River networks: An analysis of simulating algorithms and graph metrics used to quantify topology

Finnbar Lee | Kevin S. Simon | George L. W. Perry

School of Environment, The University of Auckland, Auckland, New Zealand

Correspondence
Finnbar Lee
Email: finnbar.lee@auckland.ac.nz

Handling Editor: Timothée Poisot

Abstract
1. River networks are frequently simulated for use in the development and testing of ecological theory. Currently, two main algorithms are used, stochastic branching networks (SBNs) and optimal channel networks (OCNs). The topology of these simulated networks and ‘real’ rivers is often quantified using graph theoretic metrics; however, to date, there has not been a comprehensive analysis of how these algorithms compare regarding graph theoretic metrics, or an analysis of metric redundancy and variability across dendritic ecological networks. We aim to provide guidance as to which algorithm and metrics should be used, and under what circumstances.

2. We performed an extensive simulation study in which we (a) identified orthogonal sets of metrics that describe the topology of real and simulated river networks, (b) analysed the relationship between algorithm hyper-parameters and node topology metrics, (c) determined whether simulated and real rivers are indistinguishable in their graph metric scores and (d) examined how patterns of species abundances compare across the three network types.

3. We identified two orthogonal sets of node metrics; those that describe centrality and those that describe neighbourhood characteristics. Both stochastic branching networks and optimal channel networks can reproduce network topology metric scores of real rivers, but this relationship is dependent on the algorithm hyper-parameters used. Finally, using a metapopulation model, we show that both SBNs and OCNs can reproduce ecological patterns of species abundances similar to those of real rivers.

4. SBNs and OCNs can replicate the node topology of real rivers. The choice of which algorithm to use will depend on the research aims, SBNs are faster to generate and more tractable, whereas OCNs can reproduce a wider variety of the characteristics of real rivers, but are more time-consuming to generate. When quantifying node topology in river networks, we recommend the orthogonal node metrics eccentricity, when interested in network centrality, and mean neighbour degree, when interested in local node importance.
1 INTRODUCTION

River networks exhibit intricate branching topologies and share properties such as critical scaling and self-similarity with other complex systems (Abed-Elmdoust et al., 2017; Albert & Barabási, 2002). There is theoretical and empirical evidence that network topology is important for maintaining metapopulation stability and influences patterns of biodiversity (Carrara et al., 2012; Fronhofer & Altermatt, 2017; Terui et al., 2018). Graph theoretical approaches have frequently been used to quantify network topology (Dale & Fortin, 2010; Minor & Urban, 2008; Newman, 2018). Of broad interest are metrics that characterise the importance of nodes in the network in terms of the flow of information, material and energy (O’Sullivan, 2014); however, a comprehensive analysis of node metric redundancy and variability across dendritic ecological networks (DENs) has not been conducted. This analysis is necessary to provide guidance as to which metrics are best suited to capturing variability in node characteristics in DENs.

Graph (network) theory provides an elegant framework for unifying and evaluating multiple aspects of network topology and can quantify either structural or functional connectivity (Minor & Urban, 2008). A graph or network is a set of nodes (or vertices) and edges (links), where nodes are the individual elements in the network and edges represent connections between nodes. Edges may be binary (connected or not) or contain additional information (e.g. weighted based on distance and can be directed or undirected). Likewise, nodes can be represented simply as points or can have additional attributes, such as size or carrying capacity (Newman, 2018). Graph theory has been used to describe many network types including transportation, social interactions and landscape structure; here, we focus strictly on spatial graphs.

Frequently used graph-based algorithms designed to simulate river networks fall into three broad categories (Figure 1). Stochastic branching networks (SBNs) depict rivers in the simplest terms (Labonne et al., 2008; Turcotte et al., 1998; Yeakel et al., 2014). This algorithm starts from an initial downstream node (the river mouth), one or two nodes are then iteratively added to a random node in the network with no upstream connections, the process is repeated until the desired number of nodes is reached (Labonne et al., 2008). SBNs have been widely used in models of metapopulation dynamics.
and ecosystem functioning (Anderson & Hayes, 2018; Fagan, 2002; Labonne et al., 2008; Terui et al., 2018). Recently, a second algorithm, optimal channel networks (OCNs), has become popular. OCNs mimic some of the geometric properties of real rivers, such as the distribution of upstream and downstream lengths and total contributing area to a point. The algorithm can also reproduce scaling features of real rivers such as Hack’s law (the relationship between the length of streams and the area of their basins) and the scaling of the probability distribution of drainage areas (Rigon et al., 1993; Rinaldo et al., 2014). Finally, networks of real rivers can be extracted from digital elevation models or satellite imagery and converted to adjacency matrices and graph objects. These approaches lie along the tractability-realism gradient; SBNs are easy to generate but are the least realistic (in terms of mimicking some of the geometric properties of ‘real’ rivers). OCNs are more realistic than SBNs but are computationally more expensive to generate (Figure S1). The SBN and OCN algorithms can be used to generate many networks with varying and fixed hyper-parameters (inputs supplied to the algorithm), which is desirable for modelling applications where replicate simulations with variable network topologies are required. Converting real networks to graph objects provides the most realistic representation, but it is not possible to produce a variety of networks with fixed input parameters (e.g. basin size, number of nodes). Real networks are typically used for place-based experiments and analyses, whereas SBNs and OCNs are more frequently used for more abstract simulation models and theoretical development (Carrara et al., 2012; Erös et al., 2012; Fields et al., 2017; Fronhofer & Altermatt, 2017; Segurado et al., 2013). While OCNs reproduce many of the characteristics of real rivers, how these networks, and SBNs, compare to real rivers regarding network topology metrics has not been evaluated.

