NETWORKS AND HETEROGENEOUS MEDIA Website: http://aimSciences.org

Volume **3**, Number **2**, June **2008** pp. **345–359**

ON THE RELATIONSHIPS BETWEEN TOPOLOGICAL MEASURES IN REAL-WORLD NETWORKS

Almerima Jamakovic

Delft University of Technology
P.O.Box 5031, Delft, 2600 GA, The Netherlands

Steve Uhlig

Delft University of Technology
P.O.Box 5031, Delft, 2600 GA, The Netherlands

Abstract. Over the past several years, a number of measures have been introduced to characterize the topology of complex networks. We perform a statistical analysis of real data sets, representing the topology of different real-world networks. First, we show that some measures are either fully related to other topological measures or that they are significantly limited in the range of their possible values. Second, we observe that subsets of measures are highly correlated, indicating redundancy among them. Our study thus suggests that the set of commonly used measures is too extensive to concisely characterize the topology of complex networks. It also provides an important basis for classification and unification of a definite set of measures that would serve in future topological studies of complex networks.

1. **Introduction.** Complex network structures are common for a wide range of systems in nature and society [3, 16, 35]. Although complex systems are extremely different in their function, a proper knowledge of their topology is required to thoroughly understand and predict the overall system performance. For example, in computer networks, performance and scalability of protocols and applications, robustness to different types of perturbations (such as failures and attacks), all depend on the network topology. Consequently, network topology analysis, primarily aimed at non-trivial topological properties, has resulted in the definition of a variety of practically important measures, capable of quantitatively characterizing certain topological aspects of the studied systems [1, 4, 33]. The outcome, however, has a serious drawback that it does not ensure the mutual independence among the proposed measures. In this context, having an increasing number of measures complicates attempts to determine a definite measure set that would form the basis for analyzing any network topology [23].

In this paper we study the relationships between topological measures, with the aim of classifying a subset that would effectively characterize most real-world networks. The classification of measures in our study is based on statistical analysis methods. The presented methods reveal a clear relation between topological measures: a measure accounting for a certain network property seems to be strongly

2000 *Mathematics Subject Classification.* Primary: 05C75, 62H20; Secondary: 62H25 .
Key words and phrases. topological measures, correlation analysis, real-world networks.

associated with other measures that to our knowledge has not been previously reported as being trivial. It should be noted that the studied real-world networks stem from as various as possible domains so as to avoid correlations that are due to structural constraints of the systems under study. This paper thus establishes a path towards the identification of a definite set of topological measures that would serve in future network topology analysis.

The paper is organized as follows. Section 2 gives an overview of the current state of the related work. Section 3 describes topological measures and the considered data sets, representing the topology of various complex systems. Section 4 analyses the relationship between topological measures through three different statistical methods. Section 5 summarizes our main results on classification and unification of the set of topological measures.

2. **Related work.** We already mentioned several important works in the field of complex networks, focusing on the statistical mechanics of network topology and dynamics [1, 4, 33]. They all present and discuss the main complex network models and corresponding analytical tools, covering almost every known aspect of random graphs, small-world and scale-free networks as well as variations of those models. More relevant for our work are those papers that present the measures capable of characterizing topological properties of real-world networks. In several papers, among which [1, 4, 15, 33], the authors present an extensive survey of such measures. Furthermore, a vast majority of papers attempt to address the question of finding the most relevant topological properties by calculating a set of topological measures. Much of those research efforts, however, are posed within a particular research interest, resulting in a characterization of real-world networks from a specific domain. For example, in [24] and [26] the authors have calculated a set of measures for Internet AS- and router-level topologies. Besides the quantitative characterization, recently presented studies, e.g. [26, 27], have addressed the question of selecting a subset of measures that would effectively characterize most real-world networks: they find that joint degree distributions appear to fundamentally characterize Internet AS- and router-level topologies. Although there is a large literature on characterization of real-world networks, we are not aware of much work that attempted to study the correlation between topological measures in real-world networks. The work from [15] is also notable for using statistical analysis methods that give insight into the relationships between measures in the main complex network models. However, there is little understanding of the relationships among individual measures in real-world networks, in which we make a first fundamental step.

3. **Background.** In Subsection 3.1 we provide a set of topological measures, which is considered relevant in the networking literature [26]. In Subsection 3.2 we give data sets representing the topology of complex networks from a wide range of systems in nature and society.

3.1. **Topological measures of networks.** A graph theoretic approach is used to model the topology of a complex system as a network with a collection of nodes and a collection of links that connect pairs of nodes. A network is represented as an undirected graph $G = (\mathcal{N}, \mathcal{L})$ with $n = |\mathcal{N}|$ nodes and $m = |\mathcal{L}|$ links.

Basics

A network is connected if there exists a path between each pair of nodes. When there is no path between at least one pair of nodes, a network is said to be disconnected. A disconnected network consists of several independent components. We use the number of zero eigenvalues of the Laplacian matrix[1] to check the number of components[2] a network has. In the remainder of this paper, we only consider the networks formed by the largest connected component of our real-world networks. The computation of the topological measures is thus restricted to those largest connected components.

Degree

The node degree describes the number of neighbors a node has. The node degree distribution is the probability $\Pr(k)$ that a randomly selected node has a given degree k. The number of links that on average connect to a node is called the average node degree. The average node degree can be easily obtained from the degree distribution through $E[D] = \sum_{k=1}^{D_{\max}} k \Pr(k)$, where $D_{\max}$ is the maximum degree in a given graph.

