

# Big data from a popular app reveals that fishing creates superhighways for aquatic invaders

Jessica L Weir  <sup>a,\*</sup>, Kirsten Vacura <sup>a</sup>, Jay Bagga  <sup>b</sup>, Adam Berland  <sup>c</sup>, Kieran Hyder  <sup>d,e</sup>, Christian Skov  <sup>f</sup>, Johan Attby <sup>g</sup> and Paul A Venturelli 

<sup>a</sup>Department of Biology, Ball State University, Muncie 47306, IN, USA

<sup>b</sup>Department of Computer Science, Ball State University, Muncie, IN 47306, USA

<sup>c</sup>Department of Geography, Ball State University, Muncie, IN 47306, USA

<sup>d</sup>Center for Environment, Fisheries and Aquaculture Science (Cefas), Lowestoft, Suffolk NR33 0HT, UK

<sup>e</sup>School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR4 7TJ, UK

<sup>f</sup>National Institute of Aquatic Resources, Technical University of Denmark, Silkeborg 8600, Denmark

<sup>g</sup>Fishbrain, 118 27 Stockholm, Sweden

\*To whom correspondence should be addressed: Email: [jit.weir@gmail.com](mailto:jit.weir@gmail.com)

Edited By: Karen E. Nelson

## Abstract

Human activities are the leading cause of biological invasions that cause ecologic and economic damage around the world. Aquatic invasive species (AIS) are often spread by recreational anglers who visit two or more bodies of water within a short time frame. Movement data from anglers are, therefore, critical to predicting, preventing, and monitoring the spread of AIS. However, the lack of broad-scale movement data has restricted efforts to large and popular lakes or small geographic extents. Here, we show that recreational fishing apps are an abundant, convenient, and relatively comprehensive source of “big” movement data across the contiguous United States. Our analyses revealed a dense network of angler movements that was dramatically more interconnected and extensive than the network that is formed naturally by rivers and streams. Short-distanced movements by anglers combined to form invasion superhighways that spanned the contiguous United States. We also identified possible invasion fronts and invaded hub lakes that may be superspreaders for two relatively common aquatic invaders. Our results provide unique insight into the national network through which AIS may be spread, increase opportunities for interjurisdictional coordination that is essential to addressing the problem of AIS, and highlight the important role that anglers can play in providing accurate data and preventing invasions. The advantages of mobile devices as both sources of data and a means of engaging the public in their shared responsibility to prevent invasions are probably general to all forms of tourism and recreation that contribute to the spread of invasive species.

**Keywords:** invasive species, human dimensions, network, big data

## Significance Statement:

Information about the human networks through which aquatic invasive species are spreading is limited by a lack of extensive human movement data. Big data from a popular fishing app have revealed a dense network of short-distanced movements that combine to form invasion superhighways spanning the contiguous United States. These data also revealed potential invasion fronts and invaded hubs that may be superspreaders for two relatively common aquatic invaders. Our results provide insight into a national network of aquatic invasions, and highlight both the potential of big data sources, and the role that anglers can play in preventing aquatic invasions.

## Introduction

The growth and integration of human transportation networks has facilitated the spread of invasive species (1) that disrupt ecosystem services (2), reduce the diversity and abundance of native species (3), threaten endangered species (4), and trigger trophic cascades (2, 5). These ecological impacts have socio-economic consequences for food security, industry, and human health (6, 7) that total more than US\$26 billion globally each year (8).

Human movement data for tracking the spread of invasive species are difficult to collect. Conventional approaches include travel diaries, interviews, and surveys (9–16) that can be biased in myriad ways (17, 18), and tend to have high costs and/or low response rates that limit data in space and time (19, 20). Mobile smart devices can be an efficient source of quality, high-resolution, and individualized movement data over broad spatial scales (21, 22). A total of 85% of adults in the United States use smartphones (23) that collect geolocation data. Spatiotemporal

**Competing Interest:** We disclose the financial interests of J.A., Fishbrain CEO. He provided essential data for analysis and input for revisions but has not influenced the content of the paper to benefit Fishbrain, nor undermined the objectivity of the research. The remaining authors have no competing interests.

**Received:** December 9, 2021. **Accepted:** May 26, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Table 1.** Network terms and values that describe the shape and structure of the angler and hydrologic networks, including their relevance and interpretation to this study.

Term	Definition	Value		Relevance and Interpretation
		Hydrologic	Angler	
Node lake	The fundamental units of a network that are connected by edges and represented as the mid-point of a lake.	50,202	130,498	Lakes were represented by nodes in the network that was connected by both anglers and streamflow. The angler network was composed of more connected lakes than the hydrologic network.
Edge connection	The connections between the nodes and represented as a straight-line path between the mid-points of two connected lakes.	47,494	659,669	Angler movement connected lakes by an order of magnitude more than did rivers and streams.
Component	A group of nodes that are connected to each other by a path (a route across a network that runs along edges). Here, we present the number of components and the percent of nodes in the largest component.	4,717	3,022	More components indicated a fragmented network and a lower chance for the spread of AIS. Lakes were more evenly distributed among hydrologic components that reflected river basins, but anglers connected lakes across great distances and in one single large component.
Diameter	The longest of all shortest paths measured for the largest component of the network.	31.25	95%	The diameter for the largest component of the angler and hydrologic network were approximately equal (but the angler network was much larger and more connected, and therefore, much more traversable per lake node).
Density	A value from 0 to 1 that measures the relative number of all edges to all possible edges that could exist for a set of nodes.	$1.88 \times 10^{-5}$	$3.87 \times 10^{-5}$	A higher density network facilitates dispersal of AIS. The angler network was twice as dense as the hydrologic network with more shortcuts and redundant pathways that connect lakes more fully.
Reciprocity	The number of edges that have two-way flow.	0%	43%	More two-way flow in the network leads to a higher chance of spreading AIS between lakes. The hydrologic network facilitated passive dispersal in the downstream direction, whereas the angler network facilitated dispersal in both directions.

data from mobile smart devices and social software applications (apps) have been used to predict the spread of COVID-19 (24) and other diseases (25), discern human migration patterns (26), track species distributions (27–29), and estimate recreational demand for natural spaces (30, 31). Recent studies suggest that these data can also be used to identify the pathways along which humans spread invasive species (18, 32).

