Making intelligent topology design choices: understanding structural and physical property performance implications in optical networks

ROBIN MATZNER*, DANIEL SEMRAU, RUIJIE LUO, GEORGIOS ZERVAS, AND POLINA BAYVEL

Optical Networks Group, University College London, Department of Electronic and Electrical Engineering, Roberts Building, Torrington Place, London WC1E 7JE, UK
** Corresponding author: robin.matzner.19@ucl.ac.uk*

Compiled June 22, 2021

The key goal in optical network design is to introduce intelligence in the network and deliver capacity when and where it is needed. It is critical to understand the dependencies between network topology properties and the achievable network throughput. Real topology data of optical networks is scarce and often large sets of synthetic graphs are used to evaluate their performance including proposed routing algorithms. These synthetic graphs are typically generated via the Erdos-Renyi (ER) and Barabasi-Albert (BA) models. Both models lead to distinct structural properties of the synthetic graphs, including the degree and diameter distributions. In this paper, we show that these two commonly used approaches are not adequate for the modelling of real optical networks. The structural properties of optical core networks are strongly influenced by the internodal distances. These, in turn, impact the signal-to-noise ratio, which is distance-dependent. The analysis of optical network performance must, therefore, include spatial awareness to better reflect the graph properties of optical core network topologies. In this work, a new variant of the BA model, taking into account the inter-nodal signal-to-noise ratio, is proposed. It is shown that this approach captures both the effects of graph structure and physical properties to generate better networks than traditional methods. The proposed model is compared to spatially agnostic approaches, in terms of the wavelength requirements and the total information throughput, and highlights how intelligent choices can significantly increase network throughputs whilst saving fibre.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Optical networks are high-capacity communication networks, that make use of optical fibres to generate, transmit, process and route information. These vary in distance- and time-scales, from long-distance core networks to shorter distance metropolitan-area and access networks. More recently, optical networks research has entered the realm of data centre networking, connecting multiple data centres as well as intra-data centre, connecting 10s-100s of thousands of servers.

Wavelength division multiplexed (WDM) networks use the wavelength domain to increase transmission rates per fibre as well as for routing, establishing high capacity connections, or lightpaths, between nodes. Significant research efforts have focused on calculating the number of wavelengths required to interconnect arbitrarily connected topologies [1–4]. These aimed to solve combinatorial optimisation problems posed by wavelength routing in arbitrarily connected graphs without taking into account physical properties of the networks, such as geographical node locations and the physical distances between them. As traffic demands increased, so did the channel rates and number of wavelength channels per fibre, resulting in the nonlinear Kerr effect becoming a significant source of distortions, limiting optical fibre transmission rates. This led to studies aiming to understand and evaluate the optical fibre channel capacity in point-to-point transmission [5]. Thus optimisation of wavelength-routing for networks, in the nonlinear regime, now required physical properties to be taken into account, and a re-evaluation of which graph characteristics determine network throughput for a nonlinear optical network.

In [6] the physical properties were included to maximise the network throughput via optimal routing, however the study focused on one topology only. Similarly, most papers investigating routing and resource allocation used specific (and a relatively small number of) published 'real' network topologies as benchmarks for testing algorithms and performance metrics [7–14].

These are not sufficient for the evaluation of routing algorithms or network performance, as it is important to confirm that the performance is due to an algorithmic improvement and not due to some topological artefact.

The overarching goal in network design is, therefore, to understand which key network topology characteristics impact the achievable network throughput. This requires the analysis of sufficiently large sets of graphs to adequately capture graph structures and physical properties. This paper focuses on the different techniques to generate and analyse synthetic graphs, termed generative graph models, with properties that reflect optical network characteristics, including both graph structure and physical properties.

Most previous work using generative models did not include physical properties of optical networks; these are known as non-geometric generative models [15–18]. Although [19] analysed the maximum achievable nonlinear throughput for a range of networks using their physical properties, the process of topology generation followed that of non-geometric generative models, not adequately reflecting spatial information.

Recently [20] included physical properties in the generation of topologies using a genetic algorithm, to quantify maximum achievable nonlinear throughput. With the main focus on the comparison of heuristic versus ILP solutions for resource allocation, this work, however, did not include the necessary analysis of the generated graph structures.

To include physical properties in the graph generation process, there was a number of developments in geometric generative graph models. [21] modelled optical core networks via a geometric generative graph model, which scales the probabilities of edges via an exponential decay (known as the Waxman model). However, as shown in Section 2E the exponential decay of edge probabilities with distance is not representative of the distance-dependent penalty observed in optical fibre networks. Other geometric generative models, in turn, failed to model well connected local hubs [22], typical of optical core networks connecting major centres. In [23] graphs were generated using a variety of geometric generative models, however examined neither the structure of these graphs nor the physical properties when investigating their performance.

In [24] a novel geometric generative model with properties representative of optical core networks, was introduced, together with a comparison with the traditional non-geometric generative models, identifying the desirable physical network properties.

In this paper, we expand on this newly proposed model, by comparing the graph structures it generates - via their degree, diameter and spectral properties - to a set of 25 published example core optical networks. Our generative model is found to reflect structures of real optical core networks best and, therefore, is more realistic than the other investigated models. Therefore, this model can be used for creating large sets of unique graphs for future optical network simulation studies. In this paper, it is applied to study the impact of graph structural features and physical properties on performance, characterised by the wavelength requirements, as well as the total throughput of optical networks.

The rest of the paper is organised as follows. Section 2 describes the different generative graph models used including Erdoes-Renyi, Barabasi-Albert, Waxman and the newly proposed signal-to-noise ratio aware Barabasi-Albert model. Section 3 defines the graph properties used to compare the generative graph models; Section 4 presents the comparison of the graph properties of all the considered generative and published optical network topologies. In Section 5 the node positions of the 30-node Continental United States (CONUS) network and the ubiquitous National Science Foundation network (NSFNET) topologies are selected as the basis for the graphs generated by three of the generative models. A comparison of the generated topologies in terms of wavelength requirements and maximum achievable network throughput is described. The results allow the separation of the impact of graph structure from physical properties and show that, at distances applicable to core networks, physical properties dominate achievable network throughput.

2. GRAPH MODELS

As mentioned in the introduction, neither geometric or non-geometric generative graph models ideally capture the structure of real optical core networks, and this section describes four of the most common generative models to illustrate their shortcomings in the context of optical networks, together with the newly proposed model. This new model, which accurately describes the physical channel, is designed to capture localised hub-like connections and distance-dependent properties in optical networks.

A. Erdos Renyi (ER)

The oldest generative graph model is the ER model [25], a non-geometric model, in which the graph is created from N nodes with edges modelled as independent Bernoulli distributions, given some probability p for each edge in the graph to either exist or not. The ER graphs can be characterised by $G(N, p)$ where the number of edges are approximated by $E = \binom{N}{2} p$ [25]. To ensure robust optical networks, nodes must be bi-connected; when applying the ER model yields graphs with higher connectivity than typical for optical transport network topologies.

B. Barabasi-Albert (BA)

The Barabasi-Albert (BA) preferential attachment model [26] was developed to be more representative of different types of real networks, from social to communication networks. The model uses the 'rich gets richer' principle for modelling the edges of graphs, where nodes with higher degree (δ) are more likely to attract more edges. The BA model is popular within the graph modelling community due to resulting in degree distributions that resemble many types of real graphs. The model starts with two nodes connected by an edge, after which nodes are added sequentially and a set number of m edges are connected to every newly added node. The edges are chosen given probability $p(i, j)$ corresponding to Eq. (1), where i is the newly added node and j is an existing node in the graph. The probability $p(i, j)$ is determined only by the sum of all the degrees in the graph currently and the degree of the node j. This model creates large connected hubs within the network, which generally are the nodes added earlier on in the sequence of node additions. This non-geometric model forms a good basis for modelling communication networks, however lacks the necessary distance-dependence properties, important in optical transport networks.

$$p(i, j) = \frac{\delta_j}{\sum_{k \in N} \delta_k} \tag{1}$$

C. Waxman

The Waxman model [27] is a geometric generative model that depends only on distances and does not take into account the

degree of the nodes within the graph. Probabilities of edge connections within these graphs are governed by Eq. (2). $D(i,j)$ represents the distance between the two nodes i and j. L is the largest distance within the graph and the probabilities are scaled by the constant θ, with an inverse relationship between θ and the relative difference in probabilities between connections. In this work, we altered the Waxman model to operate as a sequential model, omitting the scaling parameter α due to its negligible impact on edge probabilities in this case. This model takes into account physical distances yet does not correctly describe the distance dependence of optical fibre transmission.

$$P_{WAX}(i,j) = \exp\left(-\frac{D(i,j)}{\theta L}\right) \tag{2}$$

D. Gabriel Graphs (GG)

The Gabriel graph model [28] is a geometric generative model that is deterministic, i.e. for a specific set of node coordinates it produces a single unique graph. This lack of flexibility makes them less suitable for large scale network investigation although they have been shown to accurately model some aspects of optical networks in, for example [22, 23]. If the distance between nodes i and j is denoted by $D(i,j)$, an edge (i,j) exists, if no node lies within a radius $r_G = \frac{D(i,j)}{2}$, drawn from the half way point between the two nodes.

E. Proposed signal-to-noise ratio aware Barabasi-Albert Graphs (SNR-BA)

Physical properties of the optical fibre and amplifiers used in transmission, as well as link (edge) lengths, define the transmission performance of optical networks. Thus, certain edges may be structurally beneficial, i.e. enabling more lightpaths to be setup, however these may be heavily physically impaired. To include the physical properties of links, and their effect on data transmission and network throughput, requires the addition of a metric which describes their impact within the network generation. The metric chosen in this work is the signal-to-noise ratio (SNR) of a given transmission link (or edge) and it is included within the probability function when choosing edges in the graph generation process.

