Network Similarity Analysis of Air Navigation Route Systems

Xiaoqian Sun[a,*], Sebastian Wandelt[b]

[a] *Institute of Air Transportation Systems, German Aerospace Center, Blohmstrasse 18, 21079 Hamburg, Germany,*
[b] *Department of Computer Science, Humboldt-University Berlin, Unter den Linden 6, 10099 Berlin, Germany*

Abstract

In this research, we propose a new methodology to assess structural similarity of air navigation route systems in 58 countries. We identify functional dependencies among network metrics through regression analysis. We build a graph for the network metrics, with each metric as a node and a link existing if there is a functional dependency between two metrics. We find that the air navigation route systems in France and Italy are most similar. The air navigation route systems in Oceania ensemble common features of most countries and should be used as representatives for analysis of new air traffic management operational procedures.

Keywords: Network similarity, functional dependency, air navigation route system, air traffic management

1. Introduction

Air transportation systems are facing new challenges regarding their competitiveness, performance, and sustainability [European Commission, 2011]. With increased demand of air traffic, the current structure of air transportation systems will reach the limit in the next years [Zanin and Lillo, 2013]. As one major stakeholder in air transportation systems, it is a challenging task for Air Traffic Management (ATM) to regulate the flow of air traffic and the use of airspace in a safe, cost-efficient, environmental-friendly way [Lillo et al., 2011, Jovanovic et al., 2014].

Complex network theory [Barabsi, 2013, Newman, 2010, von Hirschhausen and Cullmann, 2010, Kotegawa et al., 2013] can highly facilitate the understanding of the structures and the dynamics in air transportation systems. Most research focuses on airport (airline flight) networks, where nodes are airports and links exist between two airports if there are flight connections [Guimera et al., 2005, DeLaurentis et al., 2008, Zhang, 2010, Zhang et al., 2014, Cardillo and et. al., 2013]. Several research focused on how delay is propagated in airport networks [Fleurquin et al., 2013a,b, Baumgarten et al., 2014, Zou and Hansen, 2014]. Robustness of airport networks has also been widely studied [Wei et al., 2014, Zhao et al., 2014, Bruni et al., 2014, Lordan et al., 2014, Niu et al., 2014].

Airport network is only one perspective of air transportation systems. Another perspective is to consider how aircraft actually fly. As aircraft fly through the sky, they follow pre-planned routes, much like highways on the ground. In the air navigation route network, nodes are air navigation points and links exist if there are route segments between two air navigation points. Research on air navigation route network using complex network theory has been conducted only recently (see [Zanin and Lillo, 2013] for a review). Cai et. al. investigated the Chinese air navigation route network and compared its topological characteristics with the Chinese airline network [Cai et al., 2012]. Vitali et al. analyzed the Italian air navigation route network [Vitali and et. al., 2012]. Gurtner et. al. applied three community detection algorithms to the European airspace [Gurtner and et. al., 2013]. Wang and Gong proposed a node optimization model for the Chinese air navigation route network [jin Wang and hui Gong, 2014].

These works either consider the air navigation route network for a single country (China or Italy) or for a whole continent (Europe). Especially, their analysis is based on different datasets, thus, these networks are generally not comparable and it is difficult to explain the similarities and differences. In this paper, we use data from a single, consistent database (European Aeronautical Information Service Database, EAD) provided by Eurocontrol. We extract

*Corresponding author: Tel.: +49 40 42878-4508
Email addresses: xiaoqian.sun@dlr.de (Xiaoqian Sun), wandelt@informatik.hu-berlin.de (Sebastian Wandelt)

the data from EAD to build the air navigation route network. Given a consistent view on the worldwide air navigation route network, it is feasible to study the structures of different sub-networks.

The goal of this research is to investigate the structural similarity of the air navigation route systems in different countries. There exist several network metrics, different metrics often yield different similarity results for the country pairs. It is non-trivial to combine these scores into an overall measure of similarity between air navigation route networks. Therefore, it is essential to properly identify the informative and redundant metrics in complex networks [Roy, 2012].

Some research has been conducted in network comparison. Bania et. al. examined the route (airport) networks of thirteen largest US airlines in 1989 and showed that there existed wide diversities [Bania et al., 1998]. Paleari et. al. compared the structure and performance of three airport networks: US, Europe, and China. They found that connections requiring intermediate airports need longer average waiting times in the European network and the US network shows better coordination [Paleari et al., 2010]. Berlingerio et. al. introduced a three-steps approach to compare large networks: feature extraction, feature aggregation, and comparison [Berlingerio et al., 2012]. In the feature extraction step, seven local features were chosen. One of their following-on works was to compare a variety of network similarity methods in an empirical study [Soundarajan et al., 2013]. However, global level network features were not considered. Further, it is still challenging to select the features to be extracted from a graph appropriately. Li et. al. proposed a graph classification approach with features from different global topological attributes and label attributes [Li et al., 2011b]. The authors considered twenty global graph attributes. However, several important graph attributes, such as betweenness centrality and eigenvector centrality, were not considered.

Further, Jamakovic et. al. studied the relationships between topological metrics in real-world networks and showed that some topological metrics tend to be more correlated than the others, with the implication that there is redundancy between topological metrics [Jamakovic et al., 2007]. Li et. al. investigated the linear correlation coefficients between widely used network metrics in three network models and in functional brain networks [Li et al., 2011a]. Bounova and de Weck presented an overview of topology metrics and their relationships for various ensembles of synthetic and real networks [Bounova and de Weck, 2012]. The authors cautioned against studying a single measure or pulling networks from different domains and topologies together. Onnela et. al. introduced a framework for constructing taxonomies of networks based on community structure [Onnela et al., 2012]. Garcia-Robledo et. al. presented an experimental study on the correlation between commonly used network metrics on the topology of the Internet [Garcia-Robledo et al., 2013]. Meng et. al. compared the digital and face-to-face communication networks using four network similarity measures: common links, adjusted common links, Pearson correlation of the networks' degree distribution, and graphlet degree distribution [Meng et al., 2014], among which the graphlet degree distribution has been widely used for the comparison of biological networks [Przulj, 2007].

These previous works all used Pearson correlation coefficients to analyze the relationships between network metrics. However, one implicit assumption when using Pearson correlation coefficients is that two variables are bivariate normally distributed, and the association between two variables is linear [Vogiatzi, 2002]. Our aim is to remove the redundant metrics which move together based on regression analysis. One significant advantage of using regression analysis is that we can detect the non-linear relationships among the network metrics.

In this paper, we propose a new methodology to assess the structural similarity of the air navigation route systems in different countries, as illustrated in Figure 1. At first, we select the air navigation route networks of the top 58 countries regarding the number of nodes in the world. We identify and remove the functional dependencies among the network metrics through regression analysis. We select fourteen functionally independent metrics for the air navigation route networks. We compute the similarity scores for all network pairs among the 58 countries. In the end, we explain the structural similarity of the air navigation route systems.

This paper is organized as follows. Section 2 introduces a set of network metrics from complex network theory. Section 3 presents the database we used in the study. In Section 4, we analyze the structural similarity for the air navigation route systems in 58 countries. Further discussion is presented in Section 5. Finally, conclusions are drawn in Section 6.

2. Background

Topological analysis is usually one of the first steps in understanding the behavior of complex networks [Newman, 2004, Kurant and Thiran, 2006]. We briefly introduce basic concepts from complex network theory, relevant for the

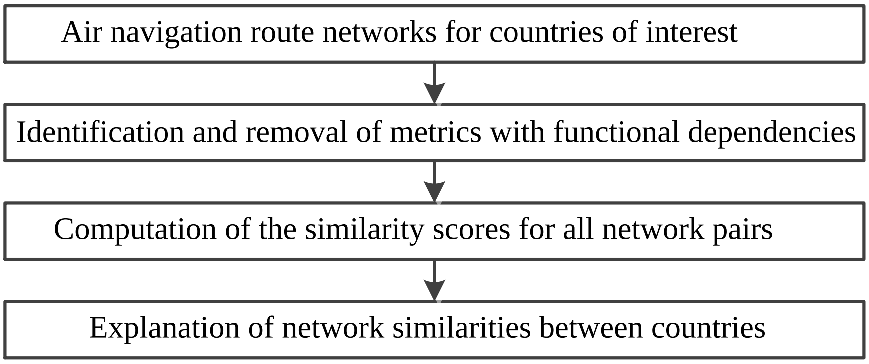


Figure 1: A new methodology to assess the structural similarity of the air navigation route systems in different countries

analysis of air navigation route networks. A network consists of nodes and links. Links can be either directed or undirected: While directed links can only be traversed from their source to the target, undirected links can be accessed in the opposite direction as well. A path is an ordered sequence of nodes and links, connecting a source node and a target node [Diestel, 2010].

In the air navigation route network, nodes are air navigation points and links are route segments connecting air navigation points. In this research, we consider the air navigation route network as an undirected network, with the assumption that the route network is approximately symmetric [Lillo et al., 2011]. We analyze a large set of network metrics: fourteen metrics at global network level and seven metrics at local node level. These metrics are listed as follows.

2.1. Metrics at global network level

1. **Assortativity:** Assortativity measures the similarity of connections in the network with respect to the node degree. The assortativity coefficient is calculated by $r = \frac{\sum_{jk} jk(e_{jk}-q_j q_k)}{\sigma_q^2}$, where e_{jk} is defined as the fraction of links that connect nodes of degree j and k, q_k is the distribution of the remaining degree, σ_q is the standard deviation of the remaining degree [Newman, 2002]. The values of r ranges from -1 to 1. Positive values of r indicate correlations between nodes with similar degree; while negative values indicate correlations between nodes with different degree.
2. **Average shortest path length:** A path is an ordered sequence of nodes and links, connecting a source node and a target node. A shortest path between two nodes is a path such that the sum of the links is minimized. The average shortest path length is calculated by $a = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)}$, where V is the set of nodes in the network, n is the number of nodes, and $d(s,t)$ is the shortest path from s to t.
3. **Cycle basis:** A basis for cycles of a network is a minimal collection of cycles such that any cycle in the network can be written as a sum of cycles in the basis. We count the number of cycles basis in the network.
4. **Degree entropy:** Entropy was originally proposed by Shannon to quantify the uncertainty in strings of text [Shannon, 1948] and is widely used as a diversity index in ecology. It is defined by $H = -\sum_{i=1}^{n} p_i ln p_i$, where p_i is the fraction of the entire population made up of specifies i, n is the number of the entire population. High value of the Entropy H indicates that the node degree is rather diversely and equally distributed in the network.
5. **Density:** The density for undirected network is defined as $d = \frac{2m}{n(n-1)}$, where n is the number of nodes and m is the number of links. This metric measures how close the network is to complete. The density of a complete graph is 1.
6. **Diameter:** The longest graph distance between any two nodes in the network, where connected nodes have graph distance 1.
7. **Link connectivity:** This metric is equal to the minimum number of links that must be removed to disconnect the network or make it trivial [Esfahanian, 2012].
8. **Node connectivity:** This metric is equal to the minimum number of nodes that must be removed to disconnect the network or make it trivial [Esfahanian, 2012].

9. **Number of nodes:** The total number of nodes in the network.
10. **Number of links:** The total number of links in the network.
11. **Percentage of central nodes:** The center is the set of nodes with eccentricity equal to the radius. We compute the ratio of the number of central nodes to the total number of nodes in the network. The eccentricity of a node is the maximum distance from this node to all other nodes in the network.
12. **Percentage of periphery nodes:** The periphery is the set of nodes with eccentricity equal to the diameter. We compute the ratio of the number of periphery nodes to the total number of nodes in the network.
13. **Radius:** The radius is the minimum value of the eccentricity.
14. **Transitivity:** The transitivity is the fraction of all possible triangles appeared in the network. It is defined as $T = \frac{3 * \text{Number of triangles}}{\text{Number of triads}}$, where triads are two links with a shared node.

2.2. Metrics at local node level

15. **Betweenness centrality:** $B_i = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$, where σ_{st} is the number of shortest paths going from node s to node t; $\sigma_{st}(i)$ is the number of shortest paths going from node s to node t and passing through node i [Freeman, 1978]. This metric indicates the number of shortest paths going through a node.
16. **Closeness centrality:** $C_i = \frac{\sum_{j \in N, j \neq i} \sigma_{ij}}{(n-1)}$, where N is the set of all nodes in the network, n is the number of nodes, σ_{ij} is the shortest path between node i and node j. This metric is the average distance from a given starting node to all other nodes in the network [Freeman, 1978].
17. **Clustering coefficient:** $CC_i = \frac{\sum_{j,k} a_{ij} a_{ik} a_{jk}}{k_i(k_i-1)}$, where a_{ik} is the connection between node i and node k, a_{jk} is the connection between node j and node k. This metric gives an overall indication of how nodes are embedded in their neighborhood.
18. **Degree:** $D_i = \sum_j a_{ij}$, where a_{ij} is the connection between node i and node j: $a_{ij} = 1$ when there is a connection existing; $a_{ij} = 0$ otherwise. For all pairs of nodes in the network, a_{ij} composes the adjacency matrix. Degree refers to the number of connections with other nodes.
19. **Eccentricity:** The eccentricity of a node is the maximum distance from this node to all other nodes in the network.
20. **Eigenvector centrality:** This metric measures the influence of a node in the network. It is the eigenvector for the largest eigenvalue of the adjacency matrix. Nodes with high eigenvector centrality also connect to other nodes which have high eigenvector centrality.
21. **Load centrality:** Load is defined as the total amount of the commodity passing through a node during all these exchanges [Brandes, 2008]. Load centrality is the fraction of all shortest paths passing through that node.