Here, we use 10 years of movement data from the popular fishing app Fishbrain to show how recreational anglers connect > 100,000 lakes across the contiguous United States, and how the resultant connectivity network provides unique insight into the current and future distribution of aquatic invasive species (AIS). Anglers are an important vector for AIS dispersal, but studies have focused on their movement at local or regional scales—often only for large or popular lakes (9, 15, 33). Two notable exceptions used movement data from apps to reveal seasonal connectivity networks among hundreds of lakes in Alberta, Canada (32), and to characterize fishing trips among ~20,000 lakes in the United States (18). These studies promote the exciting possibility of using mobile data to identify human-mediated pathways of dispersal among a much larger and more diverse subset of lakes at regional to continental scales.

Our study exemplifies how extensive, high-resolution spatiotemporal mobile data can provide an unprecedented glimpse into dispersal patterns and invasion dynamics that ultimately benefit invasive species research and management. We began by contrasting the connectivity of waterbodies in the United States through two common AIS dispersal pathways, including the novel connections that are created by angler movement and the hydrologic connections that occur naturally. We then show how the local movement of anglers combines and spans the continent to form pathways that overlap with the US interstate highway system. We assess the proximity of invaded lakes to this invasion superhighway and identify possible superspreaders for two common

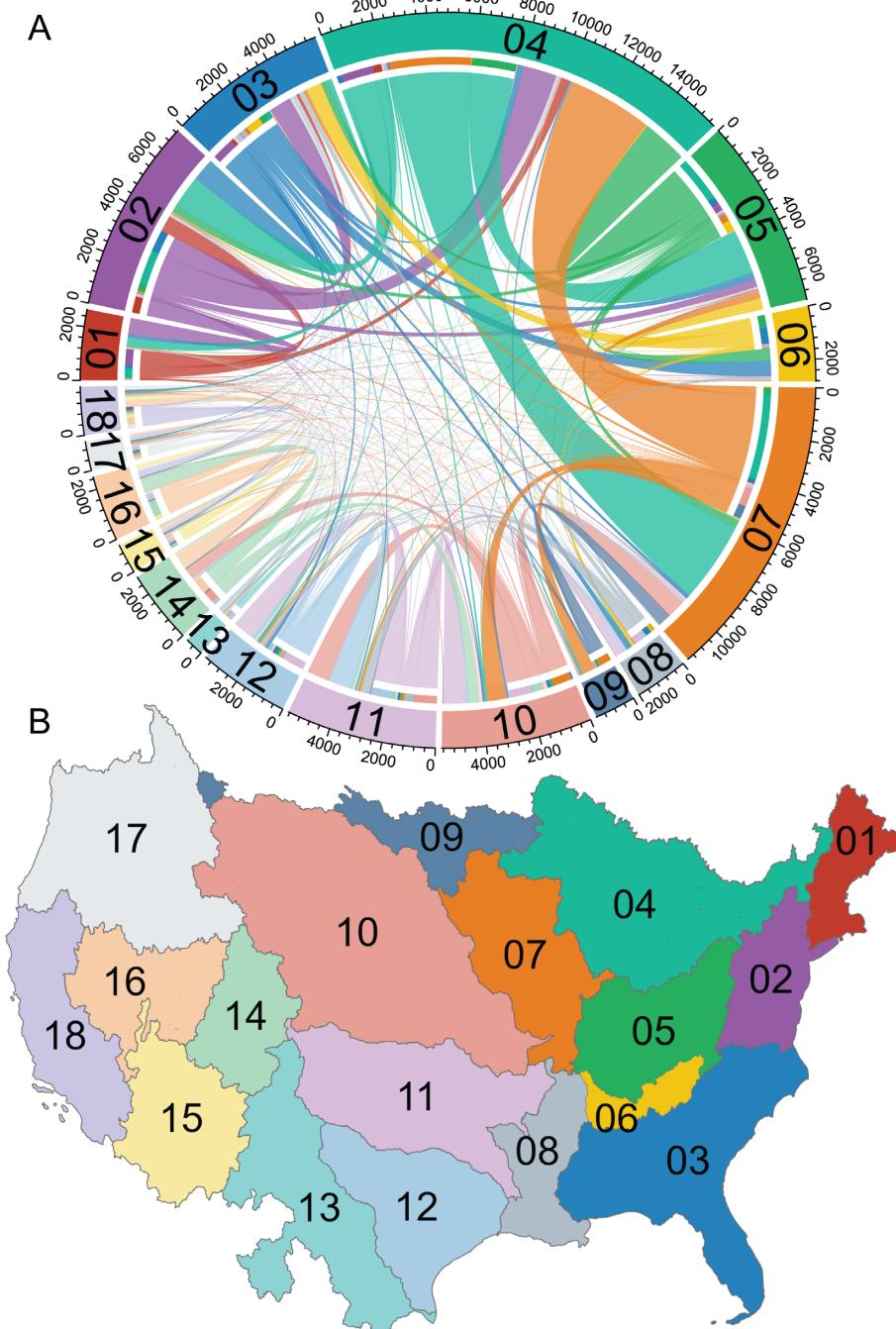
and widespread invaders of the United States. We then map areas of the United States that are presumably at high-risk of invasion because of the connections that originate at an invaded lake and terminate at an uninvaded lake.

## Results

We assigned 3,905,675 of the 4,898,603 (78%) freshwater catches that were logged in the United States to lake polygons from the National Hydrography Dataset (NHD). From these catches, sequential trips within a 99-day period resulted in 1,130,241 individual connections by 248,497 anglers. These data described an angler network that comprised 130,498 nodes (lakes) that were connected by 659,669 edges. The largest connected component in the angler network included 95% of lakes, spanned the contiguous United States, and was easy to traverse via shortcuts and redundant pathways (Table 1). In contrast, hydrologic connections formed many small, isolated components rather than a large, interconnected network (Table 1). Only 50,202 (38%) lakes in the angler network were connected by 47,494 edges that represented the downstream flow of river and stream segments. The remaining 80,296 (62%) lakes that comprised the angler network were hydrologically isolated from each other.

