ORIGINAL PAPER

Network connectivity of Minnesota waterbodies and implications for aquatic invasive species prevention

Szu-Yu Zoe Kao · Eva A. Enns · Megan Tomamichel · Adam Doll · Luis E. Escobar · Huijie Qiao · Meggan E. Craft · Nicholas B. D. Phelps

Abstract Connectivity between waterbodies influences the risk of aquatic invasive species (AIS) invasion. Understanding and characterizing the connectivity between waterbodies through high-risk pathways, such as recreational boats, is essential to develop economical and effective prevention intervention to control the spread of AIS. Fortunately, state and local watercraft inspection programs are collecting significant data that can be used to quantify boater connectivity. We created a series of predictive models to capture the patterns of boater movements across all lakes in Minnesota, USA. Informed by more than 1.3 million watercraft inspection surveys from 2014–2017, we simulated boater movements connecting 9182 lakes with a high degree of accuracy. Our predictive model accurately predicted 97.36% of the lake pairs known to be connected and predicted 91.01% of the lake pairs known not to be connected. Lakes with high degree and betweenness centrality were more likely to be infested with an AIS than lakes with low degree ($p < 0.001$) and centrality ($p < 0.001$). On average, infested lakes were connected to 1200 more lakes than uninfested lakes. In

S.-Y. Z. Kao · E. A. Enns
Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN, USA

M. Tomamichel
Odum School of Ecology, University of Georgia — Athens, Athens, GA, USA

A. Doll
Minnesota Department of Natural Resources, St. Paul, MN, USA

L. E. Escobar
Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VT, USA

H. Qiao
Key Laboratory of Animal Ecology and Conservation Biology, Chinese Academy of Sciences, Beijing, China

M. E. Craft
Department of Veterinary Population Medicine, College of Veterinary Medicine and Department of Ecology, Evolution and Behavior, College of Biological Sciences, University of Minnesota, St. Paul, MN, USA

N. B. D. Phelps (✉)
Minnesota Aquatic Invasive Species Research Center, University of Minnesota, St. Paul, MN, USA
e-mail: phelp083@umn.edu

N. B. D. Phelps
Department of Fisheries, Wildlife and Conservation Biology, College of Food, Agriculture and Natural Resource Sciences, University of Minnesota, 2003 Upper Buford Circle, St. Paul, MN 55108, USA

addition, boaters that visited infested lakes were more likely to visit other lakes, increasing the risk of AIS spread to uninfested lakes. The use of the simulated boater networks can be helpful for determining the risk of AIS invasion for each lake and for developing management tools to assist decision makers to develop intervention strategies.

Keywords Aquatic invasive species · Boater movements · Network analysis · Network features · Machine learning

Introduction

Aquatic invasive species (AIS) cause harm to native biodiversity, economies, and human health, and are widely considered among the world's greatest threats to marine and freshwater ecosystems (Bax et al. 2003; Ricciardi et al. 2017; Simberloff et al. 2013). Consequently, controlling populations and mitigating impacts of AIS has long been a priority for managers (Simberloff 2014). Keeping pace is a daunting task given the continued and rapid expansion of new invasions (Lodge et al. 2006). Preventing the introduction of new AIS into and within a region is therefore an economical and effective approach for management (Leung et al. 2002; Simberloff et al. 2013). However, designing appropriately tailored prevention efforts requires a thorough understanding of the mechanisms by which invasive species spread through human mediated and natural pathways. Many pathways exist for the translocation of AIS into new regions, such as ballast water in cargo ships (Bailey 2015), live baitfish used for angling (Drake and Mandrak 2014; Ludwig Jr and Leitch, 1996), aquaculture activities (Naylor et al. 2001), and the intentional or unintentional live release of specimens (Maki and Galatowitsch 2004; Padilla and Williams 2004). While attention to these pathways is critical, once an AIS is introduced, understanding the additional risks for secondary spread of the AIS within a region becomes a priority. One important path of secondary AIS spread is the overland movement of recreational boats from an infested waterbody to an uninfested waterbody (Rothlisberger et al. 2010; Vander Zanden and Olden 2008). Boater movement between waterbodies is a well-recognized risk factor for AIS

infestation (Buchan and Padilla 1999; De Ventura et al. 2016; Johnson and Carlton 1996; Vander Zanden and Olden 2008). Boater movements between waterbodies create an interconnected network that would otherwise be largely disconnected. The more connections a given waterbody has to and/or from other waterbodies, the greater its risk of AIS introduction and the greater its risk, if infested, of serving a source of secondary AIS introductions (Floerl et al. 2009; Muirhead and MacIsaac 2005). Understanding recreational boater movement networks is therefore critical to implementing strategic AIS prevention interventions. Quantifying the extent and patterns of recreational boater movements is challenging due to the lack of comprehensive, systematic data collection. Analyses of boater movements typically rely on survey data, which has been used previously to construct waterbody connectivity networks for a variety of contexts. In most cases, however, these studies have focused on select waterbodies where data could be most easily generated (Cole et al. 2019; Escobar et al. 2019; Haak et al. 2017). Although lacking the resolution necessary to understand landscape patterns, these studies have demonstrated the value of network approaches to informing prevention and outreach strategies for various AIS.

