Paper 152-2013

Using SAS® to Measure Airport Connectivity: An Analysis of Airport Centrality in the US Network with SAS/IML® Studio

Hector Rodriguez-Deniz, University of Las Palmas de Gran Canaria, Canary Islands, Spain
Pere Suau-Sanchez, Cranfield University, Bedfordshire, United Kingdom
Augusto Voltes-Dorta, Universitat de Barcelona, Barcelona, Spain

ABSTRACT

The U.S. Federal Aviation Administration (FAA) estimates that $52.2 billion will be available over the years 2011–2015 to fund airport infrastructure developments. Because one of the main objectives is to reduce congestion and delays, there is a need to acknowledge the importance of connectivity (measured with a centrality indicator) when establishing funding priorities. Currently, the FAA does not do this. In this paper, we exploit the capabilities of SAS/IML® Studio to implement a range of centrality measures, construct a graphical representation of the U.S. air transport network from airline ticketing data, test the algorithms to identify hub airports, and study the evolution of these indicators during the last decades in order to analyze the impact of airline decisions on airport connectivity.

INTRODUCTION

Airport classification and benchmarking is typically used for both policy and management purposes. In a context of centralized capacity development, one of the crucial aspects is the measurement of airport connectivity, especially when capacity expansions aim to reduce congestions and delays within the domestic network (e.g. the Airport Improvement Program run by the US Federal Aviation Administration). From a social perspective it seems reasonable that funding priority should be given to airports playing a central role in the network, not just because they process a significant proportion of nationwide traffic but also because passengers and airlines are connecting through them to other destinations. Hence, there is a potential for optimizing the social benefits from any public investment by introducing connectivity considerations in regulatory airport classifications. However, despite the significance of this issue[1], the existing literature does not provide an established approach to measure airport connectivity, and the choice of an appropriate indicator is still an unresolved question.

With this background, we aim to test the sensibility of several indicators to airline de-hubbing in order to assess their suitability to characterize airport connectivity. To that end, this paper uses all the available data on passenger demand from the US Department of Transportation to perform a time-series analysis of airport hubbing patterns in the US domestic network between 1993 and 2012.The well-known indicator of flow-centrality is adapted from its original social network setting to an air transport context and used to develop a novel measure of each airport's contribution to the network in terms of actual connectivity. The final indicator is directly proportional to the number of transit passengers going through each airport, and inversely proportional to the total number of passengers in those same markets. A survey of high-profile de-hubbing cases that occurred in the US during the last decades is obtained from the previous literature and the individual cases are analyzed. Besides our flow-based centrality, results are presented for other indicators that have been used in the same context such as degree centrality and betweenness centrality. The sensibility of the different indicators is established by comparing the temporal evolution of the connectivity measures immediately before and after the documented de-hubbing process. A suitable indicator should present a significant decrease in the airports' degree of connectivity. From a methodological perspective, results are expected to establish a clear difference between the concepts of airport "hubbing" and "centrality". From a policy perspective, results can be useful to improve airport classification and benchmarking within a centralized capacity management context. Finally, from a managerial perspective, results provide new insights on airport recovery patterns, not only after airline de-hubbing, but also after natural disasters or major industrial actions.

This paper is therefore an extension of the work presented by Rodriguez-Deniz (2012) at the last SAS Global Forum in Orlando, Florida. On that occasion, a market-based variation of the betweenness centrality index, implemented using SAS/IML modules, was presented and tested for the US airports using a single-year sample, as an alternative to the airport classification criteria used by the FAA in their National Plan of Integrated Airport Systems (FAA, 2011). Clearly, the scope of the present work is wider in both methodological and applied terms. Consequently, and given the amount of data involved in the calculations (flight coupon data for nearly two decades), we chose SAS, particularly SAS/IML Studio, as the leading tool for the accomplishment of our goal, once again. SAS/IML Studio

[1] The FAA estimates that $52.2 billion will be available over the period 2011-2015 under the Airport Improvement Program (AIP).

(Wicklin 2010[2]) is a flexible environment in which SAS/IML programmers can develop, run and debug their programs. Programs in SAS/IML Studio are written using IMLPlus, an upgraded version of the SAS/IML language that extends considerably the capabilities of the original SAS/IML procedure. SAS/IML Studio features includes, in addition to the full compatibility with standard SAS/IML programs, calling SAS procedures from within IMLPlus, creation and modification of statistical graphs, calling R functions and modules and interchange data between R and SAS, object-oriented programming and even multitasking, all in an integrated environment. Hence, SAS/IML Studio becomes a versatile and powerful platform for projects that requires data processing, algorithm implementation and test, and graphical presentation of the results for their subsequent analysis, as in the current case.

The rest of the paper is organized as follows. First, we describe the data sources and briefly detail the process for calling SAS procedures and creating graphs in IMLPlus. Second, degree, betweenness and market-based betweenness centrality measures are presented as well as a detailed explanation of the flow-based indicator introduced in this study. Next, we outline the data preprocessing stage and shed light on the general structure of the US airport domestic network prior to the analysis of some major de-hubbing cases from 1993 to 2012 in the results and discussion section. Finally, some concluding remarks to summarize the study, and a number of topics to be explored in future research conclude the paper.

DATABASE DESCRIPTION

Data comes from the Bureau of Transportation Statistics of the Research and Innovative Technology Administration[3] (US Department of Transportation). The Airline Origin and Destination Survey (BTS, 2012) is a sample of airline ticket information from more than 30 US carriers such as Delta, United and Southwest, among others[4]. The survey covers about 10% of tickets (not actual passengers) from reporting carriers. Detailed information of each domestic itinerary (e.g. origin and destination airport, miles flown) is provided in several tables to analyze air traffic patterns, airline dominance or passenger flows. Data is available on a quarterly basis since 1993 and, given that this study is focused on centrality, detailed stopover information between origin and destination is necessary in order to properly represent the network. The DB1BMarket[5] table contains such information, which comes in the form of a string variable that indicates the origin and destination airport for every flight segment within a single itinerary, allowing us to proceed with the calculation of centrality measures in a straightforward way. Since we are to show sharp declines in airport activity over a certain time period, we took the complete data series from 1993 to the present (first quarter 1993 to second quarter 2012[6]) available at the BTS website. Domestic US air passenger traffic has been rising steadily over the past two decades, as is shown in Figure 1, despite the 9/11 and the recent financial crisis. The resulting sample contains about 350 million records representing individual itineraries. In order to process this amount of data, several Macro definitions, DATA steps, PROC SQL and PROC IML were used throughout the data load and preparation process in SAS.

