are invariant regardless of network structure. (Generated demand, $G_i = P_i + A_i$) On the contrary, C_i and E_i are a variant type. These are

determined by the dynamics of the network. Every single flight is called a segment whereas the entire flights for an origin-destination pair is called a trip, a travel, or a path. A segment and a one-way trip becomes identical if a passenger flies on a nonstop flight. Otherwise, a trip comprises a number of segments.

2.5 Evolutionary Network Design Algorithm

In order to model the evolution of the ATN, we developed an algorithm dubbed <u>Network</u> <u>Evolution algoriThm (NET)</u>. It is an attempt to understand and unravel the fundamental dynamics of how the US ATN has evolved over history. Specifically, NET aims to emulate the whole mechanisms of network evolution in the aftermath of a myriad of adaptive behaviors of airlines under the given network circumstances.

NET is a progressive-iterative algorithm. It takes a network constructed in the previous step, ATN_{n-1} , an evolutionary environment from the evolution path, EP_n , and airlines' strategies (network policy: Pol_n) and then constructs a new ATN at *n*-th evolution step, ATN_n , illustrated in Eq. (1).

$$ATN_n = NET(ATN_{n-1}, EP_n, Pol_n)$$
(1)

The *n*-th evolution point, EP_n , consists of a set of airports, AP_n , and the *n*-th fractional demands, τ_n , as well as technology. AP_n relies on the shape of the evolution path whereas τ_n is obtained by $\tau_n = v_n \times \tau$, where τ is the entire demand matrix and y_n is a demand fraction that eventually defines an evolution path. Pol_n is the network construction policy directing airlines operations to adapt to the market. The airlines in the model have to accommodate all demands by creating new connections or using existing ones with a fleet of fitting aircraft. While evolution is in progress, various adaptive behaviors are manifested including removals, creations, or alterations of network links or segments. The network keeps constantly updated and becomes mature with growing airports and demand reflecting the past conditions while recursively performing the operations described above. Fig. 3 conceptually illustrates how NET works.

In every step of the evolution, the airlines' decisions – whether or not a direct flight between airports i and j should be made, if the two airports are supposed to be connected, which



Fig. 3. Network Evolution algoriThm (NET)

aircraft should be served, and how many times they should fly – are made based on the disutility minimization utility maximization) (i.e., principle. Those decisions are subject to be updated since everything on the next evolution point – demands, airports and technology – keeps changing to some degree. The disutility is essentially the airlines' cost with respect to aircraft operation and thus a preferential attachment to existing hubs is naturally warranted as the airlines may reduce the cost by extensively using hubs. However, passengers travel time and inconvenience increase significantly due to connections, regardless of flight distance and demand. Thus, the disutility term has another component to account for passenger disutility and therefore a trade-off process is included in the airlines decision making.

In reality, an individual airline has its own set of segments to serve but it is very common to see multiple airlines share the same segments with different fleet, operating schedules and prices. For a given origin-destination trip, every airline can offer a different route option which can be accommodated by a single or multiple operators. implies the existence of symbiotic This interactions between the airlines - competition as well as collaboration – and it should be captured by the model. Representing multiple airlines and tracking their performances during the simulation, however, are topics out of the scope of this research due to complexity. Considering this, NET assumes that a unique giant airline distributes each origin-destination demand into different viable path options depending on the fitness of each path option. NET can mimic the various and complex flight paths existing in the ATN by employing this trip distribution concept. The fitness F_i of the *i*-th path option is evaluated for every aircraft fleet calculated using a probabilistic choice model as shown in the following equation.

$$F_i = \frac{e^{-disutility_i}}{\sum_i e^{-disutility_j}}$$
(2)

where *disutility_j* denotes disutility value of *j*-th path option. As a result, the unique airline can distribute demands by $Trip_i = \tau_{OD} \times F_i$ where $Trip_i$

is the distributed enplanements based on F_i of an origin-destination demand.

