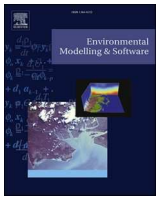
Environmental Modelling and Software 167 (2023) 105775

Contents lists available at ScienceDirect

Environmental Modelling and Software

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



Non-perennial stream networks as directed acyclic graphs: The R-package streamDAG

Ken Aho [a,*], Cathy Kriloff [b], Sarah E. Godsey [c], Rob Ramos [d], Chris Wheeler [e], Yaqi You [f], Sara Warix [c], DeWayne Derryberry [b], Sam Zipper [e], Rebecca L. Hale [g], Charles T. Bond [h], Kevin A. Kuehn [h]

[a] Department of Biological Sciences, Idaho State University, Pocatello, ID, 83209-8007, USA
[b] Department of Mathematics and Statistics, Idaho State University, Pocatello, ID, 83209-8085, USA
[c] Department of Geosciences, Idaho State University, Pocatello, ID, 83209-8072, USA
[d] Kansas Biological Survey and Center for Ecological Research, University of Kansas, Lawrence, KS, 66047, USA
[e] Kansas Geological Survey, The University of Kansas, Lawrence, KS, 66047-3724, USA
[f] Department of Environmental Resources Engineering, The State University of New York, College of Environmental Science and Forestry, Syracuse, NY, 13210, USA
[g] Smithsonian Environmental Research Center, Edgewater, MD, 21037, USA
[h] School of Biological, Environmental, and Earth Sciences, The University of Southern Mississippi, Hattiesburg, MS, 39406, USA

ARTICLE INFO

Handling Editor: Daniel P Ames

Keywords:
Non-perennial stream
Graph theory
Directed acyclic graph
Network analysis
Hydrologic connectivity
R computational software

ABSTRACT

Many conventional stream network metrics are poorly suited to non-perennial streams, which can vary substantially in space and time. To address this issue, we considered non-perennial stream networks as directed acyclic graphs (DAGs). DAG metrics allow: 1) summarization of important non-perennial stream characteristics (e.g., complexity, connectedness, and nestedness) from both local (individual segment) and global stream network perspectives, and 2) tracking of these features as networks expand and contract. We review a large number of graph theoretic metrics, and introduce a new **R** package, *streamDAG* that codifies approaches we feel are most useful. The *streamDAG* package contains procedures for handling water presence data, and functions for both local and global analyses of both unweighted and weighted stream DAGs. We demonstrate *streamDAG* using two North American non-perennial streams: Murphy Creek, a simple drainage system in the Owyhee Mountains of southwestern Idaho, and Konza Prairie, a relatively complex network in central Kansas.

Software/data Availability

Software name: *streamDAG*.
Developer: Ken Aho.
Contact information: ahoken@isu.edu.
Year first available: 2022.
Hardware required: **R**-amenable frameworks, e.g., PC, tablet, laptop;
System Software: Windows, Linux/Unix, MacOS.
Required application software: **R** https://www.r-project.org/(open source).
Program language: **R**.
Program size: 54 kb.
Repository: https://github.com/moondog1969/streamDAG (open source).

1. Introduction

Non-perennial streams currently encompass more than half of the global river network (Messager et al., 2021), and are receiving increased attention from researchers and resource managers due to their increasing spatial and temporal prevalence (Zipper et al., 2021), and their strong effects on water quantity and quality (Datry et al., 2014). By definition, non-perennial stream networks will vary in their spatial extent, complexity, and hydrologic connectedness over time. Thus, metrics for describing non-perennial streams must be amenable to spatiotemporal dynamics while providing consistent summaries of networks and network components. These efforts, however, are hampered by the lack of a consensus concerning the meaning of important descriptive terms, including hydrological connectivity (Freeman et al., 2007; Ali and Roy, 2009; Bracken et al., 2013), and a general research

* Corresponding author.
E-mail address: ahoken@isu.edu (K. Aho).

https://doi.org/10.1016/j.envsoft.2023.105775
Received 7 February 2023; Received in revised form 25 June 2023; Accepted 3 July 2023
Available online 5 July 2023
1364-8152/

and monitoring focus on perennial over non-perennial streams (Krabbenhoft et al., 2022). For example, many existing measures of hydrological connectivity are spatially explicit but time invariant because of a reliance on topography, slope, and drainage area. Examples include lumped parameter basin models (Beven and Kirkby, 1979), the field index of connectivity (Borselli et al., 2008), Hillslope-Riparian Stream connectivity (HRS, Jencso et al., 2009), the network index (Lane et al., 2009), and the Topographic Wetness Index (TWI, Sørensen et al., 2006). Several other common approaches, including Integral Connectivity Scale Length (ICSL, Western et al., 2001) and its variants (e.g., subsurface and outlet ICSL, Ali and Roy, 2010) and autocorrelation-based summaries (Knudby and Carrera, 2005; Ali and Roy, 2010), allow tracking of stream network connectivity over time, but do not quantify the relative importance of particular stream locations to whole-network functionality.

As a potential alternative, one can consider stream networks from the perspective of graph theory. This approach appears particularly useful for representing non-perennial streams because it provides straightforward, standard graphical and numerical tools for the tracking of a non-perennial stream network as its sections dry and potentially disappear. That is, graph theoretic representations of non-perennial stream networks can vary over time. Additionally, graph theoretic methods allow a broad array of *global* network level summaries as well as consideration of the importance of *local* individual stream locations to the functioning of the overall hydrologic networks.

Graph theory has been previously proposed as an important tool for describing general geological processes (Phillips et al., 2015), quantifying human impacts on the connectivity of marine and freshwater systems (Saunders et al., 2015), and studying the spatiotemporal connectivity of habitat mosaics in landscape ecology (Urban and Keitt, 2001), amphibian pond networks (Fortuna et al., 2006), and ephemeral wetlands (Bertassello et al., 2021). Several recent papers have applied graph theory perspectives specifically to stream networks. These include the use of graph betweenness centrality to identify "critical" stream network nodes (Sarker et al., 2019), the modelling of stream flow fluctuations using directed visibility graphs in time series analyses (Serinaldi and Kilsby, 2016), consideration of subsurface hydrologic connectivity using a graph-theoretic framework (Zuecco et al., 2019), conflating graph-theoretic and percolation theory perspectives to measure connectivity (Larsen et al., 2012), physics-guided graph models of stream connectivity (Jia et al., 2021), the use of nested subgraphs for measuring aquatic organism dispersal among stream reaches (Baldan et al., 2022), probabilistic models for organismal connectivity based on undirected graphs (Garbin et al., 2019), directed acyclic graph streamflow models with neural networks (Liu et al., 2022), and flow persistence models in non-perennial stream networks (Botter and Durighetto, 2020).

As a response to this growing interest, we present a formal consideration of non-perennial stream systems in the context of graph theory as a guide for future researchers, and introduce a new software package for the graph-theoretic analysis of non-perennial streams. We accomplish this over seven subsequent sections (Sections 2-8). In Section 2 we define important graph theoretic terms and demonstrate the appropriateness of directed acyclic graph representations of non-perennial stream networks. In Section 3 we introduce the *streamDAG* **R** package, and compare it to existing software. In Sections 4 and 5 we review a large number of graph theoretic approaches for unweighted and weighted directed acyclic graphs, respectively. Importantly, sections 4 and 5 contain tables that identify generally applicable metrics, recognize methods that may be particularly useful, and provide example code for *streamDAG* functions to obtain metric results. In Section 6 we use *streamDAG* functions to describe and compare two non-perennial streams with putatively distinct network characteristics. In Sections 7 and 8, we briefly discuss and summarize our work.

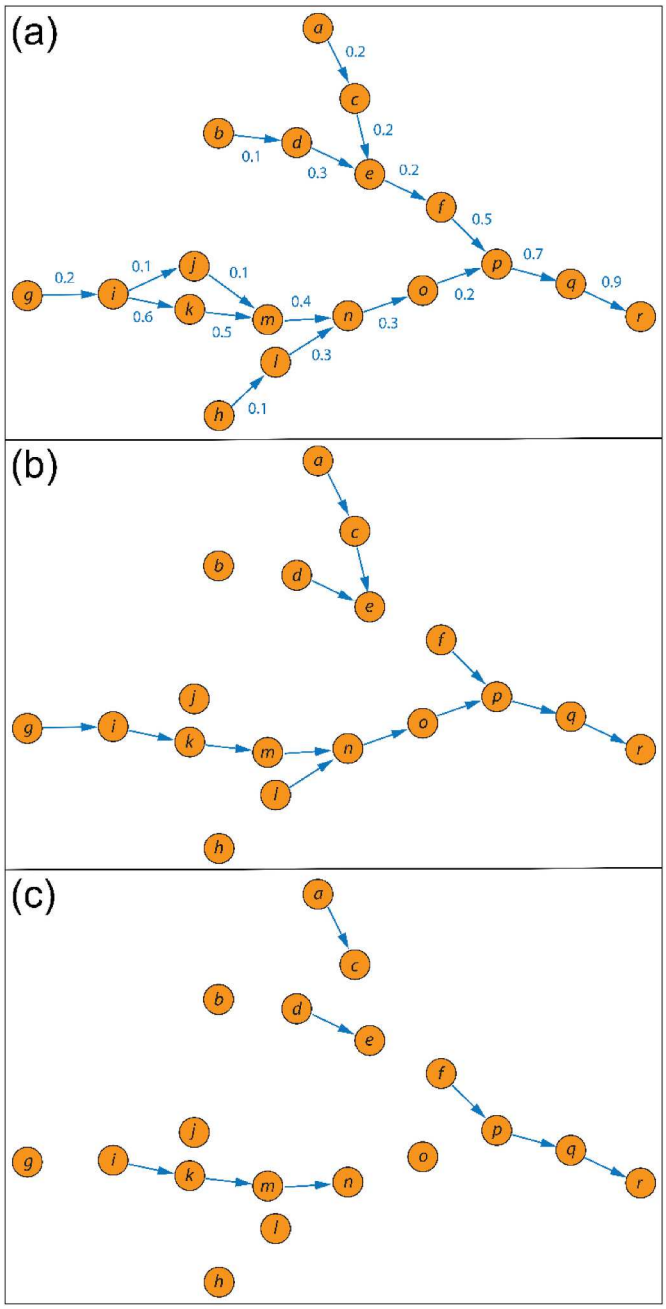


Fig. 1. A series of DAGs representing a drying stream network over time. Nodes are lettered and arcs are indicated with arrows indicating flow direction. The stream dries from (a) all arcs (segments) present, to (b) five arcs absent, to (c) ten arcs absent. Numbers in (a) are hypothetical probabilities of surface water presence which could serve as arc weights in weighted graph analyses.

2. Non-perennial streams as DAGs

2.1. Definitions and terminology

A directed graph (digraph) is an ordered pair $D = (N,A)$, where N is a set of *nodes* and A is a set of *arcs* that link the nodes. The *order* of a digraph, also called the *nodal cardinality*, is the number of digraph nodes, and is denoted as $n = |N|$, whereas the *size* of a digraph is the number of arcs. The size of a digraph, also called the *arc cardinality*, is denoted $m = |A|$. An arc from node u to node v is denoted $\overrightarrow{uv}$. This specification defines node u as the *tail* of $\overrightarrow{uv}$ and v as the *head* of $\overrightarrow{uv}$. In a digraph we can distinguish the *indegree* and *outdegree* of a node as the number of arcs

with that node as head and the number of arcs with that node as tail. The *degree* of a node is the sum of its indegree and outdegree.

Streams networks can be represented as digraphs, with streams segments constituting arcs, bounded by nodes occurring at hydrologically meaningful locations, such as sensor sites, network confluences, splits, sources, and sinks (Sarhad et al., 2014; Jin et al., 2019). Because they are strongly driven by hydrological potentials resulting from fixed elevational gradients, graphs that are most appropriate for describing passive stream network characteristics such as transport and discharge, will be both directed (with an orientation from *sources* to *sink*) and acyclic, meaning the graph will contain no directed cycles (Fig. 1). A directed graph *cycle* occurs when a directed *path* starts and ends at the same node.

The directed acyclic graph (DAG) in Fig. 1a represents a fully-wetted non-perennial stream network with 18 arcs (stream segments) and 18 nodes (stream point locations separating segments). Specifically, $N = \{a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r\}$, and $A = \{\overrightarrow{ac}, \overrightarrow{bd}, \overrightarrow{ce}, \overrightarrow{de}, \overrightarrow{ef}, \overrightarrow{gi}, \overrightarrow{hl}, \overrightarrow{ij}, \overrightarrow{ik}, \overrightarrow{jm}, \overrightarrow{km}, \overrightarrow{ln}, \overrightarrow{mn}, \overrightarrow{no}, \overrightarrow{op}, \overrightarrow{fp}, \overrightarrow{pq}, \overrightarrow{qr}\}$. The graph is acyclic because water can only flow downhill (in the direction of arc arrows). We also note that all nodes have outdegree = 1 except for the sink node (outdegree = 0), node *i,* where a split (island) occurs (as allowed in acyclic digraphs), and that all nodes except those at sources and junctions have indegree = 1. As is typical for stream networks, all confluence nodes in Fig. 1a have indegree = 2.

A digraph is *strongly connected* if every node is reachable from every other node. A digraph is *weakly connected* if every node is reachable after replacing all oriented arcs with bidirectional links between adjacent nodes. In a *disconnected* digraph, there will exist at least two nodes that cannot be connected, even with bidirectional links. In a digraph representing nonperennial stream flow, there will be no bidirectional links. Thus, these stream DAGs will never be strongly connected, and will transition from weakly connected (Fig. 1a), to disconnected (Fig. 1b and c).

2.2. The adjacency matrix

Graphs can be represented with an $n \times n$ *adjacency matrix*, $\mathbf{A}$, whose entries, A_{ij}, indicate that an arc exists from node i to j, with $A_{ij} = 1$, or that there is no arc from i to j, with $A_{ij} = 0$. The adjacency matrix can be used to describe many network characteristics. For instance, by applying the definition of matrix multiplication, the i,j entry in $\mathbf{A}^k$ will give the number of paths in the graph from node i to node j of length k. As an example, computation of $\mathbf{A}^8$ for the adjacency matrix from the stream network in Fig. 1a reveals two paths of length eight, both starting at node g and ending at the sink node, r. The paths are: (g,i,j,m,n,o,p,q,r), and (g,i,k,m,n,o,p,q,r). Other, more complex, matrix representations of graphs include the distance matrix, and the graph Laplacian and its variants (see Newman, 2018).

2.3. Weighted graphs

Nuance and realism can be enhanced in stream DAGs by adding information to nodes and/or arcs in the form of *weights*. Weighting information particularly relevant to non-perennial stream DAGs includes flow rates, instream lengths, probabilities of aquatic organism dispersal, water quality components including nutrients or sediment (i.e., *loading*; Maidment, 1996), upstream drainage area, and/or probabilities of surface and subsurface water presence. Weights can be assessed alongside the strictly topological relationships of nodes and arcs when describing DAGs. For instance, in Fig. 1a, the junction nodes e and p both have indegree = 2 and outdegree = 1. However, for the arcs $\overrightarrow{ce}$, $\overrightarrow{de}$, $\overrightarrow{ef}$ and $\overrightarrow{fp}$, $\overrightarrow{op}$, $\overrightarrow{pq}$ we have the segment surface water probabilities: 0.2, 0.3, 0.2, and 0.5, 0.2, 0.7, respectively (Fig. 1a). Viewing these numbers as arc weights, we can sum the lengths of bounding arcs to obtain a weighted nodal measure called *strength* (Barrat et al., 2004). Node e has strength 0.7, and node p has strength 1.6, potentially emphasizing the importance of node p over node e.

