

DATA ARTICLE

Lake networks and connectivity metrics for the conterminous U.S. (LAGOS-US NETWORKS v1)

Katelyn B. S. King , * Qi Wang, Lauren K. Rodriguez, Kendra S. Cheruvilil 
Michigan State University, East Lansing, Michigan

Scientific Significance Statement

Identifying lake networks and quantifying the degree of connectivity among lakes (e.g., the stream course distance between lakes) by incorporating streams can help scientists better understand and predict the movement of abiotic materials and biota within networks. LAGOS-US NETWORKS v1 fills the need for accessible and comprehensive connectivity metrics and tools at the individual lake and network scales across the conterminous United States. It includes surface connectivity metrics for a total of 898 networks comprised of 86,511 lakes ≥ 1 ha in surface area.

Abstract

Identifying lake networks and knowing the degree of surface-water connectivity among lakes can help scientists better understand and predict the movement of abiotic materials and biota within networks. Quantifying broad-scale networks that include lake and stream connections is difficult computationally. Starting from the medium resolution National Hydrography Dataset's lakes, streams, and rivers, we applied a graph theory approach to identify lake networks, a set of lakes connected by streams both upstream and downstream. The

*Correspondence: kingka21@msu.edu

Associate editor: Jordan Read

Author Contribution Statement: KBSK created the vision behind the database design and metrics, led the team in data creation, documentation, and validation processes, and drafted text and figures. QW wrote the code and created the networks and metrics. LKR created the naming conventions for the module, wrote the data source and dictionary tables, drafted figures, assisted in metric validation, and drafted text. KSC conceived of the data module and created the research team, provided the structure for the data and documentation, and mentored KBSK and LKR. All coauthors collectively framed project ideas and scope and reviewed the entire manuscript.

Data Availability Statement: Data are available in the Environmental Data Initiative repository at <https://doi.org/10.6073/pasta/98c9f11df55958065985c3e84a4fe995>.

Code availability: Wang, Q., and K. King. 2020. Code for LAGOS-US NETWORKS v1.0 (Version v1.0.0) *Zenodo*. doi:10.5281/zenodo.4383172

Measurement(s): Lake and network surface water connectivity.

Technology Type(s): Computer Science Approach.

Temporal range: 2019–2020.

Frequency or sampling interval: Single snapshot; National Hydrography Dataset medium resolution (NHDPlusV2) was downloaded on Aug 5, 2019 and is built at the 1:100,000 scale.

Spatial scale: National; conterminous United States.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

LAGOS-US NETWORKS v1 module contains four data tables, one of which includes derived surface-water connectivity metrics for lakes ($n = 86,511$ lakes ≥ 1 ha in surface area) and networks ($n = 898$) within the conterminous United States, including dams. The NETWORKS module also includes a flow table as well as a bidirectional and a unidirectional distance table that provide the stream course distances between every connected lake. Finally, this module includes a detailed User Guide.

Freshwater network structure is an important area of research for aquatic ecologists. Knowing the number of and distance to upstream and downstream lakes, the position of a lake in a network, as well as the complexity of lake networks can help scientists better understand and predict the movement of materials and biota within networks. Studies have shown that surface-water connectivity affects lake and stream characteristics such as water chemistry (Wollheim et al. 2008; Sadro et al. 2012; Soranno et al. 2015; Schmadel et al. 2018) and biotic diversity (Olden et al. 2001; Beisner et al. 2006; King et al. 2021a). Research also shows that incorporating both streams and lakes into measures of connectivity gives a more accurate representation of nutrient processing and biotic movements (Jones 2010) than using only one freshwater type (lakes or streams).

One way to characterize freshwater surface connectivity is to create metrics for surface-water networks or a series of connected lakes and stream reaches. These metrics can incorporate the number of and distance to surface-water connections as well as the waterbody position within a network (i.e., landscape position). For example, Olden et al. (2001) investigated how a suite of connectivity metrics such as upstream and downstream watercourse distances between lakes, watercourse distance through an intermediate lake, and stream gradient corresponded to fish community composition. They found that different connectivity metrics were important for different lakes. Popular stream network position metrics like stream Strahler order (Strahler 1957) and link magnitude (Shreve 1967) have been used to capture the spatial arrangement of a stream reach within a river network. Similarly, lake position within a network has been characterized with lake network number (LNN) and lake order (LO), a lake's position in a lake chain and the Strahler order of the outflowing stream, respectively (Kling et al. 2000; Riera et al. 2000; Martin and Soranno 2006). The position of a lake in the network has been shown to be correlated with both abiotic (Kling et al. 2000) and biotic (Kratz et al. 1997) properties. However, connectivity metrics that describe surface-water network structure by incorporating both streams and lakes are needed to better understand the influence of connectivity (or isolation) on biotic and abiotic lake properties.

The best approach to quantify surface-water networks depends on both the research question and focus of the study (e.g., biota vs. nutrients, streams vs. lakes) as well as the spatial scale of interest. For example, when working at broad scales (regions to continents) and including both streams and lakes, it is difficult to balance accurate estimates of surface-water

connectivity and computational challenges. Graph theory approaches, which model pairwise relationships between nodes (lakes or streams) connected to each other by edges (streams), provide ways to overcome computational challenges because they have minimal data requirements while still providing accurate estimates of connections between waterbodies (Calabrese and Fagan 2004). However, because these metrics can be computationally difficult, studies that have applied graph theory to lakes are often restricted to a few watersheds (Bishop-Taylor et al. 2015; Saunders et al. 2016).

Our research fills the need for accessible and comprehensive lake networks and connectivity metrics at the national scale. A recent study using the National Hydrography Dataset's (NHD) high-resolution lakes (> 0.5 ha) and medium resolution permanent rivers and streams classified river networks into four types based on surface connections across the conterminous United States (U.S.; Gardner et al. 2019). This study demonstrated how lake/reservoir abundance and size scale with stream order and provided a first step in incorporating lakes into river networks at the national scale (Gardner et al. 2019). Our research complements their study by focusing on lake networks and making the data and code publicly accessible for further research and applications.

This data paper presents the LAGOS-US NETWORKS v1 data module that identifies a total of 898 networks that include 86,511 lakes ≥ 1 ha in surface area (more detailed description of methods can be found in the User Guide). The number of lakes in a network ranges from 2 to 32,811 lakes, the largest network being the Mississippi River basin (Fig. 1). NETWORKS was created using a graph theory framework to generate lake networks for the conterminous U.S., where lakes and streams were the nodes and connections between them were the edges (sensu Urban and Keitt 2001; Eros et al. 2012). We defined lake networks as a set of lakes connected by ephemeral or permanent streams, regardless of the directionality of those connections (e.g., upstream, downstream, or both) and we excluded connections through the Great Lakes, oceans, and estuaries. The NETWORKS module includes all lakes that are connected to other lakes (i.e., no isolated lakes or lakes only connected to streams are included), which is about 18% of all lakes ≥ 1 ha in surface area in the study extent (Smith et al. 2021; Cheruvil et al. In press). This proportion is comparable to similar studies that found 33% (Hill et al. 2018) and 15% (Gardner et al. 2019) of NHD lakes to be in-network.

From these networks, we derived a suite of surface connectivity metrics, including metrics for connections among lakes

