FINAL REPORT

Quantifying the Stability of Fish Interaction Networks along Environmental Disturbance Gradients

SERDP Project RC-2506

DECEMBER 2016

Daniel J. McGarvey, Ph.D.
Virginia Commonwealth University

Joseph A. Veech, Ph.D.
Texas State University

Distribution Statement A
This document has been cleared for public release

SERDP
DOD ▪ EPA ▪ DOE

Page Intentionally Left Blank

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

Page Intentionally Left Blank

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*
22-12-2016

2. REPORT TYPE
Limited Scope Final Report

3. DATES COVERED *(From - To)*
01-05-2015 *to* 22-12-2016

4. TITLE AND SUBTITLE
Quantifying the Stability of Fish Interaction Networks along Environmental Disturbance Gradients

5a. CONTRACT NUMBER
W912HQ-15-P0001

5b. GRANT NUMBER
n/a

5c. PROGRAM ELEMENT NUMBER
n/a

5d. PROJECT NUMBER
15RC01-016

5e. TASK NUMBER
n/a

5f. WORK UNIT NUMBER
n/a

6. AUTHOR(S)
Daniel J. McGarvey (PI), Joseph A. Veech (Co-PI)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Environmental Studies, Virginia Commonwealth University, 1000 W. Cary St., PO Box 843050, Richmond, VA 23284-3050
Department of Biology, Texas State University, 601 University Drive, San Marcos, TX 78666

8. PERFORMING ORGANIZATION REPORT NUMBER
n/a

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Corps of Engineers Humphreys Engineer Center Support Activity, 7701 Telegraph Rd., Alexandria, VA 22315-3860

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP

11. SPONSOR/MONITOR'S REPORT NUMBER(S) n/a

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
n/a

14. ABSTRACT

This Limited Scope project used a large database of standardized stream fish samples from three biogeographical regions (East Coast, Mississippi River Basin, Pacific Northwest) to test the utility of graph theory and ecological network analysis within an applied context. Fish co-occurrence networks were first constructed in each region, using data aggregated at two spatial scales. Networks were then compared among sites that were ranked with a 3-level, ordinal scale of disturbance (good, fair, or poor condition). Network analyses were completed both for spatially-implicit data (unipartite graphs) and spatially-explicit data (bipartite graphs). Species extinction simulations were also performed to explore network responses to extinction events.

15. SUBJECT TERMS
Network analysis; graph theory; unipartite network; bipartite network; degree distribution; modularity; fish co-occurrence; environmental disturbance; community stability

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U (Unclassified)			Unclassified Unlimited	136	Daniel J. McGarvey
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U			804-828-7278

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Page Intentionally Left Blank

Table of Contents

Abstract.....1

Objective.....3

Background.....7

Materials and Methods.....10
Study Areas and Fish Data.....10
Research Tasks A_1-A_3: Building the Fish Co-occurrence Networks.....11
Research Tasks B_1-B_4: Measuring and Modeling Unipartite Network Stability.....15
Tasks B_1-B_2.....15
Task B_3.....16
Task B_4.....17
Research Tasks C_1-C_2: Bipartite Networks and Spatially-explicit Analyses.....19
Task C_1: Building Bipartite Networks.....19
Task C_2: Conducting Spatially-explicit Network Analyses.....20
Research Task D_1: Node Deletion/Extinction Experiments.....23

Results and Discussion.....25
Network Connectivity.....25
Unipartite Networks.....25
Bipartite Networks.....27
Network Modularity.....28
Functional Structure of Network Modules.....29
Node Deletion/Extinction Experiments.....32
Bipartite Networks.....32
Unipartite Networks.....35
Spatially-explicit Analyses.....39

Conclusions and Implications for Future Research.....47
Network Responses to Disturbance.....47
Dynamic Network Simulations.....54
Next Steps.....57

Literature Cited.....61

Appendix A: Exemplar Graphs and Degree Distribution Plots.....70
Appendix B: Network Metric Definitions.....95
Appendix C: Unipartite Network Summary Statistics.....98
Appendix D: Bipartite Network Summary Statistics.....116

List of Figures

Figure 1. Links between each of the proposed Research Tasks, Key Research Questions and the RCSON15-01 objectives and research needs. 5

Figure 2. Map of DoD facilities within each of the three major biogeographic regions 6

Figure 3. Three network graphs built from the same presence-absence matrix 8

Figure 4. Map of the NRSA/EMAP sampling sites 11

Figure 5. Effects of spatial scale on the unipartite networks 13

Figure 6. Weighted unipartite graph for the Ohio River Basin fish network....14

Figure 7. Spatially-explicit map of the bipartite fish network in the Kanawha River Basin 20

Figure 8. Examples of cumulative degree distributions....25

Figure 9. Examples of maximum species' degree values 26

Figure 10. Examples of maximum closeness centrality values 27

Figure 11. Examples of median closeness centrality values....27

Figure 12. Examples of modularity values 28

Figure 13. Functional hypervolume plots for fish co-occurrence networks 31

Figure 14. Plots of secondary extinctions in the four bipartite networks 33

Figure 15. Bipartite graphs illustrating sequential deletions of stream sites 34

Figure 16. Plots of species' extinction effects on a unipartite fish network 38

Figure 17. Unipartite graphs from extinction simulations 40

Figure 18. Linear regression results for HUC2 scale, spatially-explicit network analyses 43

Figure 19. Linear regression results for HUC4 scale, spatially-explicit network analyses 45

Figure 20. Weighted unipartite graphs for Ohio River Basin fish networks 48

Figure 21. Functional hypervolume plot demonstrating functional overlap among modules 52

Figure 22. Weighted unipartite graph for the Susquehanna River Basin fish network 54

Figure 23. Significantly negative associations within Ohio River Basin streams 58

List of Tables

Table 1. Summary of spatially-explicit unipartite results in Mississippi sub-drainages.....41

Table 2. Statistically significant, spatially-explicit results in Mississippi sub-drainages 42

Table 3. Species' degree values for Ohio River Basin fish networks.....50

Table 4. Lists of species 'hubs' within Ohio River Basin fish networks 51

Table 5. Species' sensitivity values for Susquehanna River Basin stream fishes 55

Acronyms

AUC – area under the curve

DoD – Department of Defense

DRN – dendritic river network

EMAP – Environmental Monitoring and Assessment Program

GIS – geographic information system

HUC – hydrologic unit code

NRSA – National Rivers and Streams Assessment

PCoA – principle coordinates analysis

SERDP – Strategic Environmental Research and Development Program

SON – statement of need

USEPA – United States Environmental Protection Agency

Keywords

Network analysis	Graph theory	Unipartite network
Bipartite network	Modularity	Fish co-occurrence
Degree distribution	Environmental disturbance	Community stability

Acknowledgements

We thank the Strategic Environmental Research and Development Program of the United States Department of Defense (SERDP contract W912HQ-15-P0001) for financial support. Additional administrative support was provided by Cara Patton (Noblis), Stephanie Lawless (Noblis), Sarah Barlow (Noblis), Eileen Spears (U.S. Army Corps of Engineers), and John Hall (U.S. Bureau of Land Management). Custom code to optimize clustering coefficient calculations for the bipartite networks (C++ code implemented within R) was graciously provided by Tore Opsahl (Bank of America). Taylor Woods (Virginia Commonwealth University) compiled functional traits data for all East Coast, Mississippi River Basin, and Pacific Northwest fish species. Miguel Aguilar and Grady Terry (Texas State University) assisted with spatially-explicit network calculations.

Abstract

Objective

In a rapidly changing, non-stationary world it is imperative for natural resource managers to understand how biotic communities may respond to a range of environmental disturbances. Predicting how these disturbances may affect the overall stability of biotic communities is of particular interest and is shifting the historical emphasis on species-level dynamics to a new focus on species' interconnections within large, complex communities. Network models are quickly becoming the ecologist's tool-of-choice for studying multi-species interconnections within biotic communities. Network graphs, which represent species as 'nodes' and interconnections as 'edges,' are powerful yet intuitive tools for documenting and visualizing species' connections at the system level. In addition, graph theory provides a mathematical framework for quantifying network structure and testing hypotheses on network assembly and stability; however, two key knowledge gaps currently limit the utility of the network approach. First, ecologists have so far focused on discrete mutualistic plant-pollinator or host-parasite interactions, or on antagonistic predator-prey interactions, but associations that are less easily classified (e.g., commensalism) or that are diffuse (i.e., realized through an intermediary) also can have strong influences on community structure and methods to incorporate them are now needed. Second, ecological network theory can predict how an acute extinction event (i.e., node removal) will influence the overall network, but it does not yet predict how ecological networks will respond to chronic environmental disturbances (i.e., those that do not cause acute extinctions). Chronic disturbances, such as physicochemical habitat degradation, are pervasive. And in an uncertain future, these types of chronic disturbances are likely to pose significant challenges to Department of Defense (DoD) managers. *The project objective was to address each of these two knowledge gaps—accounting for multiple types of species' associations and predicting network responses to chronic disturbances—using North American freshwater fishes as a model system.*

Technical Approach

The project incorporated an exceptionally large, standardized, pre-existing database of fish co-occurrences (the combined Environmental Monitoring and Assessment Program, and National Rivers and Streams Assessment datasets) within three United States (U.S.) biogeographic regions: the Mid-Atlantic coast, Mississippi River basin, and Pacific Northwest. In each region, fish co-occurrence networks were built for sampling sites that had been objectively classified, using a standardized scoring system based on benthic macroinvertebrates, as least, moderately, or severely disturbed. Different metrics of network stability, including connectivity and modularity, were then used to gauge network responses to environmental disturbance (i.e., comparisons among least, moderately, and severely disturbed sites). Species' extinction (i.e., node deletion) experiments

were also used to simulate network responses to species' extirpations and to detect secondary effects on the remaining species. Finally, a spatially-explicit framework to study and model species' associations within dendritic river networks was explored by superimposing the spatially-implicit fish networks on spatially-explicit, digital stream network maps.

Results

Network responses to chronic environmental disturbance were variable among biogeographic regions and in general, less clear than expected. Median node degree and closeness centrality did, however, change in a consistent, predictable manner with disturbance: degree decreased and closeness centrality increased with increasing disturbance. Together, these indicators suggest that fish co-occurrence networks become smaller (i.e., fewer species and fewer links among species) and more compact (i.e., shorter average distances among species' nodes) as disturbance increases. Notably, the identities of species' hubs (i.e., most highly connected species) did not change as the networks became smaller; decreasing degree values were attributable to deletions of peripheral species with low connectivity. In all networks, distinct modules of highly interconnected species were detected and the numbers of modules tended to increase with disturbance. Functional traits analyses showed that the morphological, ecological, and life-history characteristics of species within modules were highly similar, suggesting that modules may be repetitive functional 'motifs'. If these modules are functionally redundant, they may increase overall network stability; perturbations that disrupt part or all of a module may be compensated for by an independent module. Extinction simulations identified individual species that may be most vulnerable to primary and secondary extinctions, depending on the algorithm used to simulate disturbance (e.g., species least tolerant of pollution or species at sites closest to a major DoD facility).

Benefits

This Limited Scope project was a unique opportunity to generate new understanding of ecological networks within stream ecosystems and to begin to develop a process for incorporating this knowledge in conservation planning. Overall, we did not find definitive or easily generalizable results, but we did succeed in demonstrating an explicit process to quantify network structure and dynamics using widely available data on species' co-occurrences. Lessons learned are germane to most U.S. streams, as our sample data were broadly representative of stream conditions and fishes found throughout the U.S., as well as other types of ecosystems where the requisite co-occurrence data can be obtained. Results should be of particular value to DoD managers as they prepare to contend with uncertain environmental conditions in the future, given that fishes comprise the largest taxonomic group of at-risk vertebrate species on DoD property and thousands of kilometers of stream and river habitat are in close or immediate proximity to one or more DoD facilities.

Objective

The overarching goal of this Limited Scope project was to help Department of Defense (DoD) managers and the scientific community to prepare and plan for the effects of future environmental disturbances on the biota of stream and river ecosystems, using freshwater fishes as a model system. To do so, we studied fish co-occurrence networks within three major biogeographic regions of the United States. A co-occurrence network is a system of connections among species that can be used to detect interdependencies and/or to predict whether the effect of a disturbance is likely to be isolated to a single locale within the network, or to propagate throughout the network (see 'Background' below). Within ecology and environmental science, network analysis is emerging as a leading tool to study and model complex, highly interactive systems (Bascompte 2009; Proulx et al. 2005; Tylianakis et al. 2010). Our work sought to quantify the overall stability of fish co-occurrence networks, relative to a gradient of environmental disturbance, and to characterize the most critical functional roles and linkages within the networks. In the following pages, we explain how the novel, preliminary framework that we have developed may ultimately provide new tools to predict how stream and river biota may respond to future changes, and to establish specific management objectives that will protect the most essential structural and functional components of ecological networks.

Our Limited Scope project was responsive to each of the two major objectives specified in RCSON15-01: *New Paradigms for Managing Species and Ecosystems in a Non-Stationary World* (SON objectives 1 and 2). First, co-occurrence network or 'ecological network' research is a cutting-edge and rapidly growing sub-field of community ecology that can provide 'emerging theoretical and applied understanding . . . to address management challenges associated with. . .potential futures within relevant, testable, and adaptable conceptual frameworks' (SON objective 2). Second, by identifying dimensions of fish networks that have proven either robust or vulnerable to recent environmental disturbances, our methods and results can potentially help to predict and characterize the fish assemblages that DoD managers may be responsible for '25 or more years from now in a non-stationary world' (SON objective 1).

Specific *Research Objectives* of this Limited Scope project were to: (**A**) build networks of linked, co-occurring fish species within three biogeographic regions; (**B**) quantify network stability and responses to environmental disturbance via empirical observations; (**C**) create a spatially-explicit framework to study and model fish co-occurrence within dendritic river networks (DRNs); and (**D**) use extinction 'experiments' to simulate network responses to species and/or site removals. Specific *Research Tasks* performed under each Objective were to:

- **A_1**. Specify fish co-occurrence network topology in each of the three biogeographic regions.
- **A_2**. Build non-weighted, undirected, unipartite network graphs in each region.
- **A_3**. Build weighted, directed, unipartite graphs of the co-occurrence networks in each region.

- **B_1**. Stratify sample sites by environmental disturbance level ('least', 'moderate', or 'severe').
- **B_2**. Build independent fish co-occurrence networks for least, moderately, and severely disturbed sites in each biogeographic region.
- **B_3**. Compare network *connectivity* among least, moderately, and severely disturbed sites.
- **B_4**. Compare network *modularity* (i.e., the tendency for small, highly interconnected subgroups to form within the larger network) among least, moderately, and severely disturbed sites.
- **C_1**. Convert spatially-implicit, unipartite networks to spatially-explicit, bipartite networks.
- **C_2**. Reassess the previous unipartite and bipartite results within a spatially-explicit context.
- **D_1**. Perform 'node deletion' (i.e., species extinction) experiments to assess how fish co-occurrence networks may respond to extinction events.

By completing each of the above *Research Tasks*, we answered five *Key Research Questions*:

1. *Does fish network connectivity, as an index of overall network stability, change predictably when exposed to environmental disturbance in the form of chronic habitat degradation?*
2. *Does network modularity change in response to environmental disturbance?*
3. *Are modules within fish co-occurrence networks comprised of functionally similar species and if so, what are the common functional characteristics of their constituent species?*
4. *Do indicators of fish network stability have distinct spatial signatures when the unipartite fish co-occurrence networks are expressed as spatially-explicit, bipartite, DRNs?*
5. *Does the loss of a particular species or group of species pose a unique risk to network stability?*

Links between each of the above *Research Tasks*, *Key Research Questions*, and their respective RCSON15-01 objectives and research needs are illustrated in *Fig. 1*.

Ecological network analysis is a germane topic in ecological and environmental research. And the exploratory analyses detailed herein were conducted entirely with pre-existing data (see 'Methods and Materials' below). For these reasons, we did not explicitly target any particular DoD facility. However, lessons learned through this Limited Scope project may be highly relevant to DoD for at least two reasons. First, fishes comprise the largest group of vertebrate 'species at risk' on DoD property (36% of all at-risk vertebrates; NatureServe 2011). Our fish-specific results may therefore enhance DoD efforts to manage a relatively large number of high conservation priority species. Second, an abundance of stream and river ecosystems occur in close or immediate proximity to DoD facilities. In each of the three biogeographic regions that were studied – the East Coast (Chesapeake Bay to Savannah River, GA), the Mississippi River Basin, and the Pacific Northwest – there are between 34 and 145 DoD facilities (*Fig. 2*), many of which lie within a 10 km radius of fish-bearing streams/rivers. Also, a large amount of stream/river habitat lies entirely within the

boundaries of some DoD facilities. For instance, 588 km of stream channel occur within the grounds of Fort Bragg (NC) and 1,174 km occur within the Yakima Training Center (WA). Thus, the footprint of DoD activities on stream and river fishes of the U.S. is potentially very large.

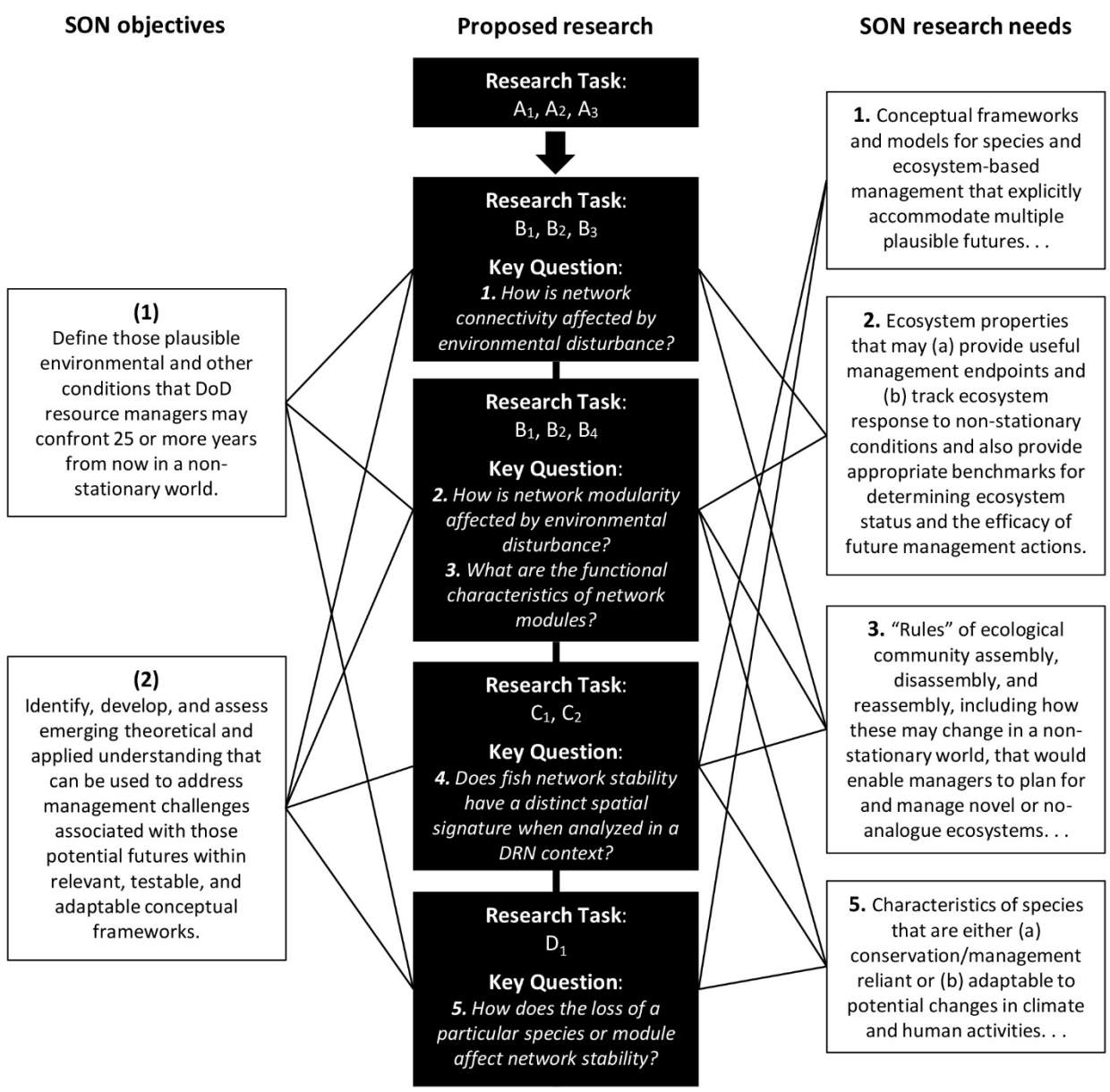


Figure 1. Links between each of the proposed Research Tasks, Key Research Questions and the RCSON15-01 objectives and research needs. Research Tasks A1, A2, and A3 are not directly linked to any of the SON objectives or research needs, but were prerequisite to each of the subsequent Research Tasks. Note that the major Research Tasks deviate slightly from the original Limited Scope Proposal; node deletion experiments were performed for *both* unipartite and bipartite networks and were therefore changed from Task B_5 to Task D_1.

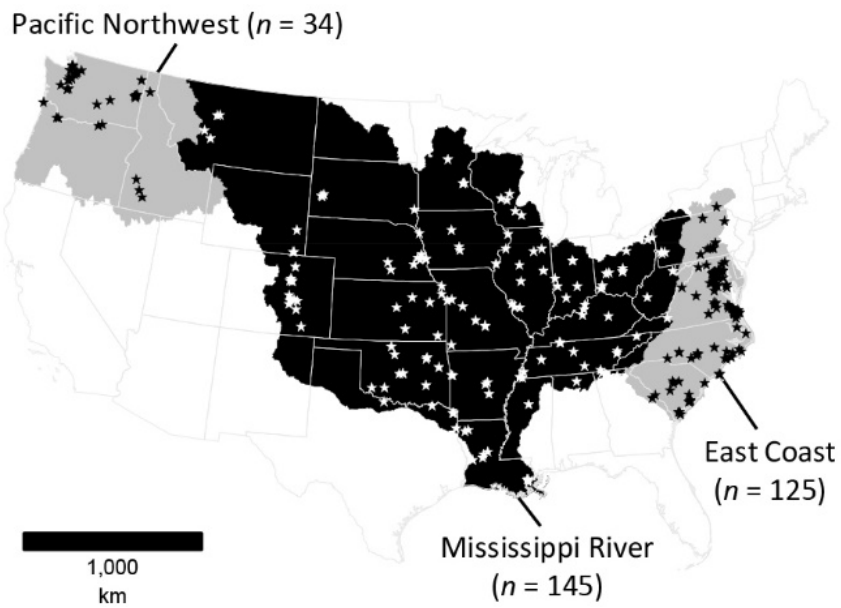


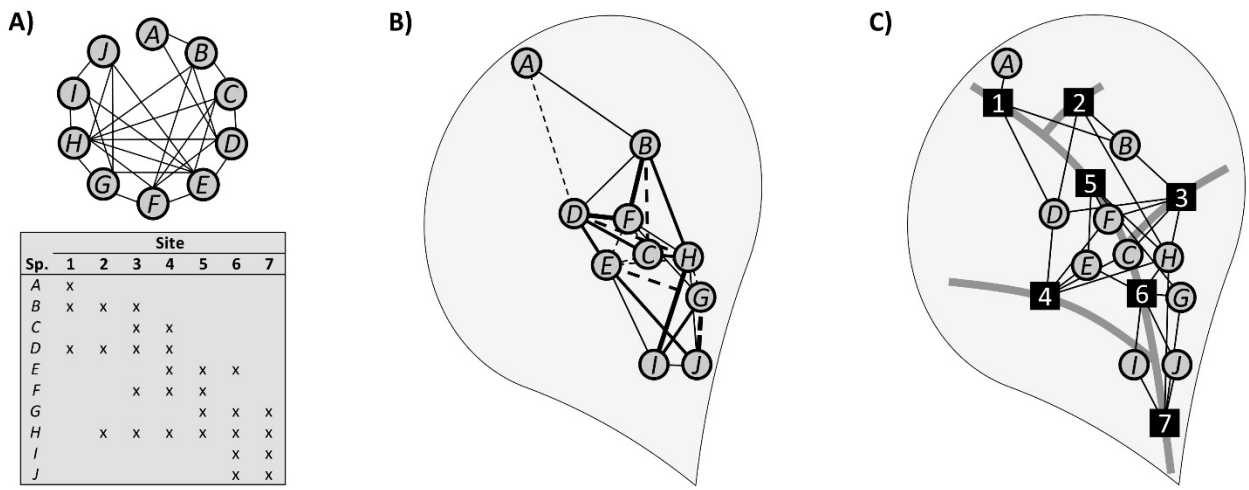
Figure 2. Map of DoD facilities (shown as stars) within each of the three major biogeographic regions included in this Limited Scope project: the East Coast, the Mississippi River Basin, and the Pacific Northwest. Total numbers of facilities are shown in parentheses for each biogeographic unit.

By demonstrating the methods used to build and analyze ecological networks, then showing that these networks can be used to detect environmental disturbances and/or to predict their potential effects on species within the networks, we believe that our Limited Scope project sets the stage for a full Strategic Environmental Research and Development Program (SERDP) proposal. To date, most ecological network research has focused on basic theory, with little attention paid to applied conservation questions (Tylianakis et al. 2010). We anticipate that a full SERDP project will allow us to achieve four future objectives: (1) confirming that pair-wise links within the unipartite networks do (or do not) reflect true biological associations; (2) testing a hypothesis of functional motif redundancy among network modules (see 'Results and Discussion'); (3) extending the simulation algorithms used in this study to better incorporate chronic or gradual disturbances and spatial dynamics within DRNs; and (4) building user-friendly tools to make the network analyses and visualizations widely accessible to managers and conservation practitioners.

Background

In a rapidly changing, non-stationary world it is important to understand how biotic communities respond to environmental disturbances (Blois et al. 2013). Efforts to enhance this understanding have recently been motivated by questions regarding the *stability* of biotic communities (Allesina and Tang 2012; Melián and Bascompte 2002; Thébault and Fontaine 2010; Tylianakis et al. 2010). These efforts are shifting the scientific community's focus from a historical emphasis on the dynamics of individual populations (Harrison 1979; Holling 1973) to a more holistic emphasis on the interconnections among large numbers of co-occurring species (Proulx et al. 2005). For example, researchers may now seek to predict whether an extinction event within a food-web is likely to trigger a cascade of secondary or tertiary extinctions, or whether functional redundancies (e.g., alternative hosts or food resources) may act as extinction buffers (Allesina et al. 2009; Dunne et al. 2002b; Lafferty and Kuris 2009; Rodriguez-Cabal et al. 2013; Saterberg et al. 2013).

The network model is a powerful tool for studying complex, multi-species interconnections within biotic communities (Bascompte and Jordano 2007; Newman 2003; Proulx et al. 2005). Network graphs, in which species are depicted as 'nodes' and co-occurrences are depicted as 'edges' or 'links' (*Fig. 3*), are an intuitive way to visualize species' connections at the system level (Fortuna and Bascompte 2008; Poisot et al. 2012). And graph theory provides a rigorous mathematical framework to quantify network structure and test hypotheses on network assembly and stability (Albert and Barabási 2002; Gross and Yellen 2006; Newman et al. 2000). For instance, the distribution of species 'degree' (the number of edges connecting a focal species to other species, also referred to as 'connectivity') can be used to gauge the overall 'robustness' (the tendency to remain a cohesive network, rather than breaking into disconnected fragments that are themselves less likely to persist in the future) of the network to extinction events (i.e., node deletions) (Albert et al. 2000). Ecological networks often exhibit power-law degree distributions, with many poorly connected species and a few highly connected 'hub' species (Camacho et al. 2002; Jordano et al. 2003; Montoya and Solé 2002). Theory predicts that these power-law networks will be robust to disturbance because random extinctions will tend to remove weakly connected nodes, leaving hub species and their links intact (Jordán 2009; Tylianakis et al. 2010). However, these same networks may quickly become unstable if a hub species is removed (Solé and Montoya 2001). Co-occurrence networks can also be used to address questions that are not amenable to more conventional statistical methods. For example, ecologists often use ordination or clustering to identify highly associated groups of species, then attempt to model group affinities as a function of independent predictor variables (Gotelli and Ellison 2004). These multivariate tools work well in some applications, such as quantifying and explaining biodiversity gradients (McGarvey and Ward 2008; Niu et al. 2012), but they do not address interconnectedness and system-level stability as effectively as some network metrics (Carstensen et al. 2013).



Sp.	Site 1	2	3	4	5	6	7
A	x						
B	x	x	x				
C			x	x			
D	x	x	x	x			
E				x	x	x	
F			x	x	x		
G					x	x	x
H		x	x	x	x	x	x
I						x	x
J						x	x

Figure 3. Three network graphs built from the same presence-absence matrix (inset lower left, with species represented as rows, sites represented as columns, and species' presences at a given site represented by an 'x'). The non-weighted, undirected, unipartite graph at upper-left (**A**) shows spatially-implicit links (co-occurrences) among species (gray circles). A weighted and directed, unipartite graph is shown for the same co-occurrence data at center (**B**); solid lines indicate (+) links, dashed lines reflect (-) links, and line thickness indicates link strength. The graph in (**B**) is also spatially-explicit; species' nodes are located at the centroids of their regional distributions (i.e., locations within the shaded river basin). A spatially-explicit, bipartite graph is shown at right (**C**); sampling sites (black numbered boxes) are located along the dendritic river network (heavy gray lines) and superimposed species' nodes are positioned as in (**B**), but links are only drawn between species and their known locations.

Yet despite much recent progress in ecological network research, two key knowledge gaps remain before network analyses can realize their full potential in natural resource management. First, most ecological network studies have focused solely on mutualistic plant-pollinator or host-parasite associations, or on antagonistic predator-prey links (Allesina and Tang 2012; Dunne et al. 2002a; Olff et al. 2009; Sauve et al. 2014; Thébault and Fontaine 2010). These discrete associations are directly observable and have therefore proven useful in empirical tests of network structure and stability. However, associations that are less conspicuous (e.g., commensalism) or that are diffuse (realized through an intermediary) can also have important influences on community structure (Callaway and Pennings 2000; Pianka 1974; Saterberg et al. 2013) and methods to incorporate them in ecological networks are needed, but currently lacking (Araújo and Rozenfeld 2014; Bascompte and Jordano 2007; Poisot et al. 2012; Thébault and Fontaine 2010). Second, a mature body of theory to predict ecological network (biotic community) responses to environmental disturbances, particularly chronic disturbances, does not yet exist (Melián and Bascompte 2002; Tylianakis et al. 2010; Tylianakis et al. 2007). Network models and graph theory are rooted in studies of communication and transmission networks, such as the internet and the spread of

infectious disease (Albert and Barabási 2002; Gross and Yellen 2006), and it is not yet clear which properties of these systems are directly applicable to ecological networks (Proulx et al. 2005; Tylianakis et al. 2010). For example, disturbance and node deletion are often synonymous concepts in network studies (Albert et al. 2000; Allesina et al. 2009; Callaway et al. 2000). But in ecological systems, richness (the number of nodes) may actually *increase* in response to disturbance (Connell 1978; Huston 1979; Olden et al. 2004; Rodriguez-Cabal et al. 2013).

Work conducted under this Limited Scope Project addressed each of these two challenges – the lack of methods to account for species' associations that cannot easily be classified as obligate mutualisms or predator-prey associations and the lack of theory to predict network responses to chronic environmental disturbance – using North American freshwater fishes as a model system. Research Tasks A1-A3 (see 'Objectives' and *Fig. 1* above) sought to determine which fish species pairs have significantly positive and/or negative associations. This was an efficient way to infer the strength and direction of species' associations at large spatial scales, when direct observation is not feasible (Araújo and Rozenfeld 2014; Veech 2006). Research Tasks B1-B4 constituted what we believe is the first empirical, large-scale study of chronic environmental disturbance effects on network structure and stability. We focused primarily on two indicators of network stability: connectivity and modularity (Bascompte and Jordano 2007; Olesen et al. 2007). Through Research Task B5, we demonstrated how dynamic network analyses could be used by DoD managers to simulate future extinction events and their effects on ecological networks. Finally, under Research Tasks C1-C2, we performed a preliminary, spatially-explicit analysis of bipartite networks that were mapped to DRNs. These spatially-explicit analyses provide both general insight to the stability of ecological networks and site-specific knowledge that can be directly applied in local management and conservation programs.

Materials and Methods

Study Areas and Fish Data

We built fish co-occurrence networks in each of three biogeographic regions: the East Coast, the Mississippi River Basin, and the Pacific Northwest. Three criteria were used in selecting these regions. First, each region has a distinct fish fauna with a unique evolutionary history (Hocutt and Wiley 1986; Smith et al. 2010). By comparing regions, we sought to distinguish fish network patterns that are specific to a given assemblage or environment from those that are common to all regions and may therefore reflect convergent underlying processes. Second, future environmental threats and challenges are expected to differ among regions. For instance, forecasted increases in annual temperature and water yield throughout much of the Upper Mississippi River Basin (Jha et al. 2004) may facilitate enhanced corn (ethanol) production, but at the cost of elevated rates of nutrient (nitrogen and phosphorous) application and runoff (Secchi et al. 2010). Urbanization impacts on upland streams and rivers (and their resident biota) throughout the East Coast region are likely to increase as sea-level rise (Sallenger et al. 2012) pushes coastal residents further inland (Rogers and McCarty 2000). And diminished snowpack has begun to alter the timing and magnitude of stream flows (higher in winter and spring, lower in summer and fall) throughout the Pacific Northwest (Miles et al. 2000; Mote et al. 2003). Thus, by including the three regions in this study, will hoped to ensure that our results are robust and relevant to managers with diverse concerns. Third, each region contains a large number of DoD facilities and an abundance of streams and rivers that are near a DoD site (see *Fig. 2*); lessons learned in this project should therefore be directly relevant to DoD managers.

Fish co-occurrence networks were built using standardized field samples from the Environmental Monitoring and Assessment Program (EMAP; USEPA 2002) and the National Rivers and Streams Assessment (NRSA; USEPA 2013) databases, which are both publicly available through the United States Environmental Protection Agency (USEPA). EMAP was the first national-scale survey of fish assemblage structure, as well as aquatic invertebrates, water chemistry, and physical habitat, within streams and rivers of the conterminous United States (Hughes et al. 2000). NRSA was the successor to EMAP (USEPA 2013). Sample site locations in both programs were selected *a priori* through a stratified random sampling design that accounted for differences among major types of river and stream ecosystems, or ‘ecoregions’; by randomly selecting sample sites within each ecoregion, USEPA was able to (i) perform statistically robust assessments of the health of the complete population of streams and rivers within each ecoregion, and (ii) scale the regional assessments up to overall, national-scale summaries of stream and river health (USEPA 2002). Importantly, standardized field protocols were prepared prior to the start of the EMAP and NRSA sampling schedules (each of which was distributed over multiple years) and field crews were rigorously trained prior to sampling events (Lazorchak et al. 2000; Lazorchak et al. 1998; USEPA

2007, 2010). A large number of preliminary samples and data analyses were also used to ensure that the final, standardized sampling methods would be adequate to fully characterize biotic and abiotic conditions at each sample site (Cao et al. 2001; Hughes et al. 2002; Reynolds et al. 2003). The combined EMAP/NRSA dataset is now the largest, most spatially extensive record of standardized fish, invertebrate, and physicochemical stream and river data ever assembled in the conterminous United States (see *Fig. 4*). Additional information on the regional distributions of individual fish species were obtained from the NatureServe database 'Digital Distribution of Native U.S. Fishes by Watershed, Version 3.0' (NatureServe 2010).

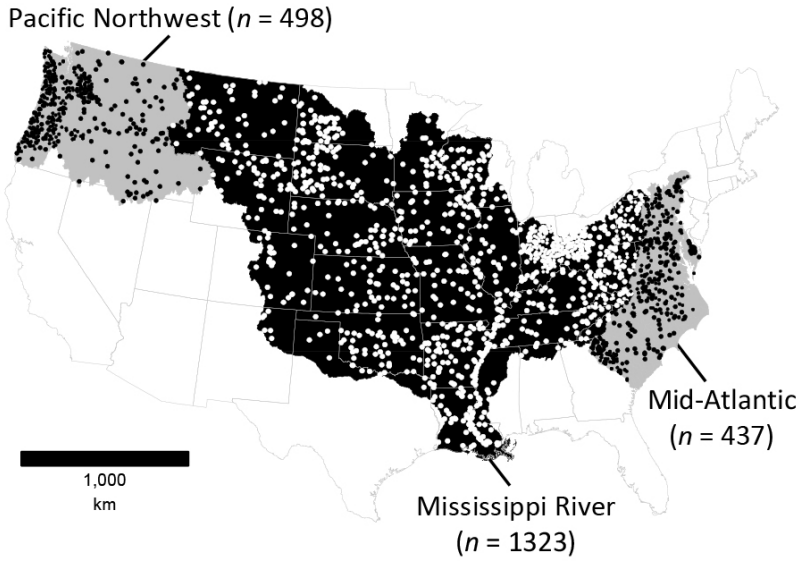


Figure 4. Map of the NRSA/EMAP sampling sites (shown as points). Major biogeographic regions are indicated by shaded polygons. The total number of sampling sites within each region is shown in parentheses.

Research Tasks A_1-A_3: Building the Fish Co-occurrence Networks

Network topology – the 'blueprint' of species co-occurrences at individual sampling sites that all network graphs were built from – was specified by combining EMAP/NRSA fish samples within presence-absence matrices.[1] Separate presence-absence matrices were compiled for each of the three biogeographic regions, with the following regional dimensions (i.e., total number of fish species × total number of sampling sites): *East Coast* = 227 species × 437 sites; *Mississippi* = 339

[1] Note that species' absences can correctly be inferred for EMAP/NRSA samples because standardized protocols with known detection efficiency were used (see *Study areas and fish data*). Failure to observe a species at a given site is therefore truly informative; the EMAP/NRSA data are not subject to the limitations of presence-only data.

species × 1323 sites; *Pacific Northwest* = 60 species × 498 sites. Presence-absence matrices were then converted to 'edge lists' (i.e., two-column tables representing associations between species pairs) for use in building unipartite networks. Unipartite networks and their graphs were then constructed from the regional edge lists in each of the three biogeographic regions. We first built non-weighted, undirected, unipartite graphs in each region. Next, we added 'weights' and 'directions' to the unipartite networks. To do so, we used the probabilistic co-occurrence model of Veech (2013, 2014).

The 'probabilistic model of species co-occurrence' (Veech 2013, 2014) provides a systematic and quantify tool to infer the direction (i.e., positive versus negative co-occurrence patterns) and overall significance or strength of every pair-wise species' association within a given dataset (i.e., edge list). Briefly, the probabilistic model of co-occurrence calculates the number of ways that two species can co-occur at exactly j number of sampling sites, given that each species occurs at N_1 and N_2 number of sites out of a total of N. From this, the model calculates the probability that two species would co-occur at more than (i.e., positive associations) or fewer than (i.e., negative associations) the observed number of sites where the two species do, in fact, co-occur (Veech 2013). All co-occurrence probability calculations were performed with the '*cooccur*' package (Griffith et al. 2014) in R (R Core Team 2016).

Importantly, the probabilistic model of co-occurrence can be influenced by spatial scaling artifacts (Veech 2006, 2013). We therefore performed a series of preliminary co-occurrence probability and unipartite network comparisons, using four different spatial scales, to identify an optimal scale (or scales) for the remainder of the network analyses. These spatial scales were represented by U.S. Geological Survey 'Hydrologic Units' within the National Watershed Boundary Dataset (USGS 2013). Using a large subset (approximately 1/3 of all samples) of the Mississippi River Basin dataset, we performed a full suite of co-occurrence and unipartite network analyses at the 'HUC0' (largest spatial units), 'HUC2', 'HUC4', and 'HUC6' (smallest spatial units) scales.

Results of the spatial scale comparisons were highly variable, confirming the importance of spatial scale in co-occurrence and network analyses. For example, we plotted three of the parameters from power-law models of network degree distribution (power-law model constants, exponents, and coefficients of determination; see below for review of the importance of network degree distributions and power-law models) and network modularity values (see below) for subsets of data that were partitioned at each of the four spatial scales (*Fig. 5*). Moreover, we repeated this comparison for networks that were built using all data, as well as networks that were built exclusively from significant paired species' associations (inferred from the probabilistic co-occurrence model). We observed large differences in overall results among all four scales (*Fig. 5*). Upon closer inspection of the fish database, however, we observed that sample sizes within the HUC6 spatial units (i.e., decreasing the size of the sub-drainages used as 'replicates' in co-occurrence and network analyses) were generally too small to support robust network analyses,

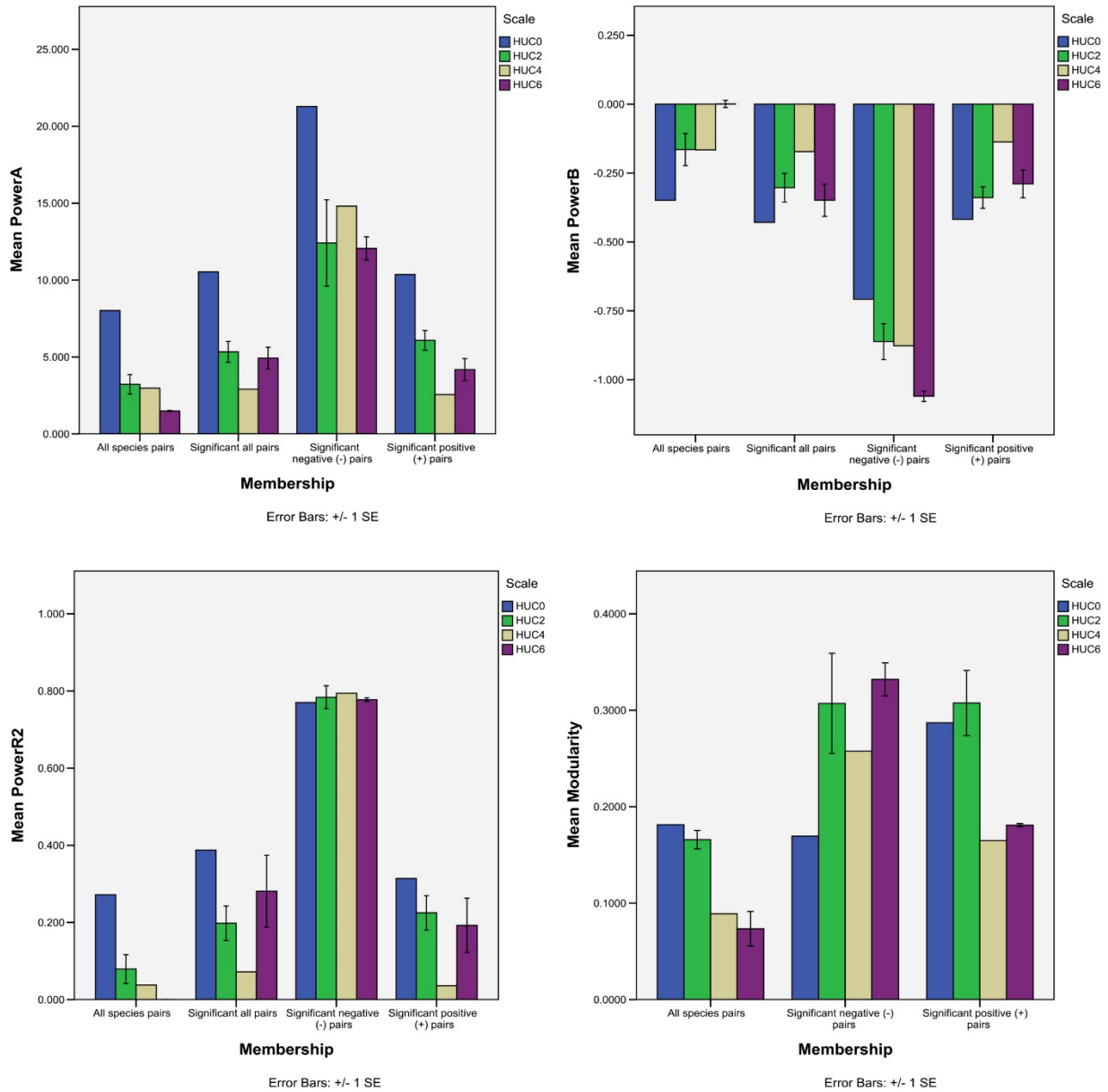


Figure 5. Effects of spatial scale on four unipartite network metrics: the constants ('PowerA'), exponents ('PowerB'), and coefficients of determination ('PowerR2') from power-law degree distribution models, and mean modularity. Results are partitioned by spatial scale (HUC0 – HUC6) and by separation into distinct networks for complete datasets ('all species pairs'), all significant species' pairs, significantly negative pairs, and significantly positive pairs using the probabilistic model of co-occurrence (see main text).

particularly when comparing networks that were further subdivided by disturbance level (see below). At the opposite extreme, we did not perform network analyses at the HUC0 scale because these spatial units were too large (e.g., the entire Mississippi River Basin or Columbia River Basin) to satisfy the core assumption of 'access to all potential co-occurrence sites' that is invoked by the

probabilistic model of co-occurrence (see Veech 2013). We therefore performed all of our analyses at the HUC2 and HUC4 spatial scales, repeating all data analysis procedures at each scale. (Results are appended with 'HUC2' or 'HUC4' labels throughout the remainder of this Final Report.) We also repeated all co-occurrence and network analyses using a 'five or more occurrences per species' criterion to minimize bias in the co-occurrence significance results that may result from including species with too few observed occurrences (i.e. < 5) to reliably estimate co-occurrence probabilities. This second set of more conservative test results is indicated throughout the Final Report by a '5+' label. An example of a weighted unipartite graph is shown at the HUC2 scale for all co-occurring fishes within the Ohio River Basin in *Fig. 6*. Additional examples of weighted unipartite graphs are shown for each of the three biogeographic regions, at both the HUC2 and HUC4 scales, in *Appendix A*.

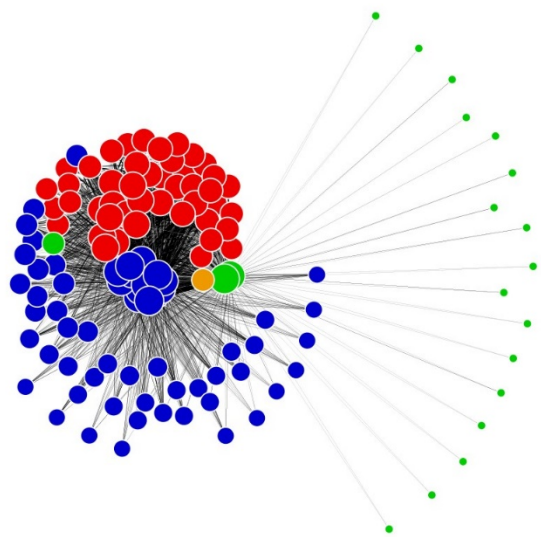


Figure 6. Weighted unipartite graph for the Ohio River Basin fish network. The size and color of each node depict its degree (i.e., connectivity) and module membership (colors are arbitrary, showing only mutual membership within modules; see below), respectively. The thickness and color of each link depict edge weights (thicker lines represent larger weights) and directions (black lines = positive associations, grey lines = negative associations).

Sixteen network metrics were then calculated for each of the unipartite networks (see *Appendix B* for definitions of all network metrics). We began by focusing on the degree values of individual nodes within the networks, calculating both the maximum and median degree within each network. Next, we fit a power-law model to the complete, cumulative degree distribution from each network (power-law degree distributions are often associated with robust networks; see 'Research Tasks B_1-B_5' below), noting the alpha (α) value and *p*-value from a Kolmogorov–Smirnov (K-S) test of model fit. Truncated power-law models (i.e. power-law models that only consider the region of

greatest curvature in cumulative distributions, effectively truncating the complete distributions at an empirically determined minimum *x*-value) were also fit to the network degree distributions, noting the minimum *x*-values, as well as the truncated α and truncated K-S *p*-values. We then calculated the maximum and median betweenness centrality of each node within the networks, as well as the maximum and median closeness centrality of each node. The global clustering coefficient was calculated for each network, in addition to the maximum and minimum local clustering coefficients. Finally, modularity and the number of distinct modules were calculated, using a 'simulated annealing' algorithm (Newman 2006; Olesen et al. 2007). All unipartite networks were built and analyzed with the '*igraph*' package (Csárdi and Nepusz 2006) in R.

