Expert Systems With Applications 107 (2018) 72–88

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Feature selection for optical network design via a new mutual information estimator

Daniela Bertolini Depizzol[a,d,*], Jugurta Montalvão[b], Fabio de Oliveira Lima[c], Marcia Helena Moreira Paiva[d], Marcelo Eduardo Vieira Segatto[d]

[a] Coordenação de Logística, Campus Cariacica, Instituto Federal do Espírito Santo (IFES), Cariacica, Espírito Santo, 29150-410, Brazil
[b] Departamento de Engenharia Elétrica, Campus São Cristóvão, Universidade Federal do Sergipe (UFS), Aracaju, Sergipe, 49100-000, Brazil
[c] Coordenação de Engenharia de Controle e Automação, Campus Serra, Instituto Federal do Espírito Santo (IFES), Serra, Espírito Santo, 29173-087, Brazil
[d] Laboratório de Telecomunicações - LabTel, Campus Goiabeiras, Universidade Federal do Espírito Santo (UFES), Vitória, Espírito Santo, 29075-910, Brazil

ARTICLE INFO

Article history:
Received 31 May 2017
Revised 12 April 2018
Accepted 12 April 2018
Available online 17 April 2018

Keywords:
Feature selection
Mutual information
Optical network design
Number of wavelengths
Topological invariants
Graph theory

ABSTRACT

An efficient design of optical networks is a complex challenge that requires knowledge of the desired performance trends. Such knowledge would have a potential impact on an expert system to this end, for instance, would help identify reliable topological parameters to characterize the desired behavior of the network. Feature selection from information theory is widely explored in many areas of expert and intelligent systems, and it is a suitable technique to choose such parameters. In optical networks, many signals are carried along the same fiber, each one with its wavelength. A possible desired performance is the minimal usage of different wavelengths, which can be influenced by many topological parameters established in the network design. However, it is difficult to determine the dependence between topological parameters and the number of wavelengths, because this latter addresses an NP-hard problem. We perform a comprehensive literature review to find topological metrics that are easier to compute and apply feature selection using a new mutual information estimator. Based on coincidence detection, this estimator is lightweight and easy-to-use and allows measuring the relevance between discrete and continuous features, without discretization nor estimating probability density functions. For this purpose, tests are performed using 315 topological parameters from graph theory and complex networks, in 15 real-world optical networks and 2.2 million random topologies that mimic real-world ones. The topological parameters are ranked based on its mutual information values, obtaining a set of the most influential for explaining the wavelength requirements. Among these parameters, as a result, the method highlights the ones derived from the edge betweenness. Moreover, some parameters proposed by the literature do not perform as expected. The results of this study can serve as a basis for new expert systems to design and expansion of optical networks, driven by the most relevant topological parameters.

1. Introduction

Optical Networks are currently the foundation of the extensive global digital communication network due to many factors such as their large traffic capacity, velocity, and reach. In this networks, many independent channels can share the same optical fiber, increasing the data rate within the same infrastructure, i.e., many signals can be transported using a same optical fiber. Each signal uses a wavelength, and this wavelength should be different for each signal transported. The totality of these different wavelengths passing through the fibers is commonly a variable of interest in the literature. Conventional optical networks, such as Optical Transport Networks (OTN), use the Wavelength Division Multiplexing (WDM) technology, which allows the implementation of Wavelength Routed Optical Networks (WRON) (Banerjee & Mukherjee, 2000). A new generation of optical networks, called Elastic Optical Networks (EON), is based on Optical Orthogonal Frequency Division Multiplexing (OOFDM). This technology allows more flexible usage of the optical spectrum, with channels of different sizes, achieving higher spectral efficiency (Christodoulopoulos, Tomkos, & Varvarigos, 2011; Tessinari, Puype, Colle, & Garcia, 2016; Zhang, De Leenheer, Morea, & Mukherjee, 2013).

* Corresponding author at: Laboratório de Telecomunicações - LabTel, Campus Goiabeiras, Universidade Federal do Espírito Santo (UFES), Vitória, Espírito Santo, 29075-910, Brazil.

E-mail addresses: ddepizzol@ifes.edu.br (D.B. Depizzol), jmontalvao@ufs.br (J. Montalvão), fabio.lima@ifes.edu.br (F.d.O. Lima), marcia.paiva@ufes.br (M.H. Moreira Paiva), segatto@ele.ufes.br (M.E. Vieira Segatto).

https://doi.org/10.1016/j.eswa.2018.04.018
0957-4174/

WRON networks involve the Routing and Wavelength Assignment (RWA) problem, which addresses the demand routing and wavelength allocation to the optical channels. The RWA can be solved to optimality, for relatively small networks using integer programming models, for instance as in Jaumard, Meyer, and Thiongane (2007) and Cousineau, Perron, Caporossi, Paiva, and Segatto (2015). There are many approaches with different objectives, but the one that separates the RWA into two clear subproblems is prevalent: demand routing, followed by wavelength allocation to the optical channels, with the objective to minimize the number of wavelengths to meet the demand for the attributed routing. This objective function is often chosen in optical network design because the wavelength requirement is related to the cost and the capacity of networks.

The RWA problem includes the wavelength continuity constraint, which makes it an NP-hard problem (Zang, Jue, & Mukherjee, 2000). In this context, each channel should use the same wavelength from the start to the end of the route. As a consequence, such constraint can generate a fragmentation of the available spectrum, which could have wavelengths available in many links, but without continuity between consecutive links, hampering to create routes with more than one link.

The called minimum number of wavelengths is the minimum number of wavelengths required to meet a given traffic demand. The term minimum number of wavelengths is called the number of wavelengths and denoted by λ. In the present study, λ is calculated in an exact way using the integer programming model proposed by Cousineau et al. (2015), as it is said in Section 2.1.

In EON networks, there is the corresponding Routing, Modulation, and Spectrum Assignment (RMSA) problem, which adds the non-overlapping constraint (Patel, Ji, Jue, & Wang, 2012) to the RWA. The optical spectrum is subdivided into slots, and channels of different sizes are created from combinations of the slots to meet the demands of different rates and requirements. The continuity constraint also applies to the channels, but the combination of channels with different sizes causes another type of fragmentation, given that only contiguous slots can form channels. A common strategy for addressing this problem is the subdivision of the spectrum into partitions, allocating to each one only channels of the same size. Hence, within each partition, the problem is reduced to the classical RWA (Wang & Mukherjee, 2014). In any case, because the RMSA is a more complex problem, the network should ideally have a low requirement for the number of wavelengths to meet the continuity constraint (Talebi et al., 2014).

Regardless of network type, an expert system to design it often involves conflicting aspects, which leads to optimize a set of features simultaneously as in Yang, Wu, Chen, and Dai (2010), or different types of network flow like in Przewoźniczek, Goścień, Walkowiak, and Klinkowski (2015). These problems have been extensively studied using all sort of artificial intelligence methods, neural networks, and genetic algorithms (Hanay, Arakawa, & Murata, 2015). For this design process, linear programming models are also frequently used, as in Antunes, Craveirinha, and Climaco (1993) and Yoon, Baek, and Tcha (1998).

The decision of which variables to include in the design model is as important as, or more important than, the model itself. Then the choice of the features to be optimized define the profile of the optical networks obtained. For instance, survivability, traffic capacity, and resources requirements such as number of wavelengths, depend on the variables chosen to be optimized in the design process.

In this work, we are particularly interested in investigating which are the topological parameters that better explain wavelength requirements. A suitable technique to obtain these parameters is the feature selection, a topic widely explored in many areas of expert and intelligent systems (Bennasar, Hicks, & Setchi, 2015). We perform a comprehensive literature review of graph topological parameters that are easier to compute than the number of wavelengths. Then, we apply a filter feature selection based on a new mutual information estimator to rank and select those parameters that lead to low wavelength usage. Bennasar et al. (2015) also works with filter method based on mutual information, but in our case, the mutual information estimator can be applied to discrete or continuous data regardlessly, without discretization nor any *a priori* knowledge concerning source distribution. Filter methods have advantages like computational efficiency and scalability in terms of the dataset dimensionality. The expected drawbacks in filter methods are the lack of information about the interaction between the parameters (the features) and the classifier, and selection of redundant parameters (Bennasar et al., 2015).

A novel entropy estimator based on coincidence detection from Montalvão, Attux, and Silva (2014) is applied to define the new mutual information estimator used in this study. This estimator is lightweight and suitable for high-dimensional spaces, similar to the Neighborhood Mutual Information from Hu et al. (2011). Our new approach uses the so-called "Method of Coincidence", a notion borrowed from Statistical Mechanics (Ma, 1985). In Section 2.3 such entropy estimator is presented, first for discrete variables, then extended to continuous ones. The resulting mutual information estimation allows measuring the relevance between discrete and continuous features, without discretization nor any a priori knowledge concerning source distribution. This technic is easy-to-use and suitable for small datasets, or where the data is difficult to reproduce. Hu et al. (2011) also estimate entropy for discrete and continuous features without discretization. However, our approach based on the method of coincidence is a more intuitive and not need to estimate probability density functions.

The topological parameters are ranked based on its mutual information values, obtaining a set of the most influential for explaining the wavelength requirements. As mutual information does not assume linearity between the variables (Bennasar et al., 2015), then relevant parameters are selected regardless of their relationship with the wavelength requirement. The results of this study serve as a basis for new expert systems to design and expansion of optical networks.

Some authors have studied topological parameters in WDM, as can be seen in Section 1.3, but with less comprehensiveness than the present work, and without any result of parameters sorting. To the best of our knowledge, no studies relating network topological features to EON parameters can be found in the literature. In the present study, optical networks are modeled as graphs, then this modeling and needed graph theory concepts are both presented in Section 1.1. In Section 1.2, we expose the problem in a concrete instance.

1.1. Graph theory in optical networks

An intuitive way to represent an optical network is through graphs. A graph is denoted by G(V, E), or just G, where V is a set of vertices and E is a set of edges (Diestel, 2016). The vertices correspond to the network nodes, and the edges correspond to the links, so the edges connect the vertices in the same way the links connect the nodes. The number of vertices is the order of the graph, denoted as $n = |V|$. The number of edges is the size of the graph and is denoted as $m = |E|$.

From the graphs, it is possible to compute the so-called graph invariants, or topological invariants, which are numeric parameters that do not change when labels of vertices or edges change. Invariants are important because they represent topological parameters of the graph and, consequently, of the optical network modeled by the graph. Hence, graph invariants are computed in this study to

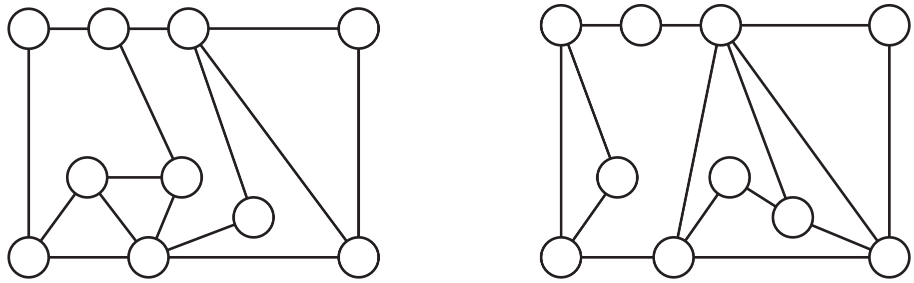


Fig. 1. Two network topologies with the same number of nodes ($n = 10$) and links ($m = 15$). Topology (a) requires 8 wavelengths, whereas topology (b) requires 12 wavelengths, i.e., 50% more wavelengths than topology (a).

understand how the network topological parameters influence the number of wavelengths (λ).

Graphs can have weights in its vertices or edges. In this case, the graph is called weighted, otherwise, it is called unweighted. A graph G(V,E) is connected if, for each pair of vertices u, v $\in$ V, there is at least one path interconnecting u and v. The shortest paths in terms of number of edges are called geodesics. In unweighted graphs, the distance between two vertices u, v $\in$ V is defined as the length of a geodesic interconnecting u and v. The mean distance of a graph is just the mean of all distances of all pair of vertices.

In addition, depending on the application, edges may or may not have a direction. When edges of a graph have a direction, the graph is said to be directed. In the context of optical networks, each link allows flows in both directions. Thus, only non-directed graphs are discussed in this work.

A simple graph is the one that (i) has at most one edge for each pair of vertices, i.e., without parallel edges, (ii) has no loops, i.e., without edges that start and end in the same vertex, and (iii) is non-directed.

Let G be a simple graph on n vertices and m edges. The maximum number of edges, m_{max}, in any simple graph on n vertices is a combination of n taken 2 at a time. Then, the edge density (α) of G is the ratio between m and m_{max}. This topological invariant can be used to estimate an amount of resources in the network.