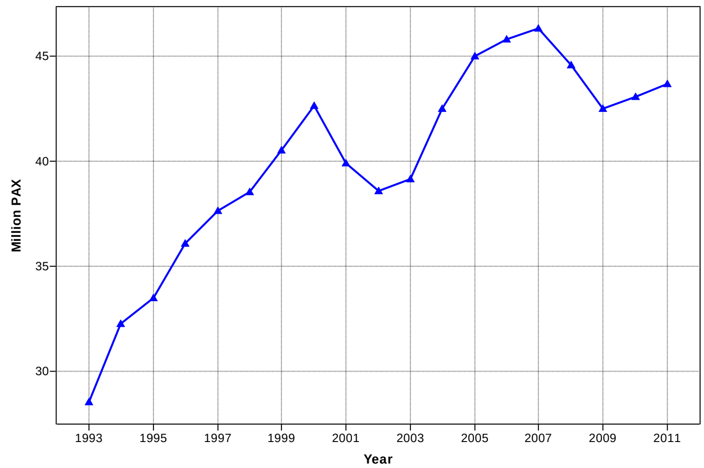

Figure 1 - US Domestic Passenger Enplanements (1993-2011) from DB1BMarket Database

---

[2] This book will be of great value for both novice and advanced SAS/IML users, as well as for those who are looking for a complete reference on the new features of SAS/IML Studio and the IMLPlus language.
[3] http://www.bts.gov
[4] For a comprehensive list of the recent reporting carriers, visit: http://www.transtats.bts.gov/ReleaseInfo.asp?tb=247&display=data
[5] http://www.transtats.bts.gov/TableInfo.asp?Table_ID=247&DB_Short_Name=Origin%20and%20Destination%20 Survey&Info_Only=0
[6] Latest data available at December 2012.

We will briefly detail the steps that are needed to create the line plot depicted in Figure 1 (the IMLPlus code is listed below). One of the key features of SAS/IML Studio is the possibility of running DATA steps or SAS procedures from an IMLPlus program, which is done by enclosing the SAS global statements in a SUBMIT block. A SUBMIT block is the set of SAS statements between a SUBMIT and ENDSUBMIT statements. In the present case, we executed a DATA step to convert the passenger number variable to millions from our IMLPlus program. The next step is to read our data set into memory using the DataObject class. DataObject is the most important class in IMLPlus since it provides fundamental methods[7] for accessing and manage your data. In order to read the dataset containing the passenger data, we declare an object (i.e. variable) of the DataObject class and then create the actual variable through the CreateFromServerDataSet method. Line plots are created in a similar way, declaring the LinePlot object and creating the variable, indicating which attributes are to be plotted, using the Create method. After the generation of the line plot, a number of methods (some inherited from the IMLPlus Plot class) are called to customize the line width and color and to add markers, labels, and so on. Finally, the plot is exported to a graphic file using the SaveToFile method.

```
/*Using IMLPlus to create a line plot to show passenger data*/
libname sasgf 'C:\PAPER_SAS_2013';

/*Data step to show the PAX traffic in millions*/
submit;
data sasgf.paxy;
set sasgf.paxy;
pax = pax / 1000000;
run;
endsubmit;

/*Create data object from SAS data set*/
declare DataObject dobj;
dobj = DataObject.CreateFromServerDataSet("sasgf.paxy");

/*Create and customize a line plot from the data object*/
declare LinePlot line;
line = LinePlot.Create(dobj,'year','pax');
line.SetLineWidth(2);
line.SetLineColor(BLUE);
line.SetLineMarkerShape(MARKER_TRIANGLE);
line.SetMarkerSize(6);
line.ShowPoints('pax');
line.SetAxisLabel(XAXIS,'Year');
line.SetAxisLabel(YAXIS,'Million Pax');
line.SetAxisNumericTicks(XAXIS,1993,2,1993,2011);
line.SetReferenceLineColor(BLACK);
line.ShowReferenceLines();

/*Save the plot to an external file*/
run GetPersonalFilesDirectory(pathname);
pathname = pathname + 'pax_year9311_BMP';
line.SaveToFile(pathname,800,600);
```

[7] See http://support.sas.com/documentation/onlinedoc/imlstudio/WebHelp/imlplus_class_reference/class_hierarchy_overview.htm or Wicklin (2010) for a complete documentation and examples about IMLPlus classes and methods.

DE-HUBBING AND CONNECTIVIY IN THE US DOMESTIC AIRPORT NETWORK

THEORETICAL BACKGROUND

The notion of node centrality is a fundamental question in network analysis, and has attracted the attention of many researchers who have been attempting to systematically identify the most important nodes within a network, over the last half century[8]. Now, in an increasingly competitive, interconnected and globalized world, this question has become paramount in a wide range of subjects from social networks to biology. In an air transport context, centrality appears as a tool to quantify the contribution of each airport to the network in terms of actual connectivity, and its use is widespread in the airline/airport networks literature. In this section, we present the methodological framework of the paper: some basic notions of graph theory and the measures employed in our study, i.e. degree and betweenness centrality (Nieminem, 1974; Freeman, 1977), a betweenness centrality index based on air traffic markets (Rodriguez-Deniz, 2012) and, finally, an adaptation of the flow centrality measure from Freeman et al. (1991). The implications of these indicators for the US domestic network will be analyzed in the next section.