As described earlier, since geometric generative models can capture the grid-like behaviour of real optical core networks, yet fail at modelling local hubs, we propose to extend the probability weights of the conventional BA model as defined in Eq. (1). This is achieved by including the SNR between any two given nodes, over a direct hypothetical link (single edge) between them, as a second weight in the conventional BA model. The SNR term attempts to make realistic link decisions via weighing shorter links more heavily and the BA term attempts to replicate local hubs in the network. A weighting parameter (θ) is used to determine how heavily to weight the physical properties within the graph generation. The probability weights of the SNR-BA are then given by

$$P_{\text{SNR-BA}}(i,j) = \left(\frac{\text{SNR}(i,j)}{\sum_{k\in N}\text{SNR}(i,k)}\right)^{\theta} \cdot \frac{\delta_j}{\sum_{k\in N}\delta_k}, \tag{3}$$

where $\text{SNR}(i,j)$ is the SNR on the direct link between nodes i and j. The SNR includes the effects of distortion arising from the optical Kerr effect. The latter can be approximated as noise, referred to as nonlinear interference noise, and amplified spontaneous emission noise from optical amplifiers. Numerous models of calculating the SNR of an optical lightpath have been proposed in the literature. Following one of the most widespread modelling approaches, namely a first-order perturbative description of the nonlinear interference noise, the SNR at optimum launch power is given by [29].

$$\text{SNR} = \frac{1}{\sqrt[3]{\frac{27}{4}P_{\text{ASE}}^2\eta n^3}}, \tag{4}$$

where P_{ASE} is the injected amplified spontaneous emission noise per amplifier, η is the nonlinear interference coefficient and n is the number of fibre spans. Eq. (4) assumes an incoherent addition of nonlinear interference across multiple spans, a common assumption in the physical layer modelling of optical networks which imposes negligible inaccuracies for C-band transmission and beyond, cf. ($\epsilon \ll 1$) in [Fig. 10-11 29]. Recalling that the number of spans is given by $n = \left\lfloor \frac{D(i,j)}{L} \right\rceil$, where L is the span length and $\lfloor x \rceil$ denotes rounding to the nearest integer, Eq. (4) can be written as

$$\text{SNR}(i,j) = \text{SNR}_1 \cdot \left\lfloor \frac{L}{D(i,j)} \right\rceil, \tag{5}$$

where SNR_1 is the SNR after a single span. Inserting Eq. (5) in Eq. (3) yields the proposed probability weights as

$$\begin{aligned} P_{\text{SNR-BA}}(i,j) &= \left(\frac{\left\lfloor \frac{L}{D(i,j)} \right\rceil}{\sum_{k\in N}\left\lfloor \frac{L}{D(i,k)} \right\rceil}\right)^{\theta} \cdot \frac{\delta_j}{\sum_{k\in N}\delta_k} \\ &\approx \frac{1}{\left(\sum_{k\in N}\frac{D(i,j)}{D(i,k)}\right)^{\theta}} \cdot \frac{\delta_j}{\sum_{k\in N}\delta_k}, \end{aligned} \tag{6}$$

where the approximation is introduced by dropping the rounding operation. The derivation of Eq. (6), assumes that the amplified spans have identical lengths throughout the network. While this is not always satisfied in practice, Eq. (6) still describes the *average* SNR scaling with respect to distance.

The next step is to compare the generative models, including the newly proposed SNR-BA model, with a set of real optical core networks, described in the next section.

F. Real Optical Core Networks

A data set of 25 core optical network topologies from the survivable network design library (SNDlib) [30] and CONUS topologies [31], are used in the comparison with the graphs generated via the different modes in the sections A-D, and are summarised in Table 2. The constraints on the choice of topologies for the set was that no network graph could be cut in two by removal of a single edge, to ensure resilience. Although most networks used are legacy networks, designed in an era where optical networks were opaque, the design goals within opaque and transparent optical networks are very similar. Opaque networks aim at minimising edge lengths and the diameter of the network, where transparent networks aim at minimising the path lengths within the network to minimise physical layer impairments.

This dataset allowed for accurate distance modelling using exact geographical node locations for each network. To model distances in these graphs, their geographical coordinates were

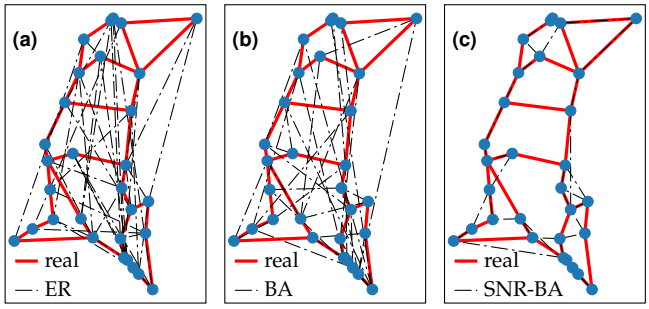

Fig. 1. Comparison between the connections made for (a) ER graph (b) BA graph (c) SNR-BA, on top of the adapated CONUS topology [31].

used in conjunction with the Harvesine formula [32]. The Harvesine formula takes into account the curvature of the earth and calculates distances over a sphere rather than a plain. We also account for realistic fibre distances as shown in Eq. (7), where D_{hav} represents the haversine distance. The fibre distances are estimated according to the European Telecommunications Standards Institute (ETSI) guidelines [33].

The next section defines the metrics used for the analysis of graph structure properties and the comparison of the generated and real optical network topologies.

$$D_{fibre} = \begin{cases} 1.5 \cdot D_{hav} & \text{if } D_{hav} < 1000\,km \\ 1500\,km & \text{if } 1000\,km \leq D_{hav} \leq 1200\,km \\ 1.25 \cdot D_{hav} & \text{if } D_{hav} > 1200\,km \end{cases} \quad \textbf{(7)}$$

3. GRAPH PROPERTIES

The generative graph models in Section 2 lead to different graph structures. This can be seen in figure 1 which shows an example North American topology (CONUS) [31] and one of the set of 25 real optical core networks, described in Section F. The published CONUS topology is plotted together with an instance of a graph generated by ER, BA and SNR-BA, defined in Section 2, for the CONUS node locations. It can clearly be seen that the SNR-BA model reflects the original structure of the graph better. Quantitatively analysing and comparing the structures of these different models is complex, so a set of comparison metrics is needed and these are defined in this section.

A. Degree & Diameter Distributions

Graphs are defined by the distribution of the individual number of connections of each node, known as the degree. It has been shown in [25] and [26] that these generative models give very distinct degree distributions.

The adjacency matrix defines the edge connections between nodes, Eq. (8).

$$A(i,j) = \begin{cases} 1 & \text{if node } i \text{ is connected to node } j \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(8)}$$

Using the adjacency matrix A, the degree of each node i is the summation of the connections to every other node j, defined in Eq. (9).

$$\delta_i = \sum_{j \in N} A(i,j) \quad \textbf{(9)}$$

The diameter of a graph is defined as the maximum eccentricity within a graph, which is the set of shortest paths from a node to any other in the graph. If the eccentricity of n is $\epsilon(n)$, then the diameter of a graph can be defined as in Eq. (10).

$$d(G) = \max_{n \in N}(\epsilon(n)) \quad \textbf{(10)}$$

B. Spectral Properties

Degree and diameter are conventional metrics for assessing networks, however are not sufficient in describing the structural properties. Spectral graph theory describes the structural properties of graphs related to the eigenvalues and eigenvectors of the matrix representations of graphs. These matrices allow for the modelling of graph structures in matrix form. The spectrum of a graph is the distribution of eigenvalues of any of the adjacency, Laplacian and normalised Laplacian matrices. Even though two graphs that are co-spectral (share the same spectrum) are not isomorphic, the graph spectrum has been shown to be a method for measuring similarity between graph structures [34–36].

The Laplacian itself is simply a matrix representation of the graph, like the adjacency matrix, where edges between nodes are denoted by values of -1 and the degree of nodes is noted in the diagonal of the matrix. This can be related back to the adjacency matrix using Eq. (11) and Eq. (12).

$$D_L^{i,j} = \begin{cases} \delta_i & \text{if i = j} \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(11)}$$

$$L = D_L - A \quad \textbf{(12)}$$

Whilst the simple Laplacian and adjacency matrix mainly describe the structures of regular graphs, the normalised form of Laplacian, and its eigenvalues, allows to generalise the network properties [36]. In this paper, we, therefore, use the normalised Laplacian as defined in Eq. (13), where I is the identity matrix.

$$L_D = D_L^{-\frac{1}{2}} L D_L^{-\frac{1}{2}} = I - D_L^{-\frac{1}{2}} A D_L^{-\frac{1}{2}} \quad \textbf{(13)}$$

Eigenvalue distribution of the normalised Laplacian help identify similarities and differences between graphs. However, a weighted spectral distribution (WSD) [34] helps highlight these differences by giving greater weights to the edges of the spectrum whilst shrinking the centre. To calculate the WSD as well, as a metric measuring the distance between two graphs $F(G_1, G_2, N)$, the following steps are taken.

L_D is a real symmetric matrix with real eigenvalues $\lambda(G) = \{\lambda_0, \lambda_1, ..., \lambda_N\}$ and real eigenvectors $v(G) = \{v_0, v_1, ..., v_N\}$. Given the eigenvalues of L_D one can construct the WSD by the probability function Eq. (14).

$$WSD(G) \rightarrow \{k \in K : (1-k)^N f(\lambda = k)\} \quad \textbf{(14)}$$

Using the determined eigenvalues of L_D one can construct its distribution over K bins. The fraction of eigenvalues falling in bin k is represented by $f(\lambda = k)$ at each equally spaced bin $k \in K$. A value of 40 bins was chosen, as it empirically showed to give interpretable WSDs for graphs of all sizes.

Using Eq. (14) one can also define a metric to measure the absolute distance between two separate graph spectra to quantify structural similarity between graphs, as

$$F(G_1, G_2, N) = \sum_{k \in K} (1-k)^N (f_1(\lambda = k) - f_2(\lambda = k))^2 \quad \textbf{(15)}$$

From Eq. (13) the normalised Laplacian yields spectra clustered around the eigenvalue this indicates duplicity in the network structure [37] - that is, if eigenvalues of nodes i and j are

close to one, then it is likely that they share some or all of their neighbours. The WSD, however, suppresses this feature to amplify other structural properties that would have otherwise been hidden.

The eigenvalues of the normalised Laplacian fall into the range $\lambda(G) \in [0, 2]$ to describe structural properties. The number of eigenvalues equal to zero indicates how many connected components exist in the graph. For optical networks the graph needs to be a single connected component and therefore consist of a single connected component, reflected by a single eigenvalue of value zero. The second smallest eigenvalue is characterised as the algebraic connectivity [36] which indicates how well connected the network is, where sparse networks have a smaller value. This value relates to the conductance (Cheeger Constant) of a graph, which describes how 'bottle-necked' a graph is [36].