Another property in graph theory is the vertex degree. For each vertex i, the number of edges connected to this vertex is called the vertex degree (or just degree) of this vertex (Diestel, 2016). The mean degree is the mean of the degrees of all vertices of a graph, and the variance degree is the variance also of the degrees of all vertices of a graph.

A graph is called 2-connected when, for each pair of vertices, there are at least two vertex-disjoint paths interconnecting the pair. This property is convenient to optical networks because, in case of single vertex or link failure, the graph remains connected.

In the next section, we present a concrete instance of the problem treated in the present study via graph modeling, which induces us to explore the optical network topologies to understand the wavelength requirements.

1.2. Exposing the problem

In optical network design, there is a huge number of possible topologies that can be generated from fixed numbers of nodes (n) and links (m), which are the most elementary topological parameters. For instance, there are 11,716,571 possible 2-connected topologies with 10 nodes. So it is not easy to find topologies that meet a set of desired parameters. To illustrate how wavelength requirements depend on topological parameters, Fig. 1 shows topologies with the same number of nodes ($n = 10$) and links ($m = 15$), but with different wavelengths requirements in uniform traffic demand, which are calculated by the method of Cousineau et al. (2015), as established in Section 1.

Efficient optical network design is a complex challenge. Among the multiple ways to connect nodes using links, the control of wavelength requirements is not easy. To reduce this complexity, instead of directly using wavelength requirement values it is suitable to explore the topological parameters of optical networks and identify how these parameter values can influence the wavelength requirements.

Thus, this study aims to find topological parameters, which can be obtained with low computational cost and that can be optimized to lead to a low wavelength requirement. Mutual information is used to select the most influential ones. The knowledge of such parameters can better guide the network topology design to expand a current optical network or create new networks.

As we model optical networks via graphs, then the topological parameters correspond to the invariants from graph theory. A complete list of the invariants analyzed in this paper can be found in Appendix A.

During the design phase of optical networks, a set of invariants can be used to characterize network topologies and identify the ones that optimize certain parameters of interest, such λ, for instance. Over time, some studies in the literature have made an effort in this direction, as can be seen in the following section.

1.3. Related works

Baroni and Bayvel (1997) are the first to analyze the wavelength requirements of real-world networks and a large number of randomly connected WRON networks for wide-area backbone applications. The networks considered in the study are simple, unweighted, 2-connected, and satisfy $0.1 < \alpha < 0.4$. The wavelength continuity constraint is assumed. It is also assumed uniform traffic demand, i.e., for each pair of nodes $\{i, j\}$ in the network, there is a single demand from i to j, where each demand uses a geodesic, and a single wavelength from the start to the end of the transmission. Through a heuristic, the number of wavelengths of the networks under study is estimated and then analyzed with respect to some invariants, such as vertices number and the edge density. Notice that in this paper the edge density is called physical connectivity. According to the observations, the average of λ requirements is practically independent of the network order but strongly decreases as edge density increases.

In Fenger, Limal, Gliese, and Mahon (2000), wavelength requirements in relation to the topological parameters of WDM optical networks are also analyzed. For this purpose, a few million random 2-connected networks with a certain number of nodes and links are generated, and the analyzed invariants of these networks are mean degree, variance degree, and number of spanning trees, all of which are described in Appendix A. General results are obtained with the average of each topological invariant studied. In this study, the authors also assume simple and unweighted graphs, uniform traffic, routing using geodesics, and wavelength continuity. It is also considered that there is no limit on the number of wavelengths that can pass through each link. The results obtained for networks with 30 vertices and 45 edges show that the average of λ increases with variance degree. The authors note that more regular networks, where the vertices tend to have closer degrees values, require fewer wavelengths. In turn, observations in networks with the number of vertices and edges equal to 10 and 20, 20 and 30, 20 and 45, and 30 and 45, respectively, show an inversely proportional relation between the number of spanning trees and the average of λ. This relation is also observed for the mean degree, with tests performed in networks with 10, 20 or 30 vertices. A considerable decrease in wavelength requirements is found as the mean degree increased to somewhere between 4 and 5, without significant advantages for networks with higher mean degrees. Thus, using the analyzed sample, the study concludes that networks with a mean degree between 4 and 5 and with low variance degree are the best possible in terms of wavelength requirements. It is also concluded that the number of spanning trees is a very accurate invariant to measure the network quality in terms of the traffic accommodation efficiency, given that the observations show a power-law relation between the number of spanning trees and the average of λ.

In the study by Châtelain, Bélanger, Tremblay, Gagnon, and Plant (2009), a set of 18 real networks with the number of nodes ranging from 11 to 53 is analyzed, as is a set with 1.5×10^6 random networks with the number of nodes ranging from 10 to 50, in 10 units steps. This study only analyzes random networks with the number of edges equal to 1.5, 2 and 2.5 multiplied by the number of nodes. For the set of real networks, it is found that $0.05 < \alpha < 0.45$, where the lowest values are associated with the largest networks. The number of wavelengths used in the study is estimated by a lower bound. For random networks, a power-law relation is established between the average of λ and the algebraic connectivity (described in Appendix A). From this relation, a concise equation is derived to predict wavelength requirements, which is tested in real long-distance networks. The authors state that the estimation of wavelength requirements based on algebraic connectivity is more accurate than the estimations performed using the variance degree, the numbers of spanning trees, and the mean distance.

Yuan and Xu (2010) also study wavelength requirements, but for optical networks with small-world and scale-free physical topologies. Two characteristics are analyzed: mean distance and 1-shell structure (Bickle, 2013). Wavelength requirements are also studied in the case of network evolution. About a hundred networks with 100 nodes and 200 links are analyzed. The traffic demand used for the RWA is not necessarily all-to-all, the links allow bidirectional flow, and the routing does not force the use of geodesics. The observations show that the mean distance traveled by the signal is directly proportional to the mean distance. Also according to the observations, networks with shorter mean distances tend to require fewer wavelengths. For these networks, the wavelength requirements growth rate is lower when traffic volume and network order increase. Additionally, the presence of 1-shell structure can increase the wavelength requirements. The higher the number of 1-shell is, the more wavelengths are required by the network.

We emphasize the fact that, in almost all works discussed above, (i) the average of the number of wavelengths is considered, and that can hide extremal cases of excessive or reduced wavelength requirements; (ii) the number of wavelengths is computed using heuristics or estimated by lower bounds, that can lead to imprecise relations between topological invariants and wavelength requirements. In contrast, our work considers extremal wavelength requirements, and compute them as optimal solutions of an ILP model (Cousineau et al., 2015). These are two important aspects of our methodology, which is described in next section.

2. Methodology

The methodology consist, in short, to generate a random sample of optical networks which are represented by graphs, and to verify in this sample which topological invariants have the greatest influence on the number of wavelengths. This inspection is performed by applying the new mutual information estimation presented in the Section 2.3. This analysis is also performed for real networks, and the results are compared.

Literature presents many methods to compute mutual information with good performance, e.g., Bennasar et al. (2015). We chose to innovate by estimating the mutual information by the method of coincidence (Ma, 1981; 1985) because it allows the computation of the mutual information of discrete and continuous features in a simple way, making it possible to differentiate the most important invariants in this scenario, while objectively measuring its importance. Another way to measure mutual information is by means of probability density functions (pdfs) of the random variables in question. However, the method of coincidence is chosen because it avoids the typical problems of pdf estimation in high-dimensional spaces, which is the case when performing invariants associations (Montalvão et al., 2014). The method of coincidence draws attention to its simplicity, both in its adaptability to high dimension spaces and in the simultaneous analysis of discrete and continuous variables combined (which is our case). This simultaneous analysis comes from the fact that it is enough to define the event "coincidence between observations" so that any set of variables (even heterogeneous) can be analyzed in the same way, and by the same method, as it is explained in Section 2.3.

To better understand the entire methodology, Section 2.1 describes how the random networks are generated and how the number of wavelengths is computed. All information about the analyzed topological invariants is presented in Appendix A. Section 2.2 describes the real networks used in the study, and Section 2.3 details the method of coincidence used to estimate the mutual information, which identifies the invariants that best provide information on wavelength requirements.

2.1. Random networks

The studies are concentrated on random graphs that mimic real optical networks, with vertices number $n = 10, 11, \ldots, 20$, where 200,000 non-isomorphic graphs are randomly generated for each n, resulting in a sample with 2.2×10^6 random graphs. Random networks of order at least 10 nodes are generated to avoid likely border effects in small graphs, and at most 20 nodes, due to the computational cost of treating the entire sample.

In addition, for this sample of random graphs to be as similar as possible to a sample of optical networks, common or desirable characteristics the of optical networks are considered. Hence, the graphs of the sample are generated so that they also have the following characteristics:

(i) simple, therefore without loops, without parallel edges and whose edges support bidirectional flow;
(ii) edges with unit weight, as a way to find patterns that could be masked when using edges with non-unit weights;
(iii) 2-connected as in Baroni and Bayvel (1997), as presented in Section 1.3. Among the generated graphs, only the 2-connected ones are considered, until obtaining 200,000 graphs per network order;

(iv) mean degree ranging from 2 to 4, as expected for real-world networks (Pavan, Morais, Ferreira da Rocha, & Pinto, 2010). The Erdös–Rényi model (Erdös & Rényi, 1960) is used to generate the random graphs. Assuming the hypothesis of mean degree ranging from 2 to 4, it is easy to see that the probability of an edge to connect each pair of vertices is $3/(n-1)$.

The RWA is solved for each graph of the sample, returning the number of wavelengths required. To solve the RWA, we consider a uniform traffic demand, which means:

(i) a bidirectional, static, logical requirement is imposed between each pair of nodes (all-to-all demand);
(ii) the routing is exclusively performed using shortest paths (geodesics), considering the weight of the edges to be unitary; and
(iii) the wavelength continuity constraint is imposed, which means that each signal should be transmitted using a fixed wavelength from the source node to the destination node.

These considerations to solve the RWA are chosen to force an extreme and fixed scenario, since the interest is to understand how network topologies can interfere in wavelength requirements. It is then required that all other characteristics remain fixed, given that variations in traffic demand or in routing process, for instance, can mask results.

The method used for solving the RWA is provided in Cousineau et al. (2015), which combines models of integer linear programming and lower bounds, among other techniques, culminating in a segmented and low-cost computational approach that allow the number of wavelengths to be obtained even for a large sample. Only 52 graphs from all 2.2×10^6 graphs are not solved within the established time limit.

For the analysis of the graph sample, a set of topological measures is selected to determine whether the different values they assume influence the wavelength requirements. These measures are presented in Appendix A. Some of them refer to vertex or edges, and other measures to the whole graph. When the measures concerning the graph *per se* do not depend on labels of vertices or edges, they are called invariants, as already mentioned. Otherwise, they are referred simply as measures.

A set of measures and invariants as diverse and as large as possible is proposed, within our best efforts. Some of these measures are exalted in the literature of optical networks. An exploratory approach is performed, where no prior trial is assumed on to the potential of the measures to explain the number of wavelengths. This course of action is selected to allow new correlations to be identified.

Given that in Appendix A there are measures which vary according to the labeling of vertices and edges, for each of these measures, the following ten invariants are adopted for each graph: (i) minimum, (ii) maximum, (iii) mean, (iv) median, (v) amplitude, (vi) interquartile distance, (vii) standard deviation, (viii) coefficient of variation, (ix) standard deviation in relation to the median,[1] and (x) coefficient of variation in relation to the median.[2]

As can be observed in Appendix A, there are 29 measures and 30 invariants, for a total of $30 + (29 \times 10) = 320$ invariants to be analyzed. The definitions of diameter, maximum distance, and maximum eccentricity coincide, so only one of them (diameter) is analyzed. Also the definitions of mean transmission and mean of adjusted betweenness centrality coincide, so only one of them (mean transmission) is analyzed. The minimum distance is not analyzed because it is clearly always equal to 1. Therefore, the amplitude of distance is not analyzed either, given that the minimum distance is fixed and the maximum distance is already analyzed. With these considerations, the number of invariants analyzed decreases to 315.

The measures in Appendix A are computed using the igraph package (Csardi & Nepusz, 2006) of the software R (R Core Team, 2016). R has good capacity to handle large sets of data, and it is open-source software, freely distributed and recognized in academia and continuously developed and updated by members of academia. The igraph package contains tens of implemented commands for computing graph measures and invariants.

The computed topological invariants result in a matrix with 315 columns (number of invariants) by $2.2 \times 10^6 - 52$ rows (number of graphs for which λ is computed).[3] This matrix is not presented here due to size issues but is further analyzed in Section 3.

Table 1
Some basic parameters of the 15 real-world networks analyzed.