The state of Minnesota, in the Upper Midwest USA, is directing considerable resources to the prevention of secondary spread once an AIS has been introduced to the state (Escobar et al. 2017). Two species of particular concern that seem to be associated with recreational boat movement include zebra mussels (*Dreisenna polymorpha*; Campbell et al. 2016; Collas et al. 2018; Johnson and Carlton 1996; McCartney and Mallez 2018) and starry stonewort (*Nitelopsis obtusa*; Larkin et al. 2018; Sleith et al. 2015). Statewide prevention and early detection efforts for these species are complicated by the sheer numbers of licensed watercraft (n = ∼ 814,000; Minnesota Department of Natural Resrouces 2019) and waterbodies (n = 25,915; Minnesota Department of Natural Resources 2014). Therefore, evaluating sustainable methods of data collection and analytical tools to inform AIS prevention and early detection at the state-level has important implications regionally and nationally. We assessed the boater movement connectivity patterns between Minnesota waterbodies by analyzing data from the Minnesota Watercraft Inspection Survey from 2014 to 2017 (Minnesota

Department of Natural Resources 2020). We described statewide and lake-specific features of the resulting waterbody network and identified lakes that may have important roles in the network. Finally, we discussed the implications of connectivity for AIS prevention and early detection, and the challenges of using boater data to infer lake connectivity patterns of AIS risk. Our analytical approach is applicable to other aquatic invaders and to other geographic regions.

Methods

We focus our analysis to boater movements between Minnesota waterbodies classified as Type 5 surface water (inland open fresh water) with a size of at least 10 acres, herein referred to as lakes. While this excludes small bodies of water and wetlands, it does include the vast majority of waterbodies relevant to recreational boating and risk of invasion through this pathway. Indeed, 96% of confirmed zebra mussel infestations are reported in lakes classified as Type 5 (Minnesota Department of Natural Resources 2020). Using data collected by the Minnesota Department of Natural Resources (MNDNR) Watercraft Inspection Program, which asks boaters about past and upcoming boating excursions at select inspected lakes, we inferred a statewide boater movement network among lakes. The weighted network captures both the overall connectivity pattern (which lakes are connected to which) as well as the strength of those connections (the number of boaters moving from one lake to another each year).

Survey data

We analyzed data from the MNDNR Watercraft Inspection Program (Minnesota Department of Natural Resources 2020), collected between April 25, 2014 and December 01, 2017. The general goals of this inspection program are to develop (i) an AIS educational opportunity for recreational boaters, (ii) an inspection opportunity of watercraft for the presence of AIS, and (iii) systematized data collection. Watercraft inspectors interview boaters at public access points during the open-water seasons as they enter or exit a lake and can require the boaters to decontaminate the watercraft if certain invasive species are found on the watercraft. The location and effort of

watercraft inspectors is determined based on waterbody popularity, infestation status, and other factors determined by local and state managers. Therefore, waterbody inspections are not evenly distributed in space or time, which is a challenge for generalizing the inspection survey data in their raw form. A total of 737 lakes had some level of inspection effort from 2014 to 2017 and appear in the survey dataset.

During inspections, boaters were asked the following questions regarding past and future boating activity: "What was the *last* waterbody visited?" and "Where are you planning to visit *next* after today's trip?" This included name and county of the relevant waterbody(-ies) visited or to visit. If the current boating activity was their first or last for the season, boaters could respond that their watercraft was coming from or heading to storage. Boaters were also given the option to not answer these questions or to respond that they did not know. Many inspections involved lake service providers (LSPs), which are organizations that rent watercraft and/or launch or remove watercraft from lakes on behalf of owners, who may not know the specific history of the given watercraft. These responses were often coded as "LSP" and were treated as missing data as no origin/destination information was provided in these inspections. Destinations outside of Minnesota were simply coded as out of state. Due to the protocol of inspecting as many watercraft as possible upon entry or exit, the same boater may be surveyed twice (both upon entry and exit) on the same trip. To avoid double counting these repeat respondents, we removed entry survey responses that matched an exit inspection survey response on date of inspection, inspection location, vehicle license plate, and watercraft type. We constructed a boater movement dataset using boater responses to past (recalled) and future (planned) boating activity on Minnesota lakes. Each lake in Minnesota can be identified by a Division of Waters (DOW) number, which are unique and permanent identifiers assigned to all waterbodies in Minnesota. Inspection locations in the survey data were identified using DOW numbers. However, boaters were asked to select the lake name and county of their past and future boating activity from drop-down menus on the tablet within the electronic survey tool. We therefore searched these lake/county pairs in the MNDNR Lake Finder Database (https://www.dnr.state.mn.us/lakefind) to identify the corresponding DOW