An eigenvalue of 2 indicates that the graph is bi-partite. Bipartite being the property of separating a graph into two subgraphs without connecting edges within those subgraphs. The closer the maximum eigenvalue is to 2 the closer it is to being bi-partite [37]. Generally, this feature is associated with sparsely connected graphs.

To understand the significance of graph spectra for the real topologies, these are plotted in figure 2 and 3 together with 3 regular 30-node topologies (5x6 grid, fully connected mesh and star) to demonstrate the differences in structure are clearly manifested in the spectrum/WSD. Figure 2 shows the empirical cumulative distribution function (ECDF) for the eigenvalues in the range $[0, 2]$. As can be seen, the fully connected mesh (diameter of 1) has the highest number of eigenvalues equal to one, clearly represented by the step-like nature of its ECDF. This is due to the regular structure of the graph, with every part of the network being identical. The star graph behaves similarly, however this type of network is less connected, with a diameter of two, however the structure of the graph is still highly repetitive and well-connected, resulting in a large clustering of eigenvalues around one.

The grid graph, however, has a spectrum that is spread more evenly, resulting in a more linear ECDF. The diameter of the grid graph is significantly larger (diameter of 9) than in the star and fully-connected mesh, and the graph is less well connected. These structural features push the second smallest and largest eigenvalues further to the edges of the spectrum, with lower algebraic connectivity and a more bi-partite graph structure. One can see that the ECDF for the real topologies is closer to the grid graph structure.

The real core topologies follow a grid structure more than that of connected hubs of star topologies, and this is reflected in the comparison of their WSDs which is shown in figure 3. It is clear that the star and fully connected mesh graphs have a flat WSD around zero, as their spectra are fully clustered around eigenvalues of one (with diameters of 1 or 2). The grid graph and real core topologies, due to their sparser connectivity and longer diameters, have more peaks at the edges of their spectrum, highlighting lower algebraic connectivity and higher bi-partite nature. In addition to the conventional graph metrics of degree and diameter, the analysis of graph spectral properties is an invaluable additional tool to understand differences in graph structures.

The next step is to compare the graph metrics of the generative models of Section 2 with those of the real core topologies.

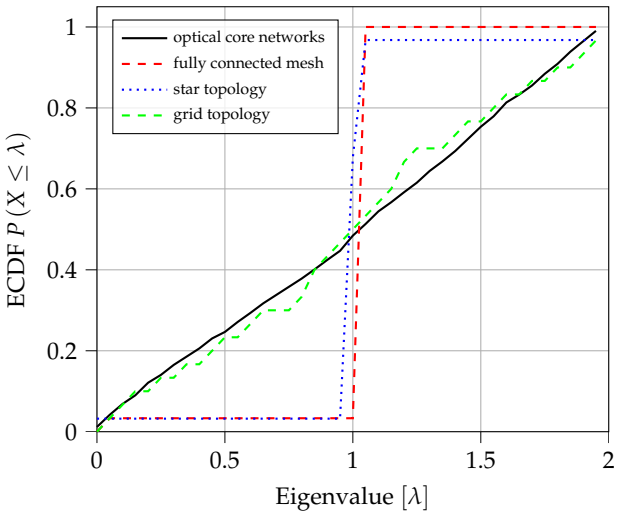

Fig. 2. Emperical cumulative distribution function of eigenvalues from the normalised laplacian - evaluated for a fully 30-node connected mesh, a 30-node star, a 5x6 grid and real optical core networks.

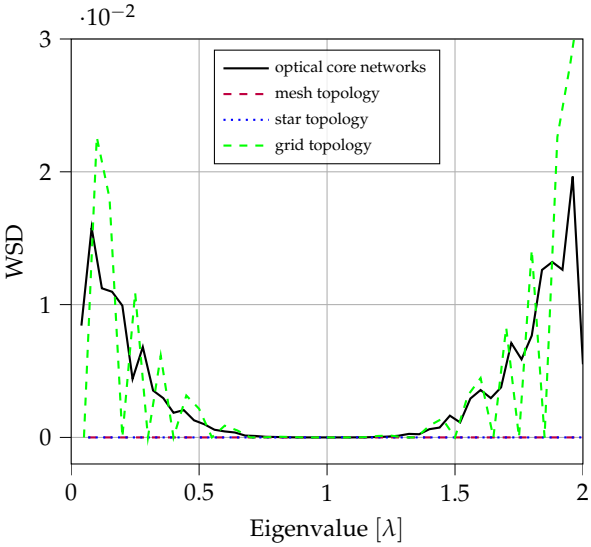

Fig. 3. Weighted spectral distribution calculated for a 30-node fully connected mesh, a 30-node star, a 5x6 grid and real optical core networks.

4. COMPARING GENERATIVE MODELS AND OPTICAL CORE NETWORKS

A. Dataset Generation

A set of 200 graphs were created, for each of the 25 real optical core networks, introduced in Section 2F, by taking the node-coordinates of the real networks, and generating the set of graphs following the rules of each of the generative models of Section 2, except for the Gabriel graphs for which only a singular graph was generated. To use this model in future simulation studies, node locations can be generated by uniformly scattering nodes on a grid of a given distance scale. To make sure the nodes are scattered evenly one can define a radius of length r, which ensures no two nodes are closer than this distance.

In total 5000 graphs were created per generative model. In

the case of the ER graphs, the graphs were first created and then the nodes assigned positions, as the node positions had no effect on the creation of the graph. In addition, since the BA graphs did not take into account distances, the coordinates did not matter, however the order in which one added nodes to the graph did, as nodes added early on in the sequence tend to act as highly connected hubs. Since the node sequence affects the graph generation, the same method for selecting the sequence of nodes for the sequential generative models (BA, Waxman, SNR-BA) was followed. The initial node selected in the sequence was the most central of all the nodes in the graph. This was determined firstly by finding the centroid of the graph via Eq. (16), where x and y refer to the latitude and longitude, respectively. We add the node with the smallest euclidean distance to the centroid at the beginning of the sequence. After this the next nodes were chosen sequentially by comparing every other node's average distance to all other nodes already present in the node sequence; each time adding the node with the smallest average distance to all other nodes in the sequence.

$$C_{x,y} = \left(\frac{x_1 + x_2 + ... + x_n}{n}, \frac{y_1 + y_2 + ... + y_n}{n}\right) \quad \textbf{(16)}$$

To select the θ parameter in Eq. (2) and Eq. (6) (the distance weighting factor) a simple sweep of the parameters was carried out to obtain values that performed well as a starting point, after which a non-gradient based optimisation was applied using the Nelson-Mead method [38]. The cost function used to evaluate the parameters was the weighted spectral distance (F), as defined in Eq. (15). It was found that the weighted spectral distance between the real graphs and those generated by SNR-BA was smallest when using $\theta \approx 5$. For Waxman graphs the smallest weighted spectral distance occurred when using $\theta \approx 8.5$. These values were used to generate the final set of graphs used for the analysis in this paper.

For the ER graphs p was chosen by evaluating $p = E/\binom{N}{2}$. If a graph was not bi-connected, then p was incremented by 0.01 and another graph generated. This is repeated until a feasible graph is found; on average 16.7% of the graphs generated were feasible graphs. For the BA model m was chosen to be $\lfloor \frac{E}{N} \rfloor$, after which edges are added to nodes until the desired number of edges is reached.

For each of these data sets , their degree, diameter, spectra and WSDs were calculated as described in Section 3. These distributions were then tested against the distributions of real topologies of Section 2F using the Kolmogorov-Smirnoff (KS) two sample test. This yields two key metrics: an absolute distance, D_{KS}, and a probability, p_{KS}. D_{KS} gives the largest variation between any two distributions. Smaller D_{KS} values indicate a small variation between the two distributions. The probability p_{KS} indicates whether the two distributions originate from the same population, with values close to one showing this, and is formally defined as the probability of a sample of size n having a D_{KS} of less or equal than the observed sample. One can evaluate p_{KS} by using the Kolmogorov cumulative distribution function defined by

$$P(D_{KS} \leq D_{obs}) = \frac{\sqrt{2\pi}}{D_{obs}} \sum_{i=1}^{\infty} e^{-(2i-1)^2\pi^2/\left(8D_{obs}^2\right)}$$

where n denotes the number of samples [39]. The weighted spectral distance, as defined in Eq. (15) indicates how different the WSD of two distributions are, where smaller values show greater similarity. The results of the KS two sample tests and the weighted spectral distances are summarised in table 1.

Model	WSD	Degree		Diameter		Spectrum	
	F	D_{KS}	p_{KS}	D_{KS}	p_{KS}	D_{KS}	p_{KS}
BA	0.00045	0.190	0	0.468	0	0.0580	0.003
ER	0.00087	0.222	0	0.397	0	0.0541	0.008
SNR-BA	**0.00019**	**0.035**	**0.177**	**0.119**	**0.827**	**0.0120**	**0.997**
Wax	0.00020	0.067	0	0.343	0.004	0.0352	0.194
GG	0.00067	0.141	0	0.252	0.327	0.0553	0.104

Table 1. Weighted spectral distances (F), Kolmogorov-Smirnov two sample test statistic (D_{KS}) and p-value (p_{KS}), calculated for the degree, diameter and spectra of the graphs generated by ER, BA, Waxman, GG and SNR-BA models.

B. Degree Distributions

The results of the comparison of the degree distributions are shown in figure 4. It can be seen that the BA graphs exhibit power law behaviour until the end of the tail, where the distribution is skewed due to the larger proportion of smaller graphs in the real optical core networks data set. The ER graphs yield degree distributions closer to Poisson distributions. This is because the ER graph generation aimed to create networks with the same number of edges as the sparse real optical core networks, and the resulting degree distribution has a sharper drop-off. Although the ER degree distribution look similar to that of real optical core networks, the beginning and end of the tail of the distribution differ. Finally comparing the SNR-BA, Waxman and GG models, it is clear these distributions fit more closely to the distributions of the real optical core networks, although still creating graphs with higher degrees than real optical core networks. The SNR-BA graphs match the start and end of the tail better than those of the Waxman/GG, and this is confirmed in table 1, where the KS two sample test for the degree distributions of the SNR-BA graphs produces the smallest absolute distance to the real optical core networks and the largest likelihood p_{KS}, indicating a larger probability that the two distributions originate from the same population than that of the other generative models. The SNR-BA model, thus, appears to generate graphs with degree properties closest to the real optical network topologies.