In order to characterize the location and the variability of the metrics' values, we aggregate the seven nodal metrics in terms of their first four moments: mean, standard deviation, skewness, and kurtosis, where skewness measures the asymmetry and kurtosis measures whether the values of the metrics are peaked or flat relative to a normal distribution [NIST, 2014]. Positive skewness indicates that most values of the metric are concentrated on the left of the mean; while negative skewness indicates that most values of the metric are concentrated on the right of the mean. A metric with high kurtosis tends to have a sharper peak around the mean and this metric has a high probability of extreme values; while a metric with low kurtosis has a flatter peak and the values of this metric are widely spread around the mean.

3. Database

We obtain the data (effective on 18 October 2012) from the centralized reference European Aeronautical Information Service Database (EAD) provided by Eurocontrol[1]. Eurocontrol is the European organization for the safety of air navigation and it coordinates and plans the air traffic control for the whole Europe. In Eurocontrol, the Aeronautical Information Services (AIS) data is managed and distributed using the Aeronautical Information Exchange Model (AIXM), a semi-structured language designed to enable the management and distribution of AIS data in digital

[1]http://www.ead.eurocontrol.int/eadcms/eadsite/index.php.html

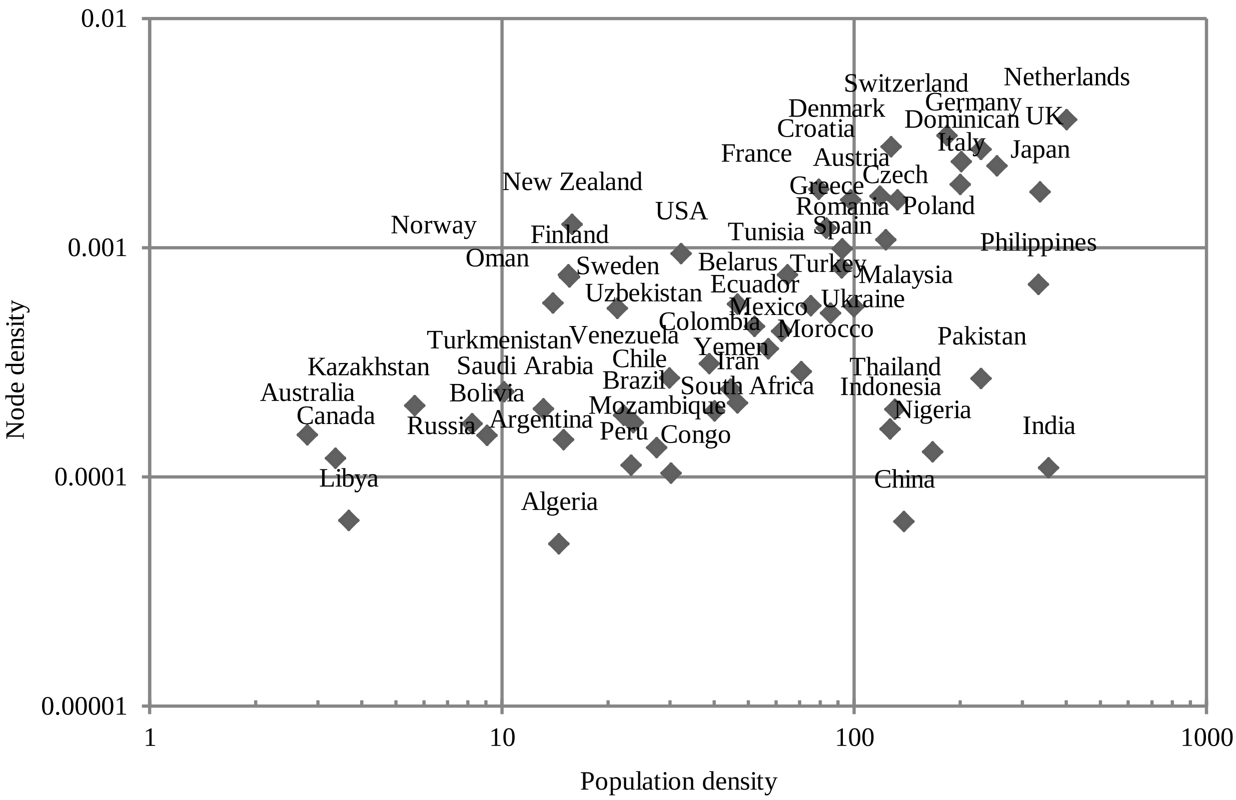


Figure 2: Population density versus air navigation node density in the 58 countries in log-log scale

format. The two components of AIXM are: AIXM Conceptual Model and AIXM XML Schema. The AIXM Conceptual Model describes the features and their properties within the aeronautical domain. The AIXM XML Schema is an exchange model for aeronautical data: It is an implementation of the conceptual model as an XML schema.

In this paper, we build air navigation route sub-networks for different countries based on the EAD. In total, we generate sub-networks for the top 58 countries according to the number of nodes from the EAD. The 58 countries account for 80% of all the nodes (38,473) and 76% of all the links (64,167) in the worldwide air navigation route network extracted from the EAD dataset. For each country, we show four attributes in Table 3 (see the Appendix): geographic area (km^2), population, number of nodes, and number of links. The bold numbers in the table are the maximum and the minimum values for each attribute. USA has the highest number of nodes and links; while Thailand and Croatia have the smallest number of nodes and Mozambique has the smallest number of links.

Note that the North Atlantic Tracks (NAT), are not included in the EAD because they are changing every day, due to the weather conditions (See [Rodionova et al., 2012] for a further study about the NAT). In this paper, we investigate the static structures of the air navigation route system. Since we build the regional air navigation route networks at a country level, the NAT are irrelevant in our study.

In order to investigate whether topological properties of the air navigation route network have connections to social-economic factors, we present the population density versus node density for the 58 countries in Figure 2. We can observe that Netherlands has the highest population density and the highest node density. Further, we find that in industrial countries, such as Germany, Italy, UK, Switzerland, and Japan, the node density scales up with the population density; while emerging countries with relatively higher population density, such as China and India, have rather low node density. With increased economic development and traffic demand, emerging countries might consider to adapt the network structure from industrial countries when they expand their air navigation route systems. It is interesting to investigate network properties and their connections to other social-economic factors in the future [Barthélemy, 2011].

Table 1: Normalization function for the size-dependent metrics

Metrics	Normalization
Average shortest path length	Divided by the longest possible path (n-1)
Betweenness centrality	Divided by the number of node pairs ((n-1)(n-2))/2
Closeness centrality	Multiply the sum of minimum possible distances (n-1)
Cycle basis	Divided by the maximum possible degree (n-1)
Density	Multiply the maximum possible degree (n-1)
Diameter	Divided by the longest possible path (n-1)
Eccentricity	Divided by the longest possible path (n-1)
Load centrality	Divided by the number of node pairs ((n-1)(n-2))/2
Radius	Divided by the longest possible path (n-1)

4. Structural similarity for the air navigation route systems

Our new methodology to assess the structural similarity for the air navigation route systems was illustrated in Figure 1. Each step of the methodology is discussed in this section. Subsection 4.1 gives an overview of the distributions of the network metrics. Subsection 4.2 presents the network evaluation using single metrics. Subsection 4.3 identifies the functional dependencies among the network metrics. We select fourteen metrics with functional independences to compare the structural similarity of the air navigation route systems in Subsection 4.4.

4.1. Metrics overview using kernel density estimation

Kernel density estimation gives us an overview of the distribution for each metric and it is defined as [Parzen, 1962]: $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$, where $(x_1, x_2, ..., x_n)$ are samples drawn from an unknown density distribution f, K is the kernel function, and h is the bandwidth. In this paper, we select the Silverman rule-of-thumb to choose the bandwidth for Gaussian kernel density estimation [Silverman, 1986].

Kernel density estimation is similar to histogram, but it does not have the data binning problem: When constructing a histogram, the interval covered by the data needs to be divided into equal sub-intervals. The size of the bins and the end points of the bins have strong influences on the shape of a histogram.

We present the kernel density estimations for the fourteen global network metrics and seven local nodal metrics with their first four moments (mean, standard deviation, skewness, and kurtosis) in Figure 3. In each sub-figure, x-axis is the value of the network metric, and y-axis is the estimated kernel density. Note that the kernel density estimations for the link connectivity and node connectivity are not shown since their values are all one. We can observe that the distributions of the skew of closeness centrality, the mean, standard deviation, and skew of degree are unimodal and can be approximated by normal distributions. We can obtain the most likely values for these metrics based on the unimodal distributions. The distributions of the other metrics are multi-modal.

4.2. Network evaluation using single metrics

In this subsection, we take two metrics as an example to show that the 58 air navigation route networks are clustered differently for different single metrics. Figure 12 (see the Appendix) presents the hierarchical clustering for the 58 air navigation route networks based on assortativity and density, respectively.

The assortativity of Germany and China is very similar; both countries are clustered together in Figure 12 (a). However, the results for density in Figure 12 (b) suggest that Germany and New Zealand are similar, while Germany is dissimilar to China and only joins with China's cluster in the top-most merge step of the clustering algorithm. Different metrics often yield different similarity scores for the countries. It is non-trivial to combine these scores into an overall measure of similarity between air navigation route networks.

4.3. Identification of functional dependencies among metrics

Recently, Bounova and de Weck found that several topological metrics are size-dependent and the correlations among the metrics would be distorted without normalization [Bounova and de Weck, 2012]. In order to analyze the air navigation route networks with size-independent metrics, we normalize the network metrics by some function of the network size [Bounova and de Weck, 2012]. The normalization functions are listed in Table 1.

After the normalization of the size-dependent metrics, we find that the network size still has influences on the network metrics. Figure 13 (see the Appendix) presents the scatter plots of the metrics after size-dependent normalization. Note that the scatter plots of the link connectivity and node connectivity are not shown since their values are all one. For the purpose of presentation, we only show the mean for the seven nodal metrics.

We can observe from Figure 13 that there is a power law relationship between the mean of closeness centrality and the mean of eccentricity. Furthermore, we also find that there exist functional dependencies among the network metrics. For example, the average shortest path length can be fitted by the diameter with a power law function. The functional dependencies indicate that there are redundancies among the network metrics.