The joint degree distribution $\Pr(k, k')$ is the probability that a randomly selected pair of nodes has degrees k and k'. A summary measure of the joint degree distribution is the average neighbor degree of nodes with a given degree k. Another summary statistics that quantifies the correlation between pairs of nodes is the assortativity coefficient r: assortative networks have $r > 0$ (disassortative, i.e. $r < 0$ resp.) and tend to have nodes that are connected to nodes with similar (dissimilar resp.) degree [32].

Distance

The distance distribution $\Pr(h)$ is the probability that the length of the shortest path between a random pair of nodes is h. From the distance distribution, the average node distance is derived as $E[H] = \sum_{h=1}^{h_{\max}} h \Pr(h)$, where $h_{\max}$ is the longest among the shortest paths between any pair of nodes. $h_{\max}$ is also referred to as the diameter of a graph. On the other hand, the eccentricity measures the largest distance between a node and any other node of a graph. The average node eccentricity is the average of eccentricities of all nodes. Obviously, the maximum eccentricity equals the diameter of a graph.

Clustering

The clustering coefficient $c_G(i)$ of a node i is the proportion of links between nodes within the neighborhood of a node i, divided by the maximum number of links that could possibly exist between those neighbors. For an undirected graph, a node i with degree d_i has at most $\frac{d_i(d_i-1)}{2}$ links among the nodes within its neighborhood. In other words, the clustering coefficient is the ratio between the number of triangles that contain node i and the number of triangles that could possibly exist if all neighbors of i were interconnected [34, 35]. The clustering coefficient for the entire graph is the average of clustering coefficients of all nodes.

The rich-club coefficient [11] is a recently introduced measure that quantifies how close subgraphs, spawned by the k largest-degree nodes, are to forming a clique. The

[1]The Laplacian matrix of a graph G with n nodes is an $n \times n$ matrix $Q = \Delta - A$ where $\Delta = diag(D_i)$, D_i is the nodal degree of node $i \in \mathcal{N}$ and A is the adjacency matrix of G [29].

[2]The multiplicity of 0 as an eigenvalue of the Laplacian matrix is equal to the number of components a network has.

rich-club coefficient ϕ is the ratio of the number of links in the subgraph induced by the k largest-degree nodes to the maximum possible links between them $(k(k-1)/2)$.

`Centrality`

Betweenness is a centrality measure of a node (link) within a graph: nodes (links) that occur on many shortest paths between other node pairs have higher node (link) betweenness than those that do not [18]. Average node (link) betweenness is the average value of the node (link) betweenness over all nodes (links).

`Coreness`

The k-core of a graph is a subgraph obtained from the original graph by the recursive removal of all nodes of degree less then or equal to k [7]. The node coreness of a given node is the maximum k such that this node is still present in the k-core but removed from the $(k+1)$-core. The average node coreness is the average value of the node coreness over all nodes.

`Connectivity`

The second smallest eigenvalue of the Laplacian matrix [17] is called the algebraic connectivity. The algebraic connectivity plays a special role in many graph theory related problems (for surveys see e.g. [10, 13, 14, 30]). The most important is its application to the overall connectivity of a graph: the larger the algebraic connectivity is, the more difficult it is to cut a graph into independent components. Two other connectivity measures are directly related to the algebraic connectivity: 1) the link connectivity is the minimal number of links whose removal would disconnect a graph, 2) the node connectivity is defined analogously (nodes together with adjacent links are removed). The latter two connectivity measures provide worst case bound on the robustness to node and link failures [17].

3.2. **Data sources of real-world networks.** We have mostly used publicly available data sets, representing the topology of complex networks from a wide range of systems in nature and society, i.e. technological, social, biological and linguistic. Technological systems we consider here include the following real-world networks:

- the Dutch road infrastructure [19];
- a European national railway infrastructure;
- a European Internet Service Provider (ISP);
- a European city area power grid;
- the western states power grid of the United States [34];
- the air transportation network representing the world wide airport connections, documented at the Bureau of Transportation Statistics database[3], and the connection between United States airports [12];
- the Internet network at the autonomous system [8] and the router level [9].

Social systems include the following real-world networks:

- the network representing soccer players association from the Dutch soccer team [20];
- the network representing actor appearance in movies [2];
- the network representing collaboration among scientists [31].

Biological systems include the following real-world networks:

- the network representing frequent associations between dolphins [25];

[3] http://www.bts.gov

- the network representing protein interaction of the yeast Saccharomyces cerevisae [11, 21].

Linguistic systems include the following real-world networks:

- the network representing common adjacencies among words in English, French and Spanish [28].

We provide in the Appendix a summary statistics of the topological measures for the considered real-world networks.

4. **Statistical analysis of topological measures.** In this section, we rely on statistical analysis methods to give insight into the relationships between measures in real-world networks. In Subsection 4.1, we relate pairs of topological measures by displaying their values as a collection of points, each having one coordinate on a horizontal and one on a vertical axis. In Subsection 4.2, we perform correlation analysis to find out which of the measures are redundant. Finally, in Subsection 4.3, we apply principal components analysis (PCA) to support the classification of correlated measures from Subsection 4.2.

4.1. **Visual comparison.** Many complex networks are characterized by a power-law node degree distribution and a relatively short path between any two nodes. However, some complex networks may lack both, the power-law as well as the small-world character. Among the considered data sets, networks representing the topology of various transportation systems and power-grids are those where the two characteristics were not entirely encountered. In Figure 1, we show the node degree distribution of networks that do not obey a power-law behavior.