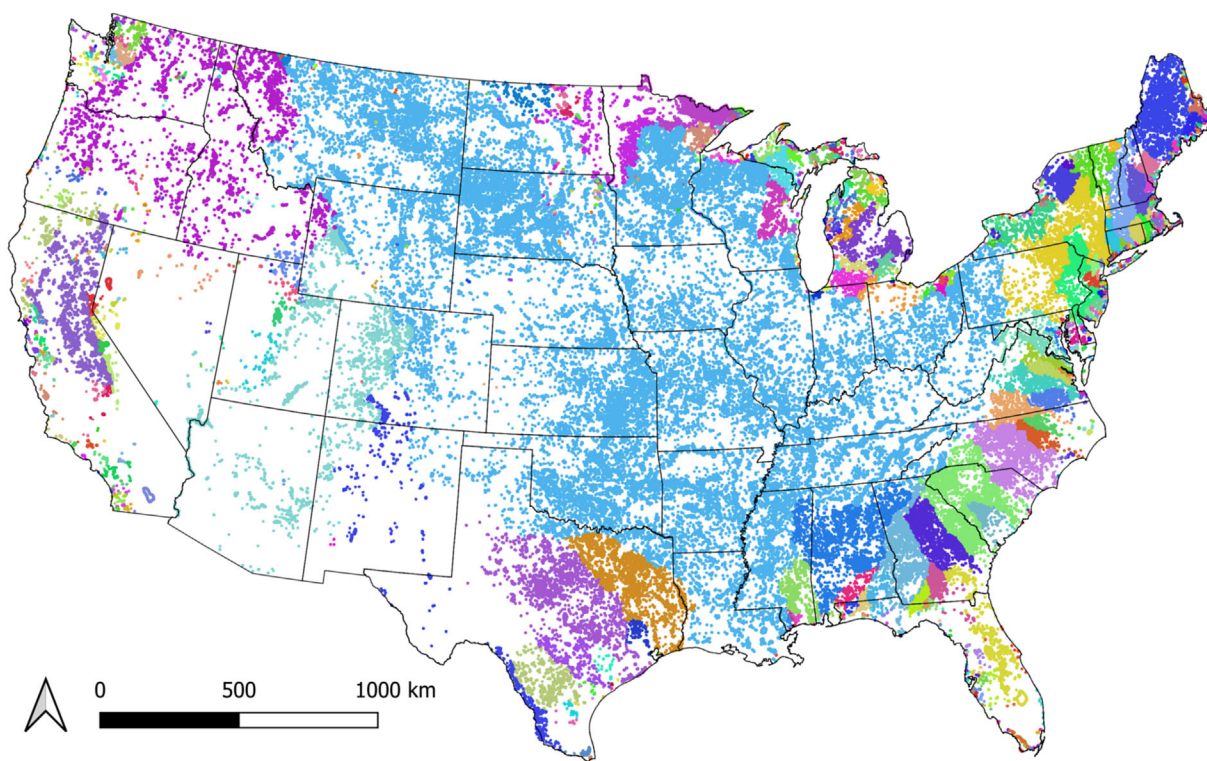


Fig. 1. Lake networks. Lakes ($n = 86,511$) in the LAGOS-US NETWORKS module, colored according to their network membership ($n = 898$ networks). NETWORKS includes lakes > 1 ha in surface area that are connected to other lakes (i.e., no isolated lakes or lakes only connected to streams are included) in the conterminous U.S.

(both upstream and downstream), dam metrics, and network position (spatial orientation of the lake within its network) for connected lakes. We also included upstream and downstream distances (bidirectional) and just downstream distance (unidirectional) between every pair of connected lakes and a flow table that describes the flow path direction (e.g., FROM and TO) between two flowlines (i.e., streams and artificial flowlines through lakes) that was used to create the networks and information in the other three tables. All data in the NETWORKS module can be linked to individual lakes in the LAGOS-US database platform via “lagoslakeid” (Smith et al. 2021; Cheruvilil et al. In press) or linked to the medium resolution NHDplusV2 via “nhdplusv2_comid” (U.S. Geological Survey 2019). Finally, we include a detailed User Guide for NETWORKS that provides more information on methods for this module.

NETWORKS is unique because it identifies lake networks for the conterminous U.S. that include lakes located on different tributaries that are connected through a downstream confluence and provides a suite of connectivity metrics at the individual lake and network scale. Thus, the NETWORKS module can be used in conjunction with other abiotic and biotic datasets to further ecological prediction, such as how nutrients or contaminants move through a network, changes in invasive species distributions, or how biota might move up or downstream in response to climate change (for further

discussion of uses, see section “Data Use and Recommendations for Reuse”). These networks will help advance our understanding of how surface-water connections and network position affect abiotic and biotic properties of lakes at regional to continental scales.

Data description

Overview of data sources

The LAGOS-US NETWORKS module was created using a variety of existing datasets. The lake networks were derived from the lake and stream flow tables of the medium resolution U.S. National Hydrography Dataset (NHDplusV2) downloaded 05 August 2019 (U.S. Geological Survey 2019). NHDplusV2 is a national geospatial surface-water dataset that integrates information from the NHD, the National Elevation Dataset, and the Watershed Boundary Dataset at a 1 : 100,000-scale. Lakes were assigned both their “nhdplusv2_comid,” which are unique identifiers for lakes from the NHDplusV2 dataset and their “lagoslakeid,” which are unique identifiers from LAGOS-US LOCUS v1 data module (Smith et al. 2021; Cheruvilil et al. In press). LAGOS-US LOCUS includes lakes and reservoirs ≥ 1 ha from the high-resolution NHD.

In order to include potential barriers to connectivity, we spatially joined the NHDplusV2 lakes and streams to dams

from a variety of data sources. The National Anthropogenic Barrier Dataset (NABD) (Ostroff et al. 2013) is a dataset of large, anthropogenic barriers that were originally spatially linked to the NHDPlusV1 data product to facilitate analyses based on the NHD and National Inventory of Dams (NID 2015). However, we used a modified NABD that was augmented by Cooper et al. (2017) with 170 additional dams from the USFWS Fish Passage Decision Support Tool and that included dam removals since the NABD was published as listed in the 2018 American Rivers dam removal database (American Rivers 2019). This modified NABD dataset was used to establish the population of dams ($n = 49,525$) that reside on streams or lakes and calculate dam metrics for all lakes and networks within the LAGOS-US NETWORKS module (Fig. 2). The NETWORKS module includes a source table that can be linked to the data tables. See the User Guide that accompanies these data for additional details (King et al. 2021b).

Overview of data tables and variables

The NETWORKS module contains two metadata tables, four data tables, and a detailed User Guide (Fig. 3). The metadata tables are (1) a data dictionary that provides a definition for each variable name or column of every table in the module and includes important information such as units, and (2) a source table that includes a description of the data sources used to create NETWORKS. The four data tables contain the key variables and include (1) a lake

connectivity metrics table (*nets_networkmetrics_medres*) that has lake identifier information, upstream and downstream connectivity metrics, upstream and downstream dam metrics, network position, and network metrics (Table 1), (b, c) two distance tables (*nets_uninetworkdistance_medres*, *nets_binetworkdistance_medres*) that include lake identifier information as well as upstream and downstream distances between pairs of connected lakes using either a unidirectional graph (Table 2) or a bidirectional graph (Table 3), and (d) the modified flow table (*nets_flow_medres*) with NHDplusV2 common identifiers for NHDFlowlines that describes the flow path direction between two flowlines (e.g., TO and FROM) and that was used to create the networks and metrics (Table 4). See the User Guide that accompanies these data for additional details (King et al. 2021b).

Figures 4–6 highlight some variables from the *nets_networkmetrics_medres* table. For example, we found that the majority of lakes are near to each other along the network (nearest bidirectional lake median 4.40 km; Table 1); however, the nearest distance to a downstream lake can be up to 200 km, with lakes in the Mississippi network even further than that (Fig. 4a). Similarly, many lakes have zero dams upstream or downstream (median 0.00; Table 1), but in the Mississippi River basin some have > 10 dams downstream (Fig. 4b). The majority of U.S. lakes have a low LNN (a lake's position in a lake chain; Martin and Soranno 2006), indicating a high amount of network branching rather than long,

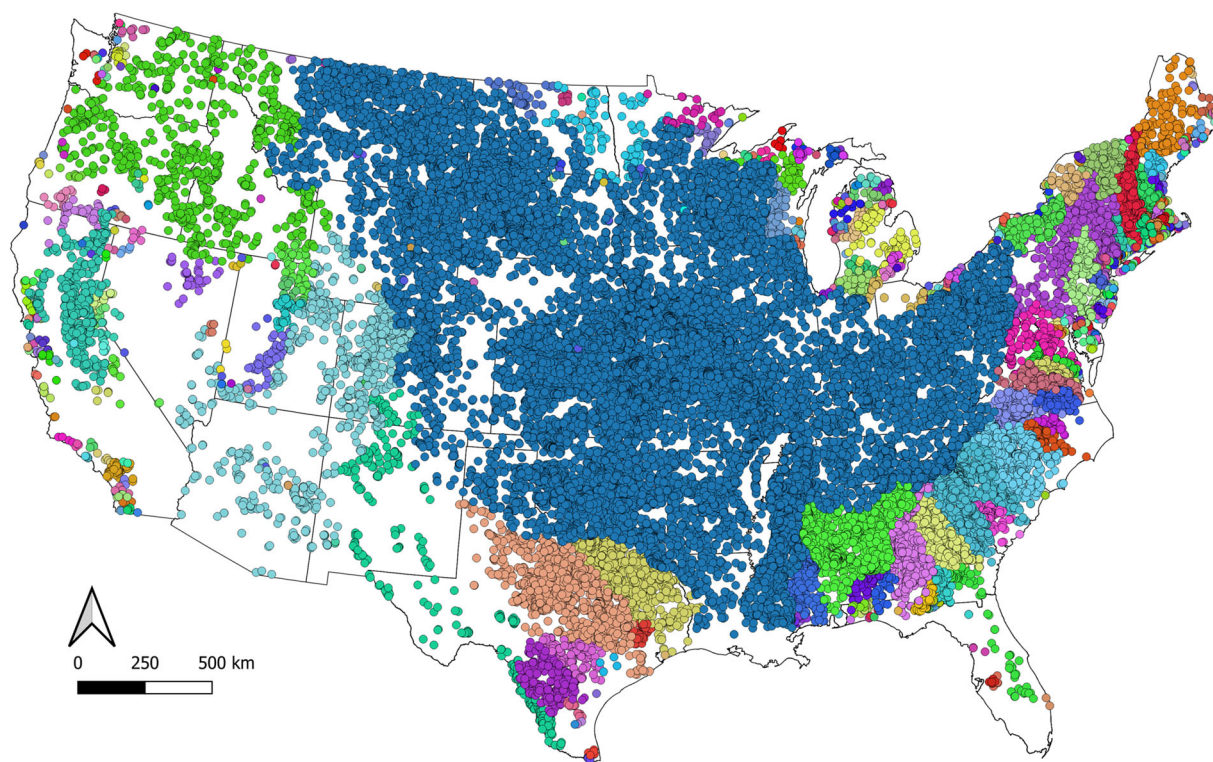


Fig. 2. Dam locations. Dam points ($n = 49,525$) in the LAGOS-US NETWORKS module colored according to their network membership ($n = 898$ networks).

