



Guiding large-scale management of invasive species using network metrics

Jaime Ashander ^{1,7}✉, Kailin Kroetz^{1,2}, Rebecca Epanchin-Niell ^{1,3}, Nicholas B. D. Phelps ⁴, Robert G. Haight⁵ and Laura E. Dee ⁶

Complex socio-environmental interdependencies drive biological invasions, causing damages across large spatial scales. For widespread invasions, targeting of management activities based on optimization approaches may fail due to computational or data constraints. Here, we evaluate an alternative approach that embraces complexity by representing the invasion as a network and using network structure to inform management locations. We compare optimal versus network-guided invasive species management at a landscape-scale, considering siting of boat decontamination stations targeting 1.6 million boater movements among 9,182 lakes in Minnesota, United States. Studying performance for 58 counties, we find that when full information is known on invasion status and boater movements, the best-performing network-guided metric achieves a median and lower-quartile performance of 100% of optimal. We also find that performance remains relatively high using different network metrics or with less information (median >80% and lower quartile >60% of optimal for most metrics) but is more variable, particularly at the lower quartile. Additionally, performance is generally stable across counties with varying lake counts, suggesting viability for large-scale invasion management. Our results suggest that network approaches hold promise to support sustainable resource management in contexts where modelling capacity and/or data availability are limited.

Complex socio-environmental interdependencies drive biological invasions at both regional and global scales^{1,2}. These invasions damage ecosystems worldwide, threatening biodiversity and ecosystem services^{3,4} and causing annual losses of over US\$46 billion and rising⁵. Over the coming decades, increases in global interconnections and associated trade are expected to drive many more invasions⁶.

Widespread biological invasions, which cover large spatial scales, are a major environmental crisis whose long-term solution will require advancing beyond existing tools for managing established invasions⁷. New, approximate methods for controlling spread at large scales are needed because existing accepted methods for finding exact optima for these spatial resource allocation problems (for example, using optimal control, mixed integer-linear and nonlinear programming^{8–14}) cannot scale efficiently to large systems¹⁵. These current methods are also often hindered by the challenge of limited data. Both the distribution of invasive species and the socio-ecological processes driving spread are at best partially observed even in established invasions^{7,16}.

Network approaches are a promising way to analyse complex socio-environmental systems (SEs)¹⁷. Over the past decade, network science has made advances in understanding how to optimally control spread in complex networks, where the ‘network protection problem’ is fundamental with applications in information security, epidemiology, politics and marketing¹⁸. Recognition that optimization approaches for network protection cannot scale to large network sizes due to problem complexity has led to intensive research effort focused on ‘network-guided management’: heuristic and approximate methods for solving this problem^{19–22} including using network metrics to prioritize management actions²³.

Despite clear parallels between network protection and invasive species management, and network approaches becoming more commonly applied to invasive species or SEs, network-guided management has not been widely adopted for managing biological invasions. Specifically, the parallel exists because an invasion can be represented as a network, with sites that an invasive species can occupy representing nodes of a network and edges representing pathways of spread (for example, via human movement using gravity models²⁴ or via biological dispersal using habitat connectivity models²⁵). As such, there is growing interest among researchers and practitioners in using network characteristics to understand invasions. Network characteristics affect spread both at landscape scales²⁶ and continental or global scales, where invasions are most often mediated by human movement and trade¹. For example, consistent with network theory, experimental work reveals that spread is facilitated by network hubs or central patches through which many dispersal pathways flow and is hindered by more clustered network structures²⁷. However, these insights are not widely used to inform management. One explanation for this is that rigorous comparisons between network-guided management and accepted methods to obtain optimal solutions have been done only for newly established invasions in small systems—a situation where management guided by network structure performs poorly⁸. Despite their poor performance for newly established invasions⁸, network-guided management has been suggested for landscape-scale management of spreading species²⁸, including aquatic invasive species (AIS) that spread via human trade and movement²⁹. To understand if network approaches provide value over accepted methods for optimal management of large networks with established invasions, however, rigorous comparisons are needed but have not yet been done.

¹Resources for the Future, Washington, DC, USA. ²School of Sustainability, Arizona State University, Tempe, AZ, USA. ³Department of Agricultural and Resources Economics, University of Maryland, College Park, MD, USA. ⁴Department of Fisheries, Wildlife, and Conservation Biology, College of Food, Agricultural, and Natural Resource Sciences, University of Minnesota, St Paul, MN, USA. ⁵Northern Research Station, USDA Forest Service, St Paul, MN, USA. ⁶Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA. ⁷Present address: Eastern Ecological Science Center, Patuxent Research Refuge (Formerly the Patuxent Wildlife Research Center), US Geological Survey, Laurel, MD, USA. ✉e-mail: jashander@usgs.gov