In most cases, and regardless of the data's source and properties, a graph representation of the air transport network, comprising airports and relationships between airports (i.e. vertices and edges), has to be generated prior to the actual analysis[9]. We represent our transportation network as a graph $G = (V, E)$, where V is the set of vertices or nodes, and E the set of edges, which represent connections between nodes. The number of vertices and edges in the network are given by $n = |V|$ and $m = |E|$, respectively. For the sake of simplicity, undirected connections are assumed, and this clearly applies to our problem since we are looking for potential hubs regardless the direction of individual flights. We denote w as the weight function of a network. In a weighted network, we assume that $w > 0$ for all $e \in E$, while $w = 1,\ e \in E,$ in the case of unweighted networks. Consequently, the length of a path between any two vertices of the network will be either the sum of the weights of its edges, or the number of steps in an unweighted network. The shortest distance $d_G(s,t)$ between two nodes s and t is the minimum length of any path in G that connects s and t. In an air transport scenario, the degree centrality represents the number of connections that an airport has, and has become a standard approach for measuring the connectivity potential of every node in the network, being strongly correlated to the airport passenger throughput. Degree centrality[10] can be formalized for an airport i as:

$$C_D(i) = \sum_j \frac{a_{ij} + a_{ji}}{2}$$

Where a_{ij} is the adjacency matrix, in which $a_{ij} = 1$ if the airport i is connected to airport j, and 0 otherwise. Betweenness centrality[11] quantifies the prominence of an actor in terms of connectivity within a network by computing how frequently a node lies on the shortest path between any other two nodes. The betweenness centrality measure is given by:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the number of minimum length paths connecting nodes $s \in V$ and $t \in V$, and $\sigma_{st}(v)$ is the number of such paths in which some $v \in V$ lies on. It is also clear that $\sigma_{st} = \sigma_{ts}$, and $\sigma_{ss} = 1$ by convention. Airports with high levels of betweenness are strategically placed close to the major airline markets and therefore they will be in a privileged, central position in comparison with the rest of their peers. From an air transport perspective, however, the betweenness centrality presents some serious drawbacks due to its strong topological motivation, e.g. well-established markets like San Francisco to Chicago and routes on which scheduled traffic is insignificant would contribute equally to the centrality index if they both lie in shortest paths between a certain origin and destination. Consequently, a well positioned but irrelevant airport in terms of passenger traffic could be highly ranked even though it lies on paths which do not represent any real market[12]. In order to overcome these limitations, Rodriguez-Deniz (2012) introduced a market-based betweenness centrality to identify key airports in an air transport network according to both their topological position (i.e. connectivity potential) and the relevance of the markets they serve in terms of traffic density, defined as:

---

[8] See e.g. Freeman (1978) for further details on the early development of the field.

[9] The graph itself is normally constructed on an ad-hoc basis, depending on which features of the network are to be emphasized. Clearly, different graph designs will result in different network representations which will affect methodological decisions and result interpretation.

[10] We calculate degree centrality for an unweighted and undirected graph representation, i.e. using a symmetric adjacency matrix.

[11] Ellis (2009) presented a PROC IML module for the calculation of an unweighted betweenness centrality, given an input adjacency matrix, using the Brandes (2001) algorithm.

[12] In this case, we define a market as a specific origin-destination journey, thus aggregating all possible routings.

$$C_{B_{mkt}}(v) = \sum_{s \neq v \neq t \in V} \frac{Q_{st}}{Q} \cdot \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where (Q_{st}) is the total number of passengers that travelled on market $s, t \in V$, and (Q) the total number of passengers in the sample. As a result, top ranked airports are likely to play an important role within the US network by combining a central location with relevant market service. Airports lacking of either characteristic will be probably mid-ranked. Airports with similar traffic levels will be classified according to their centrality.

The fourth connectivity indicator that will be included in this paper is an adaptation of the well-known flow centrality measure from Freeman et al. (1991). It was developed in a social network context and aims to quantify the proportion of the maximum directed flow of information (*m*) between two nodes (*j*,*k*) that passes through (i.e. depends on) an intermediate node (x_i). This maximum flow will depend on the capacity of the links in the network and is calculated for each pair of nodes by applying some simple rules, such as that incoming flow must equal outgoing flow for all nodes involved in the transmission of information. By aggregating all possible pairs of nodes (*j*,*k*), the measurement of flow centrality for node x_i ($i.e. C'_F(x_i)$) is easily calculated as the total directed flow that passes through x_i divided by the total flow between all pairs of nodes where x_i is neither a source of information nor its final destination. Thus, the flow centrality (valued between 0 and 1) measures the proportion of the total network flow that travels through x_i, *i.e.*

$$C'_F(x_i) = \frac{\sum_{j<k}^{n} \sum_{k}^{n} m_{jk}(x_i)}{\sum_{j<k}^{n} \sum_{k}^{n} m_{jk}}$$

Adapting this indicator to an air transport context is straightforward. The airports in the US domestic network will be defined as the nodes. The links that connect the nodes are the individual flight sectors operated by the airlines. Passenger traffic is the flow that travels through the network between a point of origin (*j*) and a final destination (*k*) using a variety of routes (either non-stop or multi-stop). Note the market-based definition of passenger flow. The capacity of the links will be defined by the total number of passengers from all different origin/destination markets that share the same individual sector. Since the available data provides information on origin, destination, and intermediate airports (if applicable) at a passenger level, it is possible to obtain both flow and capacity matrices. By incorporating all these definitions into the C'_F formula above, the degree of flow centrality for airport x_i collapses into a simple ratio between the total number of passengers that connect through x_i divided by the total network passengers that travel in all markets that do not start or terminate at x_i. This simple ratio becomes our flow-based measure of airport connectivity.

No reference values for what constitutes low-, medium- and high-level centrality can be defined, as the values depend on the size of the network and the number of airports. However, airports which, due to their privileged location and significant link capacity, are able to channel higher amounts of passenger traffic from other markets, should be expected to present higher levels of centrality that peripheral or small-capacity airports. In that context, results are typically used to compare different airports from the same network. Thus, normalizing the airport-specific estimates (e.g. with respect to the highest-valued hub) is a common practice, which can be straightforwardly performed in SAS.