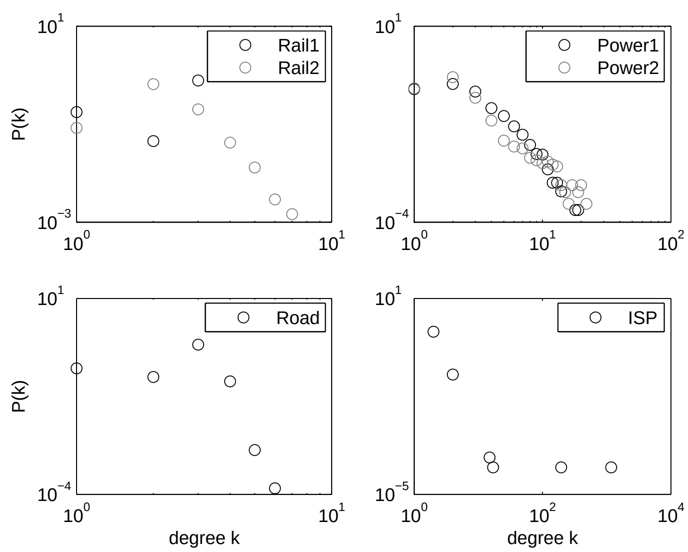

FIGURE 1. Real-world networks that do not obey a power-law degree distribution.

The average node degree is the coarsest characteristic of node interconnections. In complex networks the average node degree is typically small and independent of the network size n. In Figure 2 we show respectively the relationship between the link density and the number of nodes (and links) for various complex networks. As expected, for increasing n, the link density tends to zero and closely follows a power-law with exponent 1 (bottom of Figure 2). The link density is thus inversely proportional to the number of nodes while being inversely proportional to the square root of the number of links (top of Figure 2). From this it follows that the number of links is proportional to the number of nodes (not shown). Hence, in most complex networks, the classical assumption that $m = O(n)$ holds.

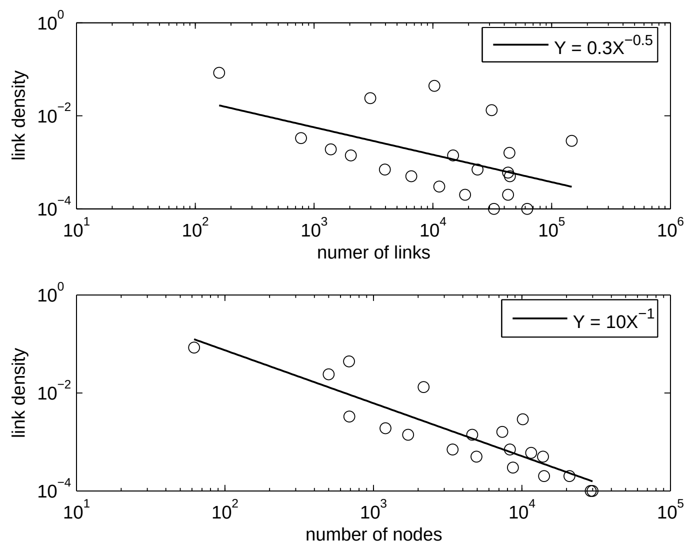

FIGURE 2. The link density as a function of the number of nodes and the number of links in real-world networks.

Node correlations play an important role in the characterization of the topology of complex networks. The most general approach to measure correlation among nodes is by means of the assortativity coefficient. On the top left scatter diagram in Figure 3 we show that disassortative networks, where high-degree nodes preferentially attach to other low-degree nodes, tend to be more clustered as their disassortativity increases. One should also notice from the ellipse on the top left scatter diagram in Figure 3 that networks are typically assortative while having almost no clustering. The latter group of networks is made of various transportation and power-grid infrastructures. In addition, we observe that assortative networks, on average, have larger distances between pairs of nodes. The relationship between the assortativity coefficient and the average node distance is shown in the upper right scatter diagram of Figure 3.

Another interesting result that confirms these conclusions is found in [15]. Here, for the scale-free graph of Barabasi-Albert [2], a negative correlation is observed between the assortativity coefficient and the clustering coefficient and a possitive between the assortativity coefficient and the average node distance. Consequently, a strong negative correlation is observed between the clustering coefficient and the average node distance. A posible explanation is that growth and preferential attachment, i.e. the two basis rules upon which the model is based, are responsible for additional links that tend to be established with the hubs, creating a more connected core and therefore contributing to higher clustering and smaller shortest paths.

Recently, it has been shown that complex networks are also characterized by the so called rich-club phenomenon [11]. The average distance between pairs of nodes as a function of the rich-club coefficient (lower left scatter diagram of Figure 3) yields that networks with smaller distance are much more likely to have high-degree nodes that form tight and well-interconnected subgraphs. As a result, one might expect that for disassortative networks, having on average smaller distance between pairs of nodes, the rich-club phenomenon would be evident as well. Nevertheless, on the lower right scatter diagram of Figure 3, we observe that the rich-club phenomenon is not trivially related to the mixing properties of networks. In other words, the rich-club phenomenon and the mixing properties express different features that are not trivially related or derived from each other.

On the other hand, topological measures associated with a certain feature, such as the shortest path length, are clearly related to each other. For example, average node betweenness increases as a function of average distance between pairs of nodes, verifying that networks that have high average distance will also have nodes that occur on many shortest paths between other node pairs, and consequently, on average, have higher node betweenness. The Internet Service Provider network is a good example of a network for which high average distance between pairs of nodes results in high average node betweenness (see summary statistics presented in the Appendix).