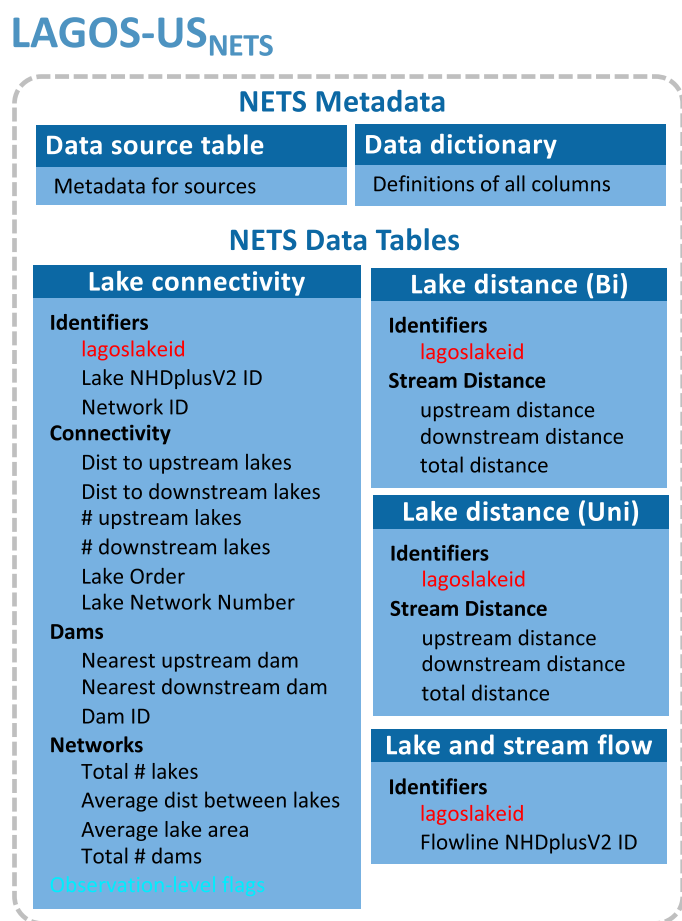


Fig. 3. The LAGOS-US NETWORKS schema. NETWORKS includes meta-data in the form of a source table and a data dictionary and four data tables (*nets_networkmetrics_medres*, *nets_binetworkdistance_medres*, *nets_uninetworkdistance_medres*, and *nets_flow_medres*). The tables are connected to each other and other LAGOS-US modules via *lagoslakeid*, depicted with red text. The *nets_networkmetrics_medres* table also includes observation-level flags, depicted with blue text. The variables in black text included in the four data tables are representative examples, see Tables 1–4 for more details on variables. Note: Not shown is the detailed User Guide.

linear lake chains (Fig. 5a). Higher values of LNN appear in the upper-midwest, west, and south-central U.S. LO (the Strahler order of the outflowing stream; Riera et al. 2000) is fairly evenly distributed across the U.S. and the majority of lakes tend to be lower order (Fig. 5b). LNN ranges from 1 to 50 and LO ranges from 0 to 9 (Table 1).

Although the Mississippi River network includes 32,811 lakes (Fig. 6a), the majority of lake networks have < 100 lakes and over a third consist of only two connected lakes (Fig. 6b; Table 1). The network average distance between lakes ranges from less than 1 to over 1500 km, with a median distance of approximately 7 km (Fig. 6c; Table 1). The network average lake area ranges from just over 1 to about 47,000 ha, with a median of approximately 18 ha (Fig. 6d; Table 1). The number

of dams in a network ranges from 0 to about 25,000 (the Mississippi River network), with the majority of networks including 1 dam (Fig. 6e; Table 1).

Overview of data access

LAGOS-US NETWORKS v1 is made up of metadata and data tables that are csv files as well as a User Guide in pdf form, all of which are available for public download via the EDI repository (King et al. 2021b). There is also code available on GitHub for those who would like to reproduce, extend, or adapt our networks (Wang and King 2020) and an R package that can be used to download and link NETWORKS with the other LAGOS-US core and extension modules (*lagosus*; Stachelek 2020). When NETWORKS data are included in analyses, users should cite them as well as this data paper that describes the motivation and context for creating the NETWORKS module.

NETWORKS v1 data and documentation: King, K., Q. Wang, L. K. Rodriguez, M. Haite, L. Danila, T. Pang-Ning, J. Zhou, and K. S. Cheruvilil. 2021. LAGOS-US NETWORKS v1.0: Data module of surface water networks characterizing connections among lakes, streams, and rivers in the conterminous U.S. Environmental Data Initiative. doi:10.6073/pasta/98c9f11df55958065985c3e84a4fe995.

Data paper: King, K., Q. Wang, L. K. Rodriguez, and K. S. Cheruvilil. 2021. Lake networks and connectivity metrics for the conterminous U.S. (LAGOS-US NETWORKS v1). *Limnol. Oceanogr.: Letters*. doi:10.1002/lol2.10204.

LAGOS-US R package: Stachelek, J. LAGOSUS: Interface to the Lake Multi-scaled Geospatial and Temporal Database. R package version 0.0.1. doi forthcoming.

Methods

This section outlines the methods used to create lake networks as well as derive connectivity metrics in LAGOS-US NETWORKS v1. We also explain how dam data from the NABD was linked to our networks to add potential barriers to connectivity. For further technical detail on this process, we have submitted data documentation in the form of a User Guide along with the metadata and datasets on EDI (King et al. 2021b) and users can consult the published code for further extension of our methods (Wang and King 2020).

Creating lake networks

Lake networks across the continental U.S. were created using the flow table from the medium resolution NHDPlusV2 database (U.S. Geological Survey 2019). The flow table from NHDPlusV2 consisted of every flowline (streams and artificial flowlines that go through lakes) either in the FROM column or TO column, denoting a direction of flow from one line to the other, as well as the distance for each connection between two flow lines. Prior to creating a graph, we removed several connections. We removed coastline connections (Fcode 56600; McKay et al. 2012) so that the connectivity

Table 1. Summary of variables in the *nets_networkmetrics_medres* table ($n = 86,511$ lakes; file size ~ 11.1 MB).

Variable name	Variable description	Units	Min	Median	Mean	Max
lagoslakeid	Unique lake identifier developed by LAGOS-US.	NA	NA	NA	NA	NA
nhdplusv2_comid	Unique lake identifier from the nhd for the medium resolution NHDplusV2.	NA	NA	NA	NA	NA
lake_nets_upstreamlake_km	Distance to nearest upstream lake using a unidirectional graph.	Kilometers	0.00	0.987	3.10	220
lake_nets_downstreamlake_km	Distance to nearest downstream lake using a unidirectional graph.	Kilometers	0.00	26.0	165	2410
lake_nets_bidirectionallake_km	Distance to the nearest lake upstream or downstream using a bi-directional graph.	Kilometers	0.00	4.40	8.40	285
lake_nets_upstreamlake_n	The number of upstream lakes directly connected through streams to a lake.	Number	0.00	0.00	1.44	7310
lake_nets_downstreamlake_n	The number of downstream lakes directly connected through streams to a lake.	Number	0.00	1.00	1.44	13.0
lake_nets_lakeorder	Lake order (LO) follows the Strahler stream order of the outflowing stream, where the higher order stream is chosen if more than one outlet occurs (Riera et al. 2000; Martin and Soranno 2006). The exceptions are that headwater lakes are 0 and terminal lakes receive the order of the highest inflowing stream.	NA	0.00	1.00	1.09	9.00
lake_nets_lnn	Lake network number (LNN) is the position of a lake within the network in reference to other lakes. The lake at the top of a network (i.e., no upstream lakes) will be 1, the next lake downstream will be 2, etc. If a lake has more than one lake upstream, it will take the higher LNN.	NA	1.00	1.00	1.65	50.0
lake_nets_nearestdamup_id	The NABD dam ID for the nearest upstream dam.	NA	NA	NA	NA	NA
lake_nets_nearestdamup_km	Distance to nearest upstream dam.	Kilometers	0.00	2.66	7.90	270
lake_nets_nearestdamdown_id	The NABD dam ID for the nearest downstream dam.	NA	NA	NA	NA	NA
lake_nets_nearestdamdown_km	Distance to nearest downstream dam.	Kilometers	0.00	0.00	37.0	1700
lake_nets_totaldamup_n	The total number of upstream dams.	Number	0.00	0.00	0.800	6030
lake_nets_totaldamdown_n	The total number of downstream dams.	Number	0.00	0.00	1.31	32.0
net_id	The unique identifier assigned by LAGOS-NETWORKS for each network ($n = 898$ networks).	NA	NA	NA	NA	NA
net_lakes_n	The total number of lakes in each network.	Number	2.00	3.00	96.3	32,800
net_averagelakedistance_km	Average distance between lakes in each network.	Kilometers	0.015	7.36	27.4	1650
net_averagelakearea_ha	Average lake area in each network.	Hectares	1.19	17.8	130	47,200
net_dams_n	The total number of dams on each network.	Number	0.00	1.00	54.3	25,000

Table 2. Summary of variables in the *nets_uninetworkdistance_medres* table ($n = 124,251$ lake pairs; file size ~ 2.6 MB).