2.6 Different Evolution Path Shapes

Since an infinite number of evolution paths may exist, a few represented ones should be created by determining exact path shapes, corresponding mathematical equations and the number of evolution points. For the linear path, an ordinary linear function is used, represented by $y_n = x_n/N$ where x_n is the number of airports, y_n is a value of the demand fraction and N is the total number of airports that are accounted for. The shapes of the convex and the concave paths are defined using the two quarter circles $(-\pi/2 < \theta < 0$ and $\pi/2 < \theta < \pi$, respectively). Table 2 summarizes the information of path shapes.

The convex path shape is the most similar to what happened to the ATN among the three paths. (although the demand and the total enplanement are related but not the same.) Each path is represented by 10 discrete points where n = 10, 20, ..., 100.

 Table 2. Notional Evolution Paths

Shape	Path Equation (x: #AP, y: demand fraction)
Linear	$y_n = x_n/N$
Convex	$y_n = -\sqrt{1 - (x_n/N)^2} + 1$
Concave	$y_n = \sqrt{1 - (x_n/N - 1)^2}$

3. Result and Discussion

3.1 Results of Different Path Shapes

By the definition of PACE breakdown, the total connection volume C can be obtained by the following equation.

$$C = \sum_{i} \sum_{j} (E_{ij} - \tau_{ij}) \tag{3}$$

Looking up SDB1B, the total amount of daily τ_{ref} is 1.287M (million) and the total reference enplanements (E_{ref}) is 1.717M. Thus the total reference connection enplanements (C_{ref}) is 0.435M. With these benchmark values, simulation and analysis were carried out for the three paths to address the research questions posed in the introduction. Fig. 4 shows how E and C evolve with regard to each evolution path shape, where the red dots in the figure represent E_{ref} and C_{ref} .

The first impression on Fig. 4 is that NET is very dynamic in a response to different evolution Different paths represent different paths. environmental changes over time. NET takes the environments and lets the airlines accommodate the demand using feasible aircraft fleet and airports, adapting to the present conditions. Therefore, it is demonstrated that NET can be used to conduct various case studies with different evolution scenarios for future and/or for other countries that have totally different developed or developing aviation histories. Moreover, it is clear that the convex evolution path gives the best match, which positively supports the conjecture posed in Section 2.2 and thus answers the last research question.

Second, in comparison to the original shape of each evolution path, a suppressed volume growth is observed for the concave and the linear cases as shown in Fig. 4 (left). The more concave, the more suppressed. In other words, a premature growth in connection enplanements is observed except for the convex case as shown in Fig. 4 (right). For example, when the number of airports reaches 60, the convex shape still has about 1M (million) demand out of 1.287M whereas the concave shape has only 0.16M to address. From this point, the convex shape creates about 0.37M connection enplanements whereas the concave shape makes only 0.06M. In the convex case, the number of airports surpasses a certain threshold and those airports are 'networked' enough as the evolution progresses such that the airlines can consider a good number of feasible hubs and viable indirect flights to handle incoming demands in the subsequent phases. Therefore, in order for C to flourish, an evolution curve should implement the following principle: spatial expansion comes first and volumetric progression follows later. The convex path shape is the one that better follows this principle and this is what has happened in the U.S. air transportation network history. (Fig. 2)

3.2 Evolution on the Convex Path

In order to obtain a deeper understanding of how the network volume evolved, the convex case is further investigated.

Fig. 5 provides geographical portrayals in an early, an intermediate, and the final phases of network showing how the network evolves. The thickness of a segment represents the magnitude of volume. It can be observed that no crosscontinent flights exist in the early phase due to the technology limitations but they begin to appear from the intermediate phase. The emergence of hub airports can be recognized as well. For example, Charlotte airport (CLT) is barely noticeable in the early phase but is becoming a strong hub airport from the intermediate phase.