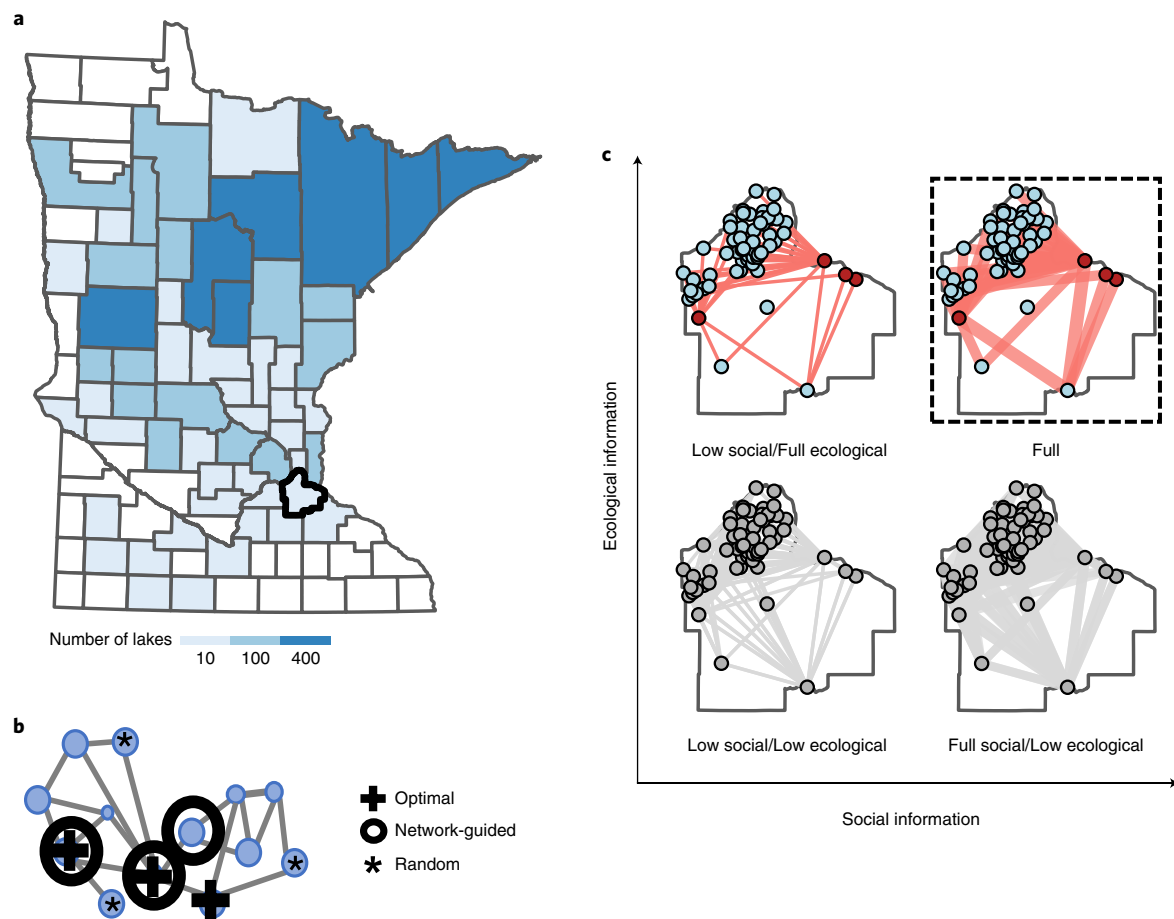


Fig. 1 | Our analytical approach. **a**, Networks of boat movement are constructed at the county level for Minnesota counties with ten or more lakes ($n=58$; Supplementary Table 1) shaded here by the number of lakes in the county, which corresponds to network size. Dakota County (bold outline) is used for illustration in **c**. **b**, Network-guided management (placement of inspection stations) using heuristics based on centrality metrics (Table 1) is compared both to optimal inspection (using integer-programming) and random management. **c**, In addition to examining performance of network-guided management with full information (dashed outline), we evaluate performance in three cases with only partial information on either or both the boater movement network (social information) and the invasion status of each lake (ecological information). See Table 2 for more detailed definitions of full and low social and ecological information.

Here, we aim to bridge the gap between advances in network science on spread prevention in complex networks and real-world management of invasions in SESs. Specifically, our approach connects recent enquiry into using network structure to guide management of invasive species with existing understanding of optimal management of species invasions. We focus on a case of managing a large-scale invasion via allocation of inspection and decontamination stations ('inspection stations' hereafter) across a network, with the objective of minimizing the number of uninspected but potentially infective connections, a common problem in invasive species management^{14,30}. We evaluate performance of management based on network structure (that is, centrality metrics, Table 1; see ref.²³) relative to the optimal management from an integer-linear programming solution^{14,30}. Since managers are often constrained by knowledge of infestation status and processes (for example, current invasion distribution and dispersal patterns), we also examine how performance degrades with reduced information on invasion status or spread magnitude, relative to the optimal solution which requires full information. Additionally, we characterize performance across a range of budgets. Thus, our analysis advances knowledge about the outstanding question of whether network metrics can effectively guide management and whether some metrics could work better

than others in more information-rich versus information-poor management settings.

Management context

We focus on the large-scale management problem of preventing zebra mussel (*Dreissena polymorpha*) spread in Minnesota, United States (Fig. 1a). Zebra mussels are one of the most costly invasive species in the United States, causing changes to ecosystem processes³¹, extirpating native mussel species³² and damaging infrastructure³³. In Minnesota, zebra mussels are a prohibited invasive species, yet have spread to more than 270 lakes or rivers since their first report in 1989, with more than half of those infestations occurring since 2016 (ref.³⁴ and Supplementary Section 1). This rapid spread has been largely human-mediated, facilitated by movement of boats and equipment^{35,36}. The objective we examine, minimizing short-term spread via optimal location of inspection stations at lakes, is implicitly aimed at achieving the more fundamental goal of reducing long-term damages from invasion spread³⁴. Given that significant state funding is delegated to the county level (US\$10 million annually for AIS prevention; ref.³⁴), decisions on where to locate most inspection stations are made by county managers³⁴. Indeed, co-authors N.B.D.P. and R.G.H. were approached by state

Table 1 | Network centrality metrics: definitions for network centrality metrics used in our three focal strategies, which prioritize nodes with highest centrality value

Method	Description
Degree	Degree is the number of direct links (incoming and outgoing) incident on a node. In weighted networks, the equivalent is weighted degree or strength.
H + A	Sum of two eigenvalue-based metrics that account for incoming links (authority score) and outgoing links (hub score). Unlike degree, the contribution of a direct link is weighted by the linking node's score.
Betweenness	The number of shortest paths between all other nodes that pass through a node.

and county managers to both characterize boat movements³⁷ and develop decision support tools for cost-effective county-level management of watercraft inspection programmes³⁰. Here, for each of the 58 counties in Minnesota with ten or more lakes, we examine the performance of network-guided approaches relative to an optimal solution obtained via integer-linear programming (Methods). Our set-up allows us to also examine how performance varies across counties that differ in system size (number of nodes).

Analytical approach

Our analytical approach centres around the optimal placement of boat inspection stations on a network of infested and uninfested lakes connected by boater movements. We compare the optimal solution from integer-linear programming to network metric-guided management with sites prioritized on the basis of the structure of the boater movement network. Our general approach involves six key steps.