Variable name	Variable description	Units
lagoslakeid	LAGOS lake identifier of the “from” lake that is connected to lagoslakeid_to using a unidirectional graph (traversing the network downstream).	Null
to_lagoslakeid	Identifier of lake 2 (lagoslakeid) connected to lake 1 using a unidirectional graph (traversing the network downstream).	Null
streamlength_down_km	Distance downstream from lake 1 (lagoslakeid) to lake 2 (lagoslakeid_to) using a unidirectional graph (traversing the network downstream).	Kilometers

networks did not connect through the ocean, estuaries, or the Great Lakes, as well as IDs associated with the Great Lakes water bodies. Artificial flowlines were linked to water bodies (*nhdplusv2_comid*), and these water bodies were linked to *lagoslakeids* using the *lake_link* table from the LAGOS_US_LOCUS module (Smith et al. 2021; Cheruvilil et al. In press). The modified version of the NHDPlusV2 flow table, including where artificial flowlines are matched to lakes from the LAGOS-US database, can be found as the *nets_flow_medres* data table.

We applied a graph theory framework to create lake networks from the *nets_flow_medres* data table. Graphs are mathematical structures made up of “nodes” and “edges” used to model pairwise relations between objects (nodes) (Eros et al. 2012). In our case, we were interested in modeling the pairs of lakes that are connected by streams (edges). We created lake networks using bidirectional graphs, which considered *both* downstream and upstream connections, using both lakes and streams as nodes (Fig. 7a). We used Dijkstra’s algorithm (Cormen et al. 2001) to traverse the graph both upstream and downstream starting at a given lake. During the traversal, if a node was a stream, we continued traversing the graph until the node was a lake. We saved the distance from the given lake to this lake and stopped traversing. If there were multiple paths to connect the same two lakes, the algorithm chose and saved the path with the shortest length. This approach produced all the connections of the given lake to its neighbor lakes. This process was repeated for every lake until the connections and stream course distances between all lakes

Table 3. Summary of variables in the *nets_binetworkdistance_medres* table ($n = 39,498,506$ lake pairs; file size ~ 1.9 GB).

Variable name	Variable description	Units
lagoslakeid	LAGOS lake identifier of the “from” lake that is connected to lagoslakeid_to using a bidirectional graph (traversing the network both downstream and upstream).	Null
to_lagoslakeid	Identifier of lake 2 (lagoslakeid) connected to lake 1 using a bidirectional graph (traversing the network both downstream and upstream).	Null
streamlength_total_km	Total stream distance from lake 1 to lake 2 (as indicated by lagoslakeid) using a bidirectional graph.	Kilometers
streamlength_up_km	Distance upstream from lake 1 to lake 2 (as indicated by lagoslakeid) using a bidirectional graph.	Kilometers
streamlength_down_km	Distance downstream from lake 1 to lake 2 (as indicated by lagoslakeid) using a bidirectional graph.	Kilometers

were known. A network includes all lakes that are connected to another lake up or downstream, thus including lakes located on different tributaries that are connected through a

Table 4. Summary of variables in the *nets_flow_medres* table ($n = 2,722,347$ rows of flowlines, streams and artificial flowlines through lakes; file size ~ 76.2 MB).

Variable name	Variable description
from_comid	Common identifier of the upstream NHDFlowline feature.
to_comid	Common identifier of the downstream NHDFlowline feature.
from_lagoslakeid	Identifier of the upstream lake as indicated by lagoslakeid.
to_lagoslakeid	Identifier of the downstream lake as indicated by lagoslakeid.

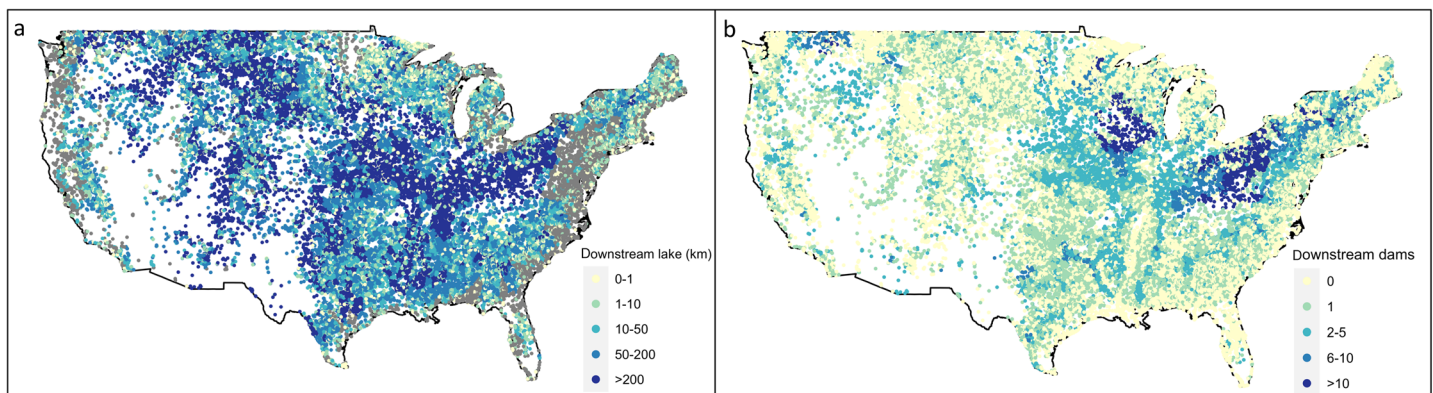


Fig. 4. Map of downstream lakes and dams. Map depicting location of lakes in NETWORKS color-coded according to (a) the distance to the nearest downstream lake (km), where gray circles mean the lake has no downstream lake and (b) the number of downstream dams from each lake within its network.

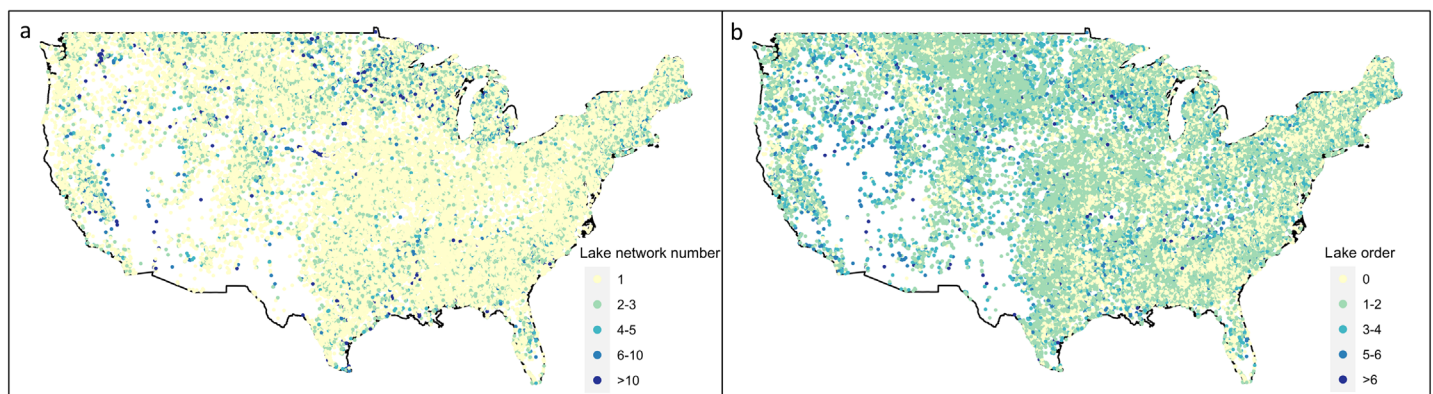


Fig. 5. Map of lake network position. Map depicting location of lakes in NETWORKS color-coded according to their network position measured as (a) lake network number and (b) lake order.

downstream confluence (Fig. 7c,d). We assigned each of these networks a unique identification number (*net_id*). All of the stream course distances between pairs of lakes can be found in the *nets_binetworkdistance_medres*. The artificial flowline distances through lakes were not included in these distances. This table includes upstream, downstream, and total distance between two lakes. The total distance may be smaller than the sum of the upstream and downstream distances due to the absence of data on stream reach intersection. For example, it is unknown whether one stream reach intersects the top, middle, or bottom of another reach; therefore, an intersecting stream reach was only counted once for the total distance, but was included in both the downstream and upstream distance columns (see the dataset User Guide, King et al. 2021b for more details).