C. Diameter Distributions

The next step was to compare the diameter distributions, shown in figure 5. The real optical core networks show longer diameters with a large range of up to 16. The BA graphs can be seen to have shorter diameters, peaking around a diameter of 4. This is mainly because BA graphs create multiple highly connected hubs that tend to make connections across the graph - regardless of distance. ER graphs also have shorter diameters peaking around a diameter of 5, however as the edges within ER graphs are modelled as independent Bernoulli distributions, they do not create highly connected hubs within the graph, resulting in slightly longer diameters. In the Waxman graphs, it is easy to see that when introducing distances into the probabilities of edge creation, the diameters of the graphs start to increase. This is best demonstrated by the SNR-BA/GG graphs that have very large diameters, as they prefer shorter edge connections and, therefore, create grid-structures with inherently longer diameters. Again it is easy to see in the table 1 that the SNR-BA graphs have the smallest D_{KS} value, as well as the largest p_{KS} for their diameters compared to the real graphs.

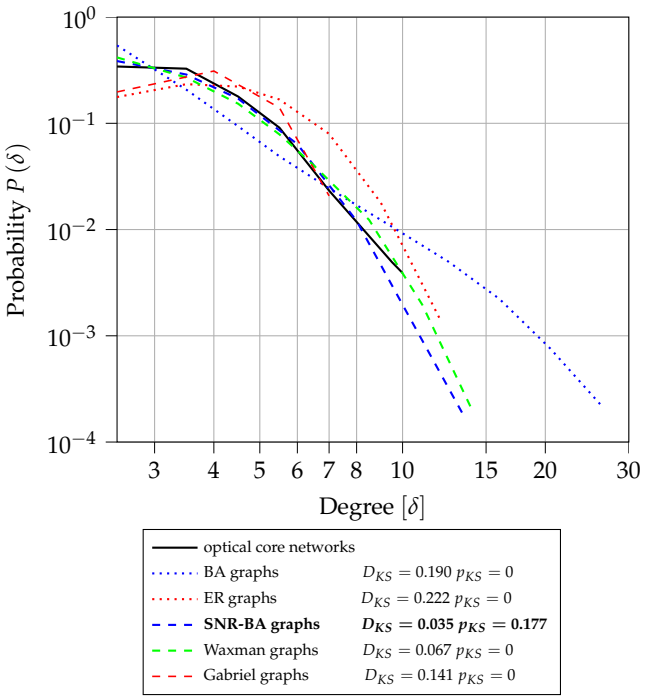

Fig. 4. Degree distributions for real optical core networks and graphs obtained from the ER, BA, Waxman, GG and SNR-BA models.

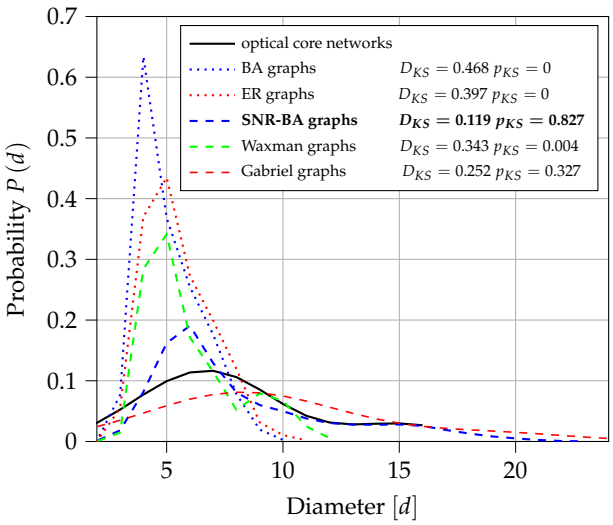

Fig. 5. Probability distributions of diameters for real optical core networks and graphs obtained from all the generative models.

D. Spectral Distributions

To give greater insight into the structure of the generated graphs, the spectra of the eigenvalues of the normalised Laplacian were compared, following Section 3. Figure 3 is re-plotted in figure 6 by adding the WSDs for all the generative models. As discussed previously, the key regions are the peaks around the smallest and largest eigenvalues.

Examining the behaviour of the left most peaks, it can be seen that the SNR-BA graphs best reflect the smaller algebraic connectivity values of real network topologies, followed by the

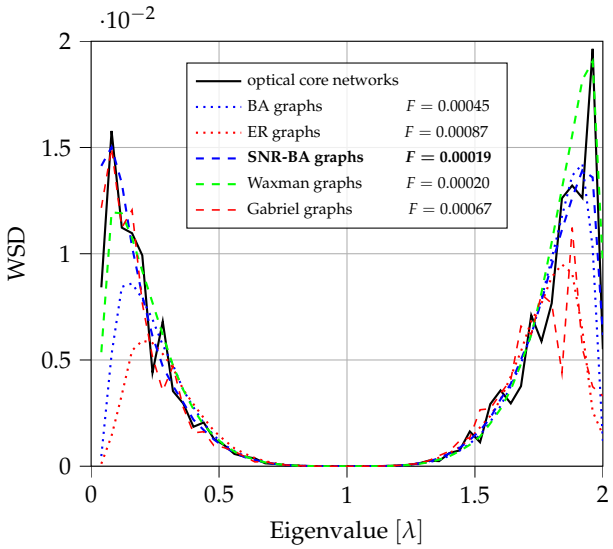

Fig. 6. Weighted spectral distribution using $N = 4$ for real optical core networks and graphs obtained from all generative graph models.

Waxman graphs. The BA and ER graphs tend to have the peaks of their spectrum closer to the centre, showing (a) algebraic connectivities are larger and (b) more of their eigenvalues are clustered around the centre of the spectrum, around the value of 1. On the right hand side, the BA graphs exhibit very similar bi-partite structure compared to SNR-BA graphs, shown by the peaks at similar positions. It is the Waxman graphs which appear to mimic the strong bi-partite nature of real optical core networks the best, indicated by the overlapping peaks, followed by the SNR-BA and then ER/GG graphs. However, the SNR-BA graphs have the smallest weighted spectral distance (F), as well as lowest D_{KS} and largest p_{KS} values of all generative models, and are thus, the most similar in structure to the real optical core networks. This comparison of the distributions of the generative models and the relative differences to the real optical core networks, highlights the importance of incorporating distances when modelling optical core networks. The proposed SNR-BA model combines the key structural features of optical network topologies together with a description which reflects their physical behaviour. This analysis can now be applied to investigate the relationship between the network structure, physical properties and performance, in terms of key optical network performance metrics, namely wavelength requirements and throughput. This is carried out in the next section using two sample optical network topologies, CONUS and NSFNET, focusing on the comparison to geometric (SNR-BA) and non-geometric (ER, BA) graphs. CONUS was selected as a representative north-American topology from the set of 25 networks used and NSFNet has been widely used in literature as a sample topology for a variety of studies [1, 20, 40]. Waxman and GG graphs were excluded to reduce computational time, as SNR-BA graphs were shown to give more structural similarities to real optical core networks when including distances in the edge creation process. Furthermore, GG graphs are deterministic and therefore cannot produce any statistically significant results in this case study.

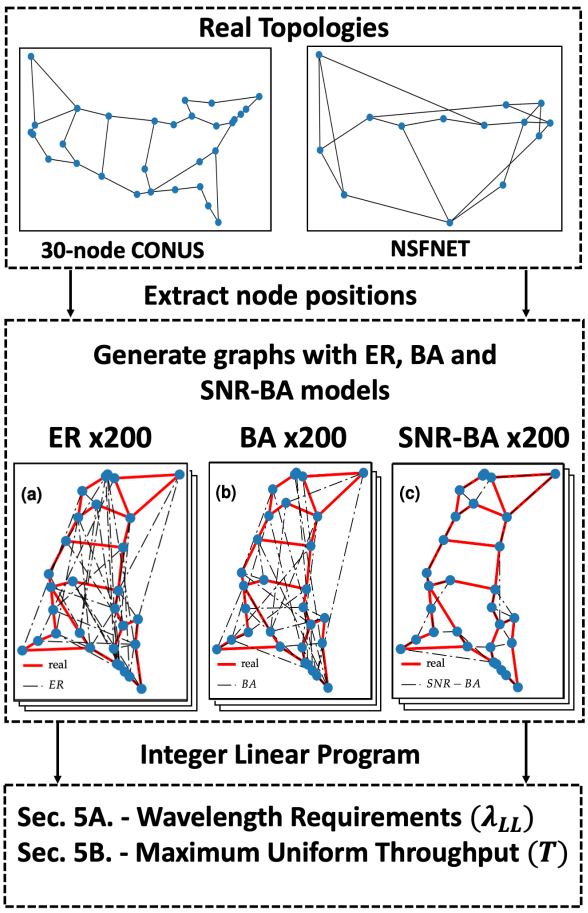

Fig. 7. Process of graph generation and optical network performance analysis.

5. NETWORK PERFORMANCE STUDY

The 30-node CONUS topology and the NSFNET [41] are shown in figure 7 together with the simulation steps in the network performance study. NSFNET was originally designed in the 1980s and focused on connecting supercomputer centres, not major city sites, and may now be less applicable. However, it has been widely used as a reference optical network, topology in the literature, and we have included it for this reason. Starting with the CONUS and NSFNET node positions, 200 graphs were generated by ER, BA and SNR-BA, respectively, with a total 600 graphs per topology. The constraint imposed on the generated topologies was that no graph could be cut in two by the removal of a single edge (bi-connectivity). The dramatically different network topologies, generated from the same set of networks nodes by different generative models, was already illustated in figure 1 but can now be analysed quantitatively. As the CONUS topology is very sparse (connectivity, defined in [1] of 0.082), the ER and BA graphs struggled to satisfy the resilience constraint, creating graphs with 25% and 19% more edges than the original network. The graphs based on the NSFNET node locations have exactly 21 edges, like the original NSFNET.

The performance metrics used to investigate the impact of structural and physical distance properties are (i) the lowest number of wavelengths needed to route all-to-all demands (denoted as the lower limit λ_{LL}) [42] and (ii) the maximum network throughput given zero blocking and uniform traffic. To calculate wavelength requirements (λ_{LL}) and throughput exactly, an integer linear program (ILP) was used, given 312 available wavelengths within a network, assuming a fully populated C-band (1530-70 nm) and 16 GBd Nyquist spaced channels (channel spacing of 0.128nm). This ILP computed λ_{LL} for all 1200 of the ER, BA and SNR-BA graphs. The aim is to compare these with the λ_{LL} value, calculated for the CONUS and NSFNET actual topologies (together with the exact link distances). The maximum throughput (T) for the generated graphs was calculated via another ILP formulation [40] and a closed form Gaussian noise (GN) physical layer impairments (PLI) model [43], explained in Section 5B, to estimate the SNR of the different lightpaths, to compare real and generated networks. Between any node pair, only paths that are the same length (in terms of hops) as the shortest path were allowed to be used.