- (1) Construct the network. Given that the primary mode of zebra mussel spread is through human movement (of boats)³⁵, we define a network with lakes that are infested or not as nodes and the number of boats moving between lakes as weighted, directed edges that mediate spread of infestation (Methods and Supplementary Section 2; ref. ³⁷). Within the overall network there is an 'infested subnetwork'—the network of potentially infective movements from infested lakes to uninfested lakes (Fig. 1b). We construct a network for each county with ten or more lakes ($n = 58$; Supplementary Table 1).
- (2) Compute the optimal inspection strategy for each county. The management objective is to maximize the inspections of potentially infective boats (boats moving from infested to uninfested lakes). We focus on county-level management, which controls 90% of the inspection stations in Minnesota. County managers decide where to locate inspection stations given a fixed budget, often with minimal coordination between counties or with the state. Thus, we solve for each county's optimal inspection pattern, subject to a fixed budget, using integer-linear programming (Methods) where the objective is to maximize the number of potentially infective boats inspected (Supplementary Section 3).
- (3) Compute network-guided inspection strategies for each county. We evaluate management strategies guided by several centrality metrics that are measures of network structure (Table 1). These include two strategies used in network protection contexts for the objective of reducing long-term spread: prioritizing by highest degree or betweenness centrality²³. These metrics also have been identified as useful predictors of node importance for invasive species spread^{1,27}. We also propose and apply a new metric that combines the hub and authority scores of ref. ³⁸,

our hub + authority score ($H + A$) favours nodes that are either sources of infestation or targets of infestation on a directed network. In total, we examine five heuristic network-guided strategies (Supplementary Sections 4 and 5), presenting results in the main text for three focal strategies that prioritize nodes with highest network centrality (Table 1 and Methods).

To evaluate management strategies guided by centrality metrics, we rank the lakes (nodes) in each county according to each centrality metric and place interventions in order of priority on the basis of these ranks until the budget runs out. We calculate the relative performance of these strategies (Fig. 1b), that is, the proportion of potentially infective boats inspected compared to optimal inspection (Supplementary Section 7).

- (4) Construct cases with reduced ecological and social information. The optimal solution provides an upper bound on the performance of network-guided management but because data are a key constraint in managing large, complex socio-ecological systems, we seek insight into whether network-guided management can perform well when planners use network-guided approaches with less detailed social or ecological information. We focus on four cases with varying levels of social and ecological information (Fig. 1c and Table 2): full social and full ecological information; full social and low ecological; low social and full ecological; and low social and low ecological. In both full and low social information cases, the directed topology of the network is known. With full social information, the edge weights are known and a weighted edge from one lake to another represents the number of boats moving from that lake to the other. With low social information, the presence or absence of edges between two lakes are known. An edge from one lake to another represents at least one boat movement from that lake to the other, while no edge represents zero boat movements. With full ecological information, lake infestation status is known. Combined with the directed topology, this means that the infested subnetwork—the network consisting of only potentially infective movements between infested lakes and uninfested lakes—is known with full ecological information. With low ecological information, only the position of each lake in the network is known, not its infestation status. See Table 2 and Supplementary Section 6 for more detail.
- (5) Evaluate median and lower-quartile performance across counties for a range of budgets. For all network metrics and information levels we consider a range of relative budgets, expressed as a proportion of the maximum budget adequate to inspect all potentially infective boats in the county. Our unit of observation is a county. To measure overall performance we examine the median (0.5 quantile) and lower-quartile (0.25 quantile) outcomes. We also computed average performance across counties within the lowest quartile and the proportion of counties where methods performed optimally or failed (defined as relative performance < 0.66 ; Methods and Supplementary Section 8).
- (6) Assess performance as a function of network size. Our last step exploits the variation in network size across counties, measured as the number of lakes within a county. Specifically, we use the results from the prior steps to examine whether performance changes with an increase in network size.

Results

Across all metrics and information levels metric-guided strategies achieved median performance of at least 0.80 of the optimal for reasonable budgets (relative budgets < 0.25 ; Supplementary Table 3). Metric-guided strategies also vastly exceeded the random benchmark, which achieved much lower performance relative to the optimal (median 0.15, lower quartile 0.10) for the same budgets

(Supplementary Table 5). As resource constraints in invasive species management mean budgets are almost never sufficient to control every site¹⁵, we focus on results with a low relative budget (0.1 the amount required to inspect all infested boats); however, the qualitative patterns described below held for all relative budgets below 0.25 (Supplementary Tables 2 and 3). Performance of metric-guided management generally improved with higher budgets (Supplementary Figs. 2 and 3). Optimal management also improved with higher budgets but showed diminishing returns (Supplementary Fig. 1). Synthesizing the results across metrics, information levels and network size leads to four main findings, which we describe in the following subsections.

Network-guided management achieves near-optimal performance. With full information, degree and H + A achieved perfect (100%) median performance for reasonable budgets. Degree achieved near-perfect performance across all counties: comparing inspected lakes between degree and the optimal solution reveals that, for relative budgets below 0.25, degree selected the same inspection patterns as the true optimum 87% of the time (Supplementary Fig. 4 and Supplementary Table 3). A ‘recalculated’ degree strategy selected the exact same inspection patterns for all relative budgets below 0.5 (Supplementary Section 5). Differences between degree and H + A emerge at the lower quartile with performance of degree (100%) exceeding H + A (97%). Still, H + A performs well across all counties with the average performance of the lower quartile of counties 89% of optimal.

Performance varies by metric. Betweenness performed worse than degree and H + A. Across the counties, its median and lower-quartile performance for a realistic budget of 0.1 of the amount required to inspect all infested boats was 93% and 84% of optimal, respectively (Supplementary Table 2). Average performance across counties in the lower quartile was 72% of optimal. Other metrics explored in the Supplementary Information perform even worse, particularly in the lower quartile, revealing that performance can be tied to the choice of metric.