Anglers reinforced 8,905 (18.75%) existing streamflow connections and created reciprocal upstream flow (i.e. in the upstream direction) for 18.8% (8,931) of hydrologic connections. Anglers created 650,764 connections between lakes that were not already connected by rivers or streams.

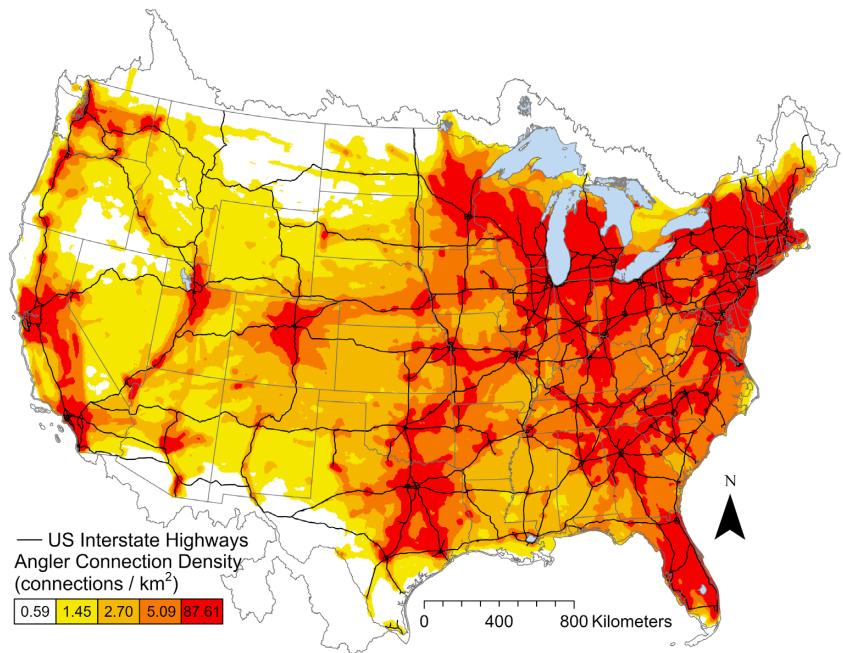
Major river basins within the contiguous United States were hydrologically isolated from each other by definition (except for terminal outlets; Figure 1A), but highly connected by angler movement (Figure 1B). Each major basin was most connected by anglers to the major basins that were adjacent to it, and the strongest interbasin angler connections were between the Great Lakes and



**Fig. 1.** The 18 major river basins of the contiguous United States are connected by anglers across natural hydrologic boundaries. (A) A map of the 18 major river basins in the contiguous United States that extend into parts of Canada and Mexico. These basins exist because waterbodies are separated by natural hydrologic boundaries, except at the basin outlet. The \* represents an area where network lakes are hydrologically connected across the river basin boundary. (B) A chord diagram showing the connections from angler movement among the 18 major river basins. The outer ring and color correspond to the region in which the angler connection originated, and the inner ring corresponds to the basin in which those connections terminate. Thicker lines indicate more movement, and therefore, higher propagule pressure. Basins: 01 New England, 02 Mid Atlantic, 03 South Atlantic-Gulf, 04 Great Lakes, 05 Ohio, 06 Tennessee, 07 Upper Mississippi, 08 Lower Mississippi, 09 Souris-Red-Rainy, 10 Missouri, 11 Arkansas-White-Red, 12 Texas-Gulf, 13 Rio Grande, 14 Upper Colorado, 15 Lower Colorado, 16 Great Basin, 17 Pacific Northwest, and 18 California.

Upper Mississippi basins (Figure 1B). The strongest connection between two nonadjacent basins originated in the South Atlantic Gulf basin and terminated in the Great Lakes basin. Angler connections were strongest among basins in the eastern United States, regardless of basin size or lake number. Only one major river basin pair was not connected by angler movement: from the Rio Grande to the Souris-Red-Rainy basin.

Half of all angler movements occurred  $\leq 7$  days apart (Figure S1, Supplementary Material) and were between lakes that were separated by  $\leq 20$  linear km (Figure S2, Supplementary Material). These frequent, short-distance movements combined to form a continuous network that connected lakes across the contiguous United States (Figure 2). These areas of high connection density overlapped with the interstate highway network in the United



**Fig. 2.** Angler connection density (edges per  $\text{km}^2$ ) across all 18 major river basins of the contiguous United States and parts of Canada and Mexico. Black lines represent the interstate highway system of the United States. The numbers on the colored scale bar represent the upper value for angler connection density quintiles.