We perform regression analysis in order to detect the functional dependencies among the network metrics. We

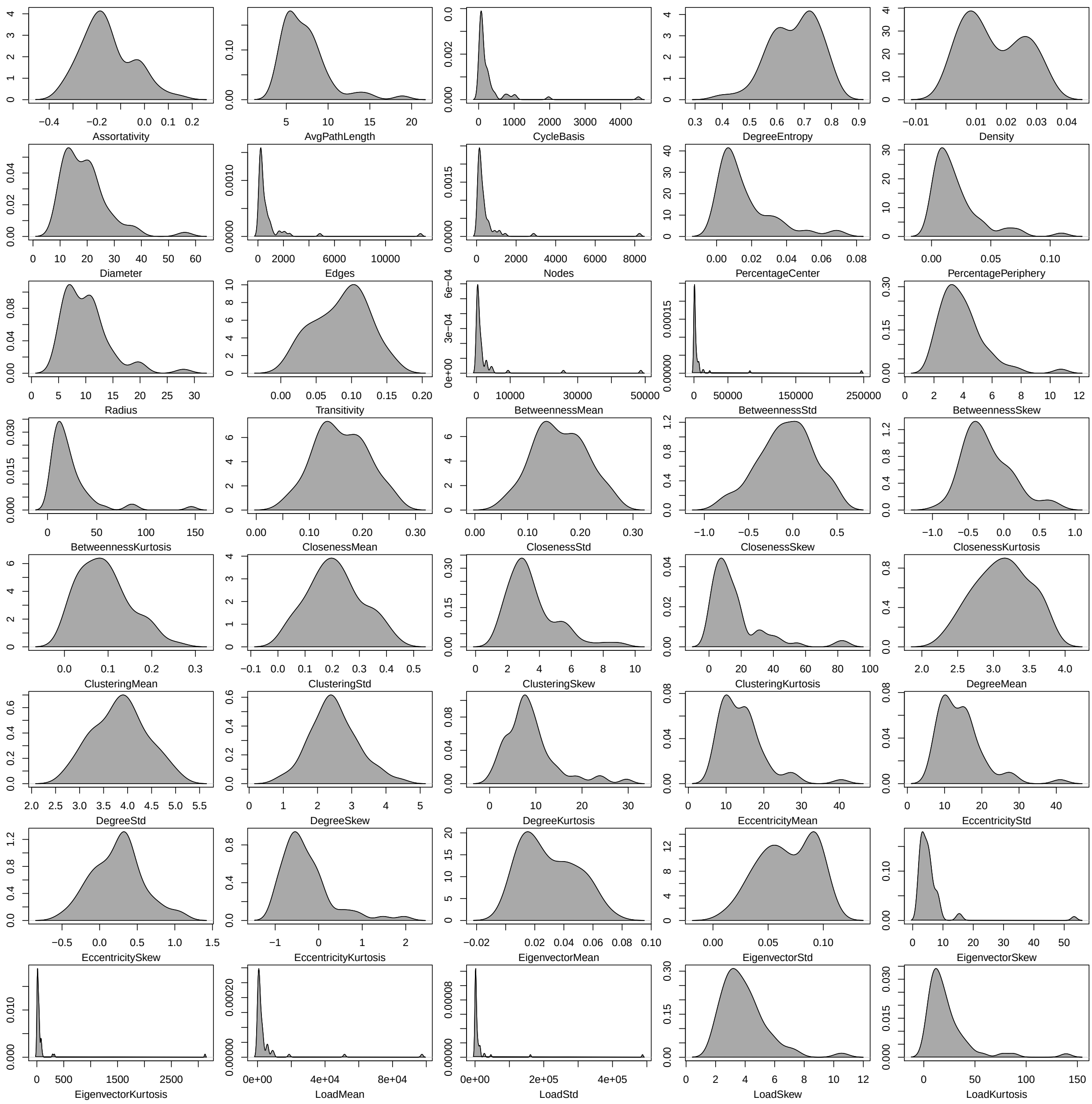


Figure 3: Kernel density estimations for the network metrics

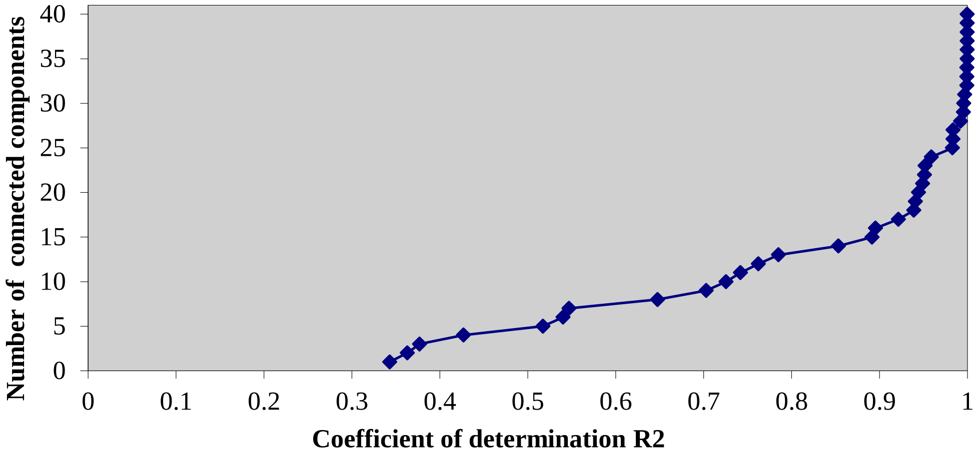


Figure 4: Sensitivity analysis for the number of connected components regarding the threshold values of R^2

use the coefficient of determination R^2 to explain how well the metrics depend on each other. In this study, we consider forty network metrics for the regression analysis[2]. For each regression, we have four function types with one predictor: exponential ($y = ae^{bx}$), linear ($y = ax + b$), logarithmic ($y = alnx + b$), and power law ($y = ax^b$). We select the regression which has the minimum square error.

For each of the forty network metrics, we perform regression analysis between this metric and all other metrics. Thus, the total number of regression analysis is ${}^{40}P_2 = \frac{40!}{(40-2)!} = 1560$. We consider each metric as independent variable and dependent variable separately. The best regression types (minimum square error) between each dependent variable and independent variable are presented in Figure 14 in the Appendix. We find that for each pair of metrics, no matter which metric is considered as independent variable, the R^2 keeps constant in the regression analysis.

Furthermore, we build a graph for the network metrics: each metric is a node and a link exists if there is a functional dependency between two metrics. The weight of a link represents the R^2. Since the R^2 keeps constant for each pair of metrics in the regression analysis, we can treat the functional dependency between two metrics is symmetric. Therefore, the graph for the network metrics is undirected and weighted.

The threshold value of R^2 has a significant influence on the reconstructed network. In the following we explain how we select the threshold value of R^2 and how changing the threshold affects metric similarity.

Figure 4 presents the number of connected components for a variable threshold value of R^2. Thus, with a decreasing threshold value, more metrics are clustered together, and the number of connected components is reduced in the network. We observe from Figure 4 that there is a (first) stationary point around $R^2 = 0.8$. Therefore, for the remaining evaluation, we set the threshold value for the $R^2 = 0.8$, i.e., metrics which have $R^2 > 0.8$ are connected in the graph.

Furthermore, in Figure 15 of the Appendix, we show how the metrics clusters are merged based on their R^2 values. If the threshold is 1.0, two metrics clusters are merged: 1) Degree and DegreeMean and 2) LoadMean and BetweennessMean. Reducing the threshold stepwise, at 0.9996, EigenvectorStd is merged with Number of Nodes. The last two clusters {DegreeKurtosis, DegreeSkew} and all the remaining metrics, are merged, when reaching a threshold of 0.34306.

With the threshold value for the $R^2 = 0.8$, Figure 5 presents the connected components for the metrics, with their functional types, where *exp* stands for exponential, *lin* stands for linear, *log* stands for logarithmic, and *pow* stands for power law. The value in the parentheses next to the function type is the coefficient of determination R^2. For example, in the connected components consisting of the DegreeSkew and the DegreeKurtosis, *pow (0.93)* indicates that there is a power law functional dependency between these two metrics.

[2]In total, we have fourteen global network metrics and seven local nodal metrics with their first four moments (mean, standard deviation, skewness, and kurtosis). The link connectivity and node connectivity are not considered in the regression analysis since their values are all one. Thus, the number of the metrics for the regression analysis is: 14+7*4-2=40.

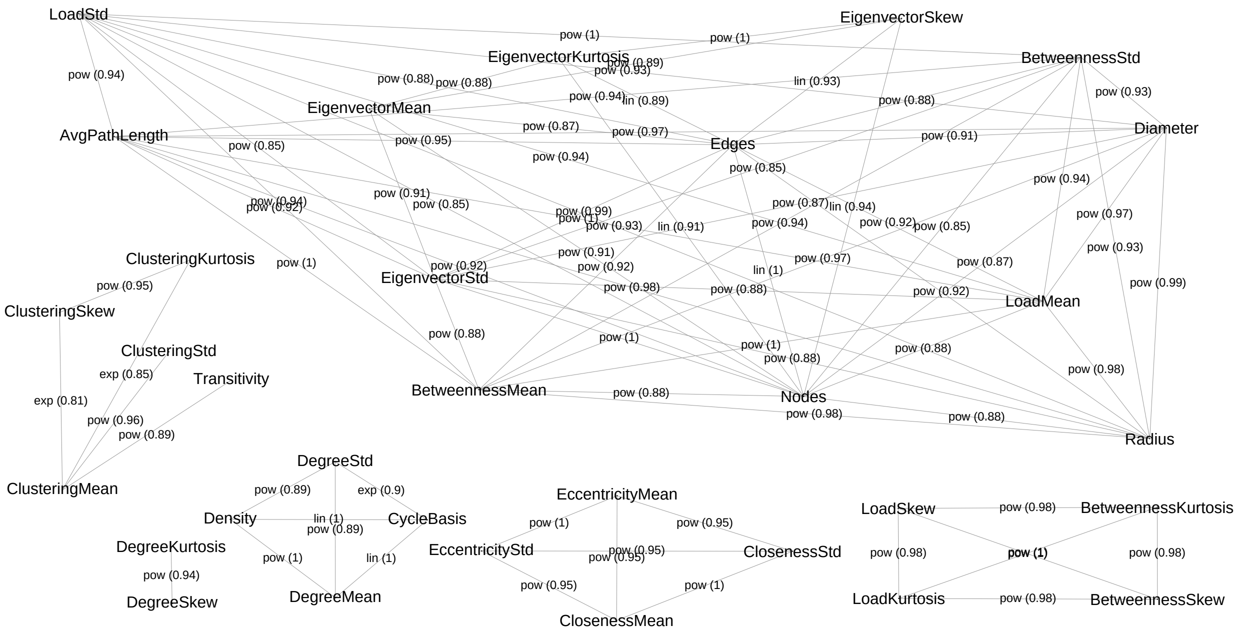


Figure 5: Connected components with functional types based on R^2

We select one metric from each of the six connected components to compare the similarity of the 58 air navigation route networks. Starting from the largest connected component, the fourteen selected metrics are listed as follows: average shortest path length, the mean of clustering coefficient, the kurtosis of degree, density, the mean of closeness centrality, the skew of betweenness centrality, and the other eight disconnected metrics.

4.4. Network evaluation using multi-metrics with functional independences

We select fourteen functionally independent metrics to compare the air navigation route networks for the 58 countries. Since the metrics have different magnitudes, we also normalize them between 0 and 1: $x_{\text{normalized}} = \frac{(x-x_{\min})}{(x_{\max}-x_{\min})}$, where x is the value of one metric.

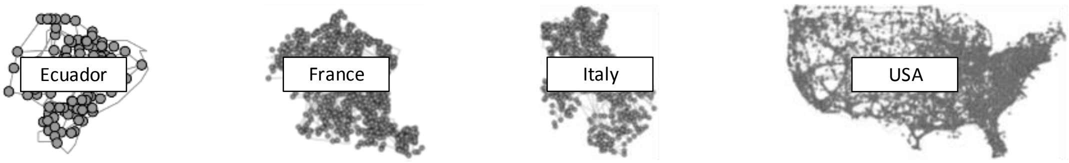
Figure 6 illustrates the flow chart how we assess the structural similarity of air navigation route networks for different countries. For the purpose of demonstration, we select the two most similar countries (France and Italy) and the two most dissimilar countries (Ecuador and USA) as an example.

At first, we compute the normalized values for the fourteen functionally independent metrics for the countries of interest. Secondly, for all network pairs, we compute the effect size for each metric: $\frac{(x_1-x_2)}{\sigma}$, where x_1 and x_2 are the metric values, σ is the standard deviation of this metric. The effect size expresses the difference between two metrics in standard deviation units. An effect size bigger than one indicates that the two networks are dissimilar regarding this metric. The similarity score is the Euclidean distance of the feature vectors between two networks. Lastly, we map the effect size matrix into a binary matrix. We assign zero to the entry when the effect size is less than one; otherwise one is assigned.

Based on the effect size matrix, for each network pair (each row), the metrics which have entry one make the network pairs distinguishable. For each metric (each column), the one which has the highest frequency is most discriminative when comparing all the network pairs. Among the fourteen metrics, we find that the average shortest path length is most discriminative (51% frequency); while the percentage of the periphery nodes is least discriminative (33% frequency).

Figure 7 presents the heat map with hierarchical clustering (dendrogram) for the 58 air navigation route networks. The clustering method is based on the group average with Euclidean distance [Manning et al., 2008]. In the hierarchical clustering, the two most similar networks are grouped together and continue to group until all networks are in the same cluster. In the heat map, the color represents the Euclidean distance between the feature vectors of two networks.