2.4. Local versus global perspectives

Both unweighted and weighted graph-theoretic approaches for describing stream DAGs can be separated into *local measures* that describe the characteristics of individual nodes or arcs, and *global measures* that summarize the characteristics of an entire digraph network. Sections 4 and 5, which address unweighted and weighted measures respectively, are each split into subsections to address these distinct perspectives.

2.5. Additional considerations

While extensive, we acknowledge that our treatment of DAGs here is intentionally simplistic and does not include all possible approaches. For example, we do not consider the vast array of methods associated with the detection of network community structures (e.g., *spectral cluster analysis*; Newman, 2006). Thorough mathematical considerations of digraphs and graphical networks are given in Bang-Jensen and Gutin (2008) and Newman (2018), respectively.

3. The *streamDAG* package

This paper serves as a formal introduction to the *streamDAG* **R** package (Aho et al., 2022), which allows visualization and analysis of non-perennial stream networks from a DAG perspective. The *streamDAG* package facilitates codification and modification of stream networks using non-perennial stream node or arc presence/absence data, and implementation of a wide variety of metrics appropriate for non-perennial stream DAGs including local and global measures for both unweighted and weighted graphs. The *streamDAG* package utilizes the programming framework of the expansive graph theory package *igraph* (Csardi and Nepusz, 2006) which can be run within **R** (R Core Team, 2022), Python, and Mathematica language environments (Igraph, 2022). The *streamDAG* package is currently maintained by the first author of this manuscript. Along with conventional graph theoretic approaches, the package contains functions not available elsewhere, including visibility algorithms (Luque et al., 2009), and Bayesian models for the probability of stream surface water presence in stream segments. Posterior distributions for this probability allow Bayesian extensions to Bernoulli stream length (Botter and Durighetto, 2020), and communication distance (see Aho et al., 2023), which are weighted DAG metrics available in *streamDAG*.

3.1. Package installation and loading

Following installation of the **R** *devtools* package, for instance, by typing: `install.packages("devtools")` at the **R** command line, the *streamDAG* package can be installed for Windows, MacOS, and Linux/Unix-alike platforms from its GitHub repository using:

```
library(devtools)
install_github("moondog1969/streamDAG")
```

And subsequently loaded, using simply:

```
library(streamDAG)
```

Installation and loading of the *streamDAG* package will result in automatic installation and loading of the **R** *igraph* package, respectively. The *streamDAG* package will be formally released to the Comprehensive **R** Archive Network (CRAN) following publication of this manuscript.

3.2. Comparisons of the streamDAG package to existing R DAG software

Aside from *igraph*, a number of **R** packages have been previously developed for DAG-like applications. For example, the package *dagitty*

Table 1
Local (generally nodal) unweighted graph metrics appropriate for stream DAGs. While all metrics listed are potentially useful for the analysis of non-perennial streams, recommended metrics (those we have found to be particularly useful) are denoted "X" in column one.

	Metric, M	Details	M applied to node p in Fig. 1a and c, i.e., $M(p_a)$ and $M(p_c)$.[a]	Type of summary	*streamDAG* code[b] G = *igraph* graph object
X	Number of nodes that can reach i th node	Largest for sink	$M(p_a) = 15$ $M(p_c) = 1$	Nodal connectivity	`local.summary(G,"n. nodes")`
	Number of nodes reachable from the ith node	Largest for sources	$M(p_a) = 2$ $M(p_c) = 2$	Nodal connectivity	`local.summary(G,"n. nodes","out")`
X	Number of paths ending at the ith node	Largest for sink	$M(p_a) = 17$ $M(p_c) = 2$	Nodal connectivity	`local.summary(G,"n. paths")`
	Number of paths beginning at the ith node	Largest for sources	$M(p_a) = 2$ $M(p_c) = 2$	Nodal connectivity	`local.summary(G,"n. paths","out")`
X	Size of the upstream network ending at (draining into) the ith node	Largest for sink	$M(p_a) = 16$ $M(p_c) = 1$	Nodal centrality	`local.summary(G,"size. intact.in")`
	Alpha centrality of ith node (Katz 1953; Bonacich and Lloyd,2001).	See Eq. (1), and Newman (2018).	$M(p_a) = 18$ $M(p_c) = 2$	Nodal centrality	`local.summary(G,"alpha. cent")`
	PageRank centrality of ith node (Brin and Page, 1998)	See details in Newman (2018).	$M(p_a) = 0.15$ $M(p_c) = 0.64$	Nodal centrality	`local.summary(G,"page. rank")`
	Nodal betweenness centrality of ith node (Freeman, 1977)	Number of paths passing through the ith node. See Eq. (2). May not accord with user conceptions of centrality.	$M(p_a) = 30$ $M(p_c) = 2$	Nodal betweenness	`local.summary (G,"betweenness")`
	Arc betweenness of the kth arc (Girvan and Newman, 2002)	The number of paths that pass through the kth arc. May not accord with user conceptions of centrality.	$M(\overrightarrow{op}_a) = 30$ $M(\overrightarrow{op}_c) = 2$	Arc betweenness	`local.summary(G,"arc. betweenness")`
X	Improved closeness centrality of ith node (Beauchamp, 1965)	See Eq. 3	$M(p_a) = 101.2$ $M(p_c) = 17.0$	Nodal centrality, connectivity	`local.summary(G,"imp. closeness")`
	Visibility of the ith node to and from other nodes	See Luque et al. (2009). May be difficult to calculate or conceptualize in disconnected graphs.	$M(p_a) = 4$ $M(p_c) = 0$	Nodal importance	`multi.path.visibility(G, source = "source.nodes", sink = "sink.node")`
X	Strahler stream order (Strahler, 1957) or Shreve stream order (Shreve,1966) of the ith node or kth arc.	See description in Section 4.1.4.	Strahler: $M(p_a) = 3$ $M(p_c) = 1$ Shreve: $M(p_a) = 4$ $M(p_c) = 1$	Node or arc nestedness	`stream.order(G, "strahler", sink = "sink. node")` `stream.order(G, "shreve", sink = "sink.node")`
	Descriptive statistics of upstream shortest in-path lengths to the ith node (recommended) or downstream out-path lengths from the ith node; e.g., mean ($\bar{x}$), variance (s^2), skew (g_1), and kurtosis (g_2).	In-path statistics will be undefined if the size of the upstream network is 0, i.e., source or disconnected nodes. Stream path maxima equal nodal eccentricities.	$M(p_a)$: $\bar{x} = 3.27$ $s^2 = 1.92$ $M(p_c)$: $\bar{x} = 1$ $s^2 = 0$	Nodal complexity, connectivity, and topological nuance	`local.summary(G, "path.len.summary")`
X	Average in-efficiency (recommended) and out-efficiency of the ith node (Latora and Marchiori, 2001).	Average of reciprocal distances.	$M(p_a) = 0.35$ $M(p_c) = 0.059$	Nodal connectivity	`local.summary(G,"avg. efficiency")`
	Descriptive statistics for indegree of nodes in upstream in-paths (recommended) or downstream out-path lengths for the ith node; e.g., mean ($\bar{x}$), variance (s^2), skew (g_1), and kurtosis (g_2).	In-path statistics will be undefined if the size of the upstream network is 0, i.e., source or disconnected nodes.	$M(p_a)$: $\bar{x} = 1.87$ $s^2 = 0.65$ $M(p_c)$: $\bar{x} = 0$ $s^2 = 0$	Nodal complexity, connectivity, and topological nuance	`local.summary(G, "path. deg.summary")`

[a] As recommended, default in-path lengths were used for examples of path length summaries (row 13) and indegree was used for degree summaries (e.g., row 15).
[b] For **R** code: `G` = *igraph* graph object. "`sink.node`" = a text string naming the sink node in `G`, "`source.nodes`" = a character vector naming the source node(s) in `G`. Most listed metrics can be obtained (for all nodes) simultaneously by typing: `local.summary(G, "all")`.

(Textor et al., 2016), its graphics extension *ggdag* (Barrett 2023), and *dagR* (Breitling et al., 2021) have been built primarily for the analysis and plotting of causal diagram structures of variables, including structural equation models (Wright 1934). This usage, however, falls outside of the scope of conventional graph theoretic considerations (e.g., Bang-Jensen and Gutin, 2008; Newman, 2018) and thus is distinct from the functionality of *streamDAG*. The **R** packages *shinyDAG* (Creed and Gerke 2018), *DiagrammeR* (Iannone 2022) and *visNetwork* (Almende et al., 2022) have been built solely for the purpose of creating network diagrams, and not for the quantitative analysis of DAGs.

3.3. General comparisons of streamDAG to existing stream network software

The goal of quantifying stream network characteristics, including network connectivity, has driven the publication of numerous computer algorithms and software packages. These include, but are not limited to the **R** packages *rtop* (Skøien et al., 2012), *SSN* (Ver Hoef et al., 2014), *riverconn* (Baldan et al., 2022), *riverdist* (Tyers, 2017), and *streamDepletr* (Zipper, 2020), the *r.stream* module (Jasiewicz and Metz, 2011) for the *GRASS* open source Geographic Information System (GIS) platform (GRASS Development Team, 2022), the *Arc Hydro* toolkit (Maidment 2002) for the commercial *ArcGIS®* platform, the *River Tool Network Toolkit* (*RivTool*; Duarte et al., 2019), and the *geostatistical connectivity* algorithm of Pardo-Igúzquiza and Dowd (2003), originally written in FORTRAN, and later codified in MATLAB (Trigg et al., 2013). The *streamDAG* package can be distinguished from these efforts in at least two ways.

First, *streamDAG* algorithms codify graph theoretic metrics relevant to non-perennial streams and classic surficial hydrological measures that can be viewed in a DAG context. In contrast, the *riverconn* package considers existing organismal dispersal connectivity metrics, with the potential for bidirectional (non-DAG) links between adjacent nodes, and with nodes as reaches and arcs as barriers or connections or splits, rather

Table 2
Global metrics appropriate for stream DAGs. While all listed metrics are potentially useful for the analysis of non-perennial streams, recommended metrics (those likely to be particularly useful) are denoted "X" in column one.

	Metric, *M*	Definition, details	*M* applied to graphs in Fig. 1a and c, i.e., $M(G_a)$ and $M(G_c)$.[a]	Type of summary	*streamDAG* code[b] G = *igraph* graph object
	Graph diameter, generally = *height* of the sink = in-eccentricity of the sink.	The length of the longest (non-infinite) path.	$M(G_a) = 8$ $M(G_c) = 3$	DAG complexity	`global.summary(G, "diameter")`
	Graph order	No. of nodes = *n*.	$M(G_a) = 18$ $M(G_c) = 18$	DAG complexity	`global.summary(G, "graph.order")`
	Size	No. of arcs = *m*, i.e., the number of wetted stream segments.	$M(G_a) = 18$ $M(G_c) = 8$	DAG complexity	`global.summary(G, "size")`
X	Number of source nodes and/or distinct stream reaches connected to sink	See Section 2.1	$M(G_a) = 4$ $M(G_c) = 1$	DAG complexity	`global.summary(G, "n.sources", sink = "sink.node")`
X	Number of paths to sink	See Section 2.1	$M(G_a) = 19$ $M(G_c) = 3$	DAG complexity/ connectivity	`global.summary(G, "n.paths.sink", sink = "sink.node")`
X	Global Strahler number (Strahler, 1957) or global Shreve stream number (Shreve, 1966)	Strahler or Shreve stream order of the sink node.	Strahler: $M(G_a) = 3$ $M(G_c) = 1$ Shreve: $M(G_a) = 4$ $M(G_c) = 1$	DAG complexity and nestedness	`global.summary(G, "strahler.num", sink = "sink.node")` `global.summary(G, "shreve.num", sink = "sink.node")`
X	Descriptive statistics of shortest upstream in-path lengths for the sink node (recommended), or entire network, and/or shortest downstream out-path lengths for the entire network; e.g., mean ($\overline{x}$), variance (s^2), skew (g_1), and kurtosis (g_2).	For descriptions of statistical estimators, see Aho (2014).	$M(G_a)$: $\overline{x} = 4.82$ $s^2 = 3.28$ $M(G_c)$: $\overline{x} = 2$ $s^2 = 1$	DAG complexity, connectivity, topological nuance	`global.summary(G, "sink.path.len.summary", sink = "sink.node")`
X	Descriptive statistics for the global indegree (recommended) or outdegree distribution e.g., mean ($\overline{x}$), variance (s^2), skew (g_1), and kurtosis (g_2).	One can consider the viability of DAG theoretical degree distributions including random (Erdős and Rényi, 1959), chaotic (Lacasa and Toral, 2010), or scale-free (Li et al., 2005).	$M(G_a)$: $\overline{x} = 1$ $s^2 = 0.44$ $M(G_c)$: $\overline{x} = 0.44$ $s^2 = 0.25$	DAG complexity, topological nuance	`global.summary(G, "deg.summary")`
X	Global efficiency (Ek et al., 2015)	The mean of all pairwise nodal efficiencies, see Eq. (4).	$M(G_a) = 0.13$ $M(G_c) = 0.03$	DAG connectivity	`global.summary(G, "global.efficiency")`
X	Harary index (Plavšić et al., 1993)	See Eq. 5	$M(G_a) = 19.3$ $M(G_c) = 5.3$	DAG connectivity	`global.summary(G, "harary")`
X	Directed first and second Zagreb index (Gutman et al., 1975)	See Section 4.2.3. Will increase with both path length and branching complexity.	1st Zagreb: $M(G_a) = 23$ $M(G_c) = 14$ ^{2nd} Zagreb: $M(G_a) = 8$ $M(G_c) = 4$	DAG complexity	`global.summary(G, "fst.zagreb")` `global.summary(G, "scd.zagreb")`
X	Directed atom-bond connectivity (Estrada et al., 1998)	See Section 4.2.3. Will only increase with increasing branching complexity.	$M(G_a) = 3.54$ $M(G_c) = 0$	DAG complexity	`global.summary(G, "ABC")`
	Assortativity index (Newman, 2002)	The correlation of the in- or outdegree of arc bounding nodes. The index will be undefined in simple path networks. The $r(+,+)$ and $r(+,-)$ bases are unlikely to be useful in stream networks.	$r(+,-)$: $M(G_a) = -0.3$ $M(G_c) = 0$ $r(-,+)$: $M(G_a) = 0.17$ $M(G_c) = \text{NaN}$	DAG assortativity	`global.summary(G, "assort.in.out")` `global.summary(G, "assort.in.in")`

[a] As recommended, default in-path lengths were used for examples of path length summaries (row 7), and indegree was used for degree summaries (row 8).
[b] For **R** code: `G` = *igraph* graph object, "`sink.node`" = a text string naming the sink node in `G`. All listed metrics can be obtained simultaneously by typing: `global.summary(G, "all", sink = "sink.node")`.

than arcs as stream segments. The *RivTool* toolkit generally focuses on the physical and topographic relationships between rivers and their surrounding basins based on GIS data, not stream networks as graph-theoretic entities. This characteristic also distinguishes *streamDAG* from the **R** package *riverdist*, which calculates instream distances using GIS shapefiles, and the GIS toolkits *Arc Hydro* and *r.stream*, although limited functional overlap occurs, including algorithms for stream order. The *SSN* package is not concerned with graph theory or hydrologic metrics, but with the development and application of stream-appropriate spatial covariance structures, including those of Cressie et al. (2006) and Ver Hoef et al. (2006), to allow the extension of spatial statistical models to streams. The *SSN* framework has been expanded by other authors to include, among other applications, Bayesian generalized linear models (the *SSNbayes* **R** package; Santos-Fernandez et al., 2022). The package *rtop* (Skøien et al., 2012) uses covariance approaches other than those in *SSN* (see Skøien et al., 2006) for the same purpose: to produce stream network spatial models. The focus of *streamDAG* on surface flow networks is also very different from *streamDepletr*, which estimates potential pumping impacts on streamflow based on inferred stream-aquifer connections (Zipper, 2020).