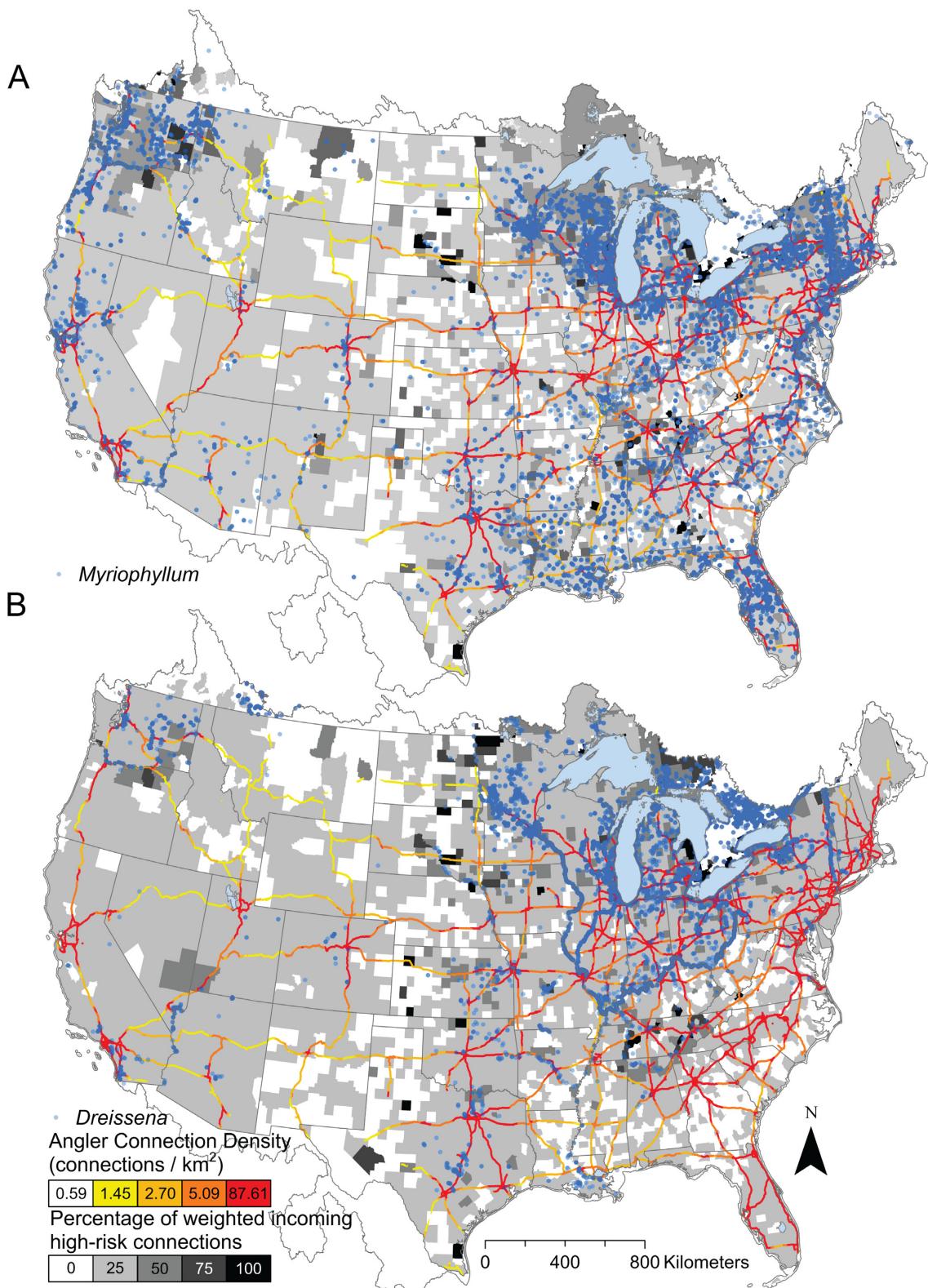
States (Figure 2). Average edge connection density was highest at  $< 20$  km from the US interstate system with an average of 7.1 connections/ $\text{km}^2$  and decreased with greater distance from the interstate. The average connection density at nonoverlapping bands at 40, 60, 80, and 100 km from the interstate highway was 4.3, 3.1, 2.4, and 1.9 connections/ $\text{km}^2$ , respectively. Parts of the interstate highway system that overlapped with the highest density of angler connectivity created the invasion superhighway (Figure 3). Lakes invaded by *Myriophyllum* (mean = 29.7 km) and *Dreissena* (mean = 40.2 km) species were closer to the invasion superhighway than all lakes (51.4 km;  $t(3298.5) = 32.477$ ,  $P < 0.001$ ;  $t(1673.1) = 7.4$ ,  $P < 0.001$ ). Half of all lakes that contained invasive milfoil or mussels were within  $\sim 19$  linear km of this “invasion highway” (Figure 4). Approximately 13% of movement was from invaded to unininvaded lakes, but this value varied spatially (Figure 3). High-risk connections, defined as angler connections from an invaded lake to an uninvaded lake, were rare in regions with few lakes and/or AIS (e.g. the center of the continent), and in regions with many lakes that contained AIS (e.g. central Florida in the case of milfoil). High-risk connections involving both milfoil and mussels were common along potential invasion fronts that spanned the center of the continent and the edge of the Appalachian Mountains in the Southeast (Figure 3). High-risk connections involving milfoil were also common in the Northwest (Figure 3A).

## Discussion

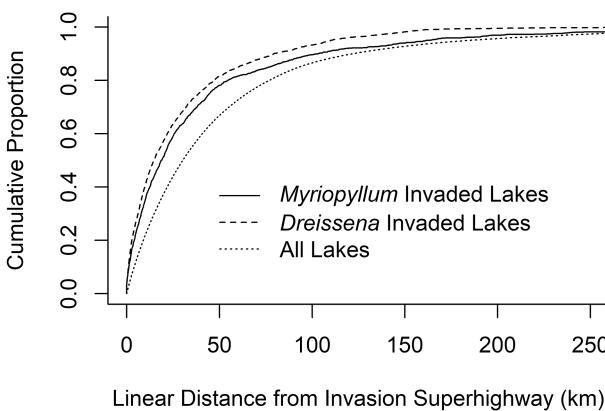
Our study is the first to reveal that movements by anglers—particularly over short-distances—combine to form an “invasion superhighway” that spans the contiguous United States, is fundamentally different from the natural network of rivers and streams, and provides important insight into the current and potential future distribution of AIS. Our results are consistent with local and regional studies that used survey, interview, and inspection data

to reveal highly connected networks that are dominated by movements over short distances and timeframes, cross natural and political boundaries, and overlap with transportation networks and/or the distribution of AIS (e.g. (9–16, 33–36). However, our study used passive data from 2 to 3 orders of magnitude more lakes to reveal this network at an unprecedented spatial scale and spatiotemporal resolution. Our work is also consistent with and complements national analyses that identified overlapping hotspots of fishing demand and AIS diversity (e.g. the Great Lakes) (37) and showed that most anglers connect lakes over short distances and timeframes (18).

