Classification of graph metrics

Javier Martín Hernández^{*} and Piet Van Mieghem[†] Faculty of Electrical Engineering, Mathematics, and Computer Science Delft University of Technology, 2628 CD Delft

November 2011

Abstract

This article aims to order and classify a wide number of metrics, proposed to characterize graphs, and the services using those graphs. The number of proposed metrics over the graph history is overwhelming. Over the years, scientists constantly introduce new metrics in order to measure specific features of specific graphs. Aiming for generality, this research will focus on the classification of unweighted, undirected, general graph metrics.

Keywords: graph metrics, topology, service, correlation.

1 Introduction

Each complex network (or class of networks) presents specific topological features which characterize its connectivity and highly influence the dynamics of processes executed on the network. The analysis, discrimination, and synthesis of complex networks therefore rely on the use of measurements capable of expressing the most relevant topological features.

Firstly, the basic structural properties of a graph can be studied by solely considering its topology. The network topology specifies how items, called nodes, are interconnected or related to other nodes by links. A graph G is a data structure consisting of a set of N nodes connected by a set of L links. The set of nodes is denoted by \mathcal{N} , and similarly the set of links by \mathcal{L} .

Secondly, each link in G can be further specified by a set of link weights (such as delay, packet loss, available bandwidth, monetary cost, etc.), and each node can be characterized by a set of node properties (such as processing time, queue length, uptime, etc.). A network is fully defined when, in addition to the topological structure, multiple protocols, dynamics and constraints are set on top of the graph. This multilayer nature of a network leads to a natural metric classification split in *topological metrics*, and *service metrics*. The first class groups network properties obtainable by processing topological information, leading to an understanding on the connectivity of the network elements. Further, *service metrics* attempt to model the dynamic processes (such as end-to-end delivery, virus spread, etc.) present in the network. This differentiation between low and high level metric introduces the concepts of service and decomposability, which are discussed in section 4.1.

The European project ResumeNet¹ aims to propose a multilevel, systemic, and systematic approach to network resilience. The project defines *resilience* as the ability of a network to provide and maintain an acceptable level of service in the face of faults to normal operation. A basic understanding of how to quantify complex networks properties is key in the process of detecting and measuring challenges to normal operation.

^{*}Email: J.MartinHernandez@tudelft.nl

[†]Email: P.F.A.VanMieghem@tudelft.nl

 $^{^{1}\}mathrm{European}$ Union Research Framework Programme 7, FP - 224619 http://www.resumenet.eu/

This document presents a first step towards the framework by extensively enumerating and classifying existing graph metrics.

The document is structured as follows. Section 2 presents a comprehensive list of low-level topological metrics. For each individual metric a closed definition is given, together with hints at how the metric affects the network functions. Section 4.1 briefly introduces high-level service metrics together with the concept of decomposability.

2 Topological metrics

For unweighted, undirected graphs, the adjacency matrix is a square matrix A consisting of elements $a_{ij} = a_{ji}$ that are either one or zero depending on whether there is a link between the node *i* and *j* or not. A metric is classified as a *topological metric* if it can be calculated by using exclusively the adjacency matrix, such as the one in Figure 1. If any additional node or link property (such as link weight) is assumed in the calculations, the resultant metric is not a *topological metric*.



Figure 1: Adjacency matrix of a graph with N = 5 nodes and L = 6 links. Row *i* describes the connectivity pattern of node *i*. For example, the first row in the table tells us that node 1 connects to nodes 2 and 4.

Topological metrics are further classified into subclasses. The three proposed classes are: distance, connection, and spectra class. The *distance class* groups the metrics that make use of the hopcount random variable, which provides information about the number of nodes a message has to cross to reach its destination. The *connection class* groups metrics related to the nodal degree random variable (i.e. the number of a node's neighbours), together with metrics that help grouping nodes into clusters or hierarchies, thus providing insights into the structure of the network. Finally, the *spectra class* includes the eigenvalues and eigenvectors of a graph.

We expect that most metrics are correlated: a single metric such as the expansion can give us information on both degree and distance. It is beyond the scope of this document to analyze the correlation map between all the different metrics.

A detailed list with all the topological metric symbols can be found in Table 1 at the end of this document.

2.1 Distance class

In communication networks, paths are basic entities in connecting two communicating parties or nodes in a graph G(N, L). A path from a node n_A to a node n_B with k hops or links is the node list $P_{A\to B}(k) = n_A \to n_2 \to n_3 \to \dots \to n_{k+1}$ where $n_{k+1} = n_B$, $n_j \neq n_i$ for each index i and j. The value k is called the hopcount of the path. Let $X_j(n_A \to n_B)$ denote the number of paths with j hops between a source node n_A and a destination node n_B . Metrics making use of the hopcount random variable are included in the distance class.

2.1.1 Hopcount

The shortest hopcount $H_{A\to B}$ between two nodes n_A and n_B is the number of hops or links in the shortest path that connects the two nodes,

$$H_{A \to B} = \min_{k \in [1..N-1]} (P_{A \to B}(k))$$
(1)

hence the hopcount of the path $P_{A\to B}(k) = n_A \to n_2 \to n_3 \to \dots \to n_{k+1}$, is $H_{A\to B} = k$.

The hopcount distribution $\Pr[H = k]$ is the probability for a random pair of nodes to be at a distance k hops from each other [1]. The hopcount of a path is often associated in physics to the distance, length, or geodesic of such path [2] [3]. The distance or length of a path $P_{A\to B}(k) = n_A \to n_2 \to n_3 \to \dots \to n_{k+1}$ is the sum of all the link weights included in such path. When all the link weights in the graph have link weight $w_l = 1$, hopcount and distance become equivalent terms.

The hopcount distribution is important for many applications, the most prominent being routing. The performance parameters of routing algorithms strongly depend on the hopcount distribution [4]. The hopcount also plays a vital role in robustness of the network to worms [5]. Worms can quickly contaminate a network that has small distances between nodes. Topology models that accurately reproduce observed distance distributions will help researchers to develop techniques to protect networks from worms.

2.1.2 Closeness

The closeness [6] of a node n_i is the average hopcount obtained from this node to all the others. The most commonly used definition is the reciprocal of the total hopcount,

$$C_i = \frac{1}{\sum_{n_j \in \mathcal{N} \setminus \{n_i\}} H_{n_i \to n_j}} \tag{2}$$

Closeness is often regarded as a measure to quantify the node's participation in a network. Nodes with low closeness scores have short hopcounts from other nodes, and so will tend to receive information sooner and disseminate information faster. Closeness has been used in biology to identify central metabolites in metabolic networks. The reciprocal of the node closeness is also known as the Wiener index W_i [7].

2.1.3 Eccentricity, Diameter, Radius

The eccentricity ε_i of a node i is defined as the longest hopcount between the node n_i and any other node in G.

$$\varepsilon_i = \max_{n_i \in \mathcal{N}} (H_{n_i \to n_j}) \tag{3}$$

The eccentricity of a graph ε is the average eccentricity over all the nodes in G. It is closely related to the flooding time [1], which is the minimum time needed to inform the last node in a network. Intuitively nodes that play an important role in a topology should be easily reachable by the rest of the nodes in a graph.

The diameter D of a graph G is the maximum node eccentricity over all the nodes in G

$$D = \max_{n_i \in \mathcal{N}} (\varepsilon_i) \tag{4}$$

The diameter [8] can also be regarded as the *longest shortest* hopcount found in a graph. This measure gives an indication on how extended a graph is. Although it can be artificially inflated by long chains of nodes.

The radius R of a graph is the minimum node eccentricity over all the nodes in G

$$R = \min_{n_i \in \mathcal{N}} (\varepsilon_i) \tag{5}$$

2.1.4 Persistence

The persistence of a graph, as introduced by Boesch *et al.* [9], is the smallest number of links whose removal increases the diameter or disconnect the graph. The persistence of a graph of diameter D is the minimum over all pairs of nonadjacent nodes of the maximum number of disjoint paths of length at most D joining them.

2.1.5 Girth

The girth γ of a graph [10] is the hopcount of the shortest cycle contained in the graph. A cycle is a closed path $P_{A\to A}$, with no other repeated nodes than the starting and ending nodes This measure has a limited use, as any graph with clustering coefficient larger than 0 will provide $\gamma = 3$. The girth of an acyclic graph, such as a tree, is defined to be infinite.

The girth can be applied to prevent network the occurrence of loops. Routing algorithms operation may sometimes induce errors, which can lead to data packets being endlessly routed in a closed loop. For this reason, graphs with high girth values are less prone to suffer self loops. However the application of the girth is limited: link-state routing protocols (e.g. OSPF) prevent self loops after a flooding, additionally distance-vector routing protocols (e.g. BGP) have built-in loop prevention algorithms.

2.1.6 Expansion

The expansion e_h of a graph [1] is the average fraction of nodes in the graph that fall within a ball of radius h (in hops) centered at a random node in the topology

$$e_h = \frac{1}{N^2} \left| C_i(h) \right| \tag{6}$$

where $C_i(h)$ is the set of nodes that can be reached in h hops from a node i. We can interpret $C_i(h)$ geometrically as a ball centered at node i with radius h.

The expansion of a node provides information about the global graph reachability from a local point of view. Minimizing the expansion of a node or a set of nodes S in a network will shorten the number of hops a message generated by S will have to cross to reach its destination.

2.1.7 Betweenness

Betweenness of a node (or a link) B_k is defined as the number of shortest paths between pairs of nodes that traverse a node or link k. Let σ_{ij} be the number of shortest paths between nodes i and j, and k be either a node or a link. Let $\sigma_{ij}(k)$ be the number of shortest paths (1) between i and j going through node (or link) k. The shortest paths betweenness for the node (or link) k is

$$B_k = \frac{\sigma_{ij}(k)}{\sigma_{ij}} \tag{7}$$

Huijuan Wang *et al.* [11] show that in overlay trees of real world complex networks with exponential link weight distribution, the probability distribution function of B_k follows a power law $\Pr[B_k = j] = c_o j^{-c}$.

Commonly shortest paths are assumed in the calculation of $\sigma_{ij}(k)$: all paths follow the shortest paths as stated by (1). However, in real networks routing protocols may route traffic through non-shortest paths subject to multiple constraints. Alternative methods have been introduced [3] to calculate the betweenness based in random walks: consider the number of times a random traveling message passes through k along his journey averaged over a large number of trials. The full random walk betweenness of a node (or link) k will then be this value averaged over all possible source/target pairs i,j.

Betweenness has been studied in the past as a measure of the centrality or influence of nodes in social networks. First, proposed by Freeman [12], B_k can be a measure of the influence of a node (or link) over the global flow of information. In communication networks, betweenness measures the potential amount of traffic that crosses a node/link. This potential traffic will be affected when the node/link's fails.

2.1.8 Central Point of Dominance

The central point of dominance [12] is defined as

$$CPD = \frac{1}{N-1} \sum_{i} (B_{\max} - B_i) \tag{8}$$

where B_{max} is the largest value of the betweenness centrality in the network. CPD is a measure of the maximum betweenness of any point in the graph: it equals 0 for complete graphs and 1 for star graphs (in which there is a central node that all paths include).

2.1.9 Distortion

Consider any spanning tree T on a graph G, and compute the average distance $t = E[H_T]$ on T between any two nodes that share a link in G. The distortion measures how T distorts links in G, i.e. it measures how many extra hops are required to go from one side of a link in G to the other, if we are restricted to using T. The distortion is defined [13] to be the smallest such average over all possible Ts.

Intuitively distortion measures how tree-like a graph is.

2.2 Connection class

One of the major concerns of network analysis lies in the identification of cohesive subgroups of actors within a network. Cohesive subgroups are subsets of nodes among whom there are relatively strong, direct, intense, frequent, or positive ties. The node degree describes the neighbors of a node and is an key property to evaluate the graph structure [14]. The degree d_i of a node n_i is the number of other nodes to which n_i is connected,

$$d_i = \sum_{j=1}^{N} a_{ij} \tag{9}$$

Metrics that focus on the node degree analysis are included in the connection class. The metrics presented in this Section intend to classify nodes into intersecting or non-intersecting sets. Nodes belonging to the same set are expected to share structural properties.

2.2.1 Degree

Let d_k be the number of nodes with degree k. The node degree distribution is the probability that the degree D of a randomly selected node equals k,

$$\Pr[D=k] = \frac{d_k}{N} \tag{10}$$

The average value of this distribution is called the average degree² E[D], and obeys the basic law [14],

$$E[D] = \frac{2L}{N} = \sum_{k=0}^{N-1} k \Pr[D=k]$$
(11)

The minimum and maximum node degree of a given graph G are denoted as d_{min} and d_{max} , respectively. The degree distribution of a random graph follows a binomial distribution [1]. On the other hand, empirical results show that in real-world networks, the degree distribution significantly deviates from a binomial distribution. In particular, for a large number of networks, including the World-Wide Web [8], Internet AS [15] level or metabolic networks [16] [17] [18], the degree distribution has a skewed distribution with a power-law tail.

Resilience can be measured in several ways, but one of the most common indicators of network resilience is the variation on the fraction of nodes in the largest connected component upon link removals. In the context of communication networks the nodes in the giant component can communicate with an extensive fraction of the entire network, whereas nodes in the small components can only communicate with a few others. Studies

 $^{{}^{2}}E[D]$ it is often represented as \bar{k} in physics.

performed on the Internet AS topology [19] [20] show that networks with power law degree distributions are relatively robust with respect to a random failures. Only a failure of central nodes is likely to cause the network to fragment. On the other hand, this type of hub-based networks is extremely vulnerable to a targeted attack, in which the most highly connected nodes are removed first. These results lead to the Internet feature known as *robust yet-fragile*.



Figure 2: Degree distribution (in logarithmic-logarithmic scale) of a topology generated by the Power Law Random Generator (PLRG) algorithm [21]. The probability distribution follows a straight line, which indicates a power law behavior.

The topology of a network has a major impact on the performance of network protocols. For this reason, network researchers often use topology generators to generate realistic graphs for their simulations. These topology generators attempt to create network topologies that capture the fundamental characteristics of real networks, being the degree distribution the simplest metric to mimic [13].

2.2.2 Joint degree distribution (JDD)

The joint degree distribution (JDD) $\Pr[D_1 = k_1, D_1 = k_2]$ is the probability that a random pair of nodes possess a degree equal to k_1 and k_2 , respectively. The joint distribution of the degree of nodes at the end points of a link equals $\Pr[D_i = k_i, D_j = k_j | a_{ij} = 1]$. When selecting a random link $l \sim (l^+, l^-)$, we can rephrase this probability as $\Pr[D_{l^+} = k_1, D_{l^-} = k_2]$. Of course, the link l is shared by both nodes $i = l^+$ and $j = l^-$ so that often, we are interested in the remaining degree at the end points of a link, which is $D_{l^+} - 1$ or $D_{l^-} - 1$. Let $m(k_i, k_j)$ be the total number of links connecting nodes of degrees k_i and k_j , then the joint probability distribution of the degree at the end points of a randomly selected link connecting k_1 and k_2 -degree nodes [16] [17] is

$$\Pr[D_{l^+} = k_1, D_{l^-} = k_2] = \mu(k_i, k_j) \frac{m(k_i, k_j)}{2L}$$
(12)

where $\mu(k_1, k_2)$ is 1 if $k_1 = k_2$ and 2 otherwise.

While the node degree distribution tells us how many nodes of a given degree are found in a network, the JDD provides information on the interconnection between these nodes, by describing correlations of degrees of nodes located at distance 1.

2.2.3 Assortativity

A straightforward way to determine the correlation among the degrees is by considering the Pearson correlation coefficient [22, 1] of the degrees at either ends of each link. This normalized value is called the assortativity

coefficient r of a graph G [23, 14],

$$r = \frac{Cov[D_{l^+}, D_{l^-}]}{\sigma_X \sigma_Y} = \frac{\frac{1}{L} \sum_{(i,j) \in \mathcal{L}} d_i d_j - \left(\sum_{(i,j) \in \mathcal{L}} \frac{1}{2L} (d_i + d_j)\right)^2}{\frac{1}{L} \sum_{(i,j) \in \mathcal{L}} \frac{1}{2} (d_i^2 + d_j^2) - \left(\sum_{(i,j) \in \mathcal{L}} \frac{1}{2L} (d_i + d_j)\right)^2}$$
(13)

where D_{l^+} and D_{l^-} are the degrees of the nodes at the end of a randomly chosen link l in the graph. The assortativity coefficient lies in the range [-1, 1]. Assortative mixing (r > 0) is defined as a preference for highdegree nodes to attach to other high-degree nodes, whereas disassortative mixing (r < 0) as the converse, where high-degree nodes attach to low-degree ones. Assortative and disassortative mixing patterns indicate a generic tendency to connect to similar or dissimilar peers respectively.

Highly connected nodes tend to be connected with other high degree nodes in social networks [23]. On the other hand, technological and biological networks typically show disassortative mixing, as high degree nodes tend to attach to low degree nodes.

2.2.4 Coreness

The k - core is the subgraph obtained from the original graph by the recursive removal of all nodes of degree less than or equal than k [24]. Hence, in a k - core subgraph all nodes have at least degree k as illustrated in Figure 3.



Figure 3: The 0, 1, 2 and 3 cores of a sample graph.

The node coreness k_i of a given node n_i is the maximum k such that this node is present in the k – core graph, but removed from the (k + 1) – core. This measure can be regarded as an indicator of node centrality, since it measures how deep within the network a node is located.

2.2.5 Cliques and n-cliques

A clique [24] of a given graph G(N, L) is a subset of nodes such that all elements in the clique $S(N_S, L_S)$, where $N_S \leq N, L_S \leq L$, are fully connected, hence, forming a full mesh. The clique number of a graph equals the largest clique S in G. Finding the clique number in a graph is NP-hard. A relaxation of the clique concept is the *n*-clique. An *n*-clique S' of a graph is the maximal set of nodes in which for all $u, v \in S'$, the shortest hopcounts $H_{u\to v} \leq n$. In other words, an *n*-clique is a set of nodes in which every node can reach every other in n or fewer steps. The set S' is maximal in the sense that no other node in the graph is at distance n or less from every other node in the subgraph. By definition, 1-clique and clique are equivalent terms.

The knowledge about subgraphs with clique features within a network can decrease the complexity of algorithms designed for such network. For example, if a network is composed of trees hanging off a densely connected component, then an algorithm can run in the center component and a second algorithm tailored specifically for trees can run in the trees.

2.2.6 Clustering coefficient

The local clustering coefficient of a node n_i in a graph G measures the cliquishness of n_i neighborhood

$$c_i = \frac{y_i}{\binom{d_i}{2}} \tag{14}$$

where y_i is the number of links between neighbors of n_i , and d_i is the degree of the node n_i .

The clustering coefficient C of the whole graph G is the average of the local clustering coefficients for all the nodes in G,

$$C = \frac{1}{N} \sum_{i \in \mathcal{N}} c_i \tag{15}$$

This number is precisely the probability that two neighbors of a node are neighbors themselves. It is 1 on a fully connected graph (everyone knows everyone else) and has typical values in the range of [0.1, 0.5] in many real-world networks [16] [1]. The clustering coefficient can be interpreted as a measure of how close a node's neighbors are to forming a *1-clique*.

A characteristic of Erdös-Rényi graphs is that the probability of loops involving a small number of vertices goes to 0 in the large network size limit [1]. This effect presents a remarkable contrast with the abundance of short loops observed in many real world networks [25] [13]. The clustering coefficient has been extensively used in network topology studies, since it is a low complexity cliquishness indicator.

2.2.7 Rich Club coefficient

For a graph G, define S_k as the subset of nodes with degree greater than k, $S_k : \{n \in G | d_n > k\}$. We call this subset S_k the k-club members. The rich-club coefficient Φ_k is defined as the ratio of the number of links connecting the club members over the maximum number of allowable links in S_k , which measures how well the rich nodes know each other,

$$\Phi_k = \frac{1}{|\mathcal{S}_k| \left(|\mathcal{S}_k| - 1\right)} \sum_{i,j \in \mathcal{S}_k} a_{ij} \tag{16}$$

A monotonic increase of Φ_k is not enough to infer the presence of the rich club phenomenon [26] since even random networks generated from the Erdös-Rényi model [27], the Molloy-Reed [28] model and the Barabási-Albert model [29] have an increasing Φ_k with respect to k. Note that the definition of the subset S_k can be altered to regard other graph properties besides the degree distribution.

The rich club phenomenon in complex networks depicts the observation that the nodes with high degrees (called *rich nodes*) are inclined to intensely connect with each other. In other words, the rich club connectivity is a measure of how close induced graphs are to cliques. In a social context, a strong rich-club phenomenon indicates the dominance of an elite of highly connected and mutually communicating entities, as opposed to a structure comprised of many loosely connected communities. In the Internet, such a feature would point to an architecture in which important hubs are much more densely interconnected than peripheral nodes in order to provide the transit backbone of the network.

2.2.8 Giant component

A strongly connected component is a maximal subgraph G_C of a *directed* graph such that for every pair of nodes $n_A, n_B \in G_C$, there exists a directed path $P_{A\to B}(k)$ and a directed path $P_{B\to A}(k)$, for any hop k. Tarjan [30] presented an efficient algorithm to find the strongly connected components.

The number of nodes of the giant component of G denoted as

m(G)

has been applied to the definition of network robustness in many studies. A network is often considered robust if the size of the giant component

m(G)

remains constant as nodes or links are randomly removed from the network.

A large number of graph metrics cannot be calculated or they lose its meaning when the network under study contains more than a single strongly connected component. For instance, the hopcount between two nodes belonging to different disconnected components is not defined. The usual way to deal with disconnected components is the isolation and exclusive study of a network's giant component, which in many real cases includes a high percentage nodes. The metrics computed for the giant component are afterwards generalized to the totality of the network.

2.2.9 Reliability

A large number of metrics have been proposed under the term *realibility*. Reliability metrics measure the number of removed elements that lead to disconnected components in a graph. Because communication between all pairs of nodes in a graph is intuitively regarded as a vital condition for robustness, reliability has historically been the classical way to define graph robustness.

• The vertex connectivity κ and edge connectivity $\lambda(G)$ of a connected graph G are the respective smallest number of nodes and links whose removal disconnects G. For any connected graph G it holds that

$$\kappa(G) \le \lambda(G) \le d_{\min} \tag{17}$$

where d_{\min} is the minimum node degree in the graph G, and it sets a higher bound for the vertex connectivity. A graph is called k-connected or k-vertex-connected if its vertex connectivity is k or greater.

In addition, the *connectivity function* [31] specifies for each i the minimum edge connectivity that can be achieved after first removing i nodes.

• The cohesion $\mu(G)$ of a graph measures the minimum number of links whose removal creates a cut node in the network [32]. Thus removal of these $\mu(G)$ links, together with a single node, will disconnect the network.

When a network becomes disconnected it is desirable to capture the extend of disruption by measuring the size and number of the remaining connected components (defined in Section 2.2.8). After all, a system that has been split into many parts represents a more severe degradation than a system that has been split into a few large parts. To aid in describing such measures suppose a set of links $S \subseteq \mathcal{L}$ are removed from the graph G yielding the network G - S with c(G - S) connected components and maximum component size m(G - S)

- The *i*th-order edge connectivity [33] is defined as the minimum |S| such that c(G S) = i + 1.
- The *i*th edge separation value [missing reference] is defined as the minimum |S| such that $m(G-S) \leq i$
- The edge integrity [34] of G is the minimum value of the sum $\{|S| + m(G S)\}$ over all $S \subseteq \mathcal{L}$.
- The edge toughness [35] is the minimum value of the ratio $\{|S|/c(G-S)\}$ over all disconnecting sets S. However, edge toughness always equals $\lambda(G)$, so a more appealing measure is the node toughness: the minimum ratio of the size of a node disconnecting set to the resulting number of components(notice that most of the link-based measures presented have appropriate node-based analogues).

For the case of probabilistic networks, in which nodes and/or links fail randomly and independently with known probabilities, a number of measures have been presented.

- The two-terminal reliability $R_{ij}(G)$ between the nodes i and j is the probability that i and j are connected by a path of operating nodes and links.
- The two-to-K-terminal reliability $R_{iK}(G)$ is the probability that there is an operative path from node i to all nodes in a specified set $K \subseteq \mathcal{N}$.
- The source-to-all-terminal reliability or reliability polynomial R(G) is the probability that there is an operative path from any node *i* to all other nodes in the network. Notice that R(G) simply expresses the probability that the graph remains connected.
- The pair-connectivity or pair connected reliability [36] is the average number of node pairs able to communicate, taken over all possible node and link failures. This measure takes into account the fact that different ways of disconnecting the network are of different severity. For example, a certain number of link failures could separate G into several connected components $G_1, ..., G_r$. All communication is then disrupted between nodes in different components, and the resulting communication capacity can be measured by the number of pairs fo nodes now able to communicate $\sum_{i=1}^{r} \binom{n_i}{2}$ where n_i is the number of nodes in component G_i .

Reliability metrics represent an intuitive measure of robustness. When a graph becomes disconnected, potentially unique information (or services) stored in the disconnected nodes becomes unreachable and unavailable. For example, a computing circuit may short circuit due to a failure leading to a disconnected component. Historically [37], reliability has been the classical way to quantify network robustness.

Numerous algorithms have been developed for calculating the various network performance measures discussed previously. The effective computation of virtually all probabilistic measures exhibit a worst-case behavior that is exponential in the size of the network [38]. For this reason, the exact computation of reliability metrics has been confined to networks of small sizes.

2.2.10 Chromatic number

The coloring [39] of a graph G is a map $c : \mathcal{N} \to \mathcal{S}$ such that $c_n \neq c_m$ whenever nodes n and m are adjacent. The elements of the set \mathcal{S} are called the available colors. All that interests us about \mathcal{S} is its size: typically, we shall be asking for the smallest integer k such that G has a k-coloring, a node coloring $c : \mathcal{N} \to \{1, ..., k\}$. This k is the chromatic number of G, denoted by \mathcal{X} . A graph G with $\mathcal{X} = k$ is called k-chromatic.

Note that a k-coloring of a graph is a node partition into k independent sets, called color classes. The non-trivial 2-colorable graphs, for example, are the bipartite graphs. The non-trivial four color theorem [40], and five color theorem [39] imply that every planar graph is 4-colorable and 5-colorable respectively.

To avoid signal interference in frequency-division multiplexing communications the channels used by the antennas are chosen so that the same channel is never concurrently used by two neighboring antennas. This problem is named channel allocation and is usually modeled as a generalized list coloring problem.

2.3 Spectra class

The algebraic eigenproblem consists in the determination of the scalars λ and the corresponding vectors $x_{Nx1} \neq 0$ of any matrix A that satisfy the equation

$$Ax = \lambda x \tag{18}$$

The scalars and vectors are called eigenvalues and eigenvectors of A, respectively. The set of distinct eigenvalues λ is called the spectrum of A. Many network properties such as the vertex connectivity are related to the spectrum of a graph [41] [1].

Along this section we will work with the spectrum of the adjacency matrix and the admittance matrix. The admittance matrix or Laplacian Q is defined as

$$Q = \Delta - A \tag{19}$$

where $\Delta = diag(d_1, d_2, ..., d_N)$ is the degree matrix. The eigenvalues of the Laplacian matrix are represented with the symbol μ . The Laplacian matrix appears in many contexts in the theory of networks, such as the analysis of diffusion and random walks on networks, Kirchoff's theorem for the number of spanning trees, etc.

2.3.1 Algebraic connectivity

Let $\mu_1 \ge \mu_2 \ge ... \ge \mu_N$ be the ordered set of eigenvalues of the Laplacian matrix Q. The algebraic connectivity [42] is the second smallest eigenvalue of the Laplacian matrix, μ_{N-1} . For large N, the distribution of the algebraic connectivity grows linearly with the minimum node degree.

The algebraic connectivity is related to the speed of solving consensus problems in networks (solve distributed decision-making problems with interacting groups of agents), and it is a lower bound for the vertex connectivity [41] [43],

$$\mu_{N-1} \le \kappa(G) \tag{20}$$

Graphs with certain high connectivity properties (such as concentrators) have been used in the construction of switching networks that exhibit high connectivity. Tanner [44] proved that connectivity properties can be analyzed by its eigenvalues. He observed that a small ratio of the algebraic connectivity to the dominant eigenvalue implies good expansion properties.

2.3.2 Spectral radius

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ be the ordered set of eigenvalues of the matrix A. The spectral radius ρ of A is

$$\rho = \max_{1 \le i \le N} |\lambda_i| \tag{21}$$

The epidemic threshold [45] is defined as follows: for effective virus spreading rates below τ the contamination in the network dies out, while for effective spreading rates above τ , the virus is prevalent, i.e. a persisting fraction of nodes remains infected as displayed in Figure 4. Piet *et al.* [46] found the relation $\tau = 1/\rho$. It follows from this result that the smaller the spectral radius, the higher the robustness of any network against the spread of viruses.

2.3.3 Fiedler vector

The eigenvector corresponding to the second smallest eigenvalue (i.e., the algebraic connectivity μ_{N-1}) of the Laplacian matrix is also called the *Fiedler vector*. Spectral partitioning methods have been developed to split the nodes of a graph into two groups, such that the number of links between the groups is minimized [47] [48]. For the specific case where the network is split into two non-intersecting groups, we can define an index vector s with N elements:

$$s_i = \begin{cases} +1 \text{ if the node } n_i \text{ belongs to group 1} \\ -1 \text{ if the node } n_i \text{ belongs to group 2} \end{cases}$$
(22)

The group selection [47] that minimizes the number of links between such groups follows



Figure 4: Fraction of infected nodes as a function of the effective infection rate τ . The epidemic threshold is denoted by τ_c .

$$s_i = \begin{cases} +1 \text{ if } x_{N-1}(i) \ge 0\\ -1 \text{ if } x_{N-1}(i) < 0 \end{cases}$$
(23)

where x_{N-1} is the *Fiedler vector*.

Community detection in large networks might prove a very useful tool. Nodes belonging to a tight-knit community are more than likely to have other properties in common. For instance, in the world wide web, community analysis has uncovered thematic clusters, in neural networks communities may be functional groups, etc.

2.3.4 Principal eigenvector

The Principal eigenvector of an adjacency matrix A is the eigenvector corresponding to the largest eigenvalue λ_{max} (i.e., the spectral radius ρ). The principal eigenvector is the vector that maximizes the eigen equation (18) such that

$$Ax_1 = \rho x_1 \tag{24}$$

which attributes useful properties related to graph theory [49] [50] and data processing [20] [51] to x_1 .

First, the components of the principal eigenvector are directly related to node centrality or relative importance of nodes in a graph. Let us illustrate this idea with an example: the hyperlink structure of the World Wide Web can be modeled as a directed graph with N nodes where each node in the graph represents a webpage and the directed links represent hyperlinks. The corresponding adjacency matrix A is called transition matrix. Let us now imagine a web surfer who at each time step visits a random webpage in Figure 5. The user randomly picks a hyperlink on the current page i and jumps to a random page j it links to with a given Markovian probability $p = a_{ij} / \sum_{j=1}^{N} a_{ij}$. The stationary vector Π (i.e. Π_i equals the chance that the user is viewing the webpage i) is then defined to be the stationary distribution of the Markov chain, or equivalently the principal eigenvector x_1 of the transition matrix. Google's PageRank algorithm [49] uses a variation of the principal eigenvector to assign *authority* weights to webpages as illustrated in Figure 5. While originally designed in the context of link analysis on the web, it can be readily applied to citation patterns in academic papers and other citation graphs.

Secondly, the Principal eigenvector is tightly related to Principal Component Analysis (PCA) [51] which aims to find patterns in high dimensional data. Where the luxury of graphical representation is not available, PCA is a way of identifying patterns in data and expressing the data in such a way as to highlight intrinsic similarities. Given a dataset of limited elements, where each element is described with P scalar measurements,



Figure 5: Numeric example of PageRank values in a small graph with 11 related pages. The principal eigenvector ranks page B the highest, therefore B is the first page displayed in a search. Note that page C has a higher PageRank than page E, even though it has fewer links to it.

the eigenvectors of the respective covariance matrix lie in the axis that maximizes the variance of the dataset. In other words, the Principal eigenvector x_1 is a projection that accounts for as much of the variability in the data as possible. Although there are many ways to apply the PCA (e.g. image compression techniques [20]), the original usage as a descriptive, dimension reducing technique is probably still the most prevalent single application.

3 Metric correlations

Given any undirected graph G with N nodes, we can build a space S(N) containing $|S(N)| \simeq 2^{N^2}$ different graphs. Additionally constraining the number of links to L reduces the space to $|S(N,L)| \simeq N^{2L}$ graphs. In order to split the graph space into classes of graphs with similar features, first we must be able to uniquely define graph features with a set of metrics. We believe that a small set of topological measures will provide enough information to characterize any given graph. The next question to answer is, what metrics are needed?

Consider the geometrical space spanned by m orthogonal vectors e_1, e_2, \ldots, e_m , where e_j represents the axis of the *j*-th metrics, and each vector has unit norm, $||e_j||_q = 1$. Every topological metric t has a projection onto $e_j, t.e_j = t_j$. For each graph G, we can compute the topology vector, denoted by t_G to explicitly refer to the graph G, and each point or vector t_G with coordinates (t_1, t_2, \ldots, t_k) uniquely represents a single graph. We are interested to find what is the minimum number of topological metrics k such that $t_{G_i} \neq t_{G_j}$ for all $G \in S(N)$, and $i \neq j$, or what is the minimum number of metrics that allows for a unique graph classification.

In reality, topological metrics are not orthogonal. For example, the minimum degree d_{\min} and the algebraic connectivity μ_{N-1} are correlated because $0 \le \mu_{N-1} \le d_{\min}$. The higher the correlation between metrics *i* and *j*, the more the vectors e_i and e_j are aligned, and the less information is reflected. Defining a minimum set of non correlated metrics is a complex task, which requires a full understanding about the metric space.

The identified correlations are listed but not limited to the following.

3.1 Hopcount set

The hopcount and the betweenness introduced in Section 2.1.1 are tightly correlated. If $H_{i\to j}$ denotes the number of hops in the shortest path from node *i* to node *j*, then the total number of hops H_G in all shortest paths in *G* is

$$H_G = \sum_{i=1}^N \sum_{j=i+1}^N H_{i \to j}$$

which also equals to $H_G = \sum_{l=1}^{L} B_l$, where B_l is the betweenness the link l in G. Taking the expectations of both relations gives the average hopcount in terms of the average link betweenness

$$E[B_l] = \frac{\binom{N}{2}}{L} E[H] \tag{25}$$

The same reasoning can be followed for the average node betweenness, resulting in

$$E[B_n] = \frac{N-1}{2}(E[H] - 1)$$
(26)

Formulas (25) and (26) prove the linear relation between the average betweenness and the average hopcount for any graph in S(N, L). Even though the first moments are linearly correlated, the distribution of the random variables $H_{i\to i}$ and B_l can be different at a node level.

3.2 Connection set

The existing correlations between sets of metrics in the connection metric class were studied by Almerima *et al.* [52]. Using a large sample of real world networks, a set of topological metrics is calculated for each network, and a correlation coefficient is computed for each metric combination. The simulation results yield a correlation coefficient of 0.81 or higher between average node degree, average coreness, and clustering coefficient. A correlation of 0.72 is shown between the algebraic connectivity and the rich club coefficient for the sample. The study concludes that topological metrics tend to be correlated, which implies the existence redundancies.

3.3 Discussion

The correlations between topological metrics strongly depend on the graph under study [52]. For some extreme cases, metrics are correlated 1 to 1. For example in a full-mesh $K_N \in S(N)$ the clustering coefficient equals the diameter. This example demonstrates that the whole graph space S(N, L) should be explored before stating that two metrics are correlated or uncorrelated. However such an exploration proves to be an NP-complete problem due to the vast size of the graph space S(N). Instead, we rely on finding a suitable metric classification with our deep understanding of the metrics at hand.

Several metric classifications can be proposed based on different criteria, for example

- The random variable used to compute the metric. Examples of random variables could be the shortest path between two nodes $P_{A\to B}(k)$, node degree d_i , or matrix spectra $\{\lambda, X\}$ (this document). This has been the criteria chosen for this document, which leads to an intuitive metric classification.
- The correlation between metrics. This criteria groups strongly correlated metrics into the same class (e.g. average hopcount and average betweenness). Given that correlations depend on the graph under study, this classification may prove to be difficult.
- Local vs. global nature of the metric. This classification splits metrics in two subclasses: local metrics which are those computed for a single node or link (e.g. clustering coefficient), and global metrics which are the ones that only make sense when computed for the whole graph (e.g. chromatic number).
- Computatinal complexity. Metrics may be classified based on the rate of the number of operations required to compute its value for a defined set of graphs. The complexity is usually a polynomial function of the number of nodes N and links L present in the network.

Category	Class	Metric	Symbol
Topological	Distance	Hopcount	$H_{A \to B}$
		Closeness	C_i
		Eccentricity	ε_i
		Diameter	D
		Radius	R
		Girth	γ
		Expansion	e_h
		Betweenness	B_i
		Ce. Pt. of Domimance	CPD
		Distortion	t
	Connection	Degree	d_i
		Entropy	H
		Joint Degree	$\Pr[d_i, d_j]$
		Assortativity	r
		Coreness	k_i
		Clique	n-clique
		Clustering C.	C
		Rich Club coefficient	Φ_k
		Giant component	G_C
		Reliability	$\kappa(G),\lambda(G),\ldots$
		Chromatic number	X
	Spectra	Algebraic connectivity	μ_{N-1}
		Spectral radius	ρ
		Spectral partitioning	s_i
		Principal eigenvector	x_1

Table 1: Classification of the topological metrics, together with the proposed symbols.

Depending on what criteria we choose, we have to strike a balance between a human friendly classification and a clear yet not so intuitive classification. For example, the fact that a metric m can be computed in $O(N^2 log N)$ seconds gives us little information (4th presented criteria) about m. On the other hand if we know that m was computed by using shortest paths we can better identify its meaning (this document's criteria) and associations. The authors of this document opted for a human friendly classification, however all the presented criteria are equally valid and should be taken into consideration for future taxonomies.

Regardless of the metric classification at hand, one should expect strong metric correlations to always be present. This is because the existing metrics are not targeted to build an orthogonal metric space, or a space where each metric independently captures a dimension of the graph. But instead metrics appeared from a subjective need to measure observed graph qualities. A long list of graph metrics emerged during the last decades [23] [16] [41] from different scientific fields, each capturing partial features of a graph. This uncentralized generation process leads to the observed correlations, or even term overloading (e.g. clustering coefficient \equiv transitivity).

All the presented metrics are intertwined in a web of correlations, being our goal to choose a minimum set that uniquely classifies each graph in S(N). Spectral analysis shines in this regard: all the eigenvectors of a matrix are orthogonal, hence forming the perfect metric space m. However, the practical implications of a matrix's eigenvectors has yet to be studied.

4 Service metrics

4.1 Introduction

We classify as a service metric any metric that requires additional information to be calculated besides the adjacency matrix A, introduced in Section 2. This additional information can be a map of information flows running through the network, the behavior of agents inside of a network, the probability of certain nodes to crash, a distribution of link weights, etc. The complexity of both the metric definitions and the correlations among them grow with the number of input variables. Nowadays some correlations between relatively simple topological metrics remain unanswered, hence we can only expect complexity to increase when dealing with service metrics. This unstudied map of correlations leads to the decomposability problem.

Decomposability of graph metrics relates to the ability to express a graph metric in terms of two or more different graph metrics, regardless of the metric nature. If we can express a metric t_1 as a function of s_1 and s_2 , then we can say that metric t_1 is decomposable. Finding the decomposability map between topological and service metrics is as hard and extensive problem graph theory is far from completing. Our final goal is not only being able to characterize a graph space S(N), but the whole multilevel space of topology and service combinations, which requires an interdisciplinary effort.

4.2 Classification

Up to date the taxonomies proposed in the literature (ReSIST, Amber, DESEREC, HIDENETS, etc.) are mainly limited to the domain dimension of the respective classification. The domains included and the detail levels differ from proposal to proposal. In a unifying effort the European Network and Information Security Agency (ENISA) proposed in 2010 a model that includes all the identified domains in a two dimensional classification: an incident dimension, and a domain-based dimension.

4.2.1 Incident Dimension

ENISA's approach of identifying resilience metrics is event-based. A service level can be compromised when a time based event such as a security incident, a system failure or a human error occurs, the incident dimension then classifies metrics around the time an incident occurred. By dividing time into phases according to the state of the graph when an incident occurs, service metrics can be classified as preparation phase, service Delivery phase, or Recovery phase.

4.2.2 Domain Dimension

Independently of the time at which an event occured, we can classify a service metric in resilience domains, as defined by the ResiliNets research initiative. Resilience subsumes a number of disciplines, many of which are tightly interrelated but have developed separately. The considered disciplines are: survivability, disruption tolerance, traffic tolerance, dependability, security and performability.

ResiliNets gathers the six identified disciplines into two major groups: those that are related to the tolerance of challenges and faults, and trustworthiness that considers aspects that can be measured. The two categories are related by the notion of robustness as displayed in Figure 7.

When using the proposed two dimensional taxonomy, service metrics can be categorized according to Figure 6. It is out of the scope of this document to define the displayed service metrics. More information regarding the taxonomy and the metrics can be found in ENISA's resilience reports.

The core of this classification as well as class definitions have been extracted from $ENISA^3$, $ResiliNets^4$

³http://www.enisa.europa.eu/act/res/other-areas/metrics/reports

 $^{^{4}} https://wiki.ittc.ku.edu/resilinets_wiki$

		Dependability		Performability
dent-based classification	Preparedness	 Mean time to Incident Discovery Mean time to Patch Patch management coverage Vulnerability scanning coverage 	 Risk assessment coverage Risk treatment plan coverage Security testing coverage Security audit deficiencies Percent of ICT systems with BC plans 	• Tolerance
	Service Delivery	 Operational mean time between failures Operational availability Operational reliability Fault report rate 	 Incident rate Illegitimate network traffic Percent of systems without known severe vulnerabilities 	 Delay variation Packet loss Bandwidth utilization
Incid	Recovery	 Mean down time Mean time to repair Maintainability 	 Mean time to incident recovery 	

Domain-based classification

Figure 6: Service metrics categorized in the ENISA taxonomy.

project and ATIS-T1 PRQC⁵ (*Network Performance Reliability and Quality of Service Committee*) technical reports.

References

- [1] P. Van Mieghem, Performance Analysis of Communications. Cambridge University Press, 2006.
- [2] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, "Characterization of complex networks: A survey of measurements," *Advances In Physics*, vol. 56, p. 167, 2007.
- [3] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Social Networks*, vol. 27, pp. 39–54, 2003.
- [4] K. Satoshi and K. Takumi, "Evaluation of routing algorithms and network topologies for mpls traffic engineering," *IEIC Technical Report*, vol. 100/458, pp. 61–66, 2000.
- [5] J. Omic, R. E. Kooij, and P. Van Mieghem, "Virus spread in complete bi-partite graphs," in *Bionetics*, 2007.
- [6] D. Koschatzki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski, "Centrality indices," *Lecture Notes in Computer Science*, vol. 3418, pp. 16–61, 2005.

 $^{^{5}\}mathrm{http://www.atis.org/0010}$



Figure 7: This diagram illustrates a possible Service metrics classification in the domain dimension.

- [7] H. Wiener, "Structural determination of paraffin boiling points," J. Am. Chem. Soc., vol. 69, pp. 1–24, 1947.
- [8] R. Albert, H. Jeong, and A.-L. Barabasi, "The diameter of the world wide web," *Nature*, vol. 401, p. 130, 1999.
- [9] F. T. Boesch, F. Harary, and J. A. Kabell, "Graphs as models of communication network vulnerability: Connectivity and persistence," *Networks*, vol. 11, no. 1, pp. 57–63, 1981.
- [10] S. Skiena, Implementing Discrete Mathematics, Combinatronics and Graph Theory with Matematica. Addison-Wesley, 1990.
- [11] H. Wang, J. M. Hernandez, and P. Van Mieghem, "Betweenness centrality in a weighted network," *Phys. Rev. E*, vol. 77, p. 046105, Apr 2008.
- [12] L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, vol. 40, pp. 35–41, March 1977.
- [13] R. G. H. Tagmunarunkit and S. Jamin, "Network topology generators: degree-based vs. structural," in SIGMCOMM, 2002.
- [14] P. Van Mieghem, Graph Spectra for Complex Networks. Cambridge, U.K.: Cambridge University Press, to appear 2010.
- [15] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, (New York, NY, USA), pp. 251–262, ACM, 1999.
- [16] P. Mahadevan, D. V. Krioukov, M. Fomenkov, B. Huffaker, X. A. Dimitropoulos, K. C. Claffy, and A. Vahdat, "Lessons from three views of the internet topology," *CoRR*, vol. abs/cs/0508033, 2005.
- [17] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, K. C. Claffy, and A. Vahdat, "The internet as-level topology: three data sources and one definitive metric," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 17–26, 2006.

- [18] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, 2000.
- [19] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, "Breakdown of the internet under intentional attack," *Phys. Rev. Lett.*, vol. 86, pp. 3682–3685, Apr 2001.
- [20] C. Clausen and H. Wechsler, "Color image compression using pca and backpropagation learning," *Pattern Recognition*, vol. 33, no. 9, pp. 1555 1560, 2000.
- [21] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing, (New York, NY, USA), pp. 171–180, ACM, 2000.
- [22] J. L. Rodgers and A. W. Nicewander, "Thirteen ways to look at the correlation coefficient," The American Statistician, vol. 42, no. 1, pp. 59–66, 1988.
- [23] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, p. 208701, 2002.
- [24] V. Batagelj, A. Ferligoj, and P. Doreian, *Generalized Blockmodeling*. Cambridge University Press, 2005.
- [25] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the internet's router-level topology," 2004.
- [26] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *NATURE PHYSICS*, vol. 2, p. 110, 2006.
- [27] P. Erdős and A. Rényi, "On random graphs," Publicationes Mathematicae, vol. 6, pp. 290–297, 1959.
- [28] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," in *Random Structures and Algorithms*, pp. 161–179, 1995.
- [29] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 5439, pp. 509–512, 1999.
- [30] R. Tarjan, "Depth-first search and linear graph algorithms," SIAM Journal on Computing, vol. 1, no. 2, pp. 146–160, 1972.
- [31] L. W. Beineke and F. Harary, "The connectivity function of a graph," Mathematika, vol. 14, pp. 197–202, 1956.
- [32] R. D. Ringeisen and M. J. Lipman, "Cohesion and stability in graphs," *Discrete Mathematics*, vol. 46, no. 2, pp. 191–198, 1983.
- [33] F. Boesch and S. Chen, "A generalization of line connectivity and optimally invulnerable graphs," SIAM J. Appl. Math., vol. 34, p. 657665, 1978.
- [34] R. E. C.A. Barefoot and H. Swart, "Vulnerability in graphs-a comparative survey," J. Comb. Math. Comput., vol. 1, pp. 13–22, 1987.
- [35] V. Chvatal, "Tough graphs and hamiltonian circuits," Discrete Mathematics, vol. 5, no. 3, pp. 215 228, 1973.
- [36] K. S. A. Amin and P. Slater, "Pair-connected reliability of a tree and its distance degree sequences," Cong. Numer., vol. 58, pp. 29–42, 1987.
- [37] E. Moore and C. Shannon, "Reliable circuits using less reliable relays," Journal of the Franklin Institute, vol. 262, no. 3, pp. 191 – 208, 1956.

- [38] D. R. Shier, Network Reliability and Algebraic Structures. Clarendon Press Oxford science publications, 1991.
- [39] R. Diestel, Graph Theory, Electronic Edition. Springer-Verlag Heidelberg, 2005.
- [40] P. D. S. N. Robertson, D. Sanders and R. Thomas., "The four-colour theorem," J. Combin. Theory B, vol. 70, pp. 2–44, 1977.
- [41] M. Fiedler, "Algebraic connectivity of graphs," Czechoslovak Math, vol. 23/98, pp. 298–305, 1973.
- [42] R.Olfati-Saber, "Ultrafast consensus in small-world networks," in American Control Conference, 2005.
- [43] A. Jamakovic and P. Van Mieghem, "On the robustness of complex networks by using the algebraic connectivity," in *Networking*, pp. 183–194, 2008.
- [44] R. M. Tanner, "Explicit concentrators from generalized n-gons," SIAM Journal on Algebraic and Discrete Methods, vol. 5, no. 3, pp. 287–293, 1984.
- [45] A. A. Jamakovic, R. E. Kooij, P. Van Mieghem, and E. R. van Dam, "Robustness of networks against viruses: the role of the spectral radius," in 13th Annual Symposium of the IEEE/CVT, 2006.
- [46] P. Van Mieghem, J. Omic, and R. E. Kooij, "Virus spread in networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 1–14, 2009.
- [47] A. Pothen, H. Simon, and K.-P. Liou, "Partitioning sparse matrices with eigenvectors of graphs," SIAM J. Matrix Anal. App., vol. 11, pp. 430–452, 1990.
- [48] D. L. Powers, "Graph partitioning by eigenvectors," *Linear Algebra and its Applications*, vol. 101, pp. 121– 133, 1988.
- [49] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," Computer Networks and ISDN Systems, vol. 30, no. 1-7, pp. 107 – 117, 1998. Proceedings of the Seventh International World Wide Web Conference.
- [50] A. Jamakovic and P. Van Mieghem, "On the robustness of complex networks by using the algebraic connectivity," in *Networking*, pp. 183–194, 2008.
- [51] I. Jolliffe, *Principal Component Analysis*. Wiley Online Library, 2002.
- [52] A. Jamakovic and S. Uhlig, "On the relationships between topological measures in real-world networks," *Networks and Heterogeneous Media*, vol. 3, no. 2, pp. 345–360, 2008.