Research Tasks B_1-B_4: Measuring and Modeling Unipartite Network Stability

Tasks B_1-B_2 – Stratify sample sites by environmental disturbance level then build independent fish co-occurrence networks for sites in 'good', 'fair', and 'poor' condition

A major objective of this Limited Scope project was to empirically document changes in network structure and stability, relative to an environmental disturbance gradient. To accomplish this, we applied the three disturbance rankings – 'good', 'fair', or 'poor' condition – that were assigned by USEPA to individual EMAP/NRSA sampling sites. This classification system was designed to provide an integrated measure of disturbance that would account for multiple types of anthropogenic activities. USEPA assigned each sample site to one of the three disturbance categories using standardized scoring rubrics for both physicochemical data (e.g., nutrient concentrations, turbidity, substrate characteristics, and riparian habitat quality; USEPA 2002, 2013) and benthic macroinvertebrate data (Herlihy et al. 2008; Stoddard et al. 2008; Stoddard et al. 2006; USEPA 2016). Initially, we planned to use the site disturbance rankings based on physicochemical data. However, upon closer inspection of the data, we learned that the physicochemical variables used to classify sites was not entirely consistent among the NRSA and EMAP surveys (see Lazorchak et al. 1998; USEPA 2007, 2016). We therefore used the benthic macroinvertebrate data to classify sites by a uniform system of disturbance ranks, following the detailed method outlined in Herlihy et al. (2008), Stoddard et al. (2008), and USEPA (2016). Notably the use of benthic invertebrates may have been an unexpected strength of our study, as invertebrate-based biological assessments a well-documented history of use in freshwater ecology and are thought to better reflect chronic disturbances through time than discrete water quality samples (e.g. Chessman 1995; Growns et al. 1997; Hawkins et al. 2000). Among East Coast sample sites, we found 0.35, 0.30, and 0.35 of all sampling sites to be in good, fair, and poor condition, respectively. Among Mississippi River Basin sites, 0.35, 0.25, and 0.40 of all sites were identified as being in good, fair, and poor condition. Among Pacific Northwest sites, 0.66, 0.16, and 0.18 of all sites were in good, fair, and poor condition.

Following procedures outlined above, we built new unipartite networks for sites in good, fair, and poor condition, in each of the three biogeographic regions, then re-calculated each of the 16 network metrics (see *Appendix B*) for the new networks using '*igraph*' in R. This process was repeated for the complete compliment of data management factors described above (i.e. by spatial scale, for significantly positive and negative co-occurrences, and for both complete species' lists and '5+ occurrences' lists). This repetitive process allowed us to examine network structure from multiple perspectives and to evaluate whether some network analysis results were unique to (or artifacts of) a particular data management scheme. However, partitioning the large yet finite fish datasets among so many competing factors did lead to many data analysis iterations (i.e., combinations of biogeographic region, spatial scale, disturbance level, etc.) where the data were no longer sufficient to complete all network analyses. These data-insufficient iterations are indicated by 'gaps' among the summary bar charts shown in *Appendix C* and *Appendix D*.

Task B_3 – Compare connectivity among 'good', 'fair', and 'poor' condition networks

Our plan to assess network connectivity was rooted in the concept of node *degree* and the importance of the cumulative *degree distribution*. At the species level, connectivity is the number of associations that a focal species shares with other species, also known as species' degree. At the network level, connectivity becomes an integrated measure of species' interconnectedness and can be expressed as the frequency distribution of edges (i.e., the degree distribution) between all co-occurring species. Connectivity is central to network stability for the following reason: theory predicts that a network with many weakly connected species and a few highly connected species, which will exhibit a power-law degree distribution, is less likely to be 'damaged' by a random extinction event (i.e., odds are that a weakly connected species will be lost, with minimal effects on remaining species) than a network with many highly connected species, which should exhibit degree distributions with truncated power-law or exponential curves (Bascompte and Jordano 2007; Jordán 2009; Tylianakis et al. 2010). More specifically, the loss of a weakly-connected species may have isolated effects, but the loss of a highly-connected hub species may split the network into fragments (Albert et al. 2000; Camacho et al. 2002; Solé and Montoya 2001).

In each biogeographic region, we compared networks in good, fair, and poor condition to assess whether connectivity, as represented by the shape of the degree distributions, changes in a consistent way with disturbance level. Because connectivity and species' degree were so central to our project, we also calculated three alternative measures of node connectivity – betweenness centrality, closeness centrality, and clustering coefficients (see *Appendix B*) – then compared these among the good, fair, and poor condition networks. Connectivity calculations were performed as above using the '*igraph*' package in R.

Task B_4 – Compare modularity among 'good', 'fair', and 'poor' condition networks

Network modularity is the tendency for nodes (i.e., species) to cluster together within small but highly interconnected sub-networks, or 'modules' (Olesen et al. 2010). It is central to ecological network stability because modules are often thought to consist of species with similar functional roles, and those roles may not exist elsewhere in the network (Dalsgaard et al. 2008; Olesen et al. 2007). If this is generally true, modules can potentially serve as fundamental conservation units: losing a complete module and its functional role(s) may cause the collapse of an entire network, and so a prudent management strategy would be to 'save a piece' of each module (Olesen et al. 2007; Tylianakis et al. 2010).

Following the logic described above for connectivity analyses, we compared modularity among networks in good, fair, and poor condition to assess whether modularity differs in a predictable way with increasing level of disturbance. Modularity was calculated for each of the networks (3 biogeographic regions × 3 disturbance levels) using the Q metric and a 'spectral partitioning' optimization algorithm (Newman 2006). All modularity calculations were performed as above using '*igraph*' in R.

We then assessed whether network modules reflected consistent themes in aggregated functional traits. Functional traits analysis, which replaces the traditional community ecology emphasis on taxonomic identity with information on the functional roles that individuals and species play within ecosystems, has become one of the most prolific and powerful tools for explaining patterns in fish assemblage structure (e.g. Blanck et al. 2007; Cilleros et al. 2016; Frimpong and Angermeier 2009; Goldstein and Meador 2004; Hoeinghaus et al. 2007; McGill et al. 2006; Meador and Brown 2015; Olden et al. 2006; Winemiller and Rose 1992). To perform the traits-based analyses, we first compiled a master database of species-level functional traits. This database included trait characters for 637 fish species, distributed throughout the continental U.S. It was assembled by first querying major repositories of fish traits data, including the *FishTraits* (Frimpong and Angermeier 2009) and *FishBase* (www.fishbase.org) databases, followed by regional atlases (e.g. *Freshwater Fishes of Virginia*; Jenkins and Burkhead 1994) and ultimately, searches in the primary and secondary literature. In total, 250 sources were consulted in preparing the master fish traits database. The complete database included 30 traits, many of which were not available across all 637 fish species (i.e., the final traits matrix included many data gaps). From these 30 traits, we selected 13 traits that were reasonably complete and represented multiple types of functional and life-history characteristics (e.g. body size, reproductive behavior, habitat use, and feeding behavior), for use in our analyses. The final list of traits included: maximum total length; maximum total length at female sexual maturity; female age at sexual maturity; maximum fecundity; egg diameter; maximum longevity; parental care (ordinal scale ranging from 0−4); typical holding depth (demersal, epibenthic, benthopelagic, pelagic); primary adult habitat (pool, channel, riffle, backwater, lake, estuary); adult feeding behavior (herbivore, detritivore, insectivore, planktivore,

piscivore, omnivore, parasitic); egg laying strategy (broadcast spawn, egg hider, nest builder, cavity spawner, attachment, live bearer); primary spawning season (spring, summer, fall, winter); and migratory behavior (anadromous, potadromous, catadromous, sedentary/non-migratory). Traits data were prepared for multivariate analysis following standard methods (e.g., Laliberté and Legendre 2010; Pla et al. 2012). The final standardized species × traits matrix in each of the three biogeographic regions was then converted to a Gower distance matrix for use in subsequent functional traits analyses. Gower distance is the preferred distance metric in functional traits analysis because it can accommodate multiple data types (continuous, ordinal, and categorical) and is robust to missing values within species × traits matrices (Laliberté and Legendre 2010; Podani et al. 1999). Gower distance matrices were calculated with the '*FD*' package (Laliberté et al. 2014) in R.

A two-stage process was then used to assess whether observed modularity patterns within the unipartite networks could reasonably be attributed to the functional traits of species within each module. First, cluster analyses (based on the Gower distance matrices) were used to partition species within datasets into 'functional clusters.' Because our goal was to assess the hypothesis that network modules reflect differences in species' functional traits, we sought to detect 'congruence' between species' clusters based on functional traits and on network modularity. Thus, we began with the assumption that the number of functional trait clusters should be equivalent to the number of modules for a given dataset. We did not, however, bias our analyses by specifying *a priori* which functional cluster a species should belong to. Rather, we used *k*-means cluster analysis, with the number of prior clusters equivalent to the observed number of modules in a given network, to organize species by functional traits and to search for similarity or congruence between the two classification methods (i.e. network modules versus functional traits clusters). *K*-means cluster analyses were performed with the '*FD*' package in R.

In the second stage of the functional traits assessment, Principle Coordinates Analysis (PCoA) was then used to ordinate and visualize network modules and their constituent species within multivariate functional traits spaces. Results from the functional traits *k*-means cluster analyses (used above to assess congruence between network module classifications and functional traits classifications) were not included in these ordinations. Rather, we used the first two PCoA axes to summarize functional traits for the species within each module and to assess whether modules had distinct or broadly overlapping functional profiles that could potentially explain observed modularity patterns within the unipartite networks. To visualize the functional traits of the modules, we plotted the constituent species of each module, as well as the centroid of each module (i.e., the average x and y coordinates of all individual species' points within a given module, calculated within a 2-dimensional Cartesian coordinate system where the first two PCoA axes are used as the x and y coordinates), along the first two PCoA axes. These plots were generated as 'hypervolumes' in which kernel density estimates (i.e., continuous probability distributions along each of the two PCoA axes) for each module were inferred (treating species within each module

as samples) and used to generate 2-dimension boundaries (essentially 95% confidence regions) about the functional traits space occupied by each network module (Blonder et al. 2014). Functional hypervolumes were built and plotted with the '*hypervolume*' package (Blonder et al. 2014) in R.

Research Tasks C_1-C_2: Bipartite Networks and Spatially-explicit Analyses

Task C_1 – Build bipartite networks

Bipartite networks and their graphs were constructed using the same data schemes described above for unipartite networks: EMAP/NRSA fish were assigned to the three biogeographic regions then partitioned among sub-drainages at the HUC2 and HUC4 spatial scales. Each of the 16 network metrics calculated for unipartite networks (see *Appendix B*) was again calculated for bipartite networks, focusing exclusively on site-level calculations, rather than species-level calculations. When working with bipartite networks (i.e., two distinct types of nodes), metrics that are calculated at the individual node-level (e.g., degree, betweenness centrality, and local clustering coefficient) can be calculated for both types of nodes. We only considered site-level metrics because our intent for the bipartite network analyses was to determine whether bipartite networks can provide novel understanding of fish assemblage structure when using site-level information to 'map' the bipartite networks to spatially-explicit, dendritic river networks (see below). All bipartite networks were built and analyzed with the '*bipartite*' package (Dormann et al. 2008) in R.

The bipartite networks were then converted to spatially-explicit network maps (e.g., *Fig. 3C*) by: (1) appending the *x* and *y* coordinates (i.e., longitude and latitude) for each sampling site to their corresponding network nodes; (2) appending the *x* and *y* coordinates interpolated for the regional (i.e., within a given HUC2 or HUC4 sub-drainage) centroid of each species' range to its respective bipartite node; and (3) superimposing the spatially-explicit, geo-referenced graphs on top of their respective drainage basin and stream networks maps (i.e., sample sites were aligned atop their true locations on the stream network maps). However, it was immediately apparent that these spatially-explicit bipartite graphs contained too much 'noise' to be useful for visualization purposes. For example, a spatially-explicit bipartite graph of all significantly positive links within the Kanawha River Basin is shown in *Fig. 7* (note that this graph was much less dense than the complete network graph with all co-occurring species, which is not shown here). The dense series of network links illustrated in this map clearly precludes any novel or intuitive visual insight. Thus, we based all of the spatially-explicit bipartite analyses on explicit 'flow-path' distance measures within the stream network itself (see 'Task C_2' below).

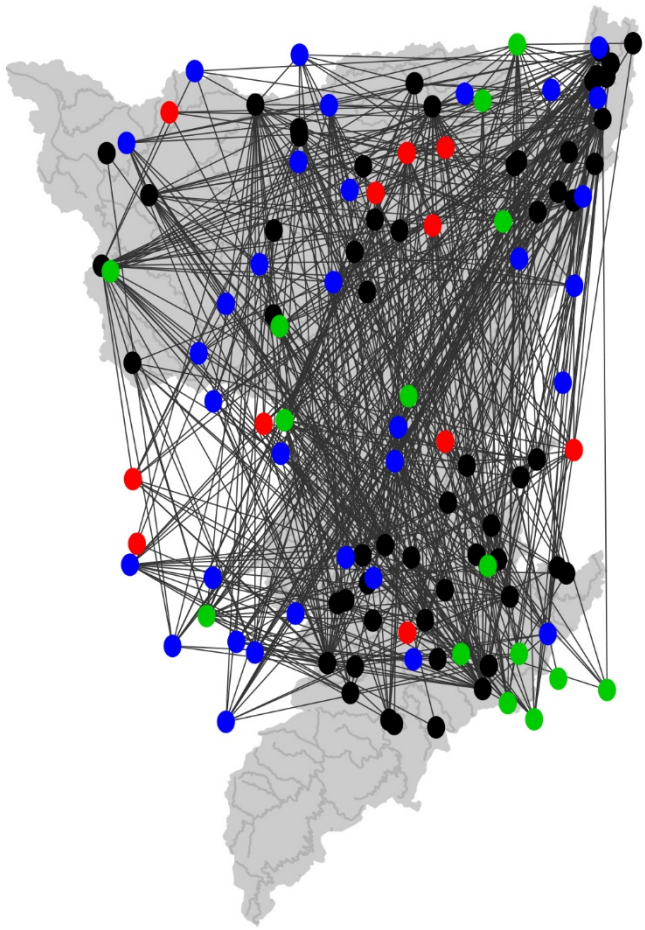


Figure 7. Spatially-explicit map of the bipartite fish network in the Kanawha River Basin (WV). Only significantly positive co-occurrences are shown. All sample sites nodes are rendered as dark grey ovals and are superimposed upon their true geographic locations with the basin's stream network (medium grey arcs/polylines within the light grey basin polygon). Species' nodes are depicted as red, blue, and green ovals, illustrating the three species' modules that were detected within this network. Note that the dense matrix of links/edges within the graph makes it difficult or to search for meaningful spatial patterns.

Task C_2 – Conduct spatially-explicit network analyses (unipartite and bipartite networks)

Individual species within a network have specific, unique characteristics that characterize their network 'positions' or roles. For example, a species' degree value is the number of co-occurring species that share a connection with it. Such species-specific characteristics may be related to the spatial location (i.e., latitude and longitude) of a species' geographic range, relative to the geographic center of the network. Thus, we expected that the degree of a given species within a unipartite network and the distance between the centroid of that species' geographic range and the

center of the physical boundary of the network (e.g., a given HUC unit) should be inversely related. That is, a species whose range is far from the geographic center of the network should have a lower degree value; it should be connected to relatively few other species because its range is either small or 'peripheral' to the geographic center of the network. We tested this hypothesis using Mississippi River Basin data at the HUC2 sub-drainage scale as a case-study. First, we delineated species' overall ranges within the Mississippi River Basin using shapefiles of each species' native range from NatureServe (http://explorer.natureserve.org/). Next, we used the NatureServe range shapefiles to interpolate the 'centroid' (i.e., geographic center) of each species' range, as well as its cumulative range size. Specifically, we calculated the geographic location of the centroid of each species' range as a discrete point in Cartesian space, using *both* the NatureServe ranges (i.e., 2-D areal polygons) and a polyline file of the stream and river network within each polygon, represented by the 1:100,000 scale National Hydrograph Dataset Plus, version 2; McKay et al. 2015). Range size estimates incorporated both the summed areas (km^2) of all polygons within a species' native range and the combined length (km) of all internal stream segment polylines within its range. Two distance measures were then used to quantify geographic distance within the unipartite networks: (1) straight-line Euclidian distance from the polyline range centroid to the polyline HUC2 centroid (i.e., the center of the entire physical river network); and (2) straight-line Euclidian distance from the polygon range centroid to the polygon HUC2 centroid (i.e., the center of the aggregated range polygon; see (e.g., see *Fig. 3B*). Two metrics of range size were also calculated: the total area of the combined range polygon and the total length of all stream segments within the range polygon.

We then tested for spatial patterns within the unipartite network for Mississippi River Basin fishes, using four of the previously described, node-specific (i.e., species-specific) unipartite network metrics: node degree, the local clustering coefficient, betweenness centrality, and closeness centrality (see *Appendix B*). To make closeness centrality scale positively with node-level connectivity (i.e., node degree), as did the local clustering coefficient and betweenness centrality, we re-scaled it as 1 – closeness centrality. Two specific predictions/hypotheses were tested. The first was that each of the four network metrics should *decrease* as the distance between a given species' range centroid and the overall center of the Mississippi Basin increases. This was a test of whether species become less well-connected within the unipartite network as their ranges move further from the geographic center of the entire river basin. Second, we tested whether each of the four network metrics was *positively* related a species' range size, anticipating that wide-ranging species will be more highly connected to other species within the network. For each combination of a species' geographic range estimate and a network metric, we used simple linear regression to test for a significant relationship. In this scheme, 16 separate regression models were tested within each of three HUC2 sub-drainages in the Mississippi Basin. All linear regression analyses were performed using the R base package (R Core Team 2016).

Spatially-explicit analyses of the bipartite network for Mississippi River Basin fishes were based on flow-path distances among sample sites (i.e., species' occurrence locations) superimposed on the stream network. For every sampling site within the Mississippi basin, we calculated the downstream flow-path distance from that site to the outlet of its parent sub-drainage, at both the HUC2 and HUC4 scales. Flow-path distance is equivalent to the one-dimensional length that a parcel of water must travel to move from a given sampling site to the basin outlet. This provided a quantitative and ecologically meaningful way for us to test whether the structure of the bipartite networks varied in a predictable manner with spatial position in the physical, dendritic stream network. Because longitudinal position within the stream network is a fundamental predictor of fish assemblage structure (e.g. Matthews 1998; McGarvey and Hughes 2008), we were confident that flow-path distance would be an ideal metric for tracking spatial effects within the dendritic river network. All flow-path distances were calculated using the National Hydrograph Dataset Plus (v. 2; 1:100,000 scale) streams network (McKay et al. 2015) and standard routing and network functions in ArcGIS version 10.3 software (Environmental Systems Research Institute, Redlands, California). Linear regression analysis was then used to test for significant relationships between flow-path distance and three bipartite network metrics, each of which was calculated for site nodes, rather than species' nodes. Focusing on site nodes, rather than species' nodes was prudent because the individual sampling sites have a true geographic location and can therefore be evaluated relative to another discrete location within the dendritic stream network (i.e., the outlet) without introducing abstract measures of space or location, such as species' centroids. The three network metrics included degree, betweenness centrality, and closeness centrality. Linear regressions were also performed using either longitude or latitude, measured at each sampling site in decimal degrees, as spatial predictor variables; this allowed us to test for general north-to-south and east-to-west patterns within the bipartite networks. All linear regression analyses were performed using the R base package (R Core Team 2016).

Because the numbers of individual sampling sites that were previously assessed as being in good, fair, or poor condition were highly variable, we did not perform separate regression analyses for subsets of the data that were exclusive to a single disturbance level. We did, however, use analysis of covariance (ANCOVA) to test for significant interaction effects between the primary spatial predictor variable (flow-path distance, longitude, or latitude) and disturbance level in each of the significant models. Thus, we tested whether the slopes of the primary relationships between predictor and response differed among the three disturbance levels. Similarly, we used ANCOVA to test for significant interaction effects with HUC identity; these tests assessed whether the slopes of the primary relationships varied among HUC2 or HUC4 scale sub-drainages within the larger Mississippi River Basin. ANCOVA was performed using the R base package (R Core Team 2016).

Research Task D_1: Node Deletion/Extinction Experiments (Bipartite and Unipartite)

Simulated extinction (i.e., node deletion) experiments were performed for bipartite and unipartite networks. Bipartite experiments were conducted in four HUC4-scale river basins: the Susquehanna River Basin (Pennsylvania, HUC-0205), the Savannah River Basin (Georgia, HUC-0306), the Kanawha River Basin (West Virginia and Virginia, HUC-0505), and the Great Miami River Basin (Ohio, HUC-0508). These four river basins were selected because each had a relatively large number of fish samples collected within it and each had a major DoD facility located within it (Susquehanna – Joint Base McGuire-Dix-Lakehurst; Savannah – Fort Gordon; Kanawha – Radford Army Ammunition Plant; Great Miami – Wright-Patterson Air Force Base). Thus, spatial proximity to a major DoD facility was included as a potential stress gradient in the bipartite extinction experiments. Notably, the bipartite experiments were only performed using *complete* networks. We did not perform independent experiments for networks that had previously been partitioned by disturbance categories (see above). This was prudent because our goal was to demonstrate a flexible and adaptable method to anticipate future changes in overall network structure, not to predict if or how future disturbance may differentially impact sites and/or species exposed to varying levels of contemporary disturbance. Furthermore, fish sample sizes in the NRSA/EMAP dataset were generally not large enough to support extinction experiments across multiple levels of contemporary disturbance.

For every sample site within each of the four bipartite networks used in node deletion experiments, we recorded the site-specific degree value (i.e., the number of fish species linked to that site) and calculated the distance to the nearest major DoD facility (see previous paragraph) as a straight-line, 'as the crow flies' Euclidian distance. Site-specific degree values were previously calculated when building the bipartite networks and stored for subsequent use. Distances between sampling sites and DoD facilities were calculated for spatially-explicit points (i.e., sample sites and the centers of each DoD facility) using ArcMap GIS software (Albers equal area projection, NAD83).

Three different 'extinction algorithms' were then used to conduct the bipartite node deletion experiments. First, we deleted sites one at a time from each of the four experimental bipartite networks according to their site-specific degree values, proceeding from the highest degree (i.e., site with the largest number of linked species) to the lowest, until all sites were eliminated and the networks were effectively destroyed. In this way, we sought to simulate worst case scenarios in which sites with the greatest potential to serve as overall network hubs were selectively eliminated. Second, we sequentially deleted sites that were closest to their respective local DoD facility. This algorithm simulated a simple disturbance effect that was generated by the respective DoD facility and projected outward in a concentric manner (i.e., an equivalent effect with increasing distance along all directional vectors). Third, we deleted sites from the bipartite networks in random order. These random deletions provided baselines for comparison with the two 'directed' extinction algorithms. For each of the four experimental networks, we performed 1000 random deletion

experiments (i.e., sequentially deleting all species within the network and repeating the process 1000 times), then averaged the results (see next paragraph). As each of the extinction simulations was performed, we documented species that were severed from all sites of known occurrence and therefore became 'secondary' extinctions (i.e., no remaining habitat for a given species within the respective bipartite network). All bipartite extinction experiments were performed using the '*bipartite*' package of Dormann et al. (2008) in R, with additional custom code written by the authors to facilitate storage and visualization of the experimental results.

Following the approach of Memmott et al. (2004), we plotted the proportion of remaining species (one minus the number of secondary extinctions) as a function of site deletions for each of the four experimental networks. This provided an intuitive assessment of the potential robustness of each network to secondary extinctions of resident fish species. When multiple site deletions are predicted to result in few or no secondary extinctions, the network is considered robust; species have sufficiently broad distributions to survive disturbances that eliminate some of their habitat. This robustness may be indicated by plots of site deletions and remaining species counts with convex shapes (i.e., secondary extinctions remain rare until a large number of sites are lost). Alternatively, a consistent linear plot suggests that secondary extinctions may be highly responsive to progressive site deletions. By this logic, the 'area under the curve' (AUC) of a site deletion vs. remaining species plot can be used as an index of network robustness; it is a direct measure of the proportion of species remaining in the network, with higher values indicating fewer extinctions (Memmott et al. 2004).

Because the unipartite networks were spatially implicit (though spatial information can be included as a secondary node attribute; see 'Task C_2' above for an example), we did not incorporate proximity to a DoD facility as a factor in unipartite extinction experiments. Instead, we appended ordinal pollution sensitivity values to each species within a given network and used the sensitivity values to determine the order of extinction events, progressing from the most to least sensitive species. Pollution sensitivity values ranged from 1 (most sensitive) to 7 (least sensitive) and were obtained for individual species from published sources (Barbour et al. 1999; Halliwell et al. 1999; Jester et al. 1992; Lyons 1992). When multiple sensitivity values were reported for a given species, we assigned it the median value, rounding to the lower (more sensitive) value for ties. This 'extinction-by-sensitivity' logic provided the first of three extinction algorithms used in the unipartite experiments. The second extinction algorithm was 'extinction-by-degree', progressing from the most highly connected species (i.e., node with the highest degree) to the least connected. The third algorithm was random extinctions. Species' degree values were previously calculated and recorded while building the unipartite networks (see above). All unipartite calculations were performed with the '*igraph*' package (Csárdi and Nepusz 2006) in R, with additional code written by the authors to generate random extinction sequences and to visualize the results.

Results and Discussion

Network Connectivity (Task B_3)

Unipartite Networks

Across all biogeographic regions and data combinations, we found little evidence for **power-law degree distributions** when complete unipartite networks (i.e., all nodes and links included) were analyzed. *P*-values from the **K-S tests of power-law model fits** were consistently < 0.01, indicating that the cumulative degree distributions did not conform to power-law curves (see *Appendix C.4*); thus, the power-law model α values (see *Appendix C.3*) were irrelevant. This lack of power-law curvature was also evident in plots of the cumulative degree distributions. Cumulative distribution plots were generally linear for complete ('all links') datasets (e.g., *Fig. 8A*) and for networks that included only significantly positive links (e.g., *Fig. 8B*; see also first and second rows in *Appendix A* graphs). Notably, the degree distributions for unipartite networks that included only significantly negative links often assumed more characteristic power-law shapes (e.g., *Fig. 8C*; see also third row in *Appendix A* graphs). But in most cases (with several notable exceptions in Mississippi Basin and Pacific Northwest streams), the K-S tests confirmed that these distributions were not true power-law curves (see *Appendix C.4*).

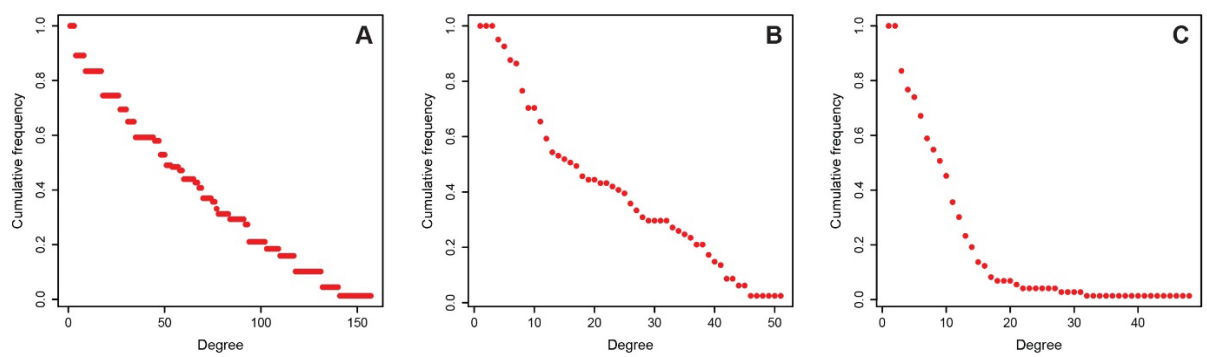


Figure 8. Examples of cumulative degree distributions from fish co-occurrence networks. Data are shown for the Chesapeake Bay rivers (HUC2-02) network when all species and links are included (panel **A**), when only significantly positive links are included (panel **B**), and when only significantly negative links are included (panel **C**).

Power-law curvature was, however, detected in the right tails of many of the cumulative degree distributions when **truncated models** were fit to the degree data. Truncated power-law tests use a 'moving window' algorithm to search for power-law curvature within isolated regions of the

complete degree distribution and they often find locally power-law behavior when only nodes with relatively high degree values are included (i.e., degree values larger than the empirically determine 'X_{min}' value; see *Appendix B*). When the X_{min} values in truncated model-fit tests were unrestricted (i.e., they were free to assume any value up to the maximum observed degree value; see *Appendix C.5*), we detected significant (K-S *p*-value > 0.05), localized power-law curvature in most of the unipartite networks (see *Appendix C.7*). Interestingly, the X_{min} values themselves co-varied with disturbance. Specifically, X_{min} tended to *decrease* with increasing disturbance. This suggests that species with relatively low degree values may be the first to be removed as disturbance increases; this tendency would tend to push the 'knee' of the power-law curve down and to the left, effectively moving X_{min} further left.

In each of the three biogeographic regions, **maximum degree** values within the unipartite networks tended to decrease with increasing disturbance levels (e.g., *Fig. 9*; see also *Appendix C.1*). This trend was strongest for HUC2 scale networks but was also observed for HUC4 scale networks, particularly in East Coast streams. Similarly, the **median degree** (50th percentile) values exhibited a strong inverse trend with disturbance level (see *Appendix C.2*). This suggested that disturbance may, in fact, have a predictable, negative effect on the numbers of species that co-occur in unipartite networks.

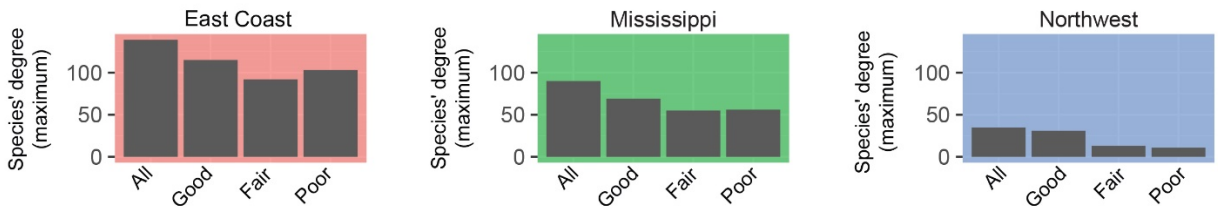


Figure 9. Examples of *maximum species' degree* values from complete (i.e., all species and links included) *unipartite* fish co-occurrence networks in East Coast, Mississippi River Basin, and Pacific Northwest streams. Results from HUC2 scale networks are shown with comparisons among networks built from all samples and partitioned by disturbance level (good vs. fair vs. poor).

Maximum betweenness did appear correlated with disturbance in many of the data combinations, though the direction (i.e., positive vs. negative) of this correlation was variable (see *Appendix C.8*). **Median betweenness** exhibited a negative correlation with increasing disturbance at the HUC2 scale in Mississippi streams when only significantly positive links were included in the network, but not for other data combinations (see *Appendix C.9*). **Closeness** appeared to be more regularly associated with disturbance. For both **maximum** and **median closeness** values and across most of the data combinations, closeness tended to increase with disturbance (e.g., *Fig. 10*; see also *Appendix C.10 and C.11*). **Global clustering coefficients** increased with disturbance for some data combinations while decreasing with disturbance in others and the overall trends were mostly

nominal (see *Appendix C.12*). **Local clustering coefficients** were almost entirely uncorrelated with disturbance (see *Appendix C.13 and C.14*).

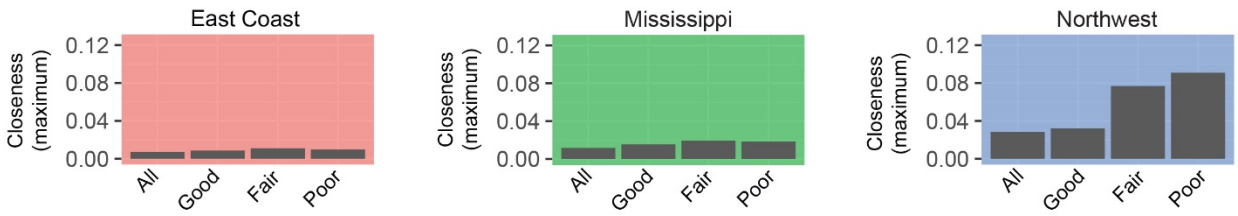


Figure 10. Examples of *maximum closeness centrality* values from complete (i.e., all species and links included) *unipartite* fish co-occurrence networks in East Coast, Mississippi River Basin, and Pacific Northwest streams. Results from HUC2 scale networks are shown with comparisons among networks built from all samples and partitioned by disturbance level (good vs. fair vs. poor).

Bipartite Networks

Indicators of node connectivity, including betweenness, closeness, and clustering coefficients were also calculated for sampling sites within the bipartite networks (see 'Building Bipartite Graphs' above). Of these, maximum betweenness and the maximum local clustering coefficient demonstrated consistent correlations with disturbance. **Maximum betweenness** *increased* at a modest yet consistent rate with disturbance in East Coast as well as Mississippi River Basin streams (see *Appendix D.1*). **Maximum local clustering coefficients** tended to *decrease* with disturbance in East Coast and Mississippi Basin streams (see *Appendix D.6*). All other bipartite metrics appeared unresponsive to disturbance (see *Appendix D*). For instance, **median closeness** was nearly constant among networks built from good, fair, and poor condition sites (*Fig. 11*).

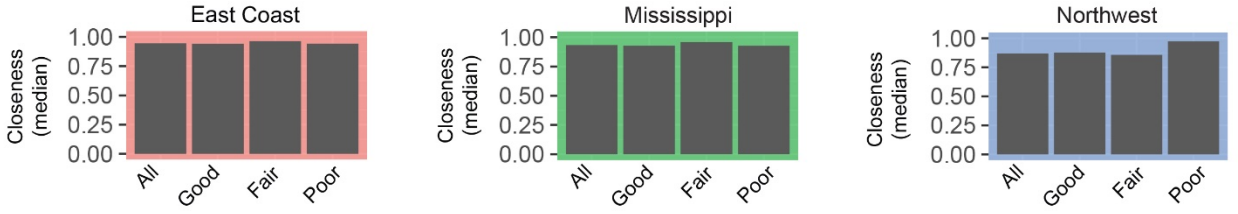


Figure 11. Examples of *median closeness centrality* values from complete (i.e., all species and links included) *bipartite* fish co-occurrence networks in East Coast, Mississippi River Basin, and Pacific Northwest streams. Results from HUC2 scale networks are shown with comparisons among networks built from all samples and partitioned by disturbance level (good vs. fair vs. poor).

Network Modularity (Task B_4)

Potential trends in **unipartite modularity** were observed among several of the data combinations (see *Appendix C.15*). For example, modularity exhibited a uniform *increase* with disturbance level at the HUC2 scale in East Coast streams when all significant links and when only significantly negative links were included (*Fig. 12A*). A similar, though less conspicuous trend (i.e., increasing modularity with increasing disturbance) was also observed at the HUC2 scale in Mississippi River Basin streams when complete datasets were included in the networks (*Fig. 12B*). However, no trend was observed in Mississippi streams when the networks were limited to significant links. In Pacific Northwest streams, we observed a uniform *decrease* in modularity with increasing disturbance at the HUC2 scale (*Fig. 12C*). In general, modularity was higher for HUC2 scale networks than for HUC4 scale networks. Modularity was also generally higher in networks that were limited to significant links ('1-SigAll', '2-SigPos', and '3-SigNeg' in *Appendix C.15*) than in networks that included all links ('0-Complete'). However, the **number of distinct modules** within networks did not vary in a predictable manner with any of the factors considered here (see *Appendix C.16*). The number of modules within a unipartite network ranged from 1−6 and these modules exhibited clear clustering patterns when identified in the unipartite graphs (see examples in *Appendix A*).

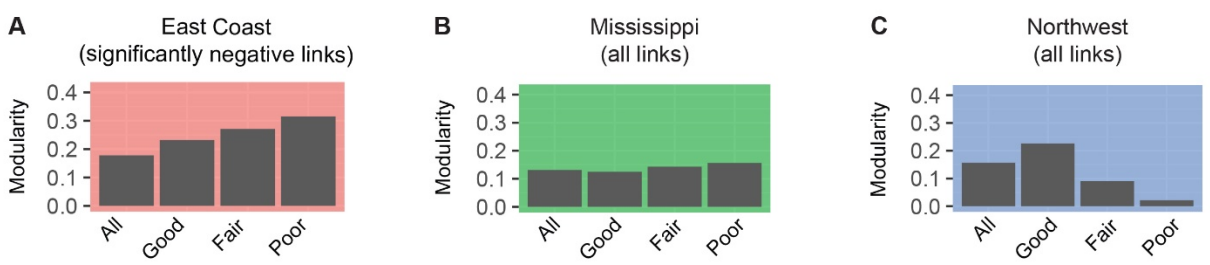


Figure 12. Examples of *modularity* values from *unipartite* fish co-occurrence networks in East Coast (**A**), Mississippi River Basin (**B**), and Pacific Northwest (**C**) streams. Results from HUC2 scale networks are shown with comparisons among networks built from all samples and partitioned by disturbance level (good vs. fair vs. poor).

Modularity within **bipartite networks** was not responsive to any of the factors included in this study (see *Appendix D.8*). Across all data combinations, site-level modularity remained nearly constant, ranging from ~0.2−0.3. Similarly, the number of distinct modules detected within bipartite networks was largely constant, ranging from 4−6 and exhibiting no consistent trend with any of the data factors (see *Appendix D.9*).

Functional Structure of Network Modules

Here we present detailed functional traits analysis results from the Susquehanna River Basin. The Susquehanna is emphasized prominently in our results because it was among the most data-rich regions in our Limited Scope project, with sufficient sample replicates to perform all network calculations independently for each of the three disturbance levels. Retrospective comparisons also showed that the Susquehanna results were often broadly representative of the other networks. Thus, the Susquehanna was a useful 'exemplar' system for examining functional structure within and among the network modules. (The Susquehanna was also used in a detailed study of extinction effects within unipartite networks; see 'Node Deletion/Extinction Experiments' below.)

Tests of congruence between network module classifications and functional traits *k*-means cluster analysis classifications (see 'Task B_4' above) in Susquehanna River Basin streams showed that agreement between the two classification systems decreased as the number of distinct modules increased, with highest congruence occurring when only two network modules were detected. However, these congruence values were rarely significant. For 5 of 6 networks that included only two modules, the permutation test *p*-value was > 0.1. (This reflected the fact that classification congruence values in the ~0.5 range are no different than random expectation in a series of 50:50 Bernoulli trials.) In only one instance was the congruence value for a network with two modules significant: the observed congruence value for 'complete network - poor condition sites' (0.67) was significant at $p = 0.01$ (lower-left panel in *Fig. 13*). When three or more modules were detected in a network, the overall congruence levels were generally lower. The 'significantly positive links - moderate condition sites' network, which included three modules and had a congruence value of 0.56, was the sole exception (3rd row, center column panel in *Fig. 13*). Nevertheless, the relatively low congruence values for networks with > 2 modules were all significant ($p \leq 0.01$; see *Fig. 13*), reflecting the exponential increase in possible outcomes that occurred when a third and/or fourth category was added to the classification system.

Functional hypervolume plots are shown for all (HUC4 scale) network data combinations in the Susquehanna River Basin in *Fig. 13*. In each plot, modules are identified by color, with large circles showing the centroid of each module. The small, dark points represent individual species within each module, as indicated by matching colors (i.e., red vs. blue vs. green). Light/faded points are random 'pseudo-samples' drawn from the probability distributions generated by the *hypervolume* program in R (i.e., the kernel density estimates; see 'Task B_4' above). Solid lines delineate confidence regions within functional traits space for each module.

In most hypervolume plots, the unipartite modules exhibited extensive overlap within functional traits space. This was particularly true for networks that consisted of only two distinct modules, including all of the complete networks (first column in *Fig. 13*) and all of the networks for poor condition sites (bottom row in *Fig. 13*). Modest separation between the two modules was, however,

noted along the first PCoA axis for networks that were restricted to poor condition sites and included only significantly positive or significantly negative links (last two plots in bottom row of *Fig. 13*). In these two networks, the modules were separated along a general gradient of size and age, with species' total length, longevity, and fecundity increasing from left-to-right (along the *x* axis/first PCoA axis). A moderate tendency for one of the three modules (i.e., the 'red' module) in the 'significantly positive links - moderate condition sites' network was also observed along the first PCoA axis (with same factor loading as above for functional traits). And in three of the networks that featured four independent modules ('significantly positive links - all sites', see red module; 'significantly negative links-all sites', see red module; and 'significantly negative links-moderately disturbed sites', see green module), we noted that one module occupied a conspicuously small volume of the total functional traits space (see red module in 'significantly positive links - all sites' and 'significantly negative links - all sites' plots; see green module in 'significantly negative links - moderately disturbed sites' plot; *Fig. 13*).

One plausible interpretation of the 'small volume' modules is that they may reflect subsets of fishes with specialized ecological roles or life-history and habitat requirements. However, the fact that the small volume modules occurred near the centers of their respective hypervolume plots may also suggest an alternative explanation: these modules may consist of species that are adapted to highly stable environments, where a consistent 'middle of the road' functional strategy maximizes fitness. Overall, we emphasize that little apparent separation was detected among network modules when plotted as hypervolumes in functional traits space (*Fig. 13*). And we suggest this surprising lack of variation may, in itself, be a novel result. If the functional traits profiles of distinct network modules exhibit *less* differentiation than expected under an appropriate null model (and we are not currently aware of a precedent or method to build that null model), it may **suggest that fish co-occurrence networks are comprised of repetitive functional 'motifs'**.

In light of these mixed results, we believe it is unlikely that functional traits act as primary determinants of the observed modular structure within the unipartite fish networks. At least two other related lines of inquiry should, however, be pursed in future research. First, *a priori* hypotheses that link specific traits or small subsets of traits to environmental conditions in ways that may confer selective advantages to individual species should be formulated and tested. We used the multivariate hypervolume method (Blonder et al. 2014) to maximize the information in our original functional traits matrix, but this indiscriminate approach assumes that all individual traits are equally important to biotic assemblages and may therefore be counterproductive, masking key functional trends, if only one or several of the functional traits are truly deterministic of the modular structure within a network. Second, a framework/method should be devised to test the idea that functional traits profiles of distinct network modules may, in fact, be so similar that they represent functionally equivalent motifs of species' characteristics.

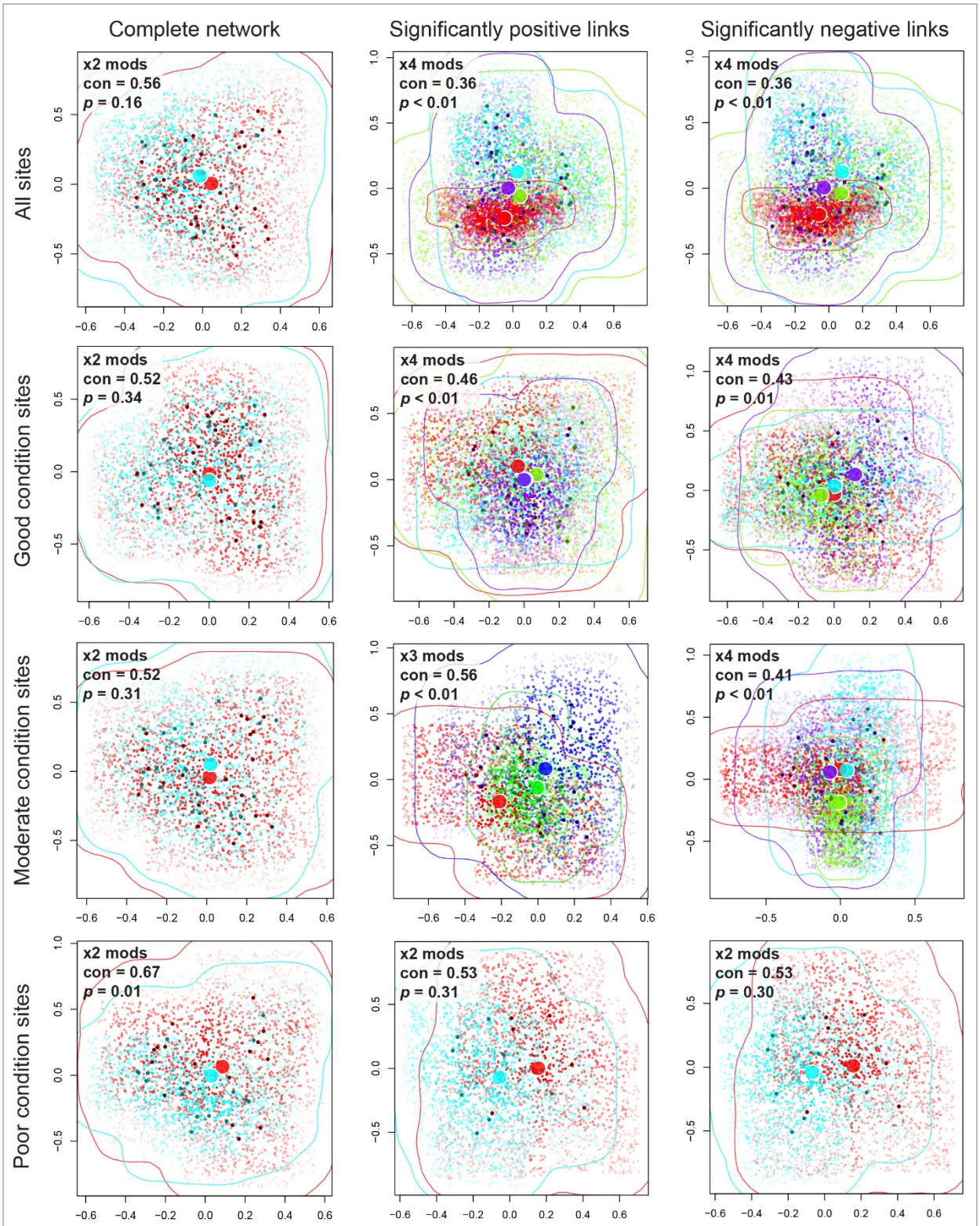


Figure 13. Functional hypervolume plots for fish co-occurrence networks in the Susquehanna River Basin. (*See next page for complete caption.*)

Figure 13 (*continued*). Functional hypervolume plots for fish co-occurrence networks in the Susquehanna River Basin. The three columns represent complete networks, networks consisting only of significantly positive links, and networks including only significantly negative links, respectively. The four rows represent networks that are inclusive of all fish samples and of sites that were previously ranked by disturbance level (i.e., good, moderate, or poor condition). In each plot, modules are identified by color, with large circles showing the centroid of each module. The corresponding (i.e., same color) small, dark points represent the individual species within each module. The light/faded points are random 'pseudo-'samples drawn from the probability distributions generated by the 'hypervolume' program (i.e., the kernel density estimates; see previous paragraph). And the solid lines/polygons are the confidence regions for each module, constructed from the kernel density pseudo-samples. Numbers of network clusters detected via simulated annealing ('× mods' values) and results from the tests of congruence between network module classifications and functional traits *k*-means cluster analysis classifications (congruence values and permutation test *p*-values) are also shown at upper-left in each plot.

Node Deletion/Extinction Experiments

Bipartite Networks

Plots of the remaining numbers of species, relative to the number of simulated site deletions, are shown for four river basins: the Susquehanna River, the Savannah River, the Kanawha River, and the Great Miami River (*Fig. 14*). A detailed visual record of the effects of progressive site deletions on the Susquehanna bipartite network is shown in *Fig. 15*. Site deletion vs. species remaining plots for the Susquehanna Basin network revealed minimal differences between the three extinction algorithms. Curves for extinction-by-site degree (red 'degree' plots), extinction-by-distance from the nearest major DoD facility (blue 'distance plots), and random extinction (green 'random' plots) simulations were similarly convex, with all AUC values close to ~0.8 (*Fig. 14*). This suggested that critical network hubs (i.e., stream sites) may be rare or nonexistent in the Susquehanna Basin. However, the extinction-by-degree plots indicated that network robustness may be lower in the Savannah, Kanawha, and Great Miami networks; each of the respective site deletion vs. species remaining plots exhibited a nearly linear effect of progressive site deletions on secondary species extinctions (i.e., species that become entirely isolated from other species via site deletions). Vulnerability to secondary extinctions appeared particularly high in the Kanawha Basin, where the loss of the six highest degree sites could potentially eliminate > 20% of all fish species within the network. This strong potential effect of site degree on secondary extinctions is intuitively consistent with conventional wisdom regarding habitat conservation: one should prioritize the protection of habitats that host the greatest numbers of species. Interestingly, no major differences were observed between plots of extinction-by-distance from DoD facilities and random extinctions (*Fig. 14*). Across all four bipartite networks, the 'distance' and 'random' plots were similar in shape with differences in the respective AUC values ranging from 0.01−0.04.