Second, *streamDAG* maintains a focus on non-perennial streams with functions capable of incorporating water presence/absence data at nodes and arcs. In contrast, *riverconn* connectivity metrics stress the importance of physical barriers to streamflow, particularly anthropogenic dams, which are unlikely to occur in non-perennial streams. The non-perennial focus of *streamDAG* is also distinct from the grid-reliant geostatistical connectivity algorithm (Pardo-Igúzquiza and Dowd, 2003), which lends itself to analysis of remotely sensed floodplain images based on continuous grids (Trigg et al., 2013; Karim et al., 2015; Chen et al., 2020).

4. Unweighted measures for non-perennial stream DAGs

This section concerns unweighted *streamDAG* measures, i.e., metrics that do not require or use ancillary arc or node weighting information. Metrics under consideration include those appropriate for local stream DAG perspectives (Section 2.1; Table 1), and those appropriate for global analyses of stream networks (Section 2.2; Table 2).

4.1. Local measures

Local DAG measures summarize particular locations in a graph, e.g., individual nodes, arcs, and subgraph regions. These approaches encompass a broad suite of potentially relevant characteristics for stream networks including the importance of a location to network function and integrity, and local connectedness, complexity, and network nestedness.

4.1.1. Centrality

Possibly the most common and widespread measures of local network importance are those that consider nodal *centrality*. Many metrics of nodal centrality have been proposed, reflecting myriad perspectives on graph centrality (Bonacich, 1987). These include *degree centrality* (i.e., the nodal degree), *eigenvector centrality* (Bonacich 1972, 1987), *authority centrality* (Kleinberg, 1999), *closeness centrality, information centrality* (Brandes and Fleischer, 2005), and *random walk betweenness* (Newman, 2005), among others (Borgatti, 2005). See Borgatti and Everett (2006) for a mathematical classification of centrality indices, Schoch and Brandis (2016) for unifying perspectives on these measures, and Boldi and Vigna (2014) for a suite of centrality axioms.

Unfortunately, many conventional centrality approaches will be uninformative, incalculable, or otherwise problematic when applied to stream networks. For instance, in a stream DAG, *degree centrality* may be largely invariant because all nodes in simple paths will have the same degree centrality (because indegree = outdegree = 1), and, while nodes at confluences joining two arcs may be common (indegree = 2), splits may be rare. *Eigenvector centrality*, the corresponding entry in the principal eigenvector of the graph adjacency matrix, extends degree centrality by accounting for a node's connection to nodes that are themselves important (Newman, 2018). However, because the adjacency matrix of a directed graph will be asymmetric, it will have distinct left- and right-hand eigenvectors. In analyses of stream DAGs, one might use the right-hand principal eigenvector because centrality measures of a node will then be based on upstream input nodes (Newman, 2018). However, other problems arise, including the fact that source nodes, which must have indegree zero, will drive all downstream nodes to have an eigenvector centrality of zero (Newman 2018, pg. 162).

One solution to this problem is *alpha* or *Katz centrality* (Katz, 1953; Bonacich and Lloyd 2001) which, following Newman (2018), is defined for all nodes simultaneously by

$$\boldsymbol{x} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1}\mathbf{1}, \tag{1}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix, $\mathbf{1} = (1, 1, \ldots, 1)$ with n entries, and α is a user-defined constant that allows weighting all nodes with a small but nonzero amount of initial centrality. Many researchers define α to be slightly less than the reciprocal of the primary eigenvalue because such a number: 1) allows the computational convergence of Eq 1, and 2) results in an outcome similar to eigenvector centrality. The *streamDAG* package utilizes the existing alpha centrality algorithm from *igraph*, which uses $\alpha = 1$ by default. The *PageRank* metric (Brin and Page, 1998), is similar to alpha centrality, but ensures that the centrality of a node is proportional to the centrality of the neighbors of the node divided by their outdegree (Newman, 2018). Newman (2018, pg. 165) describes methods for terms with outdegree zero. In stream DAGs, nodes with larger alpha-centrality and PageRank outcomes can be viewed as having greater influence and importance in the stream network.

Nodal betweenness centrality (Freeman, 1977) measures how often a node lies between other nodes. The nodal betweenness centrality of the ith node has the form:

$$B_i = \sum_{uv} \frac{n^i_{uv}}{n_{uv}} \tag{2}$$

where n^i_{uv} is the number of shortest paths from node u to node v that pass-through node i, and n_{uv} is the total number of shortest paths from u to v, which is at most one if there are no splits. In the case that $n_{uv} = 0$ (and hence $n^i_{uv} = 0$), the ratio is assumed to be 0. Thus, unlike other centrality measures, nodal betweenness centrality does not necessarily quantify how well-connected a node is, but measures of how often a node falls between other nodes. Because of its unique conceptualization of centrality, betweenness centrality may provide assessments of nodal centrality in non-perennial streams distinct from other measures, although this perspective may not accord with conventional conceptions of hydrological connectivity and importance (cf. Terui et al., 2021). In stream DAGs, nodal betweenness centrality will be highest at confluences or splits, and at locations near the middle of reaches, and lowest for source and sink nodes, which will have a betweenness centrality of zero. Betweenness centrality of arcs can also be calculated. Specifically, *arc betweenness centrality* is the number of shortest paths that pass through an arc (Girvan and Newman, 2002).

Closeness centrality (Bavelas, 1950) measures the mean shortest path distance from a node to all other nodes. This metric is also poorly suited to non-perennial stream networks because a DAG will not be strongly connected and may be disconnected (Fig. 1), causing many conceptual internodal distances to be infinite. To account for this, several modifications to closeness centrality have been proposed, including Lin's index (Lin 1976) and improved closeness centrality (Beauchamp, 1965). *Improved closeness centrality*, also called *harmonic centrality* (Rochat, 2009), and *valued centrality* (Dekker, 2005), is based on the reciprocals of nodal shortest path distances from the i th node to all other nodes, $1/\delta_{i,j}$ where $j \neq i = 1, 2, \ldots, n-1$. Specifically, the *improved closeness centrality* for the i th node is:

$$C_i = (n-1)\sum_{i \neq j} \frac{1}{\delta_{i,j}} \tag{3}$$

where, for disconnected nodes, the reciprocal of an infinite distance is taken to be zero. In a stream DAG, a node will have high *improved closeness centrality* if it has many adjacent neighboring nodes, and few disconnected internodal relationships which will not contribute to the summation in Eq. (3).

Boldi and Vigna (2014) evaluated the characteristics of 11 centrality indices, including degree centrality, alpha-centrality, betweenness centrality, Lin's centrality, and improved closeness centrality with respect to three well-reasoned axioms of centrality. Improved closeness centrality was the only index that met the requirements of all three axioms. Based on this assessment, and our own analyses of artificial and authentic stream DAGs, we recommend the use of improved closeness centrality over other local centrality measures for describing local importance of nodes to overall network function in non-perennial streams.

4.1.2. Summaries of paths and distances

The connectivity and importance of a stream node can be considered by summarizing the distribution of its path lengths using conventional descriptive statistics. Path lengths include the lengths of paths that *end* at a particular node (*in-path lengths*) and path lengths that *begin* at a particular node (*out-path lengths*). Thus, no in-paths will exist for source nodes and no out-paths will exist for sink nodes. In the summarization of stream DAGs, in-paths are likely to be of greater interest (and serve as the default for relevant *streamDAG* functions) rather than out-paths. This is because the former allows consideration of the capacity of a node to be

an intermediate or final repository of upstream information (Newman 2018, pg. 162). For stream DAGs, it is reasonable to ignore nonexistent upstream and disconnected paths (see discussion in Newman 2018, pg. 133).

The most common statistical summary of nodal path lengths is *mean path length* (Albert and Barabási, 2002). The connection of the *i*th node to distant nodes will increase the *i*th node's mean path length, emphasizing its increased importance to the network. Additional node-level topological nuances may be revealed by other statistical measures, such as the heterogeneity of path lengths (e.g., the variance) and the symmetry and peakedness of path length distributions (e.g., the skew and kurtosis) of a node. In these summaries, *streamDAG* calculates population variances. That is, for the *i*th node, $s^2 = n^{-1}\sum_{j=1}^{n}(x_i - \bar{x})^2$, where x_j is the *j*th path length for the *i*th node and $\bar{x}$ is the *i*th node path length population mean. This approach is valid because the number of possible path lengths for the *i*th node, *n*, is finite and defines the population size under consideration (Aho 2014). Ignoring impossible (disconnected and upstream) paths allows computation of the *eccentricity* of a DAG node, that is, the longest path distance between that node and all other nodes. The maximum DAG in-path length to the *i*th node is the *in-eccentricity* of the *i*th node (often called *height*), and the maximum DAG out-path length from the *i*th node is the *out-eccentricity* of the *i*th node.

The reciprocal of the distance between nodes *i* and *j* defines their *efficiency* (Latora and Marchiori, 2001). Reflecting the constraints of improved closeness centrality, efficiencies based on infinite distances are generally taken to be zero. In DAGs, *in-efficiencies* (based on in-paths) will be distinct from *out-efficiencies* (based on out-paths), allowing calculations of average in-and out-efficiency for individual nodes. The former is the default for relevant *streamDAG* functions. Note that the overall mean efficiency of the *i*th node (based on both in-efficiencies and out-efficiencies) will be the improved closeness centrality of the *i*th node (Eq (3)), times $\frac{1}{n(n-1)}$. As with improved closeness centrality, high mean efficiencies will occur for stream nodes with many close neighbors and few disconnected associations.

4.1.3. Visibility

Visibility graphs (Lacasa et al., 2008; Luque et al., 2009; Lacasa and Toral, 2010) allow summaries of nodal importance based on the *visibility* of nodes to other nodes within a sequential series. Specifically, nodes *i* and *j* will be visible to each other if, when node data are plotted as vertical bars (with bar heights designating nodal data outcomes), and bars are placed along the abscissa based on some ordering of nodes in the stream network, the tops of bars for nodes *i* and *j* can be connected with a straight line, uninterrupted by other bars (see Lacasa et al., 2008; Luque et al., 2009).

In a stream DAG, a node will always be visible from the node directly upstream (and vice versa), regardless of data outcomes, and nodes with larger data outcomes will be able to "see" more nodes and be "more visible" to other nodes than those with smaller outcomes. One potential source of nodal data for visibility graphs is the indegree or outdegree of the nodes themselves. Under this approach, high degree nodes, located at stream junctions or splits, and housed between long, simple paths, will have high visibility and will block visibility of downstream nodes from upstream nodes, and vice versa. The ordering of nodes is vitally important to the calculation of visibility. Visibility functions in *streamDAG* order nodes by identifying all paths from each source node to the sink and summing the all internodal node distances in each path. These sums are then sorted decreasingly. Visibility constitutes a unique and potentially useful method for quantifying nodal importance in stream networks. However, straightforward methods for implementation in disconnected streams are unclear and remain under development in *streamDAG*.

4.1.4. Stream nestedness and hierarchy

Several topological measures of branching complexity specific to stream networks have been proposed under the name *stream order*, not to be confused with *graph order* (the number of graph nodes). *Strahler stream order* (Strahler, 1957) is a "top down" system in which first order stream sections (and their associated nodes and arcs) occur at the outermost tributaries. A stream section resulting from the merging of tributaries of the same order will have a Strahler order one unit greater than the order of the tributaries. That is, a stream section downstream of a confluence of two first-order tributaries will be second-order. A stream section resulting from the merging of tributaries of different order will have the Strahler stream order of the tributary with the larger Strahler number. Under *Shreve stream order* (Shreve, 1966), a stream section resulting from the merging of tributaries will always have an order that is the sum of the order of those tributaries.

Some considerations are necessary when using nodal stream order in disconnected stream DAGs. One approach is to calculate stream order only for nodes in the subgraph containing the sink. This is the method used by the function `stream.order()` in *streamDAG* (Table 1). As an alternative, one could define separate subgraphs for each disconnected portion of the network and calculate nodal stream order summaries for each subgraph.

4.2. Global measures

Global DAG measures allow consideration of a stream network in its entirety. Statistical summaries (e.g., mean, median, variance) of local metrics, including degree and path lengths provide one global approach. For instance, the mean of all path lengths in a graph is a frequently used global metric. Other global path length summaries include *graph diameter* (the maximum eccentricity across all nodes) and the *graph radius* (the minimum eccentricity across all nodes).

4.2.1. Global stream order

While rarely applied for this purpose, stream order can be used to track changes in a stream's network structure by only considering network components with surface water presence, rather than the entire channel network (Godsey and Kirchner, 2014). The global Strahler stream or global Shreve stream order is the corresponding stream order of the sink node, which will be the maximum nodal stream order of the network (or the sink sub-network in disconnected stream DAGs). Extending our suggestions for local DAG measures of nestedness in Section 4.1.4, we recommend the use of stream order to describe and track the global hierarchical structure of non-perennial streams (Table 2).

4.2.2. Global efficiency

Global metrics that use sums of path distances, including the *Wiener index* (Wiener, 1947) and the *hyper-Wiener index* (Randić, 1993), are problematic for non-perennial stream DAGs, because as noted above, distances between disconnected nodes (and distances from downstream to upstream nodes) will be infinitely large. Several metrics, including *global efficiency* (Ek et al., 2015), *the Harary index* (Plavšić et al., 1993), and *Balaban's J-index* (Balaban, 1982), address this problem by considering scaled sums of nodal reciprocal distances, i.e., the nodal efficiencies. The global efficiency of a digraph *D* is simply the mean of all pairwise nodal efficiencies:

$$E(D) = \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} e_{i,j}, \tag{4}$$

where the efficiency between nodes *i* and *j*, for all $i \neq j$, is defined as $e_{i,j} = 1/\delta_{i,j}$, where $\delta_{i,j}$ is the distance from node *i* to node *j* in *D*. Global efficiency is closely related to the Harary index:

$$H(D) = \frac{1}{2} \sum_{1 \le i < j \le n} e_{i,j} = \frac{n(n-1)}{2} E(D) \tag{5}$$

Reflecting our recommendation of inverse distance metrics for local

summaries (e.g., improved closeness centrality), we recommend the Harary index and global efficiency as measures of global connectivity in non-perennial stream DAGs (Table 2).