A plethora of metrics quantify different aspects of network topology; these metrics have a long history of use in spatial ecology (Bunn et al., 2000; Cantwell & Forman, 1993; Dale & Fortin, 2010; Wagner & Fortin, 2005) and many have been co-opted from other disciplines (Urban & Keitt, 2001). Graph metrics typically focus on the node level (one value per node), the network level (one value per network) or somewhere in between, such as motifs (there are also many edge metrics, which are analogous to node metrics). Node-level metrics characterise each node in terms of their degree (the number of links the node has to other vertices), how often the node is expected to be traversed when moving between nodes, the number of shortest paths between nodes that cross through a given node, the distance from a node to other nodes, and can integrate additional information such as edge weights (Newman, 2018). A number of reviews have assessed the application of node importance metrics across a variety of network topologies (e.g. landscape and temporal networks and random graphs; Minor & Urban, 2008; Dale & Fortin, 2010; Bounova & De Weck, 2012; Nicosia et al., 2013), but not on dendritic ecological networks. In many non-DENs, such as social networks or terrestrial habitat networks, each node may be connected to many other nodes and hence there will be multiple pathways between any two nodes. By contrast in DENs there is typically just one pathway between any pair of nodes (assuming the network is constructed at the basin scale); this characteristic reduces the possible range of some metric scores and renders some topological metrics redundant. Because DENs differ in fundamental ways from the types of networks covered in previous reviews, there is a need to identify a set of metrics that enable concise characterisation of independent aspects of network topology and to help modellers determine the appropriate algorithm and parameterisation for a given problem. The primary aim of our study was to evaluate how a suite of frequently used network topology metrics vary across a gradient of river representations, ranging from networks produced via a simple stochastic branching algorithm to real rivers. Specifically, we aimed to:

1. Identify a set of orthogonal topology metrics that capture the most information about DENs produced by an SBN algorithm, OCNs and real rivers, and to determine whether the pool of metrics is consistent across the algorithms.
2. Analyse the relationship between algorithm hyper-parameters and node metrics to assess whether systematically altering hyper-parameters results in predictable shifts in node metric scores (as would be desirable for simulation applications).
3. Determine whether simulated networks (SBNs and OCNs) replicate the node metric scores of real rivers.
4. Evaluate how OCNs and SBNs compare at reproducing ecological patterns generated by a metapopulation model.

2 | MATERIALS AND METHODS

2.1 | Network generation

SBNs were generated using a stochastic branching process. The network generation process started from an initial downstream node (the river mouth). At each iteration, a random node in the network with no upstream connections was selected, and one or two nodes added upstream of it, depending on a branching probability (p). This process was repeated until some pre-determined number of nodes across the entire network was attained. Because the generation process was stochastic, two networks generated with the same branching probability could differ and, conversely, different branching probabilities could produce identical network structures (i.e. equifinality). The exception to this stochasticity is when p = 0.0 or 1.0, which always produce linear, non-branching and bifurcating networks, respectively.

OCNs are oriented spanning trees (built on rectangular lattices) that reproduce some of the scaling features characteristic of real river networks (Rinaldo et al., 2014; Rodriguez-Iiturbe et al., 2009). The algorithm starts with a feasible network configuration imposed on a lattice landscape, and the OCNs are obtained by minimisation of a function representing total energy dissipated by water flowing through the network spanning the lattice. Links in the initial network are sequentially rewired until
total energy expenditure across the network is minimised (Carraro et al., 2020). We simulated OCNs using the \texttt{R} package \texttt{OCNNet} version 0.3.2 (Carraro et al., 2020), where the network topology is determined by seven hyper-parameters. Hyper-parameters \(\text{dim}_a\) and \(\text{dim}_d\) determine the size (number of pixels) of the initial lattice, \(\text{thrA}\), the threshold value of the drainage area used to derive the aggregated network, \(\text{maxL}\), the maximum reach length allowed (in planar units). If the path length between a channel head and the downstream confluence is higher than \(\text{maxL}\), an additional node is inserted into the reach. The shape of the initial network configuration is defined by the parameter \text{state}, and \text{temperature} (a combination of two parameters, \text{initialNoCoolingPhase} and \text{coolingRate}) determines the temperature of the simulated annealing algorithm, which is used to derive the final network configuration. Finally, \text{type} determines the level at which pixels are aggregated together. Carraro et al. (2020) provide a detailed description of the implementation of the OCN algorithm.

Graphs of real river networks were generated by clipping a shapefile of global river centrelines with catchment polygons (https://www.hydrosheds.org) ranging in size from 300 to 6,000 km\(^2\) (given we did not sample the largest rivers, it is possible extrapolating the results of this study to larger systems will underestimate differences between simulated and real river networks). Each river network was converted to an adjacency matrix with a node placed at all reach terminals and confluences. In all, 150 randomly selected catchments were sampled (Figure S2). The initial 150 sampled rivers consisted of nodes with a degree of 1 or 3, whereas SBNs and OCNs can produce networks with a degree of 1, 2 and 3 (because these algorithms can discretise a river reach into a sequence of nodes). Thus, to make SBNs, OCNs and real river networks comparable, when we converted shapefiles of real rivers to graph networks, we split reaches and inserted additional nodes so that none of the obtained reaches were longer than some predefined length (\(\text{med}\)). We set \(\text{med}\) to 0.5, 1 and 1.5 multiplied by the median network reach length, this yielded a pool of 600 real river networks for analysis (including the 150 networks where no partitioning of reaches was performed). Finally, for all metric calculations, we used unweighted graphs so that we could compare metrics across the three algorithms; metric scores generated from weighted and unweighted networks were highly correlated (Figure S3).