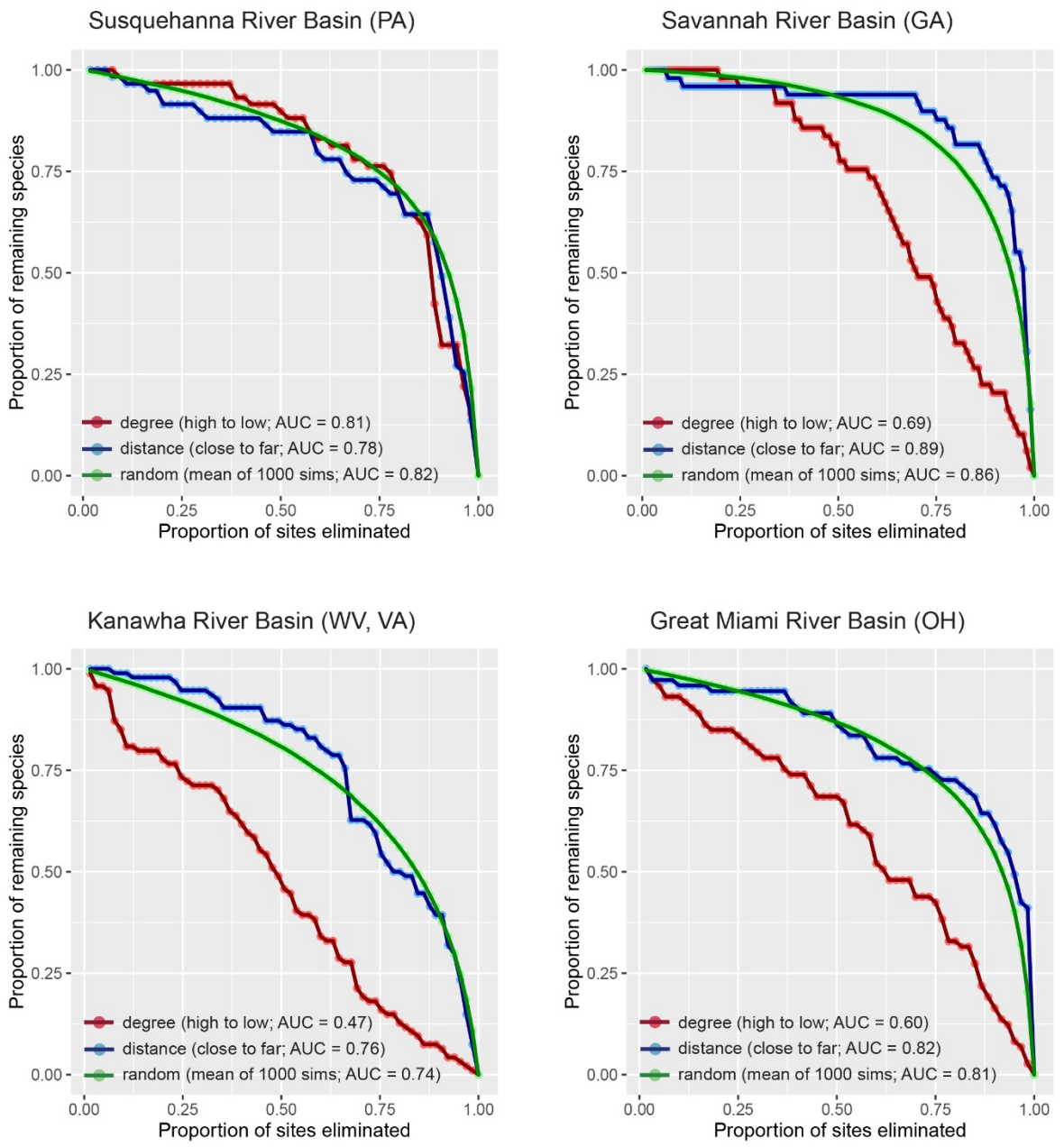


Figure 14. Plots of secondary extinctions in the Susquehanna, Savannah, Kanawha, and Great Miami Basin bipartite networks. Simulation results are shown for three extinction algorithms: extinction as a function of site connectivity (highest to lowest site-level degree; red symbols), as a function of Euclidian distance to the nearest major DoD facility (shortest to farthest distance; blue symbols), and random site deletions (average of 1000 simulations; green symbols). In each network, greater 'robustness' to secondary extinctions is indicated by plots with strongly convex shapes; these curves suggest that secondary fish species extinctions may be slow to occur as sites are removed from the network. Plots with nearly linear shapes indicate rapid loss of species with site removals. Thus, AUC (shown in parentheses for each plot) can be used as a simple index of network robustness, with higher AUC values indicating many species that are broadly distributed and therefore less vulnerable to extinction via acute habitat loss.

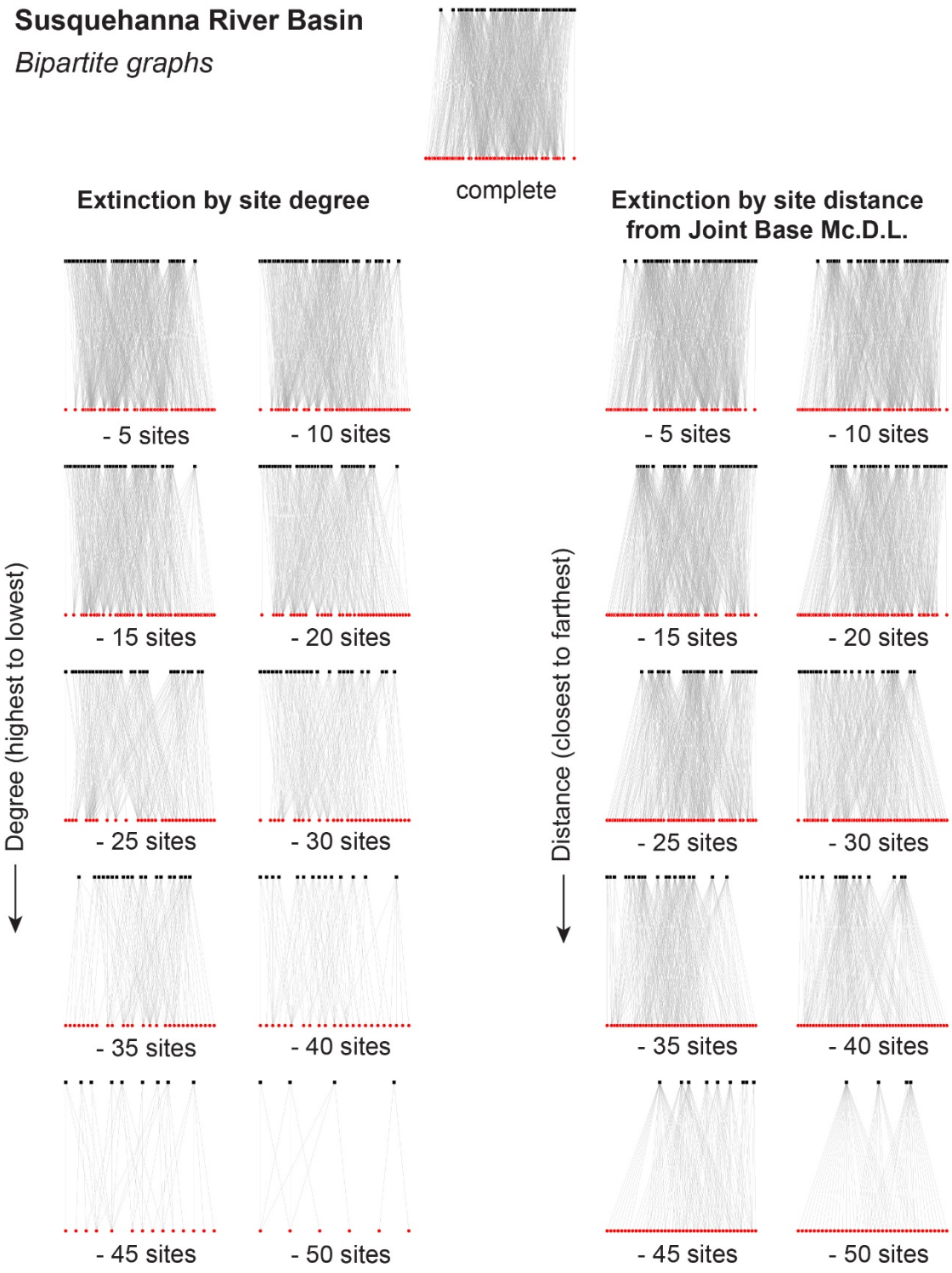


Figure 15. Bipartite graphs illustrating sequential, simulated deletions of discrete stream sites within the Susquehanna River Basin. (*See next page for complete caption*.)

Figure 15 (*continued*). Bipartite graphs illustrating sequential, simulated deletions of discrete stream sites (i.e., nodes) within the Susquehanna River basin. The original, complete fish network graph is shown at top center with sampling sites represented by black squares (54 sites) and species represented by red circles (59 species). Species' occurrences at each site are represented by grey lines. Numbers of simulated site deletions are shown below each graph (e.g., '-5 sites'), ranging from 5 to 50 site deletions. Deletions that simulated the progressive loss of the most highly connected sites (i.e., highest to lowest site-level degree) are shown in the two left columns. Deletions that simulated site losses as a function of proximity to Joint Base McGuire-Dix-Lakehurst (i.e., closest to furthest site locations) are shown in the two right columns.

One important caveat of the bipartite extinction results is the fact that secondary species' extinctions did not occur until all of the sites that a given species was linked to (i.e., that it occurred at) were removed from the network. This liberal assumption may be realistic in other types of bipartite networks where species' movements are not highly constrained, such as plant-pollinator networks that include many species of flying insects. However, fishes that are constrained to the wetted channels of dendritic river networks will necessarily have lesser abilities to 'search' for alternative nodes (i.e., sites) throughout the complete, remaining network. It is conceivable that the loss of upstream and/or downstream sites may isolate a given site from the rest of the physical stream network and if this discontinuity is maintained, it may prevent locally occurring species from reaching other sites within the network. In effect, deleting a given site from a bipartite fish network may have secondary effects on the remaining links (i.e., site-to-species connections) within the network, as well as the species losses depicted in *Fig. 14*. Thus, secondary extinctions may occur *more rapidly* than predicted in *Fig. 14*. Developing a method to simultaneously predict the effects of site deletions within traditional bipartite networks (as presented here) and within spatially-explicit dendritic river networks, to produce more realistic, conservative predictions of secondary fish species extinction rates, could therefore be an important avenue for future research.

Unipartite Networks

Here we present exemplar results from the Susquehanna River Basin. For each of the three extinction algorithms (by sensitivity, by degree, and random), six network metrics were calculated at each step during the extinction simulations (i.e., after each node/species deletion): the global clustering coefficient, median average local clustering coefficient, maximum and median species' degree within the remaining network, and maximum and median species' betweenness within the remaining network. Extinction effects on the **global clustering coefficient**, with larger values indicating dense, highly interconnected networks (see *Appendix B*), were minimal and generally similar among the three extinction algorithms (i.e., nearly parallel plots for 'sensitivity', 'degree', and 'random' data) until a threshold of ~22 species extinctions was reached (*Fig. 16A*). Beyond 22 extinctions (i.e., moving further right along the *x*-axis in *Fig. 16A*), the global clustering

coefficient began a distinct downward trend for the *extinction-by-degree* results, ultimately collapsing to a value of zero at 35 extinctions. Global clustering coefficients for the *extinction-by-sensitivity* and *random extinction* results became more erratic at ~20−22 species extinctions, but did not exhibit a clear downward trend until ~35−40 species extinctions had occurred. Notably, the central tendencies of the extinction-by-sensitivity and random extinction results did not deviate from each other; while the random points exhibited wider overall spread, the sensitivity points consistently remained near the center of the random results.

Results from the random extinction simulations were generally similar for the **median local clustering coefficients** (i.e., the number of 'neighbors' of a focal species that are themselves connected, averaged across all species within the network; see *Appendix B*). *Random* points exhibited increasing spread with an increasing number of species extinctions, ultimately displaying a rapid downward trend beyond ~40 extinctions (*Fig. 16B*). The *extinction-by-sensitivity* and *extinction-by-degree* results were, however, more dynamic. For example, extinction-by-degree results displayed a strong, downward-curving trend through ~23 species' extinctions, then rapidly increased and began a second downward trend until reaching a value of zero (i.e., none of the species connected to a given species were themselves directly connected) at 35 extinctions. Furthermore, a rapidly declining 'spike' was observed in the extinction-by-degree results at ~9 extinctions before returning to the broader trend at ~11 extinctions (*Fig. 16B*). This spike was particularly interesting because it marked the point at which secondary extinctions (i.e., species that had not yet been explicitly deleted in the simulations, but were functionally isolated with no remaining links to other species) began to appear in the unipartite networks. (This phenomenon is illustrated in the unipartite graphs shown in *Fig. 17* below; secondary extinctions, indicated by heavy black circles around species' nodes, first appear in the '-9 species' graph.) Median local clustering coefficients for the extinction-by-sensitivity results also reached a threshold at ~8−9 extinctions, but shifted in the opposite direction; they *increased* to a nearly constant value before dissociating to random behavior at ~20 extinctions, then increasing to constant, perfect clustering (i.e., median local clustering coefficients of unity) at 40 extinctions (*Fig. 16B*). This seemingly complex behavior in the extinction-by-sensitivity results may be due to the punctuated elimination of subsets of relatively rare, peripheral species with low pollution tolerances. These peripheral subsets will tend to decrease the average clustering coefficient if the constituent species are not themselves strongly interconnected. Removing them may therefore increase the average clustering coefficient. By the same logic, the perfect clustering (i.e., local clustering coefficient of unity) observed beyond 40 extinctions suggests a subgroup of highly tolerant species with a high incidence of co-occurrence.

Plots of **maximum** and **median species' degree** showed that both values are predictable, linear functions of the number of species removed from the network. These plots also confirmed that selective extinctions of the most highly connected species will tend to degrade the overall structure of the network more rapidly than selective extinctions of the most sensitive species, as well as

random extinctions. *Extinction-by-degree* plots were clearly steeper than *extinction-by-sensitivity* and *random extinction* plots, for both maximum (*Fig. 16C*) and median (*Fig. 16D*) degree values. Interestingly, the extinction-by-sensitivity results for maximum degree values were identical to the random extinction results, displaying a perfect, negative linear relationship. This suggested that progressive losses of the most sensitive species are unlikely to have secondary effects on the most highly connected species within a network.

Plots of **maximum** and **median betweenness** (i.e., the tendency for many network connections to pass through a given node; see *Appendix B*) also indicated strong differences among the *extinction-by-degree* algorithm and the *extinction-by-sensitivity* and *random extinction* algorithms. Moreover, no similarity was observed in overall trends between the maximum (*Fig. 16E*) and median (*Fig. 16F*) results. Maximum betweenness decreased with the number of extinctions in a variable, yet mostly linear manner for extinction-by-sensitivity and random extinction data, but displayed conspicuous 'wave-like' spikes for extinction-by-degree data. Tellingly, these spikes were congruent with gaps in the extinction-by-degree, maximum degree plot (compare with *Fig. 16C*) that represented short lists of highly connected species with the same degree values (i.e., eight species each with 55 connections, followed by eight species each with 48 connections, followed by six species each with 45 connections, etc.). The maximum betweenness waves spiked when only one species remained at a given maximum degree level; this result can be visually confirmed by cross-referencing the maximum betweenness and maximum degree values at the 8th, 16th, 22nd, 29th, and 33rd extinction events. This pattern may indicate that multiple, highly connected hub species occur within the network, but are themselves located in different regions of the network. In this way, maximum betweenness may increase until a single, most highly connected species remains, then rapidly decrease when that species is lost and the network resets to a new, lower maximum degree value that is exhibited by several new hub species. The extinction-by-degree plot for median betweenness indicated that at the complete network level, the median average betweenness value may track a single 'boom-and-bust wave' then settle to a nearly constant value, located between the 10th and 12th extinctions (*Fig. 16F*). Random extinction results indicated that median betweenness may be a decreasing, albeit highly variable function of the number of extinction events. In general, this negative trend in median betweenness was also observed in the extinction-by-sensitivity results. However, a clear sequence of relatively low, constant values was noted between the 8th and 20th extinction events. This puzzling gap in the overall extinction-by-sensitivity pattern may reflect a group of moderately sensitive species with uniquely low connectivity. Following the extinctions of the seven most sensitive species, median betweenness may be rapidly depressed by a large number of species (12 species in this particular instance) that co-occur with only one or several species; upon removal of these poorly connected, moderately sensitive species, median betweenness within the remaining network may rise again.

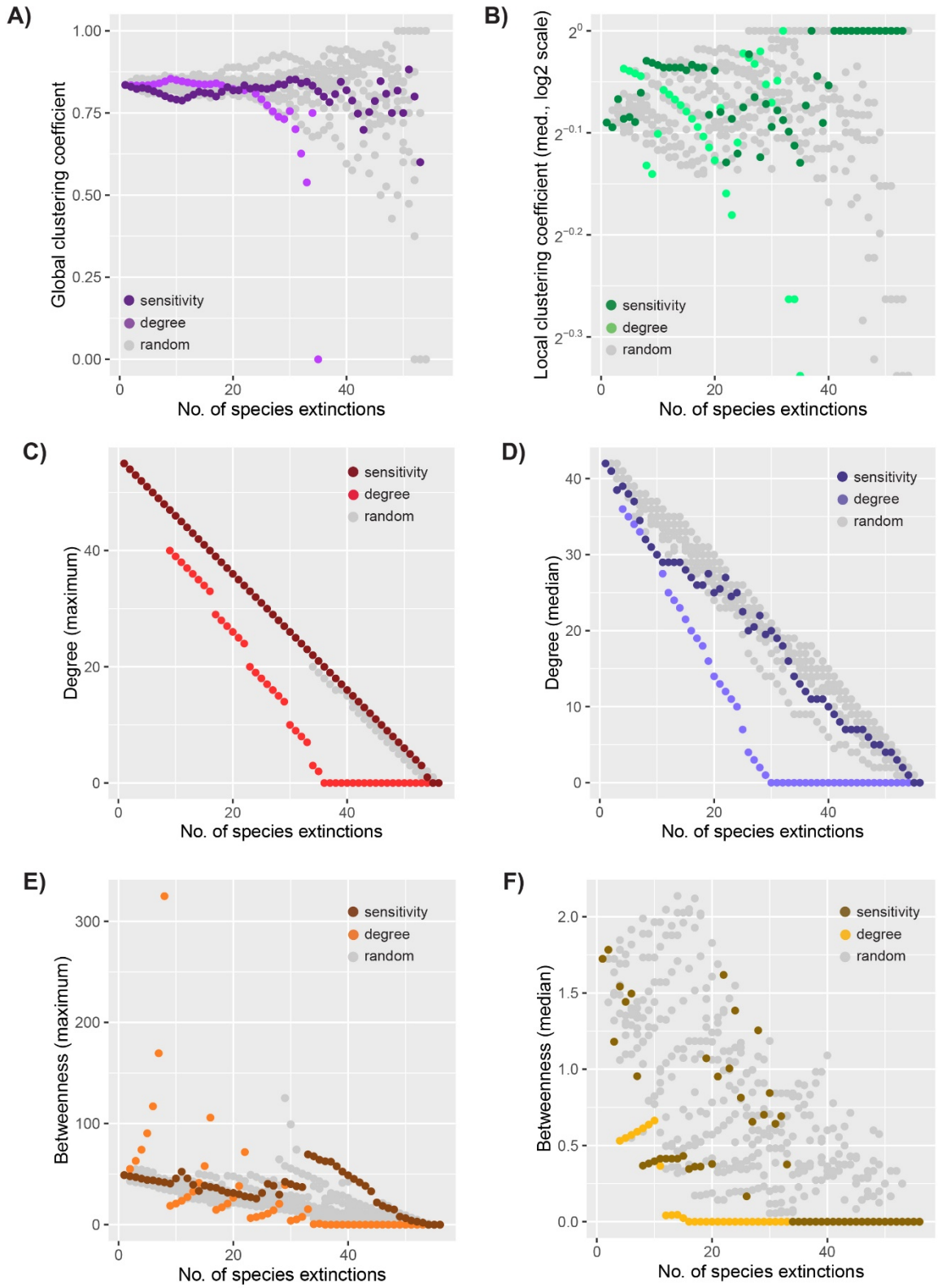


Figure 16. Plots of species' extinction effects on the unipartite fish network of the Susquehanna River basin. (*See next page for complete caption.*)

Figure 16 (*continued*). Plots of species' extinction effects on the unipartite fish network of the Susquehanna River basin. Scatterplots track the effects of sequential (i.e., one at a time) extinctions on the global clustering coefficient (**A**), median average local clustering coefficient (**B**), maximum (**C**) and median (**D**) species' degree, and maximum (**E**) and median (**F**) species' betweenness. Three node deletion algorithms are included in each plot: extinction due to species' sensitivity to pollution (i.e. most sensitive species removed first), as a function of species' connectivity/degree (i.e., species with the highest degree removed first), and random species extinctions (results from 10 random simulations are shown).

One other notable observation that emerged from the unipartite extinction experiments was the tendency for cumulative extinctions to have differential effects on modules within the network. As larger numbers of species were removed and secondary extinctions increased, these extinctions were 'confined' to one of the two initial network modules (red nodes enclosed by heavy black circles in *Fig. 17*). No species from the second module (green nodes in *Fig. 17*) became vulnerable to secondary extinctions (i.e., the first instances of green secondary extinctions occurred after 30 species were removed from the network) until *all* species from the first module had been removed. This suggests that, based solely on the topology of the unipartite network, it may be possible to identify the most and least at-risk species. For instance, drawing upon the extinction experiment illustrated in *Fig. 17*, we predict that the most vulnerable fishes in Susquehanna River Basin streams may include the Comely Shiner (*Notropis amoenus*), Swallowtail Shiner (*Notropis procne*), Fantail Darter (*Etheostoma flabellare*), and Redfin Pickerel (*Esox americanus americanus*). Least vulnerable species may include the Johnny Darter (*Etheostoma nigrum*), Brown Bullhead (*Ameiurus nebulosus*), and Slimy Sculpin (*Cottus cognatus*). (See 'Conclusions and Implications for Future Research' below for further explanation.)

Spatially-Explicit Analyses

Unipartite Networks (Mississippi River Basin)

Within each HUC2 sub-drainage of the Mississippi River Basin, we observed substantial variation among species in each of the four network metrics (*Table 1*). Variation was also observed in the geographic orientations of species' ranges and range sizes (*Table 1*). However, variation in the four network metrics was generally unrelated to the geography or spatial orientation of species' ranges. r^2 values from most regressions were < 0.1; only those combinations shown in *Table 2* exceed $r^2 = 0.1$. Linear regressions that simultaneously incorporated both range distance and range size as a single predictor variable were also explored by multiplying the area of species' range polygons by the inverse of polygon-based distance. This composite variable approach led to several additional, significant regression models (see *Table 2*). The majority (11 of 17) of all significant relationships were observed in HUC-0210 (the Missouri River sub-drainage) and of these, five included the local clustering coefficient as the significant network metric.

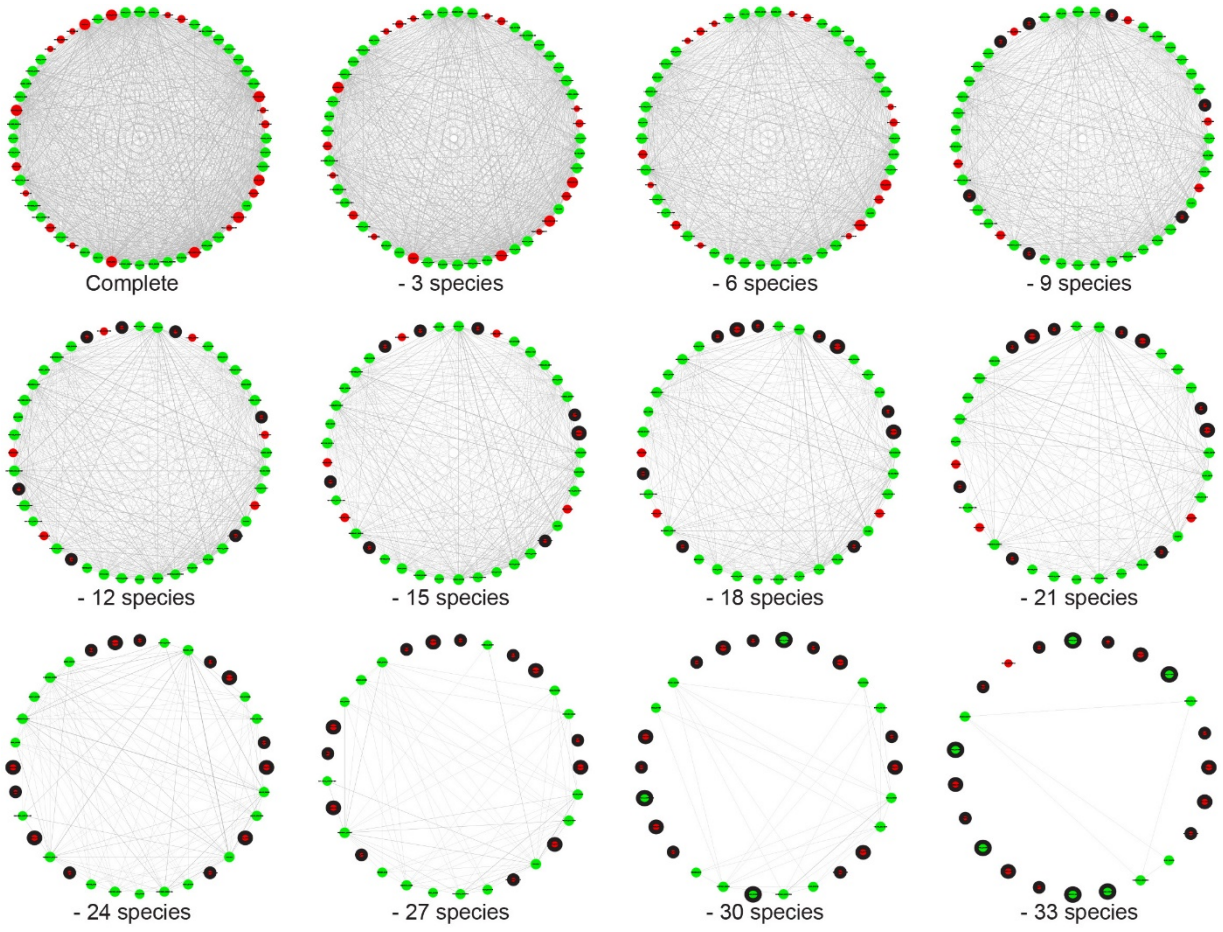


Figure 17. Unipartite graphs ('ring' layout) from the extinction-by-degree simulations. Results are shown with node colors (red and green) indicating species' membership within the two modules that were originally detected in the complete unipartite network; initial module membership was maintained for each of the remaining species as additional extinctions were simulated. Unipartite graphs are shown for the complete network (56 species) and at 'three species extinction' intervals (e.g., '- 3 species'). When the number of simulated extinctions exceeds 34, the network collapses completely with no remaining links among species.

We expected that network metrics would be negatively associated with range distance variables (i.e., *polyline* distance, or 'PLD', and *polygon* distance, or 'PGD', in *Table 1*) and positively associated with range size variables (i.e., polyline length, or 'PLL', and polygon area, or 'PGA', in *Table 1*). However, these expectations were met in only nine (~ 50%) of the significant outcomes (see *Table 2*). Thus, we were not able to determine with a high level of certainty whether species with distant and/or small ranges tend to be less well-connected in unipartite co-occurrence networks. That said, we did note that all seven of the significant regression relationships conformed to the expected direction when node degree was used as the network metric (i.e.,

dependent variable) in linear regressions. Node degree increased significantly with increasing range size (PGA and PLL in *Table 1*) and with decreasing range distance (PLD and PGD in *Table 1*). Degree also increased significantly as the composite variable increased (PGA × 1/PGD in *Table 2*). This is notable because species' degree is the most intuitive and computationally simple indicator of network connectivity that we examined and it is a truly objective measure of the connectivity that a focal species shares with its immediate neighbors in a unipartite network. The other three metrics of network connectivity that we examined– the local clustering coefficient, betweenness centrality, and closeness centrality – were more abstract than node degree and were generally not found to be affected by the spatial orientation of a species' range or its range size.

Table 1. Mean values (+/- 1 standard deviation) of species' network metrics and range characteristics for stream fish species occurring in HUC2 sub-drainages of the Mississippi River Basin.

Network Characteristics†				Range Characteristics‡			
	HUC-0207	HUC-0210	HUC-0211		HUC-0207	HUC-0210	HUC-0211
Degree	11.9 ± 7.7	9.7 ± 7.6	11.6 ± 7.0	PLD	424.7 ± 186.4	897.1 ± 315.4	508.8 ± 133.5
LCC	0.668 ± 0.204	0.703 ± 0.239	0.717 ± 0.185	PGD	445.2 ± 173.6	861.4 ± 348.2	663.1 ± 185.0
BC	32.3 ± 47.3	31.6 ± 63.7	26.8 ± 43.6	PLL	91.45 ± 48.96	91.23 ± 50.74	98.95 ± 48.81
1-CC	0.991 ± 0.002	0.992 ± 0.002	0.990 ± 0.002	PGA	123.7 ± 67.0	123.8 ± 68.1	130.7 ± 68.1

† Each of the four network characteristics (node degree, the local clustering coefficient [LCC], betweenness centrality [BC], and 1 minus closeness centrality [1-CC]) measures the connectedness of a species with other species in the unipartite network.
‡ The range characteristics refer to the distance (km) between the centroid of the species polyline range and center of the HUC polyline (PLD), distance between the centroid of the species polygon and center of the HUC polygon (PGD), total length (units of 10,000 km) of the species polyline (PLL), and total area (units of 10,000 km^2) of the species polygon (PGA).

Bipartite Networks

At the HUC2 scale, the degree value of the sample site nodes was significantly related to each of the three spatial predictors ($p < 0.01$). As expected, **site degree** decreased with flow-path distance (*Fig. 18*), reflecting the natural and well-documented tendency for fish species richness to be highest at low to mid-elevation sites, relative to smaller, more isolated high-elevation sites (e.g., McGarvey and Ward 2008; Schlosser 1987). Site degree also decreased with latitude, mirroring the well-documented latitudinal diversity gradient (e.g., Hillebrand 2004; Mittelbach et al. 2007), while increasing with longitude (i.e., from east-to-west). ANCOVA also revealed significant

interaction effects with HUC identity (indicated by '$p_{X\text{-}HUC}$' values in *Fig. 18*) in the latitude and longitude models, but not in the flow-path distance model. This interesting result suggested that 'secondary biogeographical gradients' in network structure may exist within the large Mississippi River Basin (i.e., differences among sub-drainages nested within the larger, complete basin), but that flow-path distance may be a more fundamental predictor of network structure that is invariant to latitudinal and longitudinal gradients. Significant interactions were not detected for disturbance level (see '$p_{X\text{-}disturb}$' values in *Fig. 18*).

Table 2. Significant relationships ($r^2 > 0.1$ and $p < 0.05$) between species' network metrics and range characteristics within three HUC2 sub-drainages. See *Table 1* for abbreviations.

Region	Range variable	Network variable	Std. regress. coefficient	Expected relationship	Outcome as expected?	r^2	p
HUC-0207	PGA × (1/PGD)	Degree	0.332	+	Yes	0.110	0.024
HUC-0210	PLL	Degree	0.567	+	Yes	0.320	<0.0001
	PGA	Degree	0.578	+	Yes	0.334	<0.0001
	PLD	LCC	0.435	_	No	0.190	0.002
	PGD	LCC	0.431	_	No	0.185	0.003
	PLL	LCC	-0.445	+	No	0.198	0.002
	PGA	LCC	-0.469	+	No	0.220	0.001
	PGA × (1/PGD)	LCC	-0.623	+	No	0.388	<0.0001
	PGA	BC	0.336	+	Yes	0.113	0.012
	PGA × (1/PGD)	BC	0.633	+	Yes	0.401	<0.0001
	PLL	1-CC	-0.533	+	No	0.281	<0.0001
	PGA	1-CC	-0.543	+	No	0.287	<0.0001
HUC-0211	PLD	Degree	-0.376	_	Yes	0.142	0.008
	PLL	Degree	0.356	+	Yes	0.127	0.012
	PGA	Degree	0.342	+	Yes	0.117	0.016
	PGA × (1/PGD)	Degree	0.405	+	Yes	0.164	0.004
	PLD	LCC	0.362	_	No	0.131	0.012

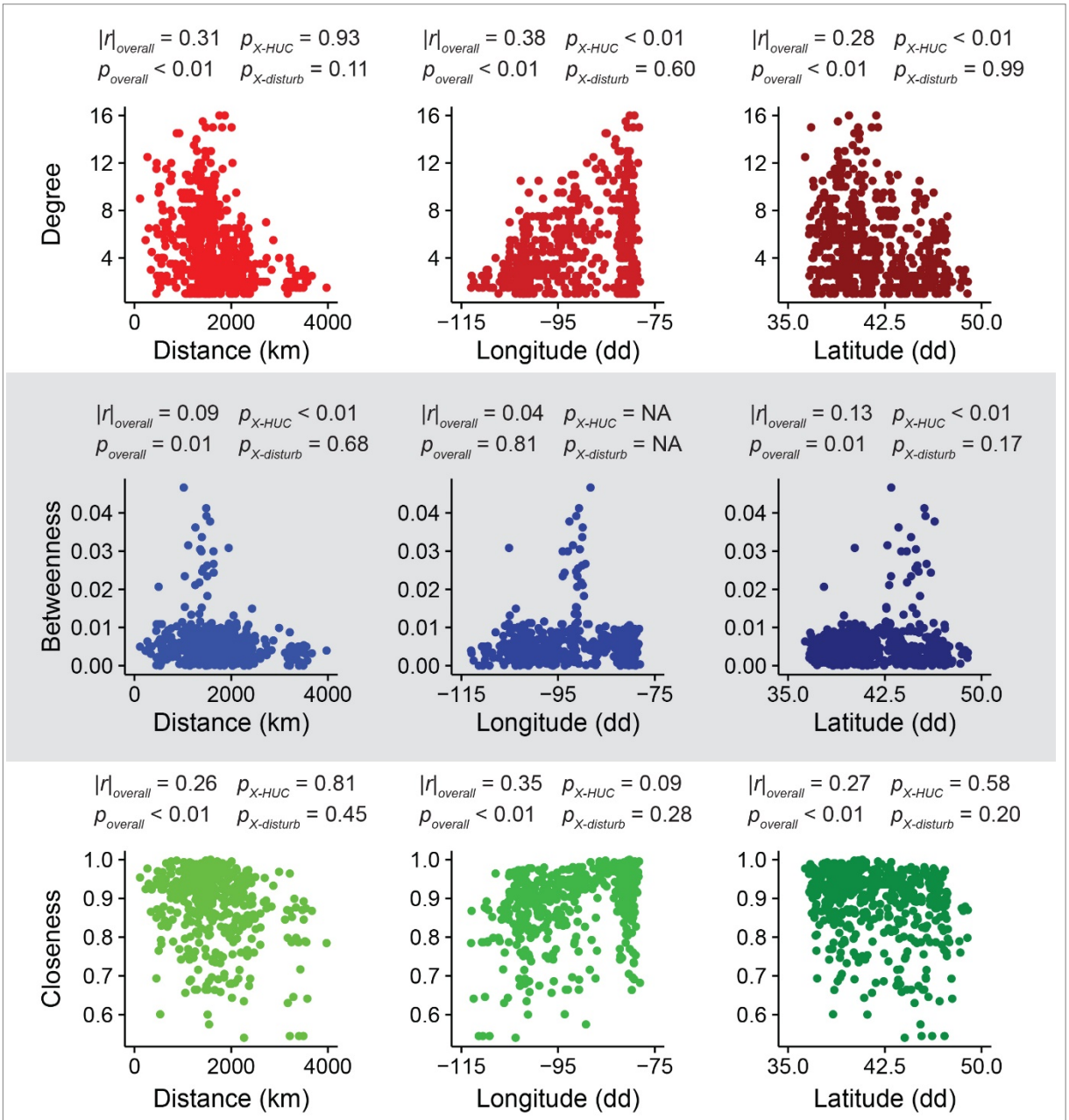


Figure 18. Linear regression results for HUC2 scale models of spatial position and bipartite network structure. Three spatial predictor variables (flow-path distance between each sampling site and the outlet of its respective river basin, longitude, and latitude) and three bipartite network metrics (degree, betweenness centrality, and closeness centrality for each site/node within the network) are shown (nine total combinations). For each model, the absolute correlation coefficient ($|r|_{\text{overall}}$) and linear model *p*-value (p_{overall}) are shown at the top of the scatterplot. Significance values for ANCOVA interactions with HUC identity ($p_{X\text{-}HUC}$) and disturbance level ($p_{X\text{-}disturb}$) are also shown at top.

Notably, while each of the site degree models was significant at the HUC2 scale, the strengths of the overall correlations were modest. The absolute values of the correlation coefficients ('$|r|_{overall}$' values) ranged from 0.28−0.38 and scatterplots of the primary effects (i.e., spatial predictor vs. degree) each exhibited a conspicuous triangle shape, with an apparent upper diagonal boundary (*Fig. 18*). These triangular shapes, when observed in univariate scatter plots, have been described as 'constraint spaces' and are hypothesized to represent upper-limits to the individual-level physiological or ecological tolerances that are ultimately reflected in large-scale, biogeographical patterns (Brown 1995). In the present context, it is tempting to infer that these constraints may be real. For instance, site-level degree within the Mississippi River Basin is clearly influenced by factors other than flow-path distance – the variability in degree is too high to reasonably conclude otherwise (see upper-left panel in *Fig. 18*) – but the acutely triangular shape of the scatterplot also begs further testing of the hypothesis that as flow-path distance (and the physiological burden that accompanies continued residency at far upstream locations within a directed stream network) increases, natural limits may be imposed on site-level degree or the numbers of species that can potentially co-occur at a given site.

At the HUC4 scale, none of the correlations ($|r|_{overall}$ values) between spatial predictors **and site-level degree** exceeded 0.1 and none of the regression models were significant (top row, *Fig. 19*). This was likely an artifact of the HUC4 sub-drainages being too small to encompass meaningful spatial gradients in degree values. For example, the triangular shape of the HUC2 scale 'flow-path distance vs. degree' plot (see *Fig. 18*) did not emerge until distance exceeded ~2,000 km; at the HUC4 scale, none of the flow-path distances were greater than 600 km. Nevertheless, the observation that perceived network structure may itself be a function of spatial scale is a useful result and the differences in regression models of spatial predictors vs. degree at HUC2 and HUC4 scales may help to clarify whether characteristic scales in fish co-occurrence network structure do, in fact, exist.

Similar to the model for degree, **betweenness centrality** at the HUC2 scale was a significant, negative function of flow-path distance (middle row, *Fig. 18*). However, the negative linear relationship was also a clear function of a modal 'spike' that occurred at a flow-path distance of ~1,500 km; without this spike, the correlation between distance and betweenness would have been much weaker. This is a notable observation because it suggests that a relatively small number of sites (i.e., nodes) located at mid-point distances along the longitudinal profile of a physical stream network (i.e., the modal sites in the middle-left plot of *Fig. 18*) may have uniquely important roles as 'habitat hubs' (given that betweenness is a measure of the tendency for many network connections to pass through a given node; see *Appendix B*). These hubs may facilitate connections among other sites that themselves have relatively low betweenness centrality values. By the same logic, these modal, high-betweenness sites may be uniquely important to the overall integrity of a network and therefore worthy of priority in conservation planning. A modal spike in betweenness centrality was also evident in the HUC2 scale plots of longitude and latitude, though the overall

regression model was not significant for longitude. In this particular scenario, however, a similar 'hub' interpretation may not be warranted because broad variation in latitude will often incorporate spatially independent stream networks. It is therefore more tenuous to assume that north-to-south shifts in betweenness centrality reflect an underlying, ecologically meaningful gradient such as the upstream-downstream dynamic of flow-path distance along a longitudinal river profile. At the HUC4 scale, longitude and latitude were both significant predictors of betweenness centrality (middle row, *Fig. 19*). However, given the diffuse nature of the respective scatterplots, we concluded that further interpretation of the results was unwarranted.

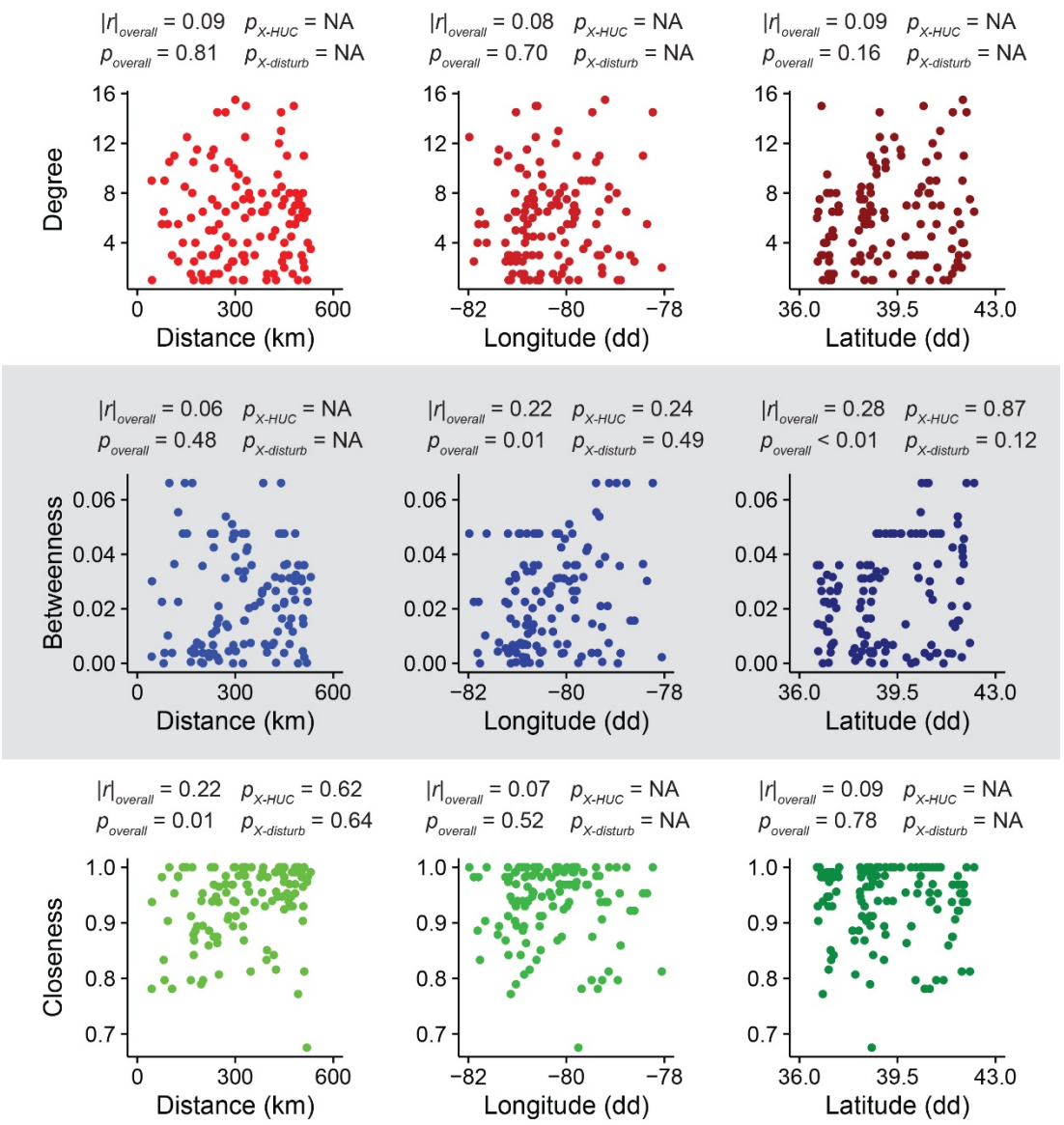


Figure 19. Linear regression results for HUC4 scale models of spatial position and bipartite network structure. All results and plots are formatted as in *Fig. 18*.

At the HUC2 scale, **closeness centrality** was significantly correlated with each of the three spatial predictor variables, while none of the secondary interactions (i.e., HUC identity and/or disturbance level) were significant (bottom row, *Fig. 18*). At the HUC4 scale, only flow-path distance was significantly correlated with closeness centrality and neither of the secondary interactions were significant (bottom row, *Fig. 19*). Because closeness centrality is a summative measure of the overall distance between a focal node and every other node within the network (i.e., large closeness centrality values indicate that a focal node is directly connected to many other nodes and/or nodes that are not directly connected are rarely far from the focal node's neighbors, such that long pathways within the network are rare; see *Appendix B*), it is intuitive that the closeness centrality results tended to mirror the node degree results. Closeness was negatively correlated with flow-path distance and latitude, but positively correlated with longitude. Unlike the other models, however, closeness centrality did not exhibit conspicuous triangular or modal patterns. Rather, the significance of the HUC2 scale closeness models (despite modest $|r|_{\text{overall}}$ values) appeared to reflect weak (but perhaps perceptible) trends within relatively large datasets (>500 sites). Nevertheless, the significant models do indicate that a relatively small number of sites (i.e., nodes) that do not share many species with other sites do exist at 'high' positions (i.e., long flow-path distances) within the physical stream networks, at more western longitudes, and at more northern latitudes. Interestingly, the direction of the significant relationship between flow-path distance and closeness centrality was reversed at the HUC4 scale; it became a positive, rather than negative, trend (bottom row, *Fig. 19*). This suggests that the dense cloud of points centered at ~1,500 km flow-path distance in the HUC2 closeness centrality model may have obscured a modal closeness value, as a locally positive trend could have occurred at lower distance values (e.g., from ~0−1,200 km; see lower-left plot in *Fig. 18*).

Conclusions and Implications for Future Research

Network Responses to Disturbance

This Limited Scope project was originally motivated by the question: "Does ecological network structure and/or stability change in a predictable way with chronic environmental disturbance?" To answer this question, we used fish co-occurrence data from East Coast, Mississippi River Basin, and Pacific Northwest streams to quantify network structure within systems that were minimally impacted ('good' condition), moderately impacted ('fair' condition), or highly impacted ('poor' condition). Overall our ability to detect and characterize network responses to disturbance was modest and most of the potential trends detected in our network data were equivocal. Indicators of connectivity and modularity within the networks were often variable among levels of disturbance, biogeographic regions, and spatial scales (HUC2 vs. HUC4 scales; see *Appendix C* and *Appendix D*). However, these differences were most often idiosyncratic and did not reflect clear, ordinal (i.e., consistently increasing or decreasing) trends with disturbance. In only several instances did the network metrics change in a consistent, predictable way among the three disturbance levels (see next paragraph). Furthermore, the large sample sizes that were necessary (and were clearly larger than we originally anticipated) to calculate the network metrics prevented us from obtaining robust results for many data combinations, particularly in the Pacific Northwest and at the smaller HUC4 scale (see gaps or missing bars in *Appendix C* and *Appendix D*). Thus, we emphasize that our results are, despite the considerable effort made to produce them, preliminary and should be interpreted with caution.

That said, several notable results with potentially broad management relevance did emerge from our analyses. First, individual species' **degree** values tended to *decrease* with increasing disturbance. This trend was observed for both maximum and median degree values (see *Appendix C1* and *C2*). Thus, we confirmed that the number of fish species that co-occur among a series of shared habitats (i.e., sampling sites) is, in most of the systems examined here, an inverse function of environmental disturbance. Intuitively, this makes sense as fish species richness has often been shown to decrease with disturbance (Matthews 1998). However, it is important to note that species' degree and species richness are not equivalent concepts. Richness is a static, aggregate reflection of overall diversity within a single sample or collection of samples, while species' degree is an individual-level measure of association within a collection of samples. Furthermore, while species' degree must ultimately be limited by overall richness (i.e., the number of species that a focal species co-occurs with cannot exceed the total number of species in the regional species pool), the opposite is not necessarily true: numbers of co-occurring species (i.e., degree) may remain low, despite high overall richness, if many species are limited to a relatively small number of localities and those localities are not broadly overlapping. Regardless, the fact that species' degree exhibited

a consistent tendency to decrease with increasing disturbance indicates that pair-wise associations among fishes are, in fact, diminished or lost as disturbance increases.

Individual species' **closeness** values showed the opposite trend: closeness tended to *increase* with disturbance, for both maximum and median closeness values (see *Appendix C10* and *C11*). This result was initially unexpected, given the negative trend between species' degree and disturbance noted above. We assumed that lost pair-wise associations would tend to make network distances among the remaining species longer (i.e., the number of links needed to connect any two nodes within a network should, on average, be greater when there are fewer links and nodes are less well-connected). But the increasing closeness values indicated that disturbance may diminish the frequency of indirect associations (i.e., decrease instances where nodes are separated by multiple intermediate links) within fish co-occurrence networks, resulting in networks where all species are either directly connected or separated by one or few intermediate links (e.g., two degrees of separation, rather than three or four).

Together the species' degree and closeness results suggest that ecological networks may become smaller (i.e., fewer species and among-species links) and more 'compact' (i.e., shorter average distances among species' nodes) as disturbance increases. This trend is illustrated in *Fig. 20* using the Ohio River Basin (HUC2 scale) weighted unipartite networks for complete species data (i.e., all links included). Total richness and the median degree values were highest for networks built from good condition sites and lowest for networks built from poor condition sites, while median closeness was highest in the poor condition network and lowest in the good condition network.