4.2.3. I(D) metrics

A large number of global DAG metrics relevant to non-perennial streams share the same formulaic basis. Specifically, for an arc $\overrightarrow{uv} \in A$, denote the outdegree of u as d_u^+, and the indegree of v as d_v^-. Now, let $I(D)$ represent a general topopological index for a digraph, D, that depends on d_u^+ and d_v^-:

$$I(D) = \frac{1}{2}\sum_{\overrightarrow{uv} \in A} \omega\left(d_u^+, d_v^-\right). \tag{6}$$

Four basic configurations of the function ω in Eq (6) can be considered (Deng et al., 2022) where x is d_u^+ or d_u^- and y is d_v^+ or d_v^-.

1. If $\omega(x,y) = (xy)^\alpha$, then $I(D)$ is the directed *Randić index* for D if $\alpha = -\frac{1}{2}$ (Randić, 1975), the directed *second Zagreb index* if $\alpha = 1$ (Gutman et al., 1975), and the directed *modified second Zagreb index* if $\alpha = -1$ (Anthony and Marr, 2021).
2. If $\omega(x,y) = (x+y)^\alpha$, then $I(D)$ is the directed *sum-connectivity index* for D if $\alpha = -\frac{1}{2}$ (Zhou and Trinajstić, 2009; Zhong, 2012), and the directed *first Zagreb index* if $\alpha = 1$ (Gutman et al., 1975). Further, if $\omega(x,y) = 2(x+y)^{-1}$, then $I(D)$ is the directed *harmonic index* of D (Favaron et al., 1993).
3. If $\omega(x,y) = \sqrt{\frac{x+y-2}{xy}}$, then $I(D)$ is the directed *atom bond connectivity* of D (Estrada et al., 1998).
4. If $\omega(x,y) = \frac{\sqrt{xy}}{\frac{1}{2}(x+y)}$, then $I(D)$ is the directed *geometric-arithmetic index* for D (Vukičević and Furtula, 2009).

In Supplemental Materials S3 we provide reasons why only the $x = d_u^+$, $y = d_v^-$ variant (as given in Eq. (6)) should be used for describing stream networks, and provide more thorough description of $I(D)$ metrics, including graphical comparisons of the performance of the metrics, and a mention of multiplicative forms of Eq. (6).

For the common case in stream networks of a digraph with no splits, a straightforward computation of $I(D)$ metrics is possible when using the recommended d_u^+, d_v^- variant. Under this framework, metrics that follow $\omega(x,y) = (xy)^\alpha$, including the directed Randić index, will equal $\frac{1}{2}\sum_{i=1}^{n-1} k_i^\alpha$, for a digraph with order n, where k_i denotes d_v^- for the ith arc $\overrightarrow{uv}$, and methods that follow $\omega(x,y) = (x+y)^\alpha$, including the directed sum-connectivity index, will equal $\frac{1}{2}\sum_{i=1}^{n-1}(1+k_i)^\alpha$. Therefore, if $\alpha < 0$, all $(xy)^\alpha$ metrics (including the directed Randić, and directed modified second Zagreb indices) and all $(x+y)^\alpha$ metrics (including the directed sum-connectivity) will decrease with increased branching complexity (increasing numbers of arcs at joins) given fixed graph order. Conversely, if $\alpha > 0$ these index families will increase with increased branching complexity. It is also possible to verify that for fixed graph order, the directed geometric-arithmetic index decreases with increasing numbers of arcs at a join and that this trend is reversed for the atom bond connectivity. For an unbranched path on n nodes the values in configurations 1 and 4 specialize to $\frac{n-1}{2}$ and the value in configuration 2 is $(n-1)2^{\alpha-1}$. Clearly, the directed atom bond connectivity numerator, $\sqrt{x+y-2}$, will equal zero when arcs are part of an unbranched path, causing the index summation to remain unchanged unless a join or split occurs.

Based on this summary. and content in Supplemental Materials S3, we recommend application of $I(D)$ metrics using the d_u^+, d_v^- basis and $\alpha > 0$ under configurations 1 and 2 to describe global connectivity and complexity in non-perennial stream DAGs (Table 2). Metrics using this framework, including the directed first and second Zagreb indices will increase with both increasing path length and increasing branch complexity in accordance with existing conceptions of hydrologic complexity (Terui et al., 2021). Nonetheless, because these metrics (along with many other DAG indices described here) can be made arbitrarily large with the addition of user-defined nodes in paths, they should only be used for tracking global changes in a single network with node locations fixed over time, or for comparing multiple networks with identical node designation criteria (see Section 7.3.1). We recommend the use of atom bond connectivity to track branching complexity, independent of path lengths (Table 2).

4.2.4. Assortativity

Graph assortativity quantifies the prevalence of network arcs with similar increasing or decreasing nodal indegree and outdgree patterns in their bounding nodes. For example, a strongly nested stream network with a high frequency of confluence nodes will have high assortativity because arcs will often have upstream bounding nodes with outdegree 1 and downstream bounding nodes with indegree 2.

The definitive measure of graph assortativity is the *assortativity coefficient*, which is Pearson's correlation of the degree of pairs of arc bounding nodes (Newman, 2002). Let $\overrightarrow{u_i v_i} \in A$ define nodes and directionality of the ith arc, $i = 1, 2, 3, \ldots, m$. Now, let $\gamma, \tau \in \{-, +\}$ index the degree type: $- = in$, $+ = out$, and let (u_i^γ, v_i^τ), represent the γ- and τ-degree of the ith arc. Then, the general form of the assortativity coefficient is:

$$r(\gamma,\tau) = m^{-1}\frac{\sum_{i=1}^{m}\left(u_i^\gamma - \overline{u}^\gamma\right)\left(v_i^\tau - \overline{v}^\tau\right)}{s^\gamma s^\tau} \tag{7}$$

where $\overline{u}^\gamma$ and $\overline{v}^\gamma$ are the arithmetic means of the u_i^γ s and v_i^τ s, i.e., $\overline{u}^\gamma = m^{-1}\sum_{i=1}^{m} u_i^\gamma$, and s^γ and s^τ are the population standard deviations of the u_i^γ s and v_i^τ s, i.e., $s^\gamma = \sqrt{m^{-1}\sum_{i=1}^{m}\left(u_i^\gamma - \overline{u}^\gamma\right)^2}$. Reflecting considerations given for $I(D)$ metrics earlier, there are four possible forms to $r(\gamma,\tau)$, based on the indegree and outdegree designations of arc head and tail nodes (Foster et al., 2010). These are: $r(+,-)$, $r(-,+)$, $r(-,-)$, and $r(+,+)$. The correlations $r(+,+)$ and $r(+,-)$ will rarely be finite for stream networks because the outdegree of u will almost always be 1, resulting in $s^\gamma = 0$. Given constraints of Pearson's correlation, $r(\gamma,\tau)$ outcomes of zero indicate no assortative mixing, whereas positive or negative values indicate assortative or disassortative mixing, respectively. In stream DAGs, the correlations $r(-,-)$ and $r(-,+)$ will generally be disassortive because of the characteristic strong convergence of stream paths from sources to sink (e.g., Fig. 1 in Foster et al., 2010) in most stream networks.

5. Weighted measures for non-perennial stream DAGs

While purely topological measures may be useful for describing local importance and global connectivity in stream DAGs, they will be strongly affected by user-defined node designations and abstracted from many important characteristics of stream networks. To increase DAG realism (and potentially decrease the effect of topological biases), one can attribute relevant weighting information to nodes and arcs, e.g., flow rates, stream segments lengths, etc. Weights can be incorporated directly into several of the unweighted measures introduced in Section 4. A number of weighted methods described here were developed outside the explicit realm of graph-theory. They are included because of their prior use in describing stream networks and their straightforward extendibility to a weighted digraph framework. As with non-weighted stream DAGs, both local (Section 5.1) and global (Section 5.2) summaries are possible for weighted stream DAGs.

5.1. Local measures

Weighted local graph metrics include strength centrality (Section

Table 3
Weighted local (nodal, arc, and subgraph) metrics. While all listed metrics are potentially useful for the analysis of non-perennial streams, recommended metrics (those likely to be particularly useful) are denoted with "X" in column one.

	Metric	Definition and details	M applied to node p in Fig. 1a and c, with specified weights, i.e., $M(p_{a(w)})$ and $M(p_{c(w)})$.[a]	Weights	Type of summary	*streamDAG* code[b] `W` = *igraph* weighted graph object. Arc weights can be set using `E(W)$weight <- c(w1, w2, ...)` where `w1` and `w2` are weights intended for the first two arcs in `W`.
	Strength centrality of the *i*th node	Sum of weights from arcs adjoining *i*th node.	$M(p_{a(w)}) = 1.4$ $M(p_{c(w)}) = 1.2$	Any arc weights	Nodal importance	`igraph::strength(W)`
	Weighted alpha centrality of the *i*th node	See description of unweighted alpha-centrality in Section 4.1.1. Highly sensitive to weights.	$M(p_{a(w)}) = 1.98$ $M(p_{c(w)}) = 1.5$	Any arc weights	Nodal importance	`local.summary(W, "alpha.cent")`
X	Length of the upstream network ending at (draining into) the *i*th node, in measured units, e.g., meters.	Sum of length weights of path arcs for a subgraph rooted at the *i*th node.	$M(p_{a(w)}) = 24$ $M(p_{c(w)}) = 1.5$	Arc lengths	Nodal importance, connectivity	`size.intact.to.node(W, node = "node.name")`
X	Statistical summary of weighted upstream shortest in-path lengths (recommended) or downstream shortest out-path lengths for the *i*th node; e.g., mean ($\bar{x}$), variance (s^2), skew (g_1), and kurtosis (g_2).	In-path statistics will be undefined if the size of the upstream network is 0, i.e., source or disconnected nodes.	$M(p_{a(w)})$: $\bar{x} = 4.9$ $s^2 = 4.04$ $M(p_{c(w)})$: $\bar{x} = 1.5$ $s^2 = 0$	Arc lengths	Nodal importance, topological nuance	`local.summary(W, "path.len.summary")`
	Weighted visibility of the *i*th node	See description of unweighted visibility in Section 4.1.3.	$M(p_{a(w)}) = 9$ $M(p_{c(w)}) = 1$	Any node weights	Nodal importance	`multi.path.visibility(G, source = "source.nodes", sink = "sink.node", weigths = wei)`
X	Average Bernoulli stream length (Botter and Durighetto, 2020) of the *k*th arc.	Probability of the presence of surface water at the *k*th arc times the length of the *k*th arc.	$M(\overrightarrow{op}_{a(w)}) = 0.45$	Arc probability (of stream activity) and arc length	Probabilistic arc length nuance.	`bern.length(lengths, pa, "local")`
X	Average communication distance (Aho et al., 2023) of the *k*th arc.	Reciprocal probability of the presence of surface water at the *k*th arc times the length of the *k*th arc.	$M(\overrightarrow{op}_{a(w)}) = 7.5$	Arc inverse probability (of stream activity) and arc length	Probabilistic arc length nuance.	`bern.length(lengths, 1/pa, "local")`

[a] For examples (column 4), weights are probabilities of stream segment surface water presence shown in Fig. 1a. Exceptions include measures requiring stream length weights, including path length summaries (row 4). In this case, universal segment lengths of 1.5 units were applied. Average Bernoulli stream length and average communication distance (rows 6 and 7) required both stream length (= 1.5) and probabilities of stream segment surface water presence (Fig. 1a). The average probability of stream segment surface water presence for bounding arcs used for *nodal* weights for the weighted visibility metric (row 5). As recommended, default in-path lengths were used for path summaries (row 4).

[b] For **R** code: `W` = *igraph* weighted graph object, `"node.name"` = name of node of interest, "`sink.node`" = a text string naming the sink node in `W`, `"source.nodes"` = a character vector naming the source node(s) in `W`, `wei` = node weights, `lengths` = vector of instream arc lengths, `pa` = vector of probabilities of water presence at arcs.

2.3), and other similar weighted variants of degree centrality (e.g., Opsahl et al., 2010), weighted alpha-centrality, and weighted path length summaries (Table 3). Two important weighted graph measures whose development was driven by non-perennial stream research are *mean Bernoulli arc length* [i.e., arc length multiplied by the probability of arc presence; Botter and Durighetto (2020)] and *mean communication distance* [(i.e., arc length multiplied by the reciprocal probability of arc presence; Aho et al. (2023)]. Local mean Bernoulli stream length measures the average length of an arc (stream segment) when considering the presence of water at that arc as a Bernoulli random variable. Thus, for the *k*th arc, this metric will *increase* and approach the actual length of the segment as the probability for surface water presence approaches one. Local mean communication measures the average *effective* length of an arc for the transportation of water-borne materials, after accounting for flow rarity. Thus, for the *k*th arc, this metric will *decrease* and approach the actual length of the segment as the probability for surface water presence approaches one.

Under an entirely different framing of stream graphs, one can define each stream reach as an individual node and define an edge (undirected link) between these nodes as a confluence or barrier or split between the reaches (Baldan et al., 2022). Then it is possible to let $p_{i,j}$ be the probability of organism dispersal or stream transport of materials from reach i to reach j. Let w_j represent a connectivity-related weighting value for the jth reach, and let W be the sum of those weights over all l reaches, then the *Reach Connectivity Index* for the ith reach (*RCI*; Baldan et al., 2022) can be defined as:

$$RCI_i = \sum_{j=1, j\neq i}^{l} p_{i,j}\frac{w_j}{W} \tag{8}$$

As Baldan et al. (2022, Eq. 2.5) point out, an undirected edge ij can be replaced by two oppositely directed (upstream and downstream) arcs to which $p_{i,j}$ and $p_{j,i}$ can be assigned as potentially distinct probabilities, resulting in a non-DAG framework. The *RCI*, along with numerous variants, is codified in the *riverconn* **R** package (Baldan et al., 2022). A large number of weighting approaches are possible for Eq. (8) that are considered briefly in the next section and are considered thoroughly by Baldan et al. (2022). Other local probabilistic metrics of stream connectivity based on an undirected graph framework include the *local connectivity* metric of Garbin et al. (2019).

5.2. Global measures

Several existing network-level connectivity metrics from the hydrological literature can be viewed as weighted digraph measures. These include *Integral Connectivity Scale Length* (*ICSL*): i.e., the average distance between wet nodes in a stream network (Western et al., 2001; Ali and Roy, 2010), average Bernoulli stream network length: i.e., the sum of average Bernoulli arc lengths, and average network-level

Table 4
Weighted global metrics for stream DAGs. While all listed metrics are potentially useful for the analysis of non-perennial streams, recommended metrics (those likely to be particularly useful) are denoted with "X" in column one.