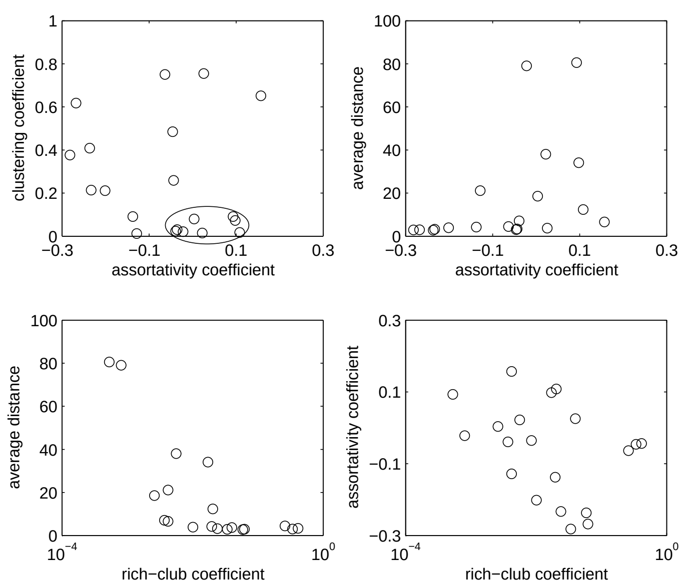

FIGURE 3. The relationship among topological measures for various real-world networks: clustering coefficient, assortativity coefficient, rich-club coefficient and average distance.

An important topological property, often ignored in the analysis of complex networks, is coreness. Node coreness refers to the degree of closeness of each node to

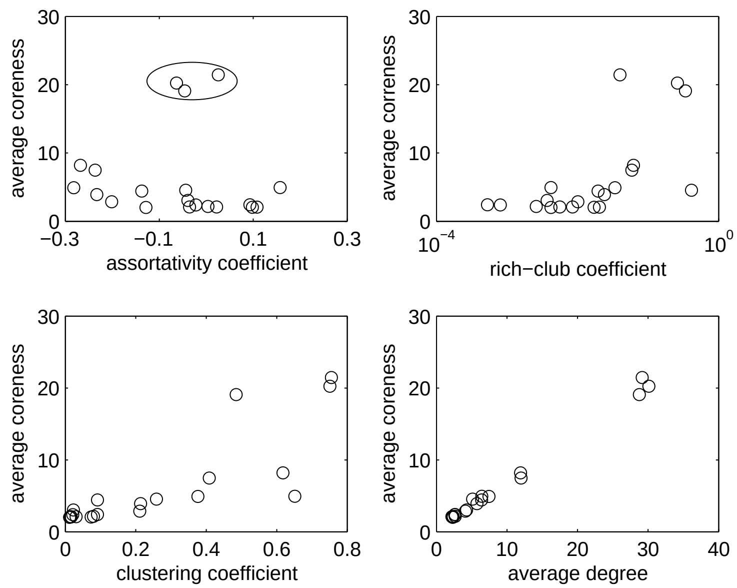

FIGURE 4. The relationship among topological measures for various real-world networks: clustering coefficient, assortativity coefficient, rich-club coefficient, average coreness and average degree.

a core of densely connected nodes, observable in the network [7]. In Figure 4 we report the relationship between average node coreness and the previously identified measures. The average node coreness as a function of the assortativity coefficient yields that social networks do not follow the generally observed trend of networks being disassortative but having, on average, higher node coreness. All three social networks are shown within an ellipse on the top left scatter diagram of Figure 4. At the same time, networks with higher average node coreness are more likely to have higher rich-club and clustering. Finally, we observe that the average node coreness is directly related to the average node degree. The former relationships are not surprising since on average, higher average node degree means higher rich-club and clustering, both for which we already perceived higher coreness.

Robustness to node and link failures is well captured by the algebraic connectivity. In essence, the algebraic connectivity quantifies the extent to which a network can accommodate an increasing number of node- and link-disjoint paths. Figure 5 shows the relationships between the algebraic connectivity and the previously identified measures. The algebraic connectivity increases with the average node degree, as networks with higher average degree are better connected and consequently, are likely to be more robust. Note that in the literature [32] it is shown that assortative networks are less vulnerable to both random failures and targeted attacks. Here, we observe that disassortative networks have larger algebraic connectivity. This is not in contradiction with the observed tendency because it is most likely to be related to the hardness to cut the graph into independent components. Moreover, the larger the algebraic connectivity, the more networks seem to have a large rich-club and hierarchical nature. This implies that they have more well-interconnected and centrally-oriented nodes that occur on many shortest paths. Still, the average node betweenness does not seem to be related to the overall connectivity of a graph.

