Expanding Graph Neural Networks for Ultra-Fast Optical Core Network Throughput Prediction to Large Node Scales

Robin Matzner[(1)], Ruijie Luo[(1)], Georgios Zervas[(1)], Polina Bayvel[(1)]

[(1)] University College London, robin.matzner.19@ucl.ac.uk

Abstract *Using maximum achievable throughput as an objective, message passing neural networks (MPNN) are applied to larger optical networks (25-100 nodes), enabling physical properties-aware large-scale topology optimisation in record time, reducing computation time by 5 orders of magnitude, with close to perfect throughput correlation ($\rho = 0.986$).*

Introduction

Optical core networks underpin the digital communications infrastructure, connecting datacentres to each other and many millions of service-users throughout the world. The maximum achievable throughput (MAT) of any optical network is mostly constrained by its physical topology. To ensure that optical networks can sustain future demands, the challenge is to maximise the network MAT, given a specific data demand matrix. Calculating this exact value for an optical network is, however, an NP-hard problem, making it impossible to include it in the design process as a cost function[1]. Integer linear programming (ILP) formulations are used to find the exact solution, however they are infeasible for larger networks (> 30)[2]–[4]. General optimisation frameworks, such as meta-heuristics, are more computationally efficient, however cannot guarantee optimal solutions and are associated with long computation times[5]–[8]. Hand-crafted routing specific heuristics are the only scalable option, yet these do not guarantee optimal solutions[9]. Therefore, there is a need for fast (order of milliseconds) and scalable (up to 100 nodes) MAT prediction, to enable its optimisation in future topology design and optimisation.

Message passing neural networks (MPNN) are a type of graph neural network (GNN), which are a form of geometric deep learning, where common deep learning concepts are adapted for relational data. Previously, MPNNs have been used to learn the relationship between the topology and the MAT of optical networks[10]. This was, however, only shown for very small topologies between 10 and 15 nodes, as an ILP was used to generate the training labels, which cannot scale to graphs with more than 30 nodes. Optical core networks, however, can reach 100+ nodes[1].

In this paper we expand the previous MPNN model to larger node scales and demonstrate its ability to learn accurate cost functions for networks up to 100 nodes. The difficulty with scaling to larger topologies however is the lack of training data that is accurate, as the ILP becomes infeasible (> 30 nodes), therefore, we first demonstrate that a routing heuristic, first-fit k-shortest paths (FF-kSP), previously proven to perform the best with regards to the ILP[9], achieves MAT trends that have a high linear correlation when compared to the ILP solutions. We then use this heuristic to produce MAT labels for 160,000 graphs in the range of 25-100 nodes. Using this training data, the MPNN is expanded to learn throughput distributions of graphs up to a 100 nodes, whilst reducing computation times by 5 orders of magnitude for the highest number of nodes.

This method allows for massive gains in throughput by allowing throughput to be directly optimised in the physical topology design, which is not possible to date. Hereby we enable future intelligent topology design taking into account the physical layer, the topology structure and their impact on the MAT of a topology.

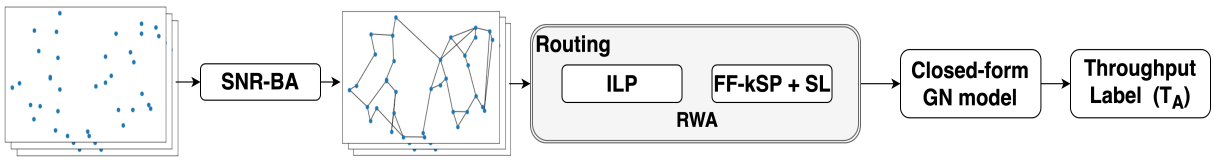

Fig. 1: Data generation process for the maximum achievable throughput labels. SL- sequential loading

Message Passing Neural Networks

The network is represented as a digraph denoted as $G(N, E)$, whereby N and E denote the sets of nodes and edges respectively. The network's nodes and edges have specific node and edge features x_n and e_{nu}, respectively, where $n, u \in N$ and $(n, u) \in E$. The node and edge features used in this work are degree, normalised node traffic and worst case noise-to-signal ratio (NSR) respectively. In addition to this, MPNNs use a set of abstract vectors, referred to as a node's hidden state, represented by h_n^t, t being a message passing iteration. The hidden states are node embeddings that record structural features from the rest of the graph.