Detailed movement data from throughout the contiguous United States are critical to predicting and preventing the spread of AIS. For example, identifying lakes that are either superspreaders (highly connected and infected) or highly vulnerable (highly connected but uninfected) can help to prioritize resources for surveillance, education, control, and enforcement (38). Connection weight and direction are also proxies for propagule pressure and colonization rate, which can be combined with hydrologic connectivity (39), establishment probability (40–42), and projected impacts (43) to assess risk, and optimize prevention and control across a landscape (44, 45). Whereas these approaches have traditionally been limited by a paucity of movement data, apps that track user location appear to be an abundant, cost-effective, and relatively comprehensive data source for any area of interest. Indeed, we were able to identify the lakes that are likely to be the largest superspreaders and most vulnerable in all 48 states of the contiguous United States (Table S1, Supplementary Material) and identify possible species-specific invasion fronts. Data of this nature enable interjurisdictional coordination that is essential to AIS control (45), but coordination can be difficult to achieve with conventional data. App-based movement data create opportunities for exciting avenues of AIS research. For example, the presence of invasion fronts in regions that have many infected lakes and high fishing effort (e.g. the Great Lakes) is an opportunity to



**Fig. 3.** The invasion superhighway combining invasion pathways along major highways in North America. Maps include the point distribution of invasive (A) *Myriophyllum* species and (B) *Dreissena* species. Overlapping points appear with greater color intensity. Counties (United States) and census units (Canada) are in grayscale to represent the percentage of weighted, incoming connections that are presumably at high-risk because they originated from an invaded lake (by one or both *Dreissena* species and *Myriophyllum* species) and terminated at an uninvaded lake.



**Fig. 4.** The cumulative proportion of all lakes > 0.06 ha, lakes invaded by *Myriophyllum*, and lakes invaded by *Dreissena* with increasing distance (km) from the invasion superhighway. Distances occurred at up to 537 km (not shown).

identify lake characteristics that contribute to resiliency to AIS. Scaling studies to a river basin could promote further exploration of the variation of activity and connectivity that occurs across the United States. Angler app data also include activities that may be important to AIS dispersal but are rarely captured by traditional methods—for example, fishing from nonmotorized boats or from shore—and connections involving small and/or private lakes and ponds (46, 47). Angler app data include catches from rivers and streams but assigning these catches to linear features that are to be treated as nodes within a network framework is a challenge that requires further study. Other types of online digital data, like photos from social media and search volumes (48, 49), may provide insight into the characteristics that make sites attractive to visitors, and thereby influence the spread of AIS.

Research is also needed to characterize movement dynamics over time (e.g. seasonal and annual trends) (32) and by angler type (e.g. avid vs. occasional) (50). Finally, although crowd-sourced, social media data appear to be a reliable proxy for lake visits (30, 51) and our results are qualitatively consistent with local and regional studies (9–16, 33–36), we encourage comparative research to evaluate, calibrate, and validate user-generated movement data relative to both traditional (e.g. survey) (32, 52–54) and emerging (e.g. cell phone location data) (55, 56) approaches. Of particular importance is the extent to which both active and passive location data from app users are representative of the overall population (18).

Our analysis of Fishbrain data also highlights the important role of anglers in generating accurate data and preventing the spread of AIS. Whether or not anglers transport AIS between lakes depends, not only on AIS type, number, and life stage, but also on mitigating behaviors (e.g. cleaning or treating equipment, trip location and frequency) (46). The presence of an invasion superhighway that spans the contiguous United States suggests that angler awareness, education, and behavior remain widespread issues (15, 57, 58). Our interactive map for viewing the spatial details of our results [<https://doi.org/10.5281/zenodo.6388121>] can stimulate AIS research and control, but also serve as a valuable tool for angler education and outreach. Apps like Fishbrain also offer unique opportunities to engage directly with anglers to emphasize their shared responsibility in preventing the spread of AIS. The advantages of mobile devices as both sources of data and a means of engaging with stakeholders are probably general to all forms of tourism and recreation that contribute to the spread of invasive species (16, 59).

## Materials and Methods

### Study area

Our study focused on the 18 major river basins (i.e. level-2 hydrologic units; Figure 1) that make up the contiguous United States and extend into parts of Canada and Mexico. The hydrography of these basins are described by the US Geological Survey's (USGS) NHD (60), which we downloaded at the highest resolution available (1:24,000). Our network analysis utilized the watershed boundary, flowline, and waterbody layers.

### Network development and comparison

#### The angler network

We obtained time, date, and geographic location data of catch records for freshwater and unknown species reported by recreational anglers between 2011 January 1 and 2021 June 14 using Fishbrain, a smartphone app that is popular among anglers. Users reported an exact catch location via a pin icon that defaulted to user location at the time of logging, and/or an approximate catch location that was based on the coordinates of the waterbody that was closest to the exact location. Users could override the approximate location by searching for the waterbody's name in the app's database. We prioritized exact locations over approximate locations in our analysis because they were more precise and made errors easier to detect.

The exact coordinate location of each catch was plotted using ArcGIS® Pro Version 2.8.2 (ESRI, Redlands, CA), and then assigned to an NHD waterbody (the fundamental unit in our network) if it occurred inside of the polygon boundary, or within 50 m of the boundary. If the exact location coordinates could not be assigned to a waterbody, then the set of coordinates for the approximate location was plotted and assigned using the same methods. We restricted our analysis to fishable lakes, defined as NHD waterbodies that were classified as a lake, pond, or reservoir with a surface area > 0.06 ha (half the size of an Olympic swimming pool). We excluded catches that could not be reliably assigned to a discrete waterbody feature like a lake. This excluded catches from linear features, such as rivers and streams.

The connections in our angler network were created by the movement of individual users between two lakes. Each lake in the resulting connectivity network was represented by a node, and lakes visited by the same angler were connected by an edge (see terms and definitions in Table 1). Each connection was composed of a direction (Lake A to Lake B or vice versa) and a weight ( $w$ ) that decreased from 1 to 0 with the number of days between lake visits ( $d$ ) according to the log-logistic curve (Figure S3, Supplementary Material):

$$w = \begin{cases} \frac{1.023}{1 + (\frac{d}{5.341})^{1.031}}, & d < 99 \\ 0, & d \geq 99 \end{cases}.$$

We derived this function from the results of a literature review of the number of days that various life stages of different invasive plants, invertebrates, and fishes are likely to survive transport by an angler (e.g. in a live well, bait bucket, or attached to a boat or trailer) regardless of temperature (Table S2, Supplementary Material). The resultant decline in the proportion of invasive life stages that we can expect to survive over time is a reasonable proxy for the time-dependent risk of transferring an invasive species. We then summed weights across all directed movements for each pair of lakes. Each connection therefore estimated propagule pressure (the cumulative sum of time-weighted movements) in either direction. Data were unavailable to further weight

according to high-risk activities such as fishing with live instead of artificial bait (61).