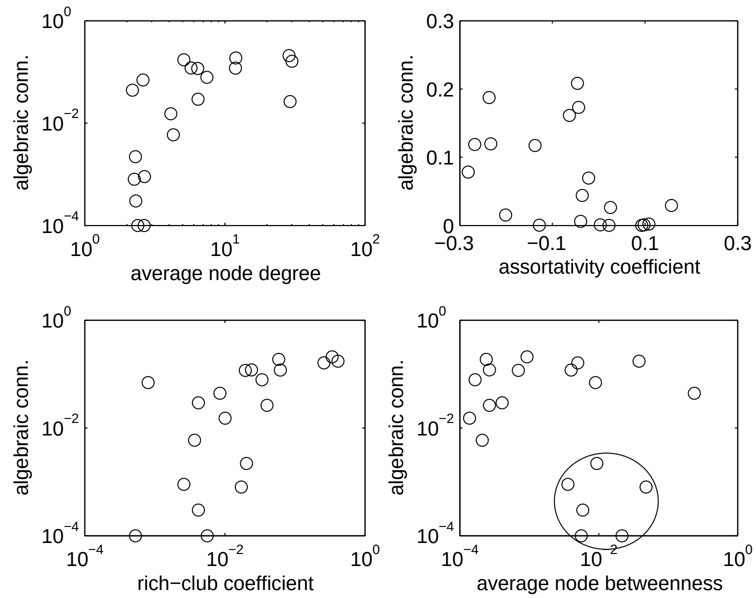

FIGURE 5. The relationship among topological measures for various real-world networks: assortativity coefficient, rich-club coefficient, algebraic connectivity, average node betweenness and average node degree.

4.2. **Correlation analysis.** Correlation analysis aims at finding out linear relationships between variables. Variables are in our case the topological measures. From the tables presented in the Appendix we derive a matrix whose columns are the different measures and the rows are the different real-world networks, denoted by $\mathbf{X}$. We then compute the correlation matrix of $\mathbf{X}$, denoted by $\mathbf{C}$. Matrix $\mathbf{C}$ is symmetric and has 1's elements on the diagonal. Each element (i, j) of $\mathbf{C}$ gives the correlation coefficient between measures i and j (rows i and j of $\mathbf{X}$). The correlation coefficient c varies between -1 and 1, and indicates whether the two variables a linearly correlated: positively if $c \sim 1$, negatively if $c \sim -1$, and uncorrelated if $c \sim 0$.

We are not interested in whether the correlation between two measures is positive or negative, but only how strongly two given measures are numerically related to each other. To ease the visualization, we show on Table 1 a symbolic encoding version of the correlation matrix. Table 1 displays the lower diagonal of the correlation matrix, using the following range of values and coding characters:

- $0 \leq |c| \leq 0.3$: “ ” (no correlation);
- $0.3 \leq |c| \leq 0.6$: “.” (mild correlation);
- $0.6 \leq |c| \leq 0.9$: “+” (significant correlation);
- $0.9 \leq |c| \leq 1$: “#” (strong correlation).

The measures on Table 1 are identified by their number at the top of each column, and by the name and number on the left of each row. As the correlation matrix is symmetric, we show only the lower diagonal. First to be noticed is that 58 among the 91 lower diagonal elements (not counting the diagonal) have a correlation coefficient less than 0.3 in absolute value. Most measures are thus weakly correlated, indicating that most of them indeed reveal different topological aspects of real-world networks. 21 among the 91 lower diagonal elements correspond to mild correlations, i.e. $0.3 \leq |c| \leq 0.6$. Only 12 among the 91 lower diagonal elements correspond to strong correlations. Based on existing correlations between measures, we can identify the following clusters (see also Figure 6):

- **Distance cluster**: average node distance, average node eccentricity, average node and link betweenness.
- **Degree cluster**: average degree, average node coreness and clustering coefficient.
- **Intra-connectedness cluster**: link density, rich-club coefficient and algebraic connectivity.
- **Inter-connectedness cluster**: average neighbor degree and assortativity coefficient.

Topological measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Number of nodes (1)	1													
Number of links (2)	.	1												
Link density (3)	.		1											
Average degree (4)		.		1										
Average neighbour degree (5)					1									
Assortativity coefficient (6)					+	1								
Rich-club coefficient (7)	.		+	.			1							
Clustering coefficient (8)		.	.	+			.	1						
Average node distance (9)	.								1					
Average node eccentricity (10)	.								#	1				
Average node coreness (11)		.		#			.	+			1			
Average node betweenness (12)	.								#	#		1		
Average link betweenness (13)	.								#	#		#	1	
Algebraic connectivity (14)	.		.	.	.	.	+	.			.			1

TABLE 1. Correlation between topological measures for 20 various real-world networks.

We labeled different measure clusters according to the type of topological information the group of measures provides. Intra- and inter- connectedness refer to the measures characterizing the observed connectivity, respectively, within and between a (sub)set of nodes in the network. All measures within each cluster are highly or partly topologically redundant. The 14 initial measures can thus be reduced to 6 (including the number of nodes and the number of links) since 8 of them are redundant with those of the same cluster. Besides the strength of the correlations within the groups, the correlation analysis shows to what extent some measures capture several topological properties of a network at once. For example, the number of nodes and the algebraic connectivity, both exhibit mild correlation to 8 other measures. The number of nodes is related to the number of links and all measures within the distance and the intra-connectedness clusters, while not related to measures within the degree or the inter-connectedness clusters. The algebraic connectivity, on the other hand, is related to all measures within the degree, intra-connectedness, and the inter-connectedness clusters, but not to any measure in the distance cluster.

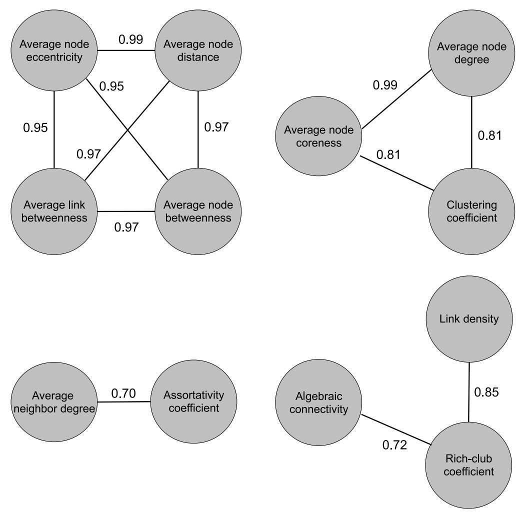

FIGURE 6. A graph in which nodes are topological measures and links the correlations that emerged from the correlation analysis. The corresponding values display the strength of the correlation between pairs of measures.