MPNNs use 3 processes to produce embeddings that are then used for regression or classification: (i) message passing (ii) update (iii) readout. The MPNN repeats T message passing iterations, where both stages (i) and (ii) are iterated over. In (i) each network node acquires messages from its local neighbourhood ($\mathcal{N}(n)$). This message is given by a message function, that uses the original edge features and hidden state vectors ($M_t(h_n^t, h_u^t, e_{nu})$). To form the final message of a node n (m_n^{t+1}), the output from the message function are summed over the neighbourhood of n. In (ii) one uses an update function ($U_t(h_n^t, m_n^{t+1})$), which iteratively updates the node hidden states (h_n^{t+1}) of each node. After finishing T - chosen in the order of the diameter of the graphs - rounds of (i) and (ii), the hidden states are aggregated and fed through a readout function which makes a graph level prediction - i.e. a throughput prediction. The exact formulation of message, update and readout functions is detailed in[10] and based on[11]. Supervised learning is then used to train these three functions, end-to-end, on a large dataset of about 160,000 graphs in the range of 25-100 nodes. The next section describes how the graphs were generated and how their MAT was calculated.

Data Generation

A dataset of about 160,000 unique graphs, with 25 to 100 nodes, was generated to train two MPNN models. The node locations were chosen uniformly over a grid the size of north-America, from which the graphs were then generated via the SNR-BA model[12],[13]. In previous work[10], an ILP formulation was used to calculate the true MAT, this however does not scale to graphs larger than 30 nodes, therefore FF-kSP in conjunction with sequential loading was investigated as a suitable replacement, due to previous work showing its good results with regards to ILP solutions[9]. Sequential loading works by sequentially adding more and more demand to the network until blocking is achieved.

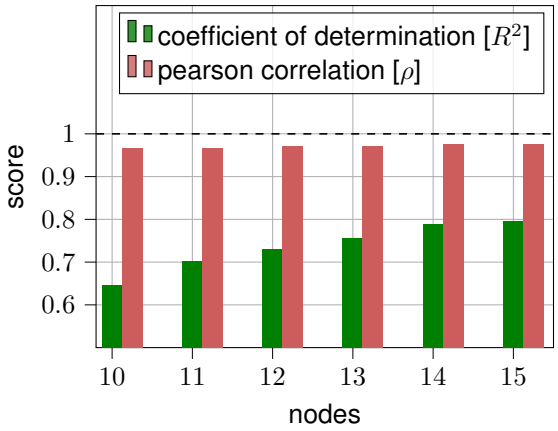

Fig. 2: Coefficient of determination and Pearson's correlation coefficient calculated between the FF-kSP and ILP throughput distributions for 10-15 nodes.

When looking at the throughput labels generated for 10 to 15 node graphs by the ILP and those generated by FF-kSP with sequential loading; high linear correlation is present. This can be realised by looking at the coefficient of determination (R^2), which measures accuracy, and the Pearson's correlation coefficient (ρ), which measures linear correlation, where 1 is a perfect score for both. Figure 2 plots both R^2 and ρ calculated between the ILP and FF-kSP calculated throughput distributions for graphs of 10-15 nodes. One can see that although the R^2 values are generally low (below 0.8 mostly), indicating low predictive accuracy; the ρ values are all close to 1, meaning high linear correlation to the original ILP data. Therefore, using FF-kSP with sequential loading can correctly predict the trend of MAT and, therefore, was chosen as the methodology for labelling graphs larger than 20 nodes. The 160,000 graphs were thus labelled with their corresponding MAT, by finding the RWA that maximises the throughput, using sequential loading and FF-kSP, given a uniform traffic distribution.