Each edge was represented by a straight-line path between the midpoints of two lakes that are connected by angler movement. Connection distance was the length of each edge in km. We removed connections that suggested same day travel with a distance > 250 km, next day travel > 500 km, and 2-day travel > 1,500 km to mitigate errors in user reported location or time. Although same-day, long-distance travel by air is possible, this mode of transportation typically involves a small amount of dry gear, and is, therefore, unlikely to be a major vector of AIS (but see (62) regarding floatplanes). We removed connections for two lakes sharing the same name due to errors that were created when a user selected the wrong lake among search results.

### The hydrologic network

We used the physical stream connections (flowlines) in the NHD to develop a weighted and directed network of hydrologic connectivity for the same set of lakes that comprised the angler network. Edge connections included streams, rivers, or culverts under a roadway. We ignored flow in ditches, small canals, karst systems, pipelines, and underground conduits because these features are nonpermanent, subterranean, impeded by grates, and/or highly artificial, and therefore unlikely to contribute significantly to the passage of AIS (63). We merged any two or more clustered lakes that were separated by < 40 m of permanent stream into a single node for the network analysis. Although we developed the merge rule to rejoin lakes that were bisected by roadways (e.g. Lake Minnetonka; 44.9443°, -93.5797°), we also applied it to natural streams (e.g. 48.0780°, -91.7863°) and lakes that were separated by narrows (e.g. 48.0414°, -91.6990°) because they are ineffective barriers to the passive or autonomous movement of AIS (64). Lakes that were merged in this step were then combined in the angler network, which removed any angler connections between the clustered lakes and moved the midpoint for distance measures to the largest lake polygon.

We identified all first-order, downstream connections between lakes, and weighted these connections according to the USGS classification-based estimate of the proportion of the year that a stream is flowing: 1.00 for perennial and unclassified streams, 0.75 for intermittent streams that flow for most weeks or months, and 0.10 for ephemeral streams that flow for hours or days in response to local rainfall. This approach assumes that the dispersal of AIS is predominantly downstream. Some aquatic organisms can disperse upstream (64), but this ability is limited to certain taxa (65), and impeded by natural and physical barriers (63).

We used the igraph package in R version 3.6.2 to calculate global network statistics for both networks including number of nodes, number of edges, % of isolates, number of components, component size, diameter, density, and reciprocity (66). We also calculated connectivity measures for each node and identified hubs. See Table 1 for a definition of each measure and its relevance to the spread of AIS.

### Invasion superhighway and the known distribution of AIS

By visualizing the density of angler connections in our network, we identified pathways along which AIS are likely to be spread through recreational fishing. We used the line density tool in ArcGIS Pro to calculate a raster layer with values corresponding to the density of edges within a 20-km search radius of each 5 km cell. We used the weight of connections to determine the number

of times the line should be counted. We excluded areas in the lowest 20% of density values to highlight the most important invasion pathways. US interstate highways (67) were overlaid with the corresponding densities from the map to form the invasion superhighways. We tested whether the density of angler movement was higher near the US interstate highway system by evaluating the mean density of angler movement in nonoverlapping polygon bands around the interstate highway that terminated at 20, 40, 60, 80, and 100 km from the interstate highway.

We evaluated the known distribution of invasive *Myriophyllum* and *Dreissena* species relative to the invasion superhighway. We compiled coordinate point location of voucherized specimens and reported sightings of these species from the USGS Nonindigenous Aquatic Species Program (NAS) the Early Detection and Distribution Mapping System (EDDMapS), and the USGS Biodiversity Information Serving Our Nation (BISON) (68–70). We combined all exotic *Dreissena* species to include zebra mussels (*D. polymorpha*) and quagga mussels (*D. rostriformis bugensis*), and all exotic *Myriophyllum* species to include Eurasian watermilfoil (*M. spicatum*), parrot feather (*M. aquaticum*), and hybrid milfoil (*M. spicatum* × *M. sibiricum*). An NHD lake was considered as invaded by one or both species if a coordinate point occurred inside or within 50 m of the lake boundary. We used the geodesic distance in the Near tool in ArcGIS Pro to calculate a straight-line distance from every invaded lake to the invasion superhighway and determined the proportion of invaded lakes near to the invasion superhighway. We tested whether invaded lakes were closer to the invasion superhighway compared to all lakes > 0.06 ha by comparing the mean distance of invaded lakes and all lakes from the invasion superhighway using a Welch two-sample t test.

We determined areas of the United States that are at a high-risk for invasions by looking at connections that were from an invaded lake going to an uninvaded lake. We used the mid-point of the receiving lake to assign it to a county in the United States (67) or census geographic units in southern Canada (71). We summed the weights of all connections that terminated in the county and summed the weights of all connections that were terminated at an uninvaded lake in which the origin lake was invaded by one or both *Dreissena* and *Myriophyllum*. We then calculated the percentage of weighted incoming connections that were at high-risk (invaded to uninvaded) for each county.

### Acknowledgments

We thank Hannah Tidbury and three anonymous reviewers for providing valuable feedback on the manuscript, and Spencer Davis and Angela Gibson for providing technical support.

### Supplementary Material

Supplementary material is available at [PNAS Nexus](#) online.

### Funding

This work utilized Ball State University's Beowulf cluster, which is supported by The National Science Foundation (MRI-1726017) and Ball State University, Muncie, Indiana.

### Authors' Contributions

Designed the research: J.W., K.V., J.B., A.B., K.H., C.S., J.A., and P.V.; performed the research: J.W., K.V., J.B., A.B., K.H., C.S., and P.V.;

analyzed the data: J.W., K.V., A.B., and P.V.; and wrote the paper: J.W., K.V., J.B., A.B., K.H., C.S., and P.V.

## Data Availability

The raw spatiotemporal data for this project were generated by users of the Fishbrain app and obtained through a data-sharing agreement. Requests for raw data should be directed to Johnan Attyby of Fishbrain at [johan@fishbrain.com](mailto:johan@fishbrain.com). Derived data supporting the findings of this study and interactive maps are openly available and hosted on Zenodo at <https://doi.org/10.5281/zenodo.6388121>.