4.3. **Principal component analysis.** Correlation analysis measures the strength of correlation between variables. Understanding correlations, however, does not give insight about the number of independent variables, possibly derived from the set of correlated variables. In this context, correlated variables are the topological measures. Principal component analysis (PCA) [22] has proven to be useful for reducing the number of variables (dimensionality) while retaining most of the original variability in the data. The number of transformed, uncorrelated variables are called principal components, which in decreasing order account for as much of the variability in the data as possible.

We denote a given data set as a matrix $\mathbf{X}$ whose p columns are the variables to be analyzed $X_i, i = 1, \ldots, p$. Each column (variable) has n elements, hence $\mathbf{X}$ is a $n \times p$ matrix. PCA performs a rotation of this matrix $\mathbf{X}$ such that

$$\mathbf{Y} = \mathbf{A}'\mathbf{X}' \tag{1}$$

where $\mathbf{A}'$ is an orthogonal matrix[4]. $\mathbf{Y}$ is the matrix of the rotated data, it is a square matrix of order n. $\mathbf{A}$ is found by constraining the covariance matrix of $\mathbf{Y}$, $\mathbf{C}_Y = \frac{1}{n-1}\mathbf{Y}\mathbf{Y}'$, to be diagonalized. A symmetric matrix can be diagonalized by the orthogonal matrix of its eigenvectors so that

$$\mathbf{C}_Y = \frac{1}{n-1}\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}' \tag{2}$$

where $\boldsymbol{\Lambda} = \mathbf{X}\mathbf{X}'$. $\mathbf{A}$ is selected so that its columns are the eigenvectors of $\boldsymbol{\Lambda}$ and the principal components of $\mathbf{X}$. The diagonal elements of $\mathbf{C}_Y$ give the variance of $\mathbf{X}$ along each principal component.

[4]A matrix is orthogonal if $\mathbf{A}'\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

The objective of PCA is to provide information about the minimal dimensionality, necessary to describe the data variability. The percentage of the total data set variance that is captured by a given number of principal components, is presented in Figure 7. The first principal component alone captures 76%, the first two components 94% and the first three components more than 99% of the total data set variance. PCA analysis shows that only 3 dimensions are enough to retain most of the original variability in the data. This, however, does not imply that the measures that are not important for the main principal components are unnecessary, but rather that they provide very specific topological information.

The reason why PCA was able to drastically reduce the dimensionality of the data set is because the principal components are a linear combination of all the measures. The first principal component is composed of two measures, i.e. the number of links and the number of nodes. All other measures have a very small weight in the linear combination of this principal component. In fact, the first principal component's measures are those missing from the four clusters we identified in the correlation analysis, presented in Subsection 4.2. The second principal component, besides the average node distance and the average node eccentricity, is also mostly made of the number of links and number of nodes. The third principal component is similar to the second in terms of which measures have the largest weights, but the sign of the weights differs as the principal components form an orthogonal basis. The fourth principal component, that captures a very small fraction of the total variance, is made almost exclusively from the average neighbor degree. PCA reveals that important measures that characterize the variations in the topological measures are the number of nodes and links and the measures within the distance and interconnectedness clusters. measures within the degree and intra-connectedness clusters are redundant with the number of nodes and the number of links, since both the average degree and the link density can be recovered from the former measures.

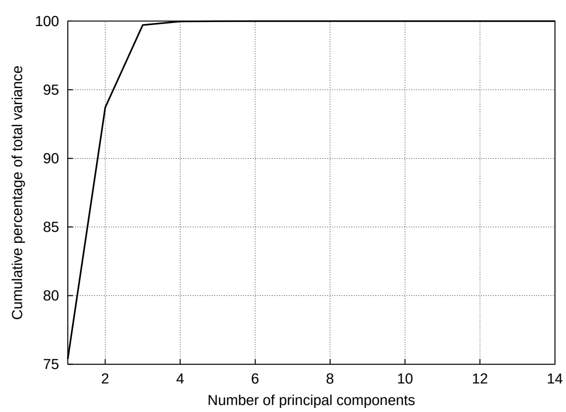

FIGURE 7. Fraction of the variance captured by the principal components.

5. **Discussions and conclusion.** In this paper, we have studied the relationships between topological measures of real-world networks. The visual analysis, presented in Subsection 4.1, revealed the following relationships among topological measures:

- The clustering coefficient increases with the increasing disassortativity. For assortative networks this relation is not trivial.

- The average node distance increases with the increasing assortativity coefficient and decreases with the increasing rich-club coefficient. Consequently, the assortativity coefficient decreases with the increasing rich-club coefficient.
- The average node coreness increases with the increasing rich-club and clustering coefficient while it decreases with the increasing assortativity coefficient. Furthermore, it is directly related to the average node degree.
- The algebraic connectivity increases with the increasing average node degree and the rich-club coefficient while it decreases with the increasing assortativity coefficient. The algebraic connectivity is not related to the average node betweenness.