Fig. 4. Network Evolution Results: Convergence Trend

AN EXPLORATION OF EVOLUTION PATHS FOR AIR TRANSPORTATION NETWORKS



Fig. 5. Network Evolution Snapshots

The result from individual airports is presented in Fig. 6 where the total enplanements (E) and connection volumes (C) of the top 23 major airports are compared to those from SDB1B. The first observation from Fig. 6 is that NET achieved a very satisfactory result at the individual airport level as well. Each airport is supposed to have its own E and C which are driven by its nature. For example, in both simulation results and reference values, New York City (@NY) and Chicago (@CH) show similar amounts of enplanements but their connection enplanements are quite different. Atlanta (ATL) and Charlotte (CLT) stand out of the crowd with high connection volumes while Los Angeles (@LA), San Francisco (@SF), Miami (@MI) have total enplanements with small contributions from connections. All these similarities and disparities are well captured by NET, which proves it handles the network dynamics as well as different characteristics of the airports with accuracy.

In order to closely examine how the evolution

unfolds for individual airports, Fig. 7 visualizes a relative dominance of generated demand (*G*) and connection enplanement (*C*) for some selected major airports on the early, intermediate, and final phases. The abscissa represents G/E and the ordinate corresponds to C/E. Because E=G+C by definition, at each evolution phase, all of the airports are located on a straight dotted line of x+y=1, where x=G/E and y=C/E. Since total volume *E* increases as evolution deploys, this differences are notionally implemented such that the line at the bottom corresponds to the early phase, that in the middle is for the intermediate phase, and the one at the top represents final evolved phase.

The square symbols are marked to show the reference data obtained from SDB1B. According to Fig. 4, the results from NET using convex path shape almost identical to that of SDB1B, which also represented in Fig. 7. As explained, there are three dotted lines represent early, intermediate, and final phase of evolution. If an airport is located close to y-axis, *C* becomes



Fig. 6. Total Enplanements and Connection Enplanements of Top 23 Major Airports

dominant. Alternately, *G* becomes dominant if an airport is close to x-axis. Therefore, Fig. 7 informs us not only on how each airport has gone through the evolution phases but also yields quantitative insights of the PACE composition. In the above airports, CLT and ATL are very strong hub airports in the ATN. One interesting observation from Fig. 7 is that whether an airport would be *G*-dominant or *C*-dominant is most likely determined in the early phase. For example, ATL have been constantly *C*-dominant from the outset. There is an exception. CLT evolves into a hub airport through the evolution. This is due to CLT has much smaller intrinsic demand.

In reality, ATL is the most significant hub airport for the Delta Airlines and CLT is the corresponding one for the US Airways. These airlines account for almost 20% of the total network volume. [17] Naturally, these airlines try to use their existing links to minimize costs and a myriad of these behaviors have been accumulated through evolution.

@LA and @SF airports in contrast are highly dominated by G. They have shown consistent evolution traces as airports working as origins or final destinations. Also, @NY showed almost same evolution trace to that of @LA (For visibility, it is not shown in Fig. 7). Being located in the coastal lines is the main reason.



Fig. 7. Evolution Traces of Some Airports

Additionally, airport capacity might also be another reason.

Dallas airport (@DA) is located on a 45degree line (not shown) which may result from the contributing factors' being balanced. It is a well-established airport in a large metropolitan area like San Francisco and Los Angeles and also it is geographically located in the central region in CONUS so as expected a mixture of these two phenomena occurs.

4. Conclusions

An evolution space was defined by employing the concept of spatial expansion and volumetric progression observed from the real history of the U.S. air transportation network. Under this hypothesis, three research questions were formulated. The network evolution algorithm ran on the three evolution paths of interest such that the research questions were *quantitatively* answered in Section 3.1. It was demonstrated that the convex path shape would yield the best match in reenacting the network evolution.

A few observations were noted from the exploration of the convex path case. First, the spatial expansion exhibited a dominant environmental influence in the early phases of evolution and the volumetric growth followed in the later phase. These opposing traits of nonlinear progression resulted in an explosive growth of the total enplanements in the later stage due to massive increase in connection enplanements. Second, as a general rule, airports that carried large connection volumes were big airports but not vice versa. Rather, connection volumes were not evenly distributed but concentrated to some airports. Those airports with a large connection percentage have noticeable differences compared to other airports in terms of such as geographical location and characteristics of metropolitan area and the airlines' preferences.