### 2.2 Metric calculation

We calculated 14 structural topology metrics for every node in each network, (Table 1: Figure S4). Each of these metrics provides information about a given node's position in the network (e.g. centrality) or its importance based on the number of connections it has to other nodes (e.g. degree). The 14 metrics we evaluated cover those

| Metric                          | Description                                                                 | Reference          |
|--------------------------------|-----------------------------------------------------------------------------|--------------------|
| Eccentricity                   | The shortest path distance from the farthest node in the graph               | West (1996)        |
| Farness centrality             | Sum of the length of the shortest paths between the node and all other nodes| Altermatt (2013)   |
| Mean neighbour degree          | Average degree of neighbouring nodes                                        | Barrat et al. (2004)|
| Degree                         | Number of edges connected to a given node                                   | Newman (2010)      |
| Harary Centrality             | Reciprocal of the shortest path distance from the farthest other node in the graph | Hage and Harary (1995)|
| Closeness centrality           | Reciprocal of the sum of the length of the shortest paths between the node and all other nodes | Newman (2010)    |
| PageRank                       | Based on the number of connected nodes and the number of connections those neighbours have | Newman (2010)   |
| Katz centrality                | Measures the number of nodes that can be connected through a path, while the contributions of distant nodes are penalised | Newman (2010)   |
| Eigenvector centrality         | Eigenvector centrality measures a node’s importance (degree) while considering the importance (degree) of its neighbours | Newman (2010)   |
| Betweenness centrality         | Number of shortest paths that pass through the node                         | Newman (2010)      |
| Harmonic centrality            | Reverses the sum and reciprocal operations in the definition of closeness centrality | Marchiori and Latora (2000)|
| Neighbours within (graph diameter \( \times d^a \)) | Number of nodes within a given distance along the network distance | West (1996) |

\(^a d = 0.05, 0.1 \text{ and } 0.2.\)
widely used in the analysis of river networks, as well as some metrics not commonly used in the analysis of rivers (e.g. PageRank) but frequently used in other fields of network analysis. We did not consider metrics do not focus on node position in the network (e.g. fragmentation metrics). Additionally, we only considered node metrics on simple, unweighted, undirected networks; that is, we did not consider metrics that require additional information, such as species dispersal kernels.

2.3 | Statistical analysis

2.3.1 | Aim 1: Identify a set of orthogonal metrics

To identify a set of orthogonal metrics that capture the most information about the topology of DENs, we used principal components analysis (PCA). SBNs and OCNs were generated in a full factorial design across hyper-parameter combinations (Table 2). Because the network generation algorithms are stochastic, 25 replicate networks were produced per hyper-parameter combination yielding 5,500 and 145,800 SBNs and OCNs, respectively. Once networks were generated, node metrics were calculated and standardised by subtracting the mean and dividing by the standard deviation; this standardisation was calculated for each node x metric matrix prior to the PCA. Each algorithm (SBN, OCN and real rivers) was analysed independently and PCAs were conducted separately for individual networks (i.e. for the SBNs 5500, PCAs were computed). PCAs for each algorithm were aligned using Procrustes rotation (Peres-Neto & Jackson, 2001) and loadings used to identify the metrics contributing most to each of the principal components; only principal components that explained more than 10% of variation in node metrics were retained. In the subsequent aims, we focus on the node metrics eccentricity (the maximum distance between a node to all other nodes) and mean neighbour degree because they had high loadings on principal components one and two and are intuitively understood. Additionally, we illustrate that the explanatory power of these two metrics depends on the processes controlling ecological patterns, and hence the utility of these distinct metrics (Figure S5).

2.3.2 | Aim 2: Relationship between algorithm hyper-parameters and node metrics

To understand how the hyper-parameterisation used to generate SBNs and OCNs relate to the resulting node metrics, the metrics generated in aim 1 were visualised using non-Metric Multidimensional Scaling (nMDS) and hyper-parameters significantly related to the ordinations identified via vector fitting [envfit in R package vegaN (v 2.5.6), a < 0.05, 999 permutations]. Prior to ordination, node metrics were averaged across each network to give a single metric value per network and distance matrices were calculated using Gower distances [the metrics were a mix of count and continuous data (Legendre & Legendre, 2012)].