number. The electronic survey tool did not restrict the selection of lake/county pairs, instead allowing the boater to independently select the lake and county from drop down menus. This resulted in many lake/-county responses that could not be matched in the MNDNR Lake Finder Database. The first type of reporting error occurred when the boater incorrectly identified the county. In this case, we hypothesized that the lake name would appear in a neighboring county, as boaters may not know the exact locations of county lines or a lake may span multiple counties but be assigned to a single county for identification purposes. To correct this potential error, we developed a 2–step matching algorithm to re-assign the erroneous boater responses. First, if the exact lake name could be found in a neighboring county, we replaced the specified county in the survey with the neighboring county. If the lake name appeared in more than one neighboring county, we selected the county containing the largest lake (in square kilometers) of the same name. For the remaining unmatched lake/county pairs, we considered two possibilities: (1) a similar lake name appears in the specified county; or (2) the exact lake name can be found in another, non-neighboring county. If more than one potential match occurred, we selected the lake/county pair corresponding to the largest lake among the matches. If lake/county pairs could not be resolved with the first type of error, we considered a second type of reporting error where the boater correctly identified the county but selected a lake in the dropdown menu that had a similar, but not identical, name as the one they intended. In this case, we hypothesized that a similar lake name would appear in the county specified or a neighboring county. The similarity of two lake names was evaluated by counting the number of matching words, excluding a set of common words, such as "the", "of", and certain adjectives like "north" or "south". For example, a lake name of "St. Louis" would have a similarity score of 1 to a lake named "Louis" and 2 to a lake named "St. Louis Bay", which could plausibly reflect the boaters intended lake if no direct name matches exist.

We summarized the boater movement dataset in terms of which lakes were reported as being connected through sequential boater activity and the number of boaters reporting that movement. For example, if a boater was surveyed at Lake A (inspection lake) and reported that the last lake they visited prior to this trip

was Lake B, then we would note a connection from Lake B to Lake A. These movement connections were directional, and the connections were weighted by the number of boaters reporting that particular movement.

Predictive models of boater movements

The boater inspection survey data only captured movements to or from inspected lakes, which constitute a small subset of the total number of lakes with recreational boating use in Minnesota. Furthermore, inspection effort and access popularity varied across lakes, making the number of boater movements to and from one inspected lake difficult to compare against the number of movements reported at a different inspected lake without some kind of adjustment. To adjust for varying levels of inspection effort and extrapolate the boater movement patterns observed at inspected lakes to all lakes in Minnesota, we constructed a series of predictive models using the eXtreme Gradient Boosting (XGBoost) machine learning algorithm (Chen and Guestrin 2016; Chuan et al. 2018; Nielsen 2016; Wang et al. 2019). XGBoost is widely used and has good performance in predictive modeling competitions on Kaggle, which is a data science community. This algorithm balances performance with over-fitting issues (Chen and Guestrin 2016; James et al. 2013). We first fit a model to predict the total arrival boater traffic at each lake (M1), followed by a model to predict the existence of boater movements between each pair of lakes (M2). We lastly fit a model to predict the number of boaters traveling between each pair of connected lakes (M3). All the predictive models were fit on an annual time scale. To predict the total boater traffic on each lake (M1), we analyzed detailed inspection hour logs for the 2017 season. Inspection effort varied among lakes, which potentially introduced bias. To mitigate this bias, we calculated the number of surveyed boaters per inspection-hour for each inspected lake and converted this value to an annual scale assuming 12 h of daily activity for a 178–day season. The total number of days in a boating season (178) was calculated by taking the maximum difference between the earliest and latest inspection dates among all inspected lakes in 2017. We fit an XGBoost model (M1) using regression with a tree-based booster to inspection data from 452 lakes that had the records of inspection hours in 2017 and used the model to predict the total boater traffic for

the rest of the 8730 lakes based on lake attributes. Attributes included in the model were the lake size, distance to the nearest highway, whether the lake had public docks or ramps, number of lakes within 30 km radius, infestation status (as of December 7, 2018), county population size, and the number of ramps (Minnesota Department of Natural Resources 2014,2018,2020; Minnesota Department of Transportation 2012; United States Census Bureau 2017). We conducted threefold cross-validation to tune the hyperparameters of the XGBoost model randomly using 90% of data as the training set and the remaining 10% as the test set to evaluate model performance. The hyperparameters are the parameters controlling for the learning process of XGBoost (Supplemental Table S1). We selected the model with the best performance according to root mean square error (RMSE). M2 was constructed to predict the existence of a boater connection between any pair of the lakes. The past and future water activities of each inspection informed whether two lakes were connected by boaters and the direction of that connection. We assigned the lake pairs connected by boaters in the inspection survey as ones and the lake pairs that were not mentioned in the survey as zeros. Therefore, we were able to construct an XGBoost model to classify binary outcomes. For M2, we included features regarding lake attributes of the originating lake, attributes of the destination lake, and many interaction terms of the attributes of both origin and destination lakes. Attributes included lake size, infestation status, number of ramps, inspection status, distance between the two lakes, and whether the two lakes were more than 30 km apart. Similarly, we conducted a grid search to tune the hyperparameters (Supplemental Table S1) using a threefold cross-validation among 80% of the boater movement data as the training set. The optimal hyperparameters were set to maximize the area under the curve (AUC) metric in predicting the existence of a boater connection between all pairs of lakes. The output of the XGBoost algorithm provides predicted probabilities of a connection existing between two lakes. As is typical in these predictive models, if the predicted probability of two lakes being connected was greater than or equal to 0.5, those two lakes were modeled as being connected through boater movements (Bishop 2006; Chen and Guestrin 2016; James et al. 2013). We did not sample based on the probabilities directly, as this would

attribute more precision to the probability estimates than would be appropriate given the limitations of the data used to produce these estimates. In particular, we know that not all lake-to-lake connections are observed in the data, even for surveyed lakes: the boater survey is only a sample of boater movements and is far from capturing the majority of boater traffic at any given lake. To predict the number of boaters traveling between two connected lakes (M3), we constructed an XGBoost model using regression and a tree-based booster accounting for the lake attributes of origin and destination lakes. The features included in the model were similar to those in M2. We also trained the model using a threefold cross validation process fit to 80% of the data set. The optimal hyperparameters were selected to minimize the RMSE. The selected models are reported in Supplemental Table S1 and the features used are presented in (Supplemental Table S2). All of the XGBoost models were constructed using the xgboost and Scikit-learn packages in Python 3.7.3 (Chen and Guestrin 2016; Papenfuss et al. 2015; van Rossum 1995).