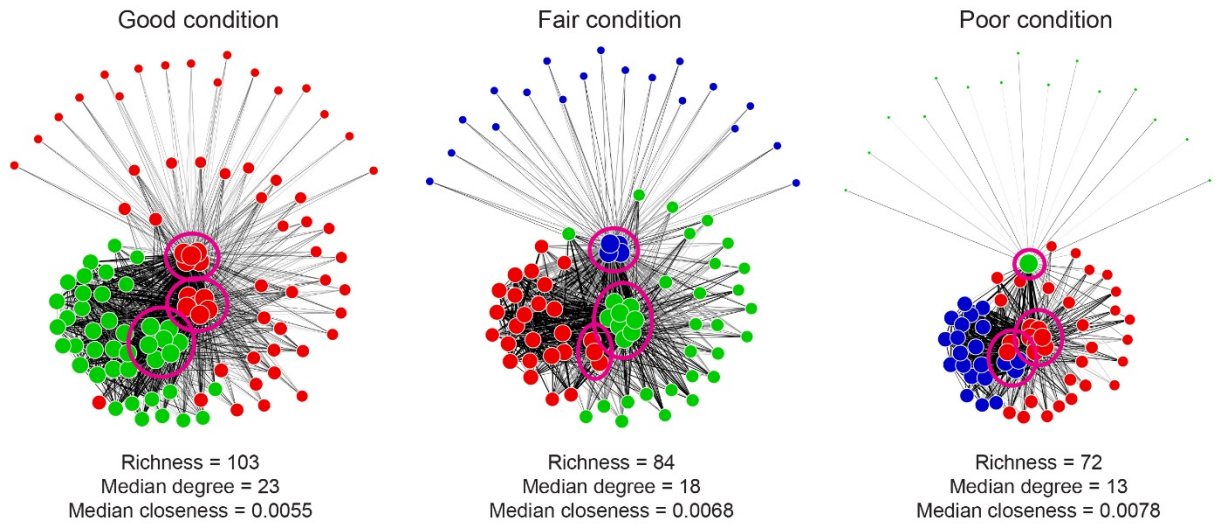


Figure 20. Weighted unipartite graphs for Ohio River Basin fish networks, built from sites in good, fair, and poor condition. Node sizes and colors represent node degree values and module membership, respectively. Potential hubs are circled. Graphs are as shown in *Appendix A*.

These trends in node degree and closeness provide some new insight to how ecological networks may respond to the types of chronic (e.g., gradual instream habitat degradation from grazing or non-point source pollution), rather than acute (e.g., toxic spills), disturbances that are most common in U.S. streams. Most notably, they suggest that stream fish networks may indeed be robust to chronic or gradual disturbances; although the networks became more compact with increasing disturbance, we found no evidence that these networks may be prone to rapid collapse. The observed changes in overall richness, node degree, and node closeness were of a gradual, incremental nature. Also, as disturbance increased, it was primarily the 'peripheral' species with low degree values that were lost from the networks. For instance, Brook Trout (*Salvelinus fontinalis*), Telescope Shiner (*Notropis telescopus*), Stonecat Madtom (*Noturus flavus*), Variegate Darter (*Etheostoma variatum*), Tonguetied Minnow (*Exoglossum laurae*), and Bigmouth Chub (*Nocomis platyrhynchus*) all had decreasing degree values among good and fair condition sites, then were lost from the networks (i.e., degree values of zero) in poor condition sites (see *Table 3*, showing degree values among good, fair, and poor condition sites for all fishes within the Ohio River Basin networks). In several instances, species became *more* highly connected as disturbance increased (e.g., Largemouth Bass [*Micropterus salmoides*], Black Bullhead [*Ameiurus melas*], and White Crappie [*Pomoxis annularis*]). But in general, we found that species-level responses to disturbance were predictable and that rapid, punctuated changes in network structure did not occur.

Furthermore, we noted that the identities of the most highly connected species within the networks were largely invariant to disturbance level. For example, Blacknose Dace (*Rhinichthys atratulus*), Central Stoneroller (*Campostoma anomalum*), Creek Chub (*Semotilus atromaculatus*), Fantail Darter (*Etheostoma flabellare*), Northern Hog Sucker (*Hypentelium nigricans*), Bluntnose Minnow (*Pimephales notatus*), Rock Bass (*Ambloplites rupestris*), and White Sucker (*Catostomus commersonii*) consistently had the highest degree values in good, fair, and poor condition networks from Ohio River Basin streams (*Table 3*). Because the K-S model fit *p*-values in tests of power-law degree distributions were generally insignificant for non-truncated degree distributions (see *Appendix C4*), we cannot characterize the fish co-occurrence networks as power-law networks *sensu stricto*. In many instances, however, we did observe a clear tendency for the most highly connected species to occur near the centers of the weighted unipartite graphs, where they provided high levels of connectivity within their respective network modules and also served as 'bridges' between modules (see below). This seemed to beg the question of whether these highly connected species may in fact serve as network hubs. For instance, three clusters of nodes with high degree values were evident in the unipartite graphs for good, fair, and poor condition sites within the Ohio River Basin (highlighted by pink circles in *Fig. 20*). Each of these highly connected, potential hub species is listed in *Table 4* with its respective module color (as shown in *Fig. 20*) for good, fair, and poor condition networks.

Table 3. Species' degree values for Ohio River Basin stream networks built from good, fair, and poor condition sites. Fish species' absences (zero degree values) are indicated by '--' marks. Species are listed by degree (highest to lowest 'good' values), then in alphabetical order.

	Degree values				Degree values		
Species	Good	Fair	Poor	Species	Good	Fair	Poor
Blacknose Dace	102	83	56	Southern Redbelly Dace	28	4	8
Central Stoneroller	102	83	56	Spotted Bass	28	32	27
Creek Chub	102	83	71	Variegate Darter	28	22	--
Fantail Darter	102	64	56	Yellow Bullhead	28	32	30
Northern Hog Sucker	102	64	46	Emerald Shiner	24	32	--
Bluntnose Minnow	85	83	56	Tonguetied Minnow	24	15	--
Greenside Darter	85	64	34	Bigeye Chub	23	4	--
Mottled Sculpin	85	44	27	Bigmouth Chub	23	12	--
Rock Bass	85	64	46	Cutlip Minnow	23	--	--
Smallmouth Bass	85	64	27	Freshwater Drum	23	21	8
White Sucker	85	64	56	Margined Madtom	23	--	--
Bluegill Sunfish	67	64	56	Whitetail Shiner	23	4	--
Johnny Darter	67	64	46	Channel Catfish	18	12	--
Rainbow Darter	67	54	34	Crescent Shiner	18	--	1
Striped Shiner	67	54	46	Gizzard Shad	18	21	8
Green Sunfish	63	54	56	Orangethroat Darter	18	18	16
River Chub	63	34	13	Redfin Shiner	18	--	1
Rosyface Shiner	63	32	8	Saffron Shiner	18	--	--
Brook Trout	54	15	--	Sharpnose Darter	18	4	--
Longear Sunfish	54	44	30	Shorthead Redhorse	18	15	8
Longnose Dace	54	18	13	White Shiner	18	--	--
Silverjaw Minnow	54	37	27	Banded Sculpin	15	12	8
Spotfin Shiner	54	32	27	Flathead Catfish	15	18	--
Logperch Darter	48	37	27	Scarlet Shiner	15	--	--
Banded Darter	46	32	30	Trout-Perch	15	--	--
Blackside Darter	46	32	20	American Brook Lamprey	11	4	1
Golden Redhorse	46	44	27	Blackstripe Topminnow	11	--	13
Telescope Shiner	46	12	--	Bluebreast Darter	11	--	--
Black Redhorse	39	32	13	Creek Chubsucker	11	--	1
Bluehead Chub	39	4	13	Golden Shiner	11	--	13
Brown Trout	39	32	16	Kanawha Minnow	11	--	--
Mountain Redbelly Dace	39	12	16	Least Brook Lamprey	11	4	13
Rosyside Dace	39	4	13	Longhead Darter	11	4	--
Silver Shiner	39	18	8	Pearl Dace	11	4	--
Stonecat Madtom	39	15	--	Redside Dace	11	21	13
Sand Shiner	38	15	27	Roanoke Darter	11	--	--
Common Carp	36	32	34	Rosefin Shiner	11	4	1
Common Shiner	36	22	1	Silver Redhorse	11	18	8
Rainbow Trout	36	4	--	Slimy Sculpin	11	4	--
Mimic Shiner	35	15	13	Spottail Shiner	11	4	--
Largemouth Bass	28	37	46	Streamline Chub	11	15	--
Pumpkinseed Sunfish	28	32	30	Suckermouth Minnow	11	--	--
Redbreast Sunfish	28	12	1	Yellow Perch	11	--	8

(*Continued on next page*)

Table 3 (*continued*). Species' degree values for Ohio River Basin stream networks built from good, fair, and poor condition sites.

Species	Degree values			Species	Degree values		
	Good	Fair	Poor		Good	Fair	Poor
Black Bullhead	5	--	13	Redear Sunfish	5	--	--
Black Crappie	5	4	--	Redfin Pickerel	5	--	1
Brook Silverside	5	4	20	River Redhorse	5	--	1
Brown Bullhead	5	15	1	Sauger	5	18	8
Fathead Minnow	5	12	13	Smallmouth Buffalo	5	18	--
Kanawha Sculpin	5	--	--	Walleye	5	12	--
New River Shiner	5	--	--	Warmouth Sunfish	5	--	--
Northern Studfish	5	--	--	White Crappie	5	--	20
Quillback Carpsucker	5	15	--				

Table 4. Lists of species 'hubs' within good, fair, and poor condition networks for the Ohio River Basin, including module membership as shown in *Fig. 20*.

Species	Module color		
	Good	Fair	Poor
Blacknose Dace	red	blue	red
Bluegill Sunfish	green	red	red
Bluntnose Minnow	red	red	red
Central Stoneroller	red	blue	red
Creek Chub	red	blue	green
Fantail Darter	red	green	red
Green Sunfish	green	red	red
Greenside Darter	red	green	--
Johnny Darter	green	blue	blue
Largemouth Bass	--	--	blue
Mottled Sculpin	red	--	--
Northern Hog Sucker	red	green	blue
Rainbow Darter	green	red	--
River Chub	green	--	--
Rock Bass	red	green	red
Rosyface Shiner	green	--	--
Smallmouth Bass	red	green	--
Striped Shiner	green	red	red
White Sucker	red	green	red

Species that stood out as potential network hubs were widely-distributed, common fishes that are often regarded as ecological 'generalists' with moderate to high tolerance to habitat alteration and other types of disturbances (e.g., Halliwell et al. 1999; Jenkins and Burkhead 1994; Trautman 1981; Wydoski and Whitney 2003). For this reason, these same species are often de-emphasized in biological assessments, which place greater emphasis on rare, sensitive species (e.g., Barbour et al. 1999; Fausch et al. 1984; Karr 1981). But our network results offer a different perspective. If species with the highest degree values serve as hubs that hold the fish co-occurrence networks together, as predicted by theory (Proulx et al. 2005; Tylianakis et al. 2010), then **these common species may be most important to the overall integrity of complete fish assemblages**.

Another observation that suggests a fundamental role for the most highly connected species is the fact that they clustered near the centers of the networks, but not in a single location. Rather, they were distributed among network modules (see *Fig. 20* and *Appendix A*). Thus, each module consisted of one or several high-degree species (i.e., hubs) that allowed more peripheral species to remain connected to the network. This is intriguing because the modules themselves did not appear to represent subsets of fishes with unique functional trait profiles (as we originally anticipated). Instead, the modules consisted, in most cases, of fishes with broadly overlapping functional traits (see *Fig. 13*). Specifically, each module was comprised of a similar mix of feeding behaviors, parental care strategies and spawning strategies (see *x*-axis in *Fig. 21*), and a similar range of body sizes, longevities, and fecundities (see *y*-axis in *Fig. 21*).

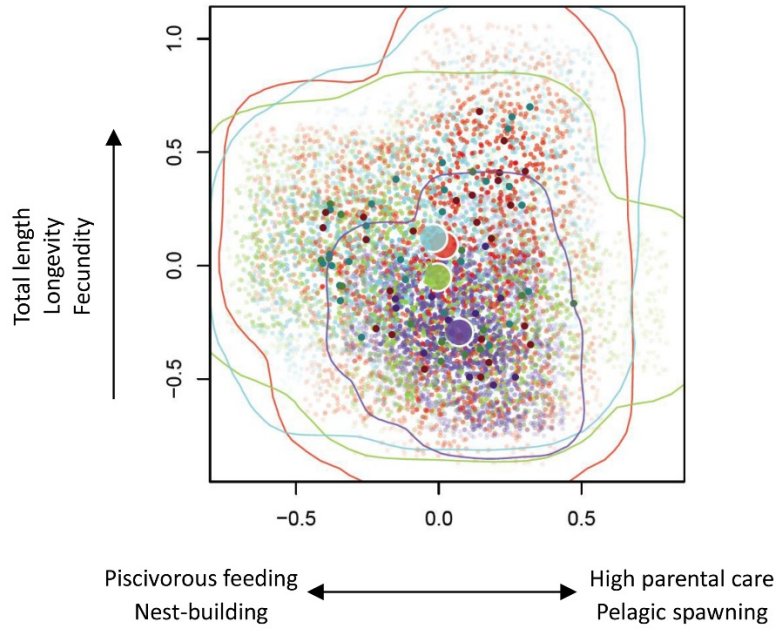


Figure 21. Functional hypervolume plot demonstrating overlap in functional traits space among fish network modules. Functional trait factors that are highly associated with each of the two major ordination axes are shown.

From a management and conservation perspective, our results may be cause for cautious optimism. Differences among networks in good, fair, and poor condition were detectable but not dramatic. With increasing disturbance, the networks became smaller and more compact, yet they did not collapse into fragments that were fundamentally different than complete, undisturbed networks. Additionally, the functional traits analyses suggested that distinct modules within the networks may constitute repeated motifs of species' life histories, feeding behaviors, and body sizes. Together, these observations point to a new question: *is the relative stability that we documented in the fish co-occurrence networks (i.e., modest change in response to disturbance) sustained by a highly level of redundancy within the networks?* In most of the networks, the putative hubs were not singular; within each module, several hubs tended to cluster together, providing potentially redundant links to more peripheral, weakly connected species (see *Fig. 20* and *Appendix A*). Similarly, the motif pattern observed in many of the functional traits ordination plots (see *Fig. 13* and *Fig. 21*) may imply that functional redundancy is present in the fish co-occurrence networks. If this latter proposition is correct, it would fundamentally challenge the current thinking on modularity within ecological networks. Consistent with recent theory, we began this project with the expectation that modules would, if detected, serve non-redundant functional roles within the larger networks (Olesen et al. 2007; Tylianakis et al. 2010). Thus, a comprehensive management strategy would need to protect the unique functional attributes of each module to ensure continued system-level function. However, our results suggest a very different possibility. If the functional characteristics of network modules are truly redundant, **an optimal management strategy might be to first ensure that one of the redundant modules will be protected in its entirety**.

At a minimum, managers should appreciate that the network approach and tools presented here can provide an alternative and perhaps more insightful framework for documenting and predicting how biotic assemblages respond to environmental disturbances than traditional biomonitoring methods. For example, multi-metric indices and 'RivPacs' models are currently two of the most widely employed tools for detecting and/or characterizing disturbance effects on freshwater biota (e.g., Barbour et al. 1995; Hawkins et al. 2000; Karr 1981; Van Sickle et al. 2005). Indeed, our disturbance rankings utilized a national-scale, multi-metric index for stream invertebrates (Herlihy et al. 2008; Stoddard et al. 2008; USEPA 2016). When properly calibrated, these tools are effective at detecting shifts in the basic structural properties of assemblages, such as changes in species richness, the prevalence of tolerant vs. intolerant taxa, and the absence of select indicator species. Unfortunately, these tools are rarely rooted in dynamic ecological or systems-level theory; they cannot predict (or even recognize) structural responses to no-analogue disturbances, nor can they be used to anticipate the secondary or tertiary effects (e.g., secondary extinction resulting from the loss of a focal species) that may propagate throughout an assemblage in response to a disturbance. We predict that network analysis and graph theory can provide a richer, more meaningful language and framework for understanding biotic communities and how they respond to environmental disturbance (Tylianakis et al. 2010). They can also provide powerful analytical tools to quantify and implement explicit regulatory standards.

Dynamic Network Simulations

Unipartite Simulations

Algorithms to simulate and predict network responses to disturbances provide an important compliment to observational studies of empirical network structure (Memmott et al. 2004; Tylianakis et al. 2010). We performed simulation experiments for both unipartite and bipartite networks, where species extinctions were simulated by sequentially removing species from the networks and re-calculating a series of network metrics. Unipartite network simulations were based on the Susquehanna River Basin (HUC4 scale; see *Fig. 22*) and led to three general conclusions. First, we observed that *extinction-by-sensitivity* (i.e., removing the most pollution intolerant species first) results did not deviate from random extinction events. Each of the six network metrics that were evaluated in the unipartite extinction simulations showed extinction-by-sensitivity results to be fully within the range or spread of the random results (see *Fig. 16*). The sole instance where sensitivity results seemed to deviate from the random results was the 'flat' sequence of median betweenness values that occurred between 8–15 extinctions (*Fig. 16E*). Closer inspection of the sensitivity values for Susquehanna River fishes did not reveal a clear explanation for this unusual result. It did, however, show that the sequence of species removals used in extinction-by-sensitivity simulations did not covary with species' degree values (see *Table 5*). Thus, the extinction-by-sensitivity removals were essentially random with respect to connectivity.

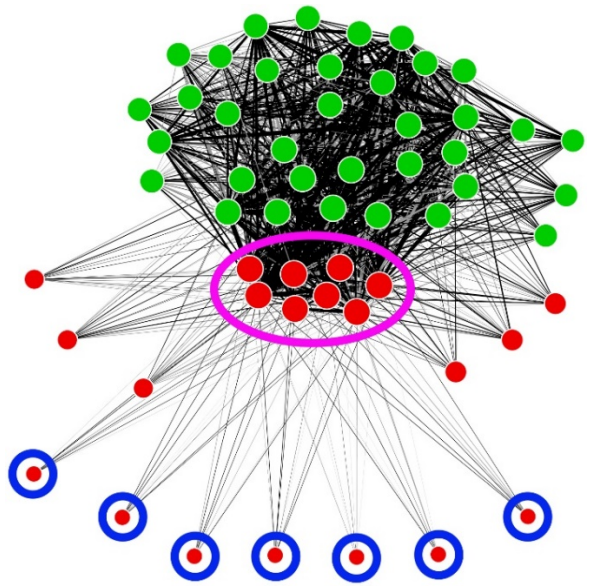


Figure 22. Weighted unipartite graph for the complete Susquehanna River Basin stream fish network. Node sizes and colors represent node degree values and module membership, respectively. Blue circles identify nodes that are lost as 'secondary extinctions' when all red hubs (circled in pink) are deleted from the network.

Table 5. Species' sensitivity values, degree values, and module membership for Susquehanna River Basin stream fishes. Species are listed by sensitivity (lower values = higher sensitivity), in the order that was used in extinction-by-sensitivity simulations. Module colors correspond to colors shown in *Fig. 22*.

Species	Sensitivity	Degree	Module	Species	Sensitivity	Degree	Module
Banded Darter	2	42	green	Sand Shiner	4	29	green
Mottled Sculpin	2	45	green	Slimy Sculpin	4	35	green
Redside Dace	2	33	green	Common Shiner	4	48	green
Rosyface Shiner	2	39	green	Fallfish	4	55	red
Northern Hog Sucker	2	48	green	Creek Chub	4	48	green
Greenside Darter	2	42	green	Walleye	4	42	green
River Chub	2	45	green	Central Stoneroller	4	48	green
Brook Trout	2	42	green	Margined Madtom	4	48	green
Longnose Dace	2	55	red	Emerald Shiner	4	29	green
Cutlip Minnow	3	55	red	White Crappie	5	22	red
Swallowtail Shiner	3	8	red	Largemouth Bass	5	42	green
Northern Pike	3	8	red	Bluegill Sunfish	5	45	green
Fantail Darter	3	8	red	Redbreast Sunfish	5	42	green
White Sucker	3	55	red	Spottail Shiner	5	45	green
Johnny Darter	3	36	green	Channel Catfish	5	16	red
Brown Trout	3	45	green	Redfin Pickerel	5	8	red
Golden Redhorse	3	16	red	Quillback Carpsucker	5	29	green
Muskellunge	3	22	red	Brown Bullhead	5	35	green
Shield Darter	3	48	green	Pumpkinseed Sunfish	5	48	green
Shorthead Redhorse	3	36	green	Black Crappie	5	8	red
Rainbow Trout	3	22	red	Yellow Perch	5	39	green
Spotfin Shiner	3	45	green	Banded Killifish	5	16	red
Chain Pickerel	3	33	green	Bluntnose Minnow	5	48	green
Smallmouth Bass	3	55	red	Common Carp	6	42	green
Tessellated Darter	3	55	red	Blacknose Dace	6	55	red
Mimic Shiner	3	33	green	Yellow Bullhead	6	39	green
Comely Shiner	4	8	red	Golden Shiner	6	8	red
Rock Bass	4	55	red	Green Sunfish	6	39	green

Second, the *extinction-by-degree* simulation results, though initially surprising, were driven by punctuated and highly predictable losses of network hubs. For example, the most highly connected species in the Susquehanna River Basin network were the eight red module nodes concentrated near the center of the unipartite graph (enclosed by pink circle in *Fig. 22*): these included the Longnose Dace (*Rhinichthys cataractae*), Cutlip Minnow (*Exoglossum maxillingua*), Smallmouth Bass (*Micropterus dolomieu*), Tessellated Darter (*Etheostoma olmstedi*), White Sucker, Fallfish (*Semotilus corporalis*), Rock Bass, and Blacknose Dace. When each of these eight species had been deleted from the network, five of the six network metrics examined in our simulations displayed a conspicuous, punctuated response (see *Fig. 16B-E*). Similarly, when the next sequential group of highly connected species (central nodes within the green module in *Fig. 22*)

was deleted from the network, the network metrics again displayed a punctuated response. This ability to explain and ultimately predict dynamic network responses to species losses will be critical in future efforts to apply network analysis and graph theory within natural resource management contexts (Tylianakis et al. 2010).

Finally, we found that it may be possible to predict the secondary effects on individual species, including secondary extinctions, of species losses within the network. When the eight most highly connected species had been deleted from the Susquehanna River Basin network (red module hubs circled in pink in *Fig. 22*), each of the seven peripheral, relatively weakly connected species within the red module (circled in blue in *Fig. 22*) was completely severed from the network (i.e., became extinct). Thus, we could predict that these peripheral species – the Comely Shiner, Swallowtail Shiner, Fantail Darter, Golden Shiner (*Notemigonus crysoleucas*), Black Crappie (*Pomoxis nigromaculatus*), Northern Pike (*Esox lucius*), and Redfin Pickerel – would likely be the most vulnerable to secondary effects of disruptions that eliminate the most highly connected species. Interestingly, the extinction-by-degree simulations also predicted that all species within the red module shown in *Fig. 22* would be lost as secondary extinctions before *any* species within the green module would become vulnerable to secondary extinctions (see nodes enclosed by heavy black circles in *Fig. 17*). This highly modular, compartmentalized vulnerability is itself notable for two reasons. First, it could aid conservation efforts by allowing managers to focus exclusively on species within one module, rather than the entire network. Second, it hints at a profound level of self-organization within in the network. If the network modules are, in fact, functionally redundant motifs (as hypothesized above) then the ability to sequester disturbance effects to a single module would confer a high level of robustness (i.e., maintaining all functional traits within undisturbed modules). This robustness would certainly be a remarkable example of evolutionary efficiency!

Bipartite Simulations

The spatially-explicit network analyses did not produce unambiguous evidence of strong spatial structure within the fish co-occurrence networks (see *Tables 1-2* and *Figs. 18-19*), but they did provide the necessary information to implement spatial extinction algorithms in the bipartite network simulations. In these simulations, we removed *sampling sites* from the DRNs then tracked the effects on resident fishes. By running these simulations in four test systems (the Susquehanna, Savannah, Kanawha, and Great Miami River Basins), we arrived at two general conclusions. First, the *extinction-by-degree* (i.e., deleting sites with the highest numbers of locally occurring fish species first) simulations (red lines in *Fig. 14*) tended to result in much higher species' extinction rates than random site deletions (green lines in *Fig. 14*). However, this was not true in all cases; in the Susquehanna River Basin, the extinction-by-degree algorithm generated species' extinctions at approximately the same rate as the random deletion algorithm (upper-left panel in *Fig. 14*). Thus, we emphasize that the spatial configuration of stream habitats and their resident fishes can

generate complex network dynamics that are not easy to anticipate. We intuitively expected that extinction-by-degree simulations would result in the most rapid species losses (i.e., linear trends in site deletion vs. remaining species plots, with low AUC values; see *Fig. 14*). But the Susquehanna River Basin results suggest that the spatial arrangement of high- and low-degree sites can influence that robustness of an ecological network to species' extinctions; otherwise, the extinction-by-degree trend for Susquehanna simulations would have been more similar to the other three systems.

Second, the *extinction-by-distance* (to the closest major DoD facility) simulations indicated that disturbances created by DoD facilities may lead to species' extinction rates that are no greater than random extinction rates. In all four test systems, the extinction-by-distance (blue lines in *Fig. 14*) and random extinction (green lines in *Fig. 14*) trajectories and AUC values in site deletion vs. remaining species plots were similar. This observation is encouraging from a management perspective because it suggests that **sites closest to the DoD facilities are not uniquely important to the overall integrity of their respective fish co-occurrence networks**. In future research, it may be prudent to replace the Euclidian distance algorithm used here (i.e., straight-line distance between the focal DoD facility and each stream site) with a river distance algorithm; literal flow-path distances between DoD facilities and stream sites could provide more ecologically meaningful measures of potential anthropogenic disturbances. For the moment, however, we suggest that DoD facilities within the four test systems should not, by simple virtue of their spatial orientations within the DRNs, pose undue threats to the regional fish networks.

Next Steps

To further advance this work and ensure that it meets its full potential in management applications, we propose a mix of modeling and field-based research in four key areas: (1) confirming that pair-wise links within the unipartite networks do (or do not) reflect true biological associations; (2) testing the hypothesis of functional motif redundancy among network modules; (3) extending the network simulations to better incorporate chronic/gradual disturbances and spatial dynamics; and (4) building user-friendly tools to make the network analyses and visualizations widely accessible.

A first step in future work will be to directly evaluate the network links among species-pairs. Testing whether large co-occurrence networks can be used to infer the strength and direction (i.e., positive versus negative) of species' interactions, across a wide range of organismal levels, is currently an active area of research (e.g., Araújo et al. 2011; Borthagaray et al. 2014; Cazelles et al. 2016; Harris 2016; Kissling et al. 2012; Morueta-Holme et al. 2016; Peura et al. 2015; Poisot et al. 2012; Poisot et al. 2015; Williams et al. 2014) and our network results clearly have potential to advance this field. While conducting this Limited Scope project, we made an initial effort to evaluate individual pair-wise species associations using literature data, but it quickly became

evident that a larger, more systematic effort would be needed to complete a meaningful analysis. We therefore propose that a series of field-trials to directly test specific network predictions is a logical next step to take in advancing ecological network research and applications.

Field-trials would begin by using the unipartite networks that consisted exclusively of significantly positive or significantly negative pair-wise associations to generate specific *a priori* hypotheses for testing. For example, *Fig. 23* illustrates the weighted unipartite network for good condition sites in the Ohio River Basin when only *significantly negative* associations (as predicted by the probabilistic model of co-occurrence; see Veech 2013) are included. In this network, the Bluehead Chub (*Nocomis leptocephalus*; circled in red in *Fig. 23*) exhibits significantly negative associations with nine species: Creek Chub, Greenside Darter (*Etheostoma blennioides*), Johnny Darter, Logperch Darter (*Percina caprodes*), Longear Sunfish (*Lepomis megalotis*), Rainbow Darter (*Etheostoma caeruleum*), River Chub (*Nocomis micropogon*), Silverjaw Minnow (*Notropis buccatus*), and Striped Shiner (*Luxilus chrysocephalus*). Using established protocols for *in situ* and laboratory trials (e.g., Echelle et al. 1972; Finger 1982; Fisher and Pearson 1987; Grossman et al. 1998; Reeves et al. 1987), we would directly test for negative interactions between the Bluehead Chub and these (infrequently) co-occurring species. This type of direct observational evidence could greatly bolster the validity of ecological network research.

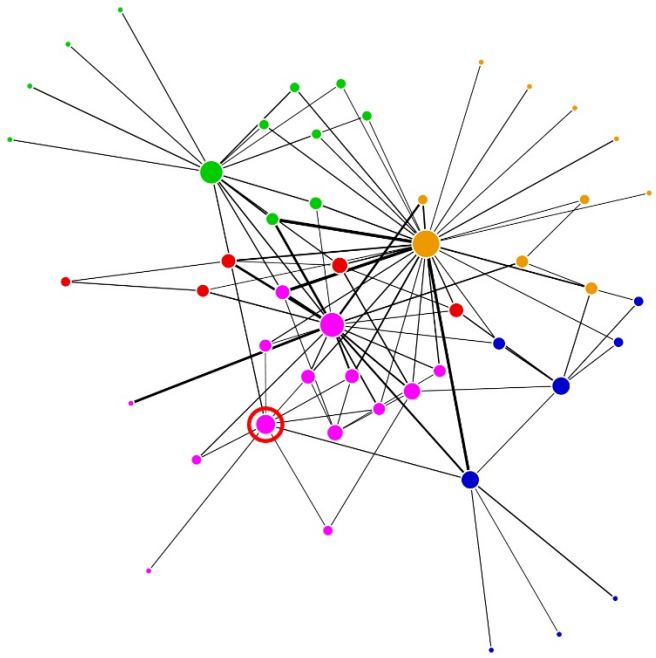


Figure 23. Weighted unipartite network of significantly negative associations within Ohio River Basin streams in good condition. The Bluehead Chub (circled in red) shares negative associations with nine other species.

Furthermore, we have already confirmed that the significantly positive and significantly negative network results tended to mirror the complete network trends among good, fair, and poor condition sites; most of the indicators of network structure that we evaluated did not change in a predictable way with disturbance, but degree values and closeness values did. For instance, unipartite networks of Ohio River Basin fishes that were built exclusively from *significantly positive* co-occurrences decreased in size (i.e., total number of nodes) from 63 to 52 to 46 species, respectively, among sites in good, fair, and poor condition. Thus, we are optimistic that field-trial results could be integrated with broader theory on ecological networks and how they respond to disturbance. Additionally, if field-trials can confirm which of the network links are truly reflective of direct, positive or negative interspecific dynamics, then managers will be able to predict, with a high level of confidence, which species are most likely to be positively or negatively influenced by specific disruptions within a network.

The second area of emphasis in future research will be a closer examination of the hypothesis that network modules are repetitive functional motifs and that these motifs may enhance network stability by providing functional redundancy. A similar theory was recently demonstrated using arthropod community data from the Atacama Desert, Chile by Borthagaray et al. (2014). They detected six modules of species, each of which included a similar mix of species' body sizes and trophic behaviors, and that responded similarly to environmental gradients. We predict that modules within the fish co-occurrence networks will, upon closer inspection, exhibit specific, predictable functional traits patterns: subsets of species with complimentary behaviors and life histories that efficiently 'fill' the ecological niche space available to them. This is a particularly exciting topic for future inquiry because research on the functional traits and life history strategies of stream fish assemblages is evolving so rapidly (e.g., Heino et al. 2013; Keck et al. 2014; Logez et al. 2013; Logez et al. 2016; Meador and Brown 2015; Mims et al. 2010; Pyron et al. 2011; Troia and Gido 2015). More importantly, research on the mechanistic links between fish functional traits and major environmental drivers, such as hydrology, can provide a rigorous foundation for predicting how fishes will respond to no-analogue environments under climate change (e.g., Buisson et al. 2013; McManamay and Frimpong 2015; McManamay et al. 2015; Olden et al. 2010; Poff and Allan 1995). Thus, **by acquiring a better understanding of the modularity and functional traits patterns described in this Limited Scope report, managers may greatly enhance their ability to forecast and plan for future disturbances**, including perturbations that are fundamentally distinct from the three classes of generalized, instream habitat disturbance that were considered herein.

A third future objective is to bolster the utility of the network simulations by incorporating more realistic disturbances in them and making better use of the spatial DRN information. Simulations used in this Limited Scope project were limited exclusively to species' extinctions and assumed that secondary effects (i.e., extinctions) within the network would not be realized until all of the links leading to a given species had been deleted (via the primary extinctions enforced under the

extinction-by-degree, extinction-by-distance, extinction-by-sensitivity, and random extinction algorithms). However, we acknowledge that this assumption is likely inaccurate; species within a network will almost certainly experience secondary effects prior to acute extinction events. At present, we are unaware of a published precedent or method to incorporate gradual, chronic disturbances within network simulations. We therefore propose that developing such capabilities would be an important step forward in basic and applied research. A potential pathway forward may be to utilize the edge weights within the networks. The probabilistic co-occurrence model allowed us to calculate edge weights among species-pairs, as well as the overall direction (positive versus negative) of the pair-wise associations, but we did not use them in the present research. If edge weights were included in the simulations, it may be possible to multiply them by 'disturbance coefficients' (e.g., an acute extinction equates to a disturbance coefficient of 1, while the coefficient for a more subtle temperature shift that impairs reproductive success is < 1) that reflect more subtle and realistic secondary effects. We also propose the addition of more complex, ecologically meaningful spatial algorithms in the network simulations. Simple 'as the crow flies' distances between DoD facilities and stream sites were used in the present simulations. This was intended to demonstrate the potential utility of spatial information and site-specific analyses. But there is clearly room for improvement. For instance, explicit flow-path distances among sites within the DRNs could be used to simulate the effects of instream pollutants. Alternatively, it should be possible to identify all segments within a DRN that must be protected to ensure that connectivity is maintained for species of special concern (Altermatt 2013). This dynamic, spatially-explicit approach to simulating disturbance effects within networks would entail a heavy computation burden. But it would also **give managers at specific locations (e.g., DoD facilities) greater ability to anticipate and plan for changes that are most likely to impact them directly**.

The fourth key objective in a future project will be to develop software tools that make the network calculations, simulations, and visualizations easier and more widely accessible. Most of the procedures described and graphics shown in this Limited Scope report required (semi-)custom programming (mostly in the R statistical programming language). Thus, the network methods and analyses that we have described are, at the moment, unavailable to much of the conservation and management community, as well as the general public. Full implementation of a network-based approach to studying and managing co-occurrence networks, in streams or elsewhere, will therefore mandate **online tools and/or custom software that can automate many of these procedures through graphical-user interfaces**. We envision the development of a web-based toolkit that will allow users to query data through a mapping interface, to specify network calculations they wish to perform via drop-down menus and toggle lists, then to visualize and download results in familiar formats (e.g., scalable vector graphics and comma-delimited text). As a conceptual example, we reference the ClimateWizard (www.climatewizard.org/).

Literature Cited

Albert, R. & Barabási, A.-L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74: 47-97.

Albert, R., Jeong, H. & Barabási, A.-L. 2000. Error and attack tolerance of complex networks. *Nature* 406: 378-382.

Allesina, S., Bodini, A. & Pascual, M. 2009. Functional links and robustness in food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 1701-1709.

Allesina, S. & Tang, S. 2012. Stability criteria for complex ecosystems. *Nature* 483: 205-208.

Altermatt, F. 2013. Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology* 47: 365-377.

Araújo, M.B. & Rozenfeld, A. 2014. The geographic scaling of biotic interactions. *Ecography* 37: 406-415.

Araújo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. 2011. Using species co-occurrence networks to assess the impacts of climate change. *Ecography* 34: 897-908.

Barbour, M.T., Gerritsen, J., Snyder, B.D. & Stribling, J.B. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, D.C.

Barbour, M.T., Stribling, J.B. & Karr, J.R. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. Pp. 63-77 in: Davis, W.S. & Simon, T.P., eds. *Biological assessment and criteria: tools for water resource planning and decision making*. CRC Press, Inc., Boca Raton, FL.

Bascompte, J. & Jordano, P. 2007. Plant-animal mutualistic networks: the architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 38: 567-593.

Blanck, A., Tedesco, P.A. & Lamouroux, N. 2007. Relationships between life-history strategies of European freshwater fish species and their habitat preferences. *Freshwater Biology* 52: 843-859.

Blois, J.L., Zarnetske, P.L., Fitzpatrick, M.C. & Finnegan, S. 2013. Climate change and the past, present, and future of biotic interactions. *Science* 341: 499-504.

Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. 2014. The *n*-dimensional hypervolume. *Global Ecology and Biogeography* 23: 595-609.

Borthagaray, A.I., Arim, M. & Marquet, P.A. 2014. Inferring species roles in metacommunity structure from species co-occurrence networks. *Proceedings of the Royal Society B: Biological Sciences* 281.

Brown, J.H. 1995. *Macroecology*. The University of Chicago Press, Chicago, IL.

Buisson, L., Grenouillet, G., Villéger, S., Canal, J. & Laffaille, P. 2013. Toward a loss of functional diversity in stream fish assemblages under climate change. *Global Change Biology* 19: 387-400.

Callaway, D.S., Newman, M.E.J., Strogatz, S.H. & Watts, D.J. 2000. Network robustness and fragility: percolation on random graphs. *Physical Review Letters* 85: 5468-5471.

Callaway, Ragan M. & Pennings, Steven C. 2000. Facilitation may buffer competitive effects: indirect and diffuse interactions among salt marsh plants. *The American Naturalist* 156: 416-424.

Camacho, J., Guimerà, R. & Nunes Amaral, L.A. 2002. Robust patterns in food web structure. *Physical Review Letters* 88: 228102.

Cao, Y., Larsen, D.P. & Hughes, R.M. 2001. Evaluating sampling sufficiency in fish assemblage surveys: a similarity-based approach. *Candian Journal of Fisheries and Aquatic Sciences* 58: 1782-1793.

Carstensen, D.W., Lessard, J.-P., Holt, B.G., Krabbe Borregaard, M. & Rahbek, C. 2013. Introducing the biogeographic species pool. *Ecography* 36: 1310-1318.

Cazelles, K., Araújo, M.B., Mouquet, N. & Gravel, D. 2016. A theory for species co-occurrence in interaction networks. *Theoretical Ecology* 9: 39-48.

Chessman, B.C. 1995. Rapid assessment of rivers using macroinvertebrates: a procedure based on habitat-specific sampling, family level identification and a biotic index. *Australian Journal of Ecology* 20: 122-129.

Cilleros, K., Allard, L., Grenouillet, G. & Brosse, S. 2016. Taxonomic and functional diversity patterns reveal different processes shaping European and Amazonian stream fish assemblages. *Journal of Biogeography* 43: 1832-1843.

Connell, J.H. 1978. Diversity in tropical rain forests and coral reefs. *Science* 199: 1302-1310.

Csárdi, G. & Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems: 1695.

Dalsgaard, B., Martín González, A.M., Olesen, J.M., Timmermann, A., Andersen, L.H. & Ollerton, J. 2008. Pollination networks and functional specialization: a test using Lesser Antillean plant–hummingbird assemblages. *Oikos* 117: 789-793.

Dormann, C.F., Gruber, B. & Fründ, J. 2008. Introducing the bipartite package: analysing ecological networks. *R News* 8: 8-11.

Dunne, J.A., Williams, R.J. & Martinez, N.D. 2002a. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences* 99: 12917-12922.

Dunne, J.A., Williams, R.J. & Martinez, N.D. 2002b. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters* 5: 558-567.

Echelle, A.A., Echelle, A.F. & Hill, L.G. 1972. Interspecific interactions and limiting factors of abundance and distribution in the Red River pupfish, *Cyprinodon rubrofluviatilis*. *American Midland Naturalist* 88: 109-130.

Fausch, K.D., Karr, J.R. & Yant, P.R. 1984. Regional application of an index of biotic integrity based on stream fish communities. *Transactions of the American Fisheries Society* 113: 39-55.

Finger, T.R. 1982. Interactive segregation among three species of sculpins (*Cottus*). *Copeia* 1982: 680-694.

Fisher, W.L. & Pearson, W.D. 1987. Patterns of resource utilization among four species of darters in three central Kentucky streams. Pp. 69-76 in: Matthews, W.J. & Heins, D.C., eds. *Community and evolutionary ecology of North American stream fishes*. University of Oklahoma Press, Norman, OK.

Fortuna, M.A. & Bascompte, J. 2008. The network approach in ecology. Pp. 371-392 in: Valladares, F., Camacho, A., Elosegui, A., Gracia, C., Estrada, M., Carles Senar, J. & Maria Gili, J., eds. *Unity in diversity: reflections on ecology after the legacy of Ramon Margalef*. Fundación BBVA, Madrid, Spain.

Frimpong, E.A. & Angermeier, P.L. 2009. FishTraits: a database of ecological and life-history traits of freshwater fishes of the United States. *Fisheries* 34: 487-495.

Goldstein, R.M. & Meador, M.R. 2004. Comparisons of fish species traits from small streams to large rivers. *Transactions of the American Fisheries Society* 133: 971-983.

Gotelli, N.J. & Ellison, A.M. 2004. *A primer of ecological statistics*. Sinauer Associates, Inc., Sunderland, MA.

Griffith, D.M., Veech, J.A. & Marsh, C.J. 2014. Cooccur: probabilistic species co-occurrence analysis in R. [https://cran.r-project.org/web/packages/cooccur/index.html.]

Gross, J.L. & Yellen, J. 2006. *Graph theory and its applications*. Chapman & Hall/CRC, Boca Raton, FL.

Grossman, G.D., Ratajczak, R.E.J., Crawford, M. & Freeman, M.C. 1998. Assemblage organization in stream fishes: effects of environmental variation and interspecific interactions. *Ecological Monographs* 68: 395-420.

Growns, J.E., Chessman, B.C., Jackson, J.E. & Ross, D.G. 1997. Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. *Journal of the North American Benthological Society* 16: 682-693.

Halliwell, D.B., Langdon, R.W., Daniels, R.A., Kurtenbach, J.P. & Jacobson, R.A. 1999. Classification of freshwater fish species of the northeastern United States for use in the development of indices of biological integrity, with regional applications. Pp. 301-337 in: Simon, T.P., ed. *Assessing the sustainability and biological integrity of water resources using fish communities*. CRC Press, Boca Raton, FL.

Harris, D.J. 2016. Inferring species interactions from co-occurrence data with Markov networks. *Ecology* 97: 3308-3314.

Harrison, G.W. 1979. Stability under environmental stress: resistance, resilience, persistence, and variability. *The American Naturalist* 113: 659-669.

Hawkins, C.P., Norris, R.H., Hogue, J.N. & Feminella, J.W. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10: 1456-1477.

Heino, J., Schmera, D. & Erős, T. 2013. A macroecological perspective of trait patterns in stream communities. *Freshwater Biology* 58: 1539-1555.

Herlihy, A.T., Paulsen, S.G., Sickle, J.V., Stoddard, J.L., Hawkins, C.P. & Yuan, L.L. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. *Journal of the North American Benthological Society* 27: 860-877.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. *The American Naturalist* 163: 192-211.

Hocutt, C.H. & Wiley, E.O. 1986. *The zoogeography of North American freshwater fishes*. John Wiley and Sons, New York, NY.

Hoeinghaus, D.J., Winemiller, K.O. & Birnbaum, J.S. 2007. Local and regional determinants of stream fish assemblage structure: inferences based on taxonomic vs. functional groups. *Journal of Biogeography* 34: 324–338.

Holling, C.S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4: 1-23.

Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Intelmann, S.S., Corbett, S.C., Arbogast, M.C. & Hjort, R.C. 2002. Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. *North American Journal of Fisheries Management* 22: 1229-1240.

Hughes, R.M., Paulsen, S.G. & Stoddard, J.L. 2000. EMAP-surface waters: a multiassemblage, probability survey of ecological integrity in the USA. *Hydrobiologia* 422: 429-443.

Huston, M.A. 1979. A general hypothesis of species diversity. *American Naturalist* 113: 81-101.

Jenkins, R.E. & Burkhead, N.M. 1994. *Freshwater fishes of Virginia*. American Fisheries Society, Bethesda, MD.

Jester, D.E., Echelle, A.A., Matthews, W.J., Pigg, J., Scott, C.M. & Collins, K.D. 1992. The fishes of Oklahoma, their gross habitats, and their tolerance of degradation in water quality and habitat. *Proceedings of the Oklahoma Academy of Science* 72: 7-19.

Jha, M., Pan, Z., Takle, E.S. & Gu, R. 2004. Impacts of climate change on streamflow in the Upper Mississippi River Basin: a regional climate model perspective. *Journal of Geophysical Research: Atmospheres* 109: D09105.

Jordán, F. 2009. Keystone species and food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 1733-1741.

Jordano, P., Bascompte, J. & Olesen, J.M. 2003. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters* 6: 69-81.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6: 21-27.

Keck, B.P., Marion, Z.H., Martin, D.J., Kaufman, J.C., Harden, C.P., Schwartz, J.S. & Strange, R.J. 2014. Fish functional traits correlated with environmental variables in a temperate biodiversity hotspot. *PLoS ONE* 9: e93237.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. 2012. Towards novel approaches to modelling biotic

interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* 39: 2163-2178.

Lafferty, K.D. & Kuris, A.M. 2009. Parasites reduce food web robustness because they are sensitive to secondary extinction as illustrated by an invasive estuarine snail. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 1659-1663.

Laliberté, E. & Legendre, P. 2010. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* 91: 299-305.

Laliberté, E., Legendre, P. & Shipley, B. 2014. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. [https://cran.r-project.org/web/packages/FD/FD.pdf.]

Lazorchak, J.M., Hill, B.H., Averill, D.K., Peck, D.V. & Klemm, D.J. 2000. *Environmental Monitoring and Assessment Program - Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-Wadeable Rivers and Streams*. U.S. Environmental Protection Agency, Cincinnati, OH.

Lazorchak, J.M., Klemm, D.J. & Peck, D.V. 1998. *Environmental Monitoring and Assessment Program - Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams*. EPA/620/R-94/004F. U.S. Environmental Protection Agency, Washington, D.C.

Logez, M., Bady, P., Melcher, A. & Pont, D. 2013. A continental-scale analysis of fish assemblage functional structure in European rivers. *Ecography* 36: 080-091.

Logez, M., Reyjol, Y. & Rodríguez, M.A. 2016. Life-history variation in freshwater fish: an assessment of the 'opportunistic–equilibrium–periodic' triangular model. *Ecology of Freshwater Fish* 25: 682-685.

Lyons, J. 1992. *Using the Index of Biotic Integrity (IBI) to measure enviromental quality in warmwater streams of Wisconsin*. North Central Forest Experiment Station, St. Paul, MN.

Matthews, W.J. 1998. *Patterns in freshwater fish ecology*. Chapman and Hall, New York, NY.

McGarvey, D.J. & Hughes, R.M. 2008. Longitudinal zonation of Pacific Northwest (U.S.A.) fish assemblages and the species-discharge relationship. *Copeia* 2008: 311-321.

McGarvey, D.J. & Ward, G.M. 2008. Scale dependence in the species-discharge relationship for fishes of the southeastern U.S.A. *Freshwater Biology* 53: 2206-2219.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. 2006. Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution* 21: 178-185.

McKay, L., Bondelid, T., Dewald, T., Rea, A., Johnston, C. & Moore, R. 2015. *NHDPlus Version 2: user guide* (Data Model Version 2.1). [ftp://ftp.horizon-systems.com/NHDplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.]

McManamay, R.A. & Frimpong, E.A. 2015. Hydrologic filtering of fish life history strategies across the United States: implications for stream flow alteration. *Ecological Applications* 25: 243-263.

McManamay, R.A., Peoples, B.K., Orth, D.J., Dolloff, C.A. & Matthews, D.C. 2015. Isolating causal pathways between flow and fish in the regulated river hierarchy. *Canadian Journal of Fisheries and Aquatic Sciences* 72: 1731-1748.

Meador, M. & Brown, L. 2015. Life history strategies of fish species and biodiversity in eastern USA streams. *Environmental Biology of Fishes* 98: 663-677.

Melián, C.J. & Bascompte, J. 2002. Complex networks: two ways to be robust? *Ecology Letters* 5: 705-708.

Memmott, J., Waser, N.M. & Price, M.V. 2004. Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London B: Biological Sciences* 271: 2605-2611.

Miles, E.L., Snover, A.K., Hamlet, A.F., Callahan, B. & Fluharty, D. 2000. Pacific Northwest regional assessment: the impacts of climate variability and climate change on the water resources of the Columbia River Basin. *Journal of the American Water Resources Association* 36: 399-420.

Mims, M.C., Olden, J.D., Shattuck, Z.R. & Poff, N.L. 2010. Life history trait diversity of native freshwater fishes in North America. *Ecology of Freshwater Fish* 19: 390-400.

Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M. & Turelli, M. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315-331.

Montoya, J.M. & Solé, R.V. 2002. Small world patterns in food webs. *Journal of Theoretical Biology* 214: 405-412.

Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B.J., Peet, R.K., Ott, J.E., Violle, C., Enquist, B.J., Jørgensen, P.M. & Svenning, J.-C. 2016. A network approach for inferring species associations from co-occurrence data. *Ecography* 39: 1139-1150.

Mote, P.W., Parson, E.A., Hamlet, A.F., Keeton, W.S., Lettenmaier, D., Mantua, N., Miles, E.L., Peterson, D.W., Peterson, D.L., Slaughter, R. & Snover, A.K. 2003. Preparing for climatic change: the water, salmon, and forests of the Pacific Northwest. *Climatic Change* 61: 45-88.

NatureServe. 2010. *Digital distribution maps of the freshwater fishes in the conterminous United States. Version 3.0*. NatureServe, Arlington, VA. [http://www.natureserve.org/sites/default/files/publications/fishdata.pdf]

Newman, M. 2003. The structure and function of complex networks. *SIAM Review* 45: 167-256.

Newman, M.E.J. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103: 8577-8582.

Newman, M.E.J., Strogatz, S.H. & Watts, D.J. 2000. *Random graphs with arbitrary degree distribution and their applications*. Santa Fe Institute, Santa Fe, NM.

Niu, S.Q., Franczyk, M.P. & Knouft, J.H. 2012. Regional species richness, hydrological characteristics and the local species richness of assemblages of North American stream fishes. *Freshwater Biology* 57: 2367-2377.

Olden, J.D., Kennard, M.J., Leprieur, F., Tedesco, P.A., Winemiller, K.O. & García-Berthou, E. 2010. Conservation biogeography of freshwater fishes: recent progress and future challenges. *Diversity and Distributions* 16: 496-513.

Olden, J.D., Poff, L.N., Douglas, M.R., Douglas, M.E. & Fausch, K.D. 2004. Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution* 19: 18-24.

Olden, J.D., Poff, N.L. & Bestgen, K.R. 2006. Life-history strategies predict fish invasions and extirpations in the Colorado River Basin. *Ecological Monographs* 76: 25-40.

Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences* 104: 19891-19896.

Olesen, J.M., Dupont, Y.L., O'Gorman, E., Ings, T.C., Layer, K., Melián, C.J., Trøjelsgaard, K., Pichler, D.E., Rasmussen, C. & Woodward, G. 2010. From broadstone to Zackenberg: space, time and hierarchies in ecological networks. *Advances in Ecological Research* 42: 1-69.

Olff, H., Alonso, D., Berg, M.P., Eriksson, B.K., Loreau, M., Piersma, T. & Rooney, N. 2009. Parallel ecological networks in ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 1755-1779.

Peura, S., Bertilsson, S., Jones, R.I. & Eiler, A. 2015. Resistant microbial co-occurrence patterns inferred by network topology. *Applied and Environmental Microbiology*.

Pianka, E.R. 1974. Niche overlap and diffuse competition. *Proceedings of the National Academy of Sciences* 71: 2141-2145.

Pla, L., Casanoves, F. & Di Rienzo, J. 2012. *Quantifying functional biodiversity*. Springer, Dordrecht, Netherlands.

Podani, J. 1999. Extending Gower's general coefficient of similarity to ordinal characters. *Taxon* 48: 331-340.

Poff, N.L. & Allan, J.D. 1995. Functional organization of stream fish assemblages in relation to hydrological variability. *Ecology* 76: 606-627.

Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. 2012. The dissimilarity of species interaction networks. *Ecology Letters* 15: 1353-1361.

Poisot, T., Stouffer, D.B. & Gravel, D. 2015. Beyond species: why ecological interaction networks vary through space and time. *Oikos* 124: 243-251.

Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. 2005. Network thinking in ecology and evolution. *Trends in Ecology and Evolution* 20: 345-353.

Pyron, M., Williams, L., Beugly, J. & Jacquemin, S.J. 2011. The role of trait-based approaches in understanding stream fish assemblages. *Freshwater Biology* 56: 1579-1592.

Reeves, G.H., Everest, F.H. & Hall, J.D. 1987. Interactions between the redside shiner (*Richarsonius balteatus*) and the steelhead trout (*Salmo gairdneri*) in western Oregon: the

influence of water temperature. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 1603-1613.

Reynolds, L., Herlihy, A.T., Kaufmann, P.R., Gregory, S.V. & Hughes, R.M. 2003. Electrofishing effort requirements for assessing species richness and biotic integrity in western Oregon streams. *North American Journal of Fisheries Management* 23: 450-461.

Rodriguez-Cabal, M.A., Barrios-Garcia, M.N., Amico, G.C., Aizen, M.A. & Sanders, N.J. 2013. Node-by-node disassembly of a mutualistic interaction web driven by species introductions. *Proceedings of the National Academy of Sciences* 110: 16503-16507.

Rogers, C.E. & McCarty, J.P. 2000. Climate change and ecosystems of the Mid-Atlantic Region. *Climate Research* 14: 235-244.

Sallenger, A.H., Doran, K.S. & Howd, P.A. 2012. Hotspot of accelerated sea-level rise on the Atlantic coast of North America. *Nature Climate Change* 2: 884-888.

Saterberg, T., Sellman, S. & Ebenman, B. 2013. High frequency of functional extinctions in ecological networks. *Nature* 499: 468-470.

Sauve, A.M.C., Fontaine, C. & Thébault, E. 2014. Structure–stability relationships in networks combining mutualistic and antagonistic interactions. *Oikos* 123: 378-384.

Schlosser, I.J. 1987. A conceptual framework for fish communities in small warmwater streams. Pp. 17-24 in: Matthews, W.J. & Heins, D.C., eds. *Community and evolutionary ecology of North American stream fishes*. University of Oklahoma Press, Norman, OK.

Secchi, S., Gassman, P.W., Jha, M., Kurkalova, L. & Kling, C.L. 2010. Potential water quality changes due to corn expansion in the Upper Mississippi River Basin. *Ecological Applications* 21: 1068-1084.

Smith, G.R., Badgley, C., Eiting, T.P. & Larson, P.S. 2010. Species diversity gradients in relation to geological history in North American freshwater fishes. *Evolutionary Ecology Research* 12: 693-726.

Solé, R.V. & Montoya, M. 2001. Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268: 2039-2045.

Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R. & Tarquinio, E. 2008. A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society* 27: 878-891.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K. & Norris, R.H. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16: 1267-1276.

R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Thébault, E. & Fontaine, C. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329: 853-856.

Trautman, M.B. 1981. *The fishes of Ohio*. Ohio State University Press, Columbus, OH.

Troia, M. & Gido, K. 2015. Functional strategies drive community assembly of stream fishes along environmental gradients and across spatial scales. *Oecologia* 177: 545-559.

Tylianakis, J.M., Laliberté, E., Nielsen, A. & Bascompte, J. 2010. Conservation of species interaction networks. *Biological Conservation* 143: 2270-2279.

Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. 2007. Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* 445: 202-205.

USEPA. 2002. *Research strategy: Environmental Monitoring and Assessment Program*. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

USEPA. 2007. *National Rivers and Streams Assessment: field operations manual*. U.S. Environmental Protection Agency, Washington, D.C.

USEPA. 2010. *National Rivers and Streams Assessment: quality assurance project plan*. EPA 841-B-07-007. U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Washington, D.C.

USEPA. 2013. *National Rivers and Streams Assessment 2008-2009: a collaborative survey (draft)*. U.S. Environmental Protection Agency, Washington, D.C.

USEPA. 2016. *National Rivers and Streams Assessment 2008-2009 Technical Report*. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, D.C.

USGS. 2013. *Federal standards and procedures for the National Watershed Boundary Dataset (WBD): techniques and methods*. 11-A3. U.S. Geological Survey and U.S. Department of Agriculture, Natural Resources Conservation Service, Reston, VA.

Van Sickle, J., Hawkins, C.P., Larsen, D.P. & Herlihy, A.T. 2005. A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society* 24: 178-191.

Veech, J.A. 2006. A probability-based analysis of temporal and spatial co-occurrence in grassland birds. *Journal of Biogeography* 33: 2145-2153.

Veech, J.A. 2013. A probabilistic model for analyzing species co-occurrence. *Global Ecology and Biogeography* 22: 252-260.

Veech, J.A. 2014. The pairwise approach to analysing species co-occurrence. *Journal of Biogeography* 41: 1029-1035.

Williams, R.J., Howe, A. & Hofmockel, K.S. 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* 5.

Winemiller, K.O. & Rose, K.A. 1992. Patterns of life-history diversification in North American fishes: implications for population regulation. *Canadian Journal of Fisheries and Aquatic Sciences* 49: 2196-2218.

Wydoski, R.S. & Whitney, R.R. 2003. *Inland fishes of Washington*. University of Washington Press, Seattle, WA.

Appendix A

Exemplar Graphs and Degree Distribution Plots

East Coast/Mid-Atlantic

Chesapeake Bay Rivers (HUC2-02) – *Combined data (all disturbance levels)*
Chesapeake Bay Rivers (HUC2-02) – *Good condition sites only*
Chesapeake Bay Rivers (HUC2-02) – *Fair condition sites only*
Chesapeake Bay Rivers (HUC2-02) – *Poor condition sites only*
Susquehanna River (HUC4-0205) – *Combined data (all disturbance levels)*
Susquehanna River (HUC4-0205) – *Good condition sites only*
Susquehanna River (HUC4-0205) – *Fair condition sites only*
Susquehanna River (HUC4-0205) – *Poor condition sites only*

Mississippi River Basin

Ohio River (HUC2-05) – *Combined data (all disturbance levels)*
Ohio River (HUC2-05) – *Good condition sites only*
Ohio River (HUC2-05) – *Fair condition sites only*
Ohio River (HUC2-05) – *Poor condition sites only*
Kanawha River (HUC4-0505) – *Combined data (all disturbance levels)*
Kanawha River (HUC4-0505) – *Good condition sites only*
Kanawha River (HUC4-0505) – *Fair condition sites only*

Pacific Northwest

Pacific Northwest Region (HUC2-17) – *Combined data (all disturbance levels)*
Pacific Northwest Region (HUC2-17) – *Good condition sites only*
Pacific Northwest Region (HUC2-17) – *Fair condition sites only*
Pacific Northwest Region (HUC2-17) – *Poor condition sites only*
Oregon-Washington Coastal (HUC4-1710) – *Combined data (all disturbance levels)*
Oregon-Washington Coastal (HUC4-1710) – *Good condition sites only*
Oregon-Washington Coastal (HUC4-1710) – *Fair condition sites only*
Oregon-Washington Coastal (HUC4-1710) – *Poor condition sites only*

General Key

Each of the following pages illustrates basic network results for an exemplar dataset. Networks are organized by biogeographic region (East Coast, Mississippi River Basin, Pacific Northwest), spatial scale (HUC2, HUC4), and level of anthropogenic disturbance (all data combined, 'good' condition sites only, 'fair' condition sites only, and 'poor' condition sites only).

For each combination of the above factors (biogeographic region, spatial scale, and disturbance level), we present results from one exemplar dataset. These include separate results for complete networks (all nodes and links combined, regardless of the significance or weights of the links), significantly positive links only, and significantly negative links only.

Each combination includes a weighted unipartite graph at left, illustrating several network parameters. The size and color of each node depict its degree (i.e., connectivity) and module membership (colors are arbitrary, showing only mutual membership within modules), respectively. The thickness and color of each link depict edge weights (thicker lines represent larger weights) and directions (black lines = positive associations, grey lines = negative associations). Unipartite network modularity values and the number of distinct modules within the network at listed at bottom-center for each data combination. Unipartite graphs were rendered using the 'force-directed' algorithm of Fruchterman and Reingold, which attempts to minimize overlap among network links.

Cumulative degree distributions from unipartite networks are plotted for each data combination and shown at center. Key power-law model statistics (power-law α values and Kolmogorov–Smirnov (K.S.) model-fit p-values) are also shown below each of the degree distribution plots, for complete power-law models and for truncated (trunc.) power-law models. X_{min} values from truncated power-law models are shown in parentheses.

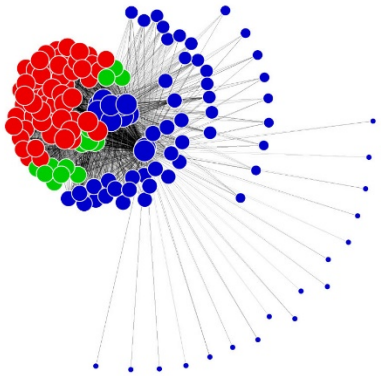
Bipartite graphs (spatially-implicit) are shown at right for each data combination. Node 'types' in the bipartite graphs are indicated by colored symbols: black squares are sampling sites, red circles are species. Links (grey lines) within the bipartite networks show which species occur at which sites. Bipartite graphs were rendered using a standard '2-mode' layout.

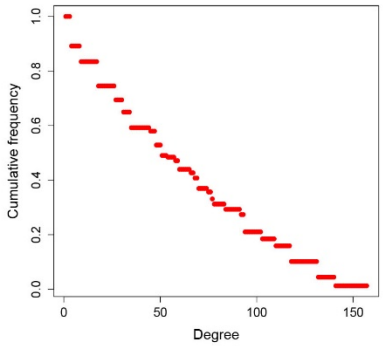
Note that in some instances (i.e., data combinations), samples sizes were insufficient to perform network analyses. These cases are noted in the following pages.

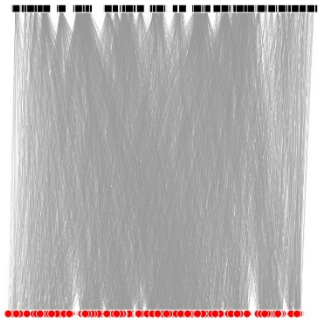
Chesapeake Bay Rivers (HUC2-02)

- Combined data (all disturbance levels) -

All links

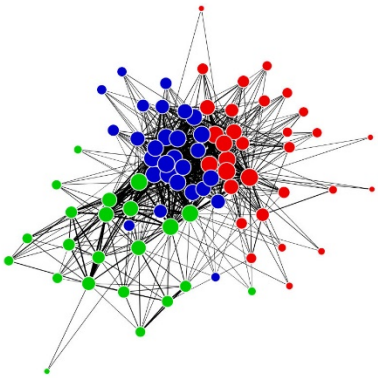


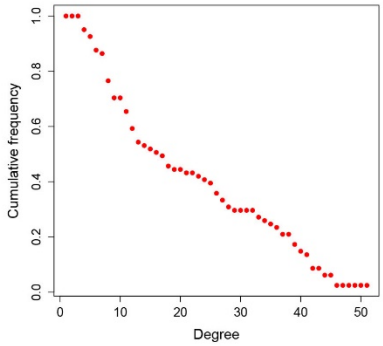


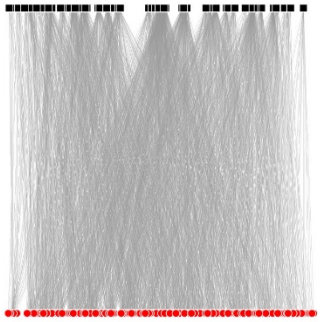


Modularity = 0.12 | Power law α = 2.00 | Trunc. (46) power law α = 2.67
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (46) power law K.S. $p < 0.01$

Significantly positive links

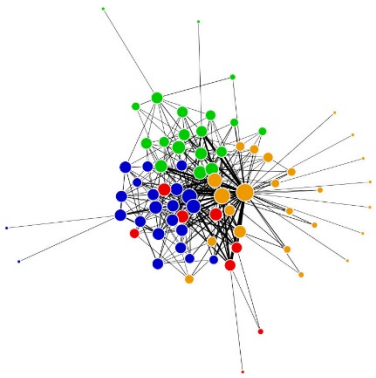


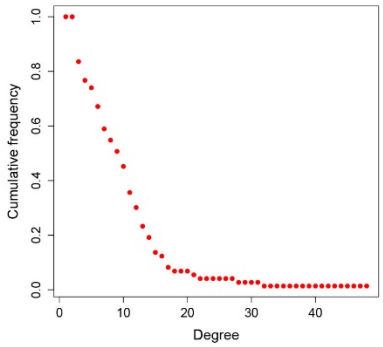


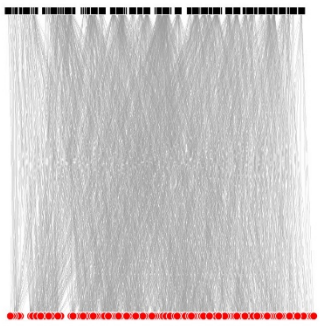


Modularity = 0.23 | Power law α = 2.00 | Trunc. (40) power law α = 12.44
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (40) power law K.S. $p = 1.00$

Significantly negative links





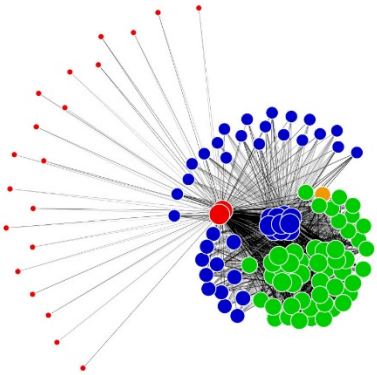


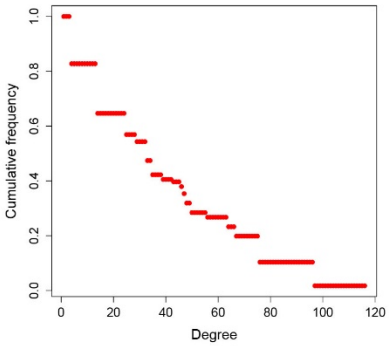
Modularity = 0.25 | Power law α = 2.00 | Trunc. (11) power law α = 3.78
Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (11) power law K.S. $p = 1.00$

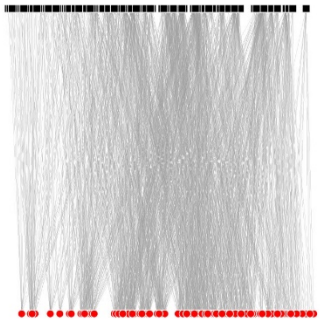
Chesapeake Bay Rivers (HUC2-02)

- GOOD condition sites -

All links

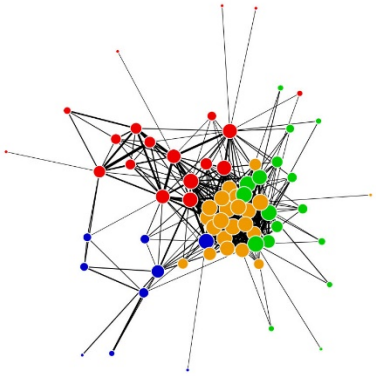


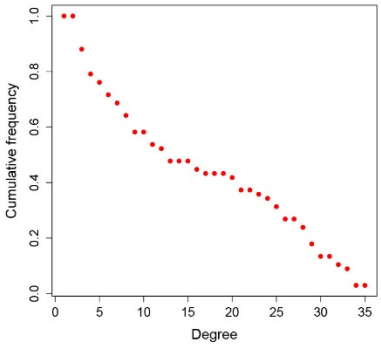


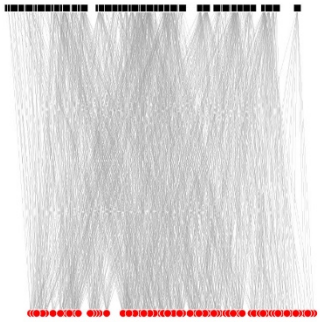


Modularity = 0.14 | Power law α = 2.00 | Trunc. (31) power law α = 2.68
Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (31) power law K.S. $p = 0.03$

Significantly positive links

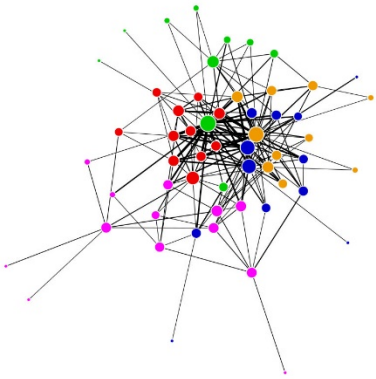


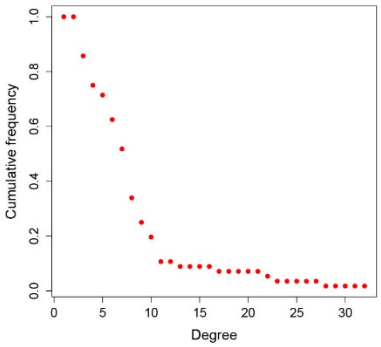


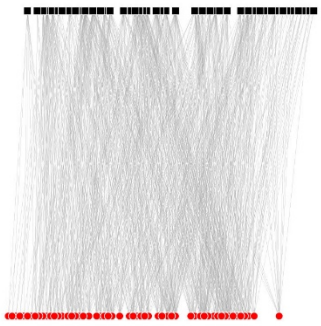


Modularity = 0.21 | Power law α = 1.37 | Trunc. (26) power law α = 8.10
Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (26) power law K.S. $p = 0.85$

Significantly negative links





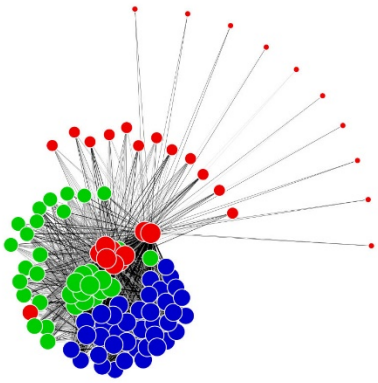


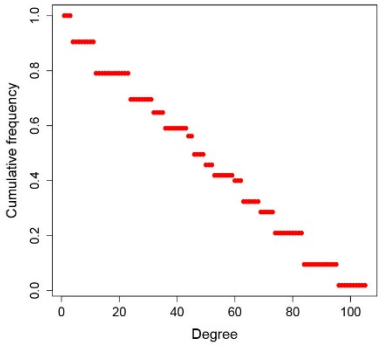
Modularity = 0.26 | Power law α = 2.00 | Trunc. (6) power law α = 3.22
Modules = 5 | Power law K.S. $p < 0.01$ | Trunc. (6) power law K.S. $p = 1.00$

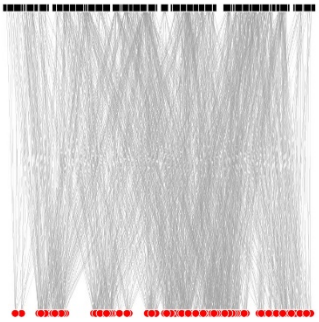
Chesapeake Bay Rivers (HUC2-02)

- FAIR condition sites -

All links



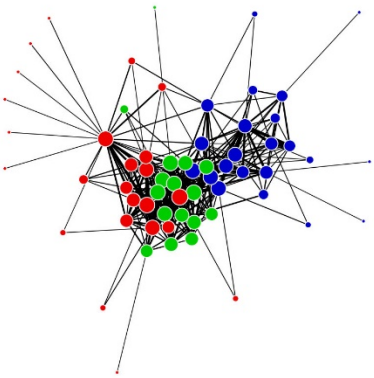


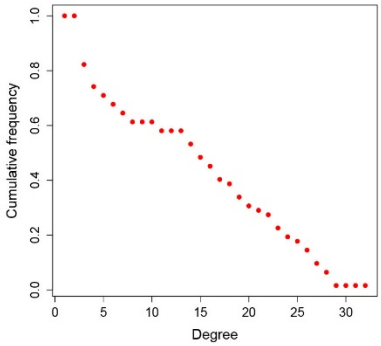


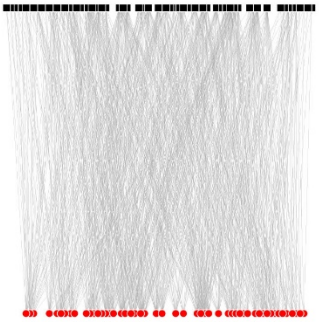
Modularity = 0.10 | Power law α = 2.00 | Trunc. (58) power law α = 4.58

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (58) power law K.S. $p = 0.02$

Significantly positive links







Modularity = 0.25 | Power law α = 1.39 | Trunc. (25) power law α = 14.05

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (25) power law K.S. $p = 1.00$

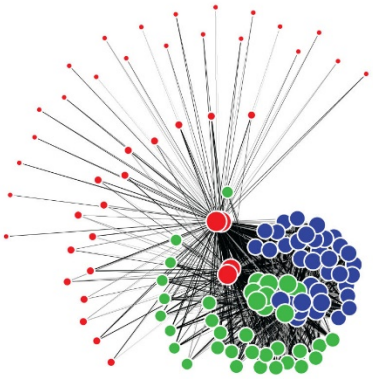
Significantly negative links

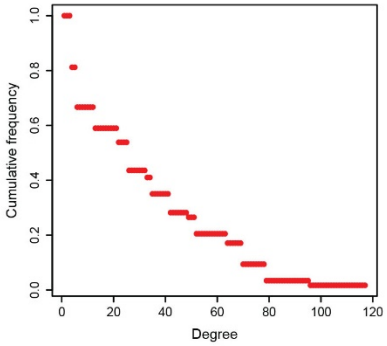
*** insufficient data to complete network analyses ***

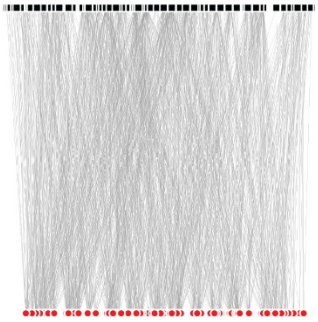
Chesapeake Bay Rivers (HUC2-02)

- POOR condition sites -

All links



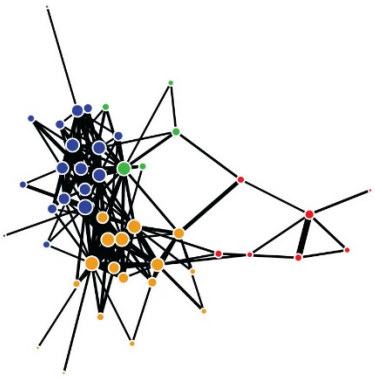


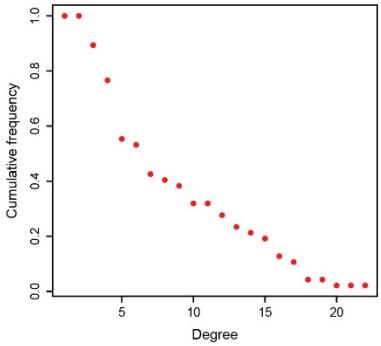


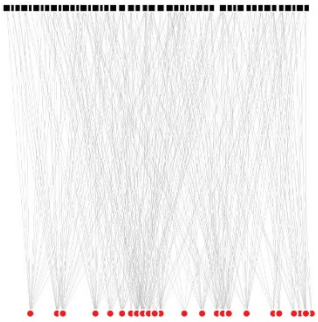
Modularity = 0.14 | Power law α = 2.00 | Trunc. (68) power law α = 8.32

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (68) power law K.S. $p = 0.76$

Significantly positive links







Modularity = 0.34 | Power law α = 1.98 | Trunc. (16) power law α = 8.53

Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (16) power law K.S. $p = 1.00$

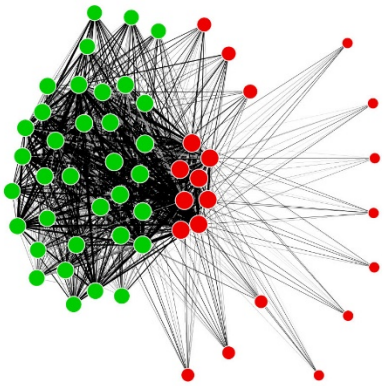
Significantly negative links

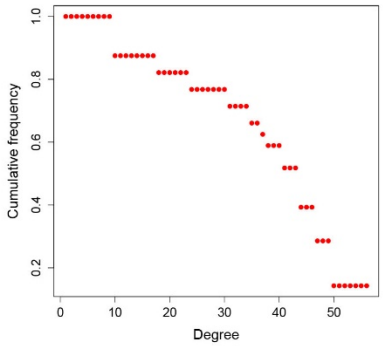
*** insufficient data to complete network analyses ***

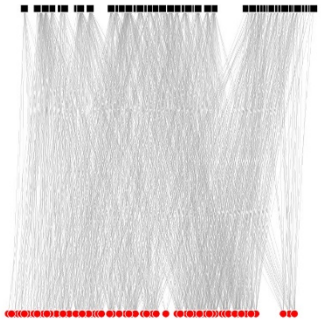
Susquehanna River (HUC4-0205)

- Combined data (all disturbance levels) -

All links

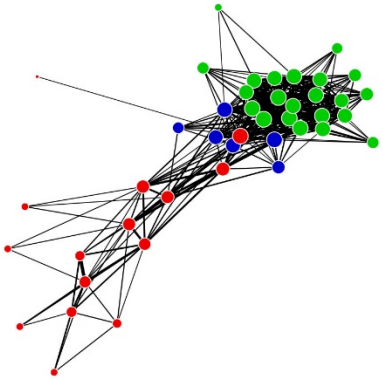


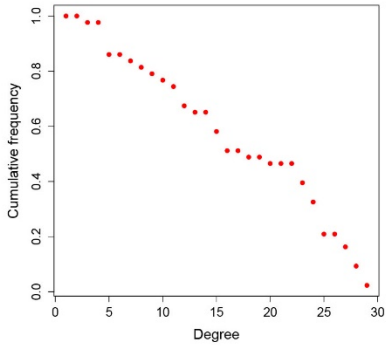


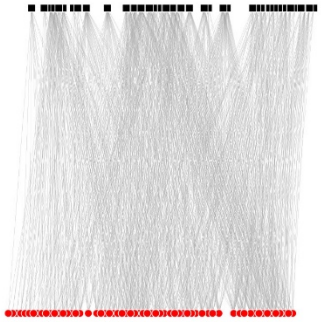


Modularity = 0.08 | Power law α = 2.00 | Trunc. (42) power law α = 8.29
Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (42) power law K.S. $p = 0.32$

Significantly positive links

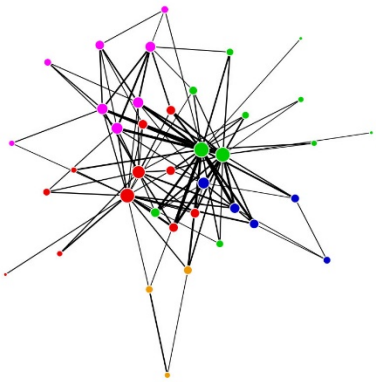


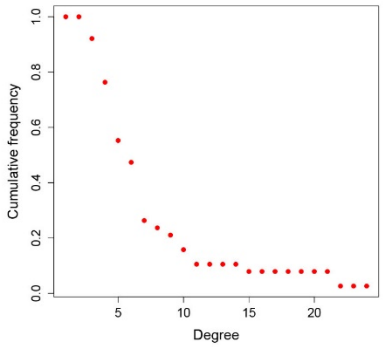


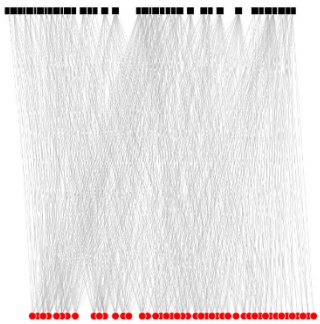


Modularity = 0.19 | Power law α = 1.98 | Trunc. (21) power law α = 7.17
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (21) power law K.S. $p = 0.74$

Significantly negative links





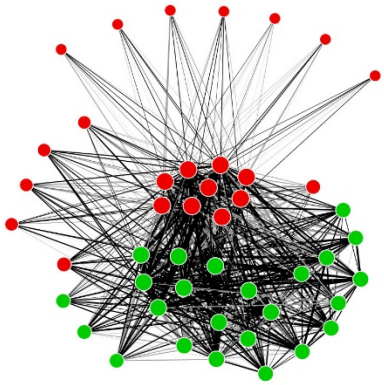


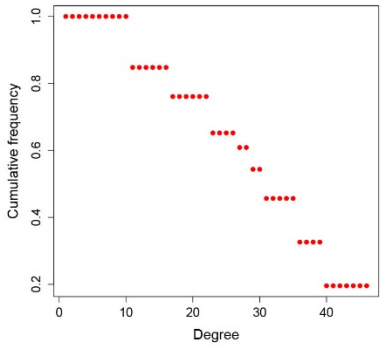
Modularity = 0.23 | Power law α = 1.97 | Trunc. (5) power law α = 2.77
Modules = 5 | Power law K.S. $p = 1.00$ | Trunc. (5) power law K.S. $p = 0.90$

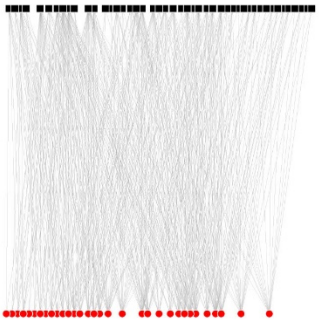
Susquehanna River (HUC4-0205)

- GOOD condition sites -

All links

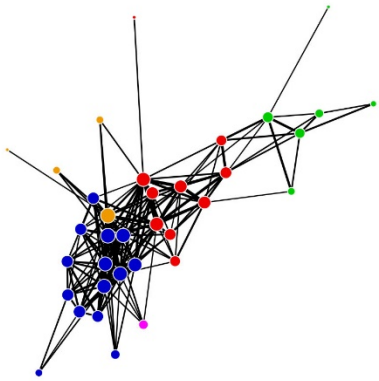


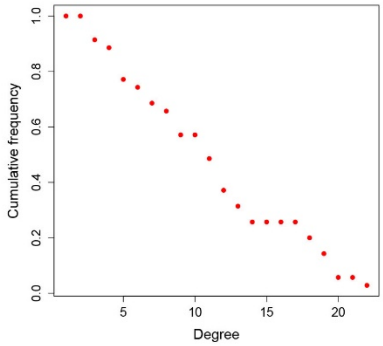


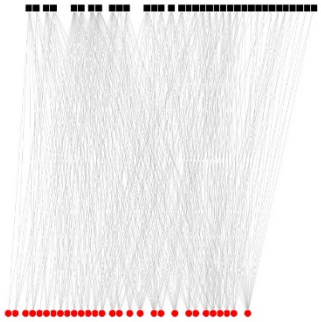


Modularity = 0.09 | Power law α = 1.97 | Trunc. (21) power law α = 3.06
Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (21) power law K.S. $p = 0.03$

Significantly positive links

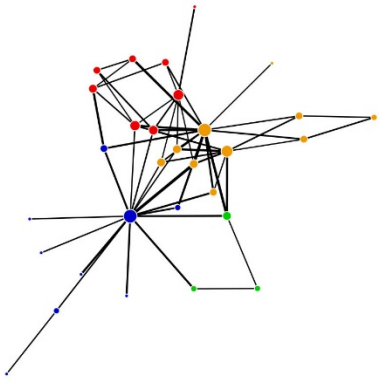


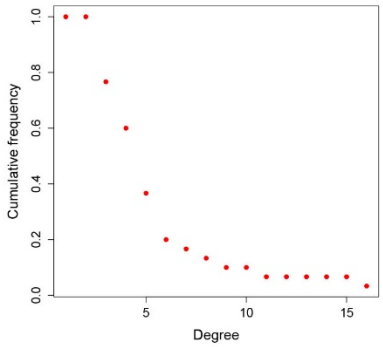


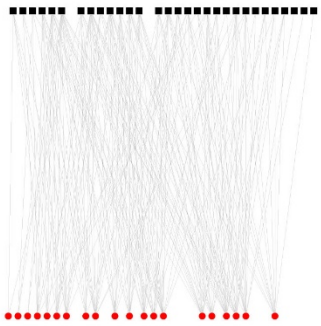


Modularity = 0.25 | Power law α = 1.39 | Trunc. (17) power law α = 9.18
Modules = 5 | Power law K.S. $p < 0.01$ | Trunc. (17) power law K.S. $p = 1.00$

Significantly negative links





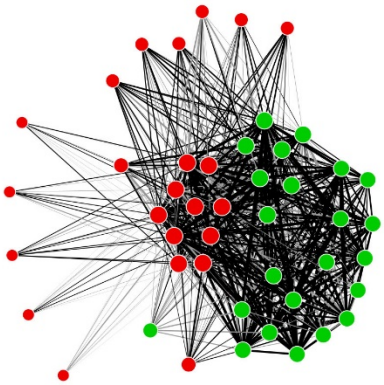


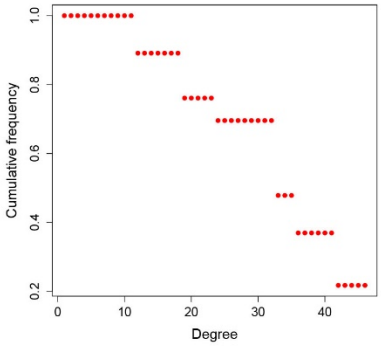
Modularity = 0.30 | Power law α = 1.97 | Trunc. (3) power law α = 2.61
Modules = 4 | Power law K.S. $p = 1.00$ | Trunc. (3) power law K.S. $p = 1.00$

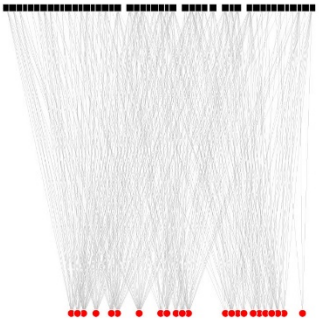
Susquehanna River (HUC4-0205)

- FAIR condition sites -

All links



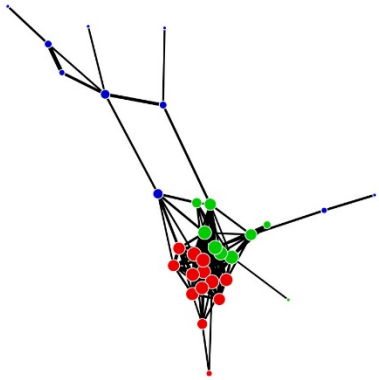


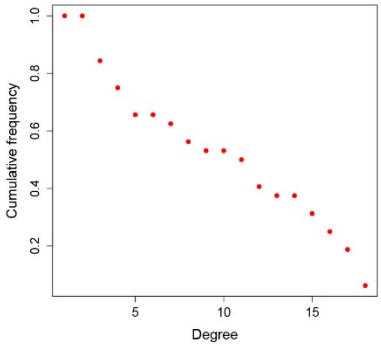


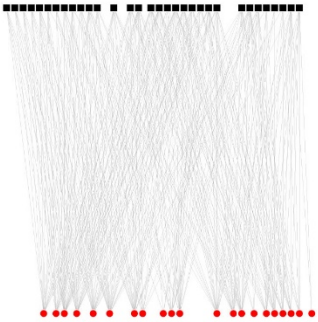
Modularity = 0.08 | Power law α = 1.98 | Trunc. (31) power law α = 5.79

Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (31) power law K.S. $p = 0.03$

Significantly positive links







Modularity = 0.19 | Power law α = 1.97 | Trunc. (13) power law α = 5.92

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (13) power law K.S. $p = 0.64$

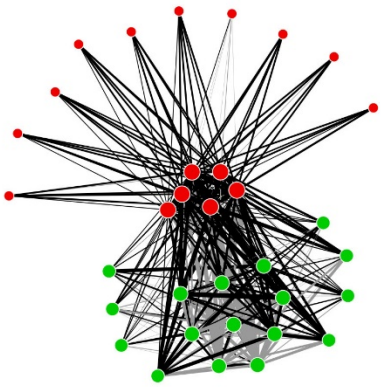
Significantly negative links

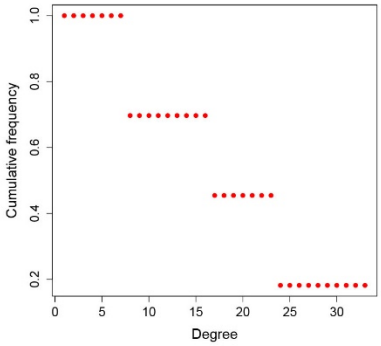
*** insufficient data to complete network analyses ***

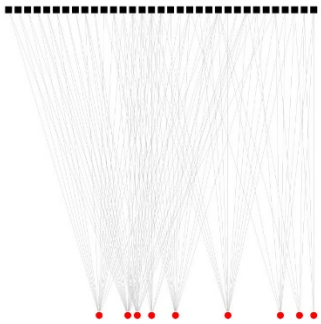
Susquehanna River (HUC4-0205)

- POOR condition sites -

All links







Modularity = 0.14 | Power law α = 1.97 | Trunc. (15) power law α = 3.52

Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (15) power law K.S. $p = 0.03$

Significantly positive links

*** insufficient data to complete network analyses ***

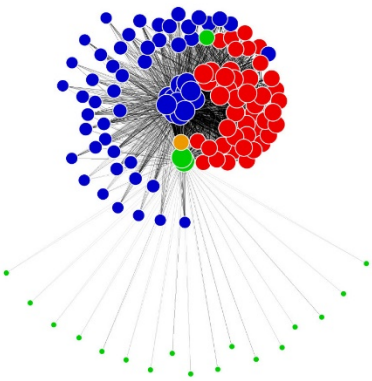
Significantly negative links

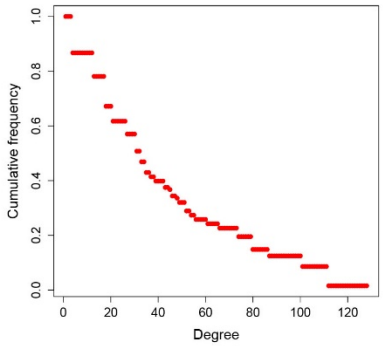
*** insufficient data to complete network analyses ***

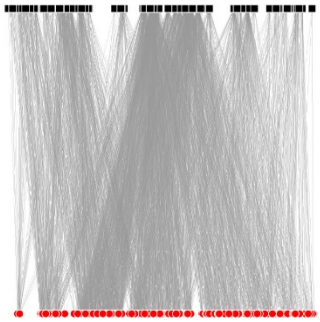
Ohio River (HUC2-05)

- Combined data (all disturbance levels) -

All links

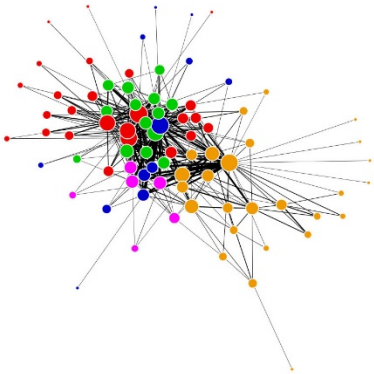


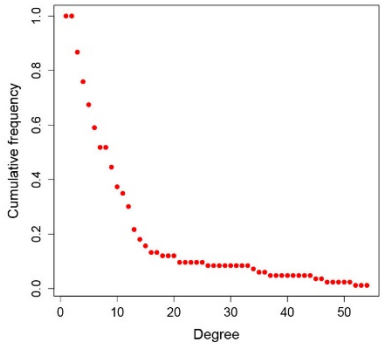


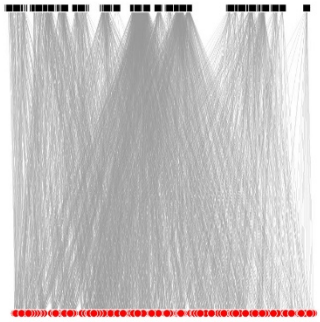


Modularity = 0.13 | Power law α = 2.00 | Trunc. (25) power law α = 2.32
Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (25) power law K.S. $p = 0.14$

Significantly positive links

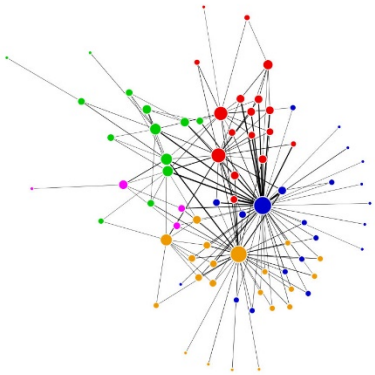


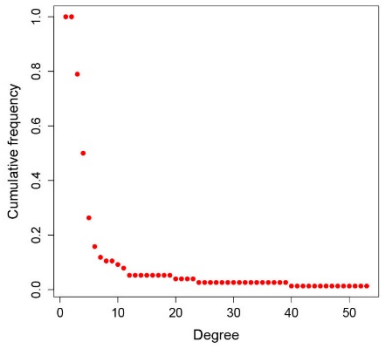


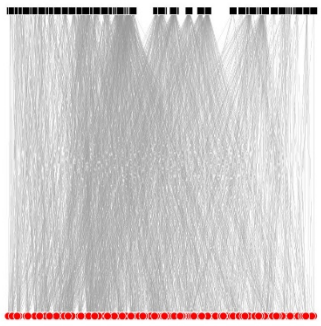


Modularity = 0.23 | Power law α = 2.00 | Trunc. (8) power law α = 2.55
Modules = 5 | Power law K.S. $p < 0.01$ | Trunc. (8) power law K.S. $p = 0.93$

Significantly negative links





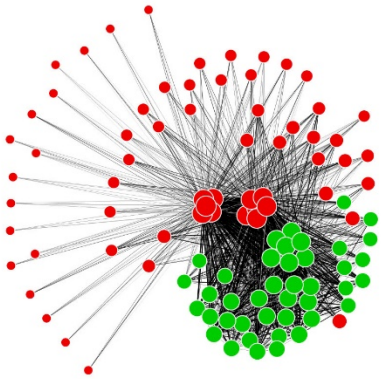


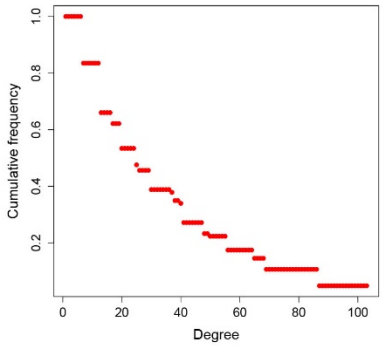
Modularity = 0.34 | Power law α = 2.00 | Trunc. (3) power law α = 2.57
Modules = 5 | Power law K.S. $p < 0.01$ | Trunc. (3) power law K.S. $p = 0.93$

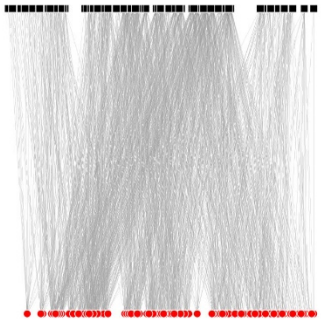
Ohio River (HUC2-05)

- GOOD condition sites -

All links

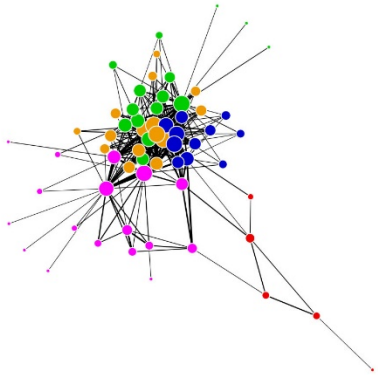


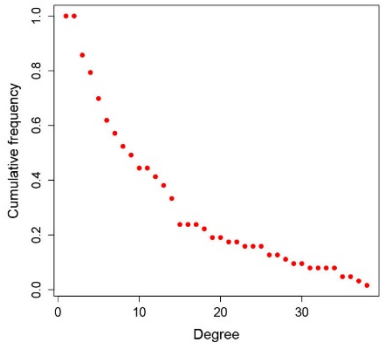


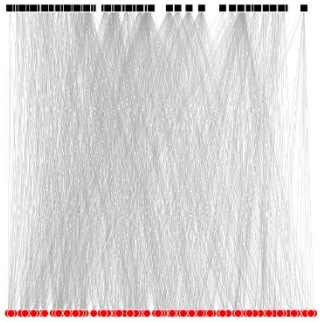


Modularity = 0.16 | Power law α = 2.00 | Trunc. (35) power law α = 2.99
Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (35) power law K.S. $p = 0.28$

Significantly positive links

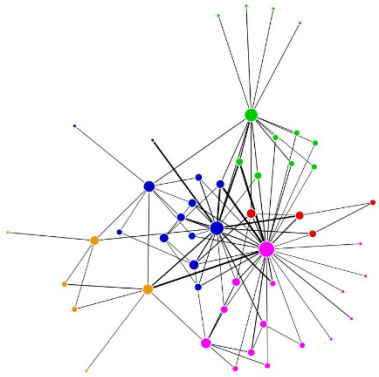


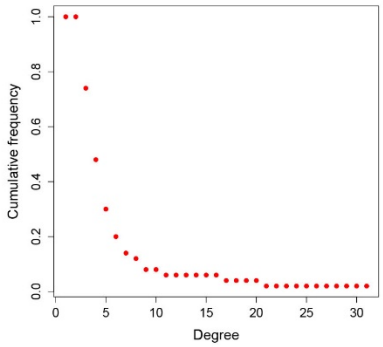


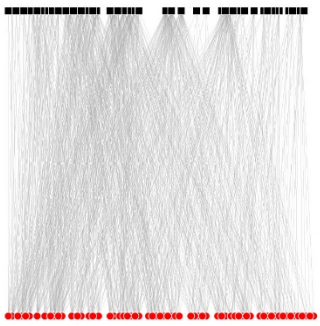


Modularity = 0.19 | Power law α = 1.83 | Trunc. (10) power law α = 2.59
Modules = 5 | Power law K.S. $p < 0.01$ | Trunc. (10) power law K.S. $p = 0.82$