## References

1. Wilson JR, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM. 2009. Something in the way you move: dispersal pathways affect invasion success. *Trends Ecol Evol.* 24:136–144.
2. Strayer DL. 2010. Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biol.* 55:152–174.
3. Gallardo B, Clavero M, Sánchez MI, Vilà M. 2016. Global ecological impacts of invasive species in aquatic ecosystems. *Glob Change Biol.* 22:151–163.
4. Dueñas MA, Hemming DJ, Roberts A, Diaz-Soltero H. 2021. The threat of invasive species to IUCN-listed critically endangered species: a systematic review. *Glob Ecol Conserv.* 26:e01476.
5. Moyle PB, Light T. 1996. Biological invasions of fresh water: empirical rules and assembly theory. *Biol Conserv.* 78:149–161.
6. Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ.* 52:273–288.
7. Bebber DP, Ramotowski MAT, Gurr SJ. 2013. Crop pests and pathogens move polewards in a warming world. *Nat Clim Change.* 3:985–988.
8. Diagne C, et al. 2021. High and rising economic costs of biological invasions worldwide. *Nature.* 592:571–576.
9. Padilla DK, Chotkowski MA, Buchan LAJ. 1996. Predicting the spread of zebra mussels (*Dreissena polymorpha*) to inland waters using boater movement patterns. *Glob Ecol Biogeogr Lett.* 5:353–359.
10. Buchan LAJ, Padilla DK. 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecol Appl.* 9:254.
11. Johnson LE, Ricciardi A, Carlton JT. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. *Ecol Appl.* 11:1789–1799.
12. Drake DAR, Mandrak NE. 2010. Least-cost transportation networks predict spatial interaction of invasion vectors. *Ecol Appl.* 20:2286–2299.
13. Murray CC, Pakhomov EA, Therriault TW. 2011. Recreational boating: a large unregulated vector transporting marine invasive species. *Diver Distrib.* 17:1161–1172.
14. Kelly NE, Wantola K, Weisz E, Yan ND. 2013. Recreational boats as a vector of secondary spread for aquatic invasive species and native crustacean zooplankton. *Biol Invas.* 15:509–519.
15. Cole E, Keller RP, Garbach K. 2019. Risk of invasive species spread by recreational boaters remains high despite widespread adoption of conservation behaviors. *J Environ Manage.* 229:112–119.
16. Smith ERC, Bennion H, Sayer CD, Aldridge DC, Owen M. 2020. Recreational angling as a pathway for invasive non-native species spread: awareness of biosecurity and the risk of long distance movement into Great Britain. *Biol Invas.* 22, 1135–1159.
17. Venturelli PA, Hyder K, Skov C. 2017. Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards. *Fish Fish.* 18:578–595.
18. Fricke RM, Wood SA, Martin DR, Olden JD. 2020. A bobber's perspective on angler-driven vectors of invasive species transmission. *NeoBiota.* 60:97–115.
19. Cessford G, Muhar A. 2003. Monitoring options for visitor numbers in national parks and natural areas. *J Nat Conser.* 11, 240–250.
20. Freeman AM, III, Herriges JA, King CL. 2014. The measurement of environmental and resource values: theory and methods. 3rd ed. Washington (DC): Resources for the Future Press. p. 6. doi: 10.4324/9781936331826.
21. González MC, Hidalgo CA, Barabási AL. 2008. Understanding individual human mobility patterns. *Nature.* 453:779–782.
22. Jurdak R, et al. 2015. Understanding human mobility from Twitter. *PLoS ONE.* 10:1–16.
23. Pew Research Center. Mobile fact sheet. Washington (DC). 2021.
24. Peixoto PS, Marcondes D, Peixoto C, Oliva SM. 2020. Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS ONE.* 15:1–23.
25. Smolinski MS, et al. 2015. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *Am J Publ Health.* 105:2124–2130.
26. Zhang W, Chong Z, Li X, Nie G. 2020. Spatial patterns and determinant factors of population flow networks in China: analysis on Tencent location big data. *Cities.* 99:102640.
27. Fox N, et al. 2020. "photosearcher" package in R: an accessible and reproducible method for harvesting large datasets from Flickr. *SoftwareX.* 12:100624.
28. Allain SJR. 2019. Mining Flickr: a method for expanding the known distribution of invasive species. *Herpetolo Bull.* 148:11–14.
29. Barve V. 2014. Discovering and developing primary biodiversity data from social networking sites: a novel approach. *Ecol Inf.* 24:194–199.
30. Keeler BL, et al. 2015. Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Front Ecol Environ.* 13:76–81.
31. Donahue ML, et al. 2018. Using social media to understand drivers of urban park visitation in the Twin Cities, MN. *Landsc Urban Plan.* 175:1–10.
32. Papenfuss JT, Phelps N, Fulton D, Venturelli PA. 2015. Smartphones reveal angler behavior: a case study of a popular mobile fishing application in Alberta, Canada. *Fisheries.* 40:318–327.
33. Muirhead JR, MacIsaac HJ. 2005. Development of inland lakes as hubs in an invasion network. *J Appl Ecol.* 42:80–90.
34. Stewart-Koster B, Olden JD, Johnson PTJ. 2015. Integrating landscape connectivity and habitat suitability to guide offensive and defensive invasive species management. *J Appl Ecol.* 52: 366–378.
35. Escobar LE, Romero-Alvarez D, Larkin DJ, Phelps NBD. 2019. Network analysis to inform invasive species spread among lakes. *J Oceanol Limnol.* 37:1037–1041.
36. Herborg LM, O'Hara P, Therriault TW. 2009. Forecasting the potential distribution of the invasive tunicate *Didemnum vexillum*. *J Appl Ecol.* 46:64–72.
37. Davis AJS, Darling JA. 2017. Recreational freshwater fishing drives non-native aquatic species richness patterns at a continental scale. *Diver Distrib.* 23:692–702.
38. Morandi MJ, Manning NF, Bossenbroek JM, Jerde CL. 2015. Assessing the influence of different inland lake management