Generating a boater movement network

We generated plausible boater movement networks between all Type 5 lakes $\geq$ 10 acres in the state of Minnesota by sampling from the predictive models fit to the boater movement data. The process of the simulated boater movement networks is presented in Fig. 1. We used M1 to predict the total incoming boater traffic on each lake. Using M2, we were able to predict whether or not a pair of lakes were connected via boater movements. If M2 predicted a probability of connection greater than 0.5 for a pair of lakes, the lakes were connected in the network model. If the predicted probability was less than 0.5, the lakes were randomly connected following a binomial process using the predicted probability from M2 as a proxy to allow the connection to exist by chance. Once the connectivity pattern was constructed, we then used M3 to predict the number of boaters moving along those connections. To maintain the level of boater traffic predicted with M1, we used the M3 predictions to calculate the proportion of a lake's incoming traffic from each connection and distributed the level of incoming traffic predicted by M1 accordingly. We repeated the sampling process to generate 20 plausible boater movement networks between lakes in the state
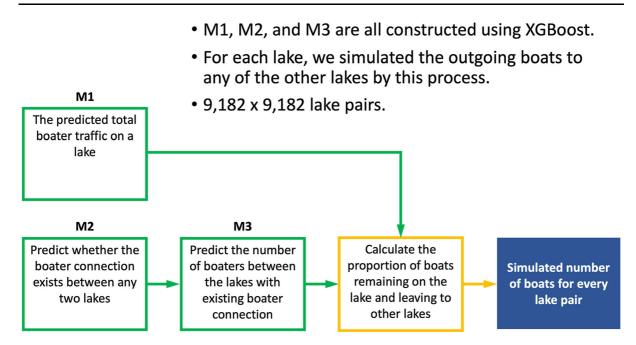
- M1, M2, and M3 are all constructed using XGBoost.
- For each lake, we simulated the outgoing boats to any of the other lakes by this process.
- 9,182 x 9,182 lake pairs.

M1

The predicted total boater traffic on a lake

M2

Predict whether the boater connection exists between any two lakes

M3

Predict the number of boaters between the lakes with existing boater connection

Calculate the proportion of boats remaining on the lake and leaving to other lakes

Simulated number of boats for every lake pair

Fig. 1 Modeling process of a simulated boater movement network. Predictive models M2 and M3 predicted the distribution of boaters remaining on a lake and leaving to each of the other lakes. To calculate the number of outgoing boaters from a lake to another lake, we multiplied the distribution informed by M1 and M2 by the predicted total boater traffic on a lake informed by predictive model M1

of Minnesota to capture some of the variability arising from the uncertainty in the predictive models.

Network analysis

We summarized lake connectivity patterns using standard network measures. We first identified whether the network was fully connected, meaning that there was a path from every lake to any other lake, or, if not, the size of the largest connected component. For each lake, we calculated a number of node-level characteristics that can reflect importance within the larger network structure. "Degree" is the number of lakes to which a given lake is connected; thus, highly connected lakes will have higher degree. "Betweenness centrality" is a numerical value that reflects the importance of a lake in maintaining the connectivity in the network. High centrality can indicate a high risk of AIS invasion or high potential for onward AIS spread. Numerically, the centrality of a given lake i is defined as:

$$\sum_{i \neq j \neq k} \frac{g_{jk}(i)}{g_{jk}} \tag{1}$$

where g_{jk} is the number of shortest paths between two nodes (j and k), and $g_{jk}(i)$ is the number of weighted shortest paths that go through node i. The edge weights in the boater movement network were set to the inverse of the number of boater movements; that is, the more boaters that move between two lakes, the shorter the "distance" between them in terms of AIS transportation risk. Centrality was calculated using the tnet R package (Opsahl 2009). We evaluated average network statistics overall, as well as stratified by infestation status. A lake was considered infested if it was listed on the MNDNR Infested Waters List as having any invasive species present as of December 7, 2018 (Minnesota Department of Natural Resources 2018).

Results

Boater inspection survey responses

From 2014 to 2017, there were 1,329,259 inspection surveys across 737 Type 5 lakes ≥ 10 acres in Minnesota. Each inspection survey yielded two

responses: one is about the previous boating activity and another is about the future activity (Table 1). There were more missing responses for future activity (n = 350,099) than for recalled activity (n = 118,058). Refusal was similarly higher for future (n = 6128) versus recalled (n = 5090) activity. There was a total of 13,093 responses from lake service providers, 28,601 reports of activity outside of Minnesota, and 3669 responses that the location of past activity was not on the list provided. A total of 41,306 boaters reported that their watercraft was coming from or heading to storage. None of these responses were used in the construction of the boater movement network.

