On the relationships between topological metrics in real-world networks

A.Jamakovic, S.Uhlig and I.Theisler

Delft University of Technology
Electrical Engineering, Mathematics and Computer Science
P.O. Box 5031, 2600 GA Delft, The Netherlands
{A.Jamakovic, S.P.W.G.Uhlig, I.Theisler}@ewi.tudelft.nl

Summary. Over the past several years, a number of metrics 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 metrics are either fully related to other topological metrics or significantly limited in the range of their possible values. Second, we observe that subsets of metrics are highly correlated, indicating redundancy among them. Our study thus suggests that the set of commonly used metrics 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 metrics 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, 15, 32]. 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 aiming at non-trivial topological properties, has resulted in the definition of a variety of practically important metrics, capable to quantitatively characterize certain topological aspects of the studied systems [1, 4, 30]. The outcome, however, has a serious drawback that it does not ensure the mutual dependence among the proposed metrics: given that new metrics are usually principally intended at characterizing the systems under study, some metrics either fully capture others or, at least, significantly limit the range of their possible values. In this context, having an increasing number of metrics complicates attempts to determine a definite metric-set that would form the basis for analyzing any network topology [22].

In this paper we study the relationships between topological metrics, with the aim of classifying a subset that would effectively characterize most real-world networks. The classification of metrics in our study is based on statistical analysis methods. The presented methods reveal a clear relation between topological metrics: a metric accounting for a certain network property seems to be strongly associated with other metrics that to our knowledge has not been previously reported as being trivial. This study thus establishes a path towards the identification of a definite set of topological metrics that would serve in future network topology analysis.

The paper is organized as follows. Section 2 describes topological metrics and the considered data sets, representing the topology of various complex systems. Section 3 analyses the relationship between topological metrics through three different statistical methods. Section 4 summarizes our main results on classification of the set of topological metrics.

2 Background

2.1 Topological metrics of networks

In this section we provide a set of topological metrics, which is considered relevant in the networking literature [24]. 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 metrics is thus restricted to those largest connected components.

`Degree`

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 metric 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 [29].

`Distance`

The distance distribution $\Pr(h)$ is the probability that the length of the shortest path (hopcount) 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 largest hopcount between any pair of nodes. $h_{\max}$ is also referred to as the diameter of a graph. On the other hand, 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 [26].

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

eccentricity measures the longest path between a random pair of nodes. The average node eccentricity is the average of eccentricities of all pairs of nodes. Obviously, the maximum eccentricity equals $h_{\max}$.

`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 [31, 32]. 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 metric that quantifies how close subgraphs, spawned by the k largest-degree nodes, are to forming a clique. The 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 [17]. 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 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.

`Robustness`

The second smallest eigenvalue of the Laplacian matrix [16] 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, 27]). The most important is its application to the robustness of a graph: the larger the algebraic connectivity is, the more difficult it is to cut a graph into independent components. Two other connectivity metrics 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 metrics provide worst case robustness to node and link failures [16].

2.2 Data sources of real-world networks

We mostly have 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 system we consider here include the following real-world networks:

- the Dutch road infrastructure [18];

- a European national railway infrastructure;
- a European Internet Service Provider;
- a European city area power grid;
- the western states power grid of the United States [31];
- the air transportation network representing the world wide airport connections, documented at the Bureau of Transportation Statistics (http://www.bts.gov) database, and the connection between United States airports [12];
- the Internet network at the autonomous [8] and the router [9] level.

Social systems include the following real-world networks:

- the network representing soccer players association to Dutch soccer team [19];
- the network representing actor appearance in movies [2];
- the network representing collaboration among scientists [28].

Biological systems include the following real-world networks:

- the network representing frequent associations between dolphins [23];
- the network representing protein interaction of the yeast Saccharomyces cerevisae [11, 20].

Linguistic systems include the following real-world networks:

- the network representing common adjacencies among words in English, French and Spanish [25].

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

3 Statistical analysis of topological metrics

In this section, we rely on statistical analysis methods to give insight on the relationships between metrics in real-world networks. In the first Subsection 3.1, we relate pairs of topological metrics by displaying their values as a collection of points, each having one coordinate on a horizontal and one on a vertical axis. In the second Subsection 3.2, we perform correlation analysis to find out which of the metrics are redundant. In the third and final Subsection 3.3, we apply principal components analysis (PCA) to support the classification of subsets of metrics that are highly correlated.