Air navigation route networks for countries of interest



Normalized values of 14 metrics for air navigation route networks

Country	Assortat	APL	DEntropy	Density	PCenter	PPeriphery	BSkew	CMean	CSkew	CKurtosis	CCMean	DKurtosis	ESkew	EKurtosis
Ecuador	0.0250	0.6154	0.6571	0.3849	0.5592	1.0000	0.1826	0.7059	0.9473	0.5584	0.3995	0.2444	0.0000	0.4472
France	0.6424	0.1109	0.7229	0.8839	0.0760	0.0264	0.3846	0.3893	0.5536	0.6336	0.4104	0.2172	0.7020	0.2963
Italy	0.6442	0.1787	0.7747	0.8782	0.0465	0.0939	0.2547	0.4370	0.3613	0.5936	0.4235	0.1839	0.6857	0.3270
USA	0.6124	0.0000	0.0000	0.5322	0.0038	0.0000	1.0000	0.1186	0.6484	0.3124	0.1164	0.6030	0.5298	0.1741

All-pairs matrix for air navigation route networks with effect size

Network pairs	Assortat	APL	DEntropy	Density	PCenter	PPeriphery	BSkew	CMean	CSkew	CKurtosis	CCMean	DKurtosis	ESkew	EKurtosis	SimilarityScore
Ecuador-France	2.7793	2.1030	0.3045	1.9394	2.0827	5.0403	1.1310	1.3246	1.5938	0.3575	0.0467	0.1418	3.2959	0.7629	1.7593
Ecuador-Italy	2.7874	1.8205	0.5441	1.9169	2.2098	4.6909	0.4041	1.1249	2.3722	0.1672	0.1030	0.3146	3.2195	0.6078	1.7401
Ecuador-USA	2.6443	2.5651	3.0412	0.5723	2.3941	5.1767	4.5757	2.4570	1.2099	1.1703	1.2124	1.8634	2.4877	1.3802	2.1246
France-Italy	0.0081	0.2825	0.2395	0.0225	0.1270	0.3494	0.7269	0.1996	0.7785	0.1903	0.0563	0.1729	0.0764	0.1550	0.2796
France-USA	0.1350	0.4621	3.3458	1.3671	0.3113	0.1364	3.4447	1.1325	0.3838	1.5278	1.2591	2.0051	0.8082	0.6173	1.2717
Italy-USA	0.1431	0.7446	3.5853	1.3446	0.1843	0.4859	4.1717	1.3321	1.1623	1.3375	1.3154	2.1780	0.7318	0.7724	1.4265

Explanation for air navigation route network similarity

Network pairs	Assortat	APL	DEntropy	Density	PCenter	PPeriphery	BSkew	CMean	CSkew	CKurtosis	CCMean	DKurtosis	ESkew	EKurtosis
Ecuador-France	1	1	0	1	1	1	1	1	1	0	0	0	1	0
Ecuador-Italy	1	1	0	1	1	1	0	1	1	0	0	0	1	0
Ecuador-USA	1	1	1	0	1	1	1	1	1	1	1	1	1	1
France-Italy	0	0	0	0	0	0	0	0	0	0	0	0	0	0
France-USA	0	0	1	1	0	0	1	1	0	1	1	1	0	0
Italy-USA	0	0	1	1	0	0	1	1	1	1	1	1	0	0

Figure 6: Flow chart of the similarity analysis for the air navigation route systems in 58 countries

Similar network pairs have light colors, while dissimilar network pairs have rather dark colors. We can observe from Figure 7 that among all the network pairs, France and Italy are closest to each other (the shortest Euclidean distance 0.2796); while Ecuador and USA are most distant to each other (the largest Euclidean distance 2.1246).

Based on the Euclidean distance, Table 2 shows the top five most similar and most dissimilar network pairs. The high similarity between France and Italy is not surprising, since both countries are members of the ECAC (European Civil Aviation Conference), Eurocontrol, and the ECAA (European Common Aviation Area). The European route network is designed by Eurocontrol in a centralized manner[3].

Regarding the most dissimilar country pairs, with approximately 20 times more population and 34 times larger geographical area, the air navigation route system in USA is most dissimilar with Ecuador. The reason of the structural dissimilarity of the two countries needs further investigation.

Further, we categorize the 58 countries into six continents: Africa, Asia, Europe, North America, Oceania, and South America. We aggregate the network similarity from the country level to the continental level as follows: First, among the 58 countries, we compute the Euclidean distance between any two countries. In total, we have 1653 pairs of the countries: ${}^{58}C_2 = \frac{58*57}{2} = 1653$. Second, we assign each country with its continent. Third, we create the pivot table based on the average of the Euclidean distance for each continent. After these three steps, we can aggregate the Euclidean distance from the country level to the continent level and the results are presented in Figure 8.

In this paper, we compute the Euclidean distance between the network metrics to measure the similarity of different air navigation route networks. Shorter Euclidean distance between two networks indicates that the two networks are closer to each other, thus, we can say that the two networks are more similar with each other. Therefore, in Figure 8 a lower score (shorter Euclidean distance) represents more similarity; while a higher score (larger Euclidean distance)

[3]http://www.eurocontrol.int/services/european-route-network-improvement-plan-ernip

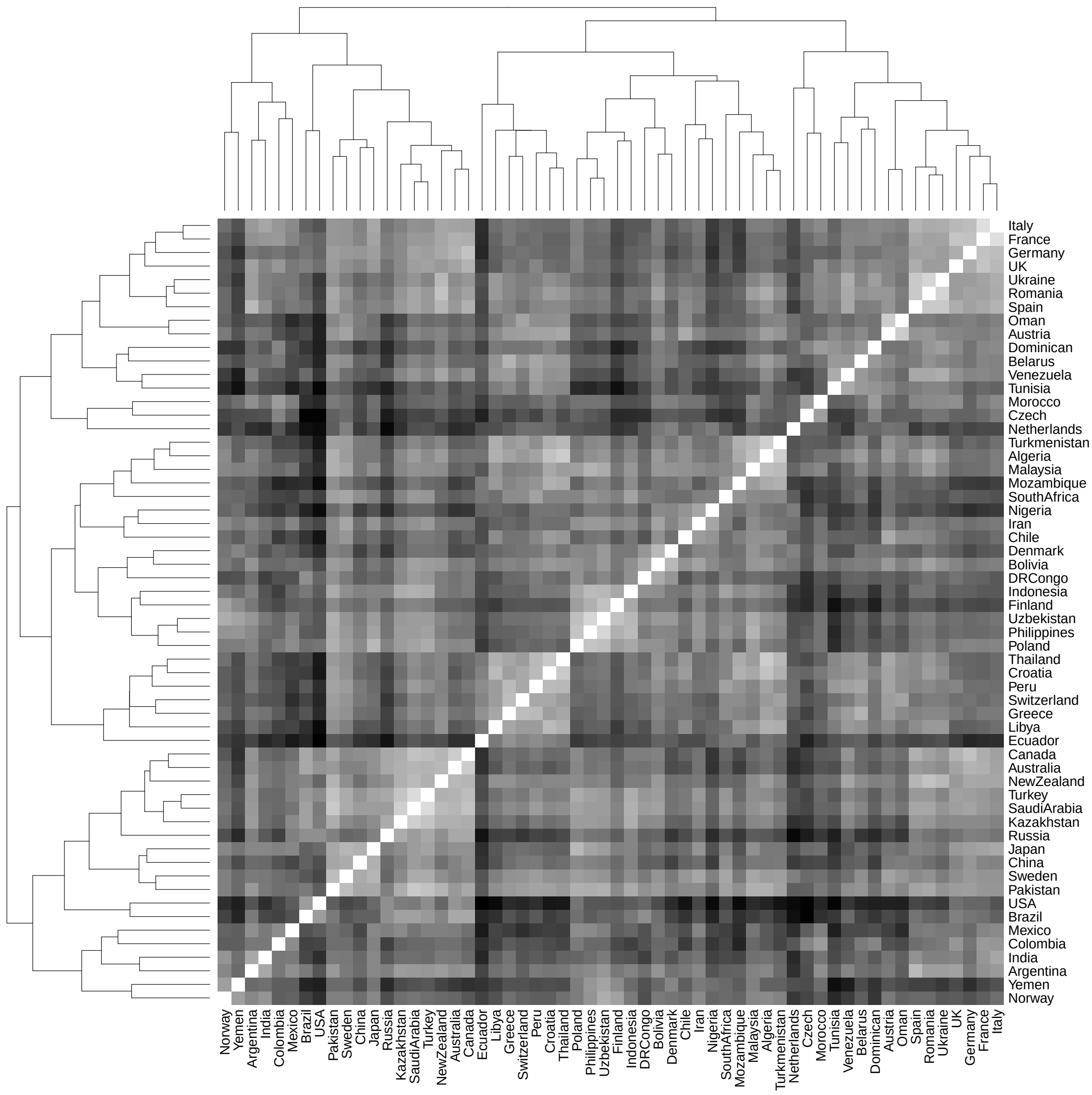


Figure 7: The dendrogram for the 58 air navigation route networks

represents more dissimilarity. The green cells stand for the most similar continent pairs; while the red cells stand for the most dissimilar continent pairs. We can observe that Oceania (Australia and New Zealand are included in the similarity analysis) is most similar to other continents; while North America (Canada, Dominican, Mexico, and USA are included in the similarity analysis) is most dissimilar to other continents.

We formulate the following hypothesis: Since the air navigation route systems in North America are most dissimilar to all other countries, they should not be used as representatives in further studies. The properties derived from the air navigation route systems in North America should not be generalized to other countries without careful consideration. Rather, the air navigation route systems in Oceania ensemble the common features of most countries. Therefore, initial analysis of new ATM operational procedures could be performed for these single countries on a small scale. After the initial assessment, further decisions could be taken on a larger scale.

Note that our current research focuses on topological properties of the air navigation route networks. Several

Table 2: Top five most similar and most dissimilar network pairs based on the Euclidean distance

Network pairs	**Euclidean distance**
Most similar	
France - Italy	0.2796
Saudi Arabia - Turkey	0.3005
Philippines - Uzbekistan	0.3392
Romania - Ukraine	0.3727
Kazakhstan - Saudi Arabia	0.4059
Most dissimilar	
Ecuador - USA	2.1246
Czech - USA	2.1031
Brazil - Czech	2.0896
Tunisia - USA	2.0620
Netherlands - Russia	2.0544

	Africa	Asia	Europe	North America	Oceania	South America
Africa	1.0752	1.1766	1.2253	**1.4908**	1.2812	1.2218
Asia	1.1766	0.9402	1.1042	1.3184	1.0599	1.2031
Europe	1.2253	1.1042	1.0714	1.2675	1.0138	1.1918
North America	1.4908	1.3184	1.2675	1.0229	0.9788	1.3385
Oceania	1.2812	1.0599	1.0138	0.9788	**0.2995**	1.1072
South America	1.2218	1.2031	1.1918	1.3385	1.1072	1.1018

Figure 8: Network similarity analysis at continental level

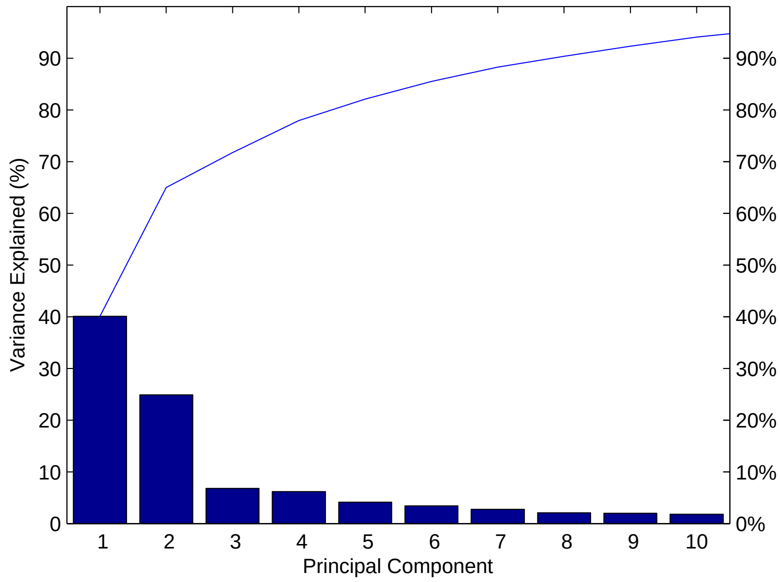


Figure 9: The first ten components which explain 95% of the total variance

issues such as loads level on links and nodes, frequencies of flights, and time schedule coordination are not included. This limitation should be kept in mind when using a network as a reference network for further studies, such as how delay is propagated in the network and what is the level of services provided by the network.

5. Discussion

In the current study, we select the functionally independent metrics to compare the similarity of the 58 air navigation route networks and functional dependencies are detected based on regression analysis. The deletion of redundant or irrelevant features is one active research area in data mining [Kantardzic, 2011]. Feature selection is an important and frequently used technique for dimension reduction [Liu and et. al., 2010]. It was shown that data mining techniques could be used for the removal of irrelevant features in complex networks [Zanin and et. al., 2012, Zanin et al., 2014]. There are several multivariate statistical techniques when examining the interplay between independent variables, one of which is Principal Component Analysis (PCA) [Shlens, 2014]. In this research, we also perform PCA for the forty network metrics.

Figure 9 shows that the first ten components explain 95% of the total variance. We can see that the first component explains 40% of the variance and the second component explains 25% of the variance. Therefore, the first two components explain 65% of the variance together.

Figure 10 presents an overview of the first two component coefficients for each of the forty metrics, where the direction and length of the vector indicate how each metric contributes to the two principal components; each singular point represents a country.

Figure 16 (see the Appendix) presents the first two component coefficients for the six *connected components*. We find out that the metrics in each of our *connected components based on the functional dependency* have also related/similar vectors in PCA: vectors which have a similar length and either 1) have a very small angle or 2) are pointing completely in the opposite direction. Similar length indicates the two metrics contribute to the first two components similarly; direction corresponds to the contribution is positive or negative.