Real networks	n	m	λ	Mean degree
BREN	10	11	12	2.2
RNP	10	12	13	2.4
CESNET	12	19	15	3.2
VBNS	12	17	19	2.8
NSFNET	14	21	13	3.0
AUSTRIA	15	22	18	2.9
MZIMA	15	19	30	2.5
ARNES	17	20	38	2.4
GERMANY	17	26	24	3.1
SPAIN	17	28	22	3.3
CANARIE	19	26	44	2.7
EON	19	37	17	3.9
LAMBDARAIL	19	23	58	2.4
MEMOREX	19	24	48	2.5
ARPANET	20	32	33	3.2

2.2. Real networks

In addition to the random networks cited in Section 2.1, a set of 15 real networks is considered for analysis in combination with the random networks, with the purpose of addressing the characteristics of both groups. This set is part of a larger set of 29 real networks used in Cousineau et al. (2015), and these 15 are selected because they have the same number of nodes, between 10 and 20, and same mean degree, between 2 and 4, as in the random sample generated.

The real networks are also represented by graphs, and for each one, λ is also computed using the same method applied to the random networks. All measures in Appendix A are also computed for the real networks, resulting in a matrix with 315 columns (number of invariants) and 15 rows (number of real networks), which is not presented here, also due to their size. However, it is analyzed in Section 3, along with the measures of the random networks. Table 1 shows some basic parameters of the 15 real networks selected for analysis.

2.3. A method to rank invariants: the method of coincidence

With the description of networks in the Sections 2.1 and 2.2 and the delimitation of measures in Appendix A, we now turn to the explanation of the method proposed to choose the best invariants for designing optical networks.

[1] Given by the median of the differences of all measure values in relation to their median.

[2] Given by the ratio of the value obtained in (ix) and the median of the measure, multiplied by 100.

[3] The total time spent in the computation of all topological invariants for all graphs is approximately 331 h using an Intel Xeon Processor E5-2430 v2 machine with 96GB DDR3 of RAM.

To understand how it works, consider the analysis of the following simple example: a fair die with C faces, where C is unknown. Even intuitively, if an observer has access to random and independent observations of the throw of this die, the observer can attempt to infer C using the enumeration of different symbols observed. For instance, in 10 throws of a die with letters drawn on the faces, the following faces are observed: l, u, p, o, w, j, t, t, l, and p, where labels are arbitrarily associated to faces, not following any alphabetic order, for instance. Based on the symbols observed, at first sight, it seems that the best thing this intuitive observer can do is to infer that the die has at least 8 faces.

In fact, the observer can do much better, although less intuitively: noting that the symbols t and p appear twice, the observer could count the occurrence of 2 coincidences, one for p and another for t. Of course, to compute the two coincidences, the observer has to, somehow, read the list of letters and compare each one of the 10 symbols to the other 9. With repetitive comparisons, there is a total of $(10 \times 9)/2 = 45$ comparisons; i.e., in 45 comparisons, 2 coincidences are observed. Analogously, it is known that, for another fair die with C faces, if only one of the faces is labeled “coincidence” and the other C − 1 faces “non-coincidence”, on average, it should be expected one occurrence of “coincidence” face every C moves. Hence, C could be estimated as $\hat{C} = 45/2 = 22.5$ faces.

It should be noted that the estimation of approximately 22 faces is based on only 10 dice throws and therefore on a number of throws smaller than the number of faces, which is somehow counterintuitive.

Although its presentation is much simplified, this principle is the essence of the method of coincidence proposed and used by Ma (1981) in the context of Statistical Mechanics, where the coincidences are defined in terms of particle trajectory and movement, as explained by Ma (1985), in Chapter 25.

One remarkable fact is that, when used with unfair dice, this method does not estimate the real number of faces but a smaller number, which can be understood as the number of effective faces or the number of faces that a fair die should have to cause the same coincidence rate. In this line of interpretation, the effective number of faces, denoted as C, can also be understood as the number of equiprobable states of a system, which can also be called the “effective cardinality” of the set of symbols, from which the random outputs observed in an experiment are drawn. In terms of information theory, the quadratic (or collision) entropy, denoted as H, associated with the random source (or, in other words, the measurement of uncertainty about which symbol is observed as output of an experiment of random nature) is simply given by the logarithm of C. Hence,

$$H = log(C).$$

In the previous illustration, the observation {l, u, p, o, w, j, t, t, l, p} clearly corresponds to a phenomenon that is discrete and finite in nature because the symbols correspond to the random draw of some of the C faces of the die. In the case of continuous random variables, there is no sense in estimating the absolute cardinality, as it always diverges to infinity. This point also makes sense in the coincidence method because, among the observations in a continuous scale, the probability of an exact coincidence tends to zero. The estimated cardinality is defined as:

$$\hat{C} = \frac{\text{number of comparisons}}{\text{number of coincidences}}.$$

Analogously, the estimated entropy is given by

$$\hat{H} = log(\hat{C}).$$

Hence, when the number of coincidences tends to zero, both $\hat{C}$ and $\hat{H}$ tend to infinity.

However, the method of coincidence can still be used with an appropriate redefinition of the meaning of coincidence for continuous measurements. For instance, it would be reasonable to assume that two continuous measurements are coincident if the absolute difference between them is smaller than a given range, called coincidence neighborhood, and denoted as Δ. In this case, $\hat{C}$ also depends on the Δ selected in the coincidence definition. When Δ tends to zero, the number of coincidences also decrease, so $\hat{C}$ tends to infinity. In order to avoid it, $\hat{C}$ is redefined as delta cardinality, $\hat{C}_\Delta$, given by:

$$\hat{C}_\Delta = \hat{C} \times \Delta.$$

When Δ tends to zero, $\hat{C}$ tends to infinity, whereas $\hat{C}_\Delta$ may converge to a real finite value if there are sufficient data.

Analogously, given that $\hat{H}$ also diverges to infinity for small Δ, a differential entropy $\hat{H}_\Delta$ is defined as

$$\hat{H}_\Delta = log(\hat{C}_\Delta) = log(\hat{C} \times \Delta),$$

which can be rewritten as

$$\hat{H}_\Delta = log(\hat{C}) + log(\Delta).$$

In the case of multidimensional observations, as in the case of the associations of different invariants, the single adaptation required for the method of coincidence is a proper definition of Δ, which can be as simple as a hypercube, for instance. The selection of the volume of coincidence should be performed with caution, as it is the most sensitive aspect of the method of coincidence. The end of this section deal with it this selection.

To explain how the measurement of H applies to estimate the number of wavelengths, it is convenient to resume the analogy with dice. In this case, two dice are considered, X and Y, one with faces {a,b,c,d}, and the other with faces {e,f}. The effective cardinalities of X and Y, C_X and C_Y, are not necessarily 4 and 2, as it could initially be expected. This statement is only true if the dice are fair (equiprobable faces). In fact, the effective cardinalities C_X and C_Y can be efficiently computed through the method of coincidence.

When throwing the two dice, the result is $4 \times 2 = 8$ possible pairs of observations. However, again, the effective joint cardinality of the pairs of observations, C_{XY}, for the same reason, not necessarily is 8. Here, C_{XY} can also be estimated using the coincidences method. Through these effective cardinalities, the mutual influence between the two dice can be measured easily because, if the dice are independent, C_{XY} should be equal to the product of the cardinalities $C_X \times C_Y$, whereas the joint cardinality, C_{XY}, always is smaller than the product of the individual cardinalities C_X and C_Y in the case of dependence between the variables (the result of die X depends on the result of die Y, and vice versa). That is, the relative cardinality is given by

$$C_R = \frac{C_X \times C_Y}{C_{XY}}.$$

C_R can assume values greater than or equal to 1, and $C_R = 1$ means independence between the dice. In turn, the higher the value of C_R is, the higher is the dependency relationship between X and Y. In Information Theory, $log(C_R) = I(X;Y)$ is called the mutual information between the random variables X and Y, where

$$\begin{aligned} I(X;Y) &= log(C_R) \\ &= log(C_X) + log(C_Y) - log(C_{XY}) \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

As C_R assumes values greater than or equal to 1, consequently $I(X;Y) = log(C_R)$ assumes values greater than or equal to zero.

As explained above, only the redefinition of the coincidence is necessary in the case of continuous variables, whereas everything else remains the same because the ranges, volumes, and hypervolumes used in the definition of coincidence cancel themselves in the computation of C_R.

Table 2
The 30 invariants which reduce at least by half the uncertainty of number of wavelengths for the $2.2 \times 10^6 - 52$ random networks analyzed. These invariants are sorted as a function of the mean $I(\lambda;k)$, where each one is represented by an identifier "k". Appendix B presents a table with all 315 invariants under study, also sorted by mean $I(\lambda;k)$.

k	Invariants	$I(\lambda;k)$			
		Min	Mean	Median	Max
1	Maximum of edge betweenness	4.92	4.9974	5.00	5.08
2	Standard deviation of edge betweenness	3.57	3.6446	3.65	3.70
3	Amplitude of edge betweenness	3.47	3.5395	3.54	3.61
4	Kirchhoff index	3.18	3.2476	3.25	3.32
5	Mean of edge betweenness	3.15	3.2081	3.21	3.28
6	Mean of vertex betweenness	3.02	3.0701	3.07	3.13
7	Mean distance	2.75	2.7981	2.80	2.84
8	Mean transmission	2.72	2.7639	2.76	2.81
9	Maximum transmission	2.70	2.7470	2.75	2.80
10	Wiener index	2.64	2.6837	2.68	2.73
11	Median transmission	2.63	2.6797	2.68	2.73
12	Coefficient of variation of communicability distance	2.57	2.6202	2.62	2.67
13	Median of edge betweenness	2.53	2.5743	2.57	2.62
14	Harary index	2.47	2.5154	2.52	2.58
15	Standard deviation of transmission	2.31	2.3529	2.35	2.40
16	Mean eccentricity	2.30	2.3411	2.34	2.37
17	Amplitude of transmission	2.30	2.3351	2.34	2.40
18	Coefficient of variation of modularity	2.26	2.3048	2.31	2.34
19	Maximum of adjusted betweenness centrality	2.26	2.2940	2.29	2.34
20	Minimum transmission	2.25	2.2909	2.29	2.33
21	Standard deviation of vertex betweenness	2.19	2.2215	2.22	2.26
22	Standard deviation of adjusted betweenness centrality	2.19	2.2215	2.22	2.26
23	Maximum of vertex betweenness	2.19	2.2198	2.22	2.26
24	Vertex betweenness centrality	2.15	2.1793	2.18	2.22
25	Mean of communicability distance	2.15	2.1771	2.18	2.20
26	Amplitude of vertex betweenness	2.12	2.1508	2.15	2.19
27	Amplitude of adjusted betweenness centrality	2.12	2.1508	2.15	2.19
28	Vertices number	2.10	2.1363	2.14	2.17
29	Interquartile distance of edge betweenness	2.06	2.0841	2.08	2.11
30	Standard deviation by median of edge betweenness	2.01	2.0379	2.04	2.07

The selection of Δ performed in this study is related to the idea proposed and justified in Montalvão et al. (2014), where the coincidence neighborhoods are defined as a function of the standard deviation of the variable (σ) and empirical values. The coincidence neighborhoods of the number of wavelengths (Δ_λ) and the invariant k (Δ_k), with $k = 1,..., 315$, are given by

$$\Delta_\lambda = 0.2 \times \max\{\sigma(\lambda), 1\}$$

and

$$\Delta_k = 0.2 \times \max\{\sigma(k), 1\}.$$

In turn, the coincidence neighborhoods for the effective joint cardinality of the pairs of observations ($\Delta_{\lambda,k}$), to combine λ and each k, is given by

$$\Delta_{\lambda,k} = 0.2 \times \max\{\sigma(k), \sigma(\lambda), 1\}.$$

Now, after this explanation about the method of coincidence, it is possible to see it in action in next section, by computation of mutual information $I(X, Y)$ with $Y = \lambda$ in function of topological invariants X.

3. Results and discussion

3.1. All networks of all orders together

The mutual information, $I(\lambda;k)$, is estimated for all $k = 1,..., 315$ invariants, with all random networks described in Section 2.1, using the methodology proposed in Section 2.3. To make this computation feasible for such a large set of data as the one in this study, the entire data matrix with 315 columns (number of invariants) by $2.2 \times 10^6 - 52$ rows (number of networks) is divided into 100 equal parts ($S_1, S_2, ..., S_{100}$) through systematic separation. The lines are numbered from 1 to $2.2 \times 10^6 - 52$, where line i deals with invariants of the network i (N_i). Thus, the 100 parts are taken in such a way that:

$$S_1 = \{N_1, N_{101}, N_{201}, \ldots\};$$
$$S_2 = \{N_2, N_{102}, N_{202}, \ldots\};$$
$$\vdots$$
$$S_{100} = \{N_{100}, N_{200}, N_{300}, \ldots\}.$$

Then, $I(\lambda;k)$ is computed for each of the 315 invariants in each of the 100 parts of the matrix. This process needs a shorter computational time than executing the computation with all rows, i.e., with all graphs at once.[4] Next, the mean of the 100 values of $I(\lambda;k)$, obtained for each of the 100 parts of the sample, is adopted to obtain a single mutual information value for each invariant. The median, maximum, and minimum are also adopted. A complete table with all these results can be seen in Appendix B.