Significantly negative links





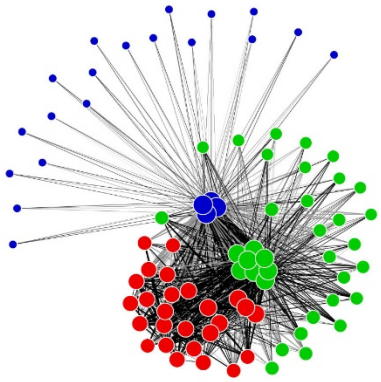


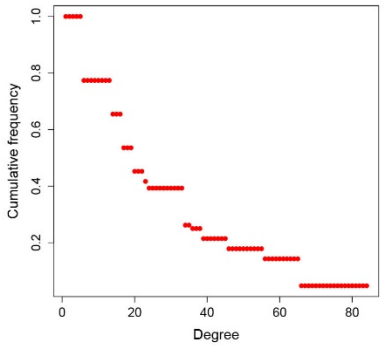
Modularity = 0.35 | Power law α = 1.70 | Trunc. (3) power law α = 2.50
Modules = 5 | Power law K.S. $p = 0.01$ | Trunc. (3) power law K.S. $p = 1.00$

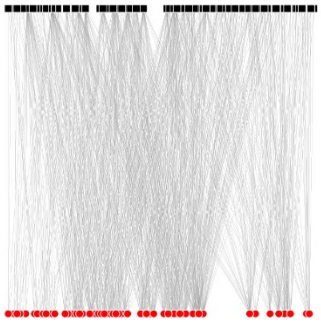
Ohio River (HUC2-05)

- FAIR condition sites -

All links

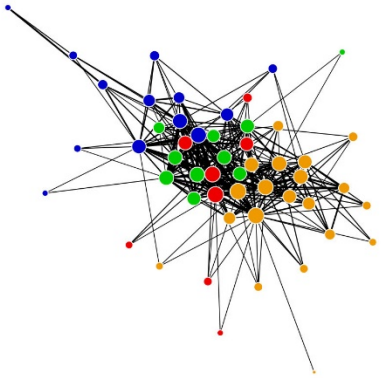


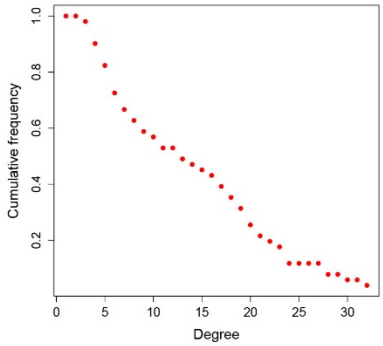


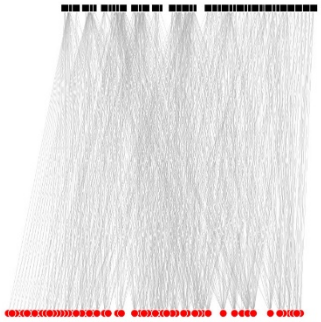


Modularity = 0.16 | Power law α = 2.00 | Trunc. (32) power law α = 3.56
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (32) power law K.S. $p = 0.15$

Significantly positive links

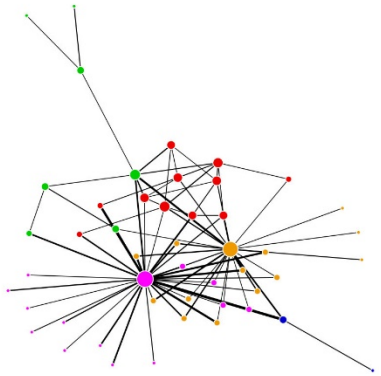


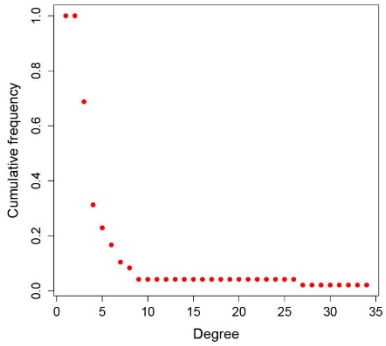


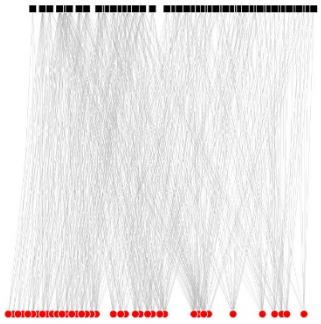


Modularity = 0.21 | Power law α = 1.37 | Trunc. (17) power law α = 4.48
Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (17) power law K.S. $p = 0.98$

Significantly negative links





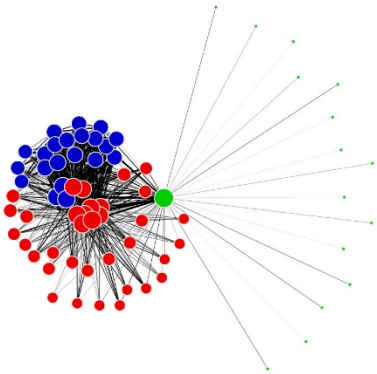


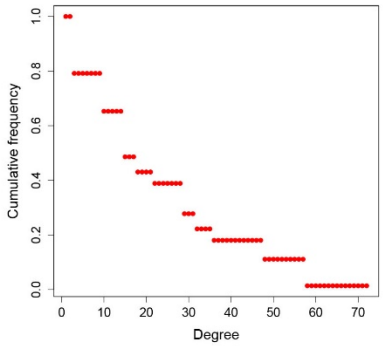
Modularity = 0.32 | Power law α = 1.78 | Trunc. (2) power law α = 2.40
Modules = 5 | Power law K.S. $p = 0.02$ | Trunc. (2) power law K.S. $p = 1.00$

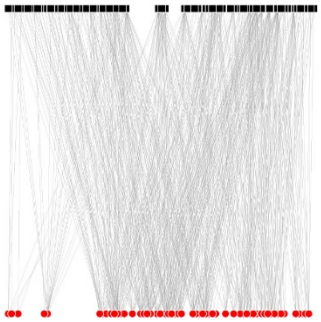
Ohio River (HUC2-05)

- POOR condition sites -

All links

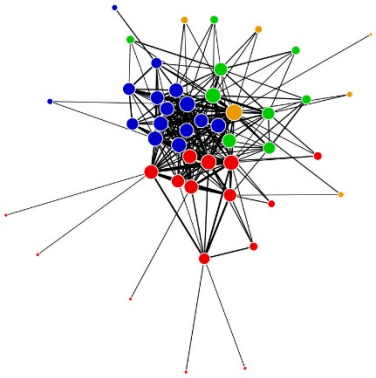


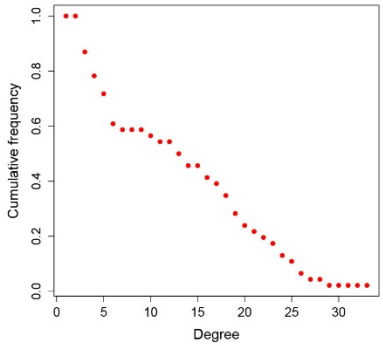


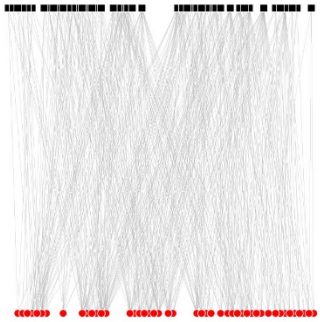


Modularity = 0.14 | Power law α = 1.34 | Trunc. (8) power law α = 1.97
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (8) power law K.S. $p = 0.01$

Significantly positive links

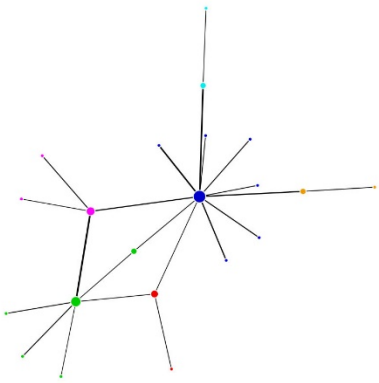


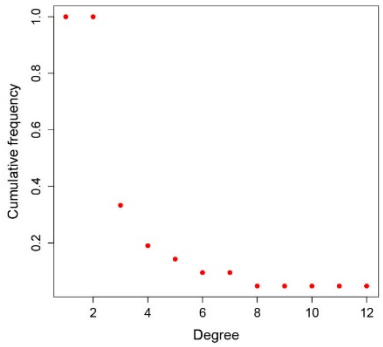


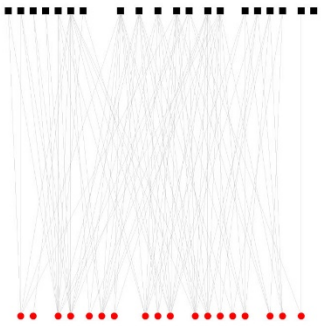


Modularity = 0.18 | Power law α = 1.39 | Trunc. (22) power law α = 7.37
Modules = 4 | Power law K.S. $p = 0.01$ | Trunc. (22) power law K.S. $p = 1.00$

Significantly negative links





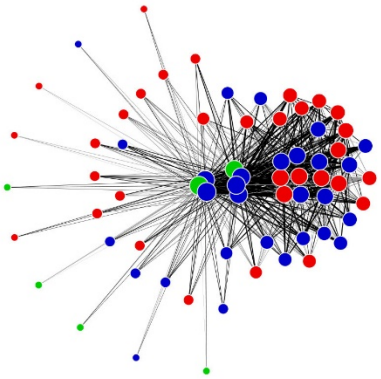


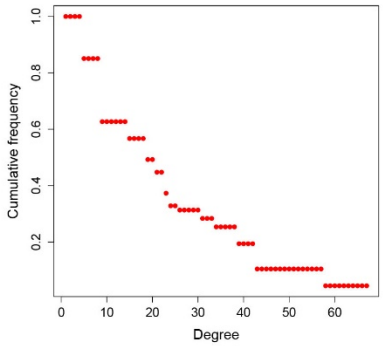
Modularity = 0.43 | Power law α = 2.16 | Trunc. (1) power law α = 2.16
Modules = 6 | Power law K.S. $p = 1.00$ | Trunc. (1) power law K.S. $p = 1.00$

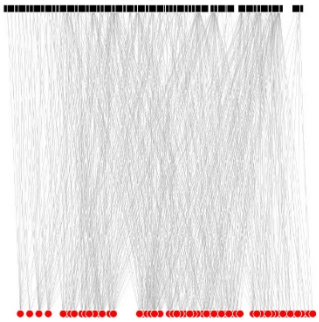
Kanawha River (HUC4-0505)

- Combined data (all disturbance levels) -

All links

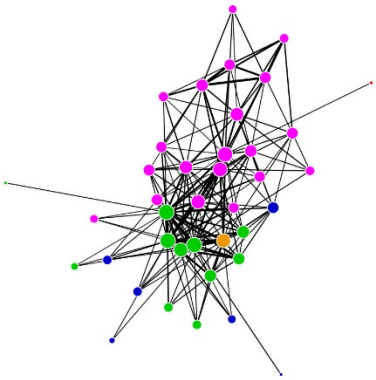


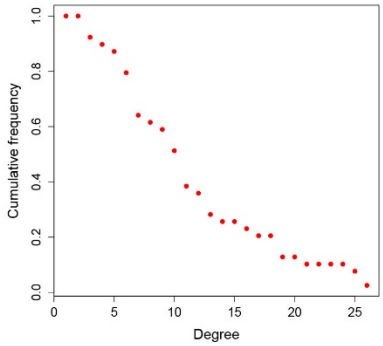


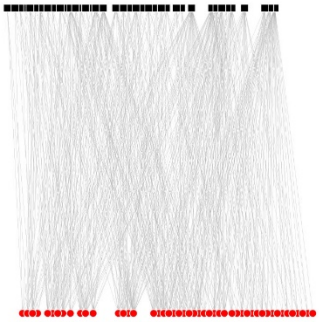


Modularity = 0.16 | Power law α = -2.00 | Trunc. (17) power law α = -2.64
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (17) power law K.S. $p = 0.19$

Significantly positive links

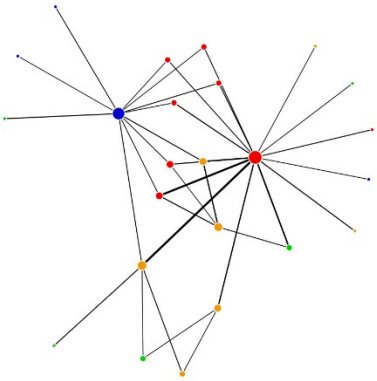


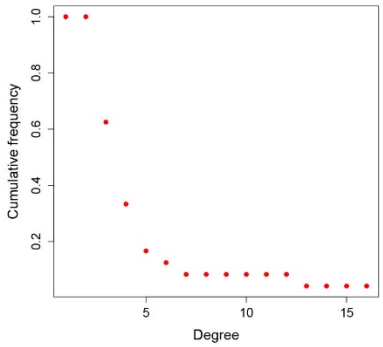


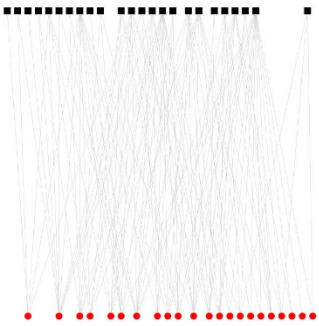


Modularity = 0.27 | Power law α = -1.85 | Trunc. (17) power law α = -9.74
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (17) power law K.S. $p = 1.00$

Significantly negative links





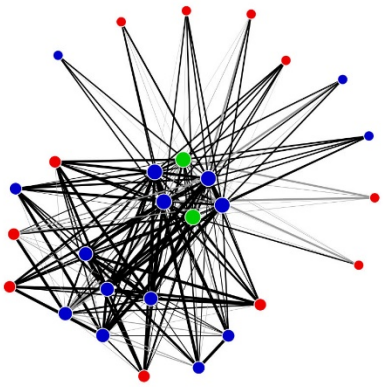


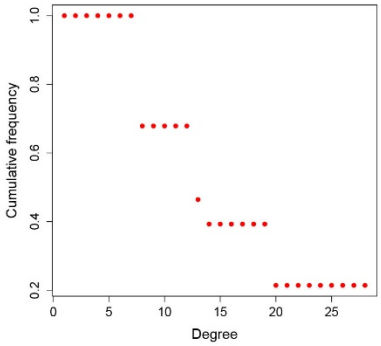
Modularity = 0.34 | Power law α = -1.93 | Trunc. (1) power law α = -1.93
Modules = 4 | Power law K.S. $p = 1.00$ | Trunc. (1) power law K.S. $p = 1.00$

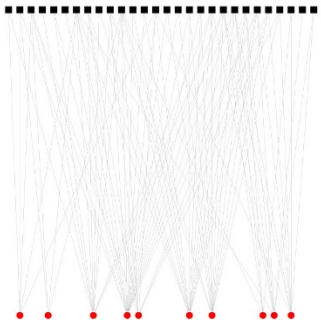
Kanawha River (HUC4-0505)

- FAIR condition sites -

All links







Modularity = 0.10 | Power law α = 1.96 | Trunc. (6) power law α = 2.24

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (6) power law K.S. $p = 0.08$

Significantly positive links

*** insufficient data to complete network analyses ***

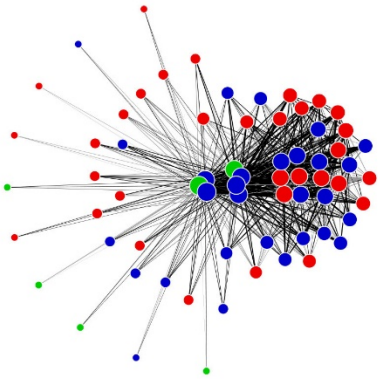
Significantly negative links

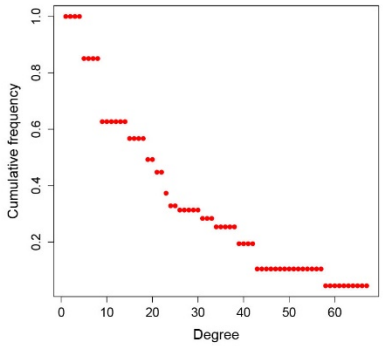
*** insufficient data to complete network analyses ***

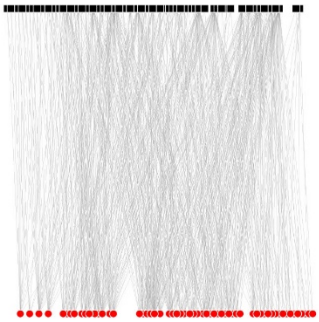
Kanawha River (HUC4-0505)

- GOOD condition sites -

All links

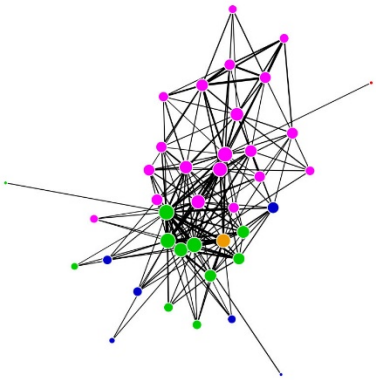


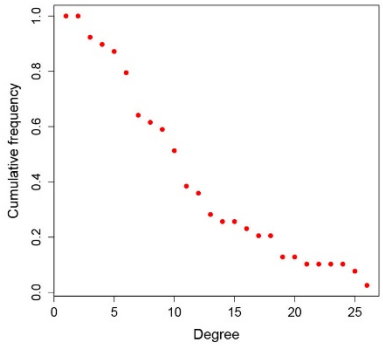


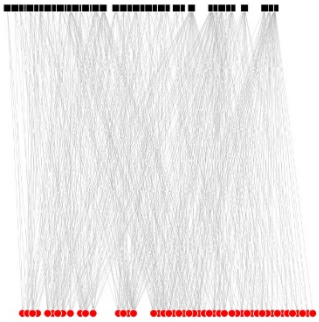


Modularity = 0.10 | Power law α = -2.00 | Trunc. (14) power law α = -2.58
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (14) power law K.S. $p = 0.16$

Significantly positive links

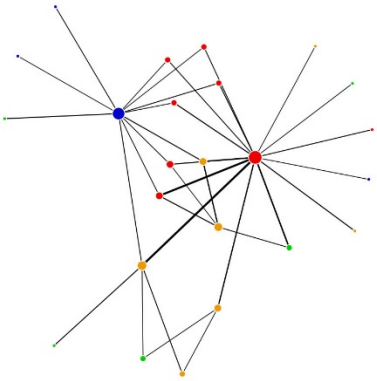


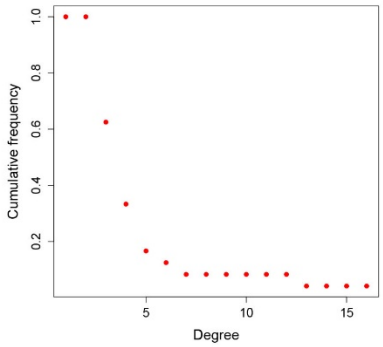


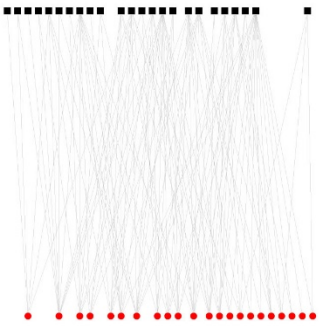


Modularity = 0.31 | Power law α = -1.97 | Trunc. (5) power law α = -2.42
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (5) power law K.S. $p = 0.93$

Significantly negative links





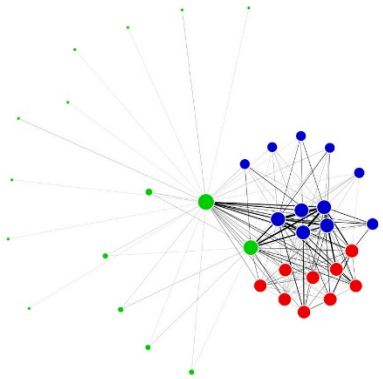


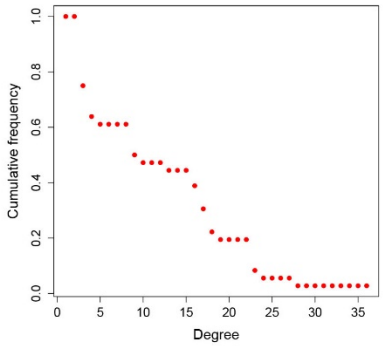
Modularity = 0.28 | Power law α = -2.29 | Trunc. (1) power law α = -2.29
Modules = 4 | Power law K.S. $p = 1.00$ | Trunc. (1) power law K.S. $p = 1.00$

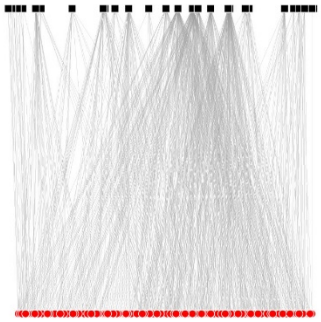
Pacific Northwest Region (HUC2-17)

- Combined data (all disturbance levels) -

All links

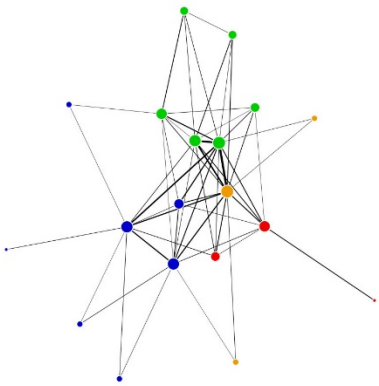


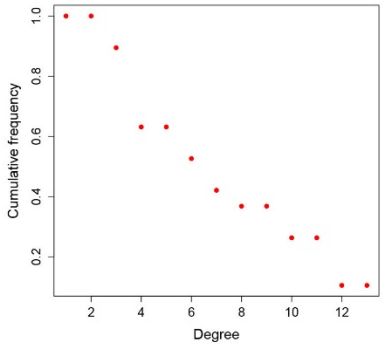


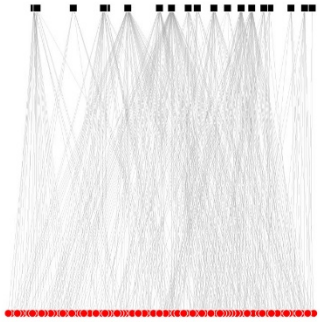


Modularity = 0.16 | Power law α = 1.97 | Trunc. (14) power law α = 4.02
Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (14) power law K.S. $p = 0.70$

Significantly positive links

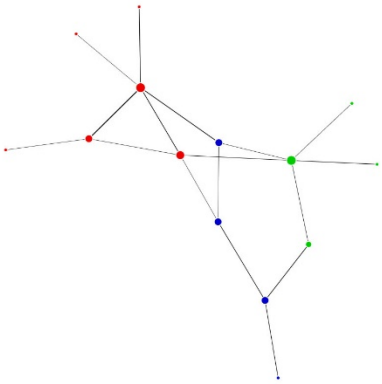


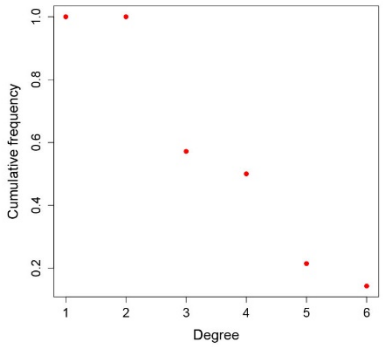


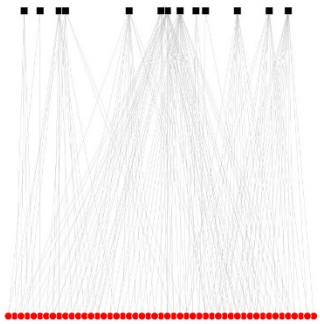


Modularity = 0.17 | Power law α = 1.95 | Trunc. (2) power law α = 1.81
Modules = 4 | Power law K.S. $p < 0.01$ | Trunc. (2) power law K.S. $p = 0.41$

Significantly negative links





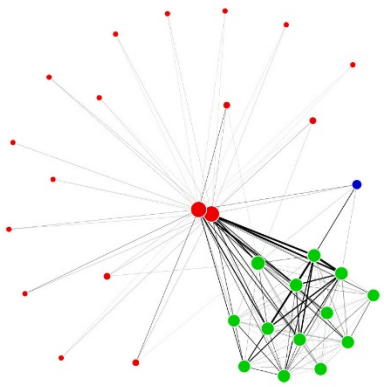


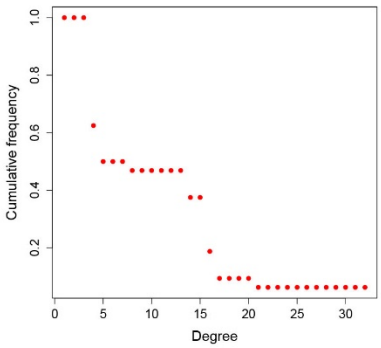
Modularity = 0.35 | Power law α = 1.81 | Trunc. (3) power law α = 3.62
Modules = 3 | Power law K.S. $p = 0.53$ | Trunc. (3) power law K.S. $p = 1.00$

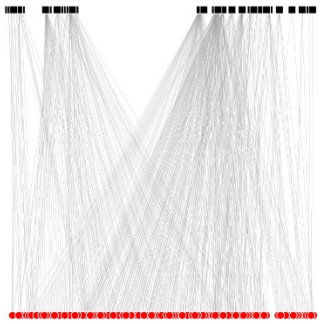
Pacific Northwest Region (HUC2-17)

- GOOD condition sites -

All links



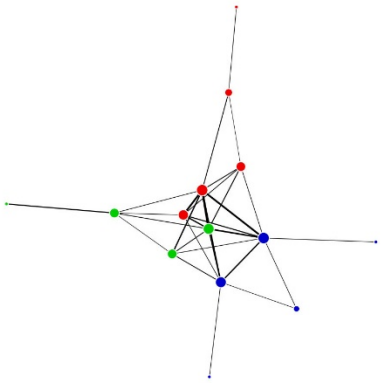


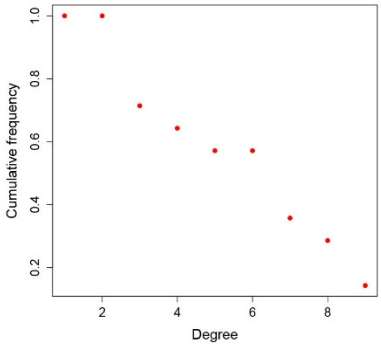


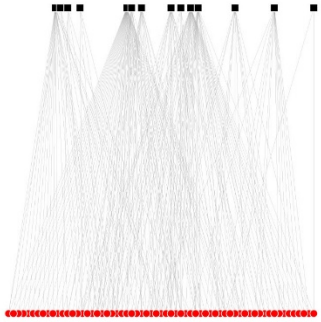
Modularity = 0.23 | Power law α = 1.97 | Trunc. (12) power law α = 4.07

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (12) power law K.S. $p = 0.53$

Significantly positive links







Modularity = 0.16 | Power law α = 1.93 | Trunc. (5) power law α = 3.77

Modules = 3 | Power law K.S. $p < 0.01$ | Trunc. (5) power law K.S. $p = 0.95$

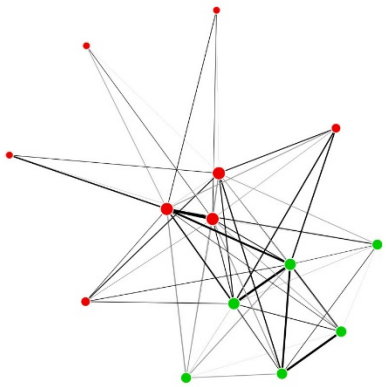
Significantly negative links

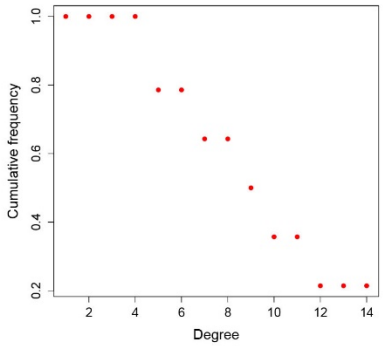
*** insufficient data to complete network analyses ***

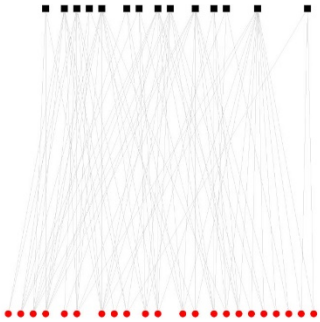
Pacific Northwest Region (HUC2-17)

- FAIR condition sites -

All links

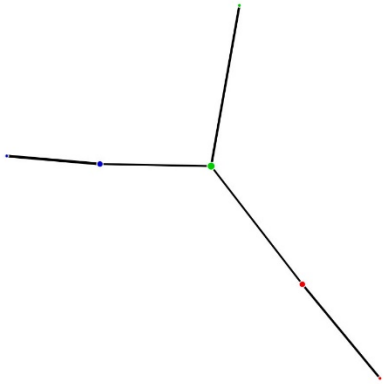


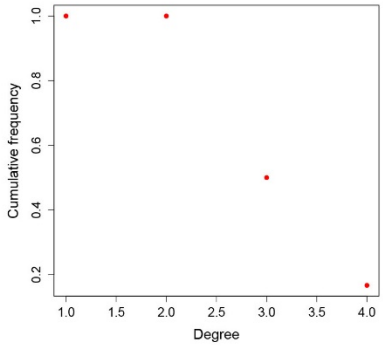


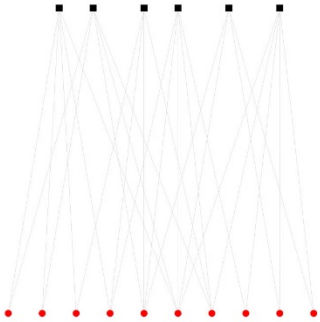


Modularity = 0.09 | Power law α = 1.67 | Trunc. (7) power law α = 3.32
Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (7) power law K.S. $p = 0.93$

Significantly positive links

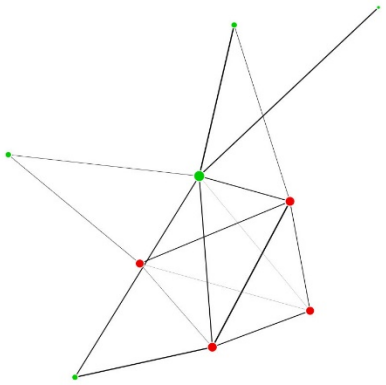


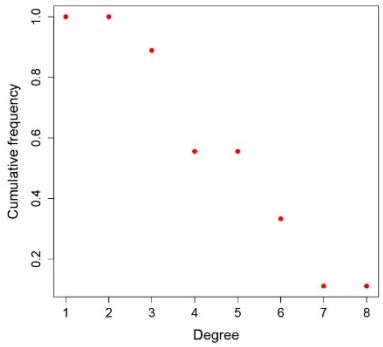


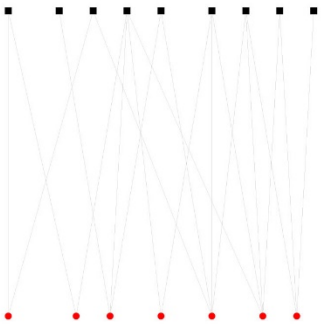


Modularity = 0.26 | Power law α = 2.02 | Trunc. (1) power law α = 2.02
Modules = 3 | Power law K.S. $p = 0.99$ | Trunc. (1) power law K.S. $p = 0.99$

Significantly negative links





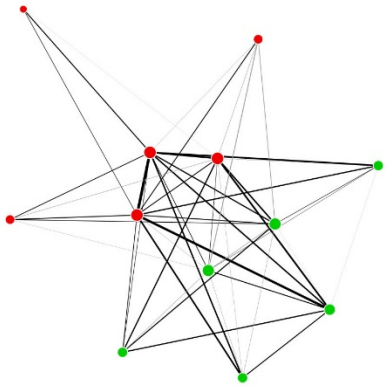


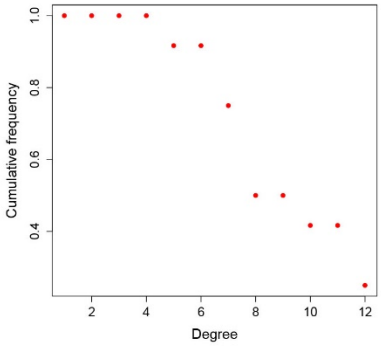
Modularity = 0.12 | Power law α = 1.89 | Trunc. (4) power law α = 3.57
Modules = 2 | Power law K.S. $p = 0.02$ | Trunc. (4) power law K.S. $p = 1.00$

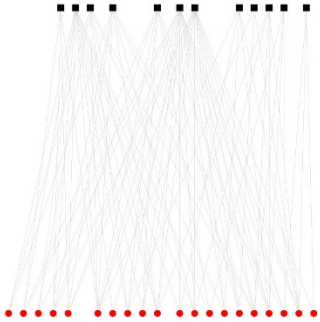
Pacific Northwest Region (HUC2-17)

- POOR condition sites -

All links







Modularity = 0.02 | Power law α = 1.37 | Trunc. (5) power law α = 2.71

Modules = 2 | Power law K.S. *p* = 0.01 | Trunc. (5) power law K.S. *p* = 0.72

Significantly positive links

*** insufficient data to complete network analyses ***

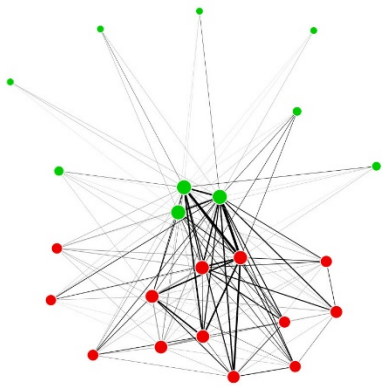
Significantly negative links

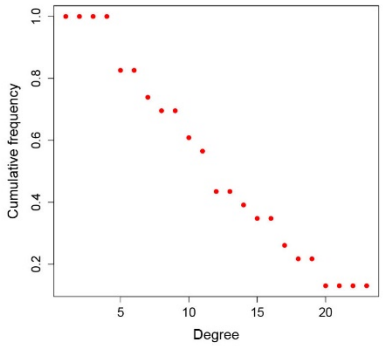
*** insufficient data to complete network analyses ***

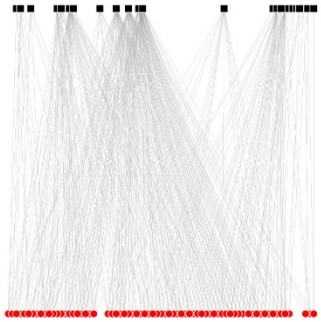
Oregon-Washington Coastal Rivers (HUC4-1710)

- Combined data (all disturbance levels) -

All links

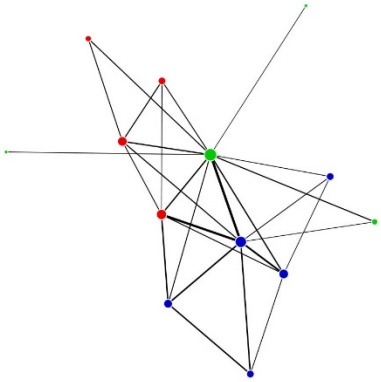


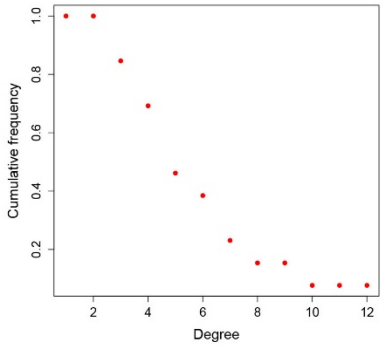


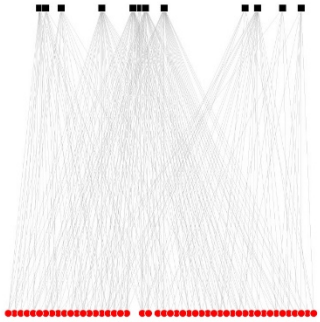


Modularity = 0.11 | Power law α = 1.35 | Trunc. (8) power law α = 2.62
Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (8) power law K.S. $p = 0.69$

Significantly positive links

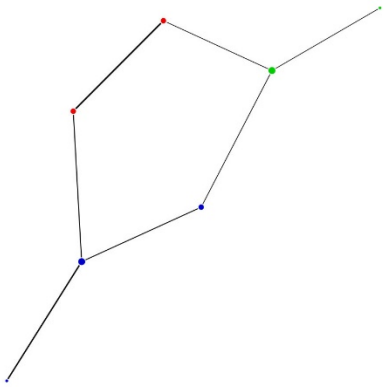


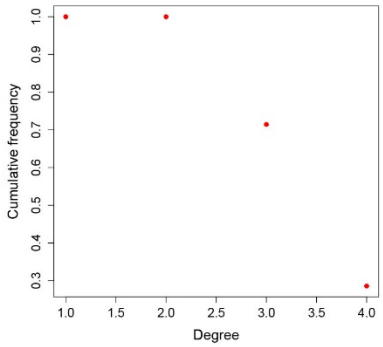


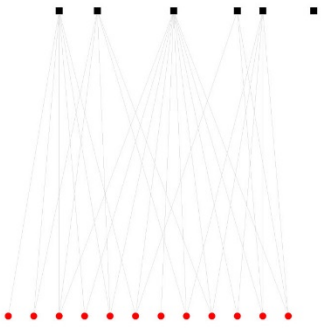


Modularity = 0.17 | Power law α = 1.92 | Trunc. (5) power law α = 3.10
Modules = 3 | Power law K.S. $p = 0.01$ | Trunc. (5) power law K.S. $p = 1.00$

Significantly negative links





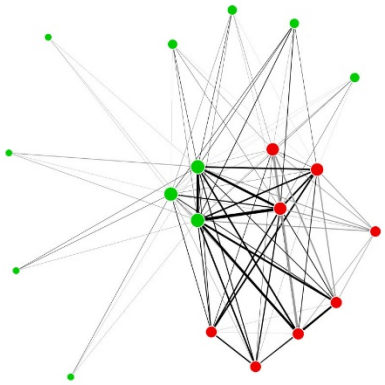


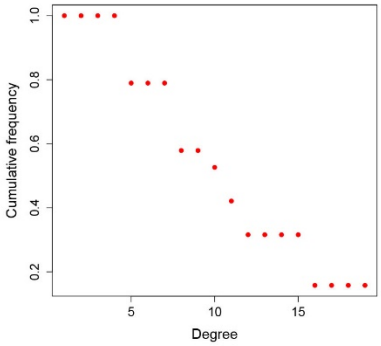
Modularity = 0.22 | Power law α = 1.82 | Trunc. (1) power law α = 1.82
Modules = 3 | Power law K.S. $p = 0.53$ | Trunc. (1) power law K.S. $p = 0.53$

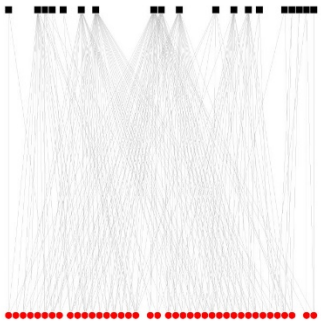
Oregon-Washington Coastal Rivers (HUC4-1710)

- GOOD condition sites -

All links



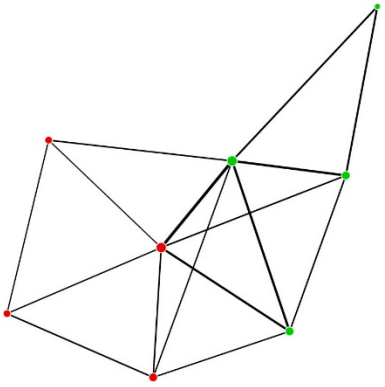


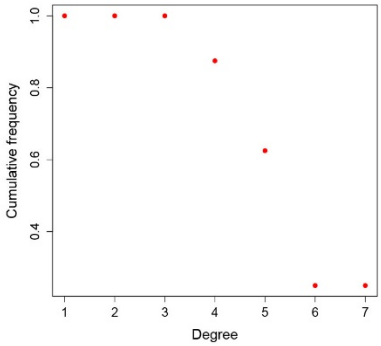


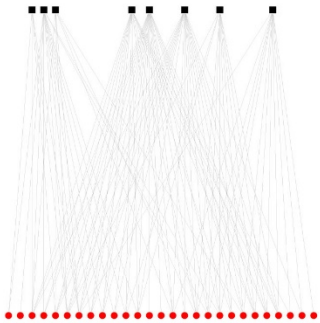
Modularity = 0.10 | Power law α = 1.37 | Trunc. (6) power law α = 2.53

Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (6) power law K.S. $p = 0.75$

Significantly positive links







Modularity = 0.13 | Power law α = 1.88 | Trunc. (3) power law α = 2.82

Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (3) power law K.S. $p = 0.90$

Significantly negative links

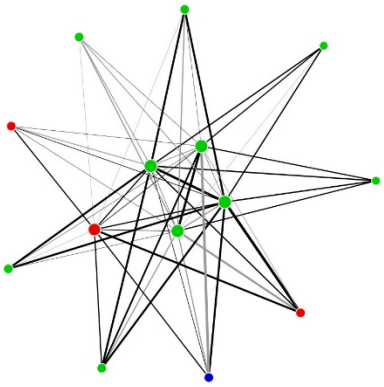
*** insufficient data to complete network analyses ***

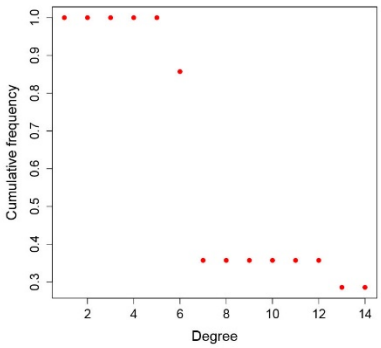
Appendix A.22

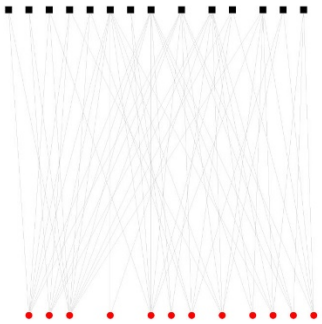
Oregon-Washington Coastal Rivers (HUC4-1710)

- FAIR condition sites -

All links







Modularity = 0.09
Modules = 3

Power law α = 1.69
Power law K.S. $p < 0.01$

Trunc. (4) power law α = 2.45
Trunc. (4) power law K.S. $p = 0.77$

Significantly positive links

*** insufficient data to complete network analyses ***

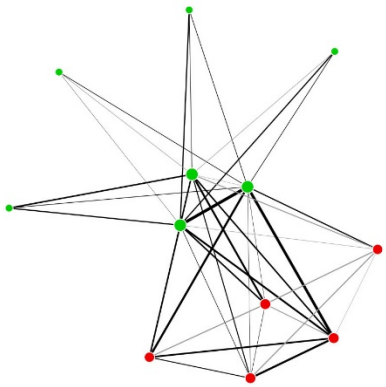
Significantly negative links

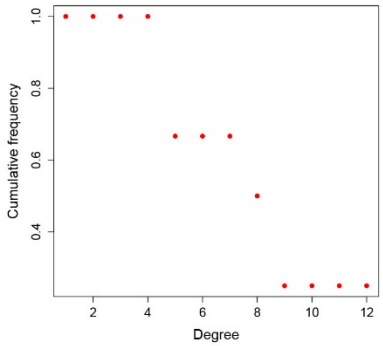
*** insufficient data to complete network analyses ***

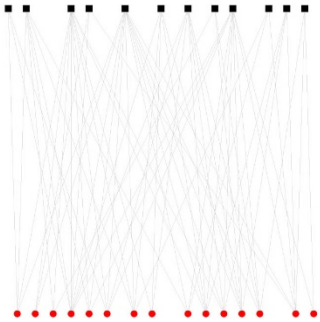
Oregon-Washington Coastal Rivers (HUC4-1710)

- POOR condition sites -

All links

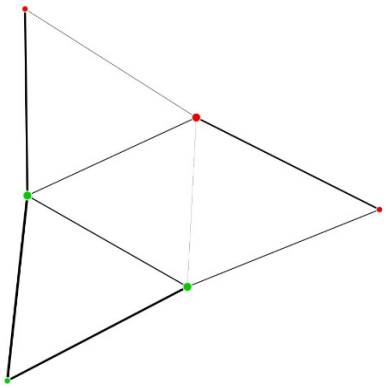


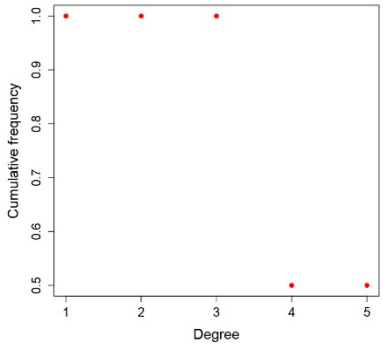


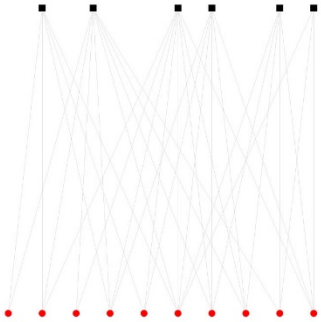


Modularity = 0.10 | Power law α = 1.92 | Trunc. (6) power law α = 3.41
Modules = 2 | Power law K.S. $p < 0.01$ | Trunc. (6) power law K.S. $p = 0.89$

Significantly positive links

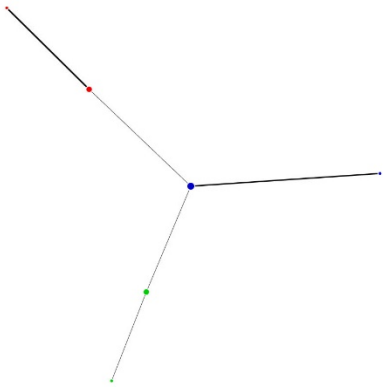


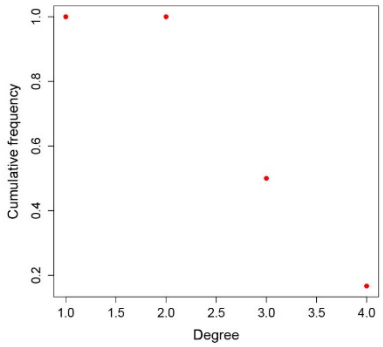


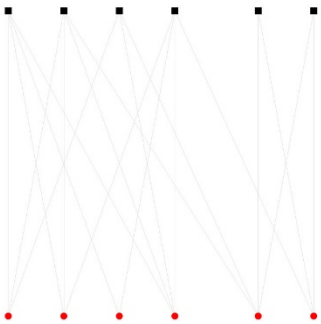


Modularity = 0.05 | Power law α = 1.83 | Trunc. (1) power law α = 1.83
Modules = 2 | Power law K.S. $p = 0.02$ | Trunc. (1) power law K.S. $p = 0.02$

Significantly negative links







Modularity = 0.26 | Power law α = 2.02 | Trunc. (1) power law α = 2.02
Modules = 3 | Power law K.S. $p = 0.99$ | Trunc. (1) power law K.S. $p = 0.99$

Appendix B

Network Metric Definitions

Node Degree – The degree of a focal node within a network is simply the number of nodes that are directly connected to it. In the present unipartite context, the degree of a node is the number of species that a focal species co-occurs with. In the bipartite context, the degree of a *site* ('node type 1') is the number of species that occur at it, while the degree of a *species* ('node type 2') is the number of sites that a species occurs at. (The same 'site vs. species' logic applies to all other network metrics that can be calculated at the individual node level within bipartite networks.) Because degree values are specific to individual nodes, they may exhibit large variation within a given network. We have therefore calculated two of the statistical moments describing populations of node degree values: ***maximum*** and ***median (50th percentile)*** *degree*.

Alpha, Power-Law Degree Distribution (Complete) – Alpha (α) is the exponent of a power-law model ($y = cx^{-\alpha}$) fit to the cumulative degree distribution of a network. In this instance, the power-law model is fit to the complete degree distribution (but see truncated power-law models below). Note that the α value in a typical power-law model of network degree distribution model is *negative*; positive α values reported here are effectively negative slopes when plotted.