2.3.3 | Aim 3: Similarity of simulated networks and real rivers

We compared node topology metrics for real rivers to those from artificial networks (SBNs and OCNs) to evaluate whether the artificial networks can adequately replicate metric patterns of real rivers. Given the computation expense of generating many large OCN networks, we restricted our analysis to networks up to 1,000 nodes in size. For real rivers, we used the networks used to address aim 1, resulting in 541 networks after dropping those larger than 1,000 nodes. SBNs were generated in a full factorial design across hyper-parameter combinations (Table 3). To reduce computation time while generating OCNs, we held the hyper-parameters state and temperature constant as they had little effect on metrics scores (Figures S6 and S7). For the hyper-parameter thrA, the number of possible settings increases with network size, so we used the OCNet R package function find_area_threshold_OCN() to identify the thrA values that would result in networks between 10 and 1,000 nodes in size and sampled 10 of these thrA values at random without replacement. We generated 10 replicate networks per hyper-parameter combination, which resulted in 100 values of thrA for each hyper-parameter combination. Since the range in possible values of thrA is a function of \( \dim_x \) and \( \dim_y \), thrA was standardised \( \left( \text{thrA} / (\dim_x \times \dim_y) \right) \) prior to visualisation. After generating the OCNs, any network with a node with a degree greater than 3 was excluded from further analysis. For each node in each network, eccentricity and mean neighbour degree were calculated and averaged at the network level. Metric scores of OCNs and SBNs were then visually compared to those of real rivers.

2.3.4 | Aim 4: Metapopulation abundance related to node metrics

Finally, we show how the node metric scores of networks can be related to ecological patterns. We used a metapopulation model to examine how patterns of abundance compare across real, OCN and SBN networks, when the networks have similar and variable node eccentricity and mean neighbour degree distributions.

The metapopulation consisted of a single species occupying a network of 55 nodes, with a carrying capacity (K) of 1,000 individuals per patch. We explored metapopulation dynamics under two different scenarios: (a) where we selected a real river network, OCN and SBN with similar eccentricity distributions and (b) where the eccentricity distributions of the OCN and SBN were similar to each other, but different from that of the real river network. We assumed that local population growth followed a discrete-time stochastic logistic model in each patch:

\[
N_{i+1} = \text{Pois} \left( N_i + r N_i \left( 1 - \frac{N_i}{K} \right) \right)
\]  

(1)
where $N_{i,t}$ population size in river $i$ at time $t$, $r$ is the intrinsic rate of population growth and $K$ the carrying capacity of patch $i$. Population growth was followed by dispersal of all individuals between patches, where individuals can move to any other network node, and the dispersal kernel (Equation 2) expresses the fraction of individuals moving from patch $i$ to any other patch $j$.

$$\frac{1}{\delta} e^{-\frac{d_{ij}}{2}}$$

where $d_{ij}$ is the along the network distance (i.e. number of nodes, links were not weighted) between patches $i$ and $j$ and $\delta$ determines the shape of the dispersal kernel.

To address aim 4, a real river network with 55 nodes was sampled from the 600 networks generated to address aim 1, two OCNs and two SBNs were then generated, where one of each was similar to the real network regarding the distribution of node eccentricity scores, and one of each were dissimilar (see Figure S8 for an analysis using mean neighbour degree). Hyper-parameters used to

### TABLE 2 Hyper-parameter values used to generate SBNs and OCNs that were subsequently used to analyse node metrics. In all, 25 replicate networks were generated for each hyper-parameter combination of SBNs and OCNs

| Network | Hyper-parameter | Hyper-parameter explanation | Values | Total networks |
|---------|-----------------|----------------------------|-----|---------------|
| SBN     | $p$             | Probability of branching   | 0–1.0 by 0.1 | 5,500         |
|         | $n_{node}$      | Number of nodes            | 50–1000 by 50 |               |
| OCN     | $dim_x$         | Width of lattice (pixels)  | 30  | 145,800       |
|         | $dim_y$         | Height of lattice (pixels) | $(0.4–2.0$ by 0.2 $) \times dim_x$ |               |
|         | $thrA$          | Threshold used to determine drainage area | 1, 2, 3, 4, 5, 10, 20, 40, 80 |               |
|         | $maxL$          | Determines maximum reach lengths | Inf, 2, 4 |               |
|         | $state$         | Configuration of the initial state of the network | V, I, T, H |               |
|         | $temperature^a$ | Used to describe the temperature of the simulated annealing algorithm | C, W, H |               |
|         | $type$          | Level at which pixels are aggregated | RN, AG |               |
| Real rivers | $Catchsize$   | Range of catchment sizes sampled from the HydroRivers database | 300–6000 km² | 600 |
|         | $Med_n$         | Value used to determine when to split reaches and insert new nodes, $Med_n = 1$: any reach longer than the network median is split. | 0.5, 1, 1.5, none |               |

$^a$temperature is a function of two sub-parameters ($initialNoCoolingPhase$ and $coolingRate$): C: $initialNoCoolingPhase = 0$, $coolingRate = 10$, W: $initialNoCoolingPhase = 0$, $coolingRate = 1$, H: $initialNoCoolingPhase = 0.1$, $coolingRate = 0.5$. See Carraro et al. (2020) for details.

### TABLE 3 Hyper-parameter values used to generate the SBNs and OCNs that were compared to the 541 real river networks. In all, 10 replicate networks were generated for each hyper-parameter combination. Hyper-parameters state and temperature had relatively little effect on eccentricity and mean neighbour degree so we fixed these values. We have analysed the effect of changing these two parameters in the Supplementary Material (Figures S6 and S7)

| Network | Hyper-parameter | Hyper-parameter explanation | Values | Total networks |
|---------|-----------------|----------------------------|-----|---------------|
| SBN     | $p$             | Probability of branching   | 0–1.0 by 0.1 | 2,200         |
|         | $n_{node}$      | Number of nodes            | 50–1000 by 50 |               |
| OCN     | $dim_x$         | Width of lattice (pixels)  | 20, 40, 60, 80, 100 | 15,000       |
|         | $dim_y$         | Height of lattice (pixels) | 20, 40, 60, 80, 100 |               |
|         | $thrA$          | Threshold used to determine drainage area | 10 values randomly sampled |               |
|         | $maxL$          | Determines maximum reach lengths | Inf, 2, 4 |               |
|         | $state$         | Configuration of the initial state of the network | V |               |
|         | $temperature^a$ | Used to describe the temperature of the simulated annealing algorithm | W |               |
|         | $type$          | Level at which pixels are aggregated | RN, AG |               |