Figure 11 shows the first two component coefficients of the eight metrics for which we could not find any functional dependencies. We can observe that these eight metric vectors have rather different directions and different length, i.e., the contributions of these eight metrics to the 58 countries are quite different. Especially, the first component has negative coefficients for assortativity and the skew of eccentricity, and positive coefficients for the remained six metrics. This indicates that the first component distinguishes among the 58 countries which have low values of

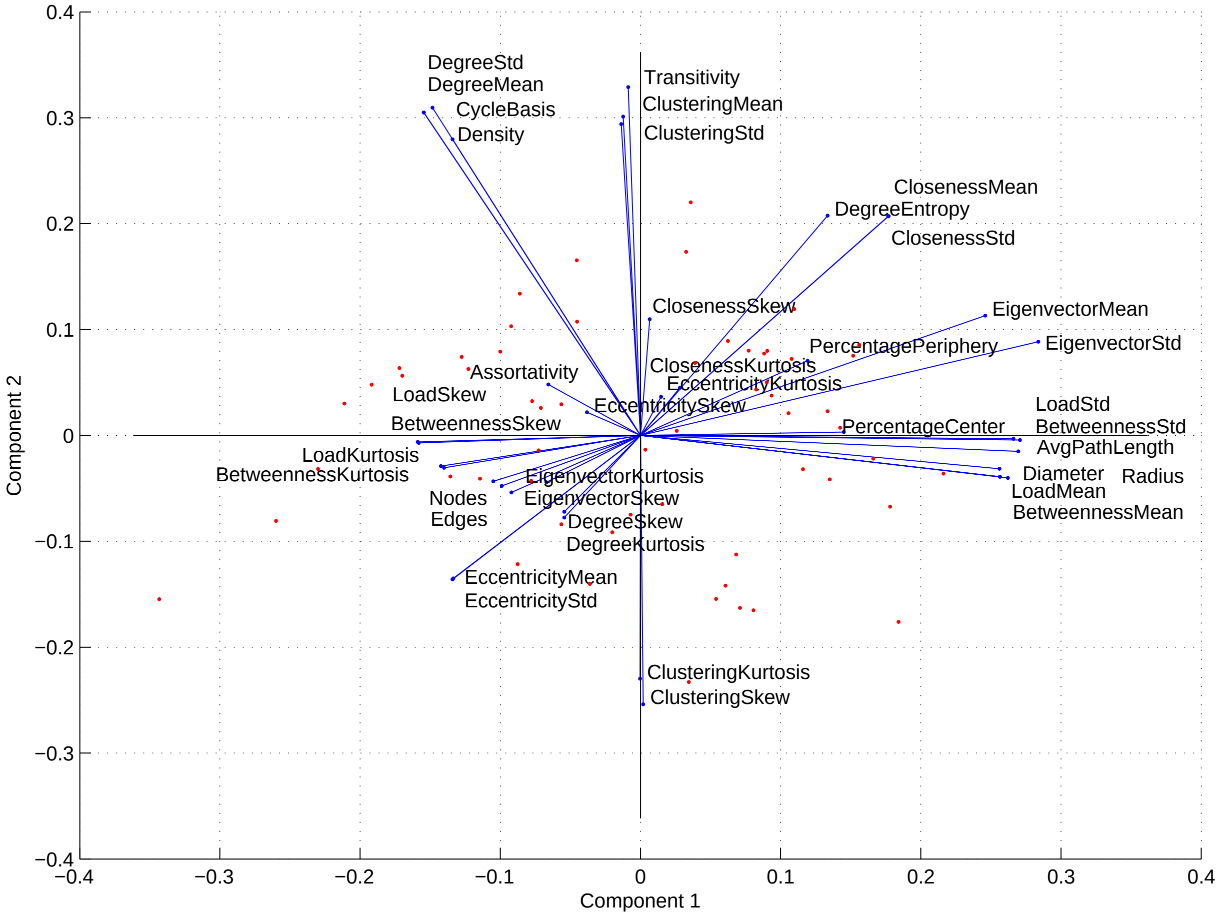


Figure 10: The first two component coefficients for each of the forty metrics

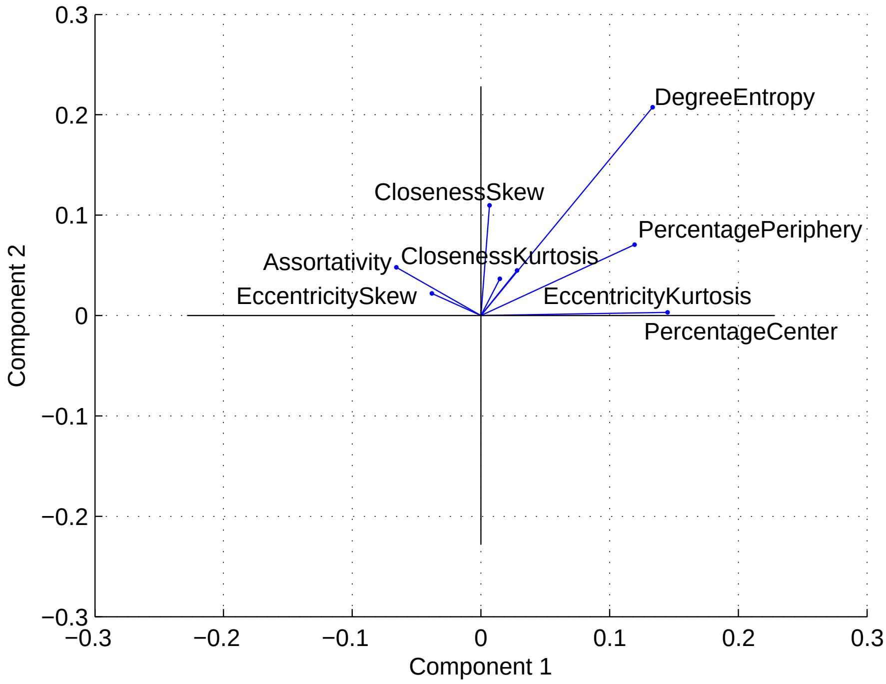


Figure 11: The first two component coefficients for the eight disconnected metrics

assortativity and the skew of eccentricity, and high values for the remained six metrics, and the countries which have high values of assortativity and the skew of eccentricity, and low values of the remained six metrics. Therefore, PCA confirms that all these eight metrics are significant for the similarity analysis of the air navigation route systems in the 58 countries.

In summary, PCA results support our own methodology to detect the functional dependencies among the network metrics.

6. Conclusions

In this research, we investigated the structural similarity of the air navigation route systems in different countries. Our new methodology was summarized in Figure 1 and the flow chart was illustrated in Figure 6. The contributions of the paper to the literature are:

1. We identified functional dependencies among the network metrics for the air navigation route systems through regression analysis, instead of the correlation coefficients, as believed in previous work on network comparison.
2. We selected fourteen functionally independent metrics to compare the structural similarity of 58 air navigation route networks.
3. We found that the air navigation route systems in France and Italy are most similar; while the ones in Ecuador and USA are most dissimilar. Among the fourteen selected metrics, the average shortest path length is most discriminative; while the percentage of the periphery nodes is least discriminative.
4. The aggregation of the network similarity from the country level to the continent level showed that the air navigation route systems in Oceania ensemble the common features of most countries and should be used as representatives for analysis of new ATM operational procedures in the future.

Our new methodology could serve as a reference to compare different types of networks, without the bias of network size and functional dependencies among network metrics. For instance, the structural similarity of different public transportation networks [Sienkiewicz and Hołyst, 2005].

Acknowledgment

The authors would like to thank Eurocontrol for providing the data in this study.

References

Neil Bania, Paul W. Bauer, and Thomas J. Zlatoper. U.s. air passenger service: a taxonomy of route networks, hub locations, and competition. *Transportation Research Part E: Logistics and Transportation Review*, 34(1):53–74, 1998. ISSN 1366-5545.

Albert-Lszl Barabsi. Network science. *Philosophical Transactions of The Royal Society*, 371:1–3, 2013.

Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

Patrick Baumgarten, Robert Malina, and Anne Lange. The impact of hubbing concentration on flight delays within airline networks: An empirical analysis of the {US} domestic market. *Transportation Research Part E: Logistics and Transportation Review*, 66(0):103 – 114, 2014. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/j.tre.2014.03.007. URL http://www.sciencedirect.com/science/article/pii/S1366554514000404.

Michele Berlingerio, Danai Koutra, Tina Eliassi-Rad, and Christos Faloutsos. Netsimile: A scalable approach to size-independent network similarity. *CoRR*, abs/1209.2684, 2012.

Gergana Bounova and Olivier de Weck. Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. *Physical Review E*, 85:016117, Jan 2012.

Ulrik Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.

M.E. Bruni, F. Guerriero, and P. Beraldi. Designing robust routes for demand-responsive transport systems. *Transportation Research Part E: Logistics and Transportation Review*, 70(0):1–16, 2014. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/j.tre.2014.06.002. URL http://www.sciencedirect.com/science/article/pii/S1366554514000970.

Kai-Quan Cai, Jun Zhang, Wen-Bo Du, and Xian-Bin Cao. Analysis of the chinese air route network as a complex network. *Chinese Physics B*, 21 (2):1–7, 2012.

Alessio Cardillo and et. al. Modeling the multi-layer nature of the european air transport network. *The European Physical Journal Special Topics*, 215(1):23–33, 2013.

Daniel DeLaurentis, En-Pei Han, and Tatsuya Kotegawa. Network-theoretic approach for analyzing connectivity in air transportation networks. *Journal of Aircraft*, 45(5):1669–1679, 2008.

Reinhard Diestel. *Graph Theory*. Springer-Verlag, Heidelberg, 2010.

Abdol Hossein Esfahanian. Connectivity algorithms. http://www.cse.msu.edu/~cse835/Papers/Graph_connectivity_revised.pdf, 2012.

European Commission. Flightpath 2050 europe's vision for aviation, 2011.

Pablo Fleurquin, Jose Ramasco, and Victor Eguiluz. Data-driven modeling of systemic delay propagation under severe meteorological conditions. In *10th USA/Europe ATM Seminar*, 2013a.

Pablo Fleurquin, Jos Ramasco, and Victor Eguiluz. Systemic delay propagation in the us airport network. *Scientific Reports*, 3:1–6, 2013b.

L. C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1978.

A. Garcia-Robledo, A. Diaz-Perez, and G. Morales-Luna. Correlation analysis of complex network metrics on the topology of the internet. In *Emerging Technologies for a Smarter World (CEWIT), 2013 10th International Conference and Expo on*, pages 1–6, Oct 2013.

R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities global roles. *PNAS*, 102(22):7794–7799, 2005.

G. Gurtner and et. al. Multi-scale analysis of the european airspace using network community detection. *arXiv:1306.3769v1*, pages 1–22, 2013.

A. Jamakovic, S. Uhlig, and I. Theisler. On the relationships between topological metrics in real-world networks. In *4th European Conference on Complex Systems, Dresden, October 1-6*, 2007.

Shi jin Wang and Yan hui Gong. Research on air route network nodes optimization with avoiding the three areas. *Safety Science*, 66(0):9–18, 2014. ISSN 0925-7535.

Radosav Jovanovic, Vojin Tosic, Mirjana Cangalovic, and Milan Stanojevic. Anticipatory modulation of air navigation charges to balance the use of airspace network capacities. *Transportation Research Part A: Policy and Practice*, 61(0):84–99, 2014. ISSN 0965-8564.

Mehmed Kantardzic. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.

Tatsuya Kotegawa, Donald Fry, Daniel DeLaurentis, and Ethan Puchaty. Impact of service network topology on air transportation efficiency. *Transportation Research Part C: Emerging Technologies*, 2013. ISSN 0968-090X.

Maciej Kurant and Patrick Thiran. Extraction and analysis of traffic and topologies of transportation networks. *Physical Review E*, 74(3):036114, 2006.

C Li, H Wang, W De Haan, CJ Stam, and P Van Mieghem. The correlation of metrics in complex networks with applications in functional brain networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(11):P11018, 2011a.

Geng Li, Murat Semerci, Bülent Yener, and Mohammed J Zaki. Graph classification via topological and label attributes. In *9th Workshop on Mining and Learning with Graphs*, 2011b.

F. Lillo, S. Micciche, R. N. Mantegna, V. Beato, and S. Pozzi. Elsa project: Toward a complex network approach to atm delays analysis. In *First SESAR Innovation Days*, 2011.

Huan Liu and et. al. Feature selection: An ever evolving frontier in data mining. *Journal of Machine Learning Research-Proceedings Track*, 10: 4–13, 2010.