The correlation analysis, presented in Subsection 4.2, resulted in several highly-correlated clusters with the following topological measures:

- Distance cluster: 1) the average node distance is strongly related to the average node eccentricity, 2) the average node (link) betweenness to the average node distance and hence 3) the average node (link) betweenness to the average node eccentricity;
- Degree cluster: 1) the average node degree is strongly related to the average node coreness and 2) the average node coreness to the clustering coefficient;
- Intra-connectedness cluster: 1) the rich-club coefficient is strongly related to the link density and 2) the algebraic connectivity to the rich-club coefficient;
- Inter-connectedness cluster: 1) the assortativity coefficient is strongly related to the average neighbor degree.

Our work showed that some topological measures tend to be more correlated than others. This observation implies redundancy between topological measures. Consequently, we have identified a significantly smaller set of topological measures that is able to characterize real-world network's structures.

Acknowledgements. This research was supported by the Netherlands Organization for Scientific Research (NWO) and the Next Generation Infrastructures programme (www.nginfra.nl). We are grateful to all persons who made possible the collection of real-world data used in this paper.

REFERENCES

[1] R. Albert and A.-L. Barabasi, *Statistical mechanics of complex networks*, Reviews of Modern Physics, **74** (2002), 47–94.

[2] A.-L. Barabasi and R. Albert, *Emergence of scaling in random networks*, Science, **286** (1999), 509–512.

[3] A.-L. Barabasi, "Linked: The New Science Of Networks," Perseus Books Group, 2002.

[4] S. Boccalettia, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwanga, *Complex networks: structure and dynamics*, Physics Reports, **424** (2006), 175–308.

[5] B. Bollobás, "Random Graphs," Cambridge University Press, 2001.

[6] B. Bollobás and O. Riordan, "Handbook Of Graphs and Networks: From the Genome to the Internet," Wiley-VCH, 2003.

[7] S. P. Borgatti and M. G. Everett, *Models of core/periphery structures*, Social Networks, **21** (1999), 375–395.

[8] CAIDA, Macroscopic Topology AS-Adjacencies Data Set, *http://www.caida.org/tools/measurement/skitter/as_adjacencies.xml*.

[9] CAIDA, Router-Level topology Measurements, *http://www.caida.org/tools/measurement/skitter/router_topology*.

[10] F. R. K. Chung, "Spectral Graph Theory," American Mathematical Society, 1997.

[11] V. Colizza, A. Flammini, M. A. Serrano and A. Vespignani, *Detecting rich-club ordering in complex networks*, Nature Physics, **2** (2006), 110–115.

[12] V. Colizza, R. Pastor-Satorras and A. Vespignani, *Reaction-diffusion processes and metapopulation models in heterogeneous networks*, Nature Physics, **3** (2007), 276–282.
[13] D. M. Cvetković, M. Doob and H. Sachs, "Spectra of Graphs, Theory and Applications," Johann Ambrosius Barth Verlag, 1995.
[14] D. M. Cvetković, M. Doob, I. Gutman and A. Torgašev, "Recent Results in the Theory of Graph Spectra," North-Holland, 1988.
[15] L. da F. Costa, F.A. Rodrigues, G. Travieso and P.R. Villas Boas, *Characterization of complex networks: A survey of measurements*, Advances of Physics, **56** (2007), 167–242.
[16] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of Networks: From Biological Nets to the Internet and WWW," Oxford University Press, 2003.
[17] M. Fiedler, *Algebraic connectivity of graphs*, Czechoslovak Mathematical Journal, **23** (1973), 298–305.
[18] P. Holme, B. J. Kim, C. N. Yoon and S. K. Han, *Attack vulnerability of complex networks*, Physical Review E, **65** (2002), 056109.
[19] A. Jamakovic, H. Wang and P. Van Mieghem, "Topological Characteristics of the Dutch Road Infrastructure," Seminar Infrastructure Reliability, Delft, 2006.
[20] A. Jamakovic, R. E. Kooij, P. Van Mieghem and E. R. Van Dam, *Robustness of networks against viruses: The role of the spectral radius*, Proceedings of the 13th Annual Symposium of the IEEE/CVT Benelux, Liège, 2006.
[21] H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai, *Centrality and lethality of protein networks*, Nature, **411** (2001), 41–42.
[22] I. T. Jolliffe, "Principal Component Analysis," Springer-Verlag, 2002.
[23] D. Krioukov, F. Chung, K. Claffy, M. Fomenkov, A. Vespignani and W. Willinger, *The workshop on Internet topology (WIT) report*, ACM SIGCOMM Computer Communication Review, **37** (2007), 69–73.
[24] L. Li, D. Alderson, W. Willinger and J. Doyle, *A first-principle approach to understanding the Internet's router-level topology*, Proceedings of ACM SIGCOMM Conference, Portland, 2004.
[25] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten and S. M. Dawsen, *The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait?*, Behavioral Ecology and Sociobiology, **54** (2003), 396–405.
[26] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. Claffy and A. Vahdat, *The Internet AS-level topology: Three data sources and one definitive metric*, ACM SIGCOMM Computer Communication Review, **36** (2006), 17–26.
[27] P. Mahadevan, D. Krioukov, K. Fall and A. Vahdat, "Systematic Topology Analysis and Generation Using Degree Correlations," Proceedings of ACM SIGCOMM Conference, Pisa, 2006.
[28] R. Milo, S. Itzkovic, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon, *Superfamilies of evolved and designed networks*, Science, **303** (2004), 1538–1542.
[29] B. Mohar, *Laplace eigenvalues of graphs: A survey*, Discrete Mathematics, **109** (1992), 171–183.
[30] B. Mohar, Y. Alavi, G. Chartrand, O. R. Oellermann and A. J. Schwenk, *The Laplacian spectrum of graphs*, Graph Theory, Combinatorics and Applications, **2** (1991), 871–898.
[31] M. E. J. Newman, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA, **98** (2001), 404–409.
[32] M. E. J. Newman, *Assortative mixing in networks*, Physical Review Letters, **89** (2002), 208701.
[33] M. E. J. Newman, *The structure and function of complex networks*, SIAM Review, **45** (2002), 167–256.
[34] D. J. Watts and S. H. Strogatz, *Collective dynamics of small-world networks*, Nature, **393** (1998), 440–442.
[35] D. J. Watts, "Small-Worlds. The Dynamics of Networks Between Order and Randomness," Princeton Univerisity Press, 1999.