	Metric	Definition and details	M applied to graphs in Fig. 1a and c, under specified weights, i.e., $M(W_a)$ and $M(W_c)$.[a]	Weights	Type of summary	*streamDAG* code[b] `W` = *igraph* weighted graph object. Arc weights can be set using `E(W)$weight <- c(w1, w2, ...)` where `w1` and `w2` are weights intended for the first two arcs in `W`.
X	Integral connectivity scale length (ICSL)	Average distance connecting wet locations (i.e., weakly connected DAG nodes) based on Euclidean distances or hydrologic distances (Ali and Roy, 2010). Includes surface ICSL (Western et al., 2001), and subsurface and outlet ICSL (Ali and Roy, 2010).	$M(W_a) = 6.53$ $M(W_c) = 2.56$	Arc lengths or Euclidean distances	DAG connectivity based on avg. distance of weakly connected nodes	`ICSL(W)`
X	Weighted Harary index (Plavšić et al., 1993), weighted global efficiency (Ek et al., 2015)	See Eqs. (4) and (5)	Harary: $M(W_a) = 12.84$ $M(W_c) = 3.56$ Global efficiency: $M(W_a) = 12.84$ $M(W_c) = 3.56$	Arc lengths (allowing computation of reciprocal distances)	DAG connectivity based on reciprocal distances of all nodes	`harary(W)` `global.efficiency(W)`
	Average strength	See description for strength centrality.	$M(W_a) = 0.66$ $M(W_c) = 0.46$	Any arc weights	DAG connectivity based on weight sums of arcs adjoining nodes	`mean(igraph::strength(W))`
	Average alpha-centrality	See description for unweighted alpha-centrality.	$M(W_a) = 1.53$ $M(W_c) = 1.33$	Any arc weights	DAG connectivity	`global.summary(W, "avg.alpha.cent")`
X	Weighted size of sink subgraph	Weights of nodes for the sink-associated graph or subgraph are summed.	$M(W_a) = 27$ $M(W_c) = 4.5$	Arc lengths	DAG connectivity, complexity	`size.intact.to.sink(W, sink = "sink.name")`
X	Average Bernoulli network length (Botter and Durighetto, 2020)	The sum of the arc lengths multiplied by either corresponding probabilities of surface water presence or (for instantaneous measures) corresponding binary surface water presence/absence outcomes.	$M(W_a) = 27$ $M(W_c) = 12$	Arc surface water probability (or stream presence/absence data) and length	DAG connectivity in units of wetted network length	`bern.length(lengths, pa, "global")`
X	Average network-level communication distance (Aho et al., 2023)	The sum of the arc lengths multiplied by either corresponding reciprocal probabilities of surface water presence or (for instantaneous measures) corresponding reciprocal binary surface water presence/absence outcomes.	$M(W_a) = 27$ $M(W_c) = \infty$	Arc surface water reciprocal probability (or reciprocal stream presence/absence data) and length	DAG *dis*connectivity in units of effective network length, due to intermittency	`bern.length(lengths, 1/pa, "global")`

[a] For examples (column 4), weights are probabilities of stream segment surface water presence shown in Fig. 1a. Exceptions include measures requiring stream length weights, including *ICSL* (row 1), the weighted Harary index, weighted global efficiency (row 2), and the weighted size of sink subgraph (row 5). In these cases, universal segment lengths of 1.5 units were applied. Stream (arc) lengths are also required for average network Bernoulli stream length and average network communication distance (rows 6 and 7), along with either probabilities or binary outcomes surface water presence/absence outcomes. The latter approach (representing instantaneous conditions) is used in the Table.

[b] For **R** code: `W` = *igraph* weighted graph object, `"node.name"` = name of node of interest, "`sink.node"` = a text string naming the sink node in `W`, `"source.nodes"` = a character vector naming the source node(s) in `W`, `wei` = *nodal* weights, `lengths` = vector of instream arc lengths, `pa` = vector of probabilities of water presence (or binary water presence/absence outcomes) at arcs.

communication distance: i.e., the sum of average arc communication distances (Table 4). Multivariate Bernoulli outcomes representing surface water presence at all arcs (1 = presence, 0 = absence) can be used in the place of probabilities to track instantaneous Bernoulli network length (cf. Durighetto and Botter, 2022). This approach is problematic for communication distance because one or more arc surface water absences will result in infinite instantaneous network communication distances (see Table 4).

Reverting to a reach-as-node perspective, the Reach Connectivity Index (Eq. (8)) can be extended to a weighted global metric, the *Catchment Connectivity Index* (*CCI*; Baldan et al., 2022):

$$CCI = \sum_{i,j=1,i\neq j}^{l} p_{i,j}\frac{w_i w_j}{W^2}. \quad (9)$$

Like RCI, CCI ranges from 0 to 1, where a zero indicates the absence of connectivity and a one indicates maximum stream connectivity. Additionally (like RCI), CCI indices generally describe dispersal constraints of aquatic animals capable of upstream travel (typically fish) in the context of stream habitat fragmentation, and thus use an undirected (non-DAG) framework.

Many CCI variants have been developed and are codified in the *riverconn* **R** package (Baldan et al., 2022). These can be distinguished by the types of weights used, and definitions of $p_{i,j}$. For instance, if weights are based on reach lengths, then Eq. (9) can be viewed as the Dendritic Connectivity Index (Cote et al., 2009; Jaeger et al., 2014), and if reach volumes are used as weights then Eq. (9) is the volume-based river connectivity index (Grill et al., 2014). Other CCI variants include the population connectivity index (Angulo-Rodeles et al., 2021), and the probability of connectivity (Pascual-Hortal and Saura, 2006). Other probabilistic global metrics of connectivity based on an undirected graph framework include path connectivity and the network connectivity metric of Garbin et al. (2019).

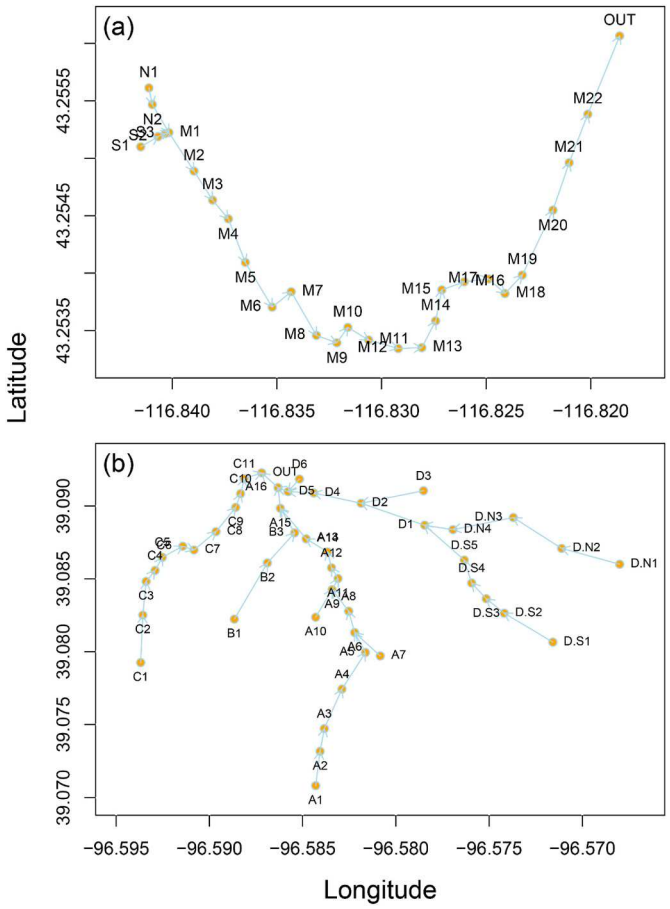


Fig. 2. Spatially explicit DAG representations of (a) the completely wetted Murphy Creek network, and (b) the completely wetted Konza Prairie network. Nodes occur at stream sensor locations. Note (user-controlled) northing exaggeration in panels.

5.3. Bayesian extensions

Bayesian extensions to Bernoulli length and communication distance are possible by viewing the probabilities of stream segment presence at arcs as random variables. This approach is useful because it allows: 1) explicit consideration of the potential variability and uncertainty in designations of probabilities of surface water presence at both local at global scales, and 2) inclusion of both current and prior information concerning those probabilities. A complete statistical background for these approaches is described in Aho et al. (2023). Briefly, given current binomial data consisting of Bernoulli {0,1} stream presence outcomes over n trials at the kth arc, and a beta-distribution prior for the probability of the presence of water at the kth arc, the conjugate posterior beta distribution for the probability of stream surface water presence for the kth arc can be expressed as:

$$\theta_k|\boldsymbol{x}_k \sim BETA\left(\omega \cdot n \cdot \tilde{p}_k + \sum \boldsymbol{x}_k,\ \omega \cdot n\left(1 - \tilde{p}_k\right) + n - \sum \boldsymbol{x}_k\right) \qquad (10)$$

where ω is the weight given to the prior relative to the current data, $\tilde{p}_k$ is the mean of the prior beta distribution, describing prior degrees of belief concerning the presence of surface water at the kth arc, and $\sum \boldsymbol{x}_k$ is the number of binary successes (stream surface water presence outcomes) at the kth arc, over n trials in the current data. The posterior distribution in Eq. (10) is an *inductive* representation of the probability of stream presence that acknowledges potential uncertainty in designation of this probability (due, for instance, to seasonal and year-to-year climatic variations), based on both current and previous information. Under

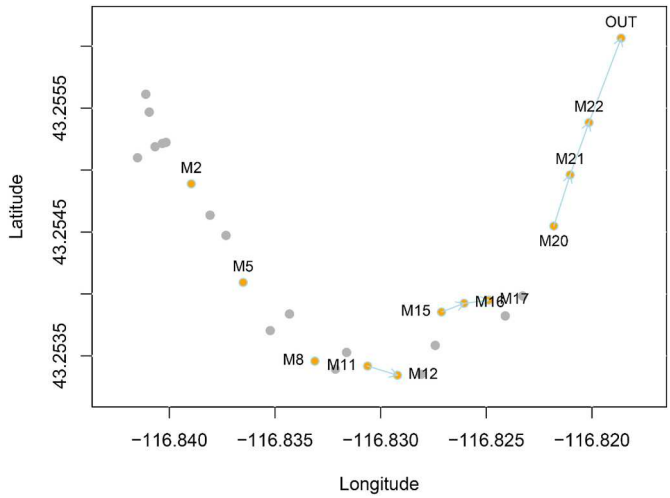


Fig. 3. Wet nodes and deduced wet arcs (blue) are distinguished from dry nodes (gray) at Murphy Creek for the timestamp: 8/9/2019 22:30.

linear transformation, multiplying the kth posterior for the probability of stream presence by the kth stream length will provide the kth posterior for Bernoulli stream length

The posterior distribution for the reciprocal probability of stream presence for the kth arc will follow an inverse beta distribution (see Aho et al., 2023) with the same parameters shown in Eq. (10). Multiplying the kth posterior for the inverse probability of stream presence by the kth stream length will provide the kth posterior for communication distance, providing both global and local estimates for the propensity of stream bottlenecking (Aho et al., 2023).

6. Application of the *streamDAG* package to authentic non-perennial stream networks

As "real-world" applications for the *streamDAG* package, we considered two non-perennial stream networks: Murphy Creek, and a portion of the south fork of Kings Creek in Konza Prairie (hereafter Konza Prairie for brevity). Murphy Creek is a simple network (two sources and a single outlet) within the larger Reynolds Creek experimental watershed in the Owyhee Mountains of southwestern Idaho, USA (43.256° N, 116.817° W). Measures of surface water presence at Murphy Creek were made at 25 nodes at 15-min intervals from March 6, 2019 to 10/3/2019. Surface water presence/absence was determined using Onset HOBO Pendant/Light 64 K Datalogger UA002-64 resistivity sensors and HOBO pressure transducers (see Warix et al., 2021). Bounding nodes were added at two theorized stream source locations and the network sink to encompass the entire length of the network. This resulted in a final Murphy Creek network with 28 nodes and 27 arcs for analysis (Fig. 2a). Konza Prairie is a relatively complex non-perennial stream network in the northern Flint Hills region of Kansas, USA (39.11394° N, 96.61153°W). Our depiction of the Konza Prairie network required 46 nodes and 45 arcs, with nine source nodes and three major reaches leading to the outlet node (Fig. 2b). Several non-perennial stream graphs, including the complete Murphy Creek and Konza networks can be called using the *streamDAG* function `streamDAGs`.

6.1. Spatial plots

Spatial representations of stream DAGs can be obtained from the *streamDAG* function `spatial.plot()` by applying node spatial coordinates to a stream DAG object. (Fig. 2). Stream shapefiles, which may capture stream segment spatial nuances (instead of arc directional arrows), can also be used by `spatial.plot()` with some loss of flexibility.

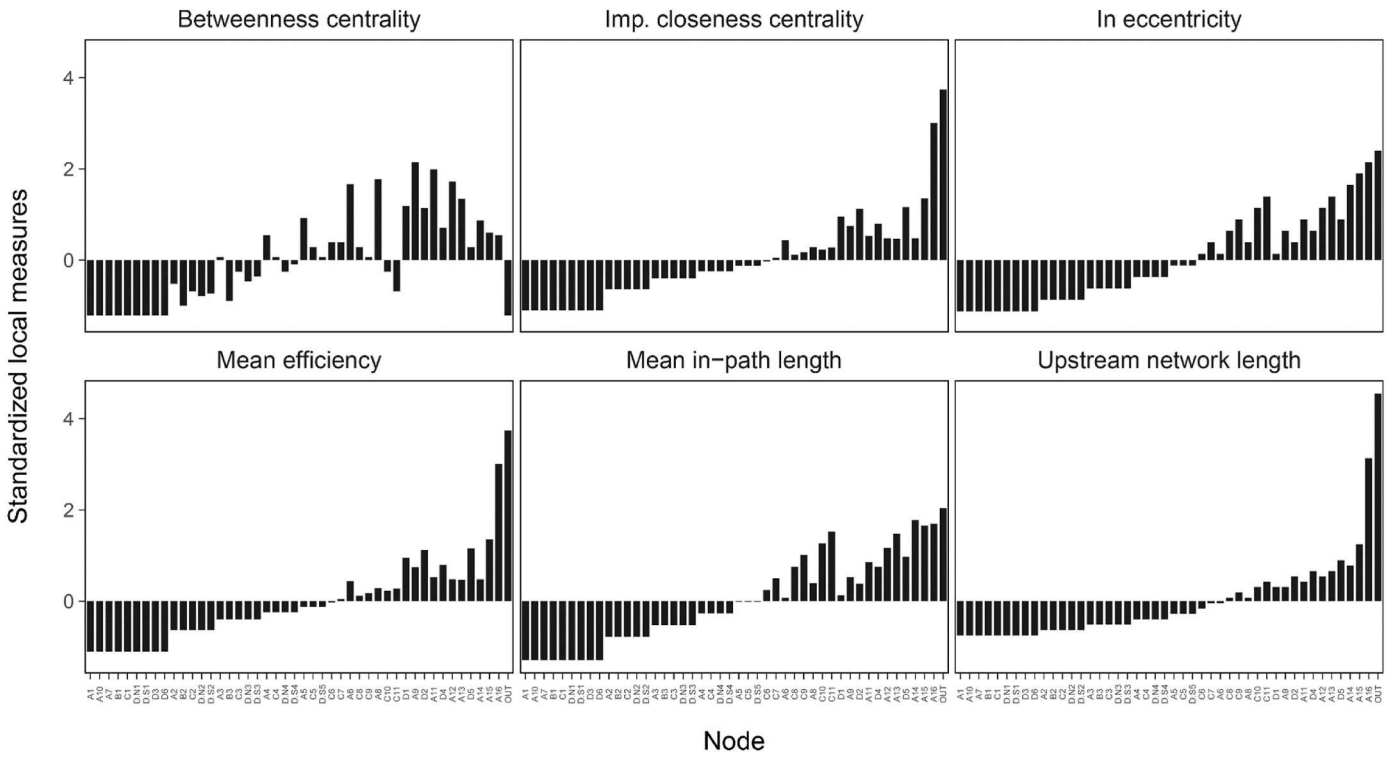


Fig. 4. Local graph-theoretic summaries for the fully wetted Konza Prairie network using `local.summary()`. Metrics are standardized so that each has mean of zero and unit variance. Nodes are organized along the *x*-axis from sources to outlet using the ordering approach for visibilities used by the function `multi.path.visibility()`. Figure code can be found in Supplementary Materials S2.

```
data(mur_coords); data(kon_coords); par(mfrow = c(2,1))
mur <- streamDAGs("mur_full"); konza <- streamDAGs("konza_full")
spatial.plot(mur, mur_coords$long, mur_coords$lat, mur_coords$Object.ID)
spatial.plot(kon, kon_coords$long, kon_coords$lat, kon_coords$Object.ID)
```