Linking dams to lake networks

The NABD is a dataset of large, anthropogenic barriers that are spatially linked to the NHDPlusV1 data product to facilitate analyses based on the NHD and National Inventory of

Dams (Ostroff et al. 2013). Cooper et al. (2017) added 170 additional dams to this database from the USFWS Fish Passage Decision Support Tool and excluded ~250 dams that were identified as having been removed since the NABD was published (American Rivers 2019). The 49,525 dams were linked to the NHDPlusV2 flowlines and were incorporated into networks. Dams were assigned to a *lagoslakeid* if they were less than 50 m from a lake (Polus et al. In preparation). Dams that were directly on (or in) a lake could not be considered as up- or downstream because they were on the node and therefore, did not have a direction in reference to that node. Therefore, these dams were assigned as upstream or downstream from a lake using two methods:

1. Using ArcGIS, lake inlets and outlets were identified using the start and end vertices associated with the artificial flowlines and extracted as points representing inlets and outlets. For each dam point location, the nearest three inlets or outlets (combined) were identified using Euclidean distance in the ArcGIS GenerateNear tool. If both inlets

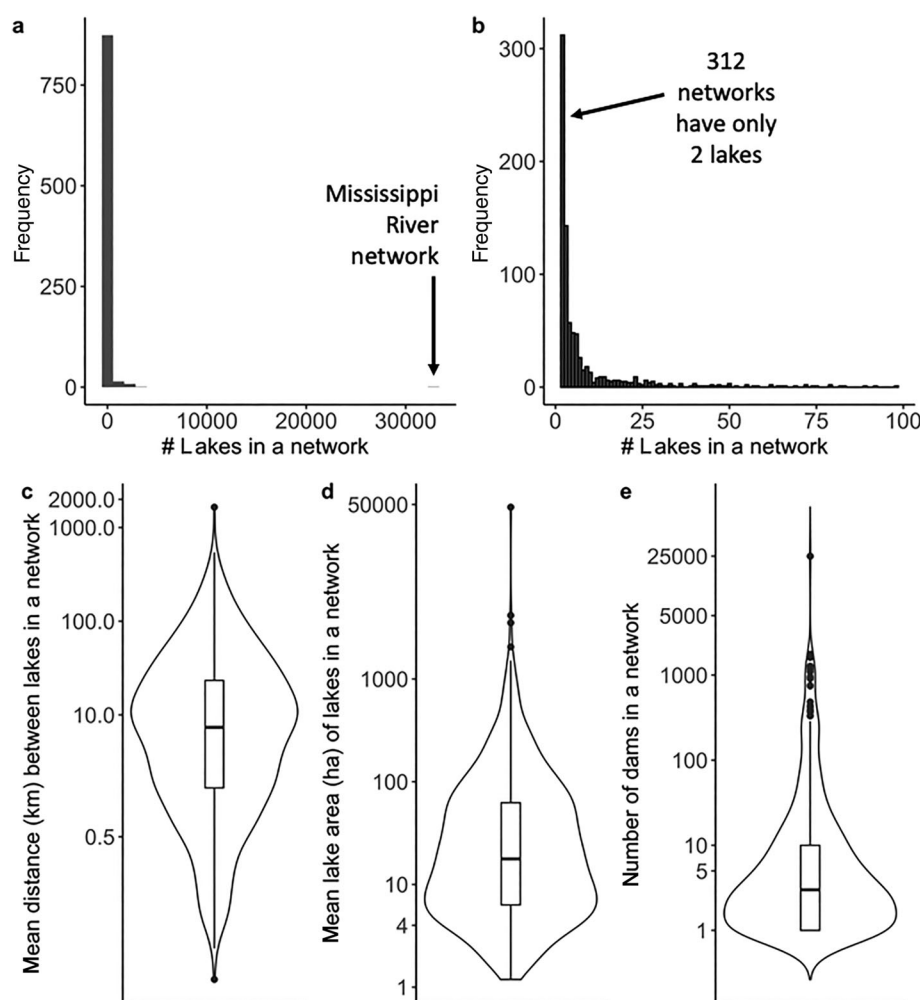


Fig. 6. Network metrics. Network metric summaries of (a) frequency distribution of the number of lakes in a network, (b) frequency distribution of the number of lakes in networks with less than 100 lakes, (c) boxplot of the average distance (km) between lakes in a network, (d) boxplot of the average lake area (ha) in a network, and (e) boxplot of the number of dams in the network. See Table 1 for summary statistics. Note that many networks have 0 dams, however, because this plot was log transformed zeros were not included in this plot.

and outlets for the same lake were very near each other or an inlet or outlet for another lake was very near, the dam position was assigned for manual review. Methods are available as Python code within the LAGOS GIS Toolbox (http://github.com/cont-limno/LAGOS_GIS_Toolbox; `national_outlets_inlets.py`, `dams_link_lake_junctions.py`). There were 11,551 dams that were assigned upstream or downstream of a lake using this method.

2. The remaining dams ($n = 1079$) that could not be identified by the automated process were then manually classified by visual inspection of the dam location in comparison to the NHD polygons and flowlines and manually assigned as either on the upstream or downstream side of a lake.

Two data flags were created during the process of linking dams to lakes and streams/ivers. These flags were for cases

when a dam fell onto an artificial flowline contained within a lake or when multiple dams fell on the same lake (Table 5; section on informational flags).

Quantifying lake and network connectivity metrics

After creating the networks, several metrics were derived at the lake scale using a unidirectional graph. Unidirectional graphs traverse the network downstream only (Fig. 7b). We used Dijkstra's algorithm (Cormen et al. 2001) to traverse the graph downstream starting at a given lake. The same process was used for the unidirectional graph that was used for the bidirectional graph described in the above section "Creating lake networks." The stream course distances between two lakes using a unidirectional graph can be found in the `nets_uninetworkdistance_medres` table.

The metrics for the nearest lake distance were determined by comparing the distance between each lake and all of its neighboring lakes and choosing the nearest distance upstream

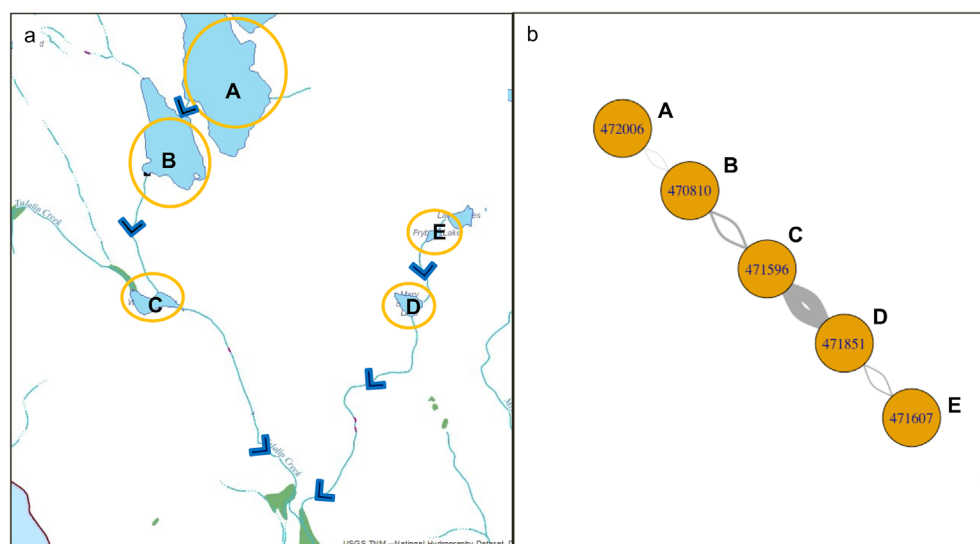


Fig. 7. Graph creation. A bidirectional graph (a) and unidirectional graph (b). An example of a lake network (c) compared to its corresponding bidirectional graph (d) to illustrate how networks were created and how upstream or downstream distances were defined in NETWORKS. The distance between lake C and lake D includes traversing the network downstream and then upstream. The stream course distance is used as a weight in panel (d); thicker connecting lines depict further distances. Panel (d) was made using the “igraph” package (Csardi and Nepusz 2006).

and the nearest distance downstream from the unidirectional graph. Note that not all lakes have both an upstream and downstream lake. The number of directly connected lakes upstream was computed as the indegree of a lake, i.e., the number of lakes upstream only connected through streams flowing into the lake. Similarly, the number of directly connected downstream lakes was calculated using the outdegree of a lake, i.e., lakes directly connected through streams flowing out of a lake. There were instances when a lake did not have any directly connected upstream or directly connected downstream lakes because the lake was only connected through the bidirectional graph to the lake network ($n = 7617$). Therefore, we also included a metric for the nearest lake using bidirectional distance. These instances are easily identifiable because these lakes only have a nearest bidirectional distance and do not have a nearest downstream or nearest upstream lake distance.