Kol.Smi. P-value, Power-Law Degree Distribution (Complete) – this is the *p*-value from a Kolmogorov-Smirnov test of the null hypothesis that the observed degree distribution (for the complete dataset, in this instance) was equivalent to a hypothetical, idealized power-law degree distribution that incorporates the same α value and sample size as the observed network. Note that small *p*-values provide evidence *against* the null hypothesis of equivalent distributions. Thus, small *p*-values suggest that the observed degree distribution does *not* conform to a power-law model (i.e. the power-law model is not a good fit).

Xmin, Power-Law Degree Distribution (Truncated) – Truncating an empirical, cumulative distribution prior to fitting one or more probability distributions can help to minimize the effect of noise within the empirical data. In this particular case – fitting a power-law model – truncating the data is intended to minimize noise in the left-side of the degree distribution, where many species with low degree values (i.e., few co-occurrences) may be aggregated. Using maximum likelihood, it is sometimes possible to identify a specific point – the $\boldsymbol{X_{min}}$ degree value – in the cumulative degree distribution at which the overall trend begins to exhibit a clear power-law tendency.

Alpha, Power-Law Degree Distribution (Truncated) – This is the exponent (α) of the power-law model fit to a *truncated* degree distribution.

Kol.Smi. P-value, Power Law Degree Distribution (Truncated) – This is equivalent to the K-S *p*-value described above, but calculated for the truncated degree distribution, rather than the complete degree distribution.

Betweenness Centrality – Betweenness centrality is a measure of a node's centrality within a network. It is the sum of all 'shortest path' connections between two neighboring nodes that intersect a focal node. Nodes with high betweenness may therefore serve as 'hubs' within a network. The concept of node betweenness is similar to, but subtly different than, node degree. We calculated and recorded the ***maximum*** and ***median (50th percentile)*** *betweenness* values within each network.

Closeness Centrality – Closeness centrality is the reciprocal of the *sum* of the distances (i.e. number of shortest-path links) from a focal node to all other nodes within a network. Large closeness values indicate that a focal node is directly connected to many or most other nodes in a given network. We calculated and recorded the ***maximum*** and ***median (50th percentile)*** *closeness* values within each network.

Local Clustering Coefficient – The local clustering coefficient is calculated by first identifying all of the neighbors of a given focal node (i.e. nodes that are directly connected to the focal node), then observing the number of direct links among those neighbors. Specifically, it is calculated by dividing the number of observed neighbor connections by the *total possible* number of neighbor connections. The local clustering coefficient ranges from 0−1and is calculated for every node in a given network. We calculated and recorded the ***maximum*** and ***median (50th percentile)*** *local clustering coefficients* within each network.

Global Clustering Coefficient – The global clustering coefficient is the number of 'closed triplets' (i.e. three nodes that are fully connected amongst each other, forming a closed triangle) in a network divided by the total number of closed *and* open triplets (i.e. pairs of nodes that are connected to a common node but not to each other). It is calculated as a single, average clustering coefficient value for an entire network and is also known as 'transitivity'. As a ratio, the global clustering coefficient is constrained to a range of 0−1, with values closer to 1 indicating dense, highly interconnected networks.

Modularity – Modularity reflects the tendency for a network consist of highly interconnected subcomponents – *modules* – that are only sparsely connected to other modules. Large modularity values indicate that two or more distinct modules are more likely to be found within a given network, but do not necessarily reflect the number of distinct modules that may exist.

No. Modules – This is the number of distinct network modules that was detected using a simulated annealing algorithm.

Appendix C

Unipartite Network Summary Statistics

Unipartite Networks – *Species' Degree, Maximum*
Unipartite Networks – *Species' Degree, 50th Percentile*
Unipartite Networks – *Alpha, Power Law Degree Distribution (Complete)*
Unipartite Networks – *Kol.Smi. P-value, Power Law Degree Distribution (Complete)*
Unipartite Networks – *Xmin, Power Law Degree Distribution (Truncated)*
Unipartite Networks – *Alpha, Power Law Degree Distribution (Truncated)*
Unipartite Networks – *Kol.Smi. P-value, Power Law Degree Distribution (Truncated)*
Unipartite Networks – *Betweenness Centrality, Maximum*
Unipartite Networks – *Betweenness Centrality, 50th Percentile*
Unipartite Networks – *Closeness Centrality, Maximum*
Unipartite Networks – *Closeness Centrality, 50th Percentile*
Unipartite Networks – *Global Clustering Coefficient*
Unipartite Networks – *Local Clustering Coefficient, Maximum*
Unipartite Networks – *Local Clustering Coefficient, 50th Percentile*
Unipartite Networks – *Modularity*
Unipartite Networks – *No. Modules*

General Key

Each of the following pages uses a common motif to illustrate a single network summary statistic, calculated and plotted in a single chart for *all* networks. For each of the summary statistics, we show the mean average, calculated across replicates (when more than one replicate was available) for each combination of data, visualized as a bar chart. Results are organized by three levels.

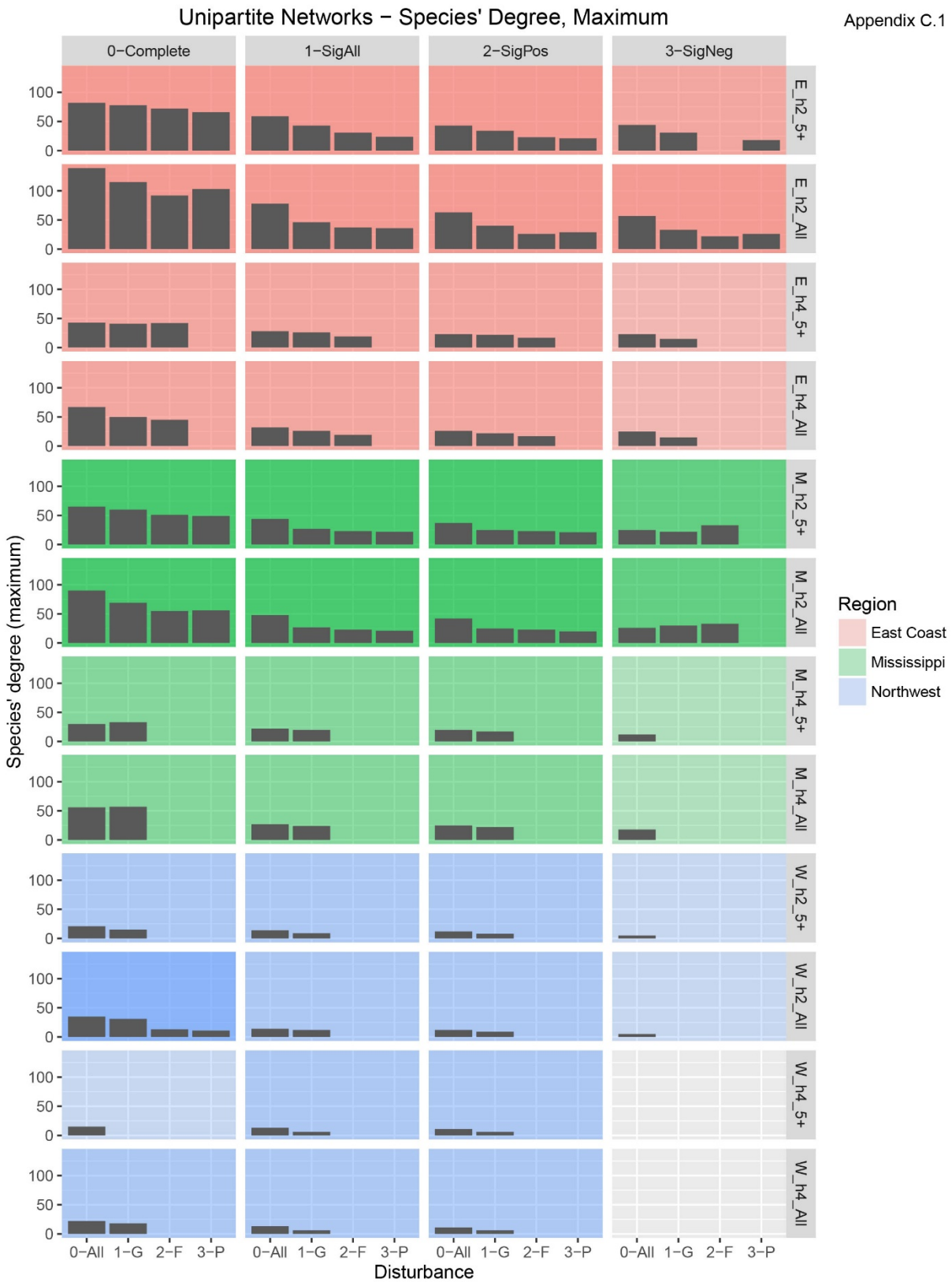
Each biogeographic region (East Coast, Mississippi River Basin, Pacific Northwest) is indicated by a background color, as shown in the 'Region' legend at right.

Within biogeographic regions, network statistics are partitioned by spatial scale (HUC2 vs. HUC4, indicated by 'h2' and 'h4' labels along the right-hand margin. Results from networks that included all species within a given data combination are also distinguished from results that used the conservative 'five or more occurrences per species' criterion (see main text) are also distinguished in the right-hand margin by 'All' and '5+' labels. The 'E', 'M', and 'W' labels in the right-hand margin are biogeographic region indicators (i.e. are redundant with the grouping by color).

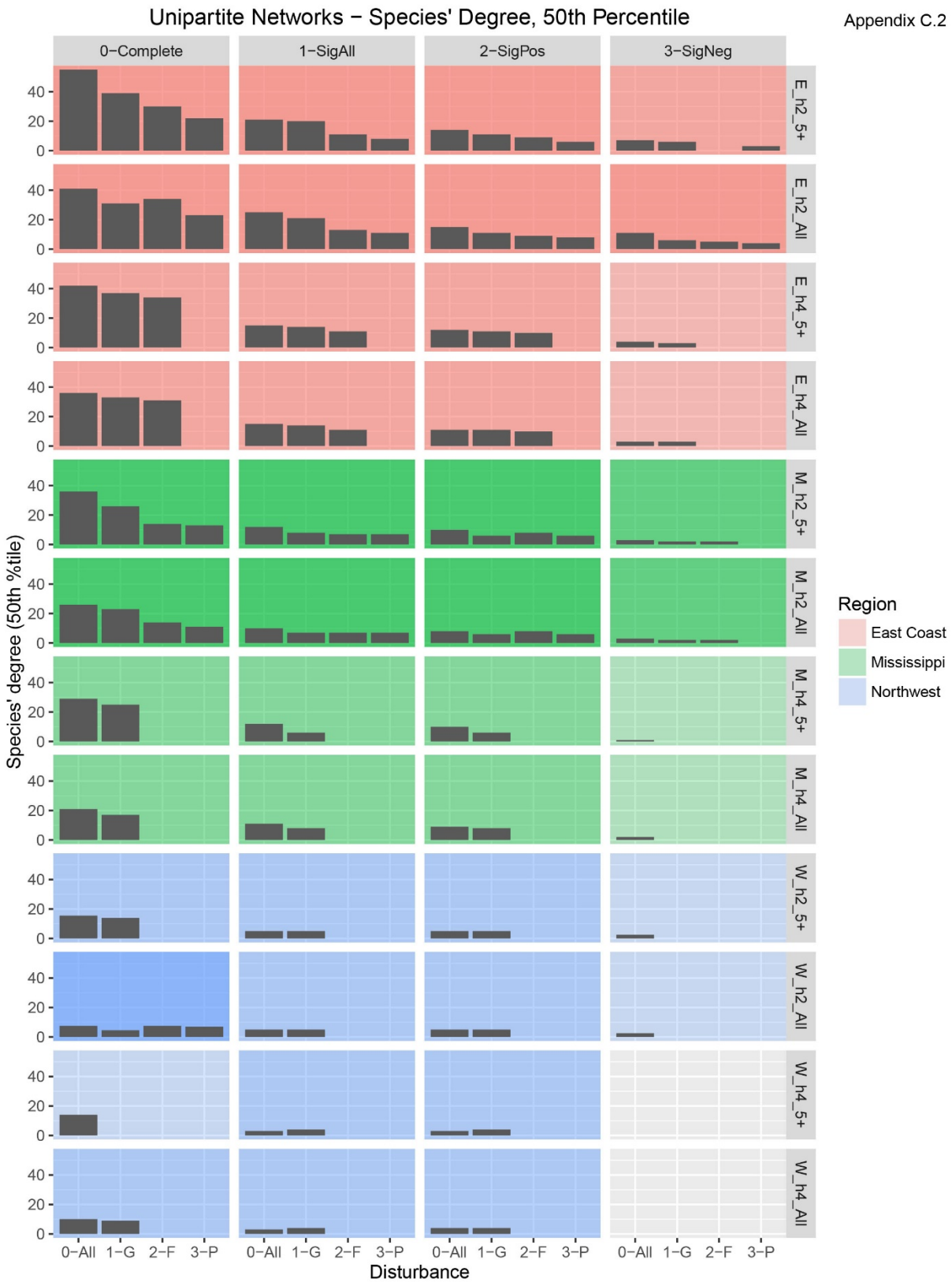
Networks that were built using all species' co-occurrences (0-Complete), combined significant positive and significantly negative co-occurrences (1-SigAll), significantly positive co-occurrences only (2-SigPos), and significantly negative co-occurrences only (3-SigNeg) are indicated by labels in the top margin.

Finally, networks that were built using data from all sites, regardless of disturbance level (0-All), data from minimally disturbed or 'good' sites (1-G), data from moderately disturbed or 'fair' sites (2-F), and data from highly disturbed or 'poor' sites (3-P) are indicated by labels at bottom.

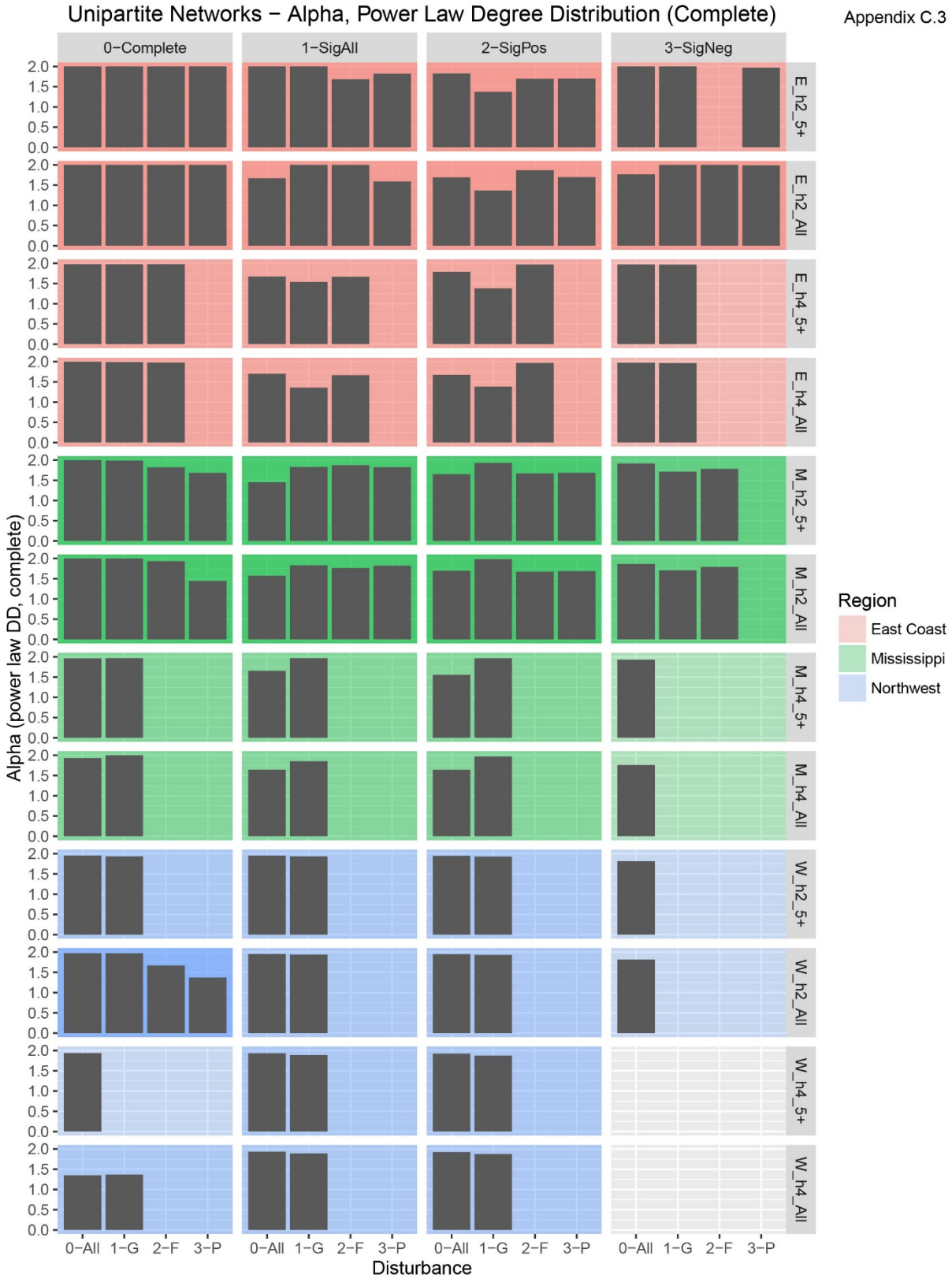
Appendix C.1


Unipartite Networks − Species' Degree, Maximum
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Species' degree (maximum)
100
50
0
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

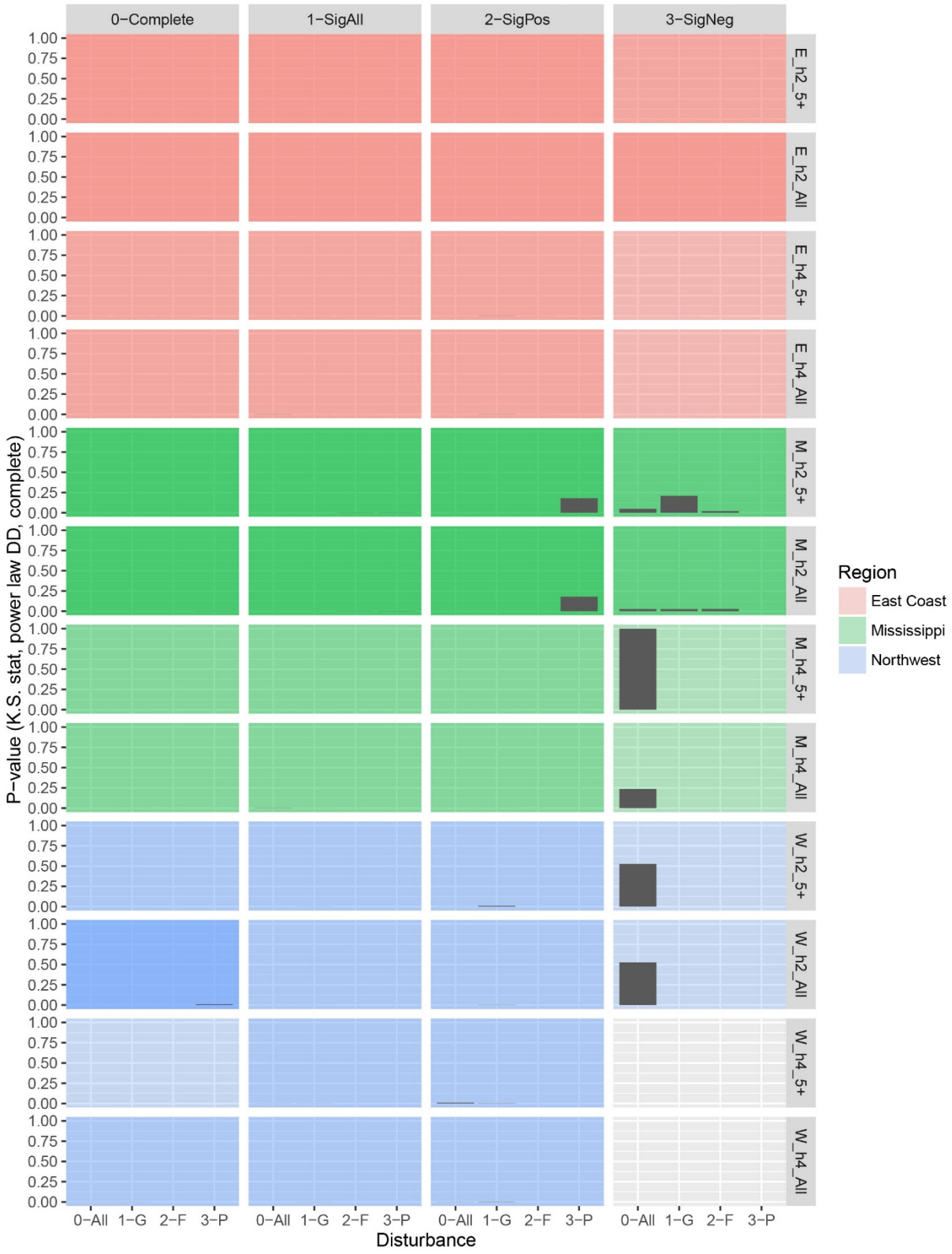
Appendix C.2


Unipartite Networks − Species' Degree, 50th Percentile
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Species' degree (50th %tile)
0
20
40
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

Appendix C.3


Unipartite Networks − Alpha, Power Law Degree Distribution (Complete)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Alpha (power law DD, complete)
2.0
1.5
1.0
0.5
0.0
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

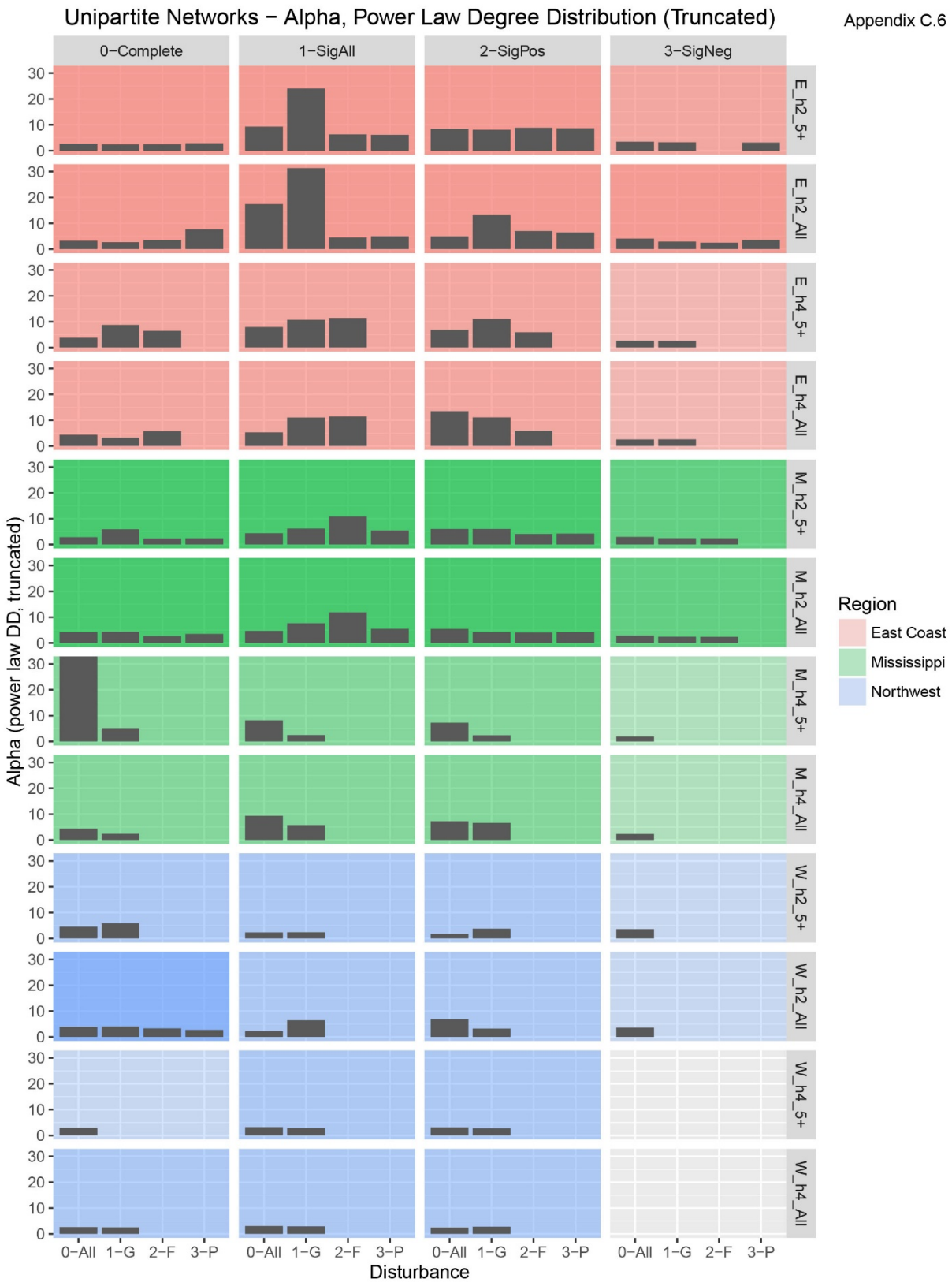
Unipartite Networks − Kol.Smi. P−value, Power Law Degree Distribution (Complete) Appendix C.4



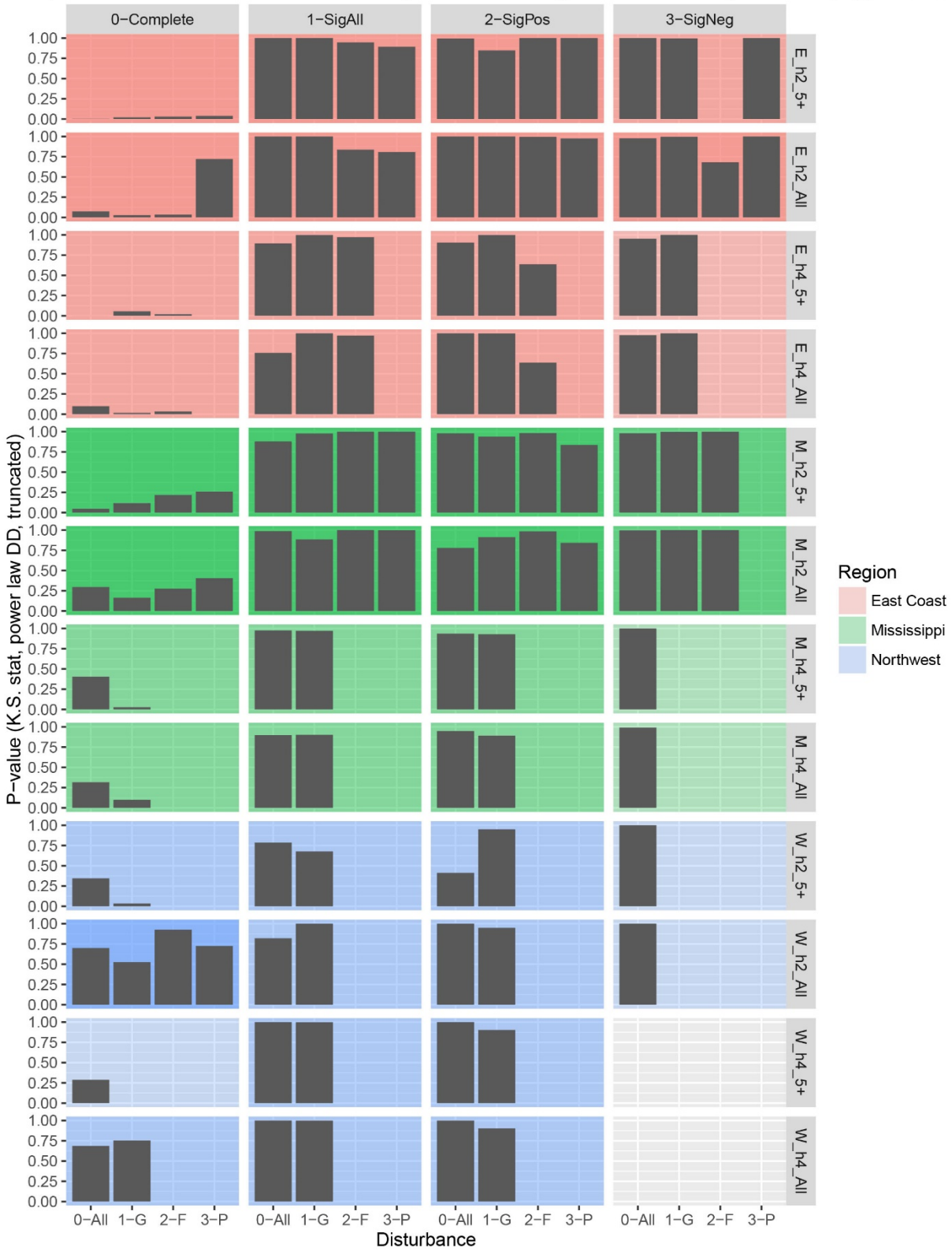
Appendix C.5

Unipartite Networks − Xmin, Power Law Degree Distribution (Truncated)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Xmin (power law DD, truncated)
0
20
40
60
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

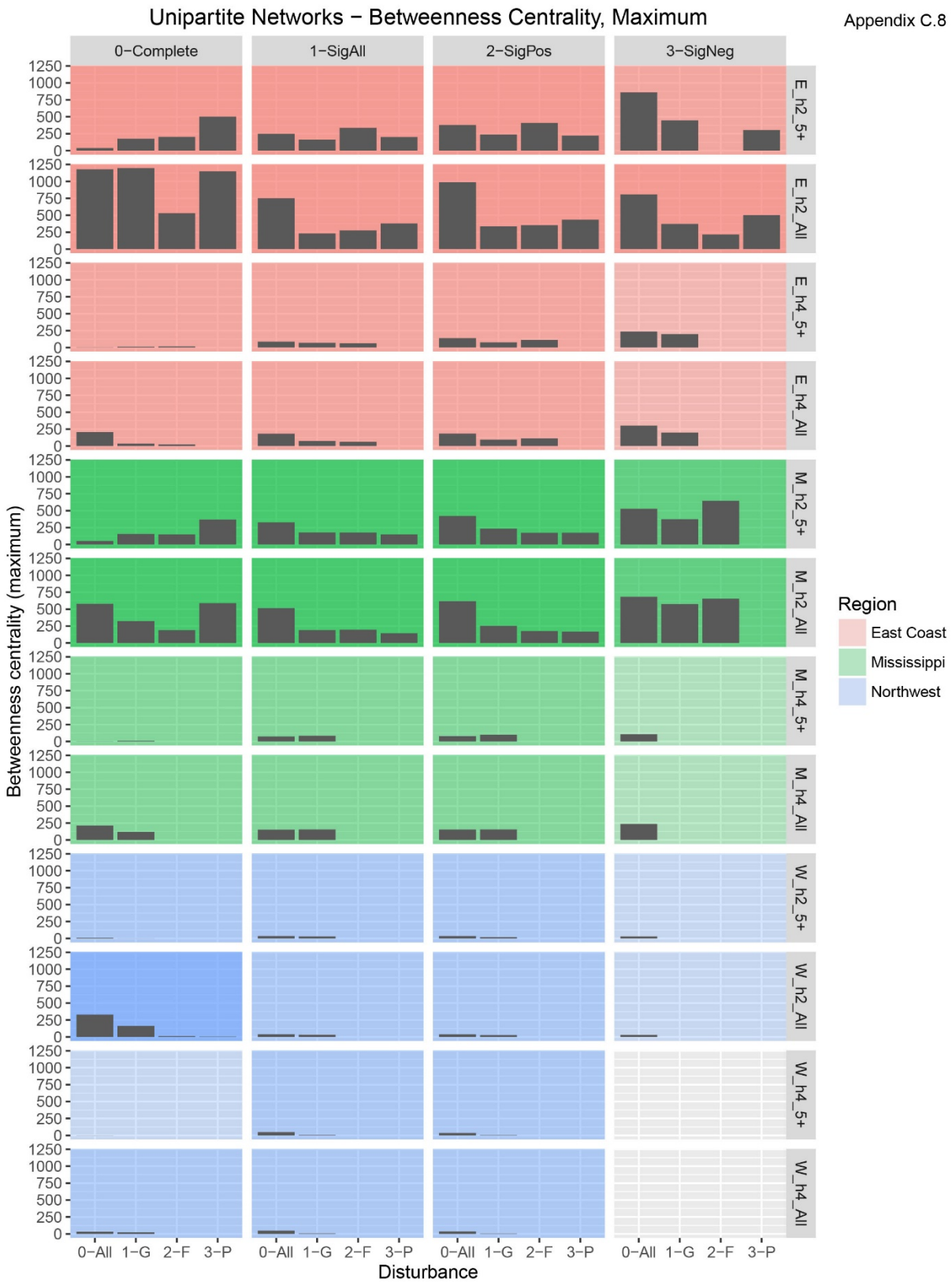
Appendix C.6


Unipartite Networks − Alpha, Power Law Degree Distribution (Truncated)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Alpha (power law DD, truncated)
0
10
20
30
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

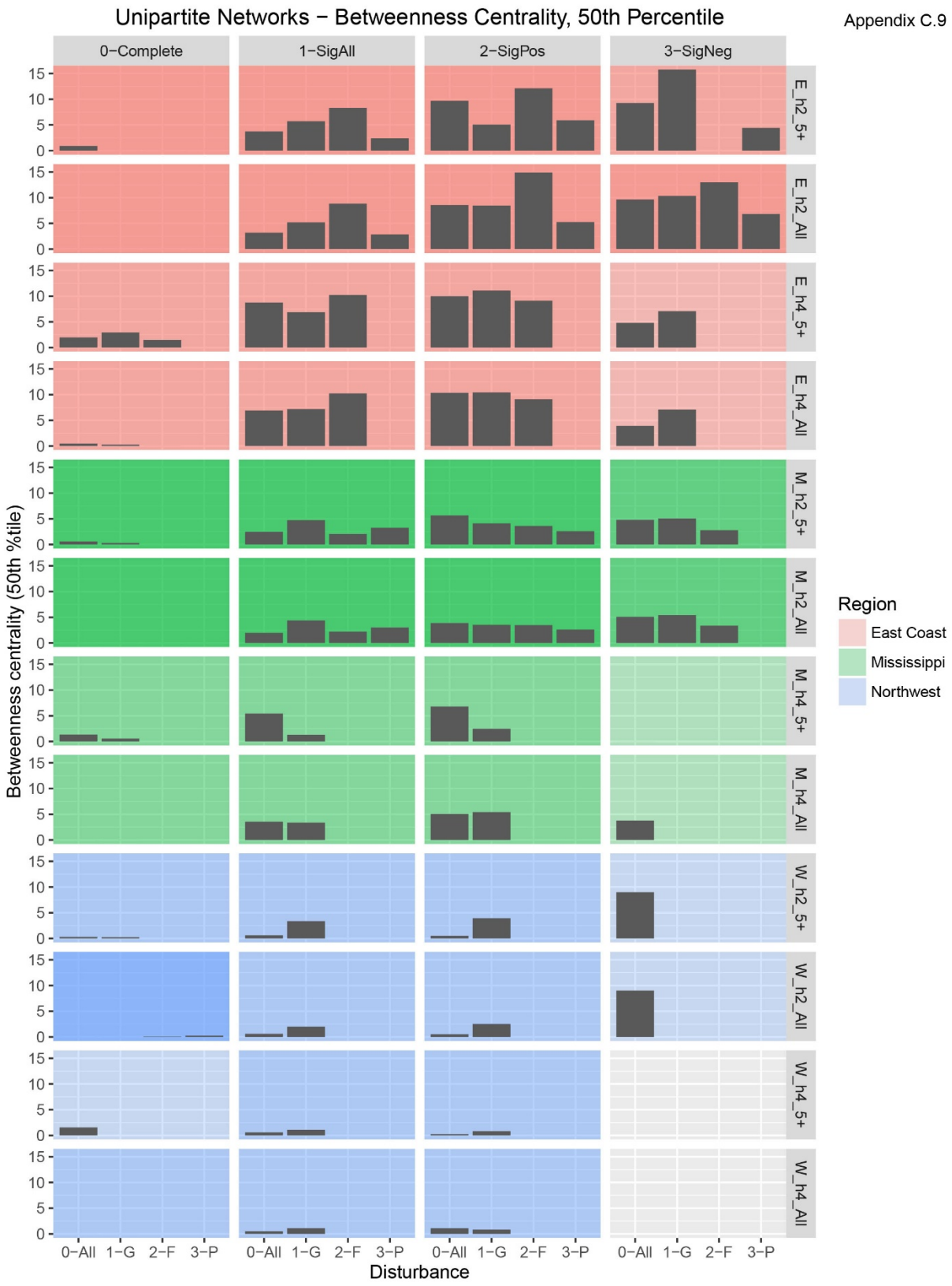
Unipartite Networks − Kol.Smi. P−value, Power Law Degree Distribution (Truncated) Appendix C.7



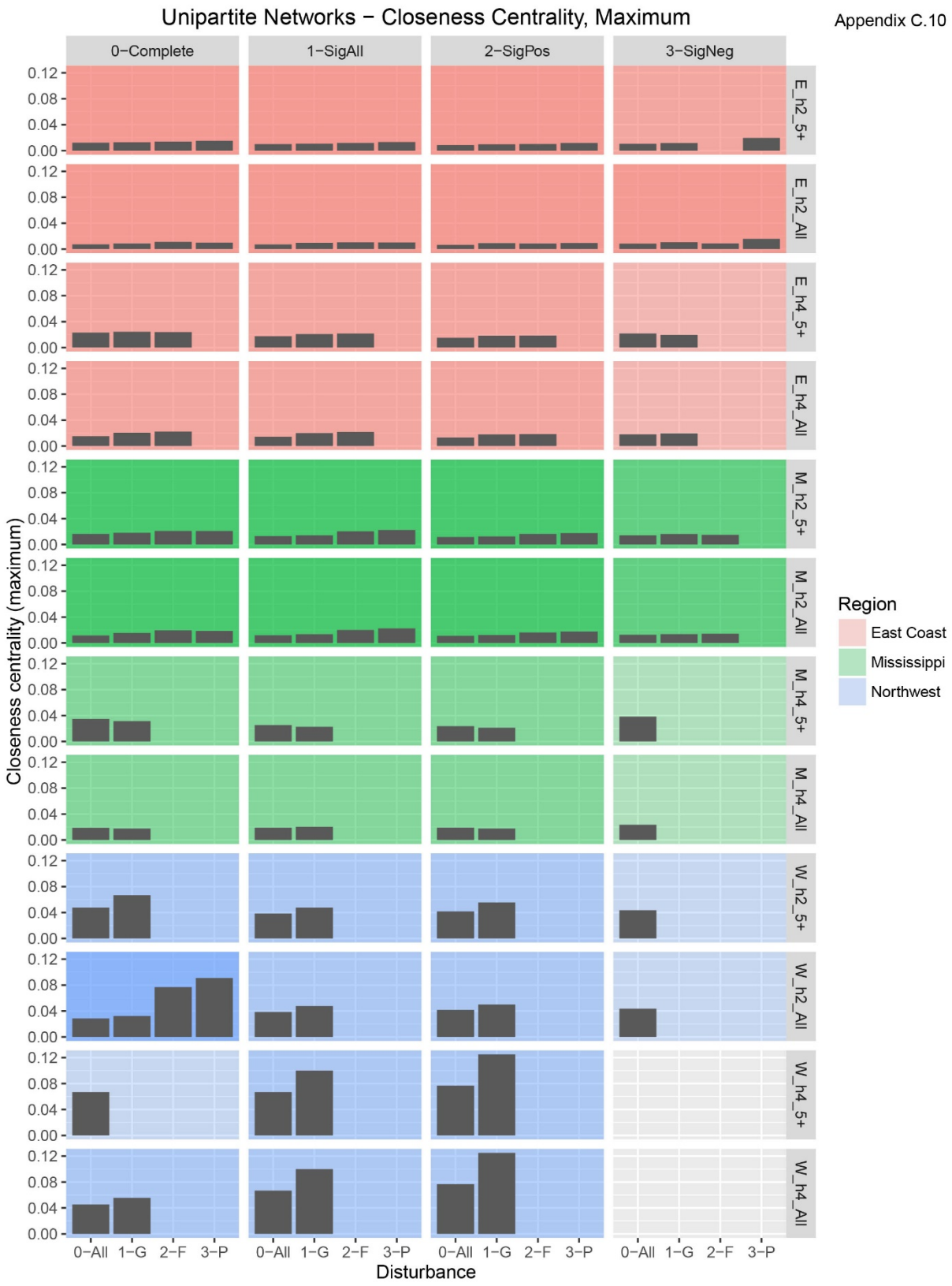
Appendix C.8


Unipartite Networks − Betweenness Centrality, Maximum
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Betweenness centrality (maximum)
0
250
500
750
1000
1250
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

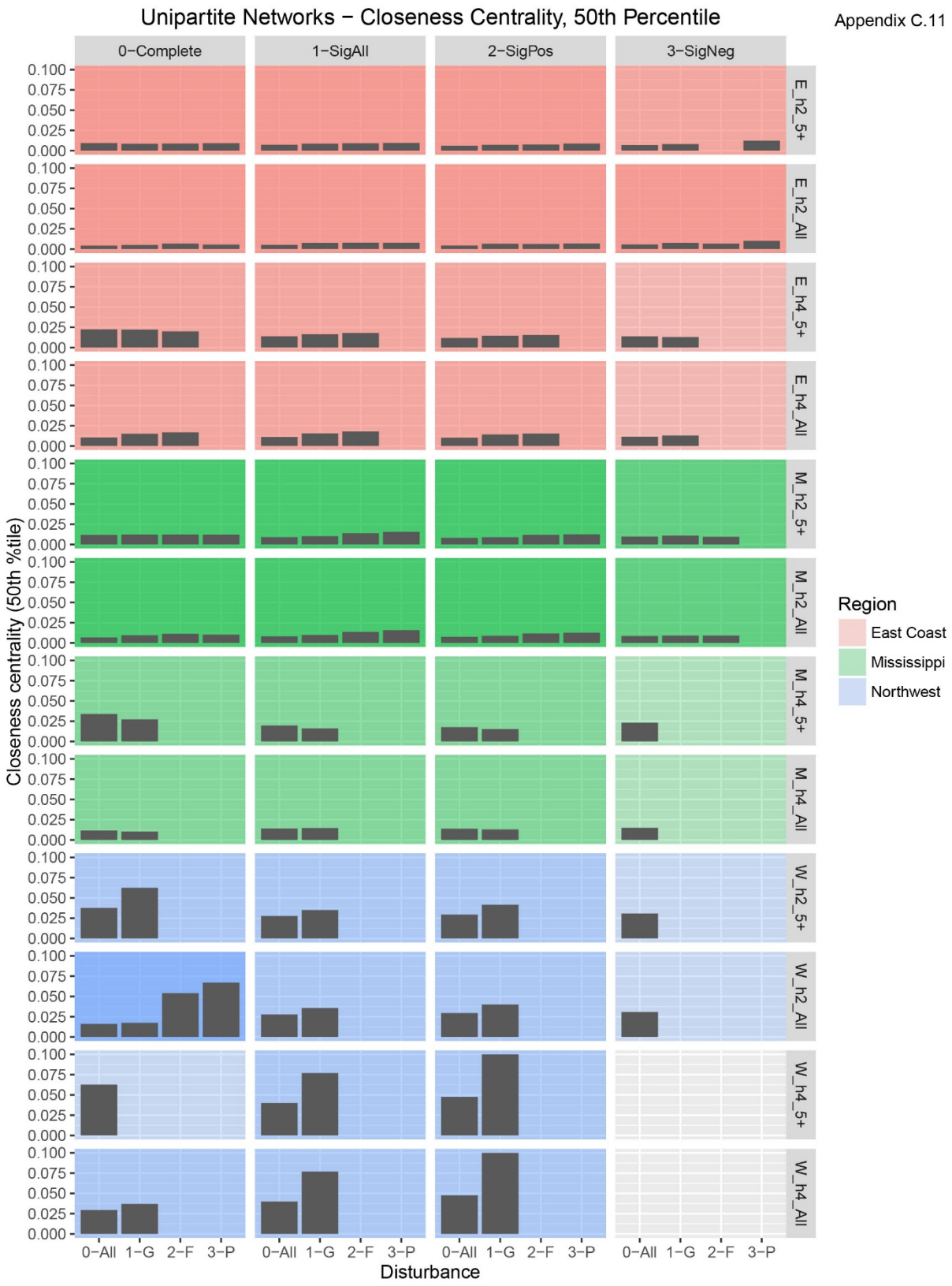
Appendix C.9


Unipartite Networks − Betweenness Centrality, 50th Percentile
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Betweenness centrality (50th %tile)
0
5
10
15
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