6.2. Tracking intermittency

Stream intermittency can be tracked using either node or arc presence/absence data. Below we create a new graph object, `G1`, consisting of the subset of wet nodes at timestamp 8/9/2019 22:30 (time point 650), from the dataframe `mur_node_pres_abs`. We then call `G1` to make a new Murphy Creek graph using `spatial.plot()` (Fig. 3). Note that arcs missing one or more wet bounding nodes are omitted by the algorithm.

```
data(mur_node_pres_abs)
G1 <- delete.nodes.pa(mur, mur_node_pres_abs[650,][,-1])
spatial.plot(G1, mur_coords$long, mur_coords$lat, mur_coords$Object.ID, xlab = "Longitude", ylab = "Latitude", plot.dry = TRUE)
```

6.3. Unweighted DAG measures

A large number of local unweighted DAG metrics can be obtained from the *streamDAG* function `local.summary()` (see Table 1). Fig. 4 summarizes nodal results for the complete Konza Prairie network (Fig. 2b). Along the *x*-axis, nodes are ordered roughly from sources (leftmost nine nodes) to the sink (rightmost node). The importance of nodes at reach convergence points, e.g., A16, and the catchment outlet, OUT, is particularly evident. Local metrics generally indicate an increase in nodal importance as distance to the sink decreases. An exception is betweenness centrality which is highest for nodes in the center of reaches, but lowest for the source and sink nodes. Note that standardized responses from improved closeness centrality and mean efficiency are essentially identical because of their shared reliance on reciprocal distances.

Table 5
Unweighted global summaries of the Murphy Creek and Konza Prairie networks.

	Murphy	Konza
Size	27.00	45.00
Diameter	25.00	14.00
Graph order	28.00	46.00
Sources	2.00	9.00
Mean α-centrality	14.29	7.37
Number of paths to sink	27.00	45.00
Sink path length summary		
$\bar{x}$	13.78	6.51
s^2	55.73	10.56
g_1	−0.11	0.25
g_2	−1.29	−0.66
DAG degree summary		
$\bar{x}$	0.96	0.98
s^2	0.11	0.37
g_1	−0.75	0.01
g_2	7.69	−0.19
Shreve number	2.00	9.00
Strahler number	2.00	3.00
First Zagreb index	28.00	53.00
Second Zagreb index	14.50	30.50
Atom bond connectivity	0.71	5.66
Harary index	40.86	53.89
Global efficiency	0.11	0.05
Assortativity: $r(+,-)$	−0.02	−0.20
Assortativity: $r(-,+)$	0.03	0.06

Table 5 provides a global metric comparative summary for the complete Murphy Creek and complete Konza Prairie networks using the

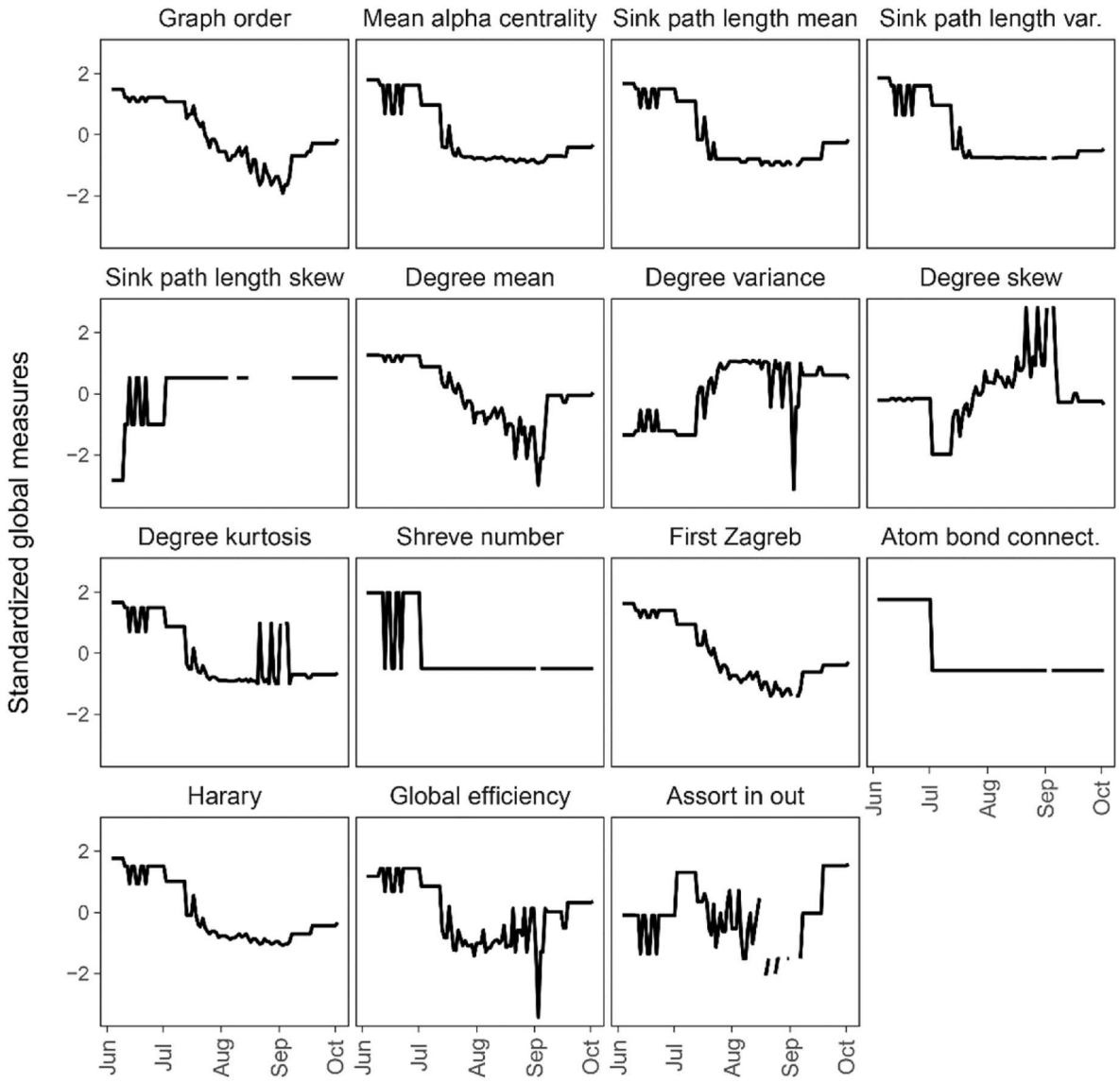


Fig. 5. Global for Murphy Creek based on stream node presence/absence data. As in Fig. 5, metrics are standardized to have a mean of zero and unit variance. Figure code can be found in Supplementary Materials S2.

`global.summary()` function. Specifically:

```
global.summary(streamDAGs("mur_full"), sink = "OUT")
global.summary(streamDAGs("konza_full"), sink = "SFM01_1")
```

Comparison of results illustrates the greater complexity of the Konza network through larger values for all metrics except mean upstream in-path length and mean α-centrality. These latter values largely reflect differences in node definitions, basin size, and density of sensor placement.

It may be informative to track changes in global metrics (and local metrics) over time. Fig. 5 shows a 100-point time series that spans the entire 2019 sampling season at Murphy Creek (Fig. 2a). Over this period, graphs were created to reflect presence or absence of water at Murphy Creek nodes, and global metrics were calculated. Note that higher scores, indicating higher network connectivity, occur for many of the metrics (e.g., graph order mean α-centrality, the mean and the variance of sink path lengths, first Zagreb index) during the spring and a re-wet period during the fall. Exceptions include measures of dispersion (network degree variance), symmetry (sink path length skew and degree skew), and assortativity.

6.4. Weighted DAG measures

As with unweighted metrics, it may be informative to track weighted global (and local) metrics for non-perennial streams over time. In Fig. 6, we calculate ICSL, intact stream length to the node, average alpha-centrality, and the Harary index for Murphy Creek, after defining instream lengths as arc weights. We consider these measures over time, based on the stream node time series data in Fig. 5. All four metrics show dramatic decreases in network connectivity from spring to summer, with a connectivity uptick in the fall due to rewetting.

6.5. Bayesian applications

Bayesian extensions to Bernoulli stream length and communication distance can be facilitated with the use of the *streamDAG* function `beta.posterior()`. As an example, assume that we wish to apply the naive Bayesian prior, $\theta_k \sim BETA(1,1)$, for the probability of stream segment surface water presence at Murphy Creek, to all stream segments. Note that the distribution $BETA(1,1)$ is equivalent to a continuous uniform distribution in 0,1, and will have mean, $E(\theta_k) = 0.5$. Assume further that we wish to give the priors 1/3 of the weight of

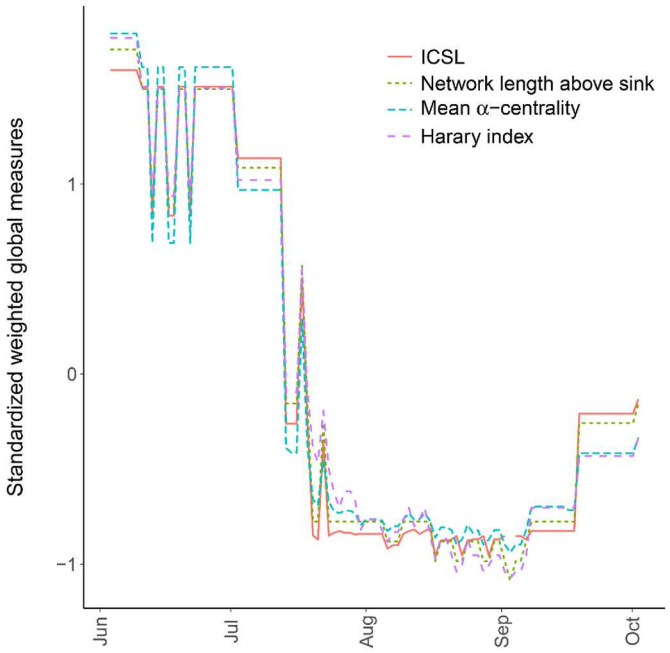


Fig. 6. Global weighted (by arc instream lengths) network connectivity measures for Murphy Creek over time. Code for creating the figure can be found in Supplementary Materials S2.

observed current binomial data outcomes (i.e., stream presence observations over n trials). The `mur_arc_pres_abs` dataframe contains 1000 multivariate Bernoulli datasets for Murphy Creek, one per row. For demonstration purposes we arbitrarily use the first 10 rows of the matrix `mur_arc_pres_abs` as an observed (current) multivariate binomial data point. We have:

```
data <- mur_arc_pres_abs[1:10,]
b <- beta.posterior(p.prior = 0.5, dat = data, length = mur_lengths[,2], w = 1/3)
```

The function `beta.posterior()` returns a list with values for shape parameters for the beta posteriors for the probability of stream presence used to create Fig. 7, and the inverse beta parameters for the reciprocal probability of stream presence (Supplementary Materials S3, Figs. S3–2).

7. Discussion

The spatiotemporal dynamism of non-perennial streams may not be well represented by common metrics of stream network complexity and connectivity, many of which are time invariant. Further, many existing stream metrics do not consider the importance of individual stream locations to stream network functionality and stability. This deficiency is particularly problematic in non-perennial streams because certain stream locations (e.g., flow bottlenecks) may have inordinately large effects on the entire network. These considerations served as primary motivators for the selection and development of tools in the *streamDAG* **R** software package.

Many measurement methodologies are possible if we consider non-perennial streams as directed acyclic graphs. This approach allows standardized graphical and numeric tracking of global stream network characteristics, and consideration of the importance of both local stream components (e.g., arcs and nodes), and global network characteristics, as stream locations dry and the network changes. To explain and justify the inclusion of particular metrics in *streamDAG*, we considered a large number of graph theoretic methods for their potential usefulness in the analysis of non-perennial streams. Notably, in Supplementary Materials S1 we identify methods that are *unlikely* to be useful for this application. We deem the latter contribution helpful given the confusing myriad of graph theoretic methods, many of which have been repeatedly "rediscovered" under different names. We emphasize that *streamDAG* has been developed to consider passive stream network processes that clearly fall under a directed acyclic framework, e.g., streamflow, and stream-borne solute transport (cf. Dodds and Rothman, 2000; Rinaldo et al., 2006). We plan on reviewing potential graph-theoretic metrics for undirected stream phenomena (e.g. bidirectional animal dispersal, hydrologic cycles) in upcoming work.

7.1. Application of the streamDAG package

The practical usefulness of graph theoretic methods for the analysis of "real world" systems has been demonstrated repeatedly. For instance, Urban and Keitt (2001) used graph theoretic analyses to develop conservation approaches for the threatened Mexican Spotted Owl (*Strix occidentalis lucida*), and Saunders et al. (2015) reviewed 42 publications in which graph theory approaches were used to measure spatial connectivity in aquatic ecosystems. In the recent analyses of stream networks, Baldan et al. (2022) identified unifying graph theoretic frameworks underlying a large number of approaches used to successfully describe the dispersal of riverine organisms, and Botter and Durighetto (2020) used a DAG framework and the concept of Bernoulli network length to effectively describe patterns of flow persistence of headwater streams in the southern Alps.

We demonstrated *streamDAG* functions in the analysis of both toy non-perennial stream examples (Fig. 1; Sections 2-5; Tables 1–4) and authentic North American non-perennial streams (Section 6). In these efforts, changes to local importance and global connectivity of networks as a consequence of drying could be clearly tracked using *streamDAG* graph-theoretic algorithms (e.g., Fig. 5). Objective inter-network comparisons are also possible using *streamDAG* algorithms (Table 5), although care should be taken in node designation to facilitate unbiased assessments (see Section 7.3.1 below). In other applications, our group recently found that microbial diversity in the Konza Prairie water column was uncorrelated with conventional hydrological descriptors of stream connectivity based solely on slope and drainage area (e.g., *TWI*), but was strongly positively correlated with several simple nodal graph theoretic metrics available in *streamDAG,* including alpha centrality, improved closeness centrality, path number, path length variance, ineccentricity, and Shreve stream order (Supplemental Materials S3, Figs. S3–3). This result suggests that Konza hydrologic connectivity (measured using DAG perspectives) affects microbial community composition and structure. Additionally, detailed Bayesian summaries for Murphy Creek have been recently completed using *streamDAG* functions for Bernoulli stream length and communication distance (Aho et al., 2023).