- strategies on human-mediated invasive species spread. *Manag Biol Invas.* 6:57–69.
39. Filipe AF, Quaglietta L, Ferreira M, Magalhães MF, Beja P. 2017. Geostatistical distribution modelling of two invasive crayfish across dendritic stream networks. *Biol Invas.* 19:2899–2912.
  40. Drake JM, Bossenbroek JM. 2004. The potential distribution of zebra mussels in the United States. *Bioscience.* 54:931–941.
  41. Poulos HM, Chernoff B, Fuller PL, Butman D. 2012. Ensemble forecasting of potential habitat for three invasive fishes. *Aqu Invas.* 7:59–72.
  42. Lovell RSL, Blackburn TM, Dyer EE, Pigot AL. 2021. Environmental resistance predicts the spread of alien species. *Nat Ecol Evol.* 5:322–329.
  43. Kulhanek SA, Ricciardi A, Leung B. 2011. Is invasion history a useful tool for predicting the impacts of the world's worst aquatic invasive species? *Ecol Appl.* 21:189–202.
  44. Vander Zanden MJ, Olden JD. 2008. A management framework for preventing the secondary spread of aquatic invasive species. *Can J Fish Aquat Sci.* 65:1512–1522.
  45. Epanchin-Niell RS, et al. 2010. Controlling invasive species in complex social landscapes. *Front Ecol Environ.* 8:210–216.
  46. Anderson LG, White PC, Stebbing PD, Stentiford GD, Dunn AM. 2014. Biosecurity and vector behaviour: evaluating the potential threat posed by anglers and canoeists as pathways for the spread of invasive non-native species and pathogens. *PLoS ONE.* 9:1–10.
  47. Niemiec RM, Ardoine NM, Wharton CB, Asner GP. 2016. Motivating residents to combat invasive species on private lands: social norms and community reciprocity. *Ecol Soc.* 21:30.
  48. Jarić I, et al. 2020. iEcology: harnessing large online resources to generate ecological insights. *Trends Ecol Evol.* 35:630–639.
  49. Lennox RJ, et al. 2022. Digital fisheries data in the Internet age: emerging tools for research and monitoring using online data in recreational fisheries. *Fish Fish.* 00:1–15.
  50. Kane DS, Kaemingk MA, Chizinski CJ, Pope KL. 2020. Spatial and temporal behavioral differences between angler-access types. *Fish Res.* 224:105463.
  51. Wood SA, Guerry AD, Silver JM, Lacayo M. 2013. Using social media to quantify nature-based tourism and recreation. *Sci Rep.* 3:1–7.
  52. Gundelund C, Skov C. 2021. Changes in angler demography and angling patterns during the Covid-19 lockdown in spring 2020 measured through a citizen science platform. *Mar Pol.* 131:104602.
  53. Jiorle RP, Ahrens RNM, Allen MS. 2016. Assessing the utility of a smartphone app for recreational fishery catch data. *Fisheries.* 41:758–766.
  54. Johnston FD, Simmons S, van Poorten BT, Venturelli PA. 2021. Comparative analyses with conventional surveys reveal the potential for an angler app to contribute to recreational fisheries monitoring. *Can J Fish Aquat Sci.* 79:31–46.
  55. Merrill NH, Atkinson SF, Mulvaney KK, Mazzotta MJ, Bousquin J. 2020. Using data derived from cellular phone locations to estimate visitation to natural areas: an application to water recreation in New England, USA. *PLoS ONE.* 15:1–19.
  56. Monz C, Mitrovich M, D'Antonio A, Sisneros-Kidd A. 2019. Using mobile device data to estimate visitation in parks and protected areas: an example from the nature reserve of orange county, California. *J Park Recreat Admin.* 37:1–19.
  57. Rothlisberger JD, Chadderton WL, McNulty J, Lodge DM. 2010. Aquatic invasive species transport via trailered boats: what is being moved, who is moving it, and what can be done. *Fisheries.* 35:121–132.
  58. Pradhananga A, Davenport MA, Seekamp E, Bundy D. 2015. Preventing the spread of aquatic invasive species: boater concerns, habits, and future behaviors. *Hum Dimen Wildlife.* 20:381–393.
  59. Anderson LG, Rocliffe S, Haddaway NR, Dunn AM. 2015. The role of tourism and recreation in the spread of non-native species: a systematic review and meta-analysis. *PLoS ONE.* 10:1–15.
  60. U.S. Geological Survey. 2019. National hydrography dataset. Reston (VA).
  61. Ludwig HR, Leitch JA. 1996. Interbasin transfer of aquatic biota via Anglers' Bait Buckets. *Fisheries.* 21:14–18.
  62. Carey MP, Sethi SA, Larsen SJ, Rich CF. 2016. A primer on potential impacts, management priorities, and future directions for *Elodea* spp. in high latitude systems: learning from the Alaskan experience. *Hydrobiologia.* 777:1–19.
  63. Kerby JL, Riley SPD, Kats LB, Wilson P. 2005. Barriers and flow as limiting factors in the spread of an invasive crayfish (*Procambarus clarkii*) in southern California streams. *Biol Conserv.* 126:402–409.
  64. Warren ML, Pardew MG. 1998. Road crossings as barriers to small-stream fish movement. *Trans Am Fish Soc.* 127:637–644.
  65. Reynold C, Miranda NAF, Cumming GS, Keller R. 2015. The role of waterbirds in the dispersal of aquatic alien and invasive species. *Diver Distrib.* 21:744–754.
  66. Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal Comp Syst.* 1695:1–9.
  67. U.S. Census Bureau. 2018. TIGER/Line shapefiles (machine-readable data files) primary roads. Suitland-Silver Hill, MD. (June 12, 2020).
  68. U.S. Geological Survey. 2021. Nonindigenous aquatic species database. Washington (DC): U.S. Department of Interior. (April 29, 2020).
  69. Center for Invasive Species and Ecosystem Health. 2021. Early detection & distribution mapping system. Athens (GA): University of Georgia. (April 24, 2020).
  70. U.S. Geological Survey. 2021. Biodiversity information serving our nation. Washington (DC): U.S. Department of Interior. (April 24, 2020).
  71. Statistics Canada. 2016. Census subdivision boundary file. Ottawa (ON).