Oriol Lordan, Jose M. Sallan, Pep Simo, and David Gonzalez-Prieto. Robustness of the air transport network. *Transportation Research Part E: Logistics and Transportation Review*, 68(0):155–163, 2014. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/j.tre.2014.05.011. URL http://www.sciencedirect.com/science/article/pii/S1366554514000805.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008. ISBN 0521865719, 9780521865715.

Lei Meng, Tijana Milenkovic, and Aaron Striegel. Systematic dynamic and heterogeneous analysis of rich social network data. volume 549, pages 25–37. Springer, 2014.

M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5):056131, November 2004.

Mark Newman. *Networks-An Introduction*. Oxford University Press, 2010.

M.E.J. Newman. Assortative mixing in networks. *Physical Review Letter*, 89:208701, 2002.

NIST. Handbook of statistical methods. http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm, 2014.

Yi-Feng Niu, William H.K. Lam, and Ziyou Gao. An efficient algorithm for evaluating logistics network reliability subject to distribution cost. *Transportation Research Part E: Logistics and Transportation Review*, 67(0):175–189, 2014. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/j.tre.2014.04.009. URL http://www.sciencedirect.com/science/article/pii/S136655451400060X.

Jukka-Pekka Onnela, Daniel J Fenn, Stephen Reid, Mason A Porter, Peter J Mucha, Mark D Fricker, and Nick S Jones. Taxonomies of networks from community structure. *Physical Review E*, 86(3):036104, 2012.

Stefano Paleari, Renato Redondi, and Paolo Malighetti. A comparative study of airport connectivity in china, europe and us: which network provides the best service to passengers? *Transportation Research Part E: Logistics and Transportation Review*, 46(2):198–210, 2010.

E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065, 1962.

Natasa Przulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):177–183, 2007.

Olga Rodionova, Mohamed Sbihi, Daniel Delahaye, and Marcel Mongeau. Optimization of aicraft trajectories in north atlantic oceanic airspace. In *5th International Conferences on Research in Air Transportation (ICRAT)*, 2012.

Soumen Roy. Systems biology beyond degree, hubs and scale-free networks: the case for multiple metrics in complex networks. *Systems and synthetic biology*, 6(1-2):31–34, 2012.

C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:623656, 1948.

Jonathon Shlens. A tutorial on principal component analysis. *arXiv:1404.1100*, 1:1–12, 2014.

Julian Sienkiewicz and Janusz A. Hołyst. Statistical analysis of 22 public transport networks in poland. *Physical Review E*, 72:046127, Oct 2005.

B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, Great Britain, 1986.

Sucheta Soundarajan, Tina Eliassi-Rad, and Brian Gallagher. Which network similarity measure should you choose: An empirical study. In *Workshop on Information in Networks, New York, USA*, 2013.

S. Vitali and et. al. Statistical regularities in atm: Network properties, trajectory deviations and delays. In *Second SESAR Innovation Days*, 2012.

Evgenia Vogiatzi. Goodness of fit and correlation coefficient. http://www.oocities.org/qecon2002/founda6.html, 2002.

Christian von Hirschhausen and Astrid Cullmann. A nonparametric efficiency analysis of german public transport companies. *Transportation Research Part E: Logistics and Transportation Review*, 46(3):436 – 445, 2010. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/j.tre.2009.11.005. URL http://www.sciencedirect.com/science/article/pii/S1366554509001380.

P. Wei, L. Chen, and D. Sun. Algebraic connectivity maximization of an air transportation network: The flight routes addition/deletion problem. *Transportation Research Part E: Logistics and Transportation Review*, 61(0):13–27, 2014. ISSN 1366-5545.

Massimiliano Zanin and et. al. Optimizing functional network representation of multivariate time series. *Scientific Reports*, 2:1–6, 2012.

Massimiliano Zanin and Fabrizio Lillo. Modelling the air transport with complex networks: A short review. *European Physical Journal Special Topics*, 215:5–21, 2013.

Massimiliano Zanin, Ernestina Menasalvas, Stefano Boccaletti, and PedroA. Sousa. Analysis of complex data by means of complex networks. volume 423, pages 39–46. Springer, 2014.

Hai-Tian Zhang, Tao Yu, Jian-Ping Sang, and Xian-Wu Zou. Dynamic fluctuation model of complex networks with weight scaling behavior and its application to airport networks. *Physica A: Statistical Mechanics and its Applications*, 393(0):590–599, 2014. ISSN 0378-4371.

Yimin Zhang. Network structure and capacity requirement: The case of china. *Transportation Research Part E: Logistics and Transportation Review*, 46(2):189–197, 2010. ISSN 1366-5545.

Chunxue Zhao, Baibai Fu, and Tianming Wang. Braess paradox and robustness of traffic networks under stochastic user equilibrium. *Transportation Research Part E: Logistics and Transportation Review*, 61(0):135–141, 2014. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/j.tre.2013.11.008. URL http://www.sciencedirect.com/science/article/pii/S1366554513001920.

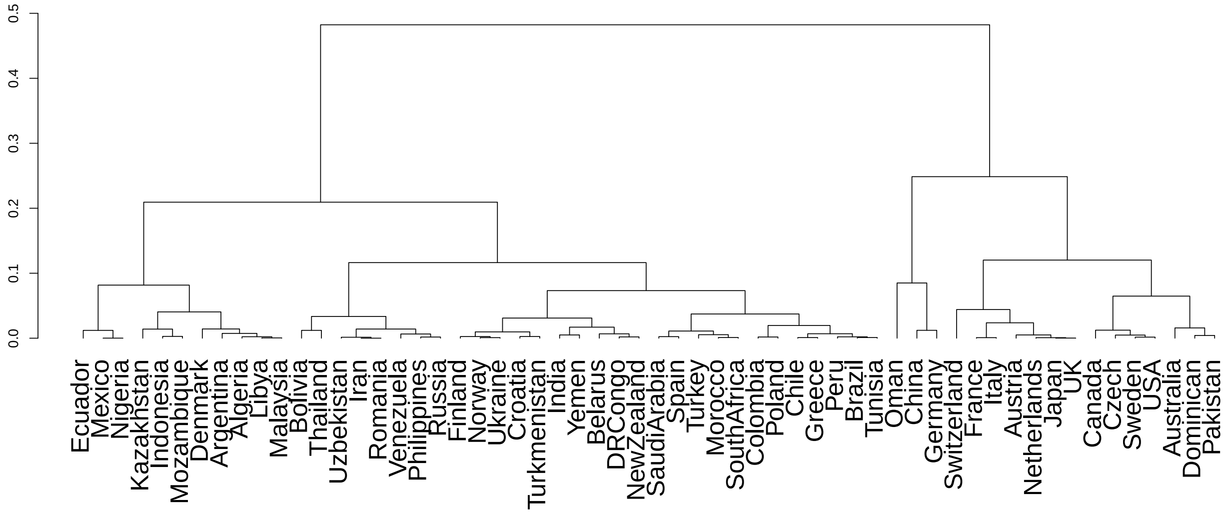
Bo Zou and Mark Hansen. Flight delay impact on airfare and flight frequency: A comprehensive assessment. *Transportation Research Part E: Logistics and Transportation Review*, 69(0):54–74, 2014.

Appendix

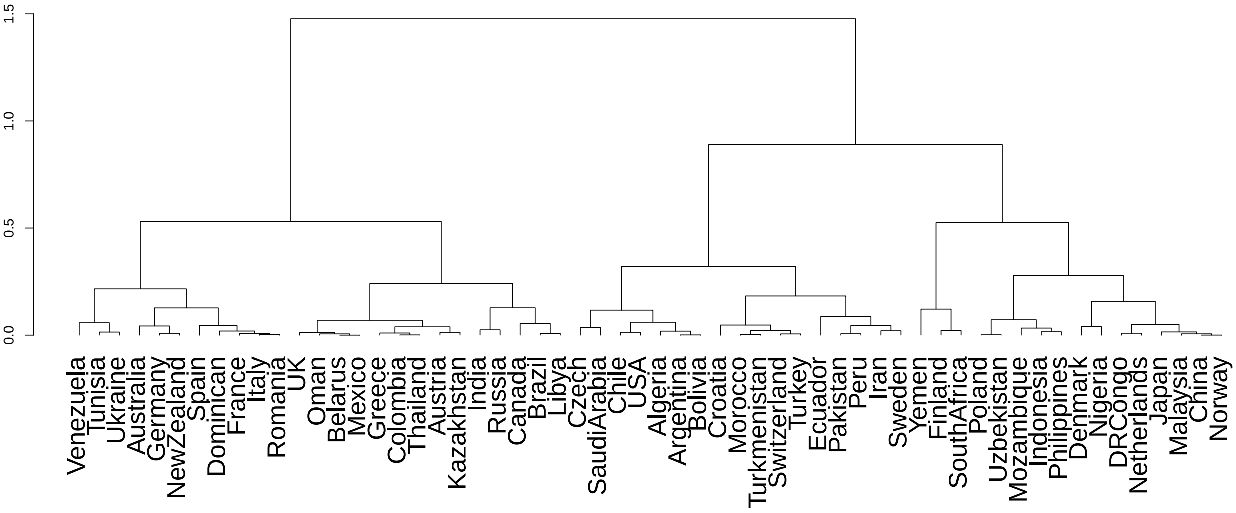
In the Appendix, we use the following abbreviations for the 41 network metrics: APL (AvgPathLength), DE(DegreeEntropy), Dens (Density), Diam (Diameter), Edg (Edges), Nod (Nodes), PercC (PercentageCenter), PercP (PercentagePeriphery), Rad (Radius), Tran (Transitivity), Asso (Assortativity), CycB (CycleBasis), BetM (BetweennessMean), BetSt (BetweennessStd), BetSk (BetweennessSkew), BetK (BetweennessKurtosis), ClosM (ClosenessMean), ClosSt (ClosenessStd), ClosSk (ClosenessSkew), ClosK (ClosenessKurtosis), LoadM (LoadMean), LoadSt (LoadStd), LoadSk (LoadSkew), LoadK (LoadKurtosis), DegrM (DegreeMean), DegrSt (DegreeStd), DegrSk (DegreeSkew), DegrK (DegreeKurtosis), EccM (EccentricityMean), EccSt (EccentricityStd), EccSk (EccentricitySkew), EccK (EccentricityKurtosis), EigM (EigenvectorMean), EigSt (EigenvectorStd), EigSk (EigenvectorSkew), EigK (EigenvectorKurtosis), ClusM (ClusteringMean), ClusSt (ClusteringStd), ClusSk (ClusteringSkew), ClusK (ClusteringKurtosis).

Table 3: Geography overview for the 58 countries

Country	Area	Population	Nodes	Links
Algeria	2,381,740	34,586,184	121	188
Argentina	2,766,890	41,343,201	400	608
Australia	7,686,850	21,515,754	1,167	2,154
Austria	83,858	8,205,000	135	214
Belarus	207,600	9,685,000	117	195
Bolivia	1,098,580	9,947,418	166	260
Brazil	8,511,965	201,103,330	1,461	2,522
Canada	9,984,670	33,679,000	1,202	2,109
Chile	756,950	16,746,491	140	217
China	9,596,960	**1,330,044,000**	610	840
Colombia	1,138,910	44,205,293	353	565
Croatia	56,542	4,491,000	**101**	143
Czech	78,866	10,476,000	127	194
Denmark	43,094	5,484,000	118	148
Dominican	48,730	9,823,821	115	206
DR Congo	2,345,410	70,916,439	243	328
Ecuador	283,560	14,790,608	128	188
Finland	337,030	5,244,000	250	298
France	547,030	64,768,389	912	1,646
Germany	357,021	81,802,257	952	1,733
Greece	131,940	11,000,000	160	264
India	3,287,590	1,173,108,018	358	605
Indonesia	1,919,440	242,968,342	309	384
Iran	1,648,000	76,923,300	344	503
Italy	301,230	60,340,328	565	1,013
Japan	377,835	127,288,000	660	913
Kazakhstan	2,717,300	15,340,000	555	910
Libya	1,759,540	6,461,454	113	198
Malaysia	329,750	28,274,729	170	223
Mexico	1,972,550	112,468,855	710	1,228
Morocco	446,550	31,627,428	128	193
Mozambique	801,590	22,061,451	107	**130**
Netherlands	41,526	16,645,000	150	183
New Zealand	268,680	4,252,277	339	619
Nigeria	923,768	154,000,000	118	165
Norway	324,220	5,009,150	246	293
Oman	212,460	**2,967,717**	121	196
Pakistan	803,940	184,404,791	215	337
Peru	1,285,220	29,907,003	144	204
Philippines	300,000	99,900,177	206	263
Poland	312,685	38,500,000	337	440
Romania	237,500	21,959,278	234	423
Russia	**17,100,000**	140,702,000	2,904	4,871
Saudi Arabia	1,960,582	25,731,776	387	631
South Africa	1,219,912	49,000,000	236	287
Spain	504,782	46,505,963	411	728
Sweden	449,964	9,555,893	244	353
Switzerland	**41,290**	7,581,000	127	190
Thailand	514,000	67,089,500	**101**	162
Tunisia	163,610	10,589,025	124	236
Turkey	780,580	77,804,122	430	653
Turkmenistan	488,100	4,940,916	114	169
Ukraine	603,700	45,415,596	335	622
UK	244,820	62,348,447	555	930
USA	9,629,091	310,232,863	**9,050**	**14,346**
Uzbekistan	447,400	27,865,738	192	244
Venezuela	912,050	27,223,228	245	453
Yemen	527,970	23,495,361	127	143