The inspection surveys yielded a total of 1,789,292 reports of recent and upcoming boating activity within Minnesota after removing duplicate entry/exit inspections on the same day at the same lake (n = 303,181 responses). There were 235,506 (13.2%) boater responses that gave incorrect pairings of lake names and counties. After applying the matching algorithm for correcting mismatched lake/county pairs, this was reduced to 112,278 (6.3%) responses, which were excluded from further analyses. In total, 1,666,704 boater responses among Type 5 lakes ≥ 10 acres were included in the analysis to inform the boater movement network. Lakes that were inspected or mentioned in survey responses included a total of 2146 lakes (23% of the 9,182 lakes included in this study). Overall, 1,163,521 (69.8%) of the boaters reported past and/or future activity at the same lake as their current activity. Within boater survey responses, an inspected lake was reported to be connected to an average of 50 (SD = 89) unique other lakes through past or future activity.

Predictive models of boater traffic and connectivity

Using M1, we predicted a mean boater traffic of 413 boats per year across all Minnesota lakes, with a minimum and maximum predictions of 105 and 91,292 boats, respectively. Across 20 simulated networks, M2 accurately predicted an average of 97.36% (SD = 0.03%, range 97.31–97.41%) of the lake pairs known to be connected and 91.01% (SD = 0.005%, range 91.00–91.02%) of the lake pairs known not to be connected. The average prediction true positive rate (% of accurately predicted boater connections among observed connections) across inspected lakes was 97.43%. The RMSEs for the selected M1 and M3 were reported in (Supplemental Table S1). The correlation between the predicted and actual number of boats was 0.53 for M1 and 0.79 for M3 in the test set. Characteristics of the inspected lakes with higher (≥ 85%) and lower (< 85%) true positive rates are summarized in Table 2. We found that the inspected lakes with lower true positive rates differ from the lakes with high true positive rates in the average number of inspections per year, size, and the number of access points on the lake. This was to be expected because lakes that had more inspections and were more accessible had more data to inform their connectivity with other lakes than the lakes with fewer inspections and less accessibility.

Table 1 Summary of inspection survey responses to past and future boating activity questions from $n = 1,329,259$ boater inspection surveys from the 2014 through the 2017 boating seasons

Type of response	Response to past activity question	Response to future activity question	Total responses
	# (%)	# (%)	# (%)
Unknown / blank	118,058 (8.9%)	350,099 (26.3%)	468,157 (17.6%)
Refuse to answer	5090 (0.4%)	6128 (0.5%)	11,218 (0.4%)
Lake service provider	6527 (0.5%)	6566 (0.5%)	13,093 (0.5%)
Not on the list	2391 (0.2%)	1278 (0.1%)	3669 (0.1%)
Storage/first time	26,141 (2.0%)	15,165 (1.1%)	41,306 (1.6%)
Out of state	16,389 (1.2%)	12,213 (0.9%)	28,602 (1.1%)
Lake within Minnesota	1,154,663 (86.9%)	937,810 (70.6%)	2,092,473 (78.7%)

Table 2 Comparison of attributes of inspected lakes with high (≥ 85%; n = 726) and low (< 85%; n = 11) average true positive rates across all simulated boater movement networks using Wilconxon rank sum test

	Lake true positive rate		
	≥ 85%	< 85%	*p*-value
Inspection Effort			
Number of inspections per year	413	6	< 0.001
Number of inspections per hour	461	76	0.266
Lake characteristics			
Acres	4,761	649	< 0.001
Number of lakes within 30 km	259	343	0.209
Number of ramps	1.55	1.00	0.334
Number of docks	1.08	.00	0.007
Number of parking lots	1.27	1.00	0.518
Number of parking spaces	18.24	45.00	0.419
Number of public access points	1.41	1.00	0.393

Boater movement network

Across 20 simulated networks, lakes were connected to an average of 267 (SD = 484) other lakes through boater movements (Table 3; data available on https://conservancy.umn.edu/handle/11299/216936). An average of 1.99 (SD = 66.89) boaters were predicted to move across each lake-lake connection per season; however, this varied greatly, with some lake-lake connections being predicted to be undertaken by over 68,000 boaters in a season. The top 10 most connected and top 10 highest centrality lakes in the boater movement networks are summarized in Table 4. Four lakes appeared on both lists. Together, these sixteen lakes accounted for 8.5% of boater movements and were collectively connected to 7392 other lakes in the network (80.5% of the total lakes in Minnesota).

Across both lists, all but two lakes were designated as infested with species listed by the MNDNR as of December 7, 2018. We present the simulated boater connections between all 9182 lakes in MN, and with the top 10 most connected and top 10 highest centrality lakes using one simulated boater network in Fig. 2. On average, lakes designated as infested were connected to 1,200 more lakes than lakes without an infestation designation (*p* < 0.001) (Table 3). The centrality of infested lakes was ∼ 72 times higher than that of uninfested lakes (*p* < 0.001). Infested lakes were less likely to have boaters remain on the lake, with an average of 17.6% (SD = 19.9%) of outgoing boater traffic returning to the same lake versus 57.4% (SD = 35.9%) for uninfested lakes. On average, 55.0% (SD = 15.8%) of outgoing boater traffic from infested lakes traveled to uninfested lakes.