Performance is high but variable with less information. Metric performance was reduced when less detailed social and ecological information was used in the metric calculation (Fig. 2a) but for a realistic budget of 0.1 these patterns differed depending on which type of data was removed (Supplementary Table 2). In the case with full social information but low ecological information, the performance of degree (median 94%, lower quartile 83%) and H + A (median 96%, lower quartile 87%) strategies were both high. The performance of betweenness was lower (median 69%, lower quartile 50%) (Supplementary Table 2). In comparison, with low social but full ecological information the ranking of metrics by their median and lower-quartile performance shifted. Betweenness (median 93%, lower quartile 84%) and H + A (median 92%, lower quartile 88%) performed best, with degree slightly lower (median 88%, lower quartile 72%). Also notable is that the average of the counties in the lower quartile was similar for H + A in each of these partial information scenarios (67% with low environmental, 77% with low social) and failure rates remained below 10%. In contrast, for betweenness and degree, average performance in the lower quartile dropped to near 50% or lower in one scenario or another and failure rates were higher. For low social/low ecological, all focal metrics performed approximately the same, with median performance above 80%, lower quartile $\approx$ 65% and average across the lower-quartile counties 44–53%. These qualitative patterns held for reasonable budgets (relative budgets <0.25, Supplementary Table 3). Very similar results to the low social/low ecological case were observed even with only undirected topology (‘minimal’ information; Supplementary Section 6).

Table 2 | Varying social and ecological information: our four focal cases of full or partial information result from varying the level of detail on ecological and social information between two levels defined in this table

Information level	Definition
Full social	Weighted-directed network of boat movements (edges represent the number of boats moving from source to target lake).
Low social	Unweighted-directed network of boat movements (edges represent one or more boats moving from source to target lake).
Full ecological	Lake position in network and invasion status.
Low ecological	Lake position in network only.

See also Fig. 1 and Supplementary Section 6.

Performance generally does not decrease in larger networks.

Metric performance generally improves or shows no significant change as network size, measured as the number of lakes, increases (Fig. 2b; Supplementary Fig. 6 for other budgets). Increases in performance with network size are common at the lowest information levels. In the low social/low ecological case, the lower (0.25 quantile) performance for all focal metrics was significantly higher ($P < 0.05$) for counties with larger networks (Fig. 2b; Supplementary Fig. 6 for other budgets). This finding also holds for even less information, when only the undirected topology is known (‘minimal’ information; Supplementary Section 6 and Supplementary Fig. 5). The one case where network size showed slight, but negative, association with performance is with full information for H + A (statistically significant decline in 0.25 quantile performance). However, the magnitude of this change over the range of network size was small compared to the larger positive association between performance and network size seen with low levels of ecological information.

Discussion

Our study demonstrates that network-based approaches can guide nearly optimal invasive species management, including in large systems with constrained resources for management. We find that by managing based on some network metrics, a manager can achieve performance close to or equivalent to optimal—particularly using degree and H + A (defined in Fig. 1, results in Fig. 2a). Performance was more variable in cases with relatively low budgets and low levels of social or ecological information. In these cases, there was relatively strong median performance across metrics but greater differences in performance of counties in the lower quartiles across metrics, including cases where average performance of the counties in the lower quartile was lower than 50% of optimal. On one hand, our results provide empirical evidence that network-guided management can achieve good performance even with limited information but the lower-quartile results suggest that appropriate metric selection is important for achieving these outcomes. Finally, we find that the performance of metric-guided management strategies generally does not decrease with network size and generally increases with network size and with low levels of information—that is, in the contexts where identification of optimal solutions can be most computationally challenging. Stability of performance across network size also suggests that metrics can be tested for performance on smaller-scale systems to provide insights into preferred metrics to use in larger systems. Together, these results suggest that network-guided management is a promising approach for managing large-scale invasions, including those characterized by limited budgets or less detailed information about the social-ecological system.