7.2. Correlations of graph-theory measures

We observed varying but often strong correspondence in the assessments of local and global metrics in the analyses of both artificial stream graphs (Fig. 1), and the Konza Prairie and Murphy Creek networks (Figs. 2–7). The correlation of local centrality measures (e.g., closeness centrality, degree, eigenvector centrality, betweenness centrality) has been considered previously (Valente et al., 2008; Batool and Niazi, 2014; Li et al., 2015). These papers generally hold that correlations of centrality measures are due to similarities in the formal definitions of indices and, conversely, an absence of correlations between indices is due to divergent conceptualizations of centrality (Schoch et al., 2017). However, inconsistencies in some empirical findings and a re-consideration of graphs with respect to their *neighborhood inclusion preorder* indicate that underlying directed network structures may strongly affect the strength of correlations among local centrality measures (see Schoch et al., 2017). Empirical assessments of the correlation of global graph measures are largely lacking, although relevant ancillary summaries are given in a number of papers including Foo et al. (2021).

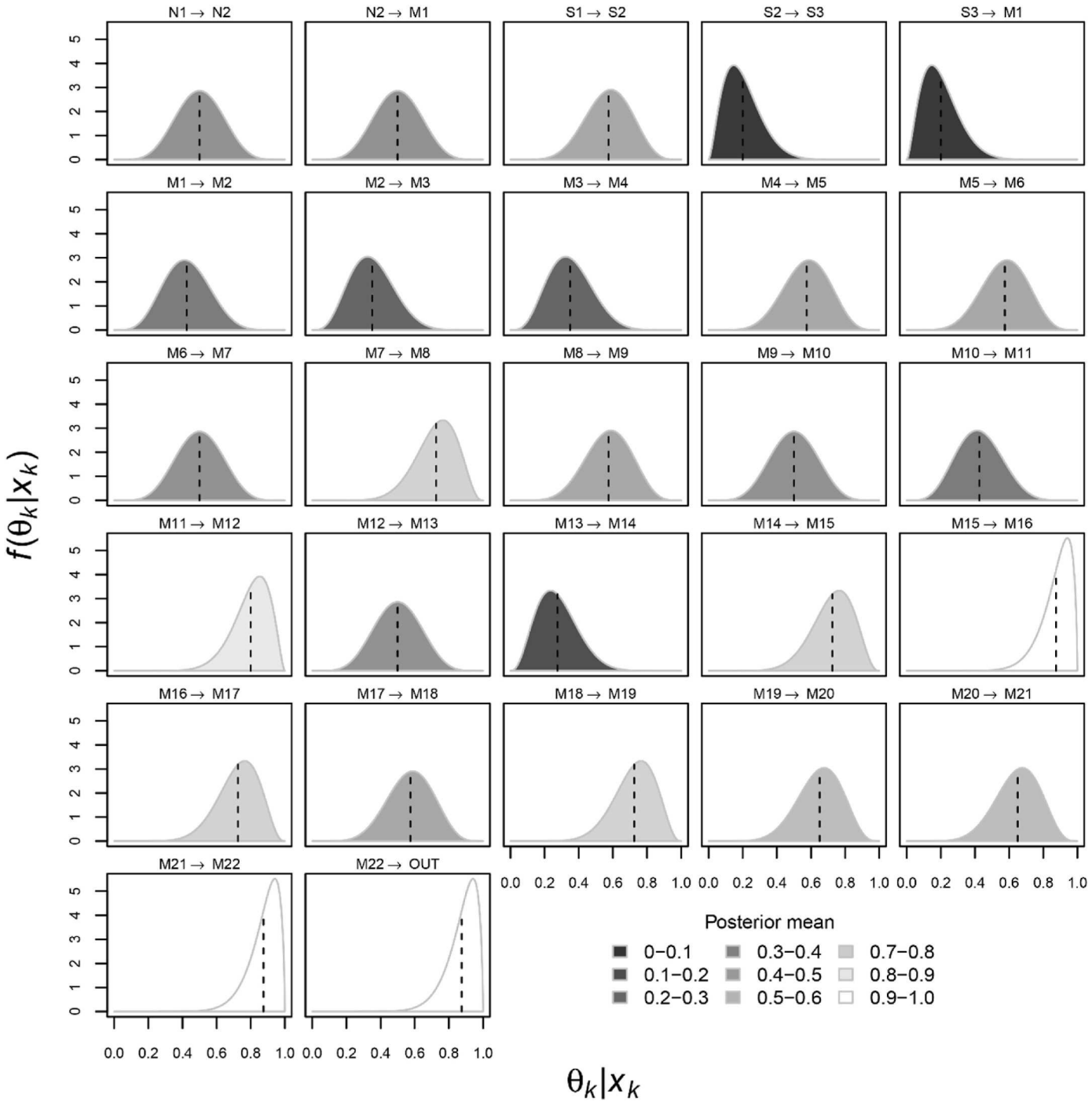


Fig. 7. Graphical summaries of posterior beta distributions representing the probabilities of stream surface water for Murphy Creek stream arcs from 06/01/2019 to 10/01/2019. Arc distributions are colored by their mean values (darker distributions have smaller means). Posterior means are overlain on the distributions with dashed lines. Code for creating the figure can be found in Supplementary Materials S2.

7.3. Uncertainties and extensions

Our work considers surficial stream networks. In principle one could consider both subsurface networks and subsurface to surface hydrologic fluxes (e.g., vertical connectivity). This extension of graph theoretic approaches to subsurface networks may be challenging given fundamental differences between surface water channels and groundwatersheds (Huggins et al., 2022), although see Zuecco et al. (2019). Stream vertical connectivity has received less attention from hydrologists compared to surficial connectivity due to the increased difficulty in obtaining subsurface permeability and flowpath information (Xiao et al., 2021). While potentially useful, models of vertical connectivity, based on subsurface reactive transport algorithms (Steefel et al., 2015), and hillslope models (Hopp and McDonnell, 2009), remain largely limited to time invariant perspectives (Xiao et al., 2021).

7.3.1. Nodal designations

Importantly, many stream DAG nodes and resultant arcs will be user-defined, and discrepancies in their designation criteria will strongly affect stream network topologies. As a result, node locations in stream graphs should be consistent and/or hydrologically meaningful. For instance, nodes could represent approximately equidistant points along

stream paths and/or joins, splits, sinks, sources, or replicates of particular environmental conditions. Our group recently placed nodes within nine non-perennial stream networks across the United States using a stratified sampling/apportionment design (Aho, 2014), based on *a priori* cutoffs of the Topographic Wetness Index (Beven and Kirkby, 1979) derived from digital elevation models.

The effect of biased or otherwise sub-optimal node designations can be moderated by applying reality-driven weights to arcs or nodes. For instance, unweighted graph-theoretic in-path lengths for a node, *v*, can be made arbitrarily large by simply adding more nodes to paths ending in *v*. This undesirable effect, however, can be assuaged (at least with respect to mean path length values) if arcs are weighted by their actual field-measured lengths. Weighted graph approaches also allow the incorporation of both structural (topological) and functional perspectives when describing streams (Baldan et al., 2022). Thus, these measures will be superior to unweighted approaches when node-specific information unrelated to topology is important in hydrological investigations.

8. Conclusions

Many conventional stream network metrics may poorly describe the spatiotemporal dynamism of non-perennial streams. To address this, we considered the applicability of DAG metrics and created an **R** package, *streamDAG*, for their implementation. The *streamDAG* package allows *igraph* codification and modification of stream DAGs using non-perennial stream presence/absence data, and application of a wide variety of DAG-appropriate metrics including local and global measures for both unweighted and weighted graphs. These include Bayesian extensions to the Bernoulli stream length (Botter and Durighetto, 2020) and communication distance weighted metrics (Aho et al., 2023). We applied *streamDAG* functions to both artificial and real-world non-perennial stream DAG networks and found that changes to local site importance and network connectedness, complexity, and nestedness due to drying could be tracked using codified approaches. Inter-network comparisons are also possible although biases will occur if consistent criteria are not used for the designation of nodes. We found the *streamDAG* package to be useful in analyses, and believe that other non-perennial stream researchers are likely to find it useful as well.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data used are contained in the R computational software streamDAG.

Acknowledgements

This work was made possible with grants from the National Science Foundation: RII Track-2 FEC: Aquatic Intermittency Effects on Microbiomes in Streams (AIMS) (NSF award 2019603) and NSF EAR: Active Learning Across Interfaces: Controls on Flow Intermittency and Water Age in Temporary Streams (NSF award1653998). Special thanks to Arya Legg for codifying documentation files for *streamDAG*, and the broader AIMS research group for their commentary and ideas as the *streamDAG* package developed. We also thank two anonymous ENVSOFT reviewers whose comments have greatly improved the quality of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2023.105775.

References

Aho, K.A., 2014. Foundational and Applied Statistics for Biologists Using R. CRC Press.
Aho, K., Ramos, R., Legg, A., Hale, R., 2022. *StreamDAG*: Descriptors and Methods for Stream DAGs. R Package Version 1, pp. 2–7. https://doi.org/10.5281/zenodo.7562599.
Aho, K., Derryberry, D., Godsey, S.E., Ramos, R., Warix, S., Zipper, S., 2023. The communication distance of non-perennial streams. *EarthArXiv*. https://doi.org/10.31223/X5Q367.
Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74 (1), 47.
Ali, G.A., Roy, A.G., 2009. Revisiting hydrologic sampling strategies for an accurate assessment of hydrologic connectivity in humid temperate systems. Geography Compass 3 (1), 350–374.
Ali, G.A., Roy, A.G., 2010. Shopping for hydrologically representative connectivity metrics in a humid temperate forested catchment. Water Resour. Res. 46 (12).
Almende, B.V., Thieurmel, B., et al., 2022. visNetwork: Network Visualization using 'vis.js' Library. R package version 2.1.2. https://CRAN.R-project.org/package=visNetwork.
Angulo-Rodeles, A., Galicia-Paredes, D., Miranda, R., 2021. A simple method to assess the fragmentation of freshwater fish meta-populations: implications for river management and conservation. Ecol. Indicat. 125, 107557.
Anthony, B.M., Marr, A.M., 2021. Directed Zagreb Indices. *Graphs And Combinatorial Optimization: from Theory To Applications: Ctw2020 Proceedings*, pp. 181–193.
Balaban, A.T., 1982. Highly discriminating distance-based topological index. Chem. Phys. Lett. 89 (5), 399–404.
Baldan, D., Cunillera-Montcusí, D., Funk, A., Hein, T., 2022. Riverconn: an R Package to Assess River Network Fragmentation. Environmental Modelling and Software. https://doi.org/10.2139/ssrn.4096555 (verified 11/18/2022.
Bang-Jensen, J., Gutin, G., 2008. Digraghs: Theory, Algorithms and Applications. Springer Monographs in Mathematics, Springer-Verlag London Ltd., London, p. 101.
Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. USA 101 (11), 3747–3752.
Barrett, M., 2023. Ggdag: Analyze and Create Elegant Directed Acyclic Graphs_. R package version 0.2.8. https://CRAN.R-project.org/package=ggdag.
Batool, K., Niazi, M.A., 2014. Towards a methodology for validation of centrality measures in complex networks. PLoS One 9 (4), e90283.
Bavelas, A., 1950. Communication patterns in task-oriented groups. J. Acoust. Soc. Am. 22 (6), 725–730.
Beauchamp, M.A., 1965. An improved index of centrality. Behav. Sci. 10 (2), 161–163.
Bertassello, L.E., Bertuzzo, E., Botter, G., Jawitz, J.W., Aubeneau, A.F., Hoverman, J.T., et al., 2021. Dynamic spatio-temporal patterns of metapopulation occupancy in patchy habitats. R. Soc. Open Sci. 8 (1), 201309.
Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrol. Sci. J. 24 (1), 43–69.
Boldi, P., Vigna, S., 2014. Axioms for centrality. Internet Math. 10 (3–4), 222–262.
Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. 2 (1), 113–120.
Bonacich, P., 1987. Power and centrality: a family of measures. Am. J. Sociol. 92 (5), 1170–1182.
Bonacich, P., Lloyd, P., 2001. Eigenvector-like measures of centrality for asymmetric relations. Soc. Network. 23 (3), 191–201.
Borgatti, S.P., 2005. Centrality and network flow. Soc. Network. 27 (1), 55–71.
Borgatti, S.P., Everett, M.G., 2006. A graph-theoretic perspective on centrality. Soc. Network. 28 (4), 466–484.
Borselli, L., Cassi, P., Torri, D., 2008. Prolegomena to sediment and flow connectivity in the landscape: a GIS and field numerical assessment. Catena 75 (3), 268–277.
Botter, G., Durighetto, N., 2020. The stream length duration curve: a tool for characterizing the time variability of the flowing stream length. Water Resour. Res. 56 (8), e2020WR027282.
Bracken, L., Wainwright, J., Ali, G., Tetzlaff, D., Smith, M., Reaney, S., Roy, A., 2013. Concepts of hydrological connectivity: research approaches, pathways and future agendas. Earth Sci. Rev. 119, 17–34.
Brandes, U., Fleischer, D., 2005. Centrality measures based on current flow. In: Annual Symposium on Theoretical Aspects of Computer Science. Springer, Berlin, Heidelberg, pp. 533–544.
Breitling, L.P., Duan, C., Dragomir, A.D., Luta, G., 2021. Using dagR to identify minimal sufficient adjustment sets and to simulate data based on directed acyclic graphs. Int. J. Epidemiol. 50 (6), 1772–1777.
Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. 30 (1–7), 107–117.
Chen, Y., Wu, L., Zhang, G., Xu, Y.J., Tan, Z., Qiao, S., 2020. Assessment of surface hydrological connectivity in an ungauged multi-lake system with a combined approach using geostatistics and spaceborne SAR observations. Water 12 (10), 2780.
Cote, D., Kehler, D.G., Bourne, C., Wiersma, Y.F., 2009. A new measure of longitudinal connectivity for stream networks. Landsc. Ecol. 24 (1), 101–113.
Cressie, N., Frey, J., Harch, B., Smith, M., 2006. Spatial prediction on a river network. J. Agric. Biol. Environ. Stat. 11, 127–150.
Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695 (5), 1–9.
Creed, J.H., Gerke, T., 2018. ShinyDAG. https://doi.org/10.5281/zenodo.1296477. Initial release (version 0.1.0).