DATA PREPARATION[13]

Data from the BTS Airline Origin and Destination Survey needs some preprocessing prior to the calculation of the measures detailed in the previous section. Given that DB1Market table provides demand data (the actual number of passengers that travel from/to/through a certain airport, i.e. the actual flow), the flow-based centrality can be calculated in a very simple manner using standard IML statements[14]. For the degree, betweenness and market-based betweenness we need to obtain both weighted and unweighted adjacency matrices. The adjacency matrix is a zero-one square matrix of order $n = |V|$ which indicates which vertices are connected. The weight matrix has a similar structure but having positive values representing the number of total transit passengers. The data set resulting from merging the 1993-2012 data series contains nearly 350 million records and 39 attributes (the COMPRESS Data step option is useful in order to reduce the size of the merged dataset). However, we just need four variables as the starting point to construct the input data for our air transport analysis: Origin, Destination, AirportGroup and Passengers. The AirportGroup attribute is a record of each point (i.e. airport) in the itinerary in which a passenger has stopped, and fits our purposes perfectly. Nonetheless, it comes in the form of a colon-separated concatenation of IATA codes (see Figure 2) which needs to be partitioned into individual origin-destination segments. DATA steps and IML code were employed for the most part of this process.

VIEWTABLE: Sasgf.Market20113

	DestStateFips	DestState	DestStateName	DestWac	AirportGroup
1541936	06	CA	California	91	JFK:SFO
1541937	36	NY	New York	22	SFO:ATL:JFK
1541938	06	CA	California	91	JFK:SFO
1541939	36	NY	New York	22	SFO:ATL:JFK
1541940	06	CA	California	91	JFK:SFO
1541941	36	NY	New York	22	SFO:ATL:JFK

Figure 2. Sample data from the 2011's DB1BMarket table displaying the AirportGroup attribute

Some interesting findings can be revealed if we take a closer look at the resulting networks, being the presence of the small-world phenomena the most noticeable. Small-world networks (Watts and Strogatz, 1998) are uniformly distributed networks where nodes have about the same number of connections, and are ubiquitous in many real-world applications like e.g. social networks and road maps. The average shortest path length measures the efficiency of the routes that connect any two nodes in the network. In this case, an extremely low average path length (2.03) and network diameter[15] (4) are found. Most nodes in the network share a common hub, while any node can be accessed in four or less steps. Also, the value of the network average clustering coefficient (C=0.69 - 0.15 for a random network with similar density, see Watts and Strogatz, 1998) indicates that the network is highly clustered. Furthermore, the degree distribution (Figure 3[16]) does not seem to follow a power-law (actually, the figure reveals an overabundance of high-connected components), which is one of the representative features of the so-called scale-free[17] networks. As a result of this evidence, we claim that the US domestic air transport network is small-world. Similar conclusions are obtained by Guimerà et al. (2005) and Kaluza et al. (2010) for the world wide air transportation network and the global cargo shipping network, respectively.

The IMLPlus code that creates the right plot of the Figure 3 (logged values of degree and number of airports) is listed in the next page. A matrix representing the 2011 US domestic air transport network passenger traffic, with vertices and edges representing airports and connectivity between airports, respectively, is transformed to determine the number of connections (degree) that each airport has. Then the degree vector is saved as a SAS dataset (note that an error will be raised at run-time if you don't use the CLOSE statement in this step) prior to the calculation of the frequency of each number of connections using the FREQ procedure within a SUBMIT block. After the computation of the decimal logarithms for both frequency and degree, we display the values of these variables using a scatter plot in IMLPlus, in an analogous way to that for the LinePlot class. As in the LinePlot case, a DataObject variable must be instantiated in order to access the dataset resulting from the previous PROC FREC. Finally, we call the CreatePolyCurve method to compute a polynomial least-squares estimator to the data in the scatter plot, and show the regression line.

---

[13] Refaat (2007) is an excellent reference for data preprocessing in SAS, as well as Delwiche and Slaughter (2008).

[14] If detailed traffic information is not available, network flow must be estimated (Freeman et al. 1991). The maximum flow between two nodes can be estimated in SAS using Linear Programming via the OPTMODEL procedure. The use of ad-hoc algorithms like Ford-Fulkerson (Ford and Fulkerson, 1956) using IML modules would be a feasible alternative.

[15] The diameter is the longest distance between any two nodes in the network (i.e. how far apart are the two most distant nodes).

[16] The 2011 data has been yearly aggregated to avoid seasonal effects.

[17] Scale-free networks' most notable characteristic is the presence of hubs, i.e. a few nodes that exhibit a great connectivity potential compared with the rest of the elements of the network.