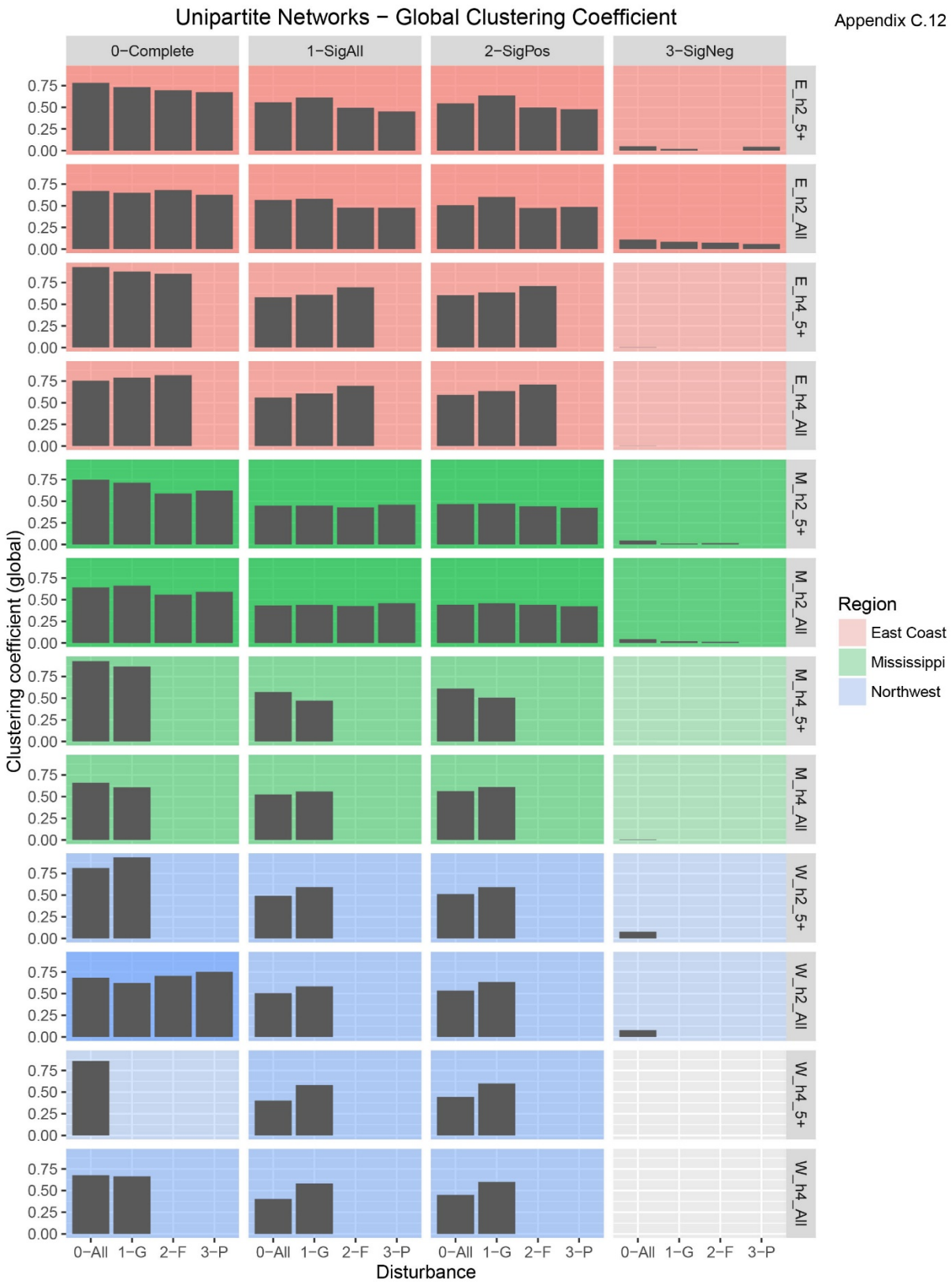
Appendix C.10



Appendix C.11

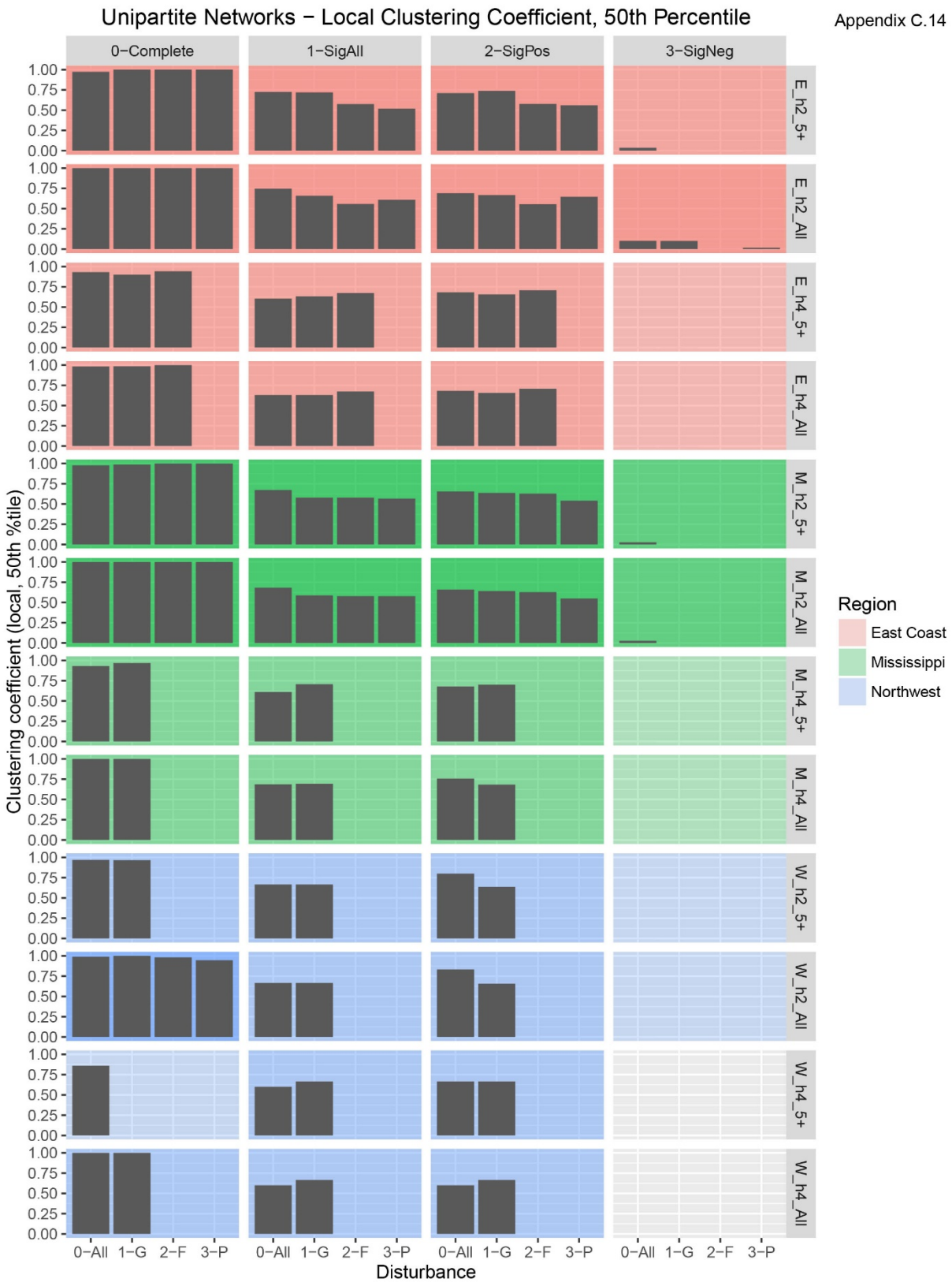


Appendix C.12

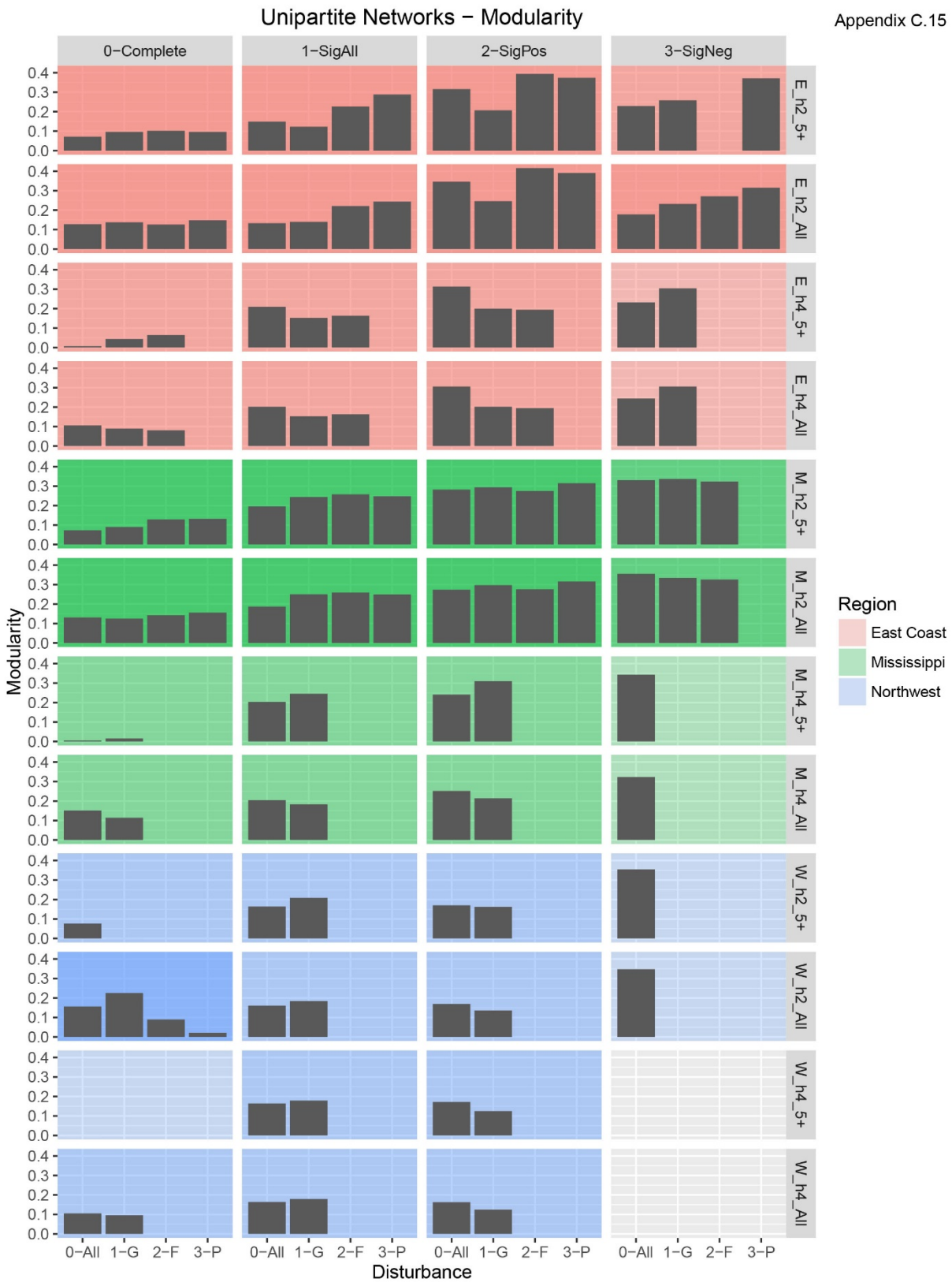


Appendix C.13

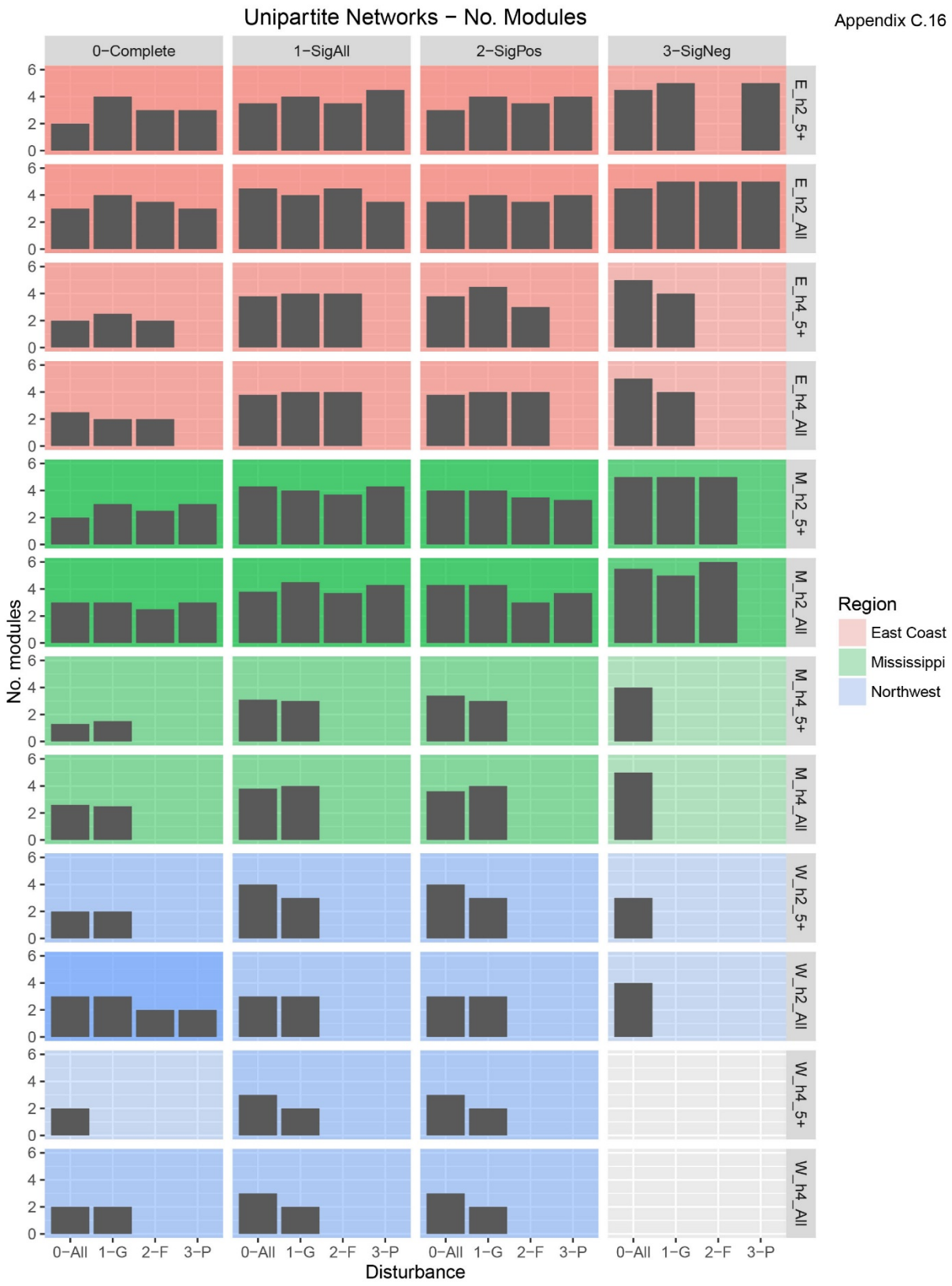
Appendix C.14



Appendix C.15



Appendix C.16



Appendix D

Bipartite Network Summary Statistics

Bipartite Networks – *Betweenness Centrality, Maximum (Sites)*
Bipartite Networks – *Betweenness Centrality, 50th Percentile (Sites)*
Bipartite Networks – *Closeness Centrality, Maximum (Sites)*
Bipartite Networks – *Closeness Centrality, 50th Percentile (Sites)*
Bipartite Networks – *Global Clustering Coefficient (Sites)*
Bipartite Networks – *Local Clustering Coefficient, Maximum (Sites)*
Bipartite Networks – *Local Clustering Coefficient, 50th Percentile (Sites)*
Bipartite Networks – *Modularity*
Bipartite Networks – *No. Modules*

General Key

Each of the following pages uses a common motif to illustrate a single network summary statistic, calculated and plotted in a single chart for *all* networks. For each of the summary statistics, we show the mean average, calculated across replicates (when more than one replicate was available) for each combination of data, visualized as a bar chart. Results are organized by three levels.

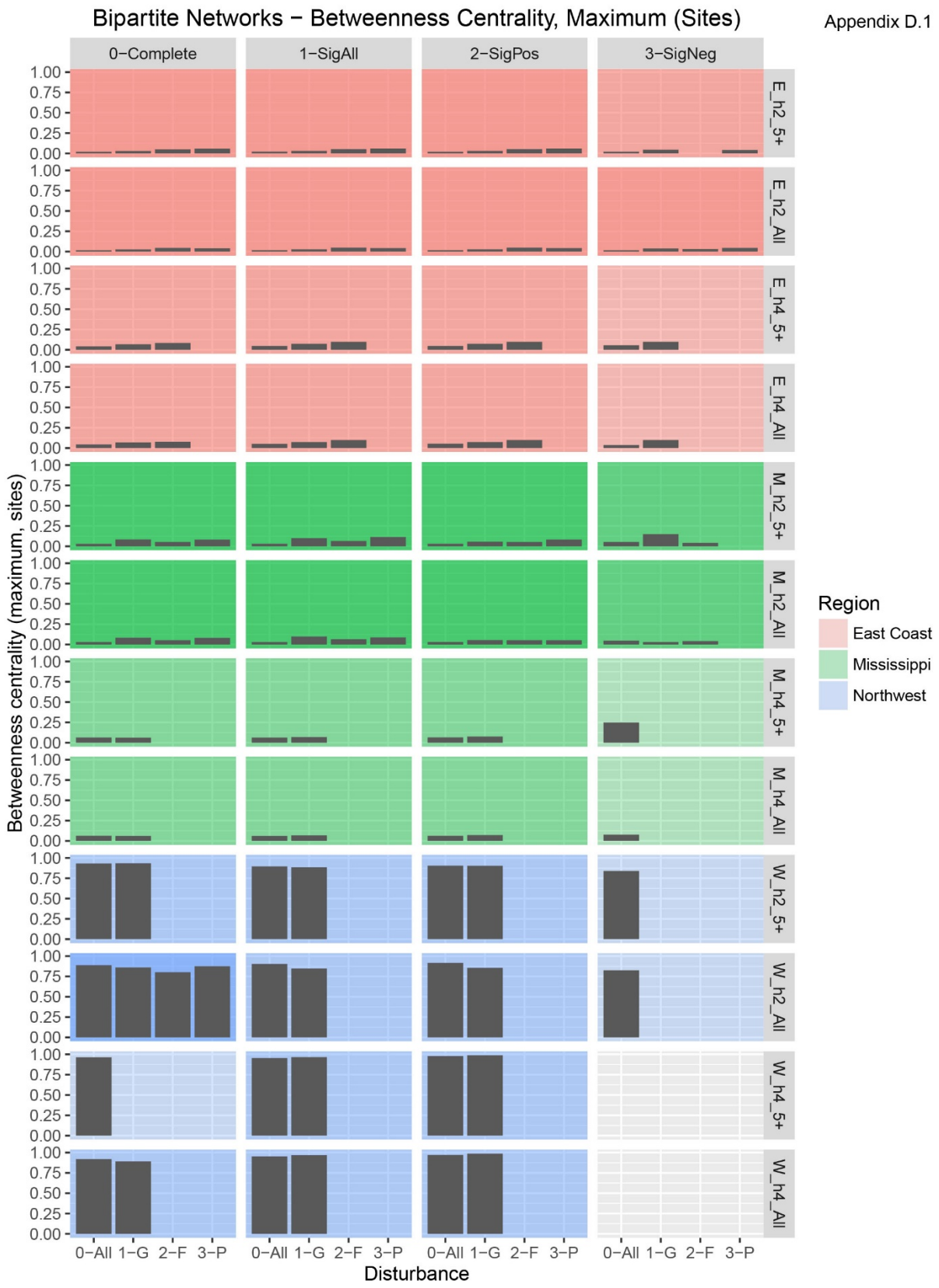
Each biogeographic region (East Coast/Mid-Atlantic, Mississippi River Basin, Pacific Northwest) is indicated by a background color, as shown in the 'Region' legend at right.

Within biogeographic regions, network statistics are partitioned by spatial scale (HUC2 vs. HUC4, indicated by 'h2' and 'h4' labels along the right-hand margin. Results from networks that included all species within a given data combination are also distinguished from results that used the conservative 'five or more occurrences per species' criterion (see main text) are also distinguished in the right-hand margin by 'All' and '5+' labels. The 'E', 'M', and 'W' labels in the right-hand margin are biogeographic region indicators (i.e. are redundant with the grouping by color).

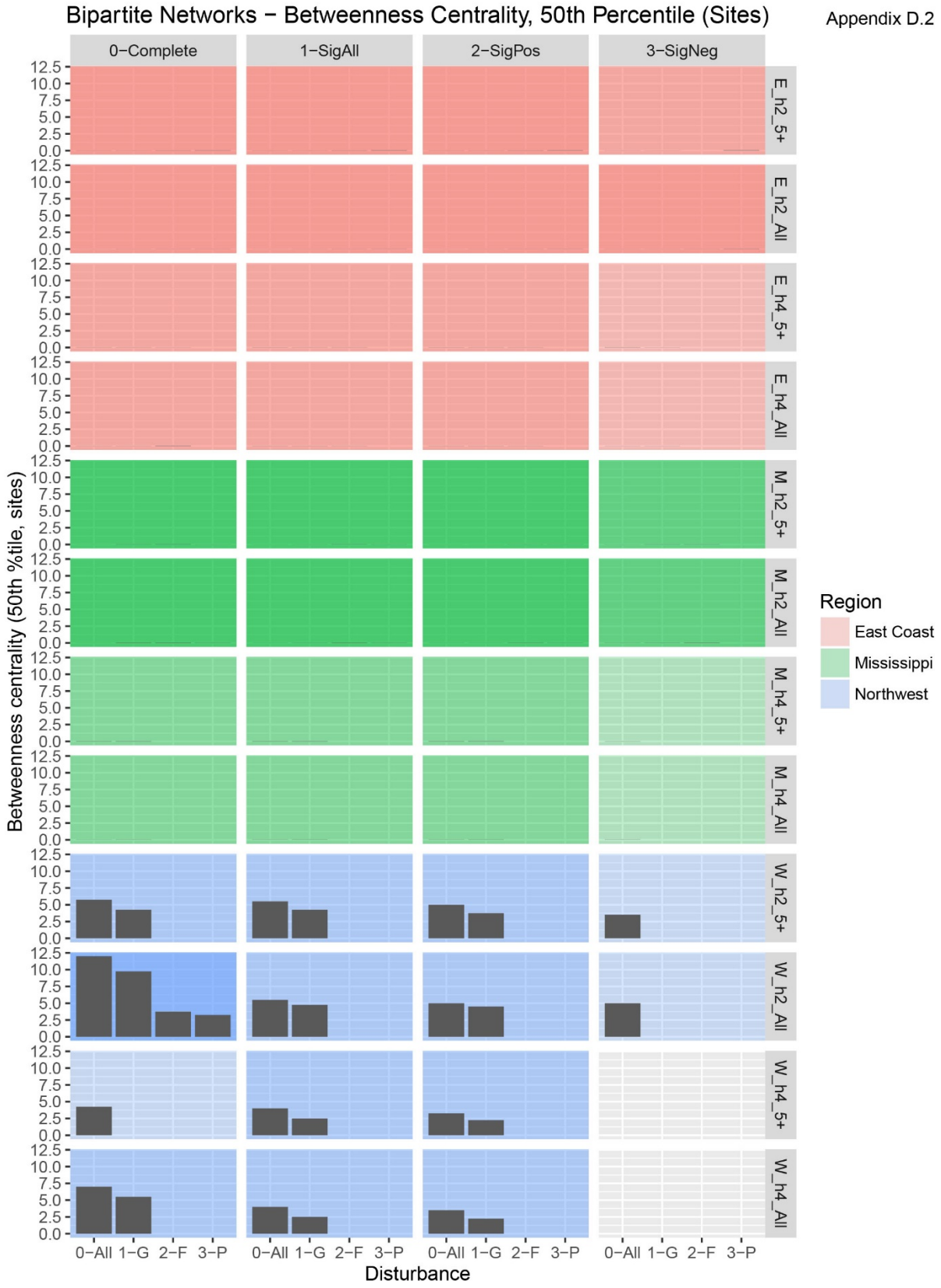
Networks that were built using all species' co-occurrences (0-Complete), combined significant positive and significantly negative co-occurrences (1-SigAll), significantly positive co-occurrences only (2-SigPos), and significantly negative co-occurrences only (3-SigNeg) are indicated by labels in the top margin.

Finally, networks that were built using data from all sites, regardless of disturbance level (0-All), data from minimally disturbed or 'good' sites (1-G), data from moderately disturbed or 'fair' sites (2-F), and data from highly disturbed or 'poor' sites (3-P) are indicated by labels at bottom.

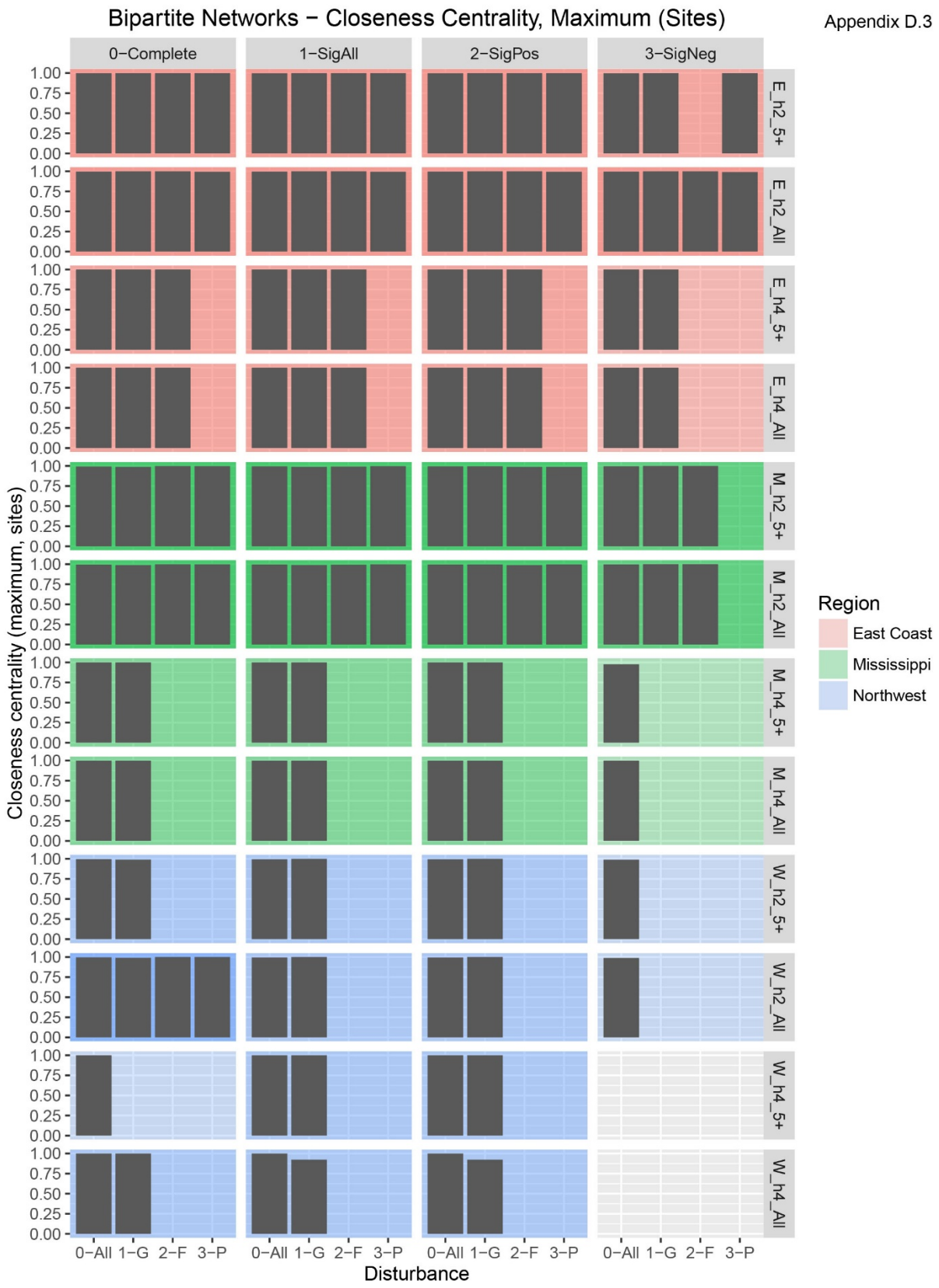
Appendix D.1


Bipartite Networks − Betweenness Centrality, Maximum (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Betweenness centrality (maximum, sites)
1.00
0.75
0.50
0.25
0.00
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

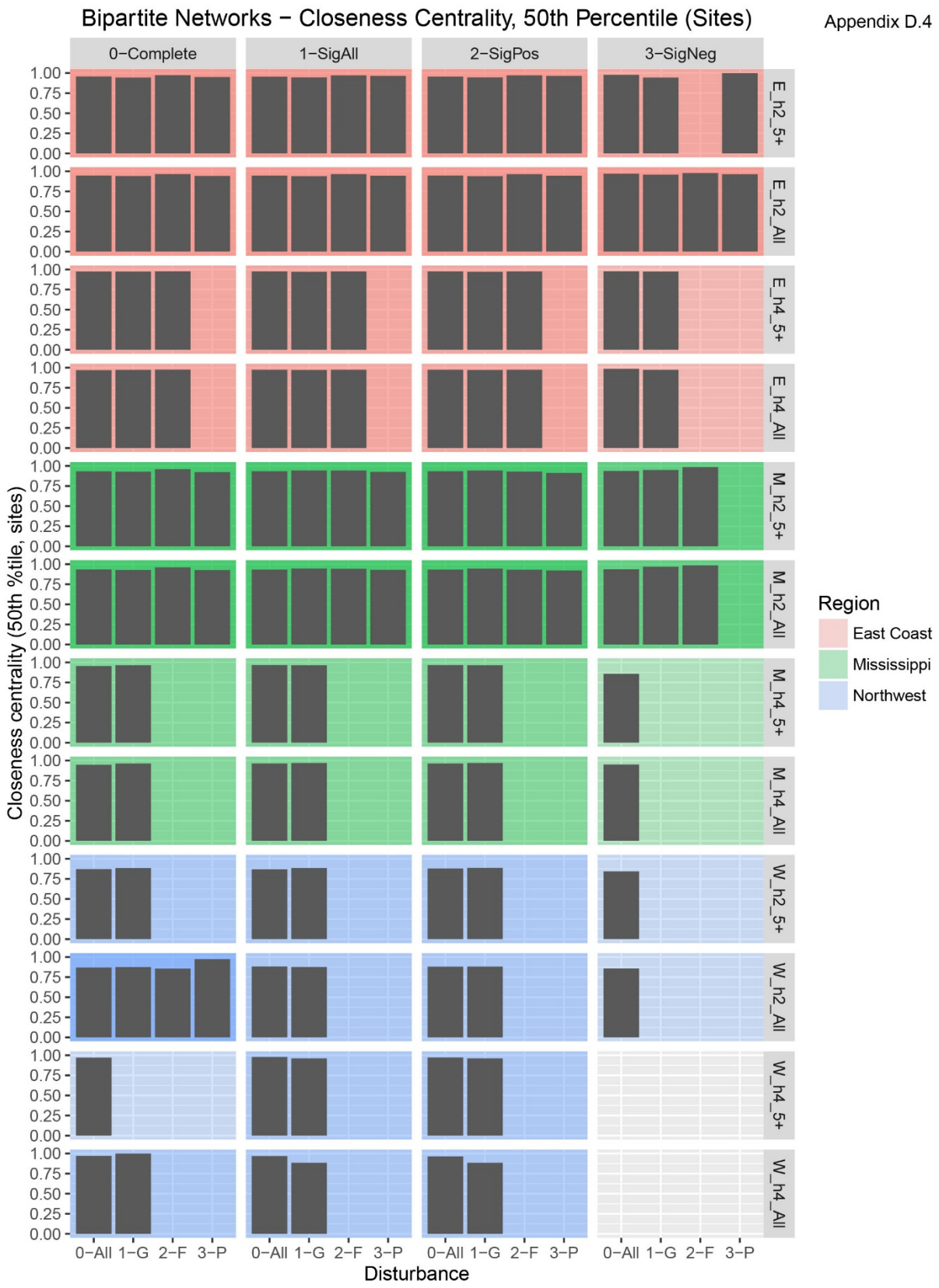
Appendix D.2


Bipartite Networks − Betweenness Centrality, 50th Percentile (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Betweenness centrality (50th %tile, sites)
12.5
10.0
7.5
5.0
2.5
0.0
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

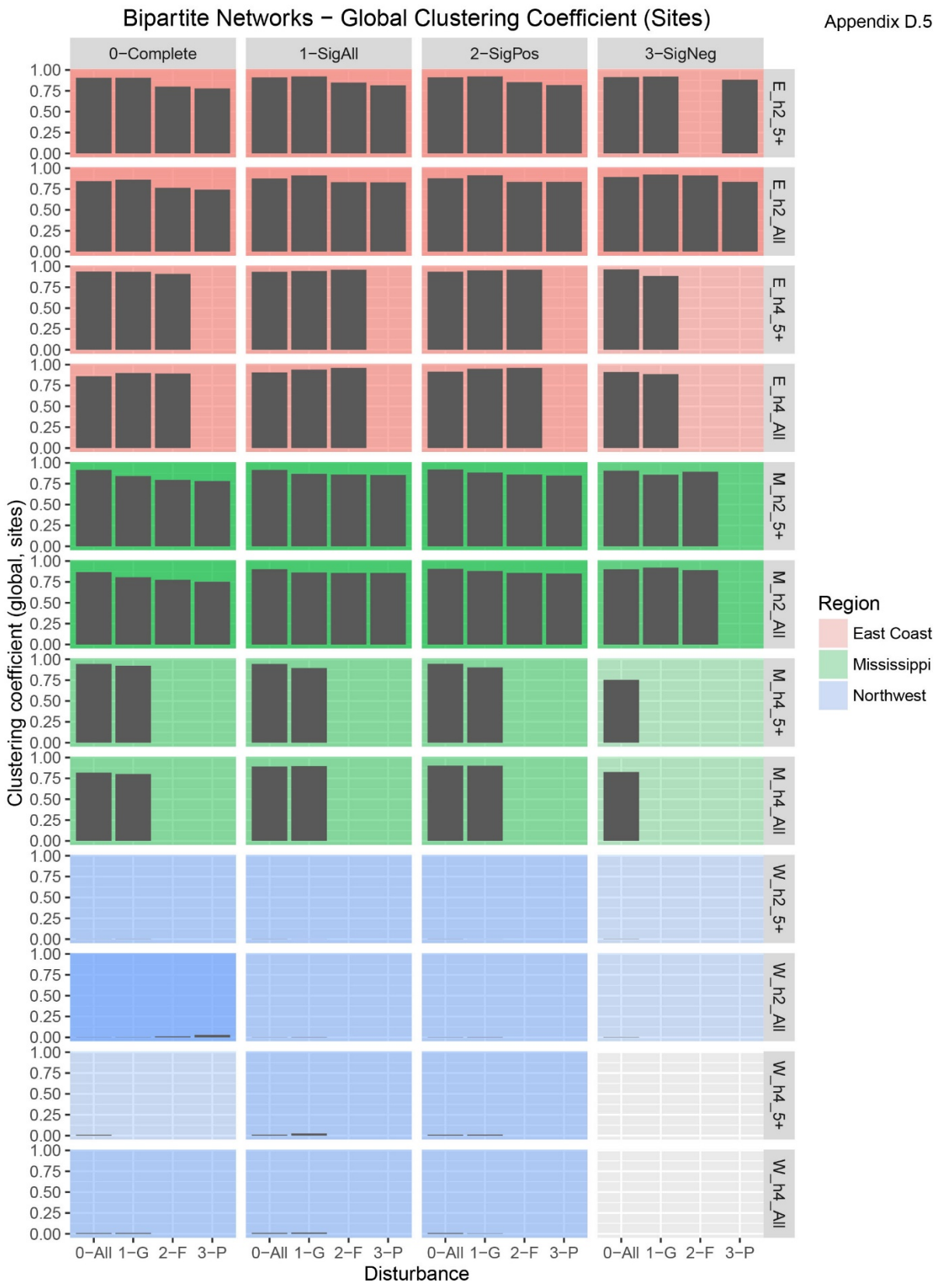
Appendix D.3


Bipartite Networks − Closeness Centrality, Maximum (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Closeness centrality (maximum, sites)
1.00
0.75
0.50
0.25
0.00
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

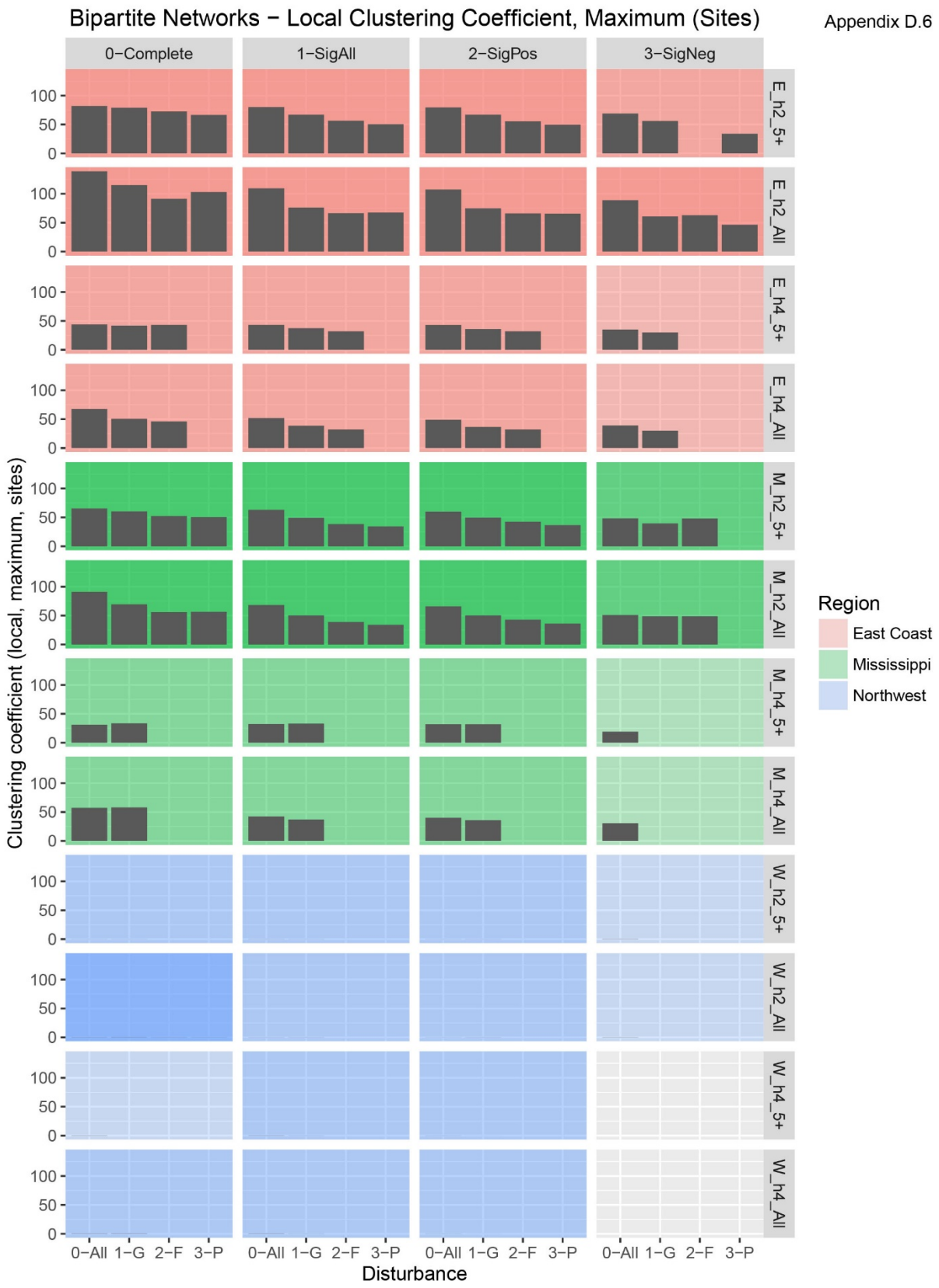
Appendix D.4


Bipartite Networks − Closeness Centrality, 50th Percentile (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Closeness centrality (50th %tile, sites)
1.00
0.75
0.50
0.25
0.00
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

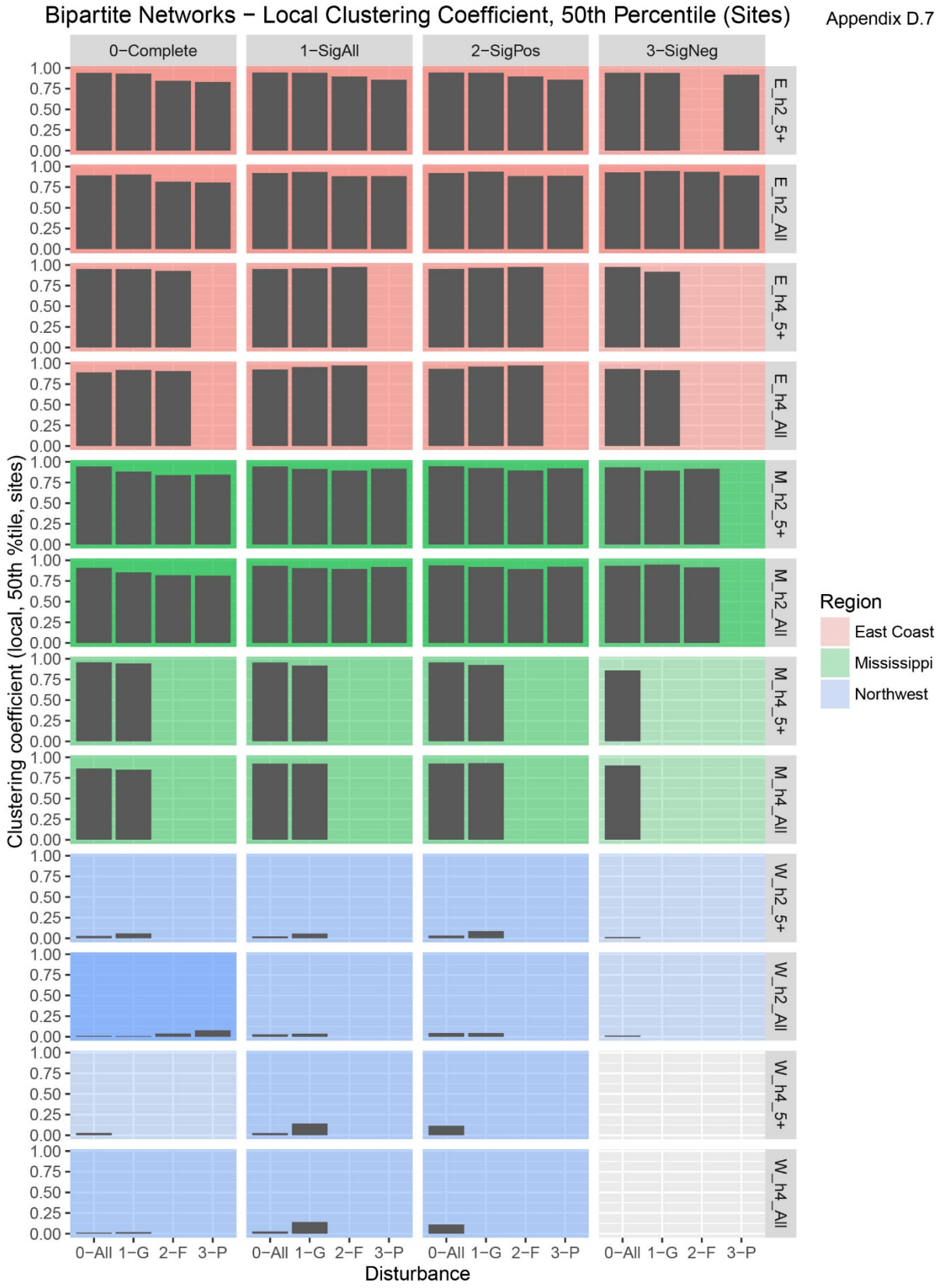
Appendix D.5


Bipartite Networks − Global Clustering Coefficient (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Clustering coefficient (global, sites)
1.00
0.75
0.50
0.25
0.00
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

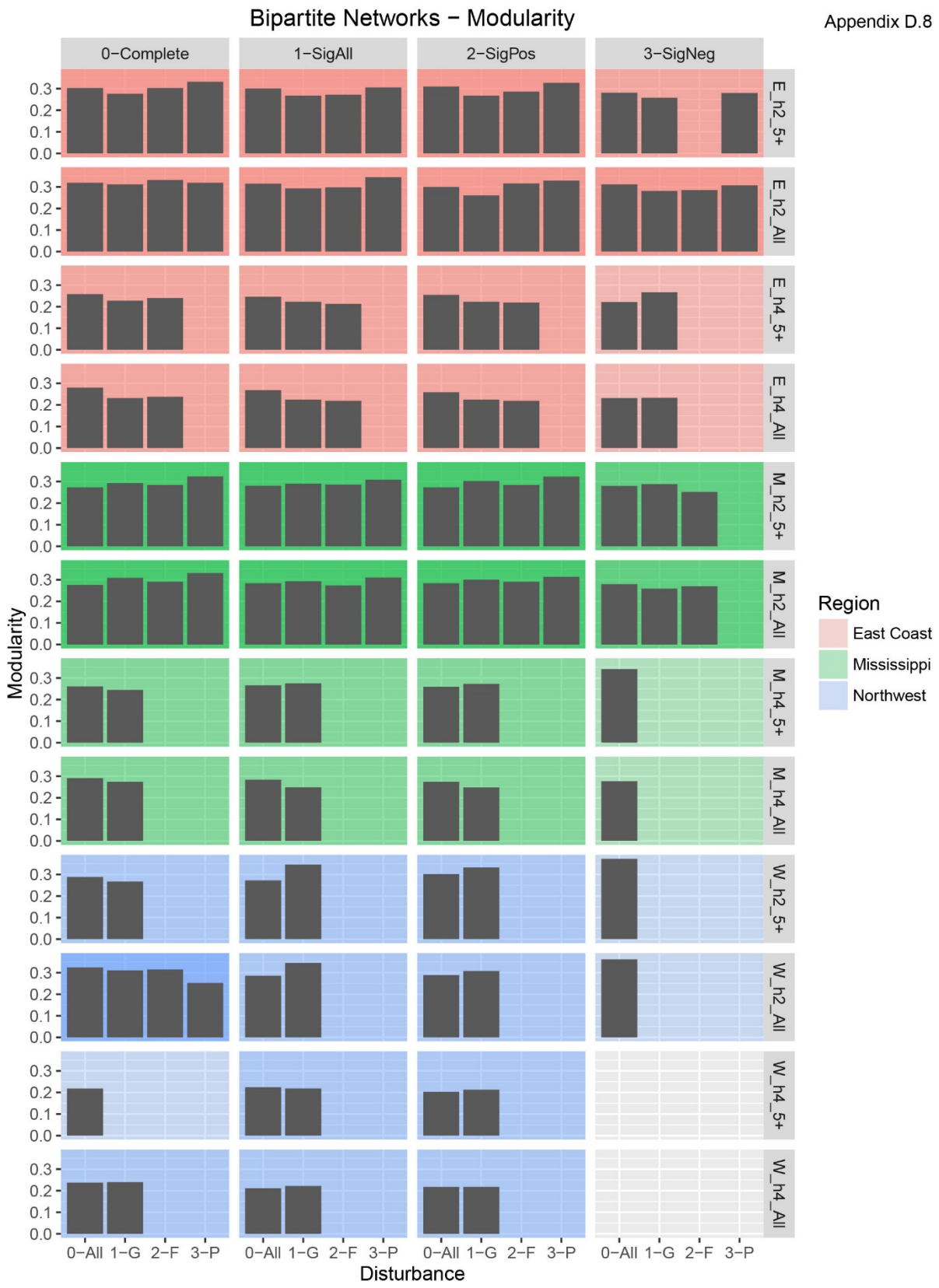
Appendix D.6


Bipartite Networks − Local Clustering Coefficient, Maximum (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Clustering coefficient (local, maximum, sites)
100
50
0
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

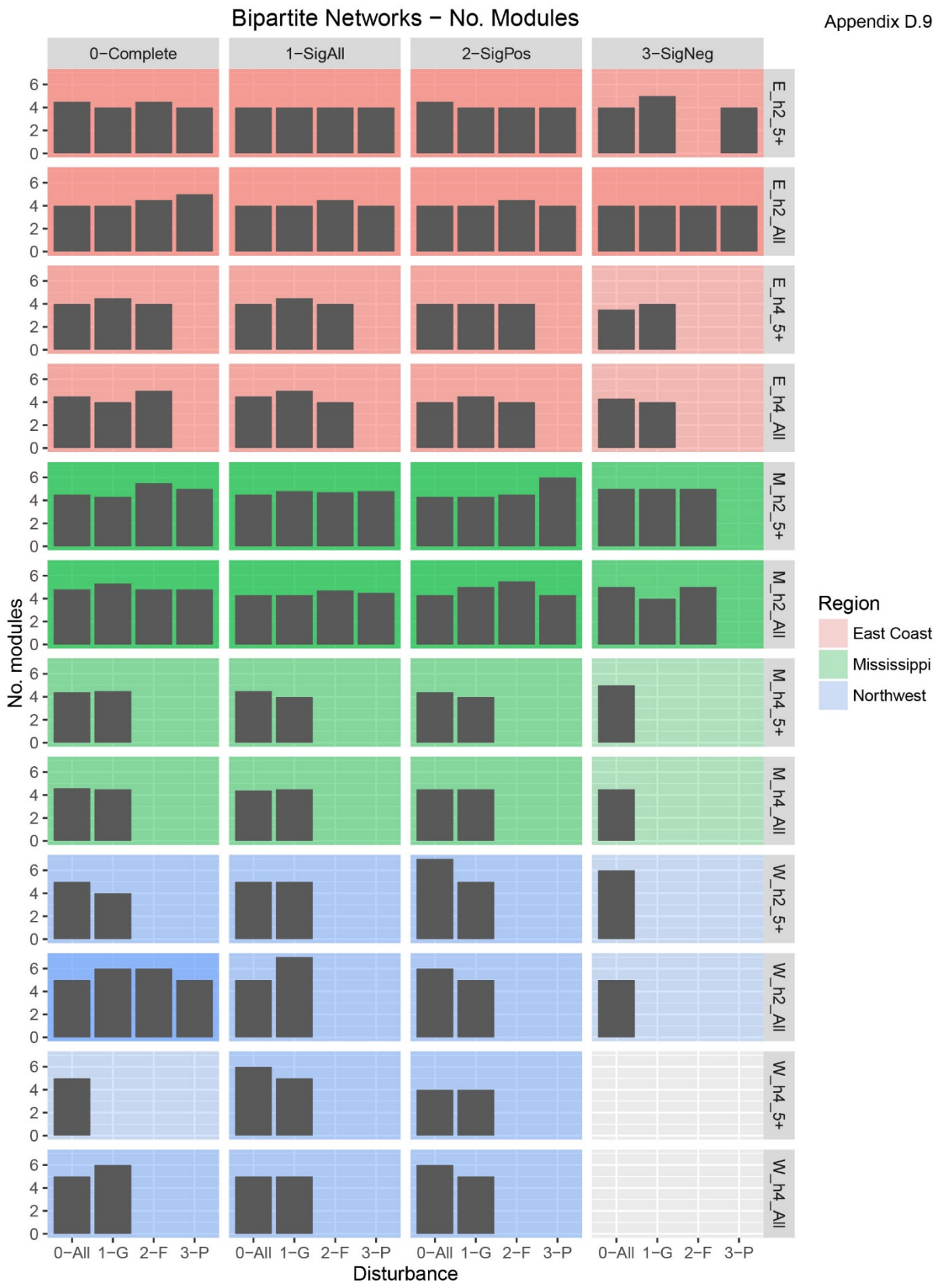
Appendix D.7


Bipartite Networks − Local Clustering Coefficient, 50th Percentile (Sites)
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Clustering coefficient (local, 50th %tile, sites)
1.00
0.75
0.50
0.25
0.00
0−All 1−G 2−F 3−P
Disturbance
Region
East Coast
Mississippi
Northwest

Appendix D.8


Bipartite Networks − Modularity
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
Modularity
0.0
0.1
0.2
0.3
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest

Appendix D.9


Bipartite Networks − No. Modules
0−Complete
1−SigAll
2−SigPos
3−SigNeg
E_h2_5+
E_h2_All
E_h4_5+
E_h4_All
M_h2_5+
M_h2_All
M_h4_5+
M_h4_All
W_h2_5+
W_h2_All
W_h4_5+
W_h4_All
No. modules
0
2
4
6
0−All
1−G
2−F
3−P
Disturbance
Region
East Coast
Mississippi
Northwest