Table 3 Table of network features overall and by infested / not infested status. The standard deviations are in parentheses

	Overall	Infested	Uninfested
N	982	471	8711
Boater movement network			
Average degree	267 (484)	1406 (919)	206 (357)
Degree distribution			
1–10	20.1%	0.2%	20.3%
10–100	37.7%	5.3%	40.3%
100–500	26.2%	11.5%	27.0%
500–1000	8.3%	20.6%	7.6%
> 1000	7.8%	62.4%	4.8%
Betweenness	16,436 (350,253)	254,698 (1,520,204)	3554 (37,017)
% boaters remaining on the lake	55.33% (36.3%)	17.6% (19.9%)	57.4% (35.9%)

Table 4 The top 10 most connected and highest centrality lakes in the simulated boater movement network

Top 10 most connected lakes				Top 10 lakes with the highest centrality		
Ranking	DOW#[a]	Lake name	Infestation status	DOW#[a]	Lake name	Infestation status
1	27013300	Lake Minnetonka	✔	27013300	Lake Minnetonka	✔
2	11020300	Leech Lake	✔	11020300	Leech Lake	✔
3	69037800	Lake Vermilion	✔	69037800	Lake Vermilion	✔
4	11030500	Gull Lake	✔	69069400	Rainy Lake	✔
5	18031200	Cross Lake Reservoir	✔	11030500	Gull Lake	✔
6	10005900	Lake Waconia	✔	48000200	Mille Lacs Lake	✔
7	18037200	North Long Lake	✔	16000100	Lake Superior	✔
8	04003500	Red Lake	✔	69011800	Burntside Lake	✔
9	29016100	Long Lake	✔	04013000	Lake Bemidji	
10	27010400	Medicine Lake	✔	31053200	Pokegama Lake	

Infestation status as of December 7, 2018. The check mark (✔) indicates that the lake was infested

[a]DOW# represents Division of Waters number for a lake

Discussion

Waterbody connectivity has implications for the spread and control of AIS. We used boater activity survey data and a machine learning algorithm to infer plausible statewide connectivity patterns between Minnesota lakes. Through this approach we uncovered connectivity patterns that were strongly associated with infestation status. The most connected and highest centrality lakes were nearly all known to be infested; of the two uninfested lakes, one (Lake Bemidji) was confirmed to be infested with zebra mussels since our analysis (Minnesota Department of Natural Resources 2018). While unfortunate, this recent infestation supports the applicability of our network modeling approach to reconstruct and forecast waterbodies at risk of AIS for management prioritization. Furthermore, our network approach revealed that boats using uninfested lakes were three times more likely to stay on the same lake for subsequent activity than boats using infested lakes. Notably, more than half of movements form infested
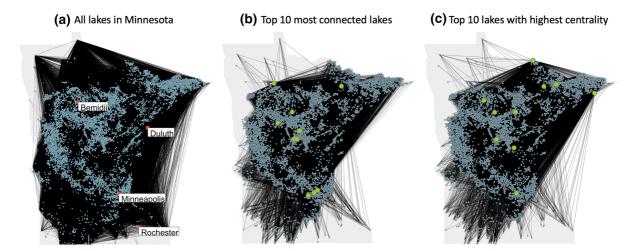

Fig. 2 Boater connectivity based on one simulated network. Panel (**a**) included the boater connection among all 9,182 lakes in Minnesota. Panel (**b**) includes the simulated boater movements from/to the top 10 most connected lakes (green dots). Panel (**c**) displays the simulated boater movements from/to the top 10 lakes with the highest centrality (green dots). We labeled four reference cities (Minneapolis, Duluth, Rochester, and Bemidji) in Minnesota in panel (**a**)

lakes move to uninfested lakes, presenting an opportunity for AIS translocation. Given the size of the Minnesota boater network and volume of boater traffic, management attention to prevent the spread through this pathway is warranted. Our findings are supported by previous studies demonstrating the value of network centrality in understanding and managing the spread of various invasive species through human-mediated pathways (Floerl et al. 2009; Kvistad et al. 2019). However, our findings of the network differences between infested and uninfested lakes should be interpreted with caution. A robust and systematic statewide AIS surveillance effort is lacking in Minnesota. Instead, new AIS invasions are often passively detected by lakeshore property owners or boaters, or opportunistic sampling by lake managers. Infestations in lakes with low boater traffic or population density may therefore be less likely to be detected and reported (Kanankege et al. 2018).