A. Wavelength Requirements

To calculate the λ_{LL} value of the graphs, an ILP based on [42] was used. For a network, a set K of k-shortest paths was found using Yen's algorithm [44], which iteratively finds alternate routes between source and destination nodes of varying lengths. The value of k was set to 100, although in most cases a significantly smaller number was achieved and used; the resultant paths were filtered so that only paths of the same lengths (number of hops) as the shortest path were used. A set of Z node pairs (equipped with a set of maximum W wavelengths) needs to be connected via a routing and wavelength assignment (RWA) via a set of K paths. The decision variable $\delta_{w,k,z}$ - with $w \in W$, $k \in K$, $z \in Z$ - is able to fully define the RWA of a network following the definition in Eq. (17).

$$\delta_{w,k,z} = \begin{cases} 1 & \text{if (k, w) are the RWA assignment} \\ & \text{for node pair } z \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

To define whether a wavelength is needed for routing within the network, the variable u_w is defined as in Eq. (18).

$$u_w = \begin{cases} 1 & \text{if wavelength w is used in any} \\ & \text{assignment in } \delta_{w,k,z} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

Constraining $u_w \geq \delta_{w,k,z}\ \forall w \in W\ \forall k \in K\ \forall z \in Z$.

Using u_w, an objective function is defined to minimise the sum of u_w over all $w \in W$, as defined in Eq. (19).

$$\lambda_{LL} = \min\left(\sum_{w \in W} u_w\right) \quad (19)$$

The ILP solution set needs to be constrained, so as to only give solutions that are feasible for optical networking. Firstly, by only assigning a single path and wavelength per node pair $z \in Z$ as defined in Eq. (20).

$$\sum_{w \in W} \sum_{k \in K} \delta_{w,k,z} = 1 \quad \forall z \in Z \quad (20)$$

Secondly, no two node-pair path assignments can share a wavelength on any given edge j. Therefore, the variable $I(j \in k)$ is defined to be 1 when edge j is in path k and 0 otherwise. Using this the wavelength uniqueness can be constrained as in Eq. (21).

$$\sum_{z \in Z} \sum_{k \in K} \delta_{w,k,z} I(j \in k) \leq 1 \quad \forall j \in E \quad \forall w \in W \quad (21)$$

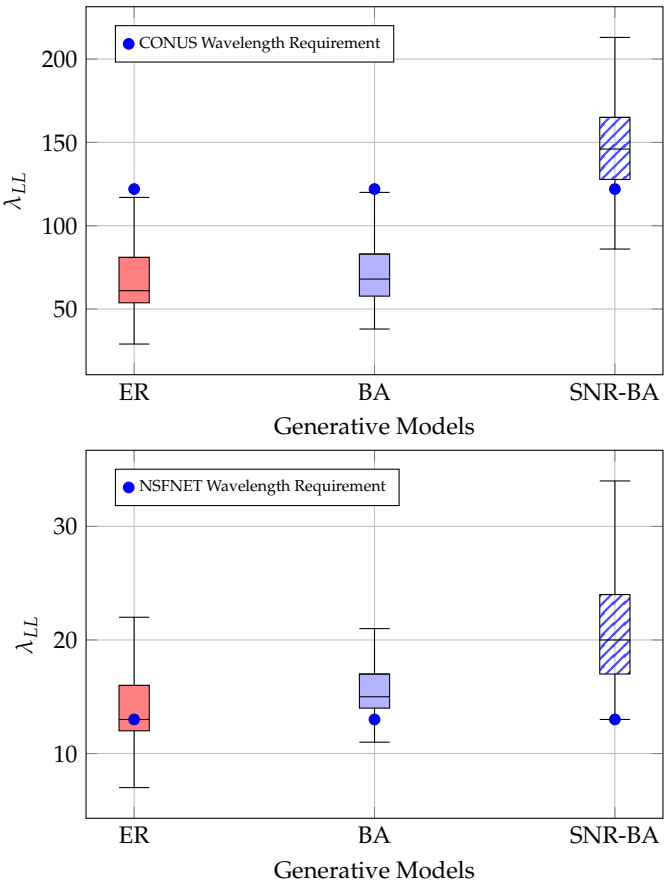

Fig. 8. Wavelength requirements shown for the CONUS and NSFNET, as well as box-plots illustrating the distribution of wavelength requirements for the graphs generated by ER, BA and SNR-BA models.

The ILP yielded the minimum wavelength requirements, that is the minimum number of wavelengths needed to route the $N(N-1)/2$ demands between all node pairs, for the CONUS network as 122, for the NSFNET this number is 13, same as determined in [1]. These are shown in the box-plots of figure 8, together with the values for the ER, BA and SNR-BA graphs. The box-plot shows the distribution of the data together with the median, interquartile range and the minimum and maximum values.

Figure 8 shows that the CONUS-based ER and BA graphs have 52% and 51% lower wavelength requirements than the SNR-BA graphs. Similarly, NSFNET-based ER and BA graphs have 31% and 23% lower wavelength requirements, than the SNR-BA graphs. The ER and BA graphs appear to have a structural advantage, in terms of wavelength requirements, over the SNR-BA graphs, because of their smaller diameters and edge connections spanning larger parts of the graph.

Wavelength requirements used to be a key performance metric in networks, due to wavelengths being a scarce resource. From this point of view one can see that, in general, the ER and BA graphs for both the CONUS and NSFNET node locations *perform well*, whereas the SNR-BA graphs have larger wavelength requirements. The CONUS topology, published more recently and now considered more representative of real networks [31], has a wavelength requirement that agrees more with that of the SNR-BA graphs. NSFNET which has lower wavelength requirements, in agreement with graphs generated with ER graphs, is now considered to be less applicable. Indeed, wavelength requirements, are generally only explained by the structural part of the network, neglecting the physical properties in the network. The overarching goal of network design is to maximise throughput for the entire network, which requires physical transmission impairments to be taken into account.

B. Maximum Achievable Throughput

To calculate the throughput of optical networks, fibre linear and nonlinear transmission impairments need to be taken into account. Unlike the SNR derived in Section 2E, here we do not use optimal launch powers, but apply a uniform launch power of 0dBm (see for example [43]) for all wavelength channels. To calculate the SNR a closed form GN model was used [43], where the SNR of a lightpath i is given by Eq. (22) . Here P_i is the launch power, η_n is the nonlinear coefficient that can be calculated following equation 5 in [43] and P_{ASE} is the power of the amplified spontaneous emission.

$$SNR_i \approx \frac{P_i}{P_{ASE} + \eta_n P_i^3} \tag{22}$$

For this set of simulations, all links were assumed to be multiples of 80km standard single mode fibre spans with $\beta = 0.2\frac{dB}{km}$, $D = 18\frac{ps}{mm\cdot km}$ and $\gamma = 1.2\frac{1}{W\cdot km}$. In between spans, erbium-doped fibre amplifiers, each with a noise figure of 4dB was used (as in [19]), although any practical value larger than the fundamental limit of 3dB can be used for this analysis. As before, Nyquist-spaced wavelength division multiplexed (WDM) channels of 16 GBd were used. They were interfaced with colourless, directionless and contentionless, reconfigurable optical add-drop multiplexers (CDC-ROADM) over a constrained C-band (1530-1570nm) optical bandwidth. The losses, filtering effects and amplification needs of the ROADMs were not considered in this work. This is a limitation of the physical model of the network and will be investigated in future work to give a more accurate representation of signal degradation.

To maximise the nonlinear throughput, an ILP based on [40] was used. This maximises the throughput given a uniform bandwidth across all node-pairs. To solve the ILP, the k-shortest paths were calculated between all node-pairs, after which the SNR for these paths was calculated assuming the worst case SNR: that of a centre channel in a fully loaded link, with 0 dBm launch power per channel, following Eq. (22). Using this SNR, their achievable capacity is calculated via Shannon's theorem [45]. Although the Shannon capacity is the upper bound and there are many forward error correction (FEC) codes that represent what is achievable, it is used here to illustrate the difference in achievable throughputs between networks [46, 47]. This set of capacities is referred to as $C_{z,k}$, where $z \in Z$ is a node pair (i, j) and $k \in K$ is a path. T_z is the normalised traffic matrix, in our case simply kept at uniform across the node pairs and c is a continuous integer variable over which one tries to find an RWA using $\delta_{z,k,w}$ such that every node-pair is able to route proportionally the same amount of bandwidth determined by T_z. Eq. (23) ensures that every node pair routes at least the product of the traffic matrix and the throughput multiplier c and Eq. (24) describes the objective of maximising c. All the previous constraints need to be followed, to ensure a feasible RWA set. Using this ILP the RWA throughput can be maximised given zero-blocking and uniform bandwidth demand.

$$\sum_w \sum_k \delta_{z,k,w} C_{z,k} - cT_z \geq 0 \quad \forall \quad z \in Z \tag{23}$$

$$c_{max} = max(c) \tag{24}$$

To calculate the resultant throughput for the RWA, we calculate the accumulated SNR for each lightpath assignment. A lightpath $i = \{p_i, \lambda_i\} \in R$ has a path p_i and a wavelength λ_i associated with it and is part of the set of lightpaths for a routing R. To calculate the capacity for this lightpath, one first needs to take into account the edges along which the lightpath travels and their respective SNR values. Using Eq. (22) and the state of the network i.e. the knowledge of which wavelengths are present on which links, one can calculate the $SNR_{(i,e)}$ value on each of the links $e \in p_i$. The total SNR of that path is then calculated by taking the inverse sum of the NSR (inverse of SNR) values of each link traversed by the path p_i, shown in Eq. (25)

$$SNR_i = \left(\sum_{e \in p_i} \frac{1}{SNR_{(i,e)}} \right)^{-1} \tag{25}$$

This SNR is then used to calculate the maximum achievable data rate over this lightpath using Eq. (26) [45]. It can be seen that the capacity of a lightpath mainly depends on the SNR of that path, which, in turn, depends on the length and congestion along the path. B_{CH} represents the channel bandwidth used, which is kept constant at 16 GHz for all channels.

$$C_i = 2B_{CH} log_2(1 + SNR_i) \tag{26}$$

$$C = \sum_{i \in R} C_i \tag{27}$$

The throughputs for all the lightpaths, allocated to satisfy the demand, were calculated and summed, as in Eq. (27) to give the total achievable throughput of the RWA for a particular network.