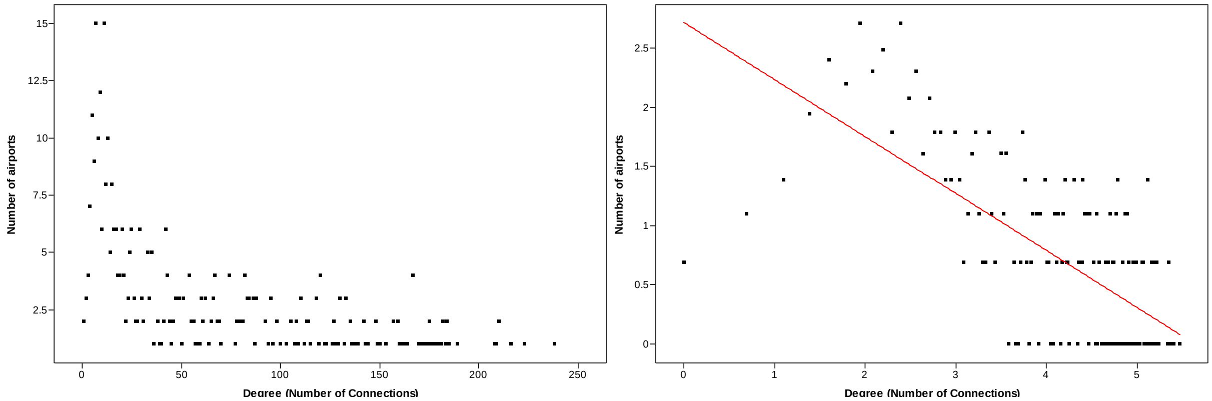

Figure 3. Degree versus cumulated number of airports for 2011. Linear (left) and logged (right) values

```
/*We represent airport connections using scatter plots in IMLPlus*/
libname sasgf 'C:\PAPER_SAS_2013';

/*Read a (weighted) adjacency matrix*/
use sasgf.M2011;
read all var _num_ into M;
close sasgf.M2011;

/*Convert weights to 1s and 0s, and sum by columns*/
M=J(nrow(M),ncol(M),1)#(M>0);
Degree = M[,+];

/*We need the frequency of each number of connections*/
create sasgf.degree from Degree;
append from Degree; close sasgf.degree;

submit;
proc freq data=sasgf.degree;
tables col1 / out=sasgf.frecDegree;
quit;

data sasgf.lfrecDegree;
set sasgf.frecDegree;
ldegree=log(col1);
lcount=log(count);
run;
endsubmit;

/*Scatterplot for the logged values*/
declare DataObject dobj;
dobj = DataObject.CreateFromServerDataSet('sasgf.lfrecDegree');

declare ScatterPlot p;
p = ScatterPlot.Create(dobj,'ldegree','lcount');
p.SetMarkerSize(5);
p.SetAxisLabel(XAXIS,'Degree (Number of Connections)');
p.SetAxisLabel(YAXIS,'Number of Airports');
p.CreatePolyCurve(1);

/*Exporting the plot to a file*/
run GetPersonalFilesDirectory( pathname );
pathname = pathname + 'smallWorld2011_log.emf';
p.SaveToFile( pathname, 800, 600 );
```

RESULTS AND DISCUSSION

The suitability of the different indicators to measure airport connectivity will be proxied by their sensibility to airline de-hubbing. Table 1 shows the percentage loss of centrality for a selection of US airports that have suffered de-hubbing in the last decades, as measured by four different indicators. The list of affected airports and the duration of the de-hubbing process was obtained from Redondi et al. (2012). Time-series data was adjusted for seasonality in the calculations. Results vary widely across the four indicators, illustrating the various ways in which centrality is measured and the impact of these conceptual differences on their characterization of airport connectivity. Unsurprisingly, degree centrality (C_D), which depends solely on the airport's number of connections without taking into account route density, is the indicator that shows the least variability. This is explained by the practice of de-hubbed carriers and alliances to keep a minimum service in order to prevent re-hubbing by rival alliances (Redondi et al., 2012). Weighted and unweighted betweeness centrality ($C_{B_{mkt}}$ and C_B, respectively) are also highly dependent on the aiports' geographical location and route structure, but in this case, results are much more erratic and unpredictable. While airports such as Cincinnati and Washington Reagan show the expected negative signs linked to the closure of direct air routes, it is unclear why the likes of Pittsburgh, Colorado Springs or Nashville experienced a significant increase in betweeness centrality during their de-hubbing period. Further investigation is required. As expected, flow-based centrality (C'_F) is the only indicator that presents the expected negative signs in all cases.

Table 1. Percentage loss of centrality for a selection of de-hubbing cases

Start	End	Airport	CODE	Carrier	Main cause	C_D	C_B	$C_{B_{mkt}}$	C'_F
05-4	10-4	Cincinnati	CVG	Delta-Northwest	Merge	-16.55%	-37.49%	-32.82%	-82.42%
05-3	06-3	New Orleans	MSY	-	Hurricane Katrina	-2.62%	-5.95%	-0.80%	-52.31%
01-3	05-3	Pittsburgh	PIT	US Airways	Network Restructuring	0.94%	17.10%	4.21%	-84.41%
03-3	04-3	Saint Louis	STL	American-TWA	Merge	-0.52%	4.73%	4.79%	-76.52%
01-2	02-2	Raleigh-Durham	RDU	Midway	Bankruptcy	-0.91%	1.32%	2.35%	-75.60%
01-3	02-3	Reagan	DCA	US Airways	9/11 Security Restrictions	-7.24%	-18.29%	-0.65%	-38.14%
97-1	98-1	Colorado Springs	COS	Western Pacific	Network Restructuring	-8.82%	29.46%	9.24%	-87.01%
95-1	96-1	Nashville	BNA	American	Network Restructuring	2.72%	42.76%	1.02%	-73.11%

Source: Own elaboration.

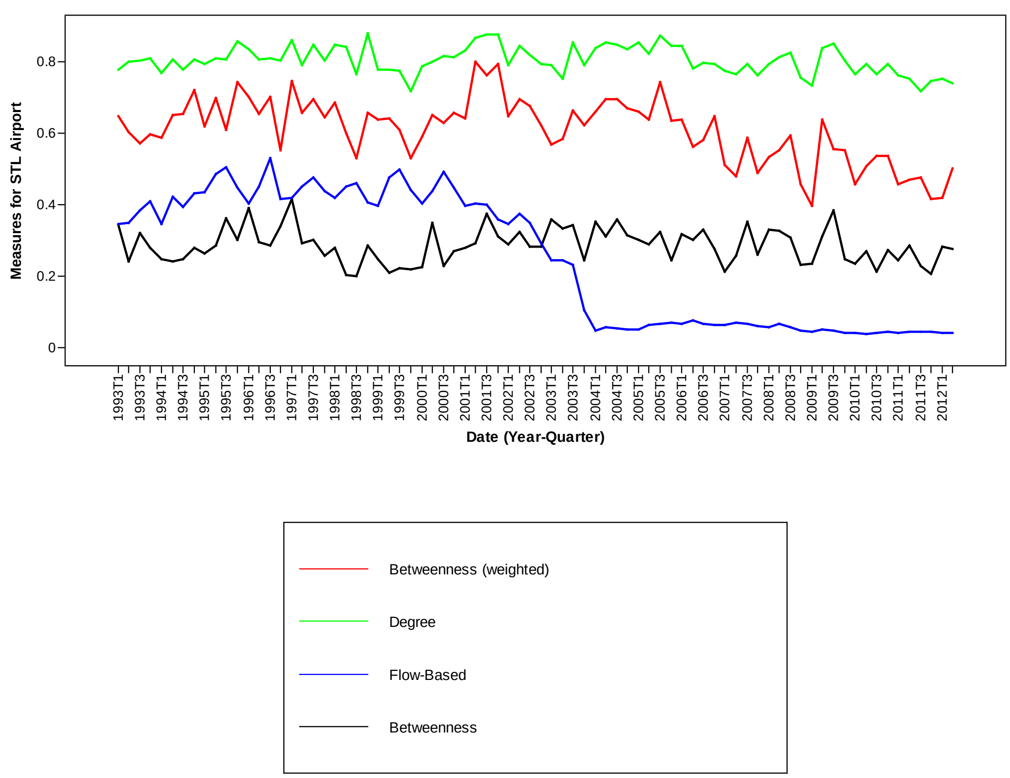

Figure 4. Evolution of centrality measures at St Louis International Airport (STL) 1993-2012