3.1 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.

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

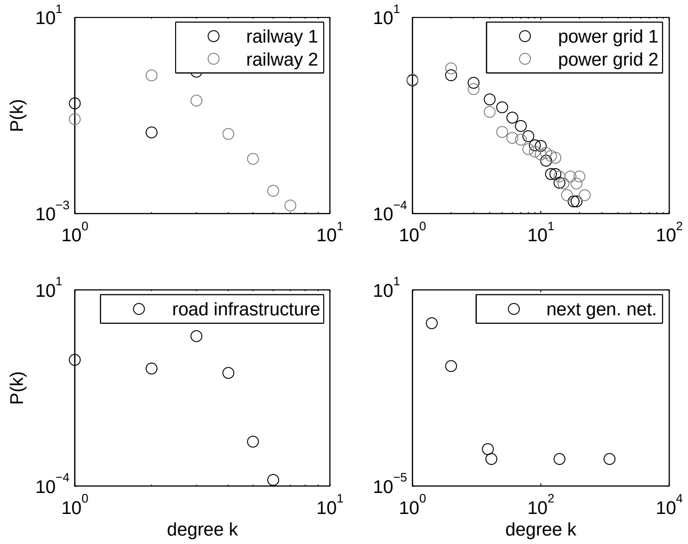

Fig. 1. Real-world networks that do not obey a power-law degree distribution.

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.

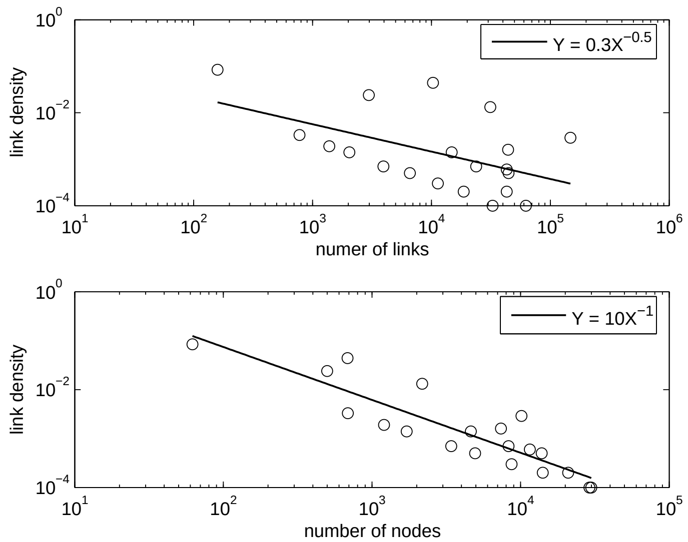

Fig. 2. The link density as a function of the number of nodes and the number of links in real-world networks.

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 high-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.

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 show 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 metrics associated with a certain feature, such as the shortest path length, are clearly related to each other. For example, average distance between pairs of nodes increases as a function of average node betweenness, verifying that networks that have many shortest paths between pairs of nodes, on average, have higher node betweenness and distance. 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).

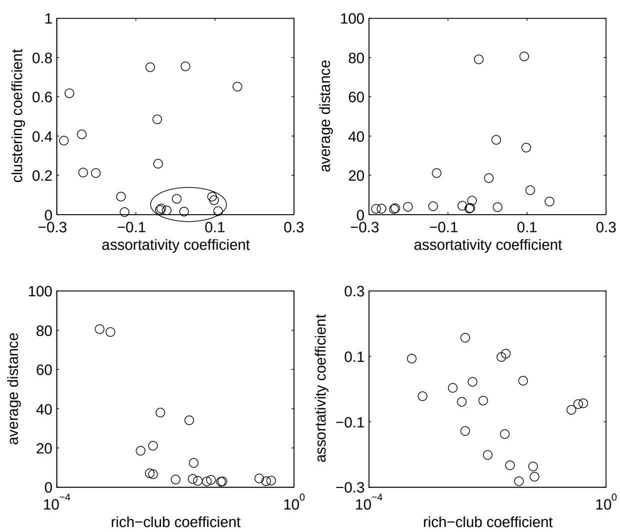

Fig. 3. The relationship among topological metrics for various real-world networks