The maximum uniform throughput (T) of the ER, BA and SNR-BA graphs based on the CONUS and NSFNET topologies was calculated and is shown in figure 9. It can be seen that it is now the SNR-BA graphs that, on average, perform 48% better than the ER graphs and 43% better than the BA graphs based on the CONUS topology, despite the greater number of edges in the ER and BA graphs. For the NSFNET-based graphs, the SNR-BA graphs, on average, outperformed the BA and ER graphs by 46% and 27%, respectively. Therefore, it is clear that the ER and BA graphs, on average, perform worse than the SNR-BA graphs for both example networks.

This drop in performance between the ER and BA graphs compared to the SNR-BA graphs is the result of longer path lengths in the former. Figure 10a shows the distribution of average path lengths for all the generative models based on the CONUS network and 10b for the NSFNET graphs. The paths in the CONUS-based ER and BA graphs are on average 215% and 187%, respectively, longer than those taken over the SNR-BA graphs. For the NSFNET-based graphs, although shorter, the signals travel 95% and 98% further over the ER and BA graphs compared to the SNR-BA graphs. This difference in distances, and the associated transmission penalties, dominate the achievable throughput, and at these distances the structural advantages of the ER and BA graphs do not translate into larger throughputs.

It should be noted that this difference in edge lengths between the SNR-BA graphs and the ER/BA graphs, results in the difference in the total deployed fibre lengths. The NSFNET-based SNR-BA graphs use 53% and 47% less in total fibre compared to their ER and BA counterparts, respectively. For the CONUS-based SNR-BA graphs, this difference is even larger, saving 72%

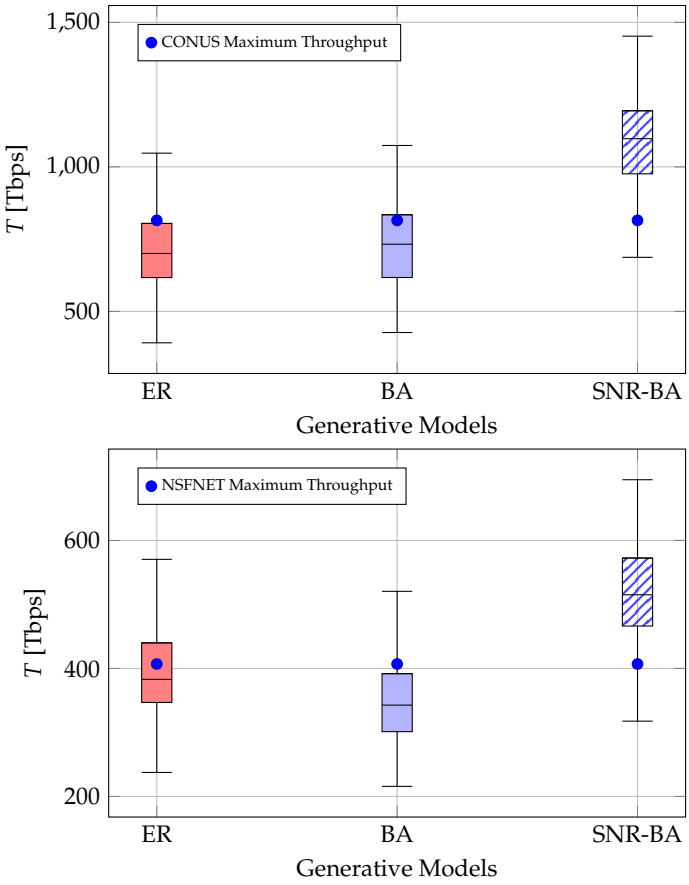

Fig. 9. Maximum uniform throughput (T) of NSFNET and CONUS based topologies for the graphs generated by ER, BA and SNR-BA generative graph models.

and 68% of total fibre compared to the ER and BA graphs respectively. The SNR-BA graphs are on average able to achieve higher throughput, which directly translates into lower blocking probability, whilst deploying less fibre.

In calculating throughput, uniform bandwidth demand was assumed, while a uniform set of connections is assumed when calculating the wavelength requirements. This difference is significant as longer lightpaths with lower SNR values, will inherently need to route more connections to satisfy a given bandwidth demand. This difference in the number of connections required to satisfy a given bandwidth demand is driven by the physical properties of the network. This is one of the factors that reduces the structural advantage of the ER and BA graphs, observed when evaluating wavelength requirements.

Due to this, the number of lightpaths established in the RWAs (λ) is not significantly different between the SNR-BA graphs and those of ER and BA. Namely, for the NSFNET based graphs, the SNR-BA graphs established on average only 8% fewer lightpaths compared to the ER graphs and 7% more lightpaths compared to the BA graphs. For the CONUS based graphs, the SNR-BA graphs established only 11% and 10% fewer lightpaths, on average, than the ER and BA graphs. This is significant, when looking at figure 11, where the distributions of throughput per lightpath ($\frac{T}{\lambda}$) is shown for each generative model. Here the same trend as before can be seen, where for the CONUS based graphs, the SNR-BA graphs achieve 66% and 59% greater throughput per lightpath than the ER and BA graphs respectively. This is a

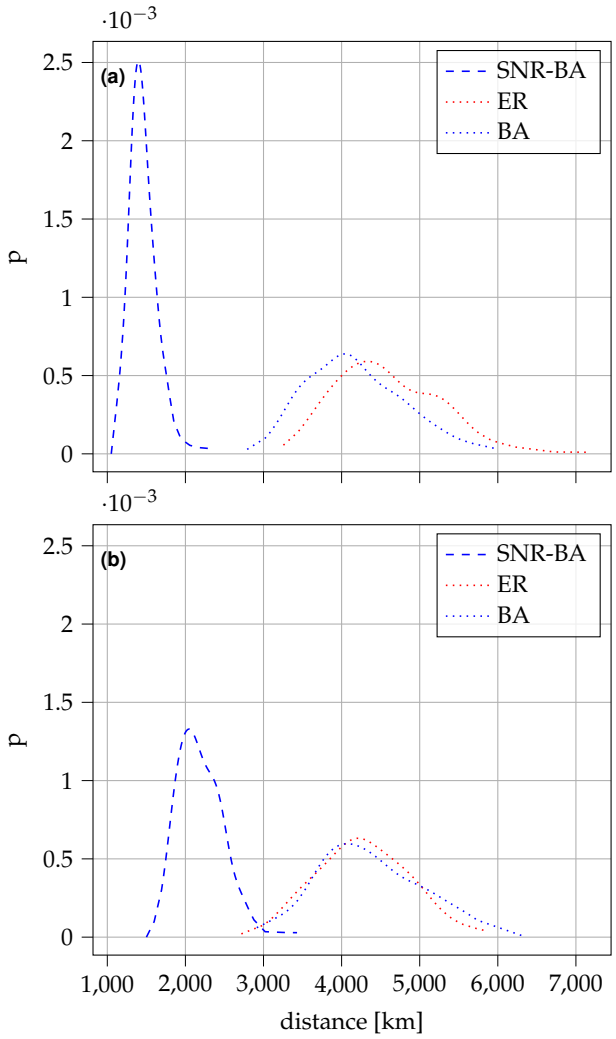

Fig. 10. Probability distributions of the average path lengths in each of the solved RWAs for graphs generated by the ER, BA and SNR-BA models, using the node-positions from (a) 30-node CONUS topology and (b) NSFNET.

significant difference, solely down to the physical properties of the network.

In summary, the structural advantage of smaller wavelength requirements seen in ER and BA graphs, does not equate to higher throughputs due to the increased path lengths, and associated transmission penalties. SNR-BA graphs, however, choose shorter edges and minimise the path lengths and, therefore, help maximise throughput when the distance dominates the achievable throughput in the network. This is summarised in figure 12, where the average values of the edge distances, total deployed fibre length, number of edges, lightpath lengths, throughput per lightpath and total throughput are plotted for each of the generative models and the corresponding network they are based on. For distance and edge values, the scales are reversed so that points on the peripheral are better (shorter/smaller), whereas for throughput values the points on the peripheral are larger. It is clear that the SNR-BA graphs achieve similar edge numbers, total deployed fibre, edge lengths compared to the original networks, yet lower average lightpath lengths which achieve higher total throughput and throughput per lightpath, than the original and ER/BA networks.

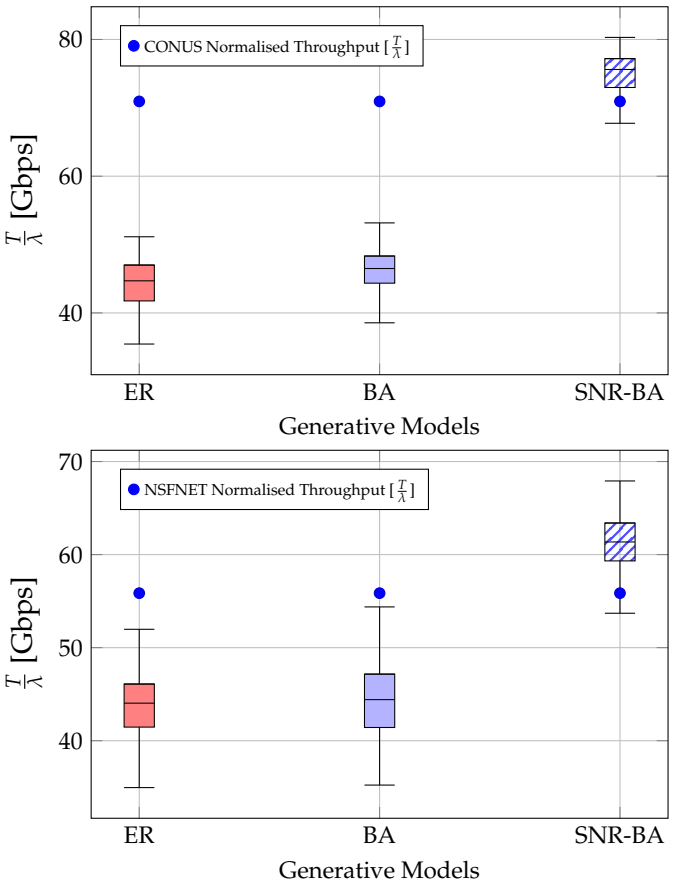

Fig. 11. Throughput (T) per lightpath established (λ) in RWA of NSFNET and CONUS based topologies for the graphs generated by ER, BA and SNR-BA generative graph models.