Datry, T., Larned, S.T., Tockner, K., 2014. Intermittent rivers: a challenge for freshwater ecology. Bioscience 64 (3), 229–235.
Deng, H., Yang, J., Tang, Z., Yang, J., You, M., 2022. On the vertex-degree based invariants of digraphs. Discrete Math. Lett. 9, 2–9.
Dekker, A., 2005. Conceptual distance in social network analysis. J. Soc. Struct. 6 (3), 31.
Dodds, P.S., Rothman, D.H., 2000. Geometry of river networks. I. Scaling, fluctuations, and deviations. Phys. Rev. 63 (1), 016115.
Duarte, G., Segurado, P., Oliveira, T., Haidvogl, G., Pont, D., Ferreira, M.T., Branco, P., 2019. The river network toolkit–RivTool. Ecography 42 (3), 549–557.
Durighetto, N., Botter, G., 2022. On the relation between active network length and catchment discharge. Geophys. Res. Lett. 49 (14), e2022GL099500.
Ek, B., VerSchneider, C., Narayan, D.A., 2015. Global efficiency of graphs. AKCE International Journal of Graphs and Combinatorics 12 (1), 1–13.
Erdős, P., Rényi, A., 1959. Some further statistical properties of the digits in Cantor's series. Acta Math. Acad. Sci. Hungar. 10, 21–29.
Estrada, E., Torres, L., Rodriguez, L., Gutman, I., 1998. An Atom-Bond Connectivity Index: Modelling the Enthalpy of Formation of Alkanes. NISCAIR-CSIR, India.
Favaron, O., Mahéo, M., Saclé, J.-F., 1993. Some eigenvalue properties in graphs (conjectures of graffiti—II). Discrete Math. 111 (1–3), 197–220.
Foo, H., Thalamuthu, A., Jiang, J., Koch, F., Mather, K.A., Wen, W., Sachdev, P.S., 2021. Age-and sex-related topological organization of human brain functional networks and their relationship to cognition. Front. Aging Neurosci. 897.
Fortuna, M.A., Gómez-Rodríguez, C., Bascompte, J., 2006. Spatial network structure and amphibian persistence in stochastic environments. Proceedings of the Royal Society B: Biological Sciences 273 (1592), 1429–1434.
Foster, J.G., Foster, D.V., Grassberger, P., Paczuski, M., 2010. Edge direction and the structure of networks. Proc. Natl. Acad. Sci. USA 107 (24), 10815–10820.
Freeman, L.C., 1977. A Set of Measures of Centrality Based on Betweenness. Sociometry, pp. 35–41.
Freeman, M.C., Pringle, C.M., Jackson, C.R., 2007. Hydrologic connectivity and the contribution of stream headwaters to ecological integrity at regional scales 1. JAWRA Journal of the American Water Resources Association 43 (1), 5–14.
Garbin, S., Celegon, E.A., Fanton, P., Botter, G., 2019. Hydrological controls on river network connectivity. R. Soc. Open Sci. 6 (2), 181428.
Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA 99 (12), 7821–7826.
Godsey, S.E., Kirchner, J.W., 2014. Dynamic, discontinuous stream networks: hydrologically driven variations in active drainage density, flowing channels and stream order. Hydrol. Process. 28 (23), 5791–5803.
GRASS Development Team, 2022. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation, Version 8.2. https://grass.osgeo.org.
Grill, G., Dallaire, C.O., Chouinard, E.F., Sindorf, N., Lehner, B., 2014. Development of new indicators to evaluate river fragmentation and flow regulation at large scales: a case study for the Mekong River Basin. Ecol. Indicat. 45, 148–159.
Gutman, I., Ruscić, B., Trinajstić, N., Wilcox Jr., C.F., 1975. Graph theory and molecular orbitals. XII. Acyclic polyenes. J. Chem. Phys. 62 (9), 3399–3405.
Hopp, L., McDonnell, J.J., 2009. Connectivity at the hillslope scale: identifying interactions between storm size, bedrock permeability, slope angle and soil depth. J. Hydrol. 376 (3–4), 378–391.
Huggins, X., Gleeson, T., Serrano, D., Zipper, S., Jehn, F., Rohde, M.M., Abell, R., Vigerstol, K., Hartmann, A., 2022. Groundwatersheds of protected areas reveal globally overlooked risks and opportunities. Down Earth. https://doi.org/10.31223/X59354 (verified 2/9/2023.
Iannone, R., 2022. DiagrammeR: Graph/Network Visualization. R Package version 1.0.9. https://CRAN.R-project.org/package=DiagrammeR.
Igraph, 2022. The Network Analysis Package. https://igraph.org/ (verified 11/18/2022).
Jaeger, K.L., Olden, J.D., Pelland, N.A., 2014. Climate change poised to threaten hydrologic connectivity and endemic fishes in dryland streams. Proc. Natl. Acad. Sci. USA 111 (38), 13894–13899.
Jasiewicz, J., Metz, M., 2011. A new GRASS GIS toolkit for Hortonian analysis of drainage networks. Comput. Geosci. 37 (8), 1162–1173.
Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Wondzell, S.M., Bencala, K.E., Marshall, L.A., 2009. Hydrologic connectivity between landscapes and streams: transferring reach-and plot-scale understanding to the catchment scale. Water Resour. Res. 45 (4).
Jia, X., Zwart, J., Sadler, J., Appling, A., Oliver, S., Markstrom, S., et al., 2021. Physics-guided recurrent graph model for predicting flow and temperature in river networks. In: Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics, pp. 612–620.
Jin, Y., Peng, R., Shi, J., 2019. Population dynamics in river networks. J. Nonlinear Sci. 29, 2501–2545.
Karim, F., Dutta, D., Marvanek, S., Petheram, C., Ticehurst, C., Lerat, J., et al., 2015. Assessing the impacts of climate change and dams on floodplain inundation and wetland connectivity in the wet–dry tropics of northern Australia. J. Hydrol. 522, 80–94.
Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18 (1), 39–43.
Kleinberg, J.M., 1999. Hubs, authorities, and communities. ACM Comput. Surv. 31 (4es), 5–es.
Knudby, C., Carrera, J., 2005. On the relationship between indicators of geostatistical, flow and transport connectivity. Adv. Water Resour. 28 (4), 405–421.
Krabbenhoft, C.A., Allen, G.H., Lin, P., Godsey, S.E., Allen, D.C., Burrows, R.M., DelVecchia, A.G., Fritz, K.M., Shanafield, M., Burgin, A.J., Zimmer, M.A., Datry, T., Dodds, W.K., Jones, C.N., Mims, M.C., Franklin, C., Hammond, J.C., Zipper, S., Ward, A.S., Costigan, K.H., Beck, H.E., Olden, J.D., 2022. Assessing placement bias of the global river gauge network. Nat. Sustain. 1 https://doi.org/10.1038/s41893-022-00873-0. –7.
Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuno, J.C., 2008. From time series to complex networks: the visibility graph. Proc. Natl. Acad. Sci. USA 105 (13), 4972–4975.
Lacasa, L., Toral, R., 2010. Description of stochastic and chaotic series using visibility graphs. Phys. Rev. 82 (3), 036120.
Lane, S.N., Reaney, S.M., et al., 2009. Representation of landscape hydrological connectivity using a topographically driven surface flow index. Water Resour. Res. 45 (8).
Larsen, L.G., Choi, J., Nungesser, M.K., Harvey, J.W., 2012. Directional connectivity in hydrology and ecology. Ecol. Appl. 22 (8), 2204–2220.
Latora, V., Marchiori, M., 2001. Efficient behavior of small-world networks. Phys. Rev. Lett. 87 (19), 198701.
Li, C., Li, Q., Van Mieghem, P., Stanley, H.E., Wang, H., 2015. Correlation between centrality metrics and their application to the opinion model. Eur. Phys. J. B 88 (3), 1–13.
Li, L., Alderson, D., Doyle, J.C., Willinger, W., 2005. Towards a theory of scale-free graphs: definition, properties, and implications. Internet Math. 2 (4), 431–523.
Lin, N., 1976. Foundations of Social Research. McGraw-Hill, New York.
Liu, Y., Hou, G., Huang, F., Qin, H., Wang, B., Yi, L., 2022. Directed graph deep neural network for multi-step daily streamflow forecasting. J. Hydrol., 127515
Luque, B., Lacasa, L., Ballesteros, F., Luque, J., 2009. Horizontal visibility graphs: exact results for random time series. Phys. Rev. 80 (4), 046103.
Maidment, D.R., 1996. GIS and hydrologic modeling-an assessment of progress. In: Third International Conference on GIS and Environmental Modeling, Santa Fe, New Mexico. https://www.caee.utexas.edu/prof/maidment/gishydro/meetings/santafe/santafe.htm (Verified 2/9/2023).
Maidment, D.R., 2002. Arc Hydro: GIS for Water Resources. ESRI, Inc.
Messager, M.L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., et al., 2021. Global prevalence of non-perennial rivers and streams. Nature 594 (7863), 391–397.
Newman, M.E., 2002. Assortative mixing in networks. Phys. Rev. Lett. 89 (20), 208701.
Newman, M.E., 2005. A measure of betweenness centrality based on random walks. Soc. Network. 27, 39–54.
Newman, M.E., 2006. Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. 74 (3), 036104.
Newman, M.E., 2018. Networks, *second ed.* Oxford University Press.
Opsahl, T., Agneessens, F., Skvoretz, J., 2010. Node centrality in weighted networks: generalizing degree and shortest paths. Soc. Network. 32 (3), 245–251.
Pardo-Igúzquiza, E., Dowd, P.A., 2003. CONNEC3D: a computer program for connectivity analysis of 3D random set models. Comput. Geosci. 29 (6), 775–785.
Pascual-Hortal, L., Saura, S., 2006. Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation. Landsc. Ecol. 21 (7), 959–967.
Phillips, J.D., Schwanghart, W., Heckmann, T., 2015. Graph theory in the geosciences. Earth Sci. Rev. 143, 147–160.
Plavšić, D., Nikolić, S., Trinajstić, N., Mihalić, Z., 1993. On the Harary index for the characterization of chemical graphs. J. Math. Chem. 12 (1), 235–250.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/ (verified 11/18/2022).
Randić, M., 1975. Characterization of molecular branching. J. Am. Chem. Soc. 97 (23), 6609–6615.
Randić, M., 1993. Novel molecular descriptor for structure—property studies. Chem. Phys. Lett. 211 (4–5), 478–483.
Rochat, Y., 2009. Closeness Centrality Extended to Unconnected Graphs: the Harmonic Centrality Index. *Applications Of Social Network Analysis*. ASNA.
Rinaldo, A., Banavar, J.R., Maritan, A., 2006. Trees, networks, and hydrology. Water Resour. Res. 42 (6).
Santos-Fernandez, E., Ver Hoef, J.M., McGree, J.M., Isaak, D.J., Mengersen, K., Peterson, E.E., 2022. SSNbayes: an R Package for Bayesian Spatio-Temporal Modelling on Stream Networks. https://doi.org/10.48550/arXiv.2202.07166 (verified 11/18/2022.
Sarhad, J., Carlson, R., Anderson, K.E., 2014. Population persistence in river networks. J. Math. Biol. 69, 401–448.
Sarker, S., Veremyev, A., Boginski, V., Singh, A., 2019. Critical nodes in river networks. Sci. Rep. 9 (1), 1–11.
Saunders, M.I., Brown, C.J., Foley, M.M., Febria, C.M., Albright, R., Mehling, M.G., et al., 2015. Human impacts on connectivity in marine and freshwater ecosystems assessed using graph theory: a review. Mar. Freshw. Res. 67 (3), 277–290.
Schoch, D., Brandes, U., 2016. Re-conceptualizing centrality in social networks. Eur. J. Appl. Math. 27 (6), 971–985.
Schoch, D., Valente, T.W., Brandes, U., 2017. Correlations among centrality indices and a class of uniquely ranked graphs. Soc. Network. 50, 46–54.
Serinaldi, F., Kilsby, C.G., 2016. Irreversibility and complex network behavior of stream flow fluctuations. Phys. Stat. Mech. Appl. 450, 585–600.
Shreve, R.L., 1966. Statistical law of stream numbers. J. Geol. 74 (1), 17–37.
Skøien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging-geostatistics on stream networks. Hydrol. Earth Syst. Sci. 10 (2), 277–287.
Skøien, J.O., Laaha, G., Koffler, D., Blöschl, G., Pebesma, E., Parajka, J.D., Viglione, A., 2012. Rtop - an R package for interpolation along the stream network. In: Abstracts and Programme. *Geophysical Research Abstracts*, p. 1. http://hdl.handle.net/20.500.12708/60971.

Sørensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol. Earth Syst. Sci. 10 (1), 101–112.
Steefel, C.I., Appelo, C.A.J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., et al., 2015. Reactive transport codes for subsurface environmental simulation. Comput. Geosci. 19 (3), 445–478.
Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Eos, Transactions American Geophysical Union 38 (6), 913–920.
Terui, A., Kim, S., Dolph, C.L., Kadoya, T., Miyazaki, Y., 2021. Emergent dual scaling of riverine biodiversity. Proc. Natl. Acad. Sci. USA 118 (47), e2105574118.
Textor, J., Van der Zander, B., Gilthorpe, M.S., Liśkiewicz, M., Ellison, G.T., 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int. J. Epidemiol. 45 (6), 1887–1894.
Trigg, M.A., Michaelides, K., Neal, J.C., Bates, P.D., 2013. Surface water connectivity dynamics of a large-scale extreme flood. J. Hydrol. 505, 138–149.
Tyers, M., 2017. riverdist: river network distance computation and applications. R package version, 0.15. 0.
Urban, D., Keitt, T., 2001. Landscape connectivity: a graph-theoretic perspective. Ecology 82 (5), 1205–1218.
Valente, T.W., Coronges, K., Lakon, C., Costenbader, E., 2008. How correlated are network centrality measures? Connections 28 (1), 16.
Ver Hoef, J.M., Peterson, E., Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environ. Ecol. Stat. 13 (4), 449–464.
Ver Hoef, J., Peterson, E., Clifford, D., Shah, R., 2014. SSN: an R package for spatial statistical modeling on stream networks. J. Stat. Software 56, 1–45.
Vukičević, D., Furtula, B., 2009. Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges. J. Math. Chem. 46 (4), 1369–1376.
Warix, S.R., Godsey, S.E., Lohse, K.A., Hale, R.L., 2021. Influence of groundwater and topography on stream drying in semi-arid headwater streams. Hydrol. Process. 35 (5), e14185.
Western, A.W., Blöschl, G., Grayson, R.B., 2001. Toward capturing hydrologically significant connectivity in spatial patterns. Water Resour. Res. 37 (1), 83–97.
Wiener, H., 1947. Structural determination of paraffin boiling points. J. Am. Chem. Soc. 69 (1), 17–20.
Wright, S., 1934. The method of path coefficients. Ann. Math. Stat. 5 (3), 161–215.
Xiao, D., Brantley, S.L., Li, L., 2021. Vertical connectivity regulates water transit time and chemical weathering at the hillslope scale. Water Resour. Res. 57 (8), e2020WR029207.
Zhong, L., 2012. The harmonic index for graphs. Appl. Math. Lett. 25 (3), 561–566.
Zhou, B., Trinajstić, N., 2009. On a novel connectivity index. J. Math. Chem. 46 (4), 1252–1270.
Zipper, S.C., Hammond, J.C., Shanafield, M., Zimmer, M., Datry, T., Jones, C.N., , et al. Allen, D.C., 2021. Pervasive changes in stream intermittency across the United States. Environ. Res. Lett. 16 (8), 084033.
Zipper, S., 2020. *streamDepletr*: Estimate Streamflow Depletion Due to Groundwater Pumping. R package version 0.1.1. https://CRAN.R-project.org/package=streamDepletr (verified 11/18/2022).
Zuecco, G., Rinderer, M., Penna, D., Borga, M., Van Meerveld, H.J., 2019. Quantification of subsurface hydrologic connectivity in four headwater catchments using graph theory. Sci. Total Environ. 646, 1265–1280.