The lack of sensibility of the degree and betweenness centrality indicators to airline de-hubbing is clearly illustrated by Figure 4, which shows the results for STL Airport between 1993 and 2012. Clearly, Figure 4 shows the differences between “topological” centrality and “hubbing” centrality. Both degree and unweighted betweenness centrality do not appear to change significantly across the whole sample period. This is consistent with the topological nature of both indicators, i.e. they are heavily dependent on the airport’s fixed location and route structure. Only when market-based weights are applied to the topological indicators it is possible to see a slight long-term decrease in centrality, which, however, is not even located around the period where the de-hubbing took place. The direct relationship between connecting passengers and flow-based “hubbing” centrality leads to a very sensitive measure of airport connectivity, as the strong impact of the dismantled routes significantly reduced the importance of STL as passenger hub within the US domestic network in less than four years.

Once the appropriateness of flow-based centrality to characterize airport connectivity has been discussed, we will now focus on its applications for airport benchmarking and, particularly, on its suitability to classify airports according to their connectivity in a context of centralized capacity management. A good case study is the National Plan of Integrated Airport Systems (NPIAS) that is used by the FAA in administering the Airport Improvement Program. In the NPIAS, investment requirements and funding priorities are set according to a decades-old airport typology based on the proportions over total US passenger enplanements. Large hubs are those airports that each account for at least 1% of total US passenger enplanements. Medium and Small hubs are defined as airports that each account for between 0.25-1 and 0.05-0.25 %, respectively (FAA, 2011).While the merit (and simplicity) of such an approach are not questioned, the existing literature widely agrees that the importance of a single airport within a network needs to take into account its hubbing potential (i.e. connectivity), which the FAA currently does not.

We propose to use our measure of flow-based centrality (proportion of connecting passengers to total US passengers excluding the base airport) as alternative classification criterion. Note the simplicity and similarity with the FAA method and the availability of data to make the calculations for US airports. The only requirement for the regulator is to set the thresholds that define the airport categories, as in the case above. The application of data clustering techniques for that end is left for future research. However, Figure 5 clearly shows that airport classification, at least for the most relevant airports, could be done at-a-glance. The line chart shows the evolution of flow-based centrality at selected airports between 1993 and 2012. Note the dynamic nature of an airport classification based on this measure: at any given time, between 2 and 3 clusters can be identified. In 1993, Atlanta, Dallas Fort-Worth, and Chicago O’Hare could all be classified as first-class hubs (above a tentative threshold of 4% centrality). After the 1996 Olympics, however, the importance of Atlanta has steadily increased to a point in which the airport is now established as the only top-tier one (above 5% centrality), significantly above the other two “second-class” hubs. With regard to the smaller airports in Figure 5 (St Louis, Cincinnati, and Pittsburgh), it is worth noting the similar evolution in flow-centrality driven by successive de-hubbing events. IMLPlus code for the generation of the Figure 5 is presented below.

```
/*Line plot with a classification variable in IMLPlus*/
libname sasgf 'C:\PAPER_SAS_2013';

/*We focus on a subset of airports*/
submit;
data panel;
set sasgf.panel_big;
where airport in ('ATL','CVG','DFW','ORD','PIT','STL');
run;
endsubmit;

use panel;
read all var {'DATE','AIRPORT','FCENT'};
close panel;

/*Now we define our LinePlot using airport codes as classification (group) variable*/
declare LinePlot line;
line = LinePlot.CreateWithGroup('Line',DATE,FCENT,AIRPORT);

line.SetLineWidth(2);
line.ShowObs(false);
line.SetAxisLabel(XAXIS,'Date (Year-Quarter)');
line.SetAxisLabel(YAXIS,'Flow-Based Centrality');
```

```
/*Drawing a legend using the DrawLegend module*/
Labels = {'ATL','CVG','DFW','ORD','PIT','STL'};
LabelSize = 8;
LineColor = BLACK || RED || GREEN || BLUE || ORANGE || PINK;
LineStyle = SOLID;
Symbol = MARKER_CIRCLE;
BGColor = -1;
Location = 'ORC';
run DrawLegend(line,Labels,LabelSize,LineColor,LineStyle,Symbol,BGColor,Location);

/*Adding some text comments to the plot*/
line.DrawSetTextColor(RED);
line.DrawText (38,15,"Comair Pilots\nStrike"J);
line.DrawSetTextColor(PINK);
line.DrawText (55,2,"AAirlines\ndehub"J);
line.DrawSetTextColor(ORANGE);
line.DrawText (50,12,"USAirways\ndehub"J);
line.DrawSetTextColor(RED);
line.DrawText (66,16,"Delta\ndehub"J);

/*Exporting the plot to a file*/
run GetPersonalFilesDirectory( pathname );
pathname = pathname + 'results_9312_big_paper_.emf';
p.SaveToFile( pathname, 1027, 768 );
```

A second advantage from this methodology is linked to the nature of the AIP. From a social perspective it seems reasonable that funding priority should be given to airports playing a central role in the network, not just because they process a significant proportion of nationwide traffic but also because passengers and airlines are connecting through them to other destinations. Hence, there is a potential for optimizing the social benefits from any public investment by introducing connectivity considerations in regulatory airport classifications.

The evolution of the connectivity indicator can also be used to analyze airport recovery after de-hubbing. Looking at Figure 5, the main conclusion is that, for a major commercial airport, full recovery is not common after significant de-hubbing. Figure 6, however, shows a short-term recovery for the double-dip de-hubbing process of Raleigh-Durham Airport. Further research on that case study, and others similar in nature, should determine the conditions that determine airport recovery, including, but not limited to, airport size and airline dominance.