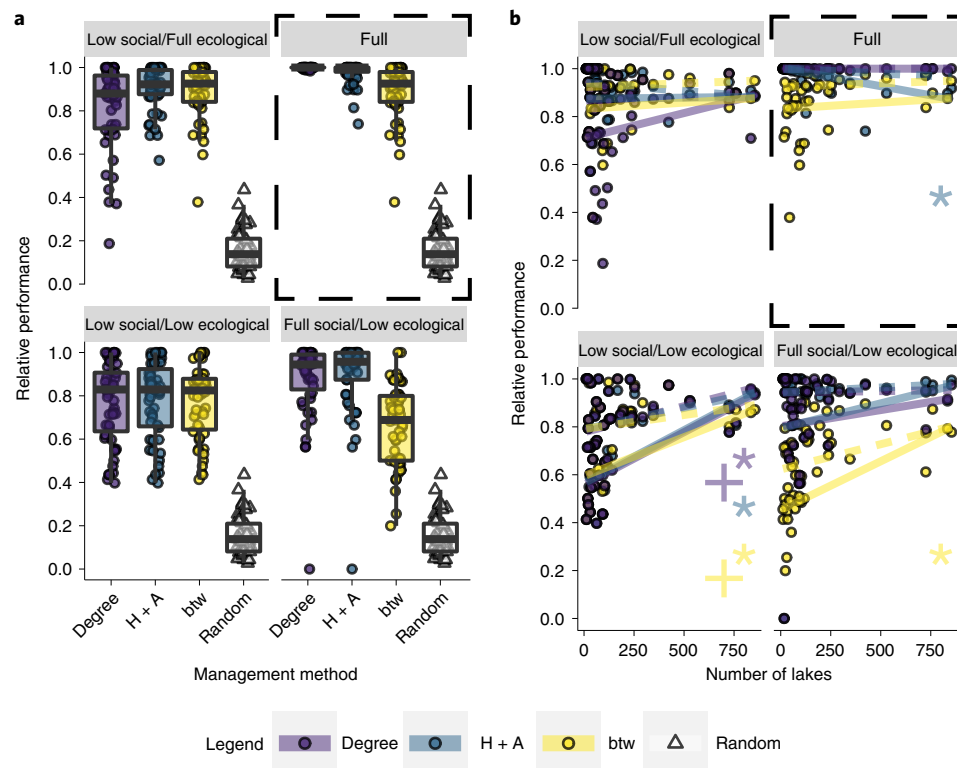


Fig. 2 | Relative performance of network-guided management—measured as the number of infective boats inspected using metric-guided management as a proportion of those inspected in the integer-programming solution—for relative budget 0.1. Performance is shown for full information (dashed outline) and three cases with only partial information (lower levels of social and ecological information) for three metrics (Table 1; 'btw' is betweenness). The legend colours apply to both panels. **a**, The distribution of performance across counties ($n=58$) illustrated with boxplots (centre line, median; box limits, upper and lower quartiles; whiskers, $1.5\times$ interquartile range) overplotted on county values (circular points). Random inspection also is included for comparison; there, triangular points represent mean performance across many random inspection placements (Methods). See Supplementary Table 2 for numeric values of median performance (0.5 quantile of the full performance distribution) and mean lower-quartile performance. **b**, Relationship between performance and network size across 58 counties (circular points); lines shown are quantile regression results for the 0.50 and 0.25 quantiles. For clarity, random inspection is not included. *, Significant regression slope of the 0.25 quantile (solid line); +, significant regression slope of the 0.50 quantile (dashed line) (no overlap of 95% confidence interval with 0).

Our research adds to a growing body of work applying network science to the management of complex SESs. The utility of centrality metrics for guiding invasion management, as found in our study, is expected from longstanding network theory (for example, ref. ²¹) but their potential to guide management has not been realized for invasive species management at large scales. For small and newly invaded systems, others⁸ found that centrality metrics can actually mislead management. However, large-scale invasions in real management contexts are inherently a large-network problem. While prior studies on invasive species in large networks assessed how well network metrics predict spread (for example, refs. ^{1,28}) and examined performance of metrics in prioritizing management actions (for example, ref. ²⁹), they did not rigorously compare the performance of network-based management to known-optimal interventions as we do here. In another conservation context (conserving species in a food web), at least one prior study compared the performance of network metric-guided management to an optimal approach³⁹ but did not test the performance for invasion management and under varying information, budgets and network size, as in this study.

Here, we test network-guided management's efficacy for the objective of minimizing the number of uninspected but potentially infective boats (boats moving from infested to uninfested lakes) but other management problems could be explored in future work. Our objective captures the current means objective in our study region and is equivalent to minimizing one-period spread, which is

implicitly aimed at achieving a more fundamental goal of reducing long-term damages from invasion spread³⁴. Invasive species management involves many approaches to reducing long-term impacts, with diverse management activities including surveillance, containment and removal, and with goals spanning prevention to eradication¹⁵. Each option represents a distinct resource allocation problem, involving different time horizons, tools and management objectives. Assessing the performance of network metrics for an objective like long-term damages would require integrated, dynamic SES models that describe changes in boater movement over time. A fuller picture of performance could also include empirical ex-post evaluation of management efficacy.

The relative performance of different metrics for network-guided management probably depends on the management objective. Here, the higher-performing metrics (degree and H + A in Fig. 2a) select nodes with a high weighted degree, which corresponds directly to high numbers of inspected boats. In contrast, the lower-performing betweenness centrality depends on all paths in the network and therefore selects nodes that may not have a high degree and thus may not correspond to a high number of inspected boats. Work on the network protection problem offers further support to this idea that the best-performing metric will depend on the objective. For example, when controlling spread on a network via protecting nodes¹⁸ but with the objective of minimizing the rate of spread in the long-term, betweenness centrality outperforms strategies based on

degree centrality²³. Thus, our empirical results highlight the potential importance of selecting metrics that ‘match’ the management objective and support a broader literature exploring this idea. For future research on network-guided management of invasive species or other applications of network-guided conservation, leveraging and translating predictions from network epidemiology (reviewed in ref. ¹⁸) is a promising approach.

Another outstanding topic for future research is how the performance of metric-based heuristics depends on interactions between invasive species’ dispersal characteristics and the data used to construct the network. The ordering of metric performance in cases with low levels of network information may be especially sensitive to these features. In our application, the spectral H + A metric had the most consistent performance across varying information levels. This indicates that in the partial information cases we considered (that is, low social/full ecological or full social/low ecological; Fig. 2) H + A captures details in the network structure that result in a high ranking for nodes that turn out to be the highest degree in the full information case—that is, on the weighted infested subnetwork. Future work could investigate if this finding is specific to our context, including the network definition (from data collected over 2014–2017; ref. ³⁷), the definition of invasion state (invaded in 2019), and the ecological properties (for example, temporal and spatial dispersal abilities) of this invasive species. In our application, centrality metrics on the network are predictive of invaded status³⁷. While this has been observed for AIS on other human transport networks²⁹, it may not be the case for some invasions. Whether our findings extend to more general cases should be investigated, potentially through simulation approaches from network epidemiology (for example, ref. ¹⁸). Another promising extension would be applying our methods to networks derived from other sources. Gravity models have often been applied to AIS (for example, refs. ^{40,41}) and these models could be used to create an approximate network²⁴ that could then be used for network-guided management.

Network approximations may be particularly valuable for spatial prioritization in very large systems, including for providing decision support for management, for several reasons. For one, integer-linear programming (ILP) approaches that have been successfully applied to small- and medium-sized problems in conservation planning (for example, conservation prioritization with MARXAN⁴², sparing-sharing in tropical forests with ILP⁴³) have computational and data constraints that limit their use for large-scale systems. Second, complementary work by others⁴⁴ demonstrates the effectiveness of network-guided management for real-world decision support by embedding near-optimal prioritization based on a network metric in a web-tool developed with intensive stakeholder engagement. Finally, output from network approaches such as ours could be integrated with broader frameworks for assessing invasion risk on the basis of factors such as habitat suitability, stream connections and impacts^{35,45}. Ongoing engagement between managers and researchers, such as in Minnesota, provides a promising context for pursuing such extensions⁴⁶.

In summary, our results suggest that network-guided management could provide an important tool for addressing management challenges posed by widespread invasions, which are becoming common due to globalization and other global changes. Our study demonstrates the potential value of methods developed in other areas of network science with a richer history of working in large complex systems for invasion management. These approaches also have relevance for other environmental contexts, such as fire, fisheries or endangered species management, where management occurs within complex systems that test the limits of traditional optimal control tools (for example, ref. ⁴⁷). Overall our results show that network approaches hold promise for improving management outcomes in contexts where modelling and/or data resources are limited.