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 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 metrics. 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

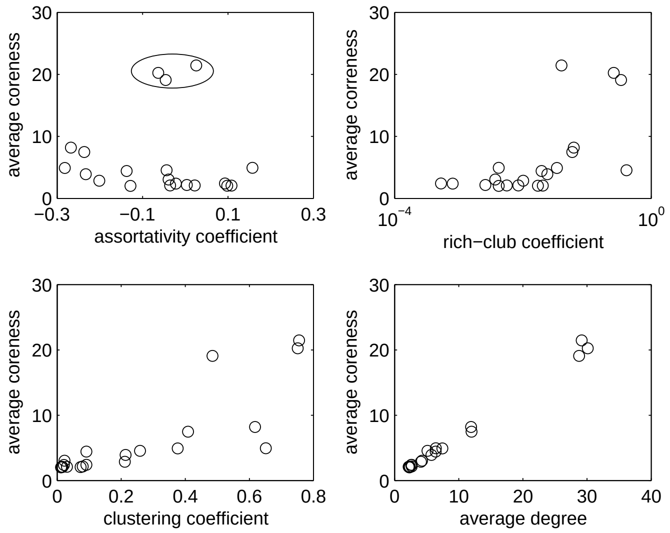

Fig. 4. The relationship among topological metrics for various real-world networks

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 metrics. 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. Contrary to the literature [29] where it is shown that assortative networks are less vulnerable to both random failures and targeted attacks, we find the opposite tendency. We observe that disassortative networks have larger algebraic connectivity. The previous observation 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 robustness of a graph.

3.2 Correlation analysis

Correlation analysis aims at finding out linear relationships between variables. Variables are in our case the topological metrics. From Figures presented in the Appendix we derive a matrix whose columns are the different metrics 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 metrics 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

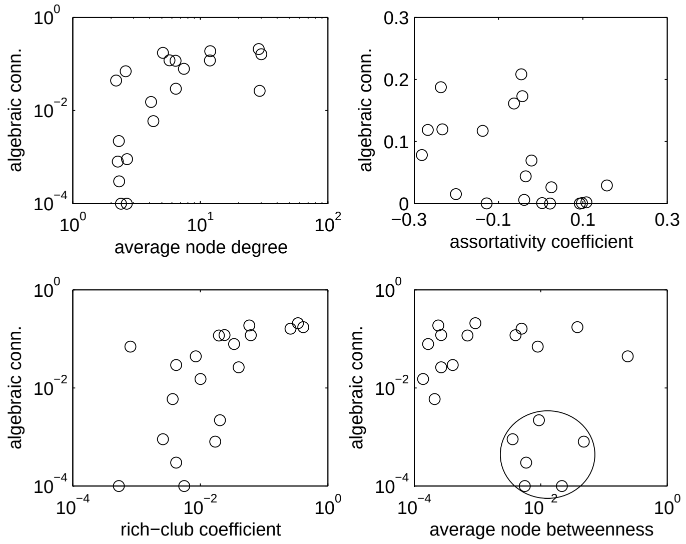

Fig. 5. The relationship among topological metrics for various real-world networks

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 metrics is positive or negative, but only how strongly two given metrics 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 metrics 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 metrics 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 metrics, 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.

We labeled different metric clusters according to the type of topological information the group of metrics provides. Intra- and inter- connectedness refer

Topological metrics	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 metrics.

to the metrics characterizing the observed connectivity, respectively, within and between a (sub)set of nodes in the network. All metrics within each cluster are highly or partly topologically redundant. The 14 initial metrics 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 metrics 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 metrics. The number of nodes is related to the number of links and all metrics within the distance and the intra-connectedness clusters, while not related to metrics within the degree or the inter-connectedness clusters. The algebraic connectivity, on the other hand, is related to all metrics within the degree, intra-connectedness, and the inter-connectedness clusters, but not to any metric in the distance cluster.