C. Scaling distance & throughput

By looking at different distance scales of the generated graphs, one can analyse at which point the structural advantages of the ER and BA graphs are beneficial to the total achievable throughput of a network. To do this the distances of all edges were scaled by some factor $x \in [0.01, 1]$, after which the throughput was recalculated. For each value of x another 200 values of throughput was calculated per generative model. These were then averaged and plotted in figure 13, along with the throughput scaling results of the original CONUS and NSFNET graphs, to understand at which point structural properties benefit total achievable throughput in optical networks.

In figure 13a one can see that the BA and ER graphs outperform the SNR-BA and original CONUS graph when looking at distance scales where $x < 0.1$. This is reasonable since BA and ER graphs are, on average, able to establish more lightpaths than the SNR-BA and CONUS graphs, however the achievable throughput of these lightpaths suffers due to long edge distances. Once these distances become too small to effectively diminish the throughput of these lightpaths, the ER/BA start outperforming the SNR-BA and CONUS network. This is down to the structure of the ER/BA graphs as they have more and better connected edges than the SNR-BA graphs and original CONUS network.

Graphs generated for NSFNET node locations were able to achieve the same number of edges due to higher connectivity of the original network. Figure 13b shows the average throughput

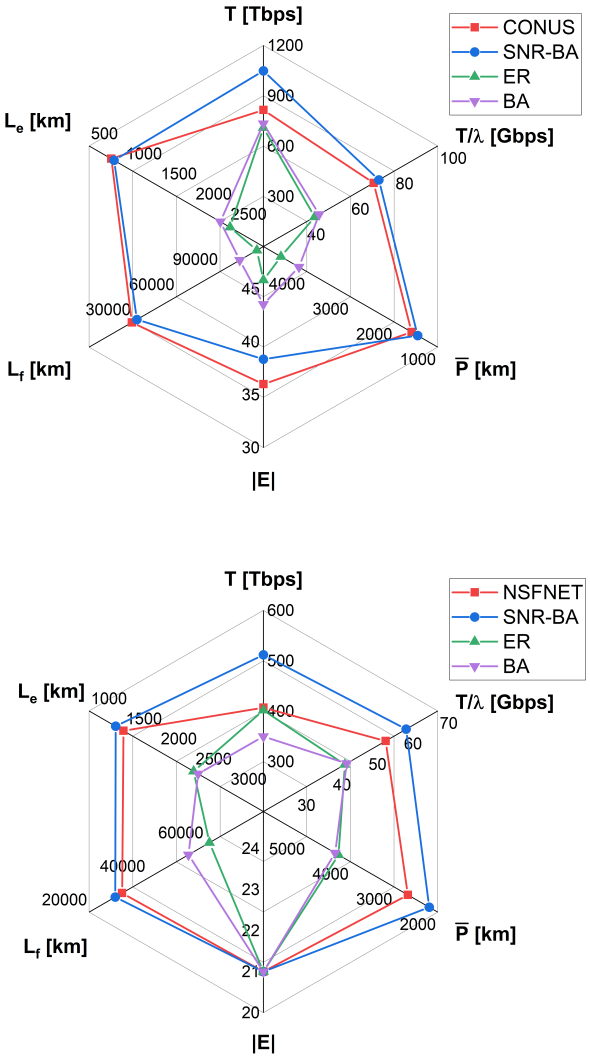

Fig. 12. Radar plots showing (clockwise) throughput (T), throughput per lightpath (T/λ), average lightpath length ($\bar{P}$), number of edges ($|E|$), total fibre deployed (L_f), the averages of the edge length (L_e).

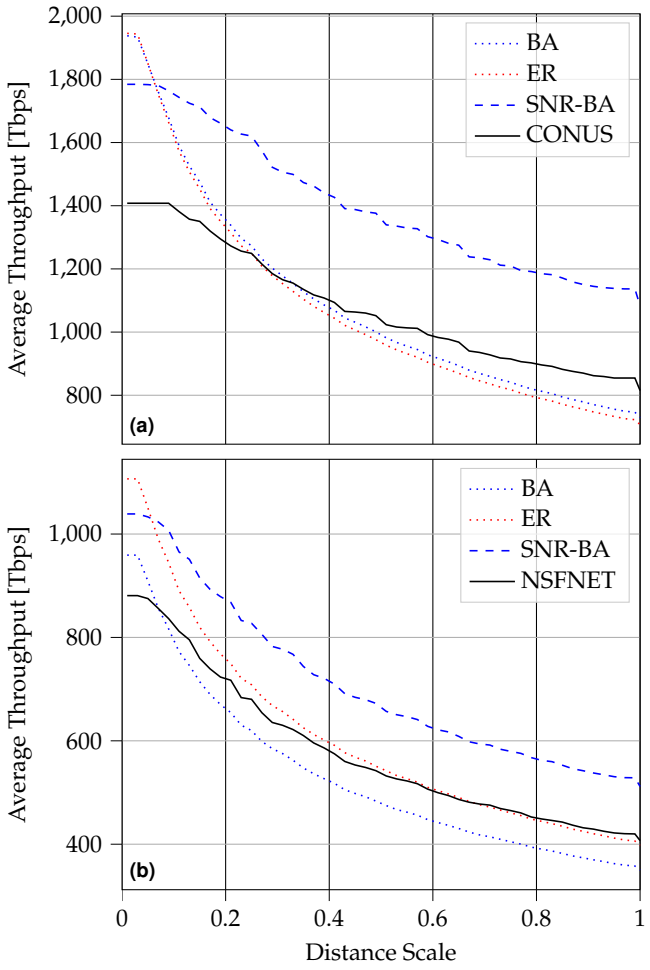

Fig. 13. Average throughput ($\bar{T}$) calculated for the distance scale x of the graphs generated by the ER, BA and SNR-BA generative graph models based on node-positions taken from (a) 30-node CONUS network and (b) NSFNET network.

scaling of the networks generated for the various generative models and for the original NSFNET. Here the BA graphs are the worst performing graphs, only outperforming the original NSFNET at smaller distance scales. The ER graphs however outperform the SNR-BA graphs again at scales of $x < 0.1$. This is due to the ER graphs being more uniformly connected across the graphs, leading to lower wavelength requirements and the ability to establish more lightpaths due to structural advantages. This comes at the cost of large link and path lengths. However, at smaller scales the distance penalties have a significantly lower or no impact on throughput. As the edges are scaled, the path lengths grow quicker for the ER and BA graphs due to larger edge lengths. This in turn reduces their average performance when looking at the original distance scales - causing SNR-BA graphs to on average out perform them in the end.

Therefore, when looking at optical core network distance scales, the SNR-BA graphs that use edge probabilities derived from the SNR between nodes, create topologies that perform better, on average, than the ER and BA graphs, due to their shorter path lengths that are more resistant to scaling of distances and nonlinearities. However, BA and ER graphs tend to create graphs with lower wavelength requirements, which create better performing networks for smaller distance scales. This may not be applicable to core optical networks but may be valuable when designing data centre or access networks.

6. CONCLUSIONS

This paper investigated the application of generative graph models for the analysis of structural and physical properties of optical networks. The paper compared the conventional Erdos-Renyi, Barabasi-Albert and Waxman models and a newly proposed model, SNR-BA using 5000 graphs for each model. The SNR-BA is designed to take into account the physical properties of real networks, and the associated linear and nonlinear optical fibre distortions, by incorporating the signal-to-noise ratio term in graph generation. It was found that the SNR-BA model resulted in graphs which were structurally most similar to the real optical core networks for every metric. This is significant as it highlighted the importance of including physical properties in graph generation for optical core network simulation studies.

A study of two sample core topologies, CONUS and NSFNET,

was carried out to compare their performance with that of the generative models. Graphs were generated based on these networks following the ER, BA and SNR-BA models. The metrics of wavelength requirements and network throughput were used to assess the performance. The results showed that on average the ER and BA models tended to have lower wavelength requirements, as they were graphs with shorter diameters and higher algebraic connectivities, whereas the SNR-BA graphs had higher mean wavelength requirements. The maximum achievable throughput was evaluated for all of the graphs, which showed that the SNR-BA graphs tended to outperform the ER and BA graphs. Overall, the proposed SNR-BA model was shown to be most accurate in reflecting core optical network structures, and offered throughput increases of 29% and 26%, against the original CONUS and NSFNET topologies. Throughput increases of 48% and 27% were observed compared to the ER graphs, and 43% and 46% compared to the BA graphs. These increases are explained by the edges and path lengths of ER and BA graphs which are, on average, more than a factor of two longer than that of the SNR-BA graphs. The SNR-BA graphs, thus, could lead to significant savings in total installed fibre for core networks. Using different distance scales, it was shown that for smaller networks, the structural advantage of the ER and BA graphs led to them outperforming the SNR-BA graphs. As the distances grew, however, the SNR-BA graphs tend to perform better due to their shorter edge distances.

In summary, the structural advantage of lower minimum wavelength requirements in ER and BA graphs, does not translate to higher throughput due to the increased path lengths, and associated transmission penalties. The proposed model of SNR-BA, however, chooses shorter edges, minimising path lengths throughout the graph, helping to maximise throughput when the distance dominates the achievable throughput in the network. This research provides insight into edge choices and their effect on final achievable throughput which could be harnessed for future intelligent optical network design. Further investigation of the relationship between structure, physical properties and non-uniform traffic distributions and their effect on performance is vital to the design of optimal optical network topologies.

Network	N	E	connectivity
Abilene	12	15	0.227
Atlanta	15	22	0.209
Cost266	37	57	0.085
France	25	45	0.150
Geant	22	36	0.155
Germany50	50	88	0.071
Giul39	39	86	0.116
india35	35	80	0.134
janos-us-ca	26	42	0.129
janos-us	39	61	0.082
nobel-eu	28	41	0.108
nobel-germany	17	26	0.191
nobel-us	14	21	0.230
norway	27	51	0.145
pioro40	40	89	0.114
polska	12	18	0.272
sun	27	51	0.145
ta1	24	51	0.1845
ta2	65	108	0.051
zib54	54	80	0.055
CORONET-CONUS	75	99	0.035
CORONET-GLOBAL	100	136	0.027
CORONET-CONUS-60-ONDP	60	77	0.043
CORONET-CONUS-30-ONDP	30	36	0.082
CORONET-CONUS-60	60	79	0.044

Table 2. Table of real networks used in structural comparison of generative graph models.