$^a$temperature is a function of two sub-parameters ($initialNoCoolingPhase$ and $coolingRate$): W: $initialNoCoolingPhase = 0$, $coolingRate = 1$. See Carraro et al. (2020) for details.
generate the SBM were \( n_{\text{nodes}} = 55 \) and \( p = 0.23 \) and \( p = 0.9 \) for the similar and not-similar networks, respectively, and hyper-parameters to generate the OCNs were \( \text{dim}_x = \text{dim}_y = 20 \), \( \text{state} = V, \text{temp} = W \), \( \text{thrA} = 3 \), \( \text{maxL} = \text{Inf} \) and \( \text{type} = \text{AG} \) for the similar network, for the not similar network the same settings were used, except \( \text{type} = \text{RN} \) and \( \text{thrA} = 13 \).

At the start of each simulation, all 55 patches were populated at 100% of carrying capacity and \( r \) was set to 0.2. After population growth dispersal occurred, with the proportion of individuals dispersing from patch \( i \) to any patch \( j \) determined by the dispersal kernel (where \( \delta = 0.3 \); Equation 2). We ran 10 replicate simulations for each dispersal scenario, with each simulation lasting 100 generations (the model converged after 10–20 generations). At each timestep, we calculated the coefficient of variation of the proportion of \( K \) occupied (as a measure of spatial variability in abundance) and at the end of each simulation, patch abundance was recorded and averaged across the last 10 timesteps. Rank abundance curves and the coefficient of variation of the proportion of \( K \) occupied were compared across real rivers, OCNs and SBNs.

All network production and analyses were conducted using R v 4.0.4 (R Core Team, 2020). SBNs were produced using a custom script and the \textsc{sbn} package (Lee, 2022) and OCNs were produced using the \textsc{OCNet} v 0.3.2 (Carraro et al., 2020) package. Network metrics were calculated using the \textsc{igraph} v 1.2.4.2 (Csardi & Nepusz, 2006), \textsc{centiserve} v 1.0.0 (Jalili, 2017), \textsc{brainGraph} v 2.7.3 (Watson, 2019), \textsc{CINNA} v 1.1.53 (Ashtiani, 2019) and \textsc{sna} v 2.5 (Butts, 2019) packages. The \textsc{vegan} v 2.5.6 (Oksanen et al., 2010) package was used for data analysis. Finally, figures were produced using \textsc{ggplot2} v 2.3.3.0 (Wickham, 2016) and \textsc{patchwork} v 1.0.0 (Pedersen, 2021) packages.

3 | RESULTS

3.1 | Aim 1: Identify a set of orthogonal metrics

The variation in node topology explained by principal components analysis was relatively consistent across SBNs, OCNs and real rivers, as were the loadings (Figure 2). For all network types, PC one was dominated by eccentricity and farness centrality, while PC two was dominated by mean neighbour degree, degree, Harary and closeness centrality and PageRank, and Katz centrality. Eccentricity and farness centrality are distance-based metrics calculated using the distance from the target node to (a) the farthest node (eccentricity) or (b) all other nodes (farness centrality). The majority of metrics dominating PC two are associated with node degree and, therefore, number of local connections rather than centrality.

3.2 | Aim 2: Relationship between algorithm hyper-parameters and node metrics

For SBN networks, ordination of metric scores revealed clear structuring for the hyper-parameters \( n_{\text{nodes}} \) (total number of nodes) and \( p \) (branching probability). As \( n_{\text{nodes}} \) increased eccentricity increased, indicating distances between nodes were increasing [although relative eccentricity decreased as the number of nodes in the network increased (Figure S9)]. As \( p \) increased, mean neighbour degree increased, indicating that on average networks had a high number of links (Figure 3). For OCN networks, there was a clear separation in ordination space for the hyper-parameter type, which determines

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2}
\caption{Variance explained (%) by the first two principal components for SBN, OCN and real river networks (a). Factor loadings (absolute) for the three network algorithms (b). SBN, stochastic branching network; OCN, optimal channel network; Real, networks generated from shapefiles of rivers.}
\end{figure}
the scale at which pixels are aggregated, where networks were simulated at the river network level (type = RN) networks had lower mean neighbor degree values, suggesting nodes had fewer links compared to networks simulated at the aggregated level (type = AG). Metric scores for the hyper-parameter maxL exhibited clustering, but this was not consistent with the gradients associated with either mean neighbor degree or eccentricity. It is likely for maxL and the other hyper-parameters that interactions between hyper-parameters obscure the patterns (Figure 4).

3.3 | Aim 3: Similarity of simulated networks and real rivers

SBNs were able to reproduce eccentricity scores of real rivers, regardless of the real river discretisation method (medn) used, or the size of the network. The ability of SBNs to reproduce eccentricity scores of real rivers depended on the hyper-parameter p, with medium values (p = 0.5) replicating eccentricity scores of non-discretised real rivers and the required value of p decreasing as medn got smaller (Figure 5). Values of p below 0.1 and greater than 0.6 resulted in eccentricity scores substantially departing from those of the real rivers (Figure 5: Figure S10). However, the eccentricity values derived from real river networks were obtained for arbitrarily selected values of medn, and further increasing or decreasing medn will change these results. For mean neighbor degree, the SBNs could reproduce metric scores of real rivers when p was greater than 0.1 (Figure 5).