(a) Assortativity



(b) Density

Figure 12: Evaluation of the 58 air navigation route networks using single metrics

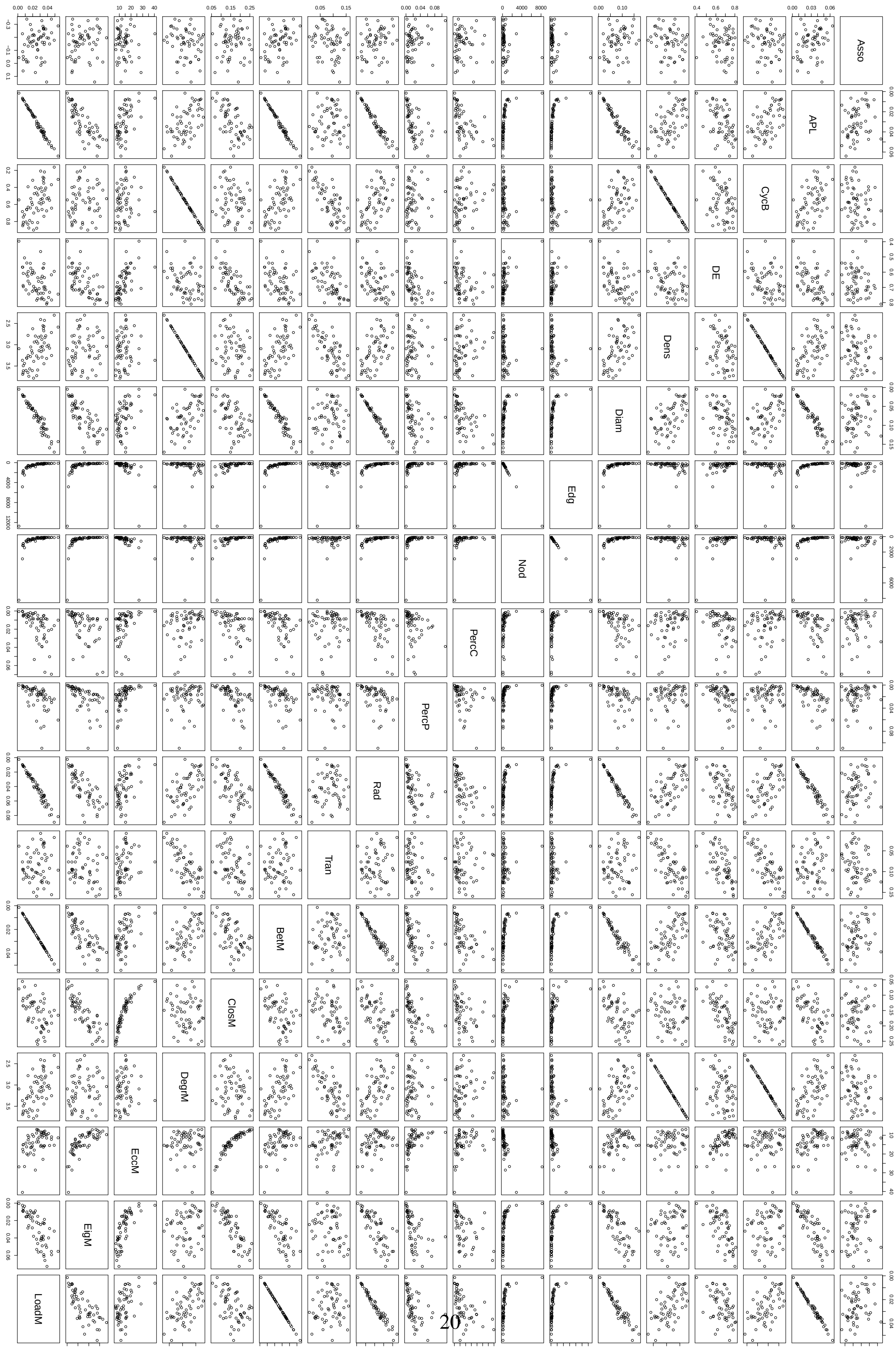


Figure 13: Scatter plots for the network metrics, after the size-dependent normalization.

	Independent variable																																							
Dependent variable	Assortativity	AvgPathLength	BetweennessKurtosis	BetweennessMean	BetweennessSkew	BetweennessStd	ClosenessKurtosis	ClosenessMean	ClosenessSkew	ClosenessStd	ClusteringKurtosis	ClusteringMean	ClusteringSkew	ClusteringStd	CycleBasis	DegreeEntropy	DegreeKurtosis	DegreeMean	DegreeSkew	DegreeStd	Density	Diameter	EccentricityKurtosis	EccentricityMean	EccentricitySkew	EccentricityStd	Edges	EigenvectorKurtosis	EigenvectorMean	EigenvectorSkew	EigenvectorStd	LoadKurtosis	LoadMean	LoadSkew	LoadStd	Nodes	PercentageCenter	PercentagePeriphery	Radius	Transitivity
Assortativity	=	exp	log	exp	log	exp	exp	log	log	log	pow	log	pow	log	exp	log	log	exp	log	lin	exp	exp	log	pow	exp	pow	exp	lin	exp	lin	exp	log	exp	log	exp	exp	log	log	exp	log
AvgPathLength	log	=	log	pow	log	pow	pow	pow	pow	pow	pow	exp	pow	exp	lin	pow	pow	lin	pow	lin	lin	pow	pow	log	lin	log	pow	pow	pow	pow	pow	log	pow	log	pow	pow	pow	pow	pow	exp
BetweennessKurtosis	exp	exp	=	exp	pow	exp	pow	exp	exp	exp	exp	pow	exp	pow	log	exp	pow	log	pow	log	log	exp	pow	lin	pow	lin	lin	lin	exp	lin	exp	pow	exp	pow	exp	lin	exp	exp	exp	pow
BetweennessMean	log	pow	log	=	log	pow	exp	pow	pow	pow	pow	exp	lin	exp	lin	pow	pow	lin	pow	lin	lin	pow	pow	log	lin	log	pow	pow	pow	pow	pow	log	=	log	pow	pow	pow	pow	pow	exp
BetweennessSkew	exp	exp	pow	exp	=	exp	pow	exp	log	exp	exp	pow	log	pow	log	exp	pow	pow	pow	log	pow	exp	pow	lin	log	lin	exp	exp	exp	lin	exp	pow	exp	pow	exp	exp	exp	exp	exp	pow
BetweennessStd	log	pow	log	pow	log	=	pow	pow	pow	pow	pow	log	pow	log	lin	pow	pow	lin	pow	lin	lin	pow	pow	pow	lin	pow	pow	pow	pow	pow	pow	log	pow	log	pow	pow	pow	pow	pow	log
ClosenessKurtosis	log	pow	pow	log	pow	pow	=	exp	exp	exp	pow	log	pow	log	pow	exp	pow	pow	pow	pow	pow	pow	exp	lin	exp	lin	log	exp	exp	lin	pow	pow	log	pow	pow	log	log	exp	pow	pow
ClosenessMean	exp	pow	log	pow	log	pow	log	=	pow	pow	exp	lin	exp	lin	exp	lin	log	exp	log	exp	exp	pow	exp	pow	exp	pow	pow	pow	lin	pow	lin	log	pow	log	pow	pow	pow	pow	pow	lin
ClosenessSkew	exp	pow	log	pow	exp	pow	log	pow	=	pow	pow	log	pow	log	exp	pow	pow	exp	log	lin	exp	pow	pow	pow	exp	pow	log	log	pow	log	pow	log	pow	exp	pow	pow	pow	pow	pow	exp
ClosenessStd	exp	pow	log	pow	log	pow	log	pow	pow	=	exp	lin	exp	lin	exp	lin	log	exp	log	exp	exp	pow	exp	pow	exp	pow	pow	pow	lin	pow	lin	log	pow	log	pow	pow	pow	pow	pow	lin
ClusteringKurtosis	pow	pow	log	pow	log	pow	pow	log	pow	log	=	log	pow	log	exp	exp	lin	exp	lin	pow	exp	pow	pow	pow	exp	pow	log	pow	pow	pow	pow	log	pow	log	pow	pow	pow	pow	pow	log
ClusteringMean	exp	log	pow	log	pow	exp	exp	lin	exp	lin	exp	=	exp	pow	pow	log	log	pow	log	pow	pow	log	pow	exp	exp	exp	lin	exp	exp	lin	exp	pow	log	pow	exp	lin	exp	exp	log	pow
ClusteringSkew	pow	pow	log	lin	exp	pow	pow	log	pow	log	pow	log	=	log	exp	exp	lin	exp	lin	pow	exp	pow	pow	pow	exp	pow	log	log	pow	log	log	log	lin	exp	pow	log	pow	pow	lin	exp
ClusteringStd	exp	log	pow	log	pow	exp	exp	lin	exp	lin	exp	pow	exp	=	pow	log	log	pow	log	pow	pow	log	pow	exp	exp	exp	lin	lin	exp	lin	exp	pow	log	pow	exp	lin	exp	exp	log	pow
CycleBasis	log	lin	exp	lin	exp	lin	pow	log	log	log	log	pow	log	pow	=	log	log	lin	log	exp	lin	lin	exp	pow	pow	pow	exp	pow	pow	pow	lin	exp	lin	exp	lin	exp	exp	pow	lin	pow
DegreeEntropy	exp	pow	log	pow	log	pow	log	lin	pow	lin	log	exp	log	exp	exp	=	exp	exp	exp	exp	exp	pow	exp	exp	lin	exp	pow	pow	pow	pow	pow	log	pow	log	pow	pow	pow	pow	pow	lin
DegreeKurtosis	exp	pow	pow	pow	pow	pow	pow	exp	pow	exp	lin	exp	lin	exp	exp	log	=	exp	pow	log	exp	pow	pow	exp	pow	exp	lin	exp	log	exp	pow	pow	pow	pow	pow	pow	exp	log	pow	exp
DegreeMean	log	lin	exp	lin	pow	lin	pow	log	log	log	log	pow	log	pow	lin	log	log	=	log	pow	=	lin	exp	pow	pow	pow	exp	pow	pow	pow	lin	exp	lin	exp	lin	exp	exp	pow	lin	pow
DegreeSkew	exp	pow	pow	pow	pow	pow	pow	exp	exp	exp	lin	exp	lin	exp	exp	log	pow	exp	=	log	exp	pow	exp	exp	pow	exp	lin	exp	log	lin	exp	pow	pow	pow	pow	lin	exp	exp	pow	exp
DegreeStd	lin	lin	exp	lin	exp	lin	pow	log	lin	log	pow	pow	pow	pow	log	log	exp	pow	exp	=	pow	lin	log	pow	pow	pow	exp	log	exp	log	exp	exp	lin	exp	lin	exp	exp	pow	lin	pow
Density	log	lin	exp	lin	pow	lin	pow	log	log	log	log	pow	log	pow	lin	log	log	=	log	pow	=	lin	exp	pow	pow	pow	exp	pow	pow	pow	lin	exp	lin	exp	lin	exp	exp	pow	lin	pow
Diameter	log	pow	log	pow	log	pow	pow	pow	pow	pow	pow	exp	pow	exp	lin	pow	pow	lin	pow	lin	lin	=	pow	log	log	log	pow	pow	pow	pow	pow	log	pow	log	pow	pow	pow	pow	pow	exp
EccentricityKurtosis	exp	pow	pow	pow	pow	pow	log	log	pow	log	pow	pow	pow	pow	log	log	pow	log	log	exp	log	pow	=	pow	pow	pow	log	pow	pow	pow	pow	pow	pow	pow	pow	log	log	pow	pow	pow
EccentricityMean	pow	exp	lin	exp	lin	pow	lin	pow	pow	pow	pow	log	pow	log	pow	log	log	pow	log	pow	pow	exp	pow	=	pow	pow	exp	exp	exp	exp	log	lin	exp	lin	pow	pow	exp	exp	exp	log
EccentricitySkew	log	lin	pow	lin	exp	lin	log	log	log	log	log	log	log	log	pow	lin	pow	pow	pow	pow	pow	exp	pow	pow	=	pow	exp	exp	log	exp	lin	pow	lin	exp	lin	exp	exp	lin	exp	log
EccentricityStd	pow	exp	lin	exp	lin	pow	lin	pow	pow	pow	pow	log	pow	log	pow	log	log	pow	log	pow	pow	exp	pow	pow	pow	=	exp	exp	exp	exp	log	lin	exp	lin	pow	pow	exp	exp	exp	log
Edges	log	pow	lin	pow	log	pow	exp	pow	exp	pow	exp	lin	exp	lin	log	pow	lin	log	lin	log	log	pow	exp	log	log	log	=	lin	pow	lin	pow	lin	pow	log	pow	lin	pow	pow	pow	lin
EigenvectorKurtosis	lin	pow	lin	pow	log	pow	log	pow	exp	pow	pow	log	exp	lin	pow	pow	log	pow	log	exp	pow	pow	pow	log	log	log	lin	=	pow	pow	pow	log	pow	log	pow	lin	pow	pow	pow	log
EigenvectorMean	log	pow	log	pow	log	pow	log	lin	pow	lin	pow	log	pow	log	pow	pow	exp	pow	exp	log	pow	pow	pow	log	exp	log	pow	pow	=	pow	pow	log	pow	log	pow	pow	pow	pow	pow	lin
EigenvectorSkew	lin	pow	lin	pow	lin	pow	lin	pow	exp	pow	pow	lin	exp	lin	pow	pow	log	pow	lin	exp	pow	pow	pow	log	log	log	lin	pow	pow	=	pow	lin	pow	lin	pow	lin	pow	pow	pow	log
EigenvectorStd	log	pow	log	pow	log	pow	pow	lin	pow	lin	pow	log	exp	log	lin	pow	pow	lin	log	log	lin	pow	pow	exp	lin	exp	pow	pow	pow	pow	=	log	pow	log	pow	pow	pow	pow	pow	lin
LoadKurtosis	exp	exp	pow	exp	pow	exp	pow	exp	exp	exp	exp	pow	exp	pow	log	exp	pow	log	pow	log	log	exp	pow	lin	pow	lin	lin	exp	exp	lin	exp	=	exp	pow	exp	lin	exp	exp	exp	pow
LoadMean	log	pow	log	=	log	pow	exp	pow	pow	pow	pow	exp	lin	exp	lin	pow	pow	lin	pow	lin	lin	pow	pow	log	lin	log	pow	pow	pow	pow	pow	log	=	log	pow	pow	pow	pow	pow	exp
LoadSkew	exp	exp	pow	exp	pow	exp	pow	exp	log	exp	exp	pow	log	pow	log	exp	pow	log	pow	log	log	exp	pow	lin	log	lin	exp	exp	exp	lin	exp	pow	exp	=	exp	exp	exp	exp	exp	pow
LoadStd	log	pow	log	pow	log	pow	pow	pow	pow	pow	pow	log	pow	log	lin	pow	pow	lin	pow	lin	lin	pow	pow	pow	lin	pow	pow	pow	pow	pow	pow	log	pow	log	=	pow	pow	pow	pow	exp
Nodes	log	pow	lin	pow	log	pow	exp	pow	pow	pow	pow	lin	exp	lin	log	pow	pow	log	lin	log	log	pow	exp	pow	log	pow	lin	lin	pow	lin	pow	lin	pow	log	pow	=	pow	pow	pow	lin
PercentageCenter	exp	pow	log	pow	log	pow	exp	pow	pow	pow	pow	log	pow	log	log	pow	log	log	log	log	log	pow	exp	log	log	log	pow	pow	pow	pow	pow	log	pow	log	pow	pow	=	pow	pow	lin
PercentagePeriphery	exp	pow	log	pow	log	pow	log	pow	pow	pow	pow	log	pow	log	pow	pow	exp	pow	log	pow	pow	pow	pow	log	lin	log	pow	pow	pow	pow	pow	log	pow	log	pow	pow	pow	=	pow	log
Radius	log	pow	log	pow	log	pow	pow	pow	pow	pow	pow	exp	lin	exp	lin	pow	pow	lin	pow	lin	lin	pow	pow	log	log	log	pow	pow	pow	pow	pow	log	pow	log	pow	pow	pow	pow	=	exp
Transitivity	exp	log	pow	log	pow	exp	pow	lin	log	lin	exp	pow	log	pow	pow	lin	log	pow	log	pow	pow	log	pow	exp	exp	exp	lin	exp	lin	exp	lin	pow	log	pow	log	lin	lin	exp	log	=