Methods

We compare optimal versus network-guided invasive species management at a landscape scale, considering siting of boat decontamination stations within 58 counties, jointly containing 9,182 lakes, in Minnesota, United States.

Data and network representation. We use a network generated from ≈ 1.6 million reported boater movements between lakes in Minnesota over the period 2014 to 2017 and zebra mussel infestation status from 2019³⁷. This network consists of lakes (nodes) connected by directed movements of boaters (edges). Edge weights represent estimated numbers of boaters moving between two connected lakes³⁷. On the basis of lake infestation status, we categorize all network edges as either potentially infective (a movement from an infested body of water to an uninfested body of water based on observed infestation status) or non-infective (Supplementary Sections 1 and 2).

Defining county-level networks and infested subnetworks. We construct 58 county-level networks from the Minnesota-wide data, including all counties with greater than ten lakes. We account for out-of-county and in-to-county boater movement by adding two nodes to each county network that serve as synthetic lakes: the first corresponding to all non-county lakes that are uninfested and the second corresponding to all non-county lakes that are infested. We aggregate all edge-weight that crosses a county line into the incoming and outgoing links to these two synthetic nodes. In this way, we define a county-level network as an adjacency matrix $A^{(k)}$ (for k in 58, the number of counties that we examine). These county-level networks vary widely in summary statistics related to the network and infestation status, including number of edges, number of trips and proportion of potentially infective trips (Supplementary Table 1).

For each county, we further define an infested subnetwork, which is a subgraph of each full county network $A^{(k)}$, consisting of only the potentially infective edges and their connected lakes. Each edge in this subnetwork is a directed edge from an infested lake to an uninfested lake. The subnetwork connects a subset of lakes in the county network as an adjacency matrix $\tilde{A}^{(k)}$. The entry $\tilde{A}_{ij}^{(k)}$ corresponds to the number of potentially infective trips from lake i to lake j .

Problem formulation. Although boater movement connects lakes across counties, resource allocation decisions primarily occur within counties, so we study a decision problem at that scale. The decision problem for county k is to determine the location of inspection stations, given a limited budget $B^{(k)}$, to maximize the number of boats inspected that are moving from infested lakes anywhere to uninfested lakes within the county. We assume uniform costs to inspect any given lake (the cost of inspection stations are the same for each lake).

Optimal solution and random-inspections benchmark. As a reference point and the upper bound on the performance of our network metrics we solve for the optimal solution for a fixed budget as an integer-linear programme, an approach that has previously been applied to optimize inspection for invasive species^{14,30}. Indeed, the ILP formulation for the optimal solution that we use here was developed to support AIS management in several Minnesota counties (ref. ³⁰; Supplementary Section 3). For each county k , we use the adjacency matrix of the infested subnetwork $\tilde{A}^{(k)}$ to define the infested matrix $N^{(k)}$ with rows corresponding to the infested lakes, within the county and the infested synthetic lake and columns corresponding to the uninfested lakes only in the county. Each entry N_{ij} gives the number of boats moving from infested lake i to uninfested lake j . We define the decision problem of locating inspection stations as a sum across entries in this matrix, subject to constraints for the budget and to disallow inspection of synthetic lakes as an integer-linear programme (Supplementary Section 3).

As a lower bound on performance, we also computed the performance of randomly located inspection stations, where for a budget of $B^{(k)}$, we chose $B^{(k)}$ lakes at random from all lakes in county k . To determine the mean performance of a random strategy for a given budget, we initially compute the mean performance for 100 replicates, then add replicates in increments of five until mean performance across all replicates changes by less than one inspected boat. Plots for the random strategy show the final mean performance; for most county-by-budget combinations, this procedure resulted in a final mean taken over 105 to 570 replicates.

Network metric solution. We compute metric-guided strategies degree, H + A and betweenness (Table 1) as described in the Supplementary Sections 4 and 5). For each metric and each information level, we select nodes for inspection sequentially until the budget (total number of nodes that can be inspected) is reached. For all metrics we select nodes on the basis of the ranking from the initial state of the network (Supplementary Algorithm 1). We also examined a ‘recalculated’ method²³: compute the metric for all nodes, select the node with the highest metric score for inspection, update the network to exclude the selected node, recalculate the metric for all remaining uninspected nodes, then select the remaining node that has the highest score, continuing with this process until the budget is reached (Supplementary Algorithm 2). For the objective and metrics we examine here, the ‘recalculated’ method shows little improvement over the initial state method (Supplementary Section 5 and Supplementary Fig. 7).

Computing relative performance of non-optimal strategies. To compare the performance of network-guided management (using heuristics based on centrality metrics) to the optimal allocation, we compute the number of potentially infective boats that are inspected under each strategy. We calculate relative performance as the number of infective boats inspected using metric-guided management divided by the number of infective boats inspected using the optimal integer-programming approach (Supplementary Section 7). We apply this approach for each county k and for each budget up to the minimum budget required to inspect all potentially infective boats given full information and solving the inspection problem for the optimal solution. We denote this budget $B_{\text{cover}}^{(k)}$ (Supplementary Section 3). We use the maximum budget to define each of the k county's relative budget $b^{(k)} = B^{(k)}/B_{\text{cover}}^{(k)}$, which facilitates comparing amongst counties of different sizes. We evaluated performance by calculating the median (0.5 quantile) relative performance across all counties for each budget as well as mean lower-quartile (0.25 quantile) performance. We also computed the proportion of counties where the performance is perfect, that is, matching the optimal exactly and the proportion of failures (defined as relative performance <0.66 ; Supplementary Section 8). These values are reported in Supplementary Tables 2–5.

Relationship between performance and network characteristics. Quantile regression estimates relationships between predictors and parts of the outcome distribution other than the mean⁴⁸. To test associations between relative performance both at the median (0.5 quantile) and lower quartile (0.25 quantile) and network characteristics (size or the number of nodes; average degree) we computed quantile regressions, using R package quantreg. We performed linear quantile regressions for the median and lower quartile (quantiles 0.5, 0.25) and tested for significance using confidence intervals produced via inverting a rank test⁴⁹. This method quantifies whether network characteristics show a significant association with increases or decreases in both average and lower-end performance.