OCNs could also reproduce the eccentricity scores of real rivers, but the relationship was more complicated given the large number (7) of hyper-parameters required to generate OCNs. The hyper-parameter type was the main determinant of the similarity between OCNs and real networks in terms of their eccentricity, with type = RN resulting in networks similar to real networks with discretisation method medn = 0.5 (Figures S11–S16), while OCNs generated with type = AG were similar to real rivers either with no discretisation or discretisation method medn = 1.5 (Figure 6: Figures S11–S16), although this relationship was dependent on maxL and thrA. Hyper-parameter thrA was next most important; networks generated with high thrA values were typically similar to real rivers with discretisation method of medn = 1.5 or none, as thrA decreased eccentricity scores became more similar to networks with medn = 0.5, when standardised thrA was less than 0.1 OCNs were not similar to real rivers, suggesting low values of thrA.
are better at representing real rivers which have been discretised with lower values of \( \text{med}_n \) than we evaluated here. Finally, \( \text{max}_L \) had a strong effect on the number of nodes in the network which affected metric scores, with higher values of \( \text{max}_L \) increasing the similarity between OCNs and the real rivers analysed (Figure 6). For both eccentricity and mean neighbour degree, OCNs replicated a proportion of the possible eccentricity scores of real rivers, whereas SBNs could generally replicate any of the mean neighbour degree metric scores of real rivers (Figures S7, S8, S10–S16).

3.4 | Aim 4: Metapopulation abundance related to node eccentricity

From the metapopulation model, both SBNs and OCNs were able to reproduce the ecological patterns observed in simulations using real rivers when the eccentricity distributions matched across the three networks. Similarly, when the distribution of node eccentricity scores of SBNs and OCNs differed from real rivers, the patterns of abundance and the coefficient of variation in the proportion of \( K \) occupied departed from those of real rivers (Figure 7). In this metapopulation, changes in abundance and occupancy are driven by the interaction between network topology (summarised by eccentricity) and dispersal. When comparing two networks with identical numbers of nodes, the network with lower average eccentricity will have more nodes with a degree of 1, that is more headwater nodes. Given headwater nodes only have one direction of potential immigrants to arrive from, there is a higher proportion of patches with lower (Figure 7e, SBN and OCN) and more variable (Figure 7f, SBN and OCN) abundances in these networks (Altermatt & Fronhofer, 2018).
4 | DISCUSSION

Ecological communities inhabiting river networks face significant threats from changing climates and anthropogenic activities (Herrera-R et al., 2020; Radinger & Garcia-Berthoux, 2020). These stressors commonly alter network topology and connectivity via changing flow regimes and channel modification. Therefore, understanding the role of the topological structure of river networks in shaping ecological processes is crucial (Erős et al., 2018; Erős & Lowe, 2019). Here we presented an analysis of dendritic ecological network generating algorithms and the graph-based metrics used to quantify topology of such networks. We identify a useful set of metrics for describing DEN node topology, show how topology relates to algorithm hyper-parameters and demonstrate that both OCNs and SBNs can replicate the topological metric scores of real rivers in certain circumstances.

Different metrics characterise different components of network topology. We recommend eccentricity and/or farness metrics be used to quantify node centrality and either mean neighbour degree or degree be used to quantify local node importance in DENs. Given the increasing use of graph-theoretic approaches to analyse river networks, it is important to identify how commonly used metrics respond to a variety of network topologies. Because of the constrained nature of rivers' topology, many metrics show little variability across nodes. This reduced variability is driven by (i) the low values and range in degree of nodes in DENs, which range from 1 to 3 and (ii) there only ever being one unique pathway between any pair of nodes in DENs (exceptions to this assumption are braided rivers and for species capable of overland dispersal, not considered here), which is important because many metrics are fundamentally derived from either the number of times a node is traversed or inter-node distances (Newman et al., 2001). Many metrics are calculated using the same limited set of network characteristics (degree, distance, path traversals); thus, there is considerable redundancy among metrics. This redundancy is exacerbated by the topologically constrained nature of DENs. To reduce this redundancy, we identified orthogonal sets of metrics. Our analysis suggests metrics can be split into two broad groups: (1) those that describe a node’s position in the network (node centrality) and (2) those that describe the number of local connections (node importance). Eccentricity and farness were strongly associated with PC 1, which explained a large (>50%) amount of variation in node variability; these metrics are well suited to identifying nodes most vulnerable to fragmentation and keystone patches that link the wider network (Sarker et al., 2019).

Mean neighbour degree and degree were two of the preferred node importance metrics; these metrics are determined by the number of connections a node and its neighbours have and describe how locally connected a node is rather than the node’s overall position in the network (à la centrality metrics). Importance metrics are useful for identifying locally well-connected patches and sites for invasive species control or targets for restoration (Drake et al., 2017; Ferrari et al., 2014; Perry et al., 2017).