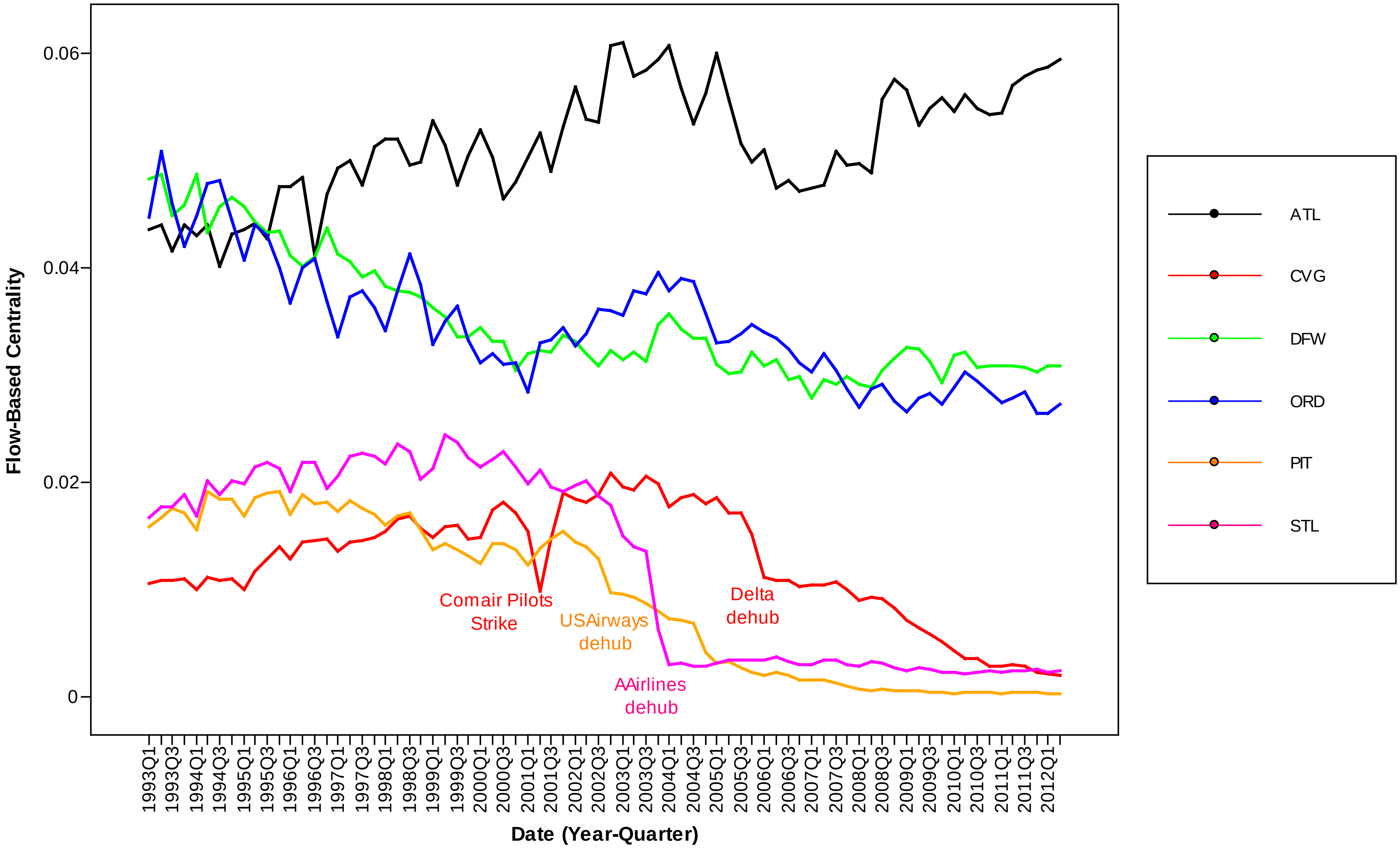

Figure 5. Evolution of flow-based centrality at selected airports, 1993-2012

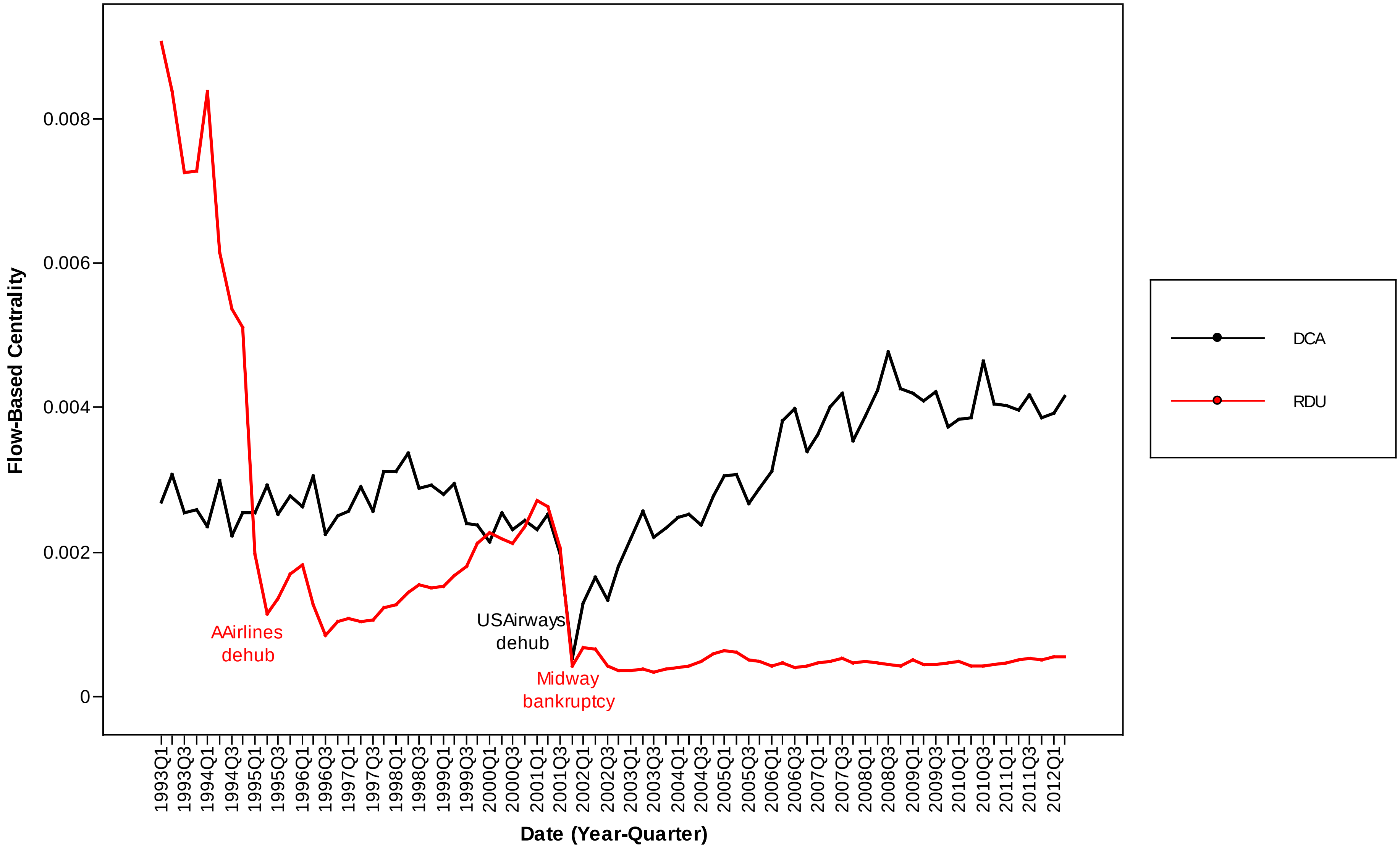

Figure 6. Evolution of flow-based centrality at small hubs 1993-2012

SUMMARY

Airport benchmarking is typically used for both policy and management purposes. In a context of centralized network management and airport capacity development, the choice of an appropriate indicator of airport connectivity is still an unresolved issue. This is mainly due to the lack of definite criteria to determine the suitability of the several indicators available in the literature. In that regard, we propose to use the sensibility to airline de-hubbing as a proxy. To that end, this paper uses quarterly data on passenger demand to perform a time-series analysis of airport hubbing patterns in the US domestic network between 1993 and 2012. The well-known indicator of flow-centrality is adapted to an air transport context and used to develop a novel measure of each airport's contribution to the network in terms of actual connectivity. From a methodological perspective, results are expected to establish a clear difference between the concepts of airport "hubbing" and "centrality". From a policy perspective, results can be useful to improve airport classification and benchmarking within a centralized capacity management context.

Several high-profile de-hubbing cases are analyzed and our flow-based measurement is shown to be much more sensitive than other indicators that have been used in the same context such as degree centrality and betweenness centrality. Thus, we conclude that flow-based centrality should be used as a standard to measure airport connectivity. The suitability of this indicator to serve as a criterion for airport classification in the US domestic network is discussed. Note the simplicity and similarity with the current FAA method and the availability of data to make the calculations. The only requirement for the regulator is to set the thresholds that define the airport categories, which can easily be obtained using data clustering techniques. Further research on the evolution of the centrality indicator should aim to identify the conditions that determine airport recovery after experiencing significant de-hubbing.

SAS/IML Studio proved to be a powerful framework in order to accomplish the essential goals of a project that required the integration and preprocessing of massive airline ticketing data, implementation and test of state-of-the-art centrality algorithms, and a variety of graphical presentations of the results for their posterior analysis. The new capabilities that this software provides through the IMLPlus programming language like e.g. calling SAS procedures, in addition to the full compatibility with standard IML statements, results in an integrated application that would make many IML programmers forget the traditional SAS environment, once they have migrated to SAS/IML Studio.