Two metrics that describe the position of a lake within the network and landscape were derived using a unidirectional graph: LNN and LO (Riera et al. 2000; Martin and Soranno 2006) (Fig. 8). LNN was computed by starting at the first lake in a network (e.g., no upstream lakes) and assigning that lake a “1,” then moving downstream to another lake and assigning that lake a “2,” and so on throughout the network. Therefore, multiple lakes in a network could be assigned a “1” if they do not have any upstream lakes. Lakes with multiple upstream lakes were assigned the larger sequential number (Martin and Soranno 2006). LO was assigned using the Strahler stream order from the NHDplusV2 attributes. LO followed the Strahler stream order of the outflowing stream, where the higher order stream was chosen if more than one outlet was

present (Riera et al. 2000; Martin and Soranno 2006). There were two exceptions to this: headwater lakes were assigned a “0” and terminal lakes received the Strahler order of the inflowing stream (Riera et al. 2000; Martin and Soranno 2006). We considered inflowing streams for LO calculations to differentiate between headwater lakes and lakes that had inflowing streams but not upstream lakes. There were instances when a loop between two lakes occurred (0.02% of all connections), for example lake A flowed to lake B and lake B flowed back to lake A. In these instances, we randomly removed one connection.

Several dam metrics were derived that characterize barriers to connectivity. The depth first search (DFS; Cormen et al. 2001) algorithm was used to traverse each lake-stream network to find all of the upstream dams and downstream dams. Dijkstra’s algorithm was used to compute the distance to the nearest upstream and downstream dams (Cormen et al. 2001). Because we used a graph to create the network, the algorithm did not have the exact location of the dam on the stream reach, just the flowline it was located on. Therefore, when deriving the metrics for the nearest dam, the entire stream reach with the dam was included in the distance calculation. Thus, there were instances when two or more dams fell on the same stream flowline (8.7% occurrence). In these instances, all dams were considered as the nearest up- or downstream dams, they were assigned the same distance from the lake, and all of the dam ids were included and separated by a semicolon. These instances are easily identifiable because more than one dam is listed in the *lake_nets_nearestdamdown_id* or *lake_nets_nearestdamup_id* column. Similarly, if multiple dams were on a lake (0.15% occurrence), all of the dams were considered the nearest dam, all dam ids were included, and dams

located on a lake were assigned the distance of 0 km. Lakes with multiple dams on the lake were assigned a flag (Table 5; section on informational flags).

At the network scale, we traversed the completed lake networks using the DFS algorithm. This process calculated total lakes in each network, the average distances between lakes in a network, and the total number of dams in each lake network. The average area of the lakes in each network was calculated using lake area values from LAGOS-US LOCUS v1.0 polygons (Smith et al. 2021; Cheruvilil et al. In press), grouping lakes by networks, and then using the Calculate Geometry tool in ArcGIS. See the User Guide for more methodological details.

Technical validation

Informational flags

During construction of the module, we created a series of data flags that convey something about a data observation that may be of interest to users. These flags are all informational flags of general relevance to the data user and *none* of these flags are cautionary flags that indicate potential concerns for inclusion of particular data observations in analysis (Table 5).

Validation and quality control/quality assurance

The validation and quality control/quality assurance (QAQC) process was intended to ensure that the procedures used to create the values for NETWORKS variables resulted in the intended outcomes. We used two methods for validation and QAQC.

First, during the creation of the metrics, a simulation graph was created to validate the code. This simulation graph included paths that were unidirectional as well as bidirectional, multiple

connections between lakes, lakes that were directly connected to other lakes without streams, and a Great Lake. Using this simulation graph, we checked that the distance between pairs of lakes was correct for downstream, upstream, and bidirectional connections. Then, we ensured that the code accurately selected the shorter distance if there were multiple connections between lakes for both the unidirectional and bidirectional connections. For lakes that did not have a stream connection between them, we ensured the code resulted in downstream and upstream distances of 0 km. Finally, we tested that the code ignored connections to the Great Lakes. Our team manually examined resulting networks and associated metrics using either ArcGIS 10.3 Desktop (ESRI 2014) or the “hydrolinks” package (Winslow et al. 2018), which downloads and traverses paths for the medium resolution NHDplusV2 data to identify potential issues with either the input data or code. All solvable issues were reconciled and the networks or metrics were regenerated and retested until no further issues were found.

After metrics were quantified, we proceeded with a second phase of QAQC. We queried the NETWORKS metrics data table (*nets_networkmetrics_medres*) to: (1) identify potential data or geoprocessing issues and (2) verify that data values were sensible (e.g., within expected ranges and expected completeness of data). These checks of individual variables assessed that the workflow generating data accurately reflected both the source data and the lake-specific values. For this process, the *nets_networkmetrics_medres* data table, in csv (comma-separated values) format, was imported by semiautomated R scripts that then summarized the data table, ensured comparability with the source GIS layer and data dictionary, summarized and mapped values for each variable, and automatically generated scores for three main evaluation criteria in a QAQC summary report

Table 5. Description and occurrence of lake informational data flags in the *nets_networkmetrics_medres* table (number of lakes = 86,511).

Flag	Value	Description	User relevance	Number of occurrences	Percent of data
lake_nets_damonlake_flag	Y, N	A value of “Y” indicates that there is at least one dam on this lake. This means that the dam point falls onto one of the artificial flowlines that flows through a lake and is therefore associated with the lake and not a stream reach. A value of “N” indicates there is not a dam on a lake.	This flag primarily serves to alert the user of the presence of a dam directly on a lake as opposed to on a connecting stream reach.	12,630	14.6%
lake_nets_multidam_flag	Y, N	A value of “Y” indicates that there are multiple dams on a lake. A value of “N” indicates there are not multiple dams.	This flag identifies lakes that have multiple dams. There may be a dam at multiple inlets or outlets or a dam at both locations.	132	0.15%

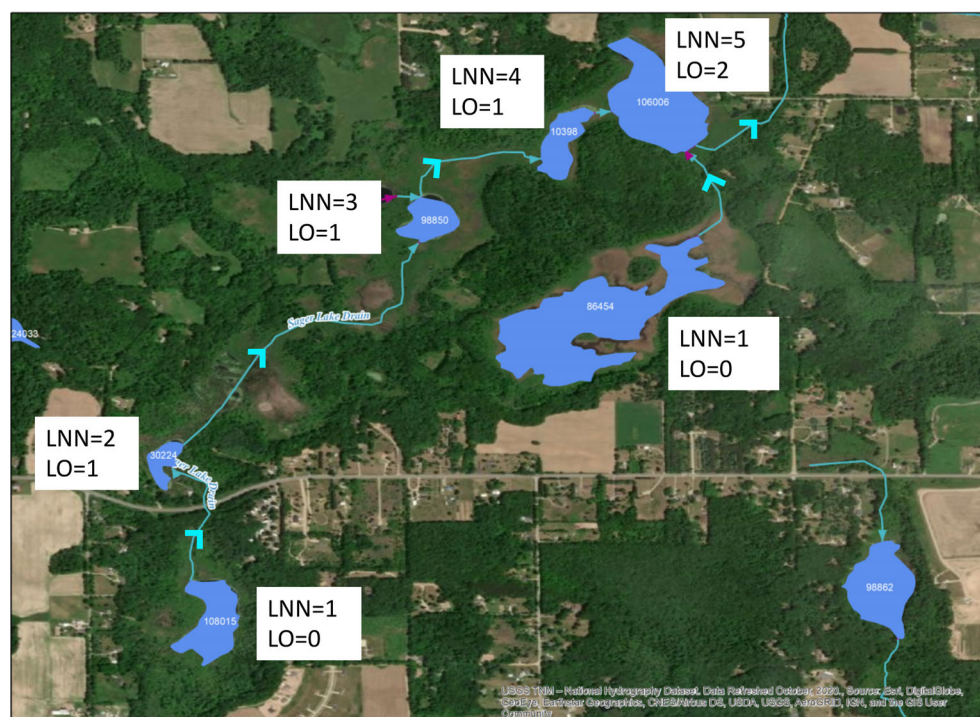


Fig. 8. Lake network number and lake order. Example of part of a lake network with lake network number (LNN) and lake order (LO) metrics for each lake.

provided in html format. Note that actions were iterative; the QAQC review feeds back into the data creation process, which then re-exports the data table and then re-runs the entire QAQC process.

Below are the three evaluation criteria and subsequent actions used in this process:

1. Match with GIS data: This check compared the list of lagoslakeids in the data table with those in the corresponding LAGOS-US LOCUS reference shapefiles maintained in an ArcGIS geodatabase (GIS_LOCUS; Smith et al. 2021; Cheruvilil et al. In press). If a “Fail” warning was generated, nonmatching lagoslakeids were manually investigated to identify the source of the mismatch between the data table and the reference GIS data layer.
2. Match with metadata: Variable names in the data table were compared with the master list of variable names maintained in the metadata table data dictionary. Where there was no match, due to missing or incorrect names in either the data dictionary or the data table, a “Fail” warning was generated and the mismatches were listed in a table in the QAQC report. Where a “Fail” warning was generated, the data dictionary and data table variable names were examined and the name(s) in error were fixed as necessary.
3. Missing value: This check counted the number of observations with missing values, listed them, and produced maps of their location. A “Warn” evaluation was created for this

criterion and variables were inspected to make sure there were no gaps in the input data.