When interrogating the influence of topology on ecological processes and their emergent patterns, it is useful to be able to generate a gradient of network structures spanning the breadth of possible network configurations. Ideally, a clear relationship would exist between algorithm hyper-parameters and the resulting network metrics. For both SBNs and OCNs, network size was most associated with consistent shifts in metric scores; relationships with other hyper-parameters were mixed. Aside from network size, for SBNs the hyper-parameter branching probability (p) had a predictable effect on node metrics. In particular, increasing p resulted in increasing mean node neighbour degree, suggesting networks become more ‘branchy’. Other than the size of the lattice used, when generating OCNs aggregation level (type) and thrA were the hyper-parameters with the most predictable effect on node metric scores; pixel aggregation at the river network level (type =RN) resulted in nodes with lower mean neighbour degree scores (i.e. nodes having fewer local connections), compared to type =AG. The lack of clear association between other OCN hyper-parameters and node metric scores makes generating networks with predictable node metric scores difficult using OCN algorithms. There is no one-size-fits-all...
solution as to which algorithm to use to simulate dendritic networks. If the goal is to create networks that perfectly represent rivers, obtaining topological descriptions of real rivers may be the best approach; however, this approach is relatively time-consuming, can only produce a limited number of networks with specific characteristics (e.g. number of nodes) and when conducted across large geographical extents computational processing becomes a challenge. Selecting between SBNs or OCNs will depend on whether the primary goal is to:

1. Represent the branching structure of rivers as summarised in graph theoretic metrics. In this case, both SBNs and OCNs could replicate eccentricity and mean neighbour degree of real rivers some of the time, regardless of the discretisation method used to generate real river networks, therefore either algorithm can be used and the decision of which to use should be made based on other requirements.

2. Replicate drainage properties of real rivers. Although not analysed here, OCNs can mimic multiple aspects of real rivers including scaling features such as Hack’s law and the scaling of the probability distribution of drainage areas (Rigon et al., 1993; Rinaldo et al., 2014). As an example, patch size is often of interest. Using an OCN algorithm, this can be estimated using the river reach width as a proxy. Patch size would be more difficult to accurately model using a network generated via an SBN algorithm as there is no inherent estimation of basin size or reach length. Therefore, if representing scaling features of rivers is important, OCNs should be used.

![Figure 6](https://example.com/figure6.png)

**Figure 6** Relationship between network size (number of nodes), hyper parameter $max_L$ and graph metrics eccentricity and mean neighbour degree, comparing OCNs (coloured points) to real rivers (grey shapes represent the four discretisation methods used to convert shapefiles of real rivers). OCN hyper-parameters kept fixed: type = AG, state = V, temperature = W. dimx = dimy = 100. See Figures S6 and S7 for an analysis of fixed hyper-parameters.
3. Generate networks that vary predictably with hyper-parameter values. In this case, we recommend SBNs as they have only two hyper-parameters (nodes and p) and node-level metrics varied reasonably predictably in response to them. For the OCNs, the metrics varied somewhat predictably with type, but not with the remaining hyper-parameters. The larger number of hyper-parameters required by OCNs and the non-predictable response of metrics to five of the hyper-parameters means the relationship between hyper-parameters and node metrics is less transparent for OCNs than for SBNs.

Finally, SBNs are much less computationally expensive than OCNs [2–3 orders of magnitude quicker (Figure S1)]; where many networks and/or large networks need to be generated, computational time may be an issue making SBNs the only feasible option.

Centrality and node importance metrics measure different aspects of network topology, and they can be used in conjunction to evaluate different hypotheses. This use of different metrics was highlighted by our demonstration of how local abundance can be predicted by node metrics in a metapopulation model (Figure S3). In our analysis, a local node importance metric (degree) better predicted abundance when dispersal was local; however, when dispersal was wide ranging, eccentricity was a better predictor. This analysis highlights how the (spatial) extent of the process of interest can be matched by topology metrics that operate at similar scales.

We recommend that authors explicitly state whether they are interested in node centrality or node importance (or both) and use the metrics highlighted above.

5 | CONCLUSIONS

Riverine ecosystems are spatially structured habitats in which network topology shapes many biotic and abiotic processes. The hierarchical and branching structure of river networks can help to explain patterns of diversity, metapopulation stability and processes affecting productivity in riverine ecosystems (Helton et al., 2018; Mari et al., 2014; Muneepeerakul et al., 2019). Graph theory provides a convenient way to represent river networks and a host of easy-to-calculate metrics for quantifying the relative importance of nodes based on their position in the network. Because of the topologically constrained nature of river networks, there are fewer suitable metrics available than for other network types (e.g. terrestrial habitat networks, social networks). Eccentricity and farness centrality are recommended for capturing node centrality, and mean neighbour degree or degree are recommended for identifying well-connected nodes. In terms of generating graph-based representations of rivers (e.g. for simulation exercises), all three methods we evaluated have their limitations, but OCNs represent a balance between ease of generation and realism. Many questions remain about how spatial node characteristics relate...
to ecological patterns and processes; identifying a suite of topological metrics that adequately describe these systems will help to answer these questions. Finally, the metapopulation modelling results presented here provide some concrete predictions about how ecological patterns may relate to network topology metrics. Empirically testing these hypotheses should be the focus of future research.

ACKNOWLEDGEMENTS
The authors wish to thank two anonymous reviewers for providing detailed, constructive feedback that has greatly improved this manuscript. F.L. was supported by a University of Auckland Doctoral Scholarship and a Faculty of Science PhD Output Award. The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high-performance computing facilities as part of this research, provided by NeSI and funded jointly by NeSI’s collaborator institutions and through the Ministry of Business, Innovation & Employment’s Research Infrastructure programme (https://www.nesi.org.nz). Open access publishing facilitated by The University of Auckland, as part of the Wiley · The University of Auckland agreement via the Council of Australian University Librarians. [Correction added on 16 May 2022, after first online publication: CAUL funding statement has been added.]