Finally, the MAT was calculated using a closed form Gaussian noise (GN) model[14] to calculate the SNR of the different lightpaths. 32 GBd Nyquist spaced channels over the C-band (1530-70 nm) were used, giving 156 possible wavelengths. Edges within the network 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 $\gamma = 1.2\frac{1}{W \cdot km}$, amplified with identical erbium-doped fibre amplifiers (noise figure of 4dB). Shan-

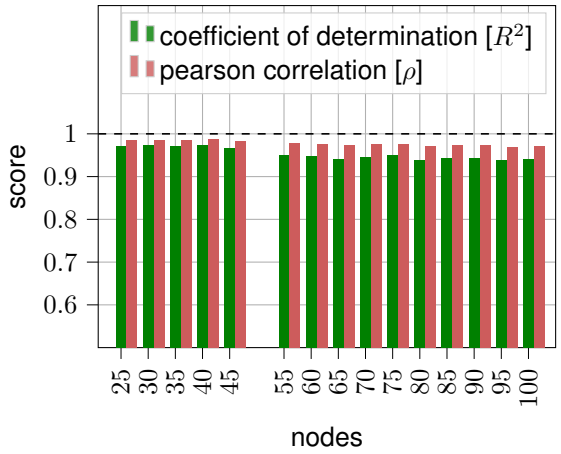

Fig. 3: Coefficient of determination and Pearson's correlation coefficient calculated between MPNN and FF-kSP throughput distributions for 25-100 nodes.

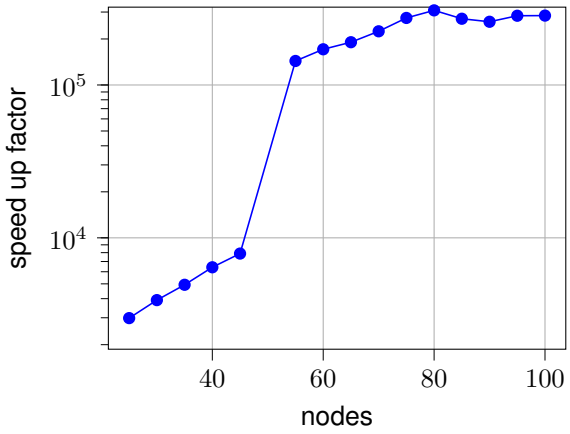

Fig. 4: Speed up factor between MAT calculation using FF-kSP with sequential loading and the MPNN model.

non's capacity formula was then used to calculate the throughput of lightpaths, which were then summed[15].

The process of data labelling for both small and large networks is visualised in figure 1. As the network scale grows, so did the sparsity of the graph. Therefore, to allow for enough capacity, every edge modelled had 4 and 16 fibres for graphs of 25-45 and 55-100 nodes respectively. Using these throughput labels, two MPNN models were trained, one for 25-45 node scales (alpha model) and one for 55-100 node scales (beta model).

Results

To test the accuracy and performance of the MPNN, a test dataset of 16,000 graphs was generated in the same manner as the training dataset. These graphs however are unseen by the training process. To measure the performance of the model, the coefficient of determination (R^2) and the Pearson's correlation (ρ) coefficient are used again. These are measured for each of the node scales tested between 25-100 nodes and plotted in figure 3. One can clearly see that both alpha and beta models have R^2 and ρ values close to one, none dropping below 0.93 and 0.96 respectively. Generally the alpha model predicts the trend and actual throughput values better, with an average R^2 and ρ score of 0.973 and 0.986 respectively, compared to 0.948 and 0.974 respectively for the beta model. This essentially boils down to the amount of data available for training for each of these models, where a larger proportion of training data was generated for the alpha model as the node scales were smaller and number of fibres per edge used were also less, therefore taking less time to generate.

The overwhelming advantage of this model is the low computational complexity required for the inference of a network's MAT via the MPNN model. To demonstrate this, the time taken for estimating the MAT via FF-kSP with sequential loading and using the MPNN was measured and the speed up factor plotted in figure 4. Here one can see that the benefit becomes apparent when the node scales approach 55 to 100 nodes. Here, the MPNN takes about 5 orders of magnitude less time than the heuristic, as the heuristic has to load the network many times before it realises the MAT. In addition, there is a big jump in computation time required for FF-kSP with the larger node scales (55-100) due to more fibres being used. These speed-ups allow for the MPNN to be used in an optimisation process as a cost function and, therefore, enable throughput optimisation of optical networks.