For further research, we would like to answer the questions such as "What if the final evolution point is not corresponding to the current state? If so, what is beyond the end of convex path?" Also, different evolution paths can be created and explored by setting up various scenarios. For example, a scenario with volumetric growth only constrained by the status quo network and another scenario with increased number of airports in the nation due to the advancement in aircraft technology would have to undoubtedly follow totally different evolution paths.

It is expected that the network evolution algorithm will enable us to retrieve meaningful insight and knowledge associated with the dynamics of the air transportation network.

References

- Guimera R, Mossa S, Turtschi A, and Amaral L A N. The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, pp. 7794-7799, 2005.
- [2] Cheung D P and Gunes M H. A Complex Network Analysis of the United States Air Transportation. *IEEE/ACM International Conference*, Istanbul, Turkey, 2012.
- [3] Xu Z and Harriss R. Exploring the Structure of the U.S. Intercity Passenger Air Transportation Network: A Weighted Complex Network Approach. *GeoJournal*, Vol. 73, pp. 87-102, 2008.
- [4] Albert R and Barabasi A –L. Topology of Evolving Networks: Local Events and University. *Physical Review Letters*, Vol. 85, No. 24, pp. 5234-5237, 2000.
- [5] Bhadra D and Hogam B. A Preliminary Analysis of the Evolution of US Air Transportation Network. *MITRE Corporation Technical Report*, 2006.
- [6] Bonnefoy P A, Hansman R Jr. Scalability and Evolutionary Dynamics of Air Transportation Networks in the United States. 7th AIAA Aviation Technology, Integration and Operations, Belfast, Northern Ireland, 2007.
- [7] Lewe J -H and Yang E S. Active Control Algorithms of the Air Transportation Network. *Collaborative Systems Engineering and Design Symposium*, Atlanta, GA, April 2008.
- [8] Yang E -S. A Design Methodology for Evolutionary Air Transportation Networks. *Ph.D. dissertation, Georgia Institute of Technology*, Atlanta, GA, August 2009.
- [9] Song K, Lewe J -H and Mavris D N. A Multi-Tier Evolution Model of Air Transportation Networks. Proceedings of the 14th AIAA Aviation Technology, Integration, and Operations Conference, Atlanta, GA, 2014.
- [10] Bureau of Transportation Statistics. Mater Coordinate. URL:http://www.transtats.bts.gov/Fields.asp?Table_ID =288
- [11] Bureau of Transportation Statistics. Historical Air Traffic Statistics, Annual 1954-1980. URL:http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/fi

les/subject_areas/airline_information/air_carrier_traffic _statistics/airtraffic/annual/1954_1980.html

- [12] Bureau of Transportation Statistics. Historical Air Traffic Statistics, Annual 1981-1995.
 URL:http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/fi les/subject_areas/airline_information/air_carrier_traffic _statistics/airtraffic/annual/1981_present.html
- [13] Bureau of Transportation Statistics. U.S. Air Carrier Traffic Statistics, January 1996 – December 2015. URL:http://www.rita.dot.gov/bts/acts/customized/table? adfy=1996&adfm=1&adty=2015&adtm=12&aos=0&ar td=1&arti&arts&asts&astns&astt=3&ascc&ascp=1
- [14] Bureau of Transportation Statistics. Airline Origin and Destination Survey (DB1B).
 URL:http://www.transtats.bts.gov/Fields.asp?Table_ID =289
- [15] Bureau of Transportation Statistics. Air Carriers: T-100 Domestic Segment (All Carriers). URL:http://www.transtats.bts.gov/DL_SelectFields.asp ?Table_ID=311
- [16] Federal Aviation Administration. National Plan of Integrated Airport Systems. URL:http://www.faa.gov/airports/planning_capacity/np ias/reports/historical/media/2011/npias_2011_narrative. pdf
 [17] The Statistics Particle Demonstration Machael Sharporf.
- [17] The Statistics Portal. Domestic Market Share of Leading US Airlines Between February 2015 and January 2016.

URL:http://www.statista.com/statistics/250577/domesti c-market-share-of-leading-us-airlines/

Contact Author Email Address

mailto: j.lewe@gatech.edu

Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.