Data use and recommendations for reuse

NETWORKS is based on the medium resolution NHDplusV2 flow data because of limitations in computing capacity at the continuous U.S. scale and because there are stream attributes in the medium resolution that are not available in the NHD high-resolution data. Although the NHD high resolution includes smaller streams that connect some lakes that are not connected in the NHD medium resolution, the NHD high resolution separates some lakes that lie very close to a stream and considers them isolated when they are connected in the NHD medium resolution. Thus, there are benefits and challenges to basing networks and connectivity metrics on either version of the NHD.

We advise users to heed caution when combining the data in NETWORKS that are based on the medium resolution NHDplusV2 flow data with other resolutions of the NHD data or with derived data using other NHD versions. For example, when users combine NETWORKS with LAGOS-US, they should be aware that connectivity metrics will differ between the LOCUS and NETWORKS modules. Because NETWORKS is based on the medium resolution NHDplusV2 flow data, whereas LOCUS used the NHD high resolution (Smith et al. 2021; Cheruvilil et al. In press), there may be lakes classified as connected in LOCUS that are not part of a network in NETWORKS or a lake classified as isolated in LOCUS might be part of a network in the NETWORKS module.

We also wish to remind users that the metrics in NETWORKS were derived only for lakes connected to other lakes. Therefore, the NETWORKS module does not include isolated lakes or lakes that are only connected to streams. Finally, networks that cross international borders (e.g., Canada) may underestimate the number of upstream lakes, nearest upstream lake, and LNN, due to the nature of the dataset being constrained to the continental U.S.

The data in NETWORKS have not yet been used in research that has been published, but are being used in several ongoing efforts that will result in publications. For example, these metrics are being used to quantify how connectivity affects both stream and lake fish communities within and across networks. These data are also being used to determine how freshwater networks best facilitate latitudinal range shifts for species under ongoing climate change and if highly connected networks reside in protected areas. Finally, these data are being used for studies of invasive species movement and species distribution modeling.

NETWORKS will be a valuable data source for building broad-scale understanding of lake networks and the role of connectivity and barriers to connectivity for movement of abiotic materials and biota. Future users can combine these data with a variety of lake abiotic and biotic data, by linking the module with other LAGOS-US modules or to their study systems by using the NHD unique identifiers. For example, lake nutrient data from LAGOS-US-LIMNO could be linked with NETWORKS to investigate if lake network position has the same influence on lake nutrient concentrations across regions of the U.S. Likewise, water temperature, habitat data, and species observations could be linked to on-network lakes to investigate the available habitat suitability for species range shifts. Further, a management agency considering barrier removal could use the NETWORKS module to identify the entire network that would be affected by removal, including quantity of habitat available for migratory species or invasive species within the network.

In addition to the derived metrics, we have published the distance table, flow table, and our code to facilitate the extension of our work through the creation of additional metrics or modeling by future users. For example, the distance table and connectivity table could be joined to calculate the lake density (number of lakes per total distance in the network) or the lake distance table could be used for similar modeling techniques as those that have been developed for stream networks (Peterson et al. 2007; Isaak et al. 2014). Moreover, the distance and flow tables act as an “edge list” that can be used in the “igraph” package (Csardi and Nepusz 2006) to calculate more graph metrics for specific lake networks (e.g., betweenness centrality, lakes that are the shortest path between many other lakes in the network). Finally, the flow table includes all individual stream reaches, so our code could be used to make similar metrics for streams or to incorporate more information on lakes or streams such as size, slope, or quality of habitat

patch that would weight the links and nodes (Eros et al. 2012) to answer a myriad of questions related to freshwater surface connectivity.

Comparison with existing datasets

Although the majority of past studies fail to address surface-water connectivity at the U.S. national scale, we provide an overview of preexisting datasets so that readers and users understand what connectivity information was available at the time of writing and how these previous methods align with or deviate from the networks and connectivity metrics in the NETWORKS module.

Several connectivity datasets and tools for the conterminous U.S. exist for streams. For example, the NHD (McKay et al. 2012) includes connectivity metrics such as a modified version of Strahler stream order. Cooper and Infante (2017) have created dam metrics for streams in the conterminous U.S., which represent network fragmentation. However, these datasets and metrics do not include lakes and the stream networks stop at dams because they were created for biotic variables that cannot move past these barriers (e.g., fish). In addition, the U.S. Geological Survey has created a tool “Hydro Network-Linked Data Index (NLDI)” (<https://labs.waterdata.usgs.gov/about-nldi/index.html>), which is a web application programming interface that can traverse upstream or downstream and link to other NHD or Water Quality Portal data. Similarly, the “nhdplusTools” R package (Blodgett 2019) can be used to download NHD data and navigate upstream or downstream from a feature in the network. These tools may be useful for small-scale studies, but would take considerable time for broad-scale research.

For lakes, there are a few broad-scale U.S. datasets that have important similarities and differences to NETWORKS. The LAGOS-US LOCUS module (Smith et al. 2021; Cheruvilil et al. In press) includes several connectivity metrics, such as connectivity classes, the number of upstream lakes, upstream lake area, and stream density within a watershed. However, this dataset lacks downstream connections because it was created for abiotic variables. LakeCAT includes some metrics such as density of streams or dams within a catchment; however, they do not quantify lake or network connectivity metrics (Hill et al. 2018). Fergus et al. (2017) provide connectivity information at the HUC 12 and HUC 8 scale, including lake, stream, and wetland densities and clusters, although this is only for the northeastern/northern mid-western region of the U.S. The “hydrolinks” package (Winslow et al. 2018), which was used for NETWORKS validation, is a tool for mapping connectivity; however, it only traverses upstream or downstream, it includes coastal lines and Great Lakes polygons, and it is best used for small extents because of computation time. Therefore, NETWORKS extends these datasets and tools by providing lake networks and connectivity metrics for the entire conterminous U.S. that include both lakes and streams and both upstream and downstream information that is useful for studying both abiotic and biotic properties of fresh waters.