Conclusions

This paper expanded the previous MPNN model from 10-15 node graphs[10] to graphs of 25-100 nodes. This was done by initially showing that FF-kSP has a high linear correlation to the original ILP data and therefore was chosen to generate training data for these larger graphs. 160,000 graphs were generated with their corresponding MAT labels (calculated via FF-kSP) and used to train two MPNN models that then were tested against unseen test sets. Both models showed high accuracy, with R^2 and ρ values of 0.973 and 0.986 respectively for the best model. In addition both models showed computation speed ups of up to 5 orders of magnitude compared to FF-kSP. This model therefore enables future intelligent optical network design, that can maximise the throughput of optical networks, previously not possible.

Acknowledgements

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

References

[1] J. M. Simmons, *Optical Network Design and Planning*, en, ser. Optical Networks. Cham: Springer International Publishing, 2014, ISBN: 978-3-319-05226-7 978-3-319-05227-4.

[2] 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 Network Communications*, vol. 29, pp. 244–256, Jun. 2015.

[3] R. Ramaswami and K. N. Sivarajan, "Routing and wavelength assignment in all-optical networks", *IEEE/ACM Transactions on Networking*, vol. 3, no. 5, pp. 489–500, Oct. 1995.

[4] S. Subramaniam and R. Barry, "Wavelength assignment in fixed routing WDM networks", in *Proceedings of ICC'97 - International Conference on Communications*, vol. 1, Jun. 1997.

[5] R. Rodriguez-Dagnino, E. Lopez-Caudana, H. Martinez-Alfaro, and J. Gonzalez-Velarde, "Simulated annealing and stochastic ruler algorithms for wavelength assignment planning in WDM optical networks", in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 6, Oct. 1999, 1015–1020 vol.6.

[6] N. Banerjee, V. Mehta, and S. Pandey, "A Genetic Algorithm Approach for Solving the Routing and Wavelength Assignment Problem in WDM Networks", in *Portuguese Conference on Artificial Intelligence*, Aug. 2017, p. 8.

[7] P. Wright, M. C. Parker, and A. Lord, "Minimum- and maximum-entropy routing and spectrum assignment for flexgrid elastic optical networking [invited]", *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 1, A66–A72, Jan. 2015.

[8] A. Rubio-Largo, M. A. Vega-Rodriguez, J. A. Gomez-Pulido, and J. M. Sanchez-Perez, "A Comparative Study on Multiobjective Swarm Intelligence for the Routing and Wavelength Assignment Problem", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, Nov. 2012.

[9] R. J. Vincent, D. J. Ives, and S. J. Savory, "Scalable Capacity Estimation for Nonlinear Elastic All-Optical Core Networks", *Journal of Lightwave Technology*, vol. 37, no. 21, pp. 5380–5391, Nov. 2019.

[10] R. Matzner, R. Luo, G. Zervas, and P. Bayvel, "Ultra-fast Optical Network Throughput Prediction using Graph Neural Networks", in *26th International Conference on Optical Network Design and Modelling*, May 2022, p. 3.

[11] K. Rusek and P. Cholda, "Message-Passing Neural Networks Learn Little's Law", *IEEE Communications Letters*, vol. 23, no. 2, pp. 274–277, Feb. 2019.

[12] R. Matzner, D. Semrau, R. Luo, G. Zervas, and P. Bayvel, "Making intelligent topology design choices: Understanding structural and physical property performance implications in optical networks [Invited]", *Journal of Optical Communications and Networking*, vol. 13, no. 8, pp. D53–D67, Aug. 2021.

[13] 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?", Nov. 2020, p. 4.

[14] 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", *Journal of Lightwave Technology*, vol. 37, no. 9, pp. 1924–1936, May 2019.

[15] C. E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal*, vol. 27, pp. 379–423, Oct. 1948.