To our knowledge, this is the largest dataset used to create a lake-level connectivity network based on more than 1.6 million reported lake-lake connections by recreational boats. This Big Data approach to network analysis provides significant value over previous studies where data was limited and relatively few lakes were included in the network analysis (Cole et al. 2019; Escobar et al. 2019). However, the tradeoff with our approach is the inherent uncertainty that comes with potential data errors and bias in sample collection. For example, the watercraft inspectors conducting the surveys used for our analysis were not randomly located across the landscape and the effort at each access point was not equal. We did, however, take steps to address these potential sources of bias by adjusting for inspection hours and constructing models to predict lake-lake connections based on lake characteristics rather than constructing a network directly from the observed survey data. The methods employed here to address data errors and bias can serve as a model for the manipulation of similar datasets. Watercraft inspections at public water access points are an important venue for interfacing with recreational boaters and an opportunity for data collection (Cole et al. 2019; Sharp et al. 2017). Watercraft inspections have traditionally been used to educate boaters about AIS, enforce required AIS prevention measures, and detect violations. However, many state and local agencies have begun to survey boaters about their behavior alongside these inspections. In particular, many surveys now ask about the location of recent and/or planned boating activity. It is worth exploring opportunities to standardize, coordinate, and share real-time data collected by inspectors at regional or national scales with the goal of developing more complete boater networks. In addition, exploring the potential to complement these surveys with other sources of data, such as movement data related to mobile phones and angler applications (Papenfuss et al. 2015; Venturelli et al. 2017) is warranted. For example, nearly 30,000 boater movements were excluded from our dataset that moved to/from out of state waterbodies, perhaps from infested waterbodies with AIS not presently in Minnesota. This gap introduces uncertainty in management efforts that aim to prevent the spread of new invaders from other jurisdictions (e.g., quagga mussel into Minnesota from Michigan; Vander Zanden and Olden 2008; Wong and Gerstenberger 2011). The use of boater networks is directly relevant to AIS prevention and management. As we observed in our study, lakes with high boater connectivity are highly correlated with the infestation status of AIS. Understanding the movement patterns of high-risk boats (e.g., those that move from infested to uninfested waterbodies) will allow managers to more strategically develop intervention strategies. For example, developing a risk-based approach to allocation of limited watercraft inspection effort would be feasible with the analysis conducted in this study (Fischer et al. 2020; Haight et al. in review). While we describe the connectivity patterns of lakes in Minnesota arising through boater activity, we do not address the likelihood of spread via boater movement or environmental suitability—both critical factors to determine risk (Robertson et al. 2020). This is a key next step for translating lake connection networks into management tools for surveillance and early detection of aquatic invasive species.

Declartions

Conflict of interest No conflicts of interest

Reference

Bailey SA (2015) An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. Aquat Ecosyst Health Manage 18(3):261–268

Bax N, Williamson A, Aguero M, Gonzalez E, Geeves W (2003) Marine invasive alien species: a threat to global biodiversity. Mar Policy 27(4):313–323

Bishop CM (2006) Pattern recognition and machine learning. Springer

Buchan LA, Padilla DK (1999) Estimating the probability of long-distance overland dispersal of invading aquatic species. Ecol Appl 9(1):254–265

Campbell T, Verboomen T, Montz G, Seilheimer T (2016) Volume and contents of residual water in recreational watercraft ballast. Manage Biol Invasions 7:281–286

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd international conference on knowledge discovery and data mining. Pp 785–794

Chuan PM, Giap CN, Bhatt C, Khang TD (2018) Enhance link prediction in online social networks using similarity metrics, sampling, and classification. In Information Systems Design and Intelligent Applications, 823–833

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

Collas FPL, Karatayev AY, Burklakova LE, Leuven R (2018) Detachment rates of dreissenid mussels after hull-mediated overland dispesal. Hydrobiologia 810:77–84

De VenturaWeissert LN, Tobias R, Kopp K, Jokela J (2016) Overland transport of recreational boats as a spreading

vector of zebra mussel Dreiseena polymorpha. Biol Invasions 18(5):1451–1466

Drake DAR, Mandrak NE (2014) Bycatch, bait, anglers, and roads: Quantifying vector activity and propagule introduction risk across lake ecosystems. Ecol Appl 24(4):877–894

Escobar LE, Mallez S, McCartney M, Lee S, Zielinski DP, Ghosal R, Bajer PG, Wagner C, Nash B, Tomamichel M, Venturelli P, Mathai PP, Kokotovich A, Escobar-Dodero J, Phelps NBD (2017) Aquatic invasive species in the Great Lakes region: an overview. Rev Fish Sci Aquacult 26:121–138

Escobar LE, Romero-Alvarez D, Larkin DJ, Phelps NBD (2019) Network analysis to inform invasive species spread among lakes. J Oceanol Limnol 37(3):1037–1041

Fischer SM, Beck, M, Herborg LM, Lewis MA (2020) Managing aquatic invasions: Optimal locations and operating times for watercraft inspection stations. ArXiv Preprint, arXiv:2003.06092

Floerl O, Inglis GJ, Dey K, Smith A (2009) The importance of transport hubs in stepping-stone invasions. J Appl Ecol 46:37–45

Haak DM, Fath BD, Forbes VE, Martin DR, Pope KL (2017) Coupling ecological and social network models to assess "transmission" and "contagion" of an aquatic invasive species. J Environ Manage 190:243–251

Haight RG, Kinsley AC, Kao S, Yemshanov D, Phelps NBD (In review) Optimizing the location of watercraft inspection stations to slow the spread of aquatic invasive species

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, Newyork

Johnson LE, Carlton JT (1996) Post-establishment spread in large-scale invasions: dispersal mechanisms of the zebra mussel Dreissena polymorpha. Ecol 77(1686):1690

Kanankege KST, Alkhamis MA, Perez AM, Phelps NBD (2018) Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota. J Great Lakes Res 44(3):458–466. https://doi.org/10.1016/j.jglr.2018.02.008

Kvistad JT, Chadderton WL, Bossenbroek JM (2019) Network centrality as a potential method for prioritizing ports for aquatic invasive species surveillance and response in the Laurentian Great Lakes. Manage Biol Invasions 10:403–427

Larkin DJ, Monfils AK, Boissezon A, Sleith RS, Skawinski PM, Welling CH, Cahill BC, Karol KG (2018) Biology, ecology, and management of a starry stonewort (Nitellopsis obtusa; Characeae): A red-listed Eurasian green alga invasive in North America. Aquat Bot 148:15–24