FUNDING

Financial support from UK EPSRC Doctoral Training Programme and the Programme Grant TRANSNET (EP/R035342/1) as well as the Microsoft 'Optics for the Cloud' programme, is gratefully acknowledged.

REFERENCES

1. S. Baroni and P. Bayvel, "Wavelength requirements in arbitrarily connected wavelength-routed optical networks," J. Light. Technol. **15**, 242–251 (1997).
2. C. Fenger, E. Limal, U. Gliese, and C. J. Mahon, "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*, G. Pujolle, H. Perros, S. Fdida, U. Körner, and I. Stavrakakis, eds. (Springer, Berlin, Heidelberg, 2000), Lecture Notes in Computer Science, pp. 168–175.
3. B. Châtelain, M. P. Bélanger, C. Tremblay, F. Gagnon, and D. V. Plant, "Topological Wavelength Usage Estimation in Transparent Wide Area Networks," IEEE/OSA J. Opt. Commun. Netw. **1**, 196–203 (2009).
4. P. Yuan and A. Xu, "The Influence of Physical Network Topologies on Wavelength Requirements in Optical Networks," J. Light. Technol. **28**, 1338–1343 (2010).
5. E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, M. Secondini, F. R. Kschischang, A. Lord, J. Prat, I. Tomkos, J. E. Bowers, S. Srinivasan, M. Brandt-Pearce, and N. Gisin, "Roadmap of optical communications," J. Opt. **18**, 063002 (2016).
6. D. J. Ives, P. Bayvel, and S. J. Savory, "Physical layer transmitter and routing optimization to maximize the traffic throughput of a nonlinear optical mesh network," in *2014 International Conference on Optical Network Design and Modeling*, (2014), pp. 168–173.
7. X. Wan, N. Hua, and X. Zheng, "Dynamic Routing and Spectrum Assignment in Spectrum-Flexible Transparent Optical Networks," J. Opt. Commun. Netw. **4**, 603–613 (2012).
8. I. Martin, S. Troia, J. A. Hernandez, A. Rodriguez, F. Musumeci, G. Maier, R. Alvizu, and O. Gonzalez de Dios, "Machine Learning-Based Routing and Wavelength Assignment in Software-Defined Optical Networks," IEEE Transactions on Netw. Serv. Manag. **16**, 871–883 (2019).
9. X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu, and S. J. B. Yoo, "DeepRMSA: A Deep Reinforcement Learning Framework for Routing, Modulation

and Spectrum Assignment in Elastic Optical Networks," J. Light. Technol. **37**, 4155–4163 (2019). ArXiv: 1905.02248.
10. F. Shirin Abkenar and A. Ghaffarpour Rahbar, "Study and Analysis of Routing and Spectrum Allocation (RSA) and Routing, Modulation and Spectrum Allocation (RMSA) Algorithms in Elastic Optical Networks (EONs)," Opt. Switch. Netw. **23**, 5–39 (2017).
11. E. Archambault, N. Alloune, M. Furdek, Z. Xu, C. Tremblay, A. Muhammad, J. Chen, L. Wosinska, P. Littlewood, and M. P. Belanger, "Routing and Spectrum Assignment in Elastic Filterless Optical Networks," IEEE/ACM Transactions on Netw. **24**, 3578–3592 (2016).
12. H. Choi, S. Subramaniam, and H.-A. Choi, "On double-link failure recovery in WDM optical networks," in *Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies,* vol. 2 (2002), pp. 808–816 vol.2. ISSN: 0743-166X.
13. M. Moharrami, A. Fallahpour, H. Beyranvand, and J. A. Salehi, "Resource Allocation and Multicast Routing in Elastic Optical Networks," IEEE Transactions on Commun. **65**, 2101–2113 (2017).
14. K. Walkowiak, M. Klinkowski, and P. Lechowicz, "Dynamic routing in spectrally spatially flexible optical networks with back-to-back regeneration," IEEE/OSA J. Opt. Commun. Netw. **10**, 523–534 (2018).
15. H. Wu, F. Zhou, Z. Zhu, and Y. Chen, "Interference-and-security-aware distance spectrum assignment in Elastic Optical Networks," in *2016 21st European Conference on Networks and Optical Communications (NOC),* (2016), pp. 100–105.
16. M. W. Ashraf, S. M. Idrus, F. Iqbal, and R. A. Butt, "On spatially disjoint lightpaths in optical networks," Photonic Netw. Commun. **36**, 11–25 (2018).
17. A. Pages, J. Perello, S. Spadaro, and G. Junyent, "Strategies for Virtual Optical Network Allocation," IEEE Commun. Lett. **16**, 268–271 (2012).
18. D. B. Depizzol, J. Montalvão, F. de Oliveira Lima, M. H. M. Paiva, and M. E. V. Segatto, "Feature selection for optical network design via a new mutual information estimator," Expert. Syst. with Appl. **107**, 72–88 (2018). Publisher: Elsevier.
19. D. Semrau, S. Durrani, G. Zervas, R. I. Killey, and P. Bayvel, "On the relationship between network topology and throughput in mesh optical networks," (2020).
20. R. J. Vincent, D. J. Ives, and S. J. Savory, "Scalable Capacity Estimation for Nonlinear Elastic All-Optical Core Networks," J. Light. Technol. **37**, 5380–5391 (2019).
21. C. Pavan, R. M. Morais, J. R. Ferreira da Rocha, and A. N. Pinto, "Generating Realistic Optical Transport Network Topologies," IEEE/OSA J. Opt. Commun. Netw. **2**, 80–90 (2010).
22. E. K. Çetinkaya, M. J. F. Alenazi, Y. Cheng, A. M. Peck, and J. P. G. Sterbenz, "A comparative analysis of geometric graph models for modelling backbone networks," Opt. Switch. Netw. **14**, 95–106 (2014).
23. J. Velinska, M. Mirchev, and I. Mishkovski, "Optical networks' topologies: costs, routing and wavelength assignment," Opt. networks p. 10 (2017).
24. P. Bayvel, R. Luo, R. Matzner, D. Semrau, and G. Zervas, "Intelligent design of optical networks: which topology features help maximise throughput in the nonlinear regime?" (2020), p. 4.
25. P. Erdos and A. Renyi, "On the Evolution of Random Graphs," in *Publication of the Mathematical Institute of the Hungarian Academy of Sciences,* (1960), pp. 17–61.
26. A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," Science **286**, 509–512 (1999). ArXiv: cond-mat/9910332.
27. B. Waxman, "Routing of multipoint connections," IEEE J. on Sel. Areas Commun. **6**, 1617–1622 (1988).
28. K. G. Ruben and R. R. Sokal, "A new statistical approach to geographic variation analysis," Syst. Zool. pp. 259–278 (1969).
29. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-Model of Fiber Non-Linear Propagation and its Applications," J. Light. Technol. **32**, 694–721 (2014).
30. S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessäly, "SNDlib 1.0–Survivable Network Design Library," in *Proceedings of the 3rd International Network Optimization Conference (INOC 2007), Spa, Belgium,* (2007). Http://sndlib.zib.de, extended version accepted in Networks, 2009.
31. J. M. Simmons, *Optical Network Design and Planning*, Optical Networks (Springer International Publishing, Cham, 2014).
32. S. De Maesschalck, "Pan-European Optical Transport Networks: An Availability-based Comparison," Photonic Netw. Commun. **5**, 203–225 (2002).
33. S. De Maesschalck, "Network Aspects (NA); Availability performance of path elements of international digital paths," Tech. Rep. REN/NA-042140 (3uo00ioo.PDF), European Telecommunications Standards Institute, France (1998).
34. D. Fay, H. Haddadi, S. Uhlig, A. Moore, R. Mortier, and A. Jamakovic, "Weighted spectral distribution," Univ. Cambridge, Comput. Lab. Tech. Rep. UCAM-CL-TR-729, Sep (2008).
35. R. C. Wilson and P. Zhu, "A study of graph spectra for comparing graphs and trees," Pattern Recognit. **41**, 2833–2841 (2008).
36. C. Fan, *Spectral Graph Theory , by Fan Chung* (AMS, 2006).
37. E. K. Çetinkaya, M. J. F. Alenazi, A. M. Peck, J. P. Rohrer, and J. P. G. Sterbenz, "Multilevel resilience analysis of transportation and communication networks," Telecommun. Syst. **60**, 515–537 (2015).
38. J. A. Nelder and R. Mead, "A simplex method for function minimization," The computer journal **7**, 308–313 (1965). Publisher: Oxford University Press.
39. J. Wang, W. Tsang, and G. Marsaglia, "Evaluating kolmogorov's distribution," J. Stat. Softw. **08** (2003).
40. D. J. Ives, P. Bayvel, and S. J. Savory, "Routing, modulation, spectrum and launch power assignment to maximize the traffic throughput of a nonlinear optical mesh network," Photonic Netw. Commun. **29**, 244–256 (2015).
41. N. M. Garcia, M. Pereira, M. M. Freire, P. P. Monteiro, and P. Lenkiewicz, "Performance of Optical Burst Switched Networks for Grid Applications," in *International Conference on Networking and Services (ICNS '07),* (IEEE, Athens, Greece, 2007), pp. 120–120.
42. S. Baroni, R. J. Gibbens, and P. Bayvel, "On the number of wavelengths in arbitrarily-connected wavelength-routed optical networks," in *Optical Networks and Their Applications,* (OSA, Washington D. C., 1998), p. MN2.
43. D. Semrau, R. I. Killey, and P. Bayvel, "A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering," J. Light. Technol. **37**, 1924–1936 (2019).
44. J. Y. Yen, "Finding the K Shortest Loopless Paths in a Network," Manag. Sci. **17**, 712–716 (1971). Publisher: INFORMS.
45. C. E. Shannon, "A Mathematical Theory of Communication," The Bell Syst. Tech. J. **27**, 379–423 (1948).
46. R.-J. Essiambre, G. Kramer, P. Winzer, G. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks," Light. Technol. J. . **28**, 662–701 (2010).
47. A. Mecozzi and R.-J. Essiambre, "Nonlinear Shannon Limit in Pseudolinear Coherent Systems," J. Light. Technol. **30**, 2011–2024 (2012). Conference Name: Journal of Lightwave Technology.