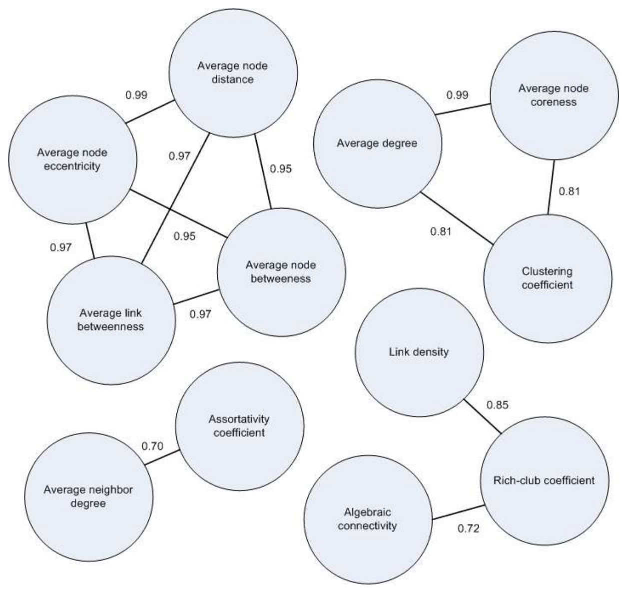

Fig. 6. A graph in which nodes are topological metric and links the correlations that emerged from the correlation analysis. The corresponding values display the strength of the correlation between pairs of metrics.

3.3 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 metrics. Principal component analysis (PCA) [21] 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.

Given a data set, denoted as a matrix $\mathbf{X}$ with the number of columns p as the number of variables to be analyzed $X_i, i = 1, \ldots, p$. Each 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[3]. $\mathbf{Y}$ 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

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

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

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 showes that only 3 dimmensions are enough to retain most of the original variability in the data. This, however, does not imply that metrics that are not important for the main principal components are unnecessary, but rather that they provide very specific topological information which does not fundamentally craracterize different networks.

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 metrics. Accordingly, the first principal component is composed of the two metrics, i.e. the number of links and the number of nodes. All other metrics have a very small weight in the linear combination of this principal component. In fact, the first principal component's metrics are those missing from the four clusters we identified in the correlation analysis, presented in Subsection 3.2. 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 metrics 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 metrics that characterize the variations in the

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

topological metrics are the number of nodes and links and the metrics within the distance and inter-connectedness clusters. Metrics 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 metrics.

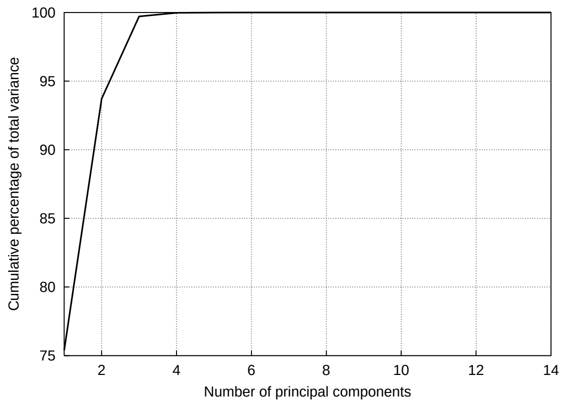

Fig. 7. Fraction of the variance captured by the principal components.

4 Discussions and Conclusion

In this paper, we have studied the relationships between topological metrics of real-world networks. The visual analysis, presented in Subsection 3.1, revealed the following relationships among topological metrics:

- 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 3.2, resulted in several highly-correlated clusters with the following topological metrics:

- 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 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 metrics tend to be more correlated than others. This observation implies redundancy between topological metrics. Consequently, we have identified a significantly smaller set of topological metrics that is able to characterize real-world network's structures. Further work comprises studying the relationships between topological metrics in real-world networks that are affected by various structural changes.

Acknowledgements

This research is partly funded by the Next Generation Infrastructures programme (www.nginfra.nl).