REFERENCES

BTS, 2012. Origin and Destination Survey (DB1B).
[online] http://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=125&Link=0
Brandes, U., 2001. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25 (2), 163-177.
Delwiche, L. D. and Slaughter, S. J., 2008. The Little SAS Book: A Primer, 4th Edition. SAS Institute Inc. Cary, NC.
Ellis, A. R., 2009. Using SAS to calculate betweenness centrality. INSNA Connections 29 (1), 26-32.
FAA, 2011. Federal Aviation Administration National Plan of Integrated Airport System (NPIAS) 2011-2015 Report.
[online] http://www.faa.gov/airports/planning_capacity/npias/reports/
Ford, L. R. and Fulkerson, D. R., 1956. Maximal flow through a network. Canadian Journal of Mathematics 8,399-404
Freeman, L. C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35-41.
Freeman, L. C., 1978. Centrality in social networks: Conceptual clarification. Social Networks 1, 215-239.
Freeman, L., Borgatti, S., and White, R., 1991. Centrality in valued graphs: a measure of betweenness based on network flow. Social Networks 13, 141-154.
Guimerà, R., Mossa, S., Turtschi, A. and Amaral, L. A. N., 2005. 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 102 (22), 7794-7799.
Kaluza, P., Kölzsch, A., Gastner, M. and Blasius, B., 2010. The Complex network of global cargo ship movements. Journal of the Royal Society Interface. [online]. http://171.66.127.193/content/7/48/1093.full.pdf
Nieminem, J., 1974. On centrality in a graph. Scandinavian Journal of Psychology 15, 322-336.
Refaat, M., 2007. Data Preparation for Data Mining Using SAS. Morgan Kaufmann. San Francisco, CA.
Redondi, R., Malighetti, P., and Paleari, S. (2012) De-hubbing of airports and their recovery patterns. Journal of Air Transport Management 18, 1-4.
Rodriguez-Deniz, H., 2012. Using SAS® to Measure Airport Connectivity: An Application of Weighted Betweenness Centrality for the FAA National Plan of Integrated Airport Systems (NPIAS). Proceedings of the SAS Global Forum 2012, Paper 162-2012.
Watts, D. J. and Strogatz, S., 1998. Collective dynamics of 'small-world' networks. Nature 393 (6684), 440–442.
Wicklin, R., 2010. Statistical Programming with SAS/IML® Software. SAS Institute Inc. Cary, NC.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Hector Rodriguez-Deniz
School of Informatics
University of Las Palmas de Gran Canaria
Canary Islands, Spain
E-mail: hrodriguez@becarios.ulpgc.es

Pere Suau-Sánchez
Department of Air Transport
Cranfield University
Cranfield, United Kingdom
E-mail: p.suausanchez@cranfield.ac.uk

Augusto Voltes-Dorta
Departament de Política Econòmica
Universitat de Barcelona
Barcelona, Spain
E-mail: avoltes@becarios.ulpgc.es