Figure 14: Best regression between each dependent and independent variable

First cluster to merge	Second cluster to merge	R2	No. of clusters after merging
{DegrM}	{Degr}	1	39
{LoadM}	{BetM}	1	38
{EigSt}	{Nod}	0.99996	37
{EccSt}	{EccM}	0.99996	36
{LoadSt}	{BetSt}	0.99994	35
{ClosSt}	{ClosM}	0.9999	34
{LoadSk}	{BetSk}	0.99946	33
{LoadK}	{BetK}	0.99939	32
{Degr;DegrM}	{CycB}	0.99932	31
{EigSt;Nod}	{Edg}	0.99693	30
{EigK}	{EigSk}	0.99577	29
{BetM;LoadM}	{APL}	0.99543	28
{Rad}	{Diam}	0.99244	27
{BetSk;LoadSk}	{BetK;LoadK}	0.98387	26
{APL;BetM;LoadM}	{Diam;Rad}	0.98369	25
{ClusSt}	{ClusM}	0.95901	24
{ClusK}	{ClusSk}	0.95211	23
{EccM;EccSt}	{ClosM;ClosSt}	0.95116	22
{APL;BetM;Diam;LoadM;Rad}	{Edg;EigSt;Nod}	0.9492	21
{APL;BetM;Diam;Edg;EigSt;LoadM;Nod;Rad}	{BetSt;LoadSt}	0.94445	20
{APL;BetM;BetSt;Diam;Edg;EigSt;LoadM;LoadSt;Nod;Rad}	{EigK;EigSk}	0.94103	19
{DegrK}	{DegrSk}	0.93924	18
{APL;BetM;BetSt;Diam;Edg;EigK;EigSk;EigSt;LoadM;LoadSt;Nod;Rad}	{EigM}	0.92155	17
{CycB;Degr;DegrM}	{DegrSt}	0.89571	16
{ClusM;ClusSt}	{Tran}	0.89176	15
{ClusM;ClusSt;Tran}	{ClusK;ClusSk}	0.85335	14
{APL;BetM;BetSt;Diam;Edg;EigK;EigM;EigSk;EigSt;LoadM;LoadSt;Nod;Rad}	{ClosM;ClosSt;EccM;EccSt}	0.78531	13
{APL;BetM;BetSt;ClosM;ClosSt;Diam;EccM;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadM;LoadSt;Nod;;Rad}	{BetK;BetSk;LoadK;LoadSk}	0.76249	12
{EccK}	{EccSk}	0.74204	11
{APL;BetK;BetM;BetSk;BetSt;ClosM;ClosSt;Diam;EccM;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;;Rad}	{PercP}	0.72566	10
{ClosSk}	{Asso}	0.70306	9
{ClusK;ClusM;ClusSk;ClusSt;Tran}	{CycB;Degr;DegrM;DegrSt}	0.64778	8
{APL;BetK;BetM;BetSk;BetSt;ClosM;ClosSt;Diam;EccM;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PercP;Rad}	{Asso;ClosSk}	0.54691	7
{APL;Asso;BetK;BetM;BetSk;BetSt;ClosM;ClosSk;ClosSt;Diam;EccM;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PercP;Rad}	{PercC}	0.54024	6
{APL;Asso;BetK;BetM;BetSk;BetSt;ClosM;ClosSk;ClosSt;Diam;EccM;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PePercP;Rad}	{DE}	0.51731	5
{APL;Asso;BetK;BetM;BetSk;BetSt;ClosM;ClosSk;ClosSt;DE;Diam;EccM;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PePercP;Rad}	{EccK;EccSk}	0.42695	4
{APL;Asso;BetK;BetM;BetSk;BetSt;ClosM;ClosSk;ClosSt;DE;Diam;EccK;EccM;EccSk;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PePercP;Rad}	{ClusK;ClusM;ClusSk;ClusSt;CycB;Degr;DegrM;DegrSt;Tran}	0.37687	3
{APL;Asso;BetK;BetM;BetSk;BetSt;ClosM;ClosSk;ClosSt;ClusK;ClusM;ClusSk;ClusSt;CycB;DE;Degr;DegrM;DegrSt;Diam;EccK;EccM;EccSk;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PePercP;Rad;Tran}	{ClosK}	0.3632	2
{APL;Asso;BetK;BetM;BetSk;BetSt;ClosK;ClosM;ClosSk;ClosSt;ClusK;ClusM;ClusSk;ClusSt;CycB;DE;Degr;DegrM;DegrSt;Diam;EccK;EccM;EccSk;EccSt;Edg;EigK;EigM;EigSk;EigSt;LoadK;LoadM;LoadSk;LoadSt;Nod;PePercP;Rad;Tran}	{DegrK;DegrSk}	0.34306	1

Figure 15: Metrics clusters merging based on the R^2 values

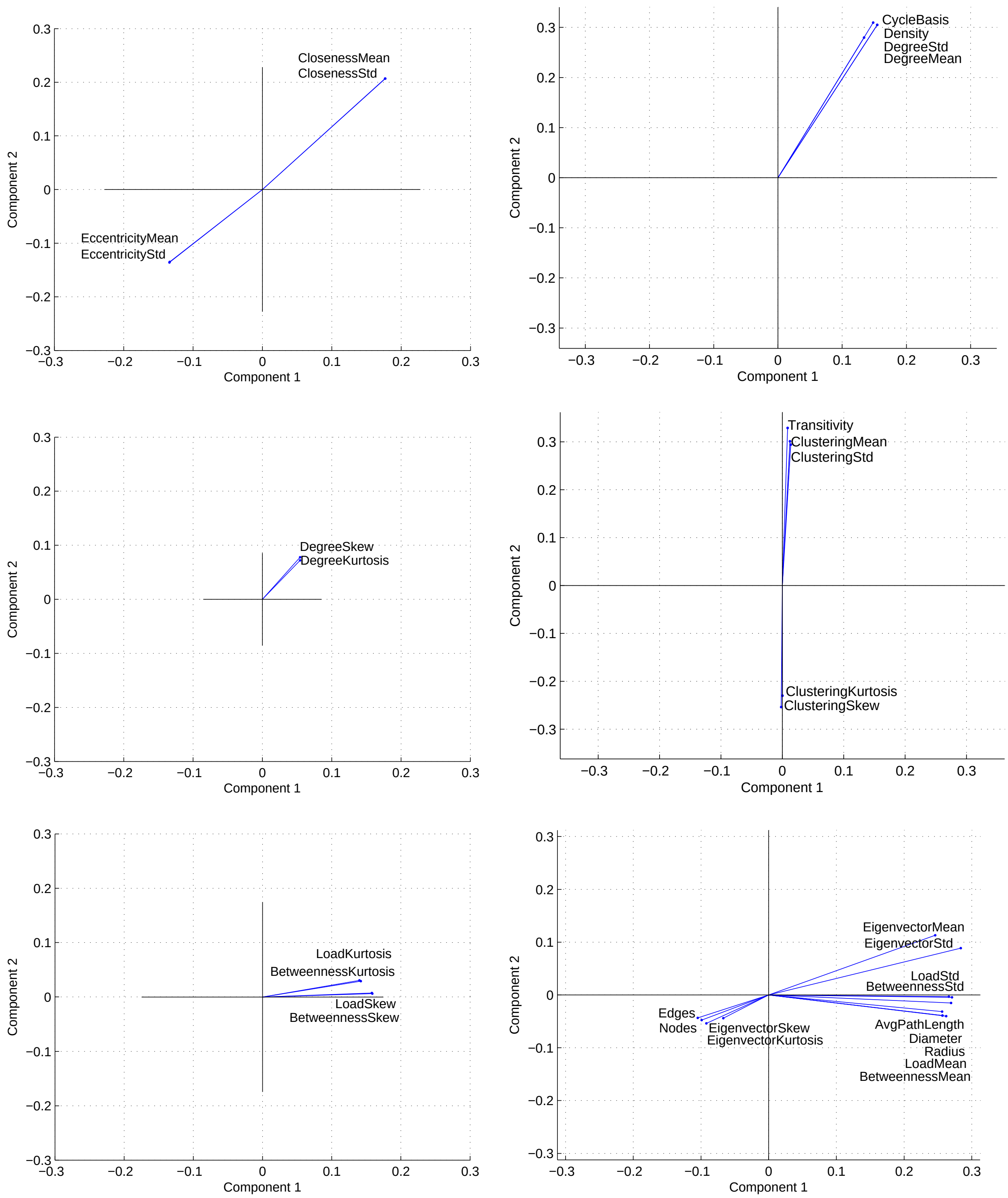


Figure 16: The first two component coefficients for the six *connected components*