References

1. Albert, R., Barabasi, A.-L., *Statistical mechanics of complex networks*, Reviews of Modern Physics 74, 47-94 (2002).
2. Barabasi, A.-L. and Albert, R., *Emergence of scaling in random networks,* Science 286, 509-512 (1999).
3. Barabasi, A.-L., *Linked, The new science of networks,* Perseus, Cambridge, MA (2002).
4. Boccalettia, S., Latora, V., Moreno, Y., Chavez, M. and Hwanga, D.-U., *Complex networks: Structure and dynamics,* Physics Reports 424, 175 – 308 (2006).
5. Bollobás, B., *Random graphs,* Cambridge University Press (2001).
6. Bollobás, B. and Riordan, O., *Mathematical results on scale-free random graphs,* Handbook of Graphs and Networks: From the Genome to the Internet, Wiley-VCH (2003).
7. Borgatti, S.P., and Everett, M.G, *Models of core/periphery structures*, Social Networks 21, 375-395 (1999).
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. Chung, F.R.K., *Spectral graph theory,* Conference Board of the Mathematical Sciences No.92, American Mathematical Society, Providence, RI (1997).
11. Colizza, V., Flammini, A., Serrano M.A. and Vespignani, A., *Detecting rich-club ordening in complex networks,* Nature Physics 2, 110-115 (2006).
12. Colizza, V., Pastor-Satorras, R. and Vespignani, A., *Reaction-diffusion processes and metapopulation models in heterogeneous networks,* Nature Physics AOP doi:10.1038/nphys560 (2007).
13. Cvetković, D.M., Doob, M., Sachs, H., *Spectra of graphs, theory and applications,* Johann Ambrosius Barth Verlag, Heidelberg, 3rd edition (1995).
14. Cvetković, D.M., Doob, M., Gutman, I. and Torgašev, A., *Recent results in the theory of graph spectra,* North-Holland, Amsterdam (1988).
15. Dorogovtsev, S. N. and Mendes, J. F. F., *Evolution of Networks: From Biological Nets to the Internet and WWW,* Oxford Univ. Press, Oxford, 2003.
16. Fiedler, M., *Algebraic connectivity of graphs,* Czechoslovak Mathematical Journal 23, 298-305 (1973).
17. Holme, P., Kim, B. J., Yoon, C. N. and Han, S. K., *Attack vulnerability of complex networks,* Phys. Rev. E 65, 056109 (2002).
18. Jamakovic, A., Wang, H. and Van Mieghem, P., *Topological Characteristics of the Dutch Road Infrastructure, Seminar Infrastructure Reliability, Delft* (2006).

19. Jamakovic, A., Kooij, R.E., Van Mieghem, P. and Van Dam, E.R., *Robustness of networks against viruses: the role of the spectral radius, Proceedings of the 13th Annual Symposium of the IEEE/CVT Benelux, Liège* (2006).
20. Jeong, H., Mason, S., Barabási, A.-L. and Oltvai, Z.N., *Centrality and lethality of protein networks,* Nature 411, 41 (2001).
21. I. Jolliffe, *Principal Component Analysis,* Springer-Verlag (2002).
22. Krioukov, D., Chung, F., Claffy, kc, Fomenkov, M., Vespignani, A. and Willinger, W., *The Workshop on Internet Topology (WIT) Report,* ACM SIGCOMM Computer Communication Review 37(1), 69-73 (2007).
23. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E. and Dawsen, S.M., *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, 396-405 (2003).
24. Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., Claffy, K. and Vahdat, A, *The Internet AS-Level Topology: Three Data Sources and One Definitive Metric,* ACM SIGCOMM Computer Communications Review 36(1), 17-26 (2006).
25. milo, R., Itzkovic, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. and Alon, U., *Superfamilies of evolved and designed networks,* Science 303, 1538-1542 (2004).
26. Mohar, B., *Laplace eigenvalues of graphs: a survey*, Discrete Mathematics 109, 171-183 (1992).
27. Mohar, B., Alavi, Y., Chartrand, G., Oellermann, O.R. and Schwenk, A.J., *The Laplacian spectrum of graphs,* Graph Theory, Combinatorics and Applications 2, 871-898 (1991).
28. Newman, M.E.J., *The structure of scientific collaboration networks,* Proc. Natl. Acad. Sci. USA 98, 404-409 (2001).
29. Newman, M.E.J., *Assortative mixing in networks,* Phys. Rev. Lett. 89, 208701 (2002).
30. Newman, M.E.J., *The structure and function of complex networks*, SIAM Review 45, 167-256 (2002).
31. Watts, D.J. and Strogatz, S.H., *Collective dynamics of small-world networks,* Nature 393, 440-442 (1998).
32. Watts, D.J., *Small-Worlds. The dynamics of networks between order and randomness,* Princeton University Press (1999).