CONFLICT OF INTEREST
The authors have no conflict of interest to declare.

AUTHORS’ CONTRIBUTIONS
All authors conceived the ideas and designed the methodology; F.L. compiled the data, developed the code, performed the data analysis and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13854.

DATA AVAILABILITY STATEMENT
The code and associated data are archived on Zenodo (Lee et al., 2022).

ORCID
Finnbar Lee https://orcid.org/0000-0002-9219-1486
Kevin S. Simon https://orcid.org/0000-0002-9537-2450
George L. W. Perry https://orcid.org/0000-0001-9672-9135

REFERENCES
Abed-Elmdoust, A., Singh, A., & Yang, Z.-L. (2017). Emergent spectral properties of river network topology: An optimal channel network approach. Scientific Reports, 7(1), 1–9. https://doi.org/10.1038/s41598-017-11579-1
Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 1–52. https://doi.org/10.1103/RevModPhys.74.47
Altermatt, F. (2013). Diversity in riverine metacommunities: A network perspective. Aquatic Ecology, 47(3), 365–377. https://doi.org/10.1007/s10452-013-9450-3
Altermatt, F., & Frohner, E. A. (2018). Dispersal in dendritic networks: Ecological consequences on the spatial distribution of population densities. Freshwater Biology, 63(1), 22–32. https://doi.org/10.1111/fwb.12951
Anderson, K. E., & Hayes, S. M. (2018). The effects of dispersal and river spatial structure on asynchrony in consumer-resource meta-communities. Freshwater Biology, 63(1), 100–113. https://doi.org/10.1111/fwb.12998
Ashitian, M. (2019). CINNA: deciphering central informative nodes in network analysis. R package version 1.1.53. Retrieved from https://CRAN.R-project.org/package=CINNA
Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11), 3747–3752. https://doi.org/10.1073/pnas.0400087101
Bounova, G., & De Weck, O. (2012). Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. Physical Review E, 85(1), 016117. https://doi.org/10.1103/PhysRevE.85.016117
Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: A conservation application of graph theory. Journal of Environmental Management, 59(4), 265–278. https://doi.org/10.1006/jema.2000.0373
Butts, C. T. (2019). sna: Tools for social network analysis. R package version 2.5. Retrieved from https://CRAN.R-project.org/package=sna
Cantwell, M. D., & Forman, R. T. (1993). Landscape graphs: Ecological modeling with graph theory to detect configurations common to diverse landscapes. Landscape Ecology, 8(4), 239–255. https://doi.org/10.1007/BF00125131
Carrara, F., Altermatt, F., Rodriguez-Iiturbe, I., & Rinaldo, A. (2012). Dendritic connectivity controls biodiversity patterns in experimental metacommunities. Proceedings of the National Academy of Sciences of the United States of America, 109(15), 5761–5766. https://doi.org/10.1073/pnas.1119651109
Carraro, L., Bertuzzo, E., Frohner, E. A., Furrer, R., Gounand, I., Rinaldo, A., & Altermatt, F. (2020). Generation and application of river network analogues for use in ecology and evolution. Ecology and Evolution, 10, 537–7550. https://doi.org/10.1002/ece3.6479
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. Inter Journal: Complex Systems, 1695(5), 1–9.
Dale, M. R. T., & Fortin, M.-J. (2010). From graphs to spatial graphs. Annual Review of Ecology, Evolution, and Systematics, 41, 21–38. https://doi.org/10.1146/annurev-ecolsys-102209-144718
Drake, J. C., Griffis-Kyle, K. L., & McIntyre, N. E. (2017). Graph theory as an invasive species management tool: Case study in the Sonoran Desert. Landscape Ecology, 32(8), 1739–1752. https://doi.org/10.1007/s10980-017-0539-2
Erős, T., & Lowe, W. H. (2019). The landscape ecology of rivers: From patch-based to spatial network analyses. Current Landscape Ecology Reports, 4(4), 103–112. https://doi.org/10.1007/s40823-019-00044-6
Erős, T., O’Hanley, J. R., & Czeglédi, I. (2018). A unified model for optimizing riverscape conservation. Journal of Applied Ecology, 55(4), 1871–1883. https://doi.org/10.1111/1365-2664.13142
Erős, T., Olden, J. D., Schick, R. S., Schmera, D., & Fortin, M.-J. (2012). Characterizing connectivity relationships in freshwaters using patch-based graphs. Landscape Ecology, 27(2), 303–317. https://doi.org/10.1007/s10980-011-9659-2
Fagan, W. F. (2002). Connectivity, fragmentation, and extinction risk in dendritic metapopulations. Ecology, 83(12), 3243–3249. https://doi.org/10.1890/0012-9658(2002)083[3243:FCFAR2.0.CO;2
Ferrari, J. R., Preisser, E. L., & Fitzpatrick, M. C. (2014). Modeling the spread of invasive species using dynamic network models. Biological Invasions, 16(4), 949–960. https://doi.org/10.1007/s10530-013-0552-6
Fields, W. R., Grant, E. H. C., & Lowe, W. H. (2017). Detecting spatial ontogenetic niche shifts in complex dendritic ecological