Leung B, Lodge DM, Finnoff D, Shogren JF, Lewis MA, Lamberti G (2002) An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proc R Soc Lond B 269(1508):2407–2413

Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A (2006) Biological invasions recommendations for US policy and management. Ecol Appl 16(6):2035–2054

Ludwig HR Jr, Leitch JA (1996) Interbasin transfer of aquatic biota via anglers' bait buckets. Fisheries 21(7):14–18

Maki K, Galatowitsch S (2004) Movement of invasive aquatic plants into Minnesota (USA) through horticulture trade. Biol Cons 118(3):389–396

McCartney MA, Mallez S (2018). The role of waterway connections and downstream drift of veliger larvae in the expanding invasion of inland lakes by zebra mussels in Minnesota, USA. Aquatic Invasions 13 3

Minnesota Department of Natural Resources. (2014). MNDNR Hydrography. Minnesota DNR – Division of Fish and Wildlife, Fisheries Unit. https://gisdata.mn.gov/dataset/water-dnr-hydrography, Accessed 1 March 2018

Minnesota Department of Natural Resources (2018) List of Infested Waters. https://www.dnr.state.mn.us/invasives/ais/infested.html, Accessed 7 December 2018

Minnesota Department of Natural Resources (2020) Watercraft Inspection Program. https://www.dnr.state.mn.us/invasives/watercraft_inspect/index.html, Accessed 16 July 2020

Minnesota Department of Natural Resources. (2019) Number of watercraft registrations. http://files.dnr.state.mn.us/rlp/licenses/total_registrations.pdf, Accessed 20 October 2020

Minnesota Department of Transportation (2012) Roads, Minnesota, 2012. https://gisdata.mn.gov/dataset/trans-roads-mndot-tis, Accessed 1 March 2018

Muirhead JR, MacIsaac HJ (2005) Development of inland lakes as hubs in an invasion network. J Appl Ecol 42:80–90

Naylor RL, Williams SL, Strong DR (2001) Aquaculture-A gateway for exotic specie. Science 294:1655–1656

Nielsen D (2016). Tree boosting with XGBoost-Why Does XGBoost Win" Every" Machine learning competition. NTNU

Opsahl T (2009). Structure and Evolution of Weighted Networks [Ph.D thesis, University of London (Queen Mary College)]. http://toreopsahl.com/publications/thesis/

Padilla DK, Williams SL (2004) Beyond ballast water: aquarium and ornamental trades as sources of invasive species in aquatic ecosystems. Front Ecol Environ 2(3):131–138

Papenfuss JT, Phelps NBD, Fulton D, Venturelli PA (2015) Coupling ecological and social network models to assess "transmission" and "contagion" of an aquatic invasive species. Fisheries 40:318–327

Ricciardi A, Blackburn TM, Carlton JT, Dick JT, Hulme PE, Iacarella J, Jeschke J, Liebhold A, Lockwood J, MacIsaac H, Pysek P, Richardson D, Ruiz G, Simberloff D, Sutherland W, Wardle D, Aldridge D (2017) Invasion science: a horizon scan of emerging challenges and opportunities. Trends Ecol Evol 32(6):464–474

Robertson JJ, Swannack M, McGarrity M, Schwalb AN (2020) Zebra mussel invasion of Texas lakes: estimating dispersal potential via boats. Biol Invasions 22:3425–3455

Rothlisberger JD, Chadderton WL, McNulty J, Lodge DM (2010) Aquatic invasive species transport via trailerd boats: What is being done, who is moving it, and what can be done. Fisheries 35:121–132

Sharp RL, Cleckner LB, DePillo S (2017) The impact of on-site educational outreach on recreational users' perceptions of aquatic invasive species and their management. Environ Educ Res 23(8):1200–1210. https://doi.org/10.1080/13504622.2016.1174983

Simberloff D (2014) Biological invasions: What's worth fighting for and what can be won? Ecol Eng 65:112–121

Simberloff D, Martin J-L, Genovesi P, Maris V, Wardle D, Aronson J, Courchamp F, Galil B, Garcia-Berthou E, Pascal M, Pysek P, Sousa R, Tabacchi E, Vila M (2013) Impacts of biological invasions: What's what and the way forward. Trends Ecol Evol 28(1):58–66

Sleith RS, Havens AJ, Stewart RA, Karol KG (2015) Distribution of Nitellopsis obtusa (Characeae) in Newyork, USA. Brittonia 67:166–172

United States Census Bureau (2017) Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016. http://factfinder2.census.gov

van Rossum G (1995) Python reference manual. CWI. CWI. https://ir.cwi.nl/pub/5008

Vander Zanden MJ, Olden JD (2008) A management framework for preventing the secondary spread of aquatic invasive species. Can J Fish Aquat Sci 65(7):1512–1522. https://doi.org/10.1139/F08-099

Venturelli P, Hyder K, Skov C (2017) Angler apps as a source of recreational fisheries data: Opportunities, challenges and proposed standards. Fish Fisheries 18:578–595

Wang W, Xu B, Liu J, Cui Z, Yu S, Kong X, Xia F (2019) Csteller: Forecasting scientific collaboration sustainability based on extreme gradient boosting. World Wide Web 22(6):2749–2770

Wong WH, Gerstenberger S (2011) Quagga mussels in the western United States: monitoring and management. Aquat Invasions 6:125–129