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

Topological metrics	Road	Rail1	Rail2	Air1	Air2	Power1	Power2	Power3	Power4	ISP
Number of nodes	14098	8710	689	500	2179	4940	3419	1713	1205	29902
Number of links	18687	11332	778	2980	31326	6594	3953	2043	1385	32707
Link density	0,0002	0,0003	0,0033	0,0239	0,0132	0,0005	0,0007	0,0014	0,0019	0,0001
Average degree	2,7	2,6	2,3	11,9	28,8	2,7	2,3	2,4	2,3	2,2
Average neighbor degree	2,9	2,8	2,5	53,8	140,5	3,9	3,8	2,9	3,1	45,7
Assortativity coefficient	0,093	-0,022	0,098	-0,268	-0,046	0,004	-0,128	0,022	0,108	-0,036
Rich-club coefficient	0,0005	0,0008	0,0172	0,0621	0,3395	0,0026	0,0042	0,0056	0,0204	0,0085
Clustering coefficient	0,0912	0,0212	0,0731	0,6175	0,4849	0,0801	0,0120	0,0145	0,0171	0,0306
Average node distance	80,6	79,0	34,1	2,9	3,0	18,5	21,1	38,0	12,3	7109,9
Average node eccentricity	177,4	158,6	65,0	5,2	5,9	34,1	38,9	71,8	22,6	14250,0
Average node coreness	2,4	2,4	2,0	8,2	19,1	2,2	2,0	2,1	2,1	2,1
Average node betweenness	0,0560	0,0090	0,0481	0,0040	0,0090	0,0036	0,0059	0,0216	0,0094	0,2377
Average link betweenness	0,00220	0,00350	0,02190	0,00050	0,00005	0,00140	0,00270	0,00930	0,00450	0,10870
Algebraic connectivity	0,0001	0,0695	0,0008	0,1186	0,2082	0,0009	0,0003	0,0001	0,0022	0,0440

Topological metrics	AS-level	Router	Protein	Soccer	Dolphins	Actor	Scientific	English	French	Spanish
Number of nodes	20906	29064	4626	685	62	10143	13861	7377	8308	11558
Number of links	42994	62260	14801	10310	159	147907	44619	44205	23832	43050
Link density	0,0002	0,0001	0,0014	0,0440	0,0841	0,0029	0,0005	0,0016	0,0007	0,0006
Average degree	4,1	4,3	6,4	30,1	5,1	29,2	6,4	11,9	5,7	7,4
Average neighbor degree	230,9	21,0	24,2	45,0	6,8	83,6	13,5	320,7	218,0	457,6
Assortativity coefficient	-0,201	-0,039	-0,137	-0,063	-0,044	0,026	0,157	-0,237	-0,233	-0,282
Rich-club coefficient	0,0101	0,0037	0,0196	0,2605	0,4127	0,0399	0,0042	0,0588	0,0240	0,0340
Clustering coefficient	0,2114	0,0232	0,0912	0,7507	0,2589	0,7551	0,6514	0,4085	0,2138	0,3764
Average node distance	3,9	7,1	4,2	4,5	3,4	3,7	6,6	2,8	3,2	2,9
Average node eccentricity	8,0	14,7	8,1	8,6	6,5	9,6	12,4	5,6	6,7	7,6
Average node coreness	2,9	3,0	4,4	20,2	4,5	21,4	4,9	7,5	3,9	4,9
Average node betweenness	0,0001	0,0002	0,0007	0,0050	0,0380	0,0003	0,0004	0,0002	0,0003	0,0002
Average link betweenness	0,00005	0,00006	0,00014	0,00022	0,01060	0,00040	0,00007	0,00003	0,00007	0,00003
Algebraic connectivity	0,0152	0,0059	0,1173	0,1612	0,1730	0,0004	0,0292	0,1875	0,1197	0,0782