If an invariant k and the number of wavelengths λ have mutual information in the range from 1 to 2, i.e.,$1 \leq I(\lambda;k) \leq 2$, the knowledge of values of the invariant k can reduce the uncertainty associated with λ by up to half. More precisely, the reduction by half occurs for $I(\lambda;k) = 2$. Among the 315 invariants analyzed, the ones that reduce the uncertainty associated with λ by at least one half are selected as relevant, i.e., the invariants with $I(\lambda;k) \geq 2$.

Table 2 lists the only 30 of 315 invariants which achieve $I(\lambda;k) \geq 2$ for the mean and the median, in decreasing order. The invariants that achieve a mean of $I(\lambda;k)$ greater than or equal to

[4] The total time spent in the computation of the mutual information, in R, for all invariants, using all of their values in all graphs, is approximately 22 h, also using Intel Xeon Processor E5-2430 v2 machine, with 96GB DDR3 of RAM.

two also achieve the same for the median, and vice versa. Still, according to Table 2, the minimum, mean, median and maximum values of the mutual information vary little from one sample to the other, even when a large volume of data is analyzed. Hence, only the mean of the mutual information of each invariant is considered for sorting effect, and the identifier k is assigned to the invariants according to this sorting, in which this assignment is considered from now on.

According to Table 2, the invariant that stood out from all is, by far, the maximum edge betweenness ($k = 1$). The edge that reaches maximum edge betweenness is the edge most demanded by the paths, so it is the edge with the largest congestion. Edge and vertex betweenness, that measure how much links and nodes are used to meet any traffic demand, are called congestion measures.

Fig. 2 shows results about maximum edge betweenness. Fig. 2a presents maximum edge betweenness versus the number of wavelengths for both random and real-world sets of networks under study. The same results separated by number of nodes are presented in Fig. 2b, as a way to confirm or to demystify behaviors observed in Fig. 2a. In both these figures, all data are presented including extremal wavelength requirements found, instead of averages. Also, the black lozenges represent real networks results. In Fig. 2c, it is shown the mutual information of maximum edge betweenness and number of wavelengths for the set of random networks under study, in function of network order. For each invariant in Table 2, these triplet graphics are presented in Appendix D.

The next two invariants in Table 2 are also related to edge congestion: the standard deviation ($k = 2$) and the amplitude ($k = 3$) of the edge betweenness. Given that these first three invariants affect directly the network wavelength requirements, it is important for optical network design to avoid excessive overload of edges. Other invariants of Table 2 which depend on edge congestion are $k = 5$, 13, 29 and 30. Thus, edge congestion shows to be an important factor in this study. Vertex congestion also shows to be important, as invariants $k = 6$, 19, 21, 22, 23, 24, 26, and 27 appear in Table 2. Because the last ones occupy lower positions in the table, it suggests that edge congestion is more critical than vertex congestion considering the number of wavelengths.

Notably, the transmission also has a representative presence, with the invariants $k = 8$, 9, 11, 15, 17 and 20 in Table 2, which highlights the importance of vertices that are good transmitters, i.e., whose sum of all distances to the other vertices is minimized. For this reason, the topological design phase should optimize this invariant. Along with the transmission, the communicability distance, represented by its coefficient of variation ($k = 12$) and its mean ($k = 25$), the Kirchhoff ($k = 4$), Wiener ($k = 10$), and Harary ($k = 14$) indices, the mean distance ($k = 7$) and the mean eccentricity ($k = 16$) emphasize the dependence of the number of wavelengths on the distance between vertices. About the modularity, it is present just with one invariant: its coefficient of variation ($k = 18$), which indicate that a more regular modularity matrix is desirable.

Finally, it is interesting to note the presence of the invariant vertices number ($k = 28$) in Table 2. The vertices number versus the number of wavelengths for all real-world and random networks under study is shown in more details is Fig. 3. According to Baroni and Bayvel (1997), the mean λ requirement is practically independent of the network order. However, our results show that the network order actually is important for both sets of networks. Larger networks can require more wavelengths than smaller networks, particularly if they have low mean degrees, as observed in Fig. 3.

In summary, it could be inferred that a good practice for the design of optical network topologies is to give as much attention as possible to congestion, with the purpose of minimizing extremes in the edge and vertex betweenness, and to make the network

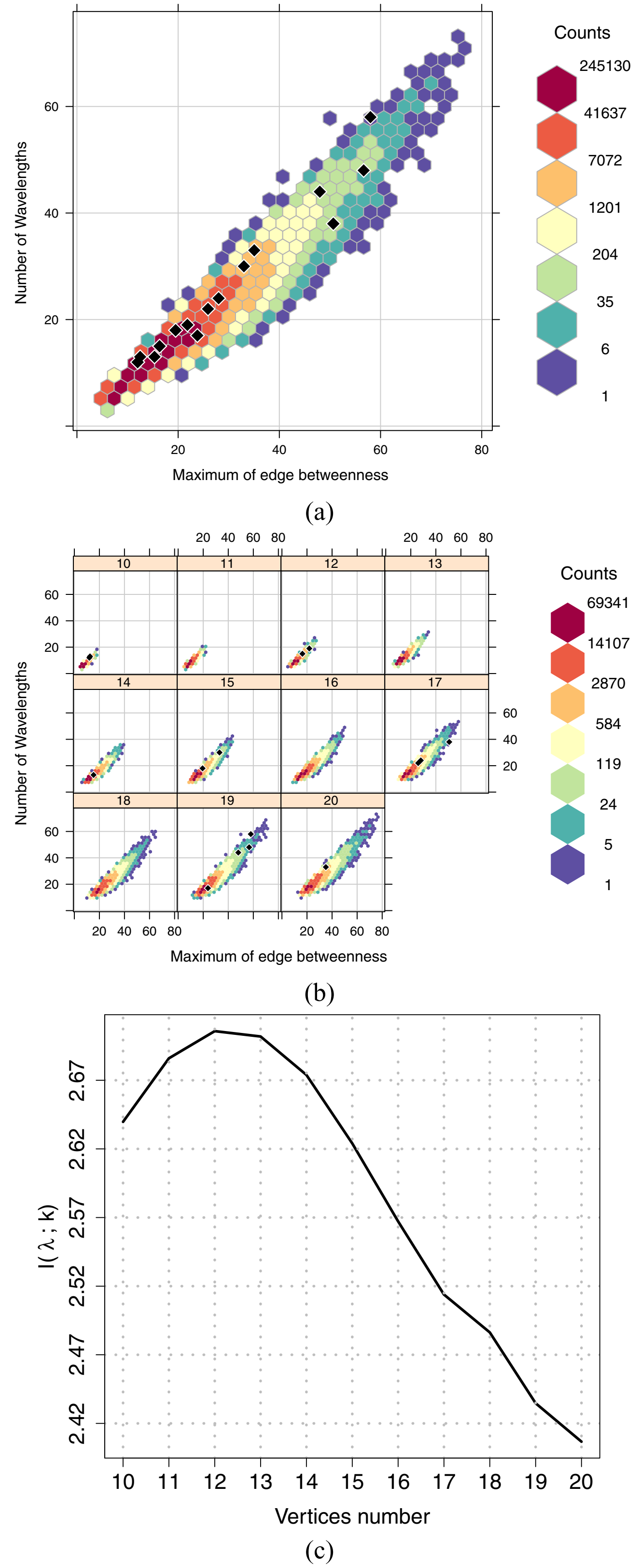


Fig. 2. Maximum of edge betweenness ($k = 1$) versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders (a), and separated by n (b). In (c), mutual information of Maximum of edge betweenness and λ, as a function of network order.

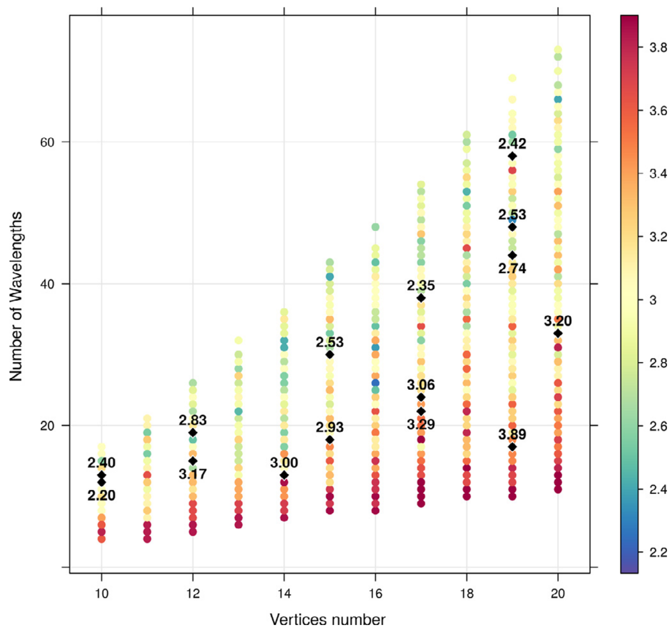


Fig. 3. Vertices number ($k = 28$) versus λ, for all real-world and random networks under study. The colors represent the mean degree, and the number with each lozenge is the mean degree of the respective real-world network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vertices good transmitters, i.e., to minimize the sum of distances of each vertex to any other.

3.2. Networks separated by order

Next, the total sample of random graphs is divided into 11 samples, where the graphs in the same sample have the same vertices number. The computation of $I(\lambda;k)$ is performed in each of these samples to verify (i) if, for each n, the relevance of the invariants could be different when compared to the one obtained with graphs of all orders together (seen in the Section 3.1) and (ii) if the relevance of the invariants changes with increased or decreased network order. Similar to the entire sample, each sample (of the 11) is also divided into 100 parts, and the mean is computed to obtain one value of $I(\lambda;k)$ for each $k = 1, ..., 315$.

The result of this process can be seen in Table 4, where the results are presented for each n independently. As observed, few invariants achieve $I(\lambda;k) \geq 2$, and therefore, the first 30 places are considered for each n, i.e., the same amount of invariants used the Table 2. A table with the mutual information of all topological invariants for all networks separated by n can be found in Appendix C, whereas the results are graphically presented in Appendix D, as it is done for the invariants of Table 2.

In Table 4, the clear cells indicate invariants that appear, with lower or higher value of mutual information, in all columns. In turn, the colored cells are the ones that do not appear in at least one column. Some cells, colored or clear, do not vary much in position among the columns as, for instance, the invariant $k = 40$. In turn, some invariants change their positions substantially, such as the invariant $k = 13$, which appears among the 30 most important invariants only in small networks.

According to Table 4, many invariants which derive from the edge betweenness are also relevant for each n, especially the maximum edge betweenness, which even is the most important in all cases of n that are individually analyzed. Fig. 2b indicates that the directly proportional relation of the maximum edge betweenness to λ is maintained for each n. However, in Fig. 2c, the mutual information value $I(\lambda;k)$ decreases with increasing n.

Table 4 shows many invariants that are also present in Table 2: $k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 25, 29$, and 30. Therefore, 19 invariants out of the 30 in Table 2 continue to be representative in the explanation of the number of wavelengths. These 19 invariants are located in the upper positions of Table 4 and are related to congestion and distance. There are also the invariants $k = 45, 55, 60, 93, 115$, and 126, which address congestion and distance issues and appear in Table 4 but not in Table 2. These facts confirm that congestion and distance are topics that deserve special attention when designing optical network topologies. Notice however that among all congestion invariants in Table 4, only one is related to vertex congestion ($k = 6$), whereas 5 invariants ($k = 6, 21, 23, 24$, and 26) appear in Table 2. It means that vertex congestion is not so relevant as edge congestion in the context of network topology design.

Among the 19 invariants mentioned above, there are 14 ($k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14$, and 16) which appear with high importance for all n, and the remaining 5 ($k = 13, 15, 25, 29$, and 30) appear in one or more (but not all) columns in Table 4. Analyzing all network orders separately, these 14 invariants prove to be more robust, keeping roughly the same level of importance for each n. Among the invariants that do not appear in all columns, $k = 13$, and 25 are important in smaller networks and $k = 15, 29$, and 30 do not have a clear pattern.