References

- American Rivers. 2019. American Rivers Dam Removal Database. Database: figshare. doi:[10.6084/m9.figshare.5234068.v6](https://doi.org/10.6084/m9.figshare.5234068.v6)
- Beisner, B. E., P. R. Peres-Neto, E. S. Lindström, A. Barnett, M. L. Longhi, E. S. Lindstr, and M. L. Longhi. 2006. The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology* **87**: 2985–2991. doi:[10.1890/0012-9658\(2006\)87\[2985:TROEAS\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2985:TROEAS]2.0.CO;2)
- Bishop-Taylor, R., M. G. Tulbure, and M. Broich. 2015. Surface water network structure, landscape resistance to movement and flooding vital for maintaining ecological connectivity across Australia's largest river basin. *Landsc. Ecol.* **30**: 2045–2065. doi:<https://doi.org/10.1007/s10980-015-0230-4>
- Blodgett, D. 2019. nhdplusTools: Tools for accessing and working with the NHDPlus. Available from <https://code.usgs.gov/water/nhdplusTools>
- Calabrese, J. M., and W. F. Fagan. 2004. A comparison-shop-per's guide to connectivity metrics. *Front. Ecol. Environ.* **2**: 529–536.
- Cheruvelil, K. S., P. A. Soranno, I. M. McCullough, K. E. Webster, L. Rodriguez, and N. J. Smith. In press. LAGOS-US LOCUS v1.0: Data module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. *Limnol. Oceanogr.: Letters*. doi:[10.1002/lol2.10203](https://doi.org/10.1002/lol2.10203)
- Cooper, A. R., and D. M. Infante. 2017. Dam metrics representing stream fragmentation and flow alteration for the conterminous United States linked to the NHDPLUSV1: U.S. Geological Survey data release. doi:[10.5066/F7FN14C5](https://doi.org/10.5066/F7FN14C5)
- Cooper, A. R., D. M. Infante, W. M. Daniel, K. E. Wehrly, L. Wang, and T. O. Brenden. 2017. Assessment of dam effects for streams and fish assemblages of the conterminous USA. *Sci. Total Environ.* **586**: 879–889. doi:[10.1016/j.scitotenv.2017.02.067](https://doi.org/10.1016/j.scitotenv.2017.02.067)
- Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein. 2001. Section 22.3: Depth-first search and section 24.3: Dijkstra's algorithm, p. 540–549, 595–601. In *Introduction to algorithms*, 2nd ed. MIT Press and McGraw-Hill.
- Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. *Int. J. Complex Syst.* **1695**: 1–9.
- Eros, T., J. D. Olden, R. S. Schick, D. Schmera, and M. J. Fortin. 2012. Characterizing connectivity relationships in freshwaters using patch-based graphs. *Landsc. Ecol.* **27**: 303–317. doi:[10.1007/s10980-011-9659-2](https://doi.org/10.1007/s10980-011-9659-2)
- ESRI. 2014. *ArcGIS Desktop: Release 10.3*. Environmental Systems Research Institute.
- Fergus, C., J. Lapierre, S. Oliver, N. Skaff, K. Cheruvelil, K. Webster, C. Scott, and P. Soranno. 2017. Integrated freshwater abundance and connectivity clusters at the Hydrologic Unit 8 scale for the Midwest and Northeast U.S.A. – freshwater metric variables and k-means cluster assignment ver 2. Environmental Data Initiative. doi:[10.6073/pasta/e69b7495674e403baa19a22ffbf40e1](https://doi.org/10.6073/pasta/e69b7495674e403baa19a22ffbf40e1)
- Gardner, J. R., T. M. Pavelsky, and M. W. Doyle. 2019. The abundance, size, and spacing of lakes and reservoirs connected to river networks. *Geophys. Res. Lett.* **46**: 2592–2601. doi:[10.1029/2018GL080841](https://doi.org/10.1029/2018GL080841)
- Hill, R. A., M. H. Weber, R. M. Debbout, S. G. Leibowitz, and A. R. Olsen. 2018. The lake-catchment (LakeCat) dataset: Characterizing landscape features for lake basins within the conterminous USA. *Freshw. Sci.* **37**: 208–221.
- Isaak, D. J., and others. 2014. Applications of spatial statistical network models to stream data. *WIREs Water*. **1**: 277–294. doi:[10.1002/wat2.1023](https://doi.org/10.1002/wat2.1023)
- Jones, N. E. 2010. Incorporating lakes within the river discontinuum: Longitudinal changes in ecological characteristics in stream–lake networks. *Can. J. Fish. Aquat. Sci.* **67**: 1350–1362. doi:[10.1139/F10-069](https://doi.org/10.1139/F10-069)
- King, K., Q. Wang, L. K. Rodriguez, M. Haite, L. Danila, T. Pang-Ning, J. Zhou, and K. S. Cheruvelil. 2021b. LAGOS-US NETWORKS v1.0: Data module of surface water networks characterizing connections among lakes, streams, and rivers in the conterminous U.S. Environmental Data Initiative. doi:[10.6073/pasta/98c9f11df55958065985c3e84a4fe995](https://doi.org/10.6073/pasta/98c9f11df55958065985c3e84a4fe995)
- King, K. B. S., M. T. Bremigan, D. Infante, and K. S. Cheruvelil. 2021a. Surface water connectivity affects lake and stream fish species richness and composition. *Can. J. Fish. Aquat. Sci.* **78**: 433–443. doi:[10.1139/cjfas-2020-0090](https://doi.org/10.1139/cjfas-2020-0090)
- Kling, G. W., G. W. Kipphut, M. M. Miller, and W. J. O'Brien. 2000. Integration of lakes and streams in a landscape perspective: The importance of material processing on spatial patterns and temporal coherence. *Freshw. Biol.* **43**: 477–497. doi:[10.1046/j.1365-2427.2000.00515.x](https://doi.org/10.1046/j.1365-2427.2000.00515.x)
- Kratz, T., K. Webster, C. Bowser, J. Magnuson, and B. Benson. 1997. The influence of landscape position on lakes in northern Wisconsin. *Freshw. Biol.* **37**: 209–217.
- Martin, S. L., and P. A. Soranno. 2006. Lake landscape position: Relationships to hydrological connectivity and landscape features. *Limnol. Oceanogr.* **51**: 801–814.
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., & Rea, A. (2012). *NHDPlus Version 2: User Guide*. Washington, DC: National Operational Hydrologic Remote Sensing Center.
- National Inventory of Dams (NID). 2015. Washington, DC: U.S. Army Corps of Engineers: Federal Emergency Management Agency. Available from <https://nid.sec.usace.army.mil/ords/f?p=105:1>
- Olden, J. D., D. A. Jackson, and P. R. Peres-Neto. 2001. Spatial isolation and fish communities in drainage lakes. *Oecologia* **127**: 572–585. doi:[10.1007/s004420000620](https://doi.org/10.1007/s004420000620)
- Ostroff, A., D. Wiefelich, A. Cooper, D. Infante, and USGS Aquatic GAP Program. 2013. 2012 National Anthropogenic Barrier Dataset (NABD): U.S. Geological Survey - Aquatic GAP Program: Denver, CO.
- Peterson, E. E., D. M. Theobald, and J. M. Ver Hoeff. 2007. Geostatistical modelling on stream networks: Developing valid covariance matrices based on hydrologic distance and

- stream flow. *Freshw. Biol.* **52**: 267–279. doi:[10.1111/j.1365-2427.2006.01686.x](https://doi.org/10.1111/j.1365-2427.2006.01686.x)
- Polus, S. M., and others. In preparation. LAGOS-US RESERVOIR: Data module classifying conterminous U.S. lakes 4 hectares and larger as natural lakes or reservoirs. Environmental Data Initiative.
- Riera, J. L., J. J. Magnuson, T. K. Kratz, and K. E. Webster. 2000. A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. *Freshw. Biol.* **43**: 301–318. doi:[10.1046/j.1365-2427.2000.00567.x](https://doi.org/10.1046/j.1365-2427.2000.00567.x)
- Sadro, S., C. E. Nelson, and J. M. Melack. 2012. The influence of landscape position and catchment characteristics on aquatic biogeochemistry in high-elevation Lake-chains. *Ecosystems* **15**: 363–386. doi:[10.1007/s10021-011-9515-x](https://doi.org/10.1007/s10021-011-9515-x)
- Saunders, M. I., and others. 2016. Human impacts on connectivity in marine and freshwater ecosystems assessed using graph theory: A review. *Mar. Freshw. Res.* **67**: 277–290. doi:[10.1071/MF14358](https://doi.org/10.1071/MF14358)
- Schmadel, N. M., and others. 2018. Thresholds of lake and reservoir connectivity in river networks control nitrogen removal. *Nat. Commun.* **9**: 2779. doi:[10.1038/s41467-018-05156-x](https://doi.org/10.1038/s41467-018-05156-x)
- Shreve, R. L. 1967. Infinite topologically random channel networks. *J. Geol.* **75**: 178–186.
- Smith, N. J., K. E. Webster, L. Rodriguez, K. S. Cheruvilil & P. A. Soranno 2021. LAGOS-US LOCUS v1.0: Module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. Environmental Data Initiative. doi:[10.6073/pasta/e5c2fb8d77467d3f03de4667ac2173ca](https://doi.org/10.6073/pasta/e5c2fb8d77467d3f03de4667ac2173ca)
- Soranno, P. A., K. S. Cheruvilil, T. Wagner, K. E. Webster, and M. T. Bremigan. 2015. Effects of land use on lake nutrients: The importance of scale, hydrologic connectivity, and region. *PLoS One* **10**: 1–22. doi:[10.1371/journal.pone.0135454](https://doi.org/10.1371/journal.pone.0135454)
- Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. *Eos Trans. AGU* **38**: 913–920.
- U.S. Geological Survey. 2019. National Hydrography Dataset (NHDPlus Version 2.1 for Hydrologic Unit (HU) 4–2001 (published 20191002)), [accessed 2019 August 5]. Available from <https://www.epa.gov/waterdata/nhdplus-national-data>
- Urban, D., and T. Keitt. 2001. Landscape connectivity: A graph-theoretic perspective. *Ecology* **82**: 1205–1218. doi:[10.1890/00129658\(2001\)082\[1205:LCAGTP\]2.0.CO;2](https://doi.org/10.1890/00129658(2001)082[1205:LCAGTP]2.0.CO;2)
- Wang, Q., and K. King. 2020. Code for LAGOS-US NETWORKS v1.0 (Version v1.0.0) *Zenodo*. doi:[10.5281/zenodo.4383172](https://doi.org/10.5281/zenodo.4383172)
- Winslow, L. A., T. H. Hahn, S. D. Princiotta, T. H. Leach, and K. C. Rose. 2018. *hydrolinks*: A new tool to link macroscale to inland water bodies. Version 0.10.0. Available from <https://cran.r-project.org/web/packages/hydrolinks/index.html>
- Wollheim, W. M., and others. 2008. Global N removal by freshwater aquatic systems: A spatially distributed, within-basin approach. *Global Biogeochem. Cycles* **22**: GB2026. doi:[10.1029/2007GB002963](https://doi.org/10.1029/2007GB002963)

Acknowledgments

We thank Dana Infante and Arthur Cooper for their dam database and expertise, Arika Hawkins for her hourly help with manual dam classification, Maggie Haite for her help with data validation and GIS maps, Nicole Smith for GIS support, Pang-Ning Pan and Jiayu Zhou for computer science expertise, Arnab Shuvo for EDI support, and Katherine Webster for QAQC support. We also thank the Continental Limnology Project team, especially Patricia Soranno, for contributions and discussions throughout the development of this data module. This study was supported by the US National Science Foundation Macrosystems Biology Program from grant numbers DEB-1638679 (MSU), DEB-1638539 (PSU), DEB-1638554 (UW), and DEB-1638550 (Missouri).

Submitted 23 December 2020

Revised 19 July 2021

Accepted 20 July 2021