Reporting summary. Further information on research design is available in the Nature Research Reporting summary linked to this article.

Data availability

The network data used in this study were previously reported³⁷ and are available at <https://conservancy.umn.edu/handle/11299/216936>. The minimal dataset supporting this study, including network data, lake metadata including infestation status and geospatial data delineating county boundaries are available⁵⁰.

Code availability

Analysis used R v.4.0.2 (2020-06-22) using packages dplyr (v.1.0.7), purrr (v.0.3.4), ggplot2 (v.3.3.3), igraph (v.1.2.5), quantreg (v.5.61) and Rglpk (v.0.6-4). Full analysis code underlying all analyses are available⁵⁰.

Received: 30 July 2021; Accepted: 6 May 2022;

Published online: 14 July 2022

References

- Banks, N. C., Paini, D. R., Bayliss, K. L. & Hodda, M. The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol. Lett.* **18**, 188–199 (2015).
- Epanchin-Niell, R. et al. Controlling invasive species in complex social landscapes. *Front. Ecol. Environ.* **8**, 210–216 (2009).
- Charles, H. & Dukes, J. S. in *Biological Invasions* (ed. Nentwig, W.) 217–237 (Springer, 2007). https://doi.org/10.1007/978-3-540-36920-2_13
- Gallardo, B., Clavero, M., Sánchez, M. & Vilà, M. Global ecological impacts of invasive species in aquatic ecosystems. *Glob. Change Biol.* **22**, 151–163 (2016).
- Diagne, C. et al. High and rising economic costs of biological invasions worldwide. *Nature* **592**, 571–576 (2021).
- Sardain, A., Sardain, E. & Leung, B. Global forecasts of shipping traffic and biological invasions to 2050. *Nat. Sustain.* **2**, 274–282 (2019).
- Epanchin-Niell, R. S. & Hastings, A. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. *Ecol. Lett.* **13**, 528–541 (2010).
- Chades, I. et al. General rules for managing and surveying networks of pests, diseases, and endangered species. *Proc. Natl. Acad. Sci. USA* **108**, 8323–8328 (2011).
- Epanchin-Niell, R. S. & Wilen, J. E. Optimal spatial control of biological invasions. *J. Environ. Econ. Manag.* **63**, 260–270 (2012).
- Epanchin-Niell, R. S. & Wilen, J. E. Individual and cooperative management of invasive species in human-mediated landscapes. *Am. J. Agric. Econ.* **97**, 180–198 (2015).
- Aadland, D., Sims, C. & Finnoff, D. Spatial dynamics of optimal management in bioeconomic systems. *Comput. Econ.* **45**, 545–577 (2015).
- Baker, C. M. Target the source: optimal spatiotemporal resource allocation for invasive species control. *Conserv. Lett.* **10**, 41–48 (2017).
- Bushaj, S., Büyüktaktakın, İ. E., Yemshanov, D. & Haight, R. G. Optimizing surveillance and management of emerald ash borer in urban environments. *Nat. Res. Model.* **34**, e12267 (2021).
- Fischer, S. M., Beck, M., Herborg, L.-M. & Lewis, M. A. Managing aquatic invasions: optimal locations and operating times for watercraft inspection stations. *J. Environ. Manag.* **283**, 111923 (2021).
- Büyüktaktakın, İ. E. & Haight, R. G. A review of operations research models in invasive species management: state of the art, challenges, and future directions. *Ann. Oper. Res.* **271**, 357–403 (2018).
- Epanchin-Niell, R. S. Economics of invasive species policy and management. *Biol. Invasions* **19**, 3333–3354 (2017).
- Bodin, Ö. et al. Improving network approaches to the study of complex social–ecological interdependencies. *Nat. Sustain.* **2**, 551–559 (2019).
- Nowzari, C., Preciado, V. M. & Pappas, G. J. Analysis and control of epidemics: a survey of spreading processes on complex networks. *IEEE Control Syst.* **36**, 26–46 (2016).
- Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).
- Kempe, D., Kleinberg, J. & Tardos, E. Maximizing the spread of influence through a social network. In *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 137–146 (ACM Press, 2003).
- Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65**, 036104 (2002).
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109 (2002).
- Muirhead, J. R. & Macisaac, H. J. Development of inland lakes as hubs in an invasion network. *J. Appl. Ecol.* **42**, 80–90 (2005).
- de la Fuente, B., Saura, S. & Beck, P. S. Predicting the spread of an invasive tree pest: the pine wood nematode in southern Europe. *J. Appl. Ecol.* **55**, 2374–2385 (2018).
- Minor, E. S. & Urban, D. L. A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv. Biol.* **22**, 297–307 (2008).
- Morel-Journel, T., Assa, C. R., Mailleret, L. & Vercken, E. Its all about connections: hubs and invasion in habitat networks. *Ecol. Lett.* **22**, 313–321 (2019).
- Perry, G. L. W., Moloney, K. A. & Etherington, T. R. Using network connectivity to prioritise sites for the control of invasive species. *J. Appl. Ecol.* **54**, 1238–1250 (2017).
- Kvistad, J. T., Chadderton, W. L. & Bossenbroek, J. M. Network centrality as a potential method for prioritizing ports for aquatic invasive species surveillance and response in the Laurentian Great Lakes. *Manag. Biol. Invasions* **10**, 403 (2019).
- Haight, R. G., Kinsley, A. C., Kao, S.-Y., Yemshanov, D. & Phelps, N. B. Optimizing the location of watercraft inspection stations to slow the spread of aquatic invasive species. *Biol. Invasions* **23**, 3907–3919 (2021).
- McEachran, M. C. et al. Stable isotopes indicate that zebra mussels (*Dreissena polymorpha*) increase dependence of lake food webs on littoral energy sources. *Freshw. Biol.* **64**, 183–196 (2019).
- Karatayev, A. Y., Burlakova, L. E. & Padilla, D. K. in *Invasive Aquatic Species of Europe. Distribution, Impacts and Management* (eds Leppäkoski, E. et al.) 433–446 (Springer, 2002).
- Prescott, T. H., Claudi, R. & Prescott, K. L. Impact of Dreissenid mussels on the infrastructure of dams and hydroelectric power plants. In *Quagga and Zebra Mussels* (eds Nalepa, T. F. & Schloesser, D. W.) 243–258 (CRC Press, 2013).
- Invasive Species of Aquatic Plants and Wild Animals in Minnesota: Annual Report for 2020* (Minnesota Department of Natural Resources, 2020).
- Kanankege, K. S., Alkhamis, M. A., Phelps, N. B. & Perez, A. M. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Front. Vet. Sci.* **4**, 231 (2018).
- Mallez, S. & McCartney, M. Dispersal mechanisms for zebra mussels: population genetics supports clustered invasions over spread from hub lakes in Minnesota. *Biol. Invasions* **20**, 2461–2484 (2018).
- Kao, S.-Y. Z. et al. Network connectivity of Minnesota waterbodies and implications for aquatic invasive species prevention. *Biol. Invasions* **23**, 3231–3242 (2021).
- Kleinberg, J. M. Authoritative sources in a hyperlinked environment. In *Proc. 9th Annual ACM-SIAM Symposium on Discrete Algorithms* 668–677 (1998).
- McDonald-Madden, E. et al. Using food-web theory to conserve ecosystems. *Nat. Commun.* **7**, 10245 (2016).
- Bossenbroek, J. M., Kraft, C. E. & Nekola, J. C. Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecol. Appl.* **11**, 1778–1788 (2001).
- Leung, B., Bossenbroek, J. M. & Lodge, D. M. Boats, pathways, and aquatic biological invasions: estimating dispersal potential with gravity models. *Biol. Invasions* **8**, 241–254 (2006).