Table 3
Linear correlation matrix of the 14 most representative invariants, i.e., the ones which appear in Table 2 and in all columns of Table 4.

Invariants	k	1	2	3	4	5	6	7	8	9	10	11	12	14	16
Maximum of edge betweenness	**1**	1.00	0.96	0.98	0.91	0.91	0.91	0.91	0.88	0.89	0.85	0.88	-0.87	0.73	0.86
Standard deviation of edge betweenness	**2**		1.00	0.97	0.85	0.85	0.85	0.89	0.81	0.83	0.77	0.81	-0.80	0.63	0.86
Amplitude of edge betweenness	**3**			1.00	0.85	0.84	0.84	0.85	0.82	0.83	0.79	0.81	-0.79	0.67	0.82
Edges number	**4**				1.00	0.98	0.99	0.92	0.99	0.98	0.98	0.98	-0.97	0.89	0.84
Mean of edge betweenness	**5**					1.00	0.98	0.97	0.96	0.94	0.93	0.96	-0.98	0.82	0.90
Mean of vertex betweenness	**6**						1.00	0.94	0.99	0.98	0.98	0.99	-0.97	0.90	0.87
Mean distance	**7**							1.00	0.90	0.90	0.85	0.90	-0.94	0.71	0.96
Mean transmission	**8**								1.00	0.99	0.99	1.00	-0.97	0.94	0.83
Maximum transmission	**9**									1.00	0.98	0.98	-0.95	0.92	0.85
Wiener index	**10**										1.00	0.99	-0.94	0.96	0.77
Median transmission	**11**											1.00	-0.97	0.94	0.82
Coefficient of variation of communicability distance	**12**												1.00	-0.87	-0.85
Harary index	**14**													1.00	0.62
Mean eccentricity	**16**														1.00

Table 4

The 30 more important invariants according to $I(\lambda\ ;k)$, for the $2.2 \times 10^6 - 52$ random networks analyzed, separated by network order. Clear cells indicate invariants presented in all columns. In turn, colored cells represent invariants that do not appear in at least one column. Appendix C presents a table with $I(\lambda;k)$ of all 315 invariants studied, separated by n.

Order	n = 10		n = 11		n = 12		n = 13		n = 14		n = 15		n = 16		n = 17		n = 18		n = 19		n = 20	
	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)
1	1	2.6240	1	2.6860	1	2.7058	1	2.7019	1	2.6739	1	2.6240	1	2.5672	1	2.5139	1	2.4863	1	2.4347	1	2.4067
2	2	2.2566	40	2.1655	40	2.1882	2	2.2270	2	2.2543	2	2.2566	2	2.2636	2	2.2649	2	2.2676	2	2.2588	2	2.2500
3	40	2.2131	2	2.1082	2	2.1600	40	2.2221	40	2.2310	40	2.2131	40	2.1902	40	2.1707	40	2.1494	40	2.1276	40	2.1058
4	3	2.0277	4	1.9846	4	1.9947	3	2.0123	3	2.0216	3	2.0277	3	2.0269	3	2.0213	3	2.0200	3	2.0069	3	2.0032
5	4	1.9584	3	1.9226	3	1.9628	4	1.9976	4	1.9880	4	1.9584	4	1.9520	4	1.9431	4	1.9434	4	1.9088	4	1.9205
6	10	1.8414	10	1.8756	10	1.8879	10	1.8786	10	1.8770	10	1.8414	10	1.8261	10	1.8106	10	1.8190	10	1.7753	10	1.7796
7	8	1.8414	8	1.8756	8	1.8879	8	1.8786	8	1.8770	8	1.8414	8	1.8261	8	1.8106	8	1.8190	8	1.7753	8	1.7796
8	6	1.8414	6	1.8517	6	1.8712	6	1.8757	6	1.8770	6	1.8414	6	1.8261	6	1.8106	6	1.8190	6	1.7753	6	1.7796
9	14	1.7399	5	1.8372	14	1.8179	14	1.7929	14	1.7740	14	1.7399	14	1.7209	14	1.6959	14	1.7191	14	1.6697	9	1.6746
10	5	1.6754	14	1.7420	5	1.7921	5	1.7374	5	1.7263	5	1.6754	5	1.6704	9	1.6466	9	1.6593	9	1.6570	14	1.6677
11	9	1.6259	7	1.5971	7	1.6193	7	1.6323	9	1.6155	9	1.6259	9	1.6379	5	1.6438	5	1.6569	5	1.6122	5	1.6188
12	7	1.6139	59	1.5813	11	1.6112	9	1.6040	11	1.6115	7	1.6139	7	1.6144	7	1.6070	7	1.6087	16	1.5901	7	1.6015
13	16	1.5527	53	1.5497	59	1.5843	126	1.5738	7	1.6067	16	1.5527	11	1.5905	16	1.5840	16	1.6043	7	1.5859	57	1.5879
14	11	1.5516	52	1.5417	9	1.5630	59	1.5651	59	1.5357	11	1.5516	16	1.5699	11	1.5451	11	1.5990	11	1.5364	63	1.5862
15	59	1.5283	167	1.5417	52	1.5197	11	1.5446	16	1.5284	59	1.5283	101	1.5152	101	1.4972	101	1.5177	57	1.5010	61	1.5861
16	126	1.5279	12	1.5417	167	1.5197	16	1.4936	101	1.4973	126	1.5279	59	1.5134	59	1.4853	93	1.4881	101	1.5003	68	1.5859
17	101	1.4916	11	1.5405	12	1.5197	52	1.4903	52	1.4752	101	1.4916	93	1.4758	93	1.4775	59	1.4793	63	1.4990	16	1.5802
18	93	1.4548	9	1.5208	126	1.5085	167	1.4903	167	1.4752	93	1.4548	52	1.4441	126	1.4762	63	1.4674	61	1.4983	11	1.5745
19	52	1.4493	126	1.4955	101	1.4925	12	1.4903	12	1.4752	52	1.4493	167	1.4441	45	1.4421	61	1.4673	68	1.4982	101	1.5288
20	167	1.4493	101	1.4811	16	1.4685	101	1.4658	93	1.4398	167	1.4493	12	1.4441	115	1.4365	68	1.4660	93	1.4896	93	1.4977
21	12	1.4493	115	1.4700	25	1.4179	93	1.4331	45	1.4395	12	1.4493	45	1.4383	60	1.4301	57	1.4636	59	1.4595	59	1.4644
22	45	1.4408	16	1.4449	115	1.4104	45	1.4268	115	1.4136	45	1.4408	115	1.4331	52	1.4282	45	1.4499	126	1.4568	115	1.4535
23	115	1.4351	25	1.4236	45	1.4100	115	1.4093	126	1.4129	115	1.4351	29	1.3957	167	1.4282	115	1.4480	115	1.4458	45	1.4479
24	29	1.3898	109	1.4166	96	1.4077	29	1.3937	29	1.3960	29	1.3898	35	1.3834	12	1.4282	52	1.4321	45	1.4442	35	1.4162
25	34	1.3736	13	1.4114	13	1.3923	96	1.3845	96	1.3851	34	1.3736	34	1.3796	29	1.3987	167	1.4321	52	1.4076	52	1.4128
26	35	1.3736	46	1.4014	93	1.3854	13	1.3789	53	1.3760	35	1.3736	53	1.3755	35	1.3915	12	1.4321	167	1.4076	12	1.4128
27	96	1.3706	45	1.3871	109	1.3842	30	1.3704	30	1.3749	96	1.3706	30	1.3739	34	1.3834	35	1.4083	12	1.4076	15	1.4082
28	30	1.3671	96	1.3850	53	1.3783	138	1.3685	34	1.3656	30	1.3671	33	1.3722	33	1.3787	29	1.4068	35	1.4024	29	1.4049
29	33	1.3649	36	1.3719	46	1.3767	139	1.3685	35	1.3623	33	1.3649	96	1.3672	30	1.3749	60	1.4003	29	1.3994	167	1.4025
30	15	1.3390	138	1.3685	29	1.3687	34	1.3529	13	1.3607	15	1.3390	55	1.3645	63	1.3730	34	1.3998	60	1.3978	60	1.4011

*__Name of invariants according to *k*:__ **1**: Maximum of edge betweenness; **2**: Standard deviation of edge betweenness; **3**: Amplitude of edge betweenness; **4**: Kirchhoff index; **5**: Mean of edge betweenness; **6**: Mean of vertex betweenness; **7**: Mean distance; **8**: Mean transmission; **9**: Maximum transmission; **10**: Wiener index; **11**: Median transmission; **12**: Coefficient of variation of communicability distance; **13**: Median of edge betweenness; **14**: Harary index; **15**: Standard deviation of transmission; **16**: Mean eccentricity; **25**: Mean of communicability distance; **29**: Interquartile distance of edge betweenness; **30**: Standard deviation by median of edge betweenness; **33**: Coefficient of variation of cocitation coupling; **34**: Coefficient of variation of dice similarity; **35**: Coefficient of variation of inverse log-weighted similarity; **36**: Mean of inverse log-weighted similarity; **40**: Algebraic connectivity; **45**: Standard deviation of distance; **46**: Mean of cocitation coupling; **52**: Edges number; **53**: Degree centralization; **55**: Median eccentricity; **57**: Coefficient of variation by median of dice similarity; **59**: Number of minimum sets that disconnects the graph; **60**: Coefficient of variation of eccentricity; **61**: Coefficient of variation by median of Jaccard similarity; **63**: Coefficient of variation by median of inverse log-weighted similarity; **68**: Coefficient of variation by median of cocitation coupling; **93**: Coefficient of variation of edge betweenness; **96**: Minimum of subgraph centrality; **101**: Minimum of average nearest neighbor degree; **109**: Number of spanning trees; **115**: Coefficient of variation of distance; **126**: Coefficient of variation by median of transmission; **138**: Coefficient of variation by median of closeness centrality; **139**: Coefficient of variation by median of closeness centrality normalized; **167**: Mean of vertex degree.

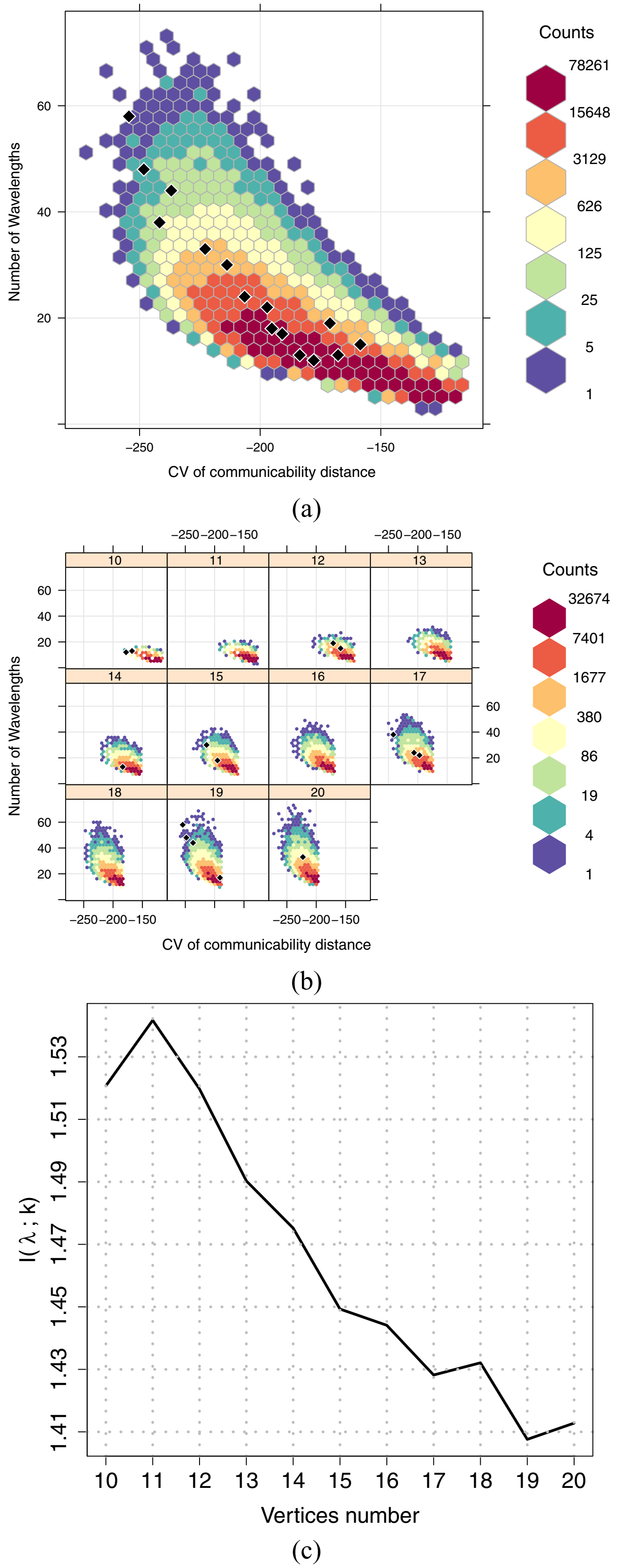


Fig. 4. Coefficient of variation of communicability distance (k = 12) versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders (a), and separated by n (b). In (c), mutual information of Coefficient of variation of communicability distance and λ, as a function of network order.

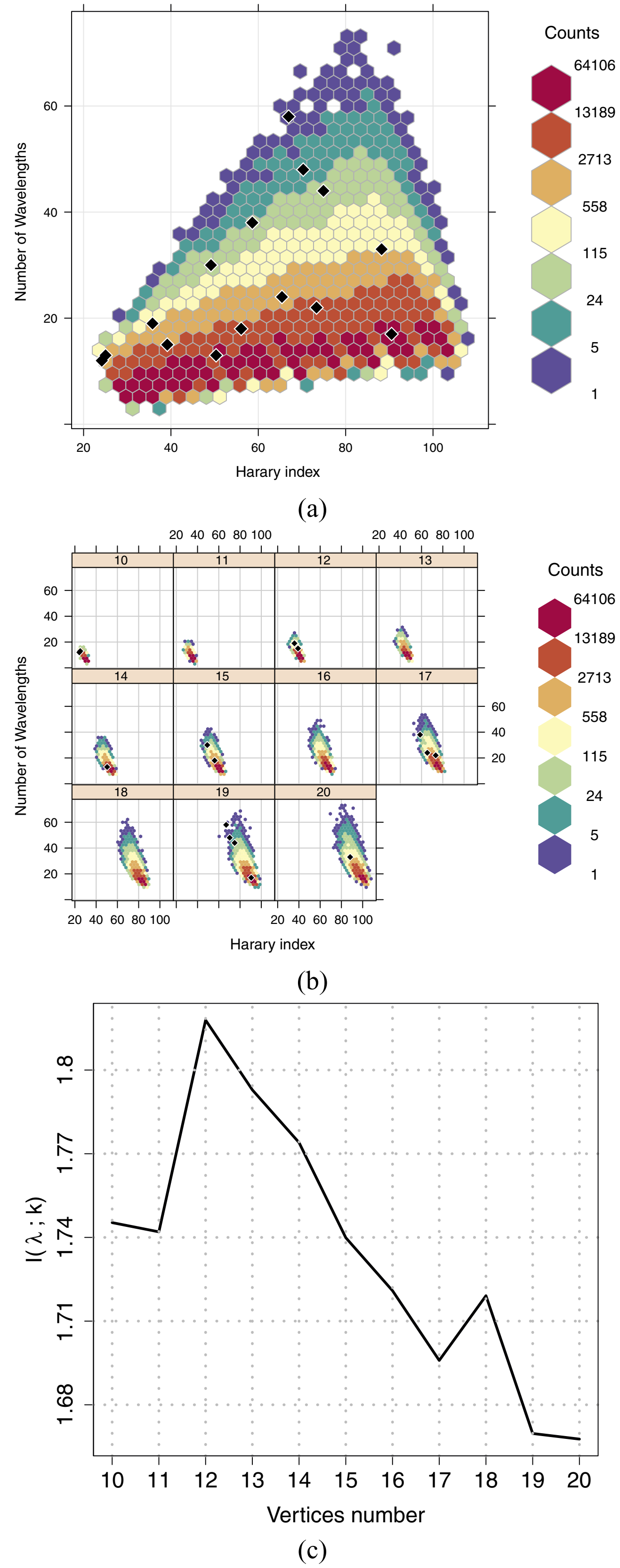
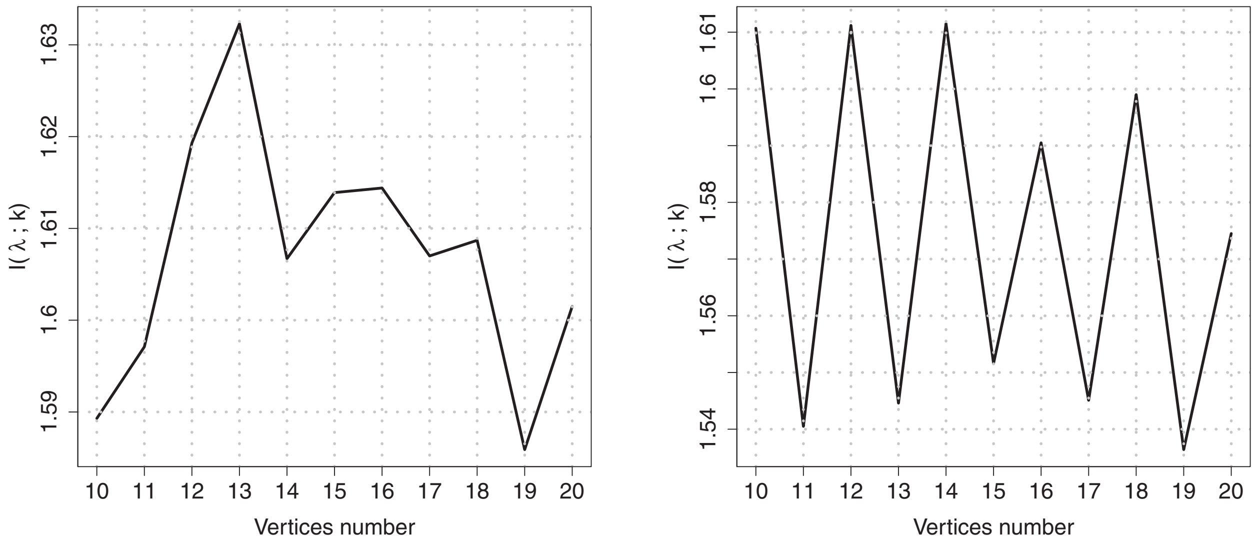


Fig. 5. Harary index (k = 14) versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders (a), and separated by n (b). In (c), mutual information of Harary index and λ, as a function of network order.



(a) (b)

Fig. 6. Mutual Information of (a) Mean distance ($k = 7$) and (b) Median transmission ($k = 11$) both with λ and in function of network order (n).

Table 3 shows a matrix of linear correlation for the 14 invariants mentioned above. Based on this table, it is possible to observe that many of these invariants are strongly correlated, especially ones that derive from the same variable, such as the pair of invariants $k = 8$ and $k = 11$, which derive from the transmission variable, or the pair $k = 1$ and $k = 3$, which derive from the edge betweenness.

When two invariants are both very important (with high mutual information), derived from the same variable, and strongly correlated, we recommend selecting only one of them, in this case, the one with the highest value of mutual information, because it avoids possible redundancies and helps to simplify the choice of invariants. For instance, in the context of optical networks, maybe there is no need to consider both maximum edge betweenness ($k = 1$) and amplitude of edge betweenness ($k = 3$) because they have linear correlation equal to 0.98.

Many other invariants do not appear in Table 2 but appear in Table 4, namely, $k = 33$, 34, 35, 36, 40, 46, 52, 53, 57, 59, 61, 63, 68, 96, 101, 109, 138, 139, 167 in a total of 19 additional invariants. Among them, 5 invariants appear in Table 4 for all network orders: algebraic connectivity ($k = 40$), edges number ($k = 52$), number of minimum sets that disconnects the graph ($k = 59$), minimum average nearest neighbor degree ($k = 101$), and mean degree ($k = 167$). Algebraic connectivity is highlighted, occupying the first positions, whereas the other invariants occupy mid-height positions in Table 4. Therefore, in addition to the edges number, invariants of connectivity and ones referent to the degree of vertices come to appear with some level of importance.

Observations also indicate invariants that become more important with increased network order, which are dispersion measures of log-weighted similarity ($k = 35$, and $k = 63$), dice similarity ($k = 57$), Jaccard similarity ($k = 61$), and cocitation coupling ($k = 68$).

Among the 19 additional invariants, 3 are related to network degree ($k = 53$, 101, and 167), where 2 of them appear in all columns of n: mean degree ($k = 167$) and minimum average nearest neighbor degree ($k = 101$). It shows the importance of the degree for explaining the number of wavelengths, as stated in Fenger et al. (2000). Notice that no degree invariant appears in Table 2.

It is interesting to note also the presence of the invariant edges number ($k = 52$) in Table 4, for all network orders, showing the intuitive importance of this invariant, which is that more edges imply in a lower λ value (e.g., a full-mesh network has the minimum λ, equal to 1).

The invariant algebraic connectivity ($k = 40$), although do not appear in Table 2, it remains among the first positions for all network orders in Table 4, showing that the relation between algebraic connectivity and λ is indeed important, as observed in Châtelain et al. (2009).

By the other hand, the invariant number of spanning trees ($k = 109$) is acclaimed by Fenger et al. (2000) but does not have great importance since it is presented only in two columns of Table 4 ($n = 11$, and 12) and in low positions. Moreover, this invariant does not appear in Table 2.

The invariant number of minimum sets that disconnects the graph - MSD ($k = 59$) seems to be important because it is presented in all columns of Table 4. Number of MSD is related to the network connectivity because a failure in a vertex of a MSD influences the chance of the graph become disconnect.

The dispersion of the closeness centrality is portrayed in two invariants of Table 4 ($k = 138$, and 139), but not in a very representative manner, as it appears only for two network orders and in much lower positions in the columns. The cocitation coupling, which reflects connections of pairs of vertices to a third pair, is represented by its mean in $k = 46$ and by its dispersion in $k = 33$ and $k = 68$, which also have low positions and appear for few network orders.

The similarity, which portrays how similar are the vertices (in aspects defined by each type of similarity), is presented in Table 4 for the invariants $k = 34$, 35, 36, 57, 61 and 63, all dispersion measurements, with the exception of 36, which is a mean. The invariants with $k = 57$, 61 and 63 seem to gain importance in larger optical networks. In turn, the invariants $k = 34$ and 35 appear in low positions, and the invariant with $k = 36$ only appear for $n = 11$. Therefore, for larger networks or ones with a tendency to grow, it is interesting to consider the invariants with $k = 57$, 61 and 63.

The invariant minimum of subgraph centrality ($k = 96$) is more representative for small networks, losing importance as n increases. The subgraph centrality of a vertex measures the number of subgraphs a vertex participates in.

Finally, there are invariants that appear in Table 2 but not in Table 4. Clearly, the invariant vertices number ($k = 28$) does not

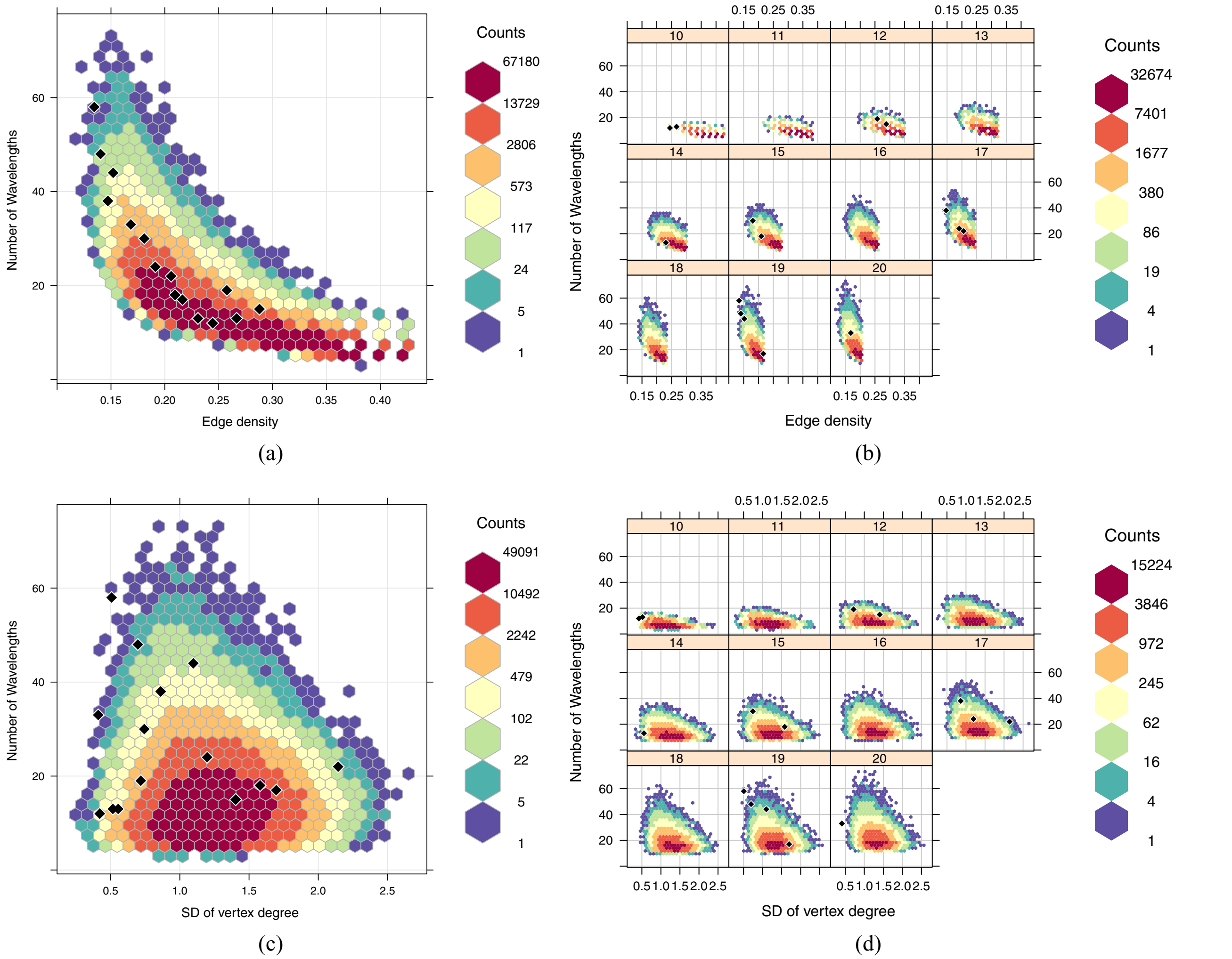


Fig. 7. Edge density ($k = 65$) and Standard deviation of vertex degree ($k = 211$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (c), and separated by n in (b) and (d).

appear in Table 4, given that it is used to separate the sample of networks in columns by n. Some invariants based on amplitude ($k = 17$, 26 and 27) also become less important in Table 4, most likely because the analyses are restricted to each network order, which could have generated amplitudes with less dispersion. The only invariant that involves the modularity ($k = 18$) appears in Table 2 but not in Table 4, showing a low importance of this invariant in this study.

In summary, Table 2 shows, and Table 4 confirms, the importance of considering congestion and transmission variables for the design of optical networks. According to Table 4, more attention should be given to the edge congestion, and it is recommended to consider the mean degree together with the edges number; the connectivity invariants algebraic connectivity and the number of minimum sets that disconnects the graph; and some similarity factors ($k = 57$, 61, and 63).

3.3. Relations between λ, k and $I(\lambda;k)$

When further analyzing the relationships of some invariants in relation to λ, on the one hand, the maximum edge betweenness ($k = 1$) tends to be directly proportional to λ as for all n together (Fig. 2a) as for each n separately (Fig. 2b), making it interesting to have small values for this invariant in optical networks. On the other hand, the coefficient of variation of communicability distance ($k = 12$) tends to be inversely proportional to λ (Fig. 4a, and b), so high values of the invariant $k = 12$ are interesting for the attempt to minimize λ. Notice however that the relation of the invariant $k = 12$ with λ is weaker for networks separated by n (Fig. 4b) than for all network orders together. In addition, the relation of invariant $k = 12$ with λ is weaker than the relation of invariant $k = 1$ with λ. Furthermore, as for the maximum edge betweenness, the importance of the coefficient of variation of communicability distance decreases with network growth (see Figs. 2c, and 4 c).

One invariant that deserves attention is the Harary index ($k = 14$). Fig. 5a may suggest a directly proportional behavior of invariant $k = 14$ with λ, even if weakly so. However, analysis of Fig. 5b indicates the opposite, i.e., an inversely proportional behavior for each n, which demonstrates the importance of considering analyses for each n. Fig. 5c indicates a decrease in the importance of the Harary index, as a function of network growth, for explaining λ.

It is possible to verify the trend of directly or indirectly proportional relation of each invariant of Tables 2 and 4 with λ by analyzing the graphics in Appendix D.

Some invariants have a relatively clear tendency of mutual information as a function of network order, such as the invariants $k = 1$, 12, and 14, seen in Figs. 2c, 4 c, and 5 c. However, some

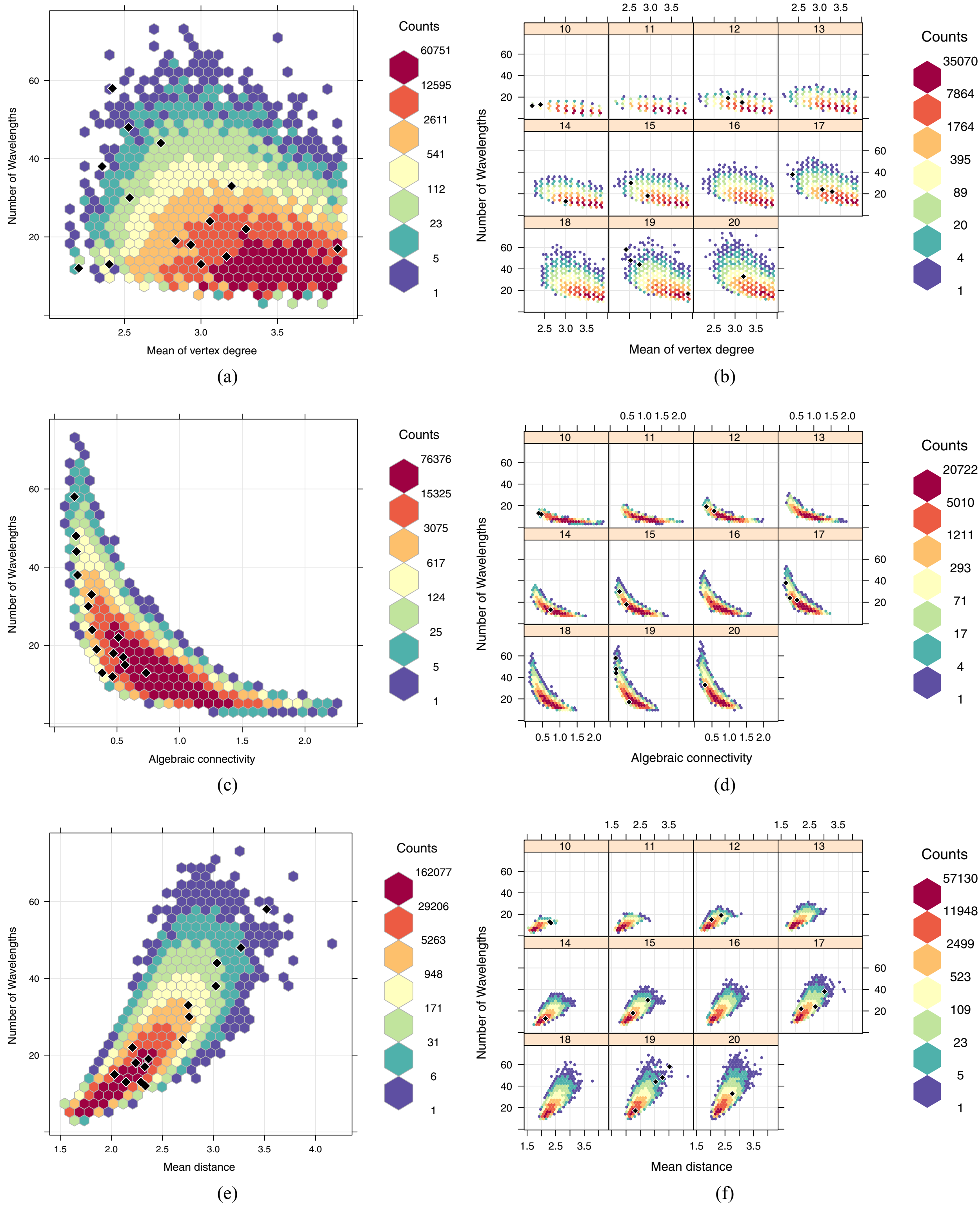


Fig. 8. Mean degree ($k = 167$), Algebraic connectivity ($k = 40$) and Mean distance ($k = 7$) all versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a), (c), and (e), and separated by n in (b), (d), and (f).

invariants do not demonstrate a predictable behavior in relation to their importance, for instance, the invariant mean distance ($k = 7$), for which the mutual information versus n can be seen in Fig. 6a.

For some invariants, the mutual information exhibits an intriguing "hand-saw" behavior, where its relevance seems to depend on the parity of the network order. One such invariant is the median transmission ($k = 8$), for which it can be seen in Fig. 6b much higher values of $I(\lambda, k)$ for even n than for odd n. As a consequence, in optical network design, it is more important to consider the median transmission for networks with even n than for odd n. The reason for this behavior is not clear and requires further analysis.

3.4. Contrasting our results with the findings in the literature

In Sections 3.1 and 3.2 it is analyzed the most important invariants with respect to wavelength requirements, for all real-world and random networks under study for all n together and each n separated, respectively. It is important to note that some invariants acclaimed in the literature (as seen in Section 1.3) do not appear in our study with the expected relevance. Then, in this section, we present results for these invariants when applying our methodology, as a way to see how they fit into the context of our study.

In Baroni and Bayvel (1997), it is stated that number of wavelengths does not depend on the number of nodes. However, as discussed in Section 3.1, the vertices number ($k = 28$) is actually important. Also in Baroni and Bayvel (1997), it is found that the edge density is strongly inversely proportional to λ. This relation does not seem to be very strong in our study, since edge density ($k = 65$) does not appear Table 2 nor in Table 4. However, the statement of Baroni and Bayvel (1997) about edge density is somehow confirmed when analyzing, more generally, all networks together in Fig. 7a. In turn, looking closer, for the networks separated by vertices in Fig. 7b, this strong relation disappears. Thus, this strong relation observed in Fig. 7a is actually caused by the combination of the weak and diverse-behavior relations presented in Fig. 7b.

According to the study by Fenger et al. (2000), there is a directly proportional relation between the variance degree and λ. In contrast, according to our observations, both in all networks together (Fig. 7c) and in the networks separated by order (Fig. 7d), there is no such relation. Notice however that, in our study, instead of analyzing the variance degree, we consider the standard deviation degree ($k = 211$). Since variance and standard deviation are proportional, it does not affect the analyses significantly. Also in Fenger et al. (2000), it is found an inversely proportional relation between mean degree and λ. But, this behavior is not clearly detected when analyzing the networks together (Fig. 8a) or separated by n (Fig. 8b). When analyzing the $I(\lambda;k)$ of the networks separated by n, the mean degree ($k = 167$) appears in all columns of Table 4, but with low importance. Therefore, it could be said that the importance of the mean degree is not as high as supposed. Furthermore, in Fenger et al. (2000) it is established a strong relation between the number of spanning trees and the number of wavelengths. However, as discussed in Section 3.2, the number of spanning trees ($k = 109$) is actually not so important as expected.

According to Châtelain et al. (2009), the algebraic connectivity is inversely proportional to λ. As already discussed in Section 3.2, our results about algebraic connectivity ($k = 40$) support this finding, and additionally this relation can be observed in Fig. 8c and d. Notice however that the relations given in these figures do not have the same behavior, and the structure provided in Fig. 8c appears to be a composition of the relations of Fig. 8d.

As observed in the study of Yuan and Xu (2010), the mean distance is directly proportional to λ in small-world and scale-free networks. Fig. 8e and f indicate that this relation also exists for real-world and random networks under study. In addition, the invariant mean distance ($k = 7$) is relevant for all network orders together, as seen in Table 2, and also for each network order separately, as seen in Table 4.

Table 5
Most relevant invariants to explain the number of wavelengths (k is the identifier of the invariant). In Appendix A it is possible to see how to compute all these invariants.

Theme	k	Invariant
Congestion	1	Maximum of edge betweenness
	2	Standard deviation of edge betweenness
	3	Amplitude of edge betweenness
	5	Mean of edge betweenness
	6	Mean of vertex betweenness
Connectivity	40	Algebraic connectivity
	59	Number of minimum sets that disconnects the graph
Transmission and Distance	4	Kirchhoff index
	7	Mean distance
	8	Mean transmission
	9	Maximum transmission
	10	Wiener index
	11	Median transmission
	12	Coefficient of variation of communicability Distance
	14	Harary index
	16	Mean eccentricity
	45	Standard deviation of distance
	115	Coefficient of variation of distance
Degree	101	Minimum average nearest neighbor degree
	167	Mean degree
Increase Network Order	35	Coefficient of variation of inverse log-weighted similarity
	57	Coefficient of variation by median of dice similarity
	61	Coefficient of variation by median of jaccard similarity
	63	Coefficient of variation by median of inverse log-weighted similarity
	68	Coefficient of variation by median of cocitation coupling

3.5. Further remarks regarding the optical network design

First of all, we emphasize the fact that our graph model well represents optical network topologies. Indeed, as can be seen in Figs. 2a, b, 4 a, b, 5 a, b, 7, and 8, both real-world and random networks under study have a similar behavior with respect to λ. This result confirms, somehow, the sufficiency of the hypotheses used to generate the random networks, in the face of the network characteristics considered in Section 2.1.

In all presented results it can be noticed a considerable range of wavelength requirements for each invariant. It means that, for a fixed value of a given invariant, it is possible to find, in this range, a network topology reaching the minimal λ. For instance, for each fixed mean degree (i.e., fixed n and m) in Fig. 8b, all real-world and random networks with the same mean degree are found in the same vertical line. For a fixed real-world network, all networks below it in this vertical line require fewer wavelengths, even with the same n and m. Thus, given n and m, there is room for designing new optical network topologies requiring few wavelengths.

We consider the most relevant invariants the ones that appear in all columns of Table 4, and we list them in Table 5. These invariants are classified into the following themed groups: “Congestion” (invariants which are related to how much nodes and links are used to meet a traffic demand), “Connectivity” (invariants which are related with the number of vertex or edge disjoint paths for each pair of vertices), “Distance” (invariants derived from the concept of distance between vertices), and “Degree” (invariants derived from the vertex degree).

There is another group called “Increase Network Order”, which brings together the invariants relevant for increasing network orders. In each group, the invariants are sorted by relevance.

Among the invariants in Table 5, to cover all mentioned aspects, we recommended to consider at least one of each group. Two highly correlated invariants are not necessarily redundant, provided they may indeed give different information Bennasar et al. (2015). It is not recommended, however, to use all invariants of Table 5, because this would make the modeling very costly and redundant, unless that a feature extraction method (Guyon & Elisseeff, 2006) is considered. In addition, within each group, the correlation between pairs of variables tends to be strong, since they reflect the same theme.

From all our investigation, the most prominent invariant is the maximum edge betweenness, as for all networks together, as for networks separated by n. It could be explained by the relation between the maximum edge betweenness and edge congestion. As observed in the study of Cousineau et al. (2015), the edge congestion is a lower bound for λ, which is often reached. Therefore, with these considerations, it is expected a strong relation between maximum edge betweenness and λ.

4. Conclusion

Expert systems to optical network design currently deal with conflicting aspects like computational and network resources (Yang et al., 2010) or different types of network flows (Przewoźniczek et al., 2015). These approaches are usually hard not merely because of their computational complexity but as well due to the immense scale of its solution space. These problems have been extensively studied using all sort of artificial intelligence methods, neural networks, and genetic algorithms (Hanay et al., 2015).

A topic widely explored in many areas of expert and intelligent systems is feature selection (Bennasar et al., 2015). To the best of our knowledge, there is no previous work in the literature resorting to feature selection to optical network design. A few works explore graph invariants from graph theory to explain the number of wavelengths, but not as comprehensive as the one we present here. Within our best efforts, we then covered the literature related to this subject and grouped a list with 315 topological invariants that are easier to compute than the number of wavelengths.

Our study explores the influence of each topological invariant analyzed on optical networks, more specifically, on the wavelength requirement. We do this using a feature selection filter method based on mutual information, similar to Bennasar et al. (2015), but in our case, we propose an estimator that can be applied to discrete and continuous data regardlessly. Our new mutual information estimator based on an entropy estimator (Montalvão et al., 2014) is suitable to obtain estimates in high-dimensional spaces, similar to the Neighborhood Mutual Information estimator from Hu et al. (2011). Both estimators can be applied to discrete and continuous features and do not require discretization. However, our entropy estimator based on the method of coincidence is a more intuitive and easy-to-use approach, and it does not require probability density function estimation.

Mutual information helps us to identify essential invariants among the 315 invariants we have analyzed. Samples of 15 real-world networks and 2.2×10^6 random topologies (which mimic real networks) are considered. We performed two kinds of analyses, one considering all networks together, and another separating the networks by order. In both cases, we observe that real-world and random networks under study have similar behavior concerning to wavelength requirements.

Overall, our results stand out the importance of considering edge congestion, connectivity, transmission, distance, and vertex degree variables to design optical network topologies with low wavelength requirements. For each one of these categories, a set of more relevant invariants is presented in a final table, which can be used in a knowledge base for optical network design. Invariants associated with the connectivity and with vertex degree appeared more important in the analysis separated by network order. Nevertheless, even if our set of invariants is as diverse and as extensive as possible, new types of invariants can emerge, which could be beneficial candidates to explain the number of wavelengths or other network metrics.

In the meantime, performing comparisons with the literature and our results, some findings are reinforced, such as the importance of the invariants: mean degree, algebraic connectivity, and the mean distance. On the other hand, we do not confirm the importance of the following exalted invariants in the literature: edge density, variance degree, and the number of spanning trees.

It is noteworthy highlight the significance of the invariants derived from edge betweenness in optical network design, the ones with the most consistent and better performance in all this investigation. This prominence explains why the routing algorithms in Cousineau et al. (2015), based in edge betweenness, have such an outstanding efficiency in optimizing congestion. Therefore, we reinforce that future works consider use edge betweenness in routing algorithms, whenever it is of interest to avoid congestion.

Lending continuity to this research, we are developing a system to design and expand optical network providing topologies with the least possible wavelength requirement. This tool uses equations obtained by appropriate regression analysis, based on the selected invariants highlighted in this study. The regression equations are used as surrogate functions to estimate the number of wavelengths in a surrogate-based optimization model. Also, a feature extraction method can be used to obtain other surrogate functions, if one wants to use a higher amount of invariants to obtain such estimation.

As future research directions, it would be interesting produce other invariants rankings to explain additional optical networks requirements also hard to compute, like Capital Expenditure and Blocking Probability (De Araújo, Bastos-Filho, & Martins-Filho, 2015), or Reliability (Pavan, de Lima, Paiva, & Segatto, 2015) and Dominating Set (El Houmaidi, Bassiouni, & Li, 2003). Once these variables are adequately explained by graph invariants, with a low computational burden, these outcomes can be combined in a single system to design optical networks topologies. Such a system could consist of a global optimization method that could optimize many optical networks features at once, finding the best topologies for the optical networks, in accordance with the considered requirements. As it might be possible that some of these requirements present trade-offs, this line of work could help decide which features to prioritize in the optical networks topology design.

A topic that can even be worked on is to consider the resilience of the networks to the failure of different nodes or links, and how this affects the network features studied that one wishes to preserve. Lastly, another contribution of this work, the new Mutual

Information estimator based on the Method of Coincidence, leaves plenty of room for application to any other fields, and also future works could compare it with similar estimators from literature.

Acknowledgment

This work is partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, with projects number 462477/2014-2, 304564/2016-8, and 304853/2015-1.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.eswa.2018.04.018 .

References

Antunes, C. H., Craveirinha, J., & Climaco, J. N. (1993). A multiple criteria model for new telecommunication service planning. *European Journal of Operational Research, 71*(3), 341–352.
Banerjee, D., & Mukherjee, B. (2000). Wavelength-routed optical networks: Linear formulation, resource budgeting tradeoffs, and a reconfiguration study. *IEEE/ACM Transactions on Networking, 8*(5), 598–607.
Baroni, S., & Bayvel, P. (1997). Wavelength requirements in arbitrarily connected wavelength-routed optical networks. *Journal of Lightwave Technology, 15*(2), 242–251.
Bennasar, M., Hicks, Y., & Setchi, R. (2015). Feature selection using joint mutual information maximisation. *Expert Systems with Applications, 42*(22), 8520–8532.
Bickle, A. (2013). Cores and shells of graphs. *Mathematica Bohemica, 138*(1), 43–59.
Châtelain, B., Bélanger, M. P., Tremblay, C., Gagnon, F., & Plant, D. V. (2009). Topological wavelength usage estimation in transparent wide area networks. *Journal of Optical Communications and Networking, 1*(1), 196–203.
Christodoulopoulos, K., Tomkos, I., & Varvarigos, E. (2011). Elastic bandwidth allocation in flexible OFDM-based optical networks. *Journal of Lightwave Technology, 29*(9), 1354–1366.
Cousineau, M., Perron, S., Caporossi, G., Paiva, M., & Segatto, M. (2015). RWA problem with geodesics in realistic OTN topologies. *Optical Switching and Networking, 15*, 18–28.
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*(5), 1–9.
De Araújo, D. R., Bastos-Filho, C. J., & Martins-Filho, J. F. (2015). An evolutionary approach with surrogate models and network science concepts to design optical networks. *Engineering Applications of Artificial Intelligence, 43*, 67–80.
Diestel, R. (2016). *Graph theory*: Vol. 173. Springer.
El Houmaidi, M., Bassiouni, M., & Li, G. (2003). Dominating set algorithms for sparse placement of full and limited wavelength converters in wdm optical networks. *Journal of Optical Networking, 2*(6), 162–177.
Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5*, 17–61.
Fenger, C., Limal, E., Gliese, U., & Mahon, C. J. (2000). Statistical study of the correlation between topology and wavelength usage in optical networks with and without conversion. In *Networking 2000 broadband communications, high performance networking, and performance of communication networks* (pp. 168–175). Springer.
Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. In *Feature extraction* (pp. 1–25). Springer.
Hanay, Y. S., Arakawa, S., & Murata, M. (2015). Network topology selection with multistate neural memories. *Expert Systems with Applications, 42*(6), 3219–3226.
Hu, Q., Zhang, L., Zhang, D., Pan, W., An, S., & Pedrycz, W. (2011). Measuring relevance between discrete and continuous features based on neighborhood mutual information. *Expert Systems with Applications, 38*(9), 10737–10750.
Jaumard, B., Meyer, C., & Thiongane, B. (2007). Comparison of ILP formulations for the RWA problem. *Optical Switching and Networking, 4*(3), 157–172.
Ma, S. (1981). Calculation of entropy from data of motion. *Journal of Statistical Physics, 26*(2), 221–240.
Ma, S. (1985). *Statistical mechanics*. World Scientific Publishing Co. Pte. Ltd.
Montalvão, J., Attux, R., & Silva, D. (2014). A pragmatic entropy and differential entropy estimator for small datasets. *Journal of Communication and Information Systems, 29*(1), 29–36.
Patel, A. N., Ji, P. N., Jue, J. P., & Wang, T. (2012). Routing, wavelength assignment, and spectrum allocation algorithms in transparent flexible optical WDM networks. *Optical Switching and Networking, 9*(3), 191–204.
Pavan, C., de Lima, L., Paiva, M., & Segatto, M. (2015). How reliable are the real–world optical transport networks? *Journal of Optical Communications and Networking, 7*(6), 578–585.
Pavan, C., Morais, R. M., Ferreira da Rocha, J. R., & Pinto, A. N. (2010). Generating realistic optical transport network topologies. *Journal of Optical Communications and Networking, 2*(1), 80–90.
Przewoźniczek, M., Goścień, R., Walkowiak, K., & Klinkowski, M. (2015). Towards solving practical problems of large solution space using a novel pattern searching hybrid evolutionary algorithm–an elastic optical network optimization case study. *Expert Systems with Applications, 42*(21), 7781–7796.
R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.
Talebi, S., Alam, F., Katib, I., Khamis, M., Salama, R., & Rouskas, G. N. (2014). Spectrum management techniques for elastic optical networks: A survey. *Optical Switching and Networking, 13*, 34–48.
Tessinari, R. S., Puype, B., Colle, D., & Garcia, A. S. (2016). ElasticO++: An elastic optical network simulation framework for OMNeT++. *Optical Switching and Networking, 22*, 95–104.
Wang, R., & Mukherjee, B. (2014). Spectrum management in heterogeneous bandwidth optical networks. *Optical Switching and Networking, 11*, 83–91.
Yang, Y., Wu, G., Chen, J., & Dai, W. (2010). Multi-objective optimization based on ant colony optimization in grid over optical burst switching networks. *Expert Systems with Applications, 37*(2), 1769–1775.
Yoon, M., Baek, Y., & Tcha, D. (1998). Design of a distributed fiber transport network with hubbing topology. *European Journal of Operational Research, 104*(3), 510–520.
Yuan, P., & Xu, A. (2010). The influence of physical network topologies on wavelength requirements in optical networks. *Journal of Lightwave Technology, 28*(9), 1338–1343.
Zang, H., Jue, J. P., & Mukherjee, B. (2000). A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. *Optical Networks Magazine, 1*(1), 47–60.
Zhang, G., De Leenheer, M., Morea, A., & Mukherjee, B. (2013). A survey on OFDM-based elastic core optical networking. *IEEE Communications Surveys & Tutorials, 15*(1), 65–87.