Received September 2007; revised November 2007.

E-mail address: a.jamakovic@ewi.tudelft.nl
E-mail address: s.p.w.g.uhlig@ewi.tudelft.nl

Appendix: Summary statistics of topological measures for various real-world networks

Topological measures	Road	Rail1	Rail2	Air1	Air2
Number of nodes	14098	8710	689	500	2179
Number of links	18687	11332	778	2980	31326
Link density	0,0002	0,0003	0,0033	0,0239	0,0132
Average degree	2,7	2,6	2,3	11,9	28,8
Average neighbor degree	2,9	2,8	2,5	53,8	140,5
Assortativity coefficient	0,093	-0,022	0,098	-0,268	-0,046
Rich-club coefficient	0,0005	0,0008	0,0172	0,0621	0,3395
Clustering coefficient	0,0912	0,0212	0,0731	0,6175	0,4849
Average node distance	80,6	79,0	34,1	2,9	3,0
Average node eccentricity	177,4	158,6	65,0	5,2	5,9
Average node coreness	2,4	2,4	2,0	8,2	19,1
Average node betweenness	0,0560	0,0090	0,0481	0,0040	0,0090
Average link betweenness	0,00220	0,00350	0,02190	0,00050	0,00005
Algebraic connectivity	0,0001	0,0695	0,0008	0,1186	0,2082

Topological measures	Power1	Power2	Power3	Power4	ISP
Number of nodes	4940	3419	1713	1205	29902
Number of links	6594	3953	2043	1385	32707
Link density	0,0005	0,0007	0,0014	0,0019	0,0001
Average degree	2,7	2,3	2,4	2,3	2,2
Average neighbor degree	3,9	3,8	2,9	3,1	45,7
Assortativity coefficient	0,004	-0,128	0,022	0,108	-0,036
Rich-club coefficient	0,0026	0,0042	0,0056	0,0204	0,0085
Clustering coefficient	0,0801	0,0120	0,0145	0,0171	0,0306
Average node distance	18,5	21,1	38,0	12,3	7109,9
Average node eccentricity	34,1	38,9	71,8	22,6	14250,0
Average node coreness	2,2	2,0	2,1	2,1	2,1
Average node betweenness	0,0036	0,0059	0,0216	0,0094	0,2377
Average link betweenness	0,00140	0,00270	0,00930	0,00450	0,10870
Algebraic connectivity	0,0009	0,0003	0,0001	0,0022	0,0440

Topological measures	AS-level	Router	Protein	Soccer	Dolphins
Number of nodes	20906	29064	4626	685	62
Number of links	42994	62260	14801	10310	159
Link density	0,0002	0,0001	0,0014	0,0440	0,0841
Average degree	4,1	4,3	6,4	30,1	5,1
Average neighbor degree	230,9	21,0	24,2	45,0	6,8
Assortativity coefficient	-0,201	-0,039	-0,137	-0,063	-0,044
Rich-club coefficient	0,0101	0,0037	0,0196	0,2605	0,4127
Clustering coefficient	0,2114	0,0232	0,0912	0,7507	0,2589
Average node distance	3,9	7,1	4,2	4,5	3,4
Average node eccentricity	8,0	14,7	8,1	8,6	6,5
Average node coreness	2,9	3,0	4,4	20,2	4,5
Average node betweenness	0,0001	0,0002	0,0007	0,0050	0,0380
Average link betweenness	0,00005	0,00006	0,00014	0,00022	0,01060
Algebraic connectivity	0,0152	0,0059	0,1173	0,1612	0,1730

Topological measures	Actor	Scientific	English	French	Spanish
Number of nodes	10143	13861	7377	8308	11558
Number of links	147907	44619	44205	23832	43050
Link density	0,0029	0,0005	0,0016	0,0007	0,0006
Average degree	29,2	6,4	11,9	5,7	7,4
Average neighbor degree	83,6	13,5	320,7	218,0	457,6
Assortativity coefficient	0,026	0,157	-0,237	-0,233	-0,282
Rich-club coefficient	0,0399	0,0042	0,0588	0,0240	0,0340
Clustering coefficient	0,7551	0,6514	0,4085	0,2138	0,3764
Average node distance	3,7	6,6	2,8	3,2	2,9
Average node eccentricity	9,6	12,4	5,6	6,7	7,6
Average node coreness	21,4	4,9	7,5	3,9	4,9
Average node betweenness	0,0003	0,0004	0,0002	0,0003	0,0002
Average link betweenness	0,00040	0,00007	0,00003	0,00007	0,00003
Algebraic connectivity	0,0004	0,0292	0,1875	0,1197	0,0782