42. Beger, M. et al. Integrating regional conservation priorities for multiple objectives into national policy. *Nat. Commun.* **6**, 8208 (2015).
43. Runting, R. K. et al. Larger gains from improved management over sparing–sharing for tropical forests. *Nat. Sustain.* **2**, 53–61 (2019).
44. Kinsley, A. C. et al. AIS Explorer: prioritization for watercraft inspections. A decision-support tool for aquatic invasive species management. *J. Environ. Manage.* **314**, 115037 (2022).
45. Vander Zanden, M. J. & Olden, J. D. A management framework for preventing the secondary spread of aquatic invasive species. *Can. J. Fish. Aquat. Sci.* **65**, 1512–1522 (2008).
46. Kanankege, K. S. et al. Lessons learned from the stakeholder engagement in research: application of spatial analytical tools in one health problems. *Front. Vet. Sci.* **7**, 254 (2020).
47. Kroetz, K. & Sanchirico, J. The bioeconomics of spatial-dynamic systems in natural resource management. *Annu. Rev. Resour. Econ.* **7**, 189–207 (2015).
48. Cade, B. S. & Noon, B. R. A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* **1**, 412–420 (2003).
49. Koenker, R. in *Asymptotic Statistics* (eds Mandl, P. & Hušková, M.) 349–359 (Springer, 1994).
50. Ashander, J. Analysis code and data for ‘Guiding large-scale management of invasive species using network metrics’. *figshare* <https://doi.org/10.6084/m9.figshare.14402447> (2021).

Acknowledgements

We thank A. Kinsley for comments on a previous draft. Funding for this research was provided by Resources for the Future and the National Socio-Environmental Synthesis

Center (SESYNC) under funding received from the National Science Foundation (NSF) DBI-1639145. The Northern Research Station, USDA Forest Service also provided support. L.E.D. acknowledges support from NSF OCE-2049360.

Author contributions

J.A., L.E.D. and K.K. conceived the study. J.A., L.E.D., R.E.-N. and K.K. designed the research. N.B.D.P. and R.G.H. contributed data or analytic tools. J.A. performed the research. J.A., L.E.D., R.E.-N. and K.K. wrote the paper and all authors edited the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41893-022-00913-9>.

Correspondence and requests for materials should be addressed to Jaime Ashander.

Peer review information *Nature Sustainability* thanks Jonathan Bossenbroek and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection N/A

Data analysis Analysis employed R version 4.0.2 (2020-06-22) using packages ggplot2 (v3.3.3), igraph (v1.2.5), quantreg (v5.61), Rglpk (v0.6-4). Full analysis code is available: 10.6084/m9.figshare.14402447.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

This study uses data on boat movement networks available from the Data Repository for the University of Minnesota: <https://hdl.handle.net/11299/216936>, and data on lake infection status from MN DNR. A compendium of data used is archived and available: 10.6084/m9.figshare.14402447. No figures have associated raw data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a computational study examining control of zebra mussel spreading between lakes via boat movements. We examine this problem using observed boat-movement networks in Minnesota and consider decisions at the county (n=58) level. We consider several network-guided methods of allocating control effort and compare them to a known-optimal method (integer-linear programming).
Research sample	This study uses an existing dataset estimating recreational boat movement in Minnesota USA (Kao et al 2020); the data is available from the Data Repository for the University of Minnesota: https://hdl.handle.net/11299/216936 .
Sampling strategy	The boat movement networks represent a unique, large-scale dataset for examining the feasibility and performance of network-guided management for invasive species. The counties (n=58) are not random but represent a real-world setting for management of an established invasive species. The replication of our problem across counties provides sufficient sample size to characterize differences in performance among network-guided management methods.
Data collection	The study uses previously-collected data: estimates of boat movement networks (Kao et al 2020; Data Repository for the University of Minnesota: https://hdl.handle.net/11299/216936) and public data on zebra mussel infestation status of Minnesota lakes..
Timing and spatial scale	The boat movement networks are based on data from user surveys 2014 to 2017. Lake infestation status (infected or not with zebra mussel) is from 2019.
Data exclusions	Data for counties with < 10 lakes were excluded from the study.
Reproducibility	All attempts to replicate the results succeeded and reproduction code is available.
Randomization	Randomization of treatments does not apply -- our study examines a unique dataset of boat movement networks for Minnesota, USA counties (n=58) and applies several potential management method to each county-level network.
Blinding	Blinding was not relevant as this is a non-experimental study.
Did the study involve field work?	☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging