THE UNIVERSITY OF CHICAGO

STRUCTURAL INFERENCES: THREE CASES OF LINKING PATTERN AND PROCESS IN ECOLOGICAL NETWORKS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY
MATTHEW J MICHALSKA-SMITH

CHICAGO, ILLINOIS
AUGUST 2018

To the giants, upon whose shoulders I rest

and to my wife who encouraged me to climb

We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance —John Archibald Wheeler

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work is due to a collaboration of many people, only a subset of which do I have space to mention here. First and foremost, this dissertation could not have been completed without the steadfast guidance of my advisor Stefano Allesina. His gift for asking the right questions and his insights into the scientific endeavor writ large have served as both inspiration and aspiration throughout my studies and I carry them with me into the next stage of my career. Second most influential has been my wife Marta, who encouraged and exposed me to science as a career, even when noone in my family had taken the path before. Her courage and strength continually push me to do more and better; to use what skills I have to make a positive impact on the world around me. My committee, consisting of Trevor Price, Mercedes Pascual, Sarah Cobey, and Amos Maritan have provided invaluable feedback on this text and many stages of my research leading up to this publication. Likewise, my fellow graduate students, especially Iuri Ventura, Marcos Vieira, Liz Sander, and Sylvia Ranjeva have pushed me to be a better student and encouraged me in times of doubt. And of course the post-docs. That ever between category, so often overlooked, but forming the backbone of modern science. Gyuri Barabás, Jacopo Grilli, Phillip Staniczenko, and others who have passed through the lab taught me so much about how to do science, but also about what it means to do science, and my heartfelt appreciation goes to them for providing a vision of a path forward. Finally I utilized an enormous amount of computing time and resources over the course of my training, and this definitely would be a far inferior dissertation without the frequent use of the computational resources of the Center for Research Informatics Tarbell and then Gardner HPC clusters at the University of Chicago.

ABSTRACT

Ecology has always been a study of interconnection, making the growth of network analysis something of a necessity as the field has continued to mature. Too often, however, network analyses are included without full appreciation for their appropriateness or requisite assumptions. This has led to a criticism of these methods and a call to better articulate their relationship to and implications for the underlying ecological system. In this dissertation, I describe three projects that seek to address this divide. In the first chapter, I look at how the structure of a network of interactions might be influenced by the type of interaction being described (e.g. parasitism or pollination). This is done through a combination of measuring a variety of network-structural properties and comparing distributions of these values between different types of networks using principal component analysis. We find that ecological interaction networks are less different than has been previously proposed. In the second chapter, we move this argument to inside a single network, in this case, a food-web. We ask if species pursuing different trophic strategies (e.g. parasitism or herbivory), have different structural roles within the network. We attempted detection of these differences by use of a network structure model known in Ecology as the group model. Though no single (or even combination of several) simple structural properties is sufficient to distinguish trophic strategies from one another, the group model consistently and significantly creates groups of species of predominantly one trophic strategy. In the final core chapter, we turn to networks of symmetric competition as might be found is a system approximated by Lotka-Volterra competitive dynamics. We ask what overarching network structure would yield the most or least stabilizing community, and then ask how inter- and intraspecific components of that structure might differentially influence that (de)stabilizing effect when varied independently of one another. Surprisingly both the most and least stabilizing configurations have clear, low-rank signatures in the network structure, with a nested configuration proving most stabilizing and an anti-modular configuration least stabilizing. In each of these three

cases, we utilize a fairly abstract method, but always with the goal of tying results to empirically observed network structures. As with all science, only time will reveal success in this endeavor.

CHAPTER 1

INTRODUCTION

Perhaps most commonly characterized through Darwin's memorable reference to an "entangled bank" [43], the realization that nature can be interpreted as a complex network of interacting beings has proven to be a driving idea throughout the history of Ecology as a discipline. Drawing on a deep history of network science stretching back more than a century before *The Origin of Species* [56], the application of network methods developed across mathematics, social sciences, physics, *etc.* has spurred ecological research around some of the most important questions of ecology up to the present day [74].

In addition to the expansion of ecological network science in terms of the sophistication and ubiquity of the mathematical tools involved, there has also been an extension into a variety of types of ecological networks. A glance at publication trends confirms that we are now in a period of exponential growth for ecological network science, but moreover, despite being dominated by the study of food-webs and competition networks for many years, the growth in the study of networks of all sorts of ecological interactions is taking up an increasing fraction of this research.

As interest in networks has grown and network representations of species interactions have become commonplace, the need for consistent ways to identify and quantify aspects of

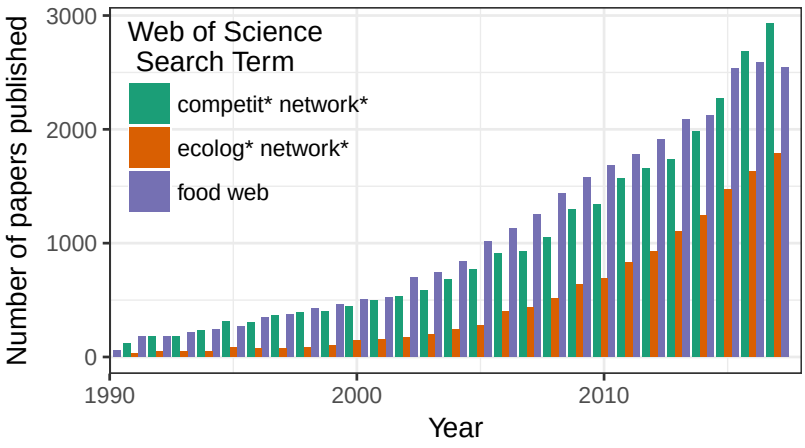


Figure 1.1: Number of publications corresponding to the given search term found on the Web of Science citation indexing service (Thomson Reuters). Bars corresponding to the search "ecolog* network*" have had results overlapping with the other two searches subtracted to minimize double-counting.

network structure has become increasingly apparent. As a response, scientists have taken to identifying metrics of network structure from the very simple (*e.g.* how many species are interacting? and what proportion of possible links are actually observed?) to the much more complicated and nuanced (*e.g.* to what degree are specialist interactions a subset of the interactions of generalists? and to what degree can species be ordered such that species interactions are interval for all species?).

In addition to developing metrics to quantify network structure, a cottage industry arose for developing models with the dual purpose of evaluating similarity to empirically observed networks and generating random networks according to simple structural constraints for use as null-models against which to compare the host of metrics one could measure for any given network.

This growth has, in some ways, outpaced the second order considerations of which methods and metrics are most appropriate for a given ecological situation. For example, centrality measures have become commonplace in initial forays into applying network approaches to a new field, yet considering which of the many centrality measures is most appropriate for a given question is often forgone in preference of listing all of those the authors cared to measure [23]. As new methods are being developed and the ease of implementing these increases, Ecology has struggled to relate metrics back to the biological processes underlying these systems, leading to some criticism of network ecology *en masse* [20, 22].

This dissertation seeks to explore how ecological network structure relates to underlying biological processes. In the next chapter, I investigate ecological networks of various provenance, describing a variety of different types of ecological interactions. I ask whether or not the disparate ecological and evolutionary forces each of these networks has undergone over the course of its formation have led to differing signatures within the network structure than can be detected and further investigated. A correct classification method could shed light on which structures provide benefits in terms of stability or persistence and contrasting

signatures between, for instance, mutualistic and antagonistic networks could provide insight into previously unrealized differences in the ways these interactions shape communities at large.

In chapter three, I frame a similar question, this time looking at individual species within a single network: are there patterns of connections which shed light on their trophic strategies (*i.e.* primary producer, herbivore, generalist consumer, parasite, *etc.*)? We utilize a model of ecological network structure termed the "Group Model" which partitions species within a network into those sharing similar predators and prey and ask whether or not the aforementioned trophic strategies, focusing in particular on the notoriously ambiguous category of parasites, are distributed nonrandomly within these groups.

Finally, in chapter four, I consider competition networks in isolation, re-evaluating the traditional dogma born out of two-species Lotka-Volterra competition that intraspecific competition must be greater than interspecific competition for coexistence. This formulation fails even with the addition of one more species, but are there informative patterns in larger competition networks that lead to greater or less stability? Moreover, what types of ecological scenarios might lead to one or the other of these structures?

I conclude with some comments on studying ecological network structure and indicate some potential avenues for future research.

CHAPTER 2

CLASSIFYING ECOLOGICAL INTERACTION NETWORKS USING QUANTIFIABLE DIFFERENCES IN NETWORK STRUCTURE: A COMPUTATIONAL CHALLENGE

Matthew J. Michalska-Smith & Stefano Allesina

Any two large networks are different, and will yield different values for a number of structural metrics; yet, whether or not these differences generalize to comparisons between classes of networks is an open question. Resolution of this question has been stymied both by the sophistication of the problem and by the number of technical choices involved in finding a solution. Here, we raise a call-to-arms for the scientific community to take up the challenge of classifying ecological interaction networks according to differences in network structure. We work to detail methods for appraising the quality and success of any proposed method and highlight some critical considerations. Finally, we give an example of a strategy that works for highly replicated non-ecological networks, but fails to distinguish ecological interaction networks based on the type of interaction they represent. This challenge, if solved, would be both a major step forward in our understanding of ecological network structure and a validation of the extolled usefulness of network analyses in Ecology.

Ecological literature [17, 90, 142, 144, 145, 147] and philosophical intuition suggest that networks depicting diverse ecological processes such as pollination, ant-plant defense mutualisms, herbivory, and parasitic infections should differ in the organization of interactions among the species involved. Yet, we show that a closer look reveals that some of the perceived differences do not stand up to scrutiny and the simplistic approach of ordinating a large number of ecological networks according to various structural metrics fails to provide

any headway in distinguishing these ecological interaction networks from one another. In this work, we lay out a computational challenge for the scientific community, explaining the desired qualities of a successful method as well as noting some of the many considerations involved. We compile a large data-base of both ecological and non-ecological networks upon which to test proposed methods and provide sample code to allow the easy addition of new metrics for comparison.

2.1 Background

Species interactions are a driving force in both ecology and evolution and are ubiquitous in nature: flowers are pollinated by insects, carnivores consume herbivores, birds disperse the seeds of trees over great distances, *etc.* Often these connections between species form complex patterns involving hundreds of species and thousands of recorded interactions, and may consist of intricate non-random arrangements across a range of scales. The use of network tools to analyze these systems is thus a natural choice that has been used with great success for well over a century [29, 74].

Interest has been levied over this time to explore the structural organization of these networks, leading to the development of new approaches and metrics for quantifying the patterns we observe. The simplest way this structure has been evaluated is through the use of summary statistics ranging from very simple metrics such as size (the number of nodes) and connectance (the proportion of all possible edges that are realized) to more complicated measures such as centrality (a family of measures related to the importance of a node in a given flow regime) and degree distribution (the number of edges connected to each node in the network) [49, 118].

In ecology, nestedness (to what degree are specialist interactions proper subsets of generalist interactions?; [149, 154]) and modularity (to what degree are species divided into subgroups, or modules, with strong/many connections to members of the same module and

weaker/fewer connections to members of other modules?; [111]) have come to the fore as two of the most commonly employed network structures. These are "global" structures, meaning that they account for every node and connection in the network. The scientific interest in nestedness and modularity has spurred a proliferation of metrics to quantify them [8, 12, 41, 111, 136, 139], and to correlate them with population-dynamical properties, such as stability [63, 100, 119], persistence [16], feasibility [125], and robustness [17, 28].

More recently, focus has shifted to "local" structures—structures that involve only a subset of nodes in the network [18, 25, 105, 120]. The most prominent of these is the measure of subgraph ("motif") frequencies. For this measure, networks are interpreted as a collection of "building blocks" each composed of a small set of nodes (usually three). Each possible triplet of three connected nodes is looked at independently and the pattern of the connections between these three nodes is noted. The frequency of each of the thirteen (in the case of directed networks) configurations is tallied [105]. This process is computationally taxing, and grows superexponentially with the size of the subgraphs being considered—extending the consideration to size four requires tallying 199 possible configurations, to five, 9364.

In all of the above measures of network structure, an important consideration is how these structures relate to the underlying system being depicted by the network. Put another way, can we tease apart the various contributions of ecological/evolutionary forces influencing the creation and maintenance of the network and artifacts of underlying mathematical tendencies surrounding stability and biological feasibility [24, 101, 132]? One way this has been approached is by comparing the observed values to an ensemble of randomized versions of the original network. This approach is not without its shortcomings, however. Most notably, it is a challenge to choose a randomization protocol which maximizes interpretability of the focal metric and network type.

2.1.1 Interaction Networks

Though it is clear that each species is involved in a variety of interactions over time (*e.g.* butterflies which pollinate the flowers are also parasitized by wasps), often, networks are constructed to depict a single type (*i.e.* pollination or parasitism) of interaction in isolation, due in part to the specialization of the scientists collecting these data [but see 57, 107, 126].

Two interaction networks, for instance one depicting pollination and another depicting parasitism, have some general features in common. Most noticeably, both are bipartite, *i.e.* the nodes can be divided into two categories such that all interactions are between individuals from different groups. For example, in the case of a pollination network, species might be divided into plants and insects, where each link connects an insect with a plant that it pollinates. Insects do not pollinate one another, nor are plants pollinated by individuals of different species. Bipartite networks are abundant in ecology and have been studied extensively over the past 20 years [103]. Another feature these two networks often have in common is that they are binary/unweighted. While more recently collected networks include some measure of the strength of interaction between species, due in part to the difficulty of collecting large-scale interaction data, most published networks only contain information on the presence or absence of an interaction between a pair of species. Moreover, as both depict ecological phenomena, one might expect them to share some biological constraints, for instance on the maximum number of interactions a particular species can participate in.

Despite these similarities, however, the processes being depicted by pollination and parasitism networks could scarcely be more different. In the case of pollination, both plant and pollinator benefit from the interaction (it is a mutualism), and thus both species have evolutionary incentives to foster the relationship. There are evolutionary forces at work in parasitic interaction webs as well, of course, but in this case the interaction is antagonistic. While the parasite is pursuing the interaction as above, the host is desperate to avoid it.

We might expect these drastically different ecological and evolutionary pressures, often

acting over long timescales, to produce detectable differences in the structure of these networks of interactions [90, 147]. Indeed, there has been some empirical work which supports this theoretical expectation. In an influential paper, **(author?)** [142] argued that differences between networks depicting mutualistic and antagonistic interactions in their levels of the global network structures of nestedness and modularity.

2.1.2 Uninformative Variation

It is important to consider that not all differences between networks are useful for distinguishing between types of interactions. Some obvious differences between networks, for example the absolute number of species or interactions observed, are uninformative with respect to distinguishing network type because they are not a property of the network type *per se*, but rather a result of sampling methodology or common to all types of interactions. This variation in size and number of connections can thus become a distraction from more useful forms of structural variation.

Indeed, the main difficulty in distinguishing networks from one another is not identifying ways in which they differ (no two ecological networks are identical), but rather in sorting through this variation to identify the strongest correlates with respect to the type of interaction being depicted. In this work, we seek to overcome this challenge through utilizing metrics which are uncorrelated with network size and fill. Moreover, we fit the model first to an "ideal" case of consistently replicated datasets spanning a wide range of possible bipartite architectures. The lack of correlation with size and fill can be confirmed through comparing the ability to separate networks in this case to the case where each of these networks has been randomized while preserving the size and fill (section 2.2.2).

2.2 Framing the Challenge

We would like to develop a method which is able to use network structure to classify networks of different interaction types, such that, by measuring a few key aspects of a novel network, we can accurately infer what type of interaction it is depicting. In our view, a successful method will possess three key qualities: generality, specificity, and scalability.

To elaborate on what we mean by these three terms, we provide a worked example as a demonstration of a solution that succeeds in some of these respects, but falls short in others. For this example, we utilize principal component analysis (PCA). PCA is a technique for projecting a high-dimensional space into fewer dimensions by defining linear combinations of the variables constituting the higher dimensions. These new dimensions are called principal components, and usually denoted by "PC#", where "#" is replaced with the number of the principal component when ordered according to the amount of variance explained (table 2.1).

Metric	Formula	PC1	PC2	PC3	PC4	PC5
Configuration model λ_1 ratio	$\log(\lambda_1/\lambda_1^{cm})$	0.149	-0.571	0.735	0.320	-0.097
Erdős Rényi λ_1 ratio	$\log(\lambda_1/\lambda_1^{er})$	0.378	-0.578	-0.212	-0.672	0.163
Marchenko–Pastur λ_2 ratio	$\log\left(\lambda_2/\lambda_2^{mp}\right)$	0.578	0.309	0.135	0.191	0.718
Marchenko–Pastur λ_3 ratio	$\log\left(\lambda_3/\lambda_3^{mp}\right)$	0.536	-0.167	-0.514	0.511	-0.399
λ_2, λ_3 eigenvalue gap ratio	$\log\left(\frac{\lambda_2-\lambda_3}{\lambda_2^{mp}-\lambda_3^{mp}}\right)$	0.462	0.465	0.365	-0.385	-0.538
Percent Explained Variance		**46.7**	**30.9**	**15.5**	**5.0**	**1.9**

Table 2.1: Metrics used in a Principal Component Analysis on 500 non-ecological bipartite networks and their correlation to the five principal components. λ here indicates an eigenvalue, with the subscript indicating its rank in terms of real part (1 is largest), and the superscript indicates that the value is an expectation given the indicated randomization protocol (Supplementary Information).

The figures in this work are intended to serve as a demonstration and are far from an optimal solution. Nevertheless, we provide here a brief outline of our approach to this problem *sensu* a worked example. First, we collected a large database of both ecological and non-ecological networks (see section 2.4). Second, for each network, we measured several aspects of their network structure (table 2.1). Third, we fit a subset of these networks

(section 2.2.1) using a PCA. Fourth, we overlaid various other networks, including randomized versions of the networks (section 2.2.2) and ecological networks (section 2.3) into this principal component space.

PCA and related techniques are popular in Ecology, and have a rich history across a range of fields. For our example, we focus on relatively few metrics of network structure as our input dimensions to the PCA and then focus on the first two principal components when assessing how well we classify networks according to type. Specifically, we calculate the three leading (*i.e.* largest real part) eigenvalues and compare these values to expectations of these values in the case of random matrices of a similar size and fill (Supplementary Information). We chose to focus on eigenvalues as measures of network structure for two main reasons: first, they are relatively fast and easy to calculate, even for large networks; second, leading eigenvalues have been associated with a variety of important network structures, including nestedness [136] and modularity [112]. For attempts with alternative metric selections, and biplots of alternative principal components, see the Supplementary Information.

In each of the following sections, there will be an associated plot where the axes are the first two principal components plotted against one another, and where each point represents a single network in principal component space. These biplots yield a metaphorical map, onto which new webs can be projected. Networks with similar structural metrics will be detectable by points that fall near one another.

2.2.1 Generality

A successful method is one that is general beyond the specific data sets provided to parameterize/test the model. Put simply, if we discover a new network, a successful method should be able to place it in the correct category without needing to re-run the entire method. Practically, we can evaluate this aspect of model performance using a technique that has become standard in machine learning: the division of the total available data into in- and out-of-fit

sets. The first data set (termed the fitting or training set) is used to fit the model, while the second is reserved until after the model is complete and is used to test how well the model works on new data (fig. 2.1).

When using PCA as the underlying model framework, this approach has a second benefit in that it also addresses the known [113] [[others]] constraint of PCA when attempting to classify into groups with differing variance or data availability. Because PCA fits axes that maximally remove variation in the data, it is affected by the data included, and will be biased towards classes with large sample sizes. By rarefying the data so that all types have the same number of networks in the fitting data, we reduce the risk of any one type of interaction overly influencing the results.

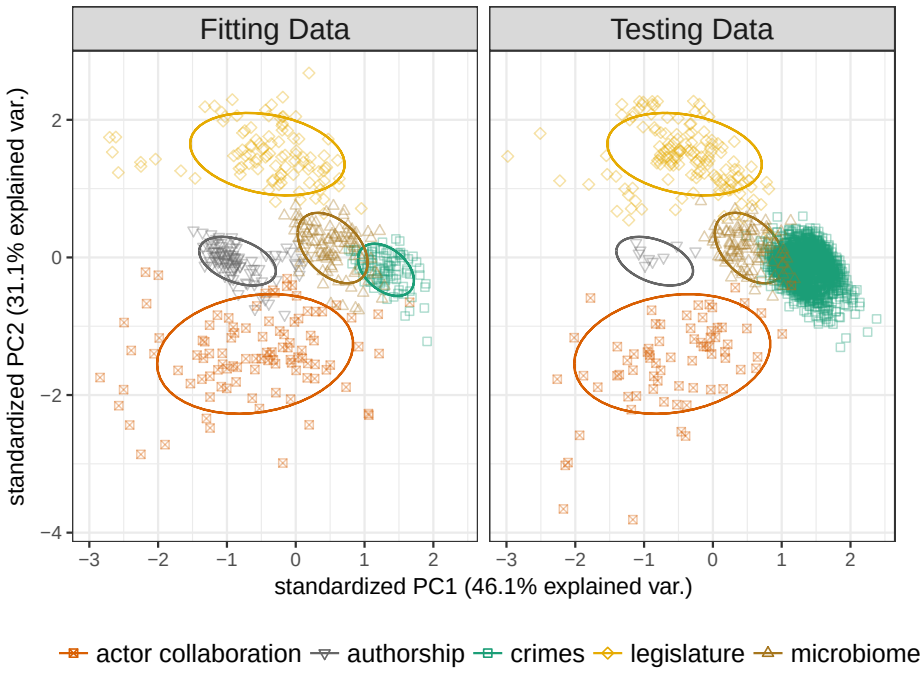


Figure 2.1: A principal component analysis (PCA) biplot of the two most explanatory principal components. Left: the 100 webs of each category used to generate the principal component functions. Right: all remaining webs of each category plotted onto the PCA space as generated in the left pane. The fact that the remaining points generally fall in the same region as those used to fit the model is an indication that the model is not too specific, only fitting incidental differences within our fitting data.

2.2.2 Specificity

While we want the method to be general enough to deal with new datasets, a successful method will also have a degree of specificity. We want the method to be able to identify important differences between ecological network types, not to be able to distinguish a collection of networks *per se*. Framed in terms of uninformative variation, the method should be flexible to variation in structural aspects dictated by sampling effort, such as network size and fill, but it should be intolerant to disruption of the structural inequalities between network types.

We evaluate a method's specificity by creating randomized versions of each of the empirical networks used in the original case and seeing where these fall within the principal component space (fig. 2.2). A successful method will fail to separate the webs following a sufficiently strong and relevant randomization as these networks would lack the informative variation for distinguishing network type.We use two well-studied randomization protocols: Erdős Rényi [55], and a configuration model. citation In the former, the only constrained metrics are the size (*e.g.* number of species) and the number of edges (*e.g.* interactions). In the latter, in addition to these two constraints, the number of connections per node (*i.e.* the degree distribution) is held constant.

in fig. 2.2, we show that, though the empirical networks separate nicely, when each network is randomized according to an Erdős Rényi protocol, the points collapse and we lose separation (fig. 2.2, middle). With a slightly more constrained randomization, such as a configuration model, we still see loss of distinction between most network types, but the authorship networks remain distinct (fig. 2.2, bottom). This suggests that important structural properties that distinguish authorship networks from, for example, crime networks is related in some way to the degree distribution. In this way, we could introduce different randomization schemes (preserving and varying different aspects of network structure) in order to further tease out important differences between network types.

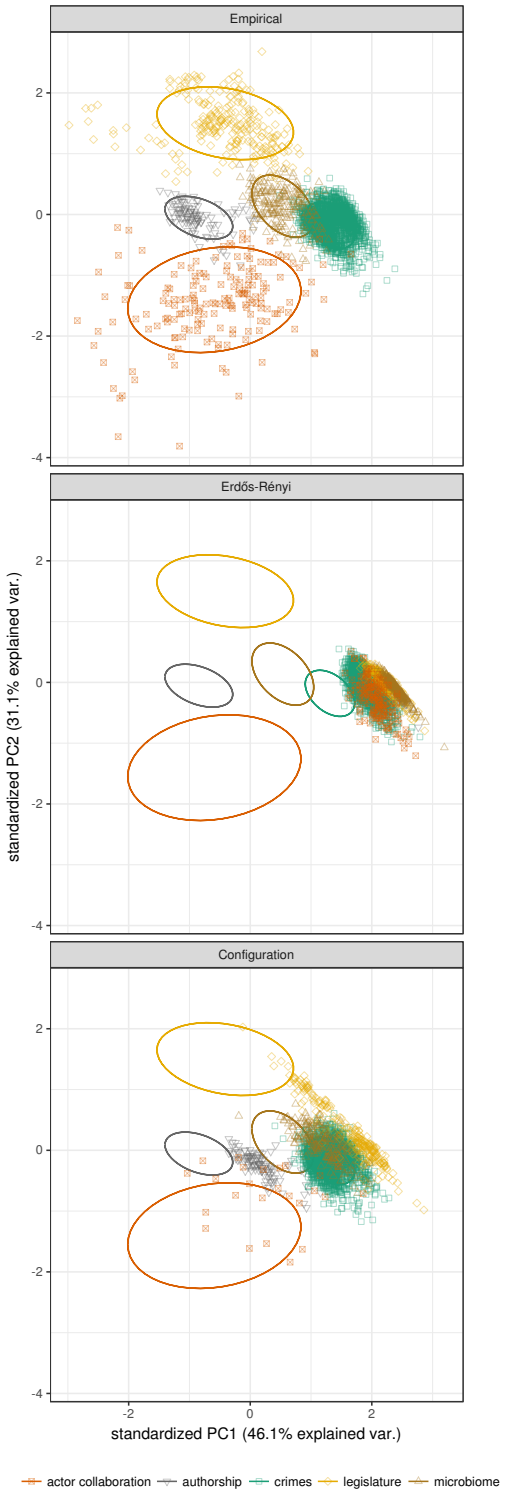


Figure 2.2: PCA biplots of the two most explanatory principal components. The top panel is identical to those in fig. 2.1, but including all data points (those from both panels in fig. 2.1). The middle panel keeps the same axes, but for each point in the top panel (for which a randomization was possible) a new point is plotted corresponding to the placement in the same principal component spaces of an Erdős Rényi randomization of that network. Likewise, the bottom panel shows a point for each configuration model randomization. This randomization is more computationally taxing, hence there are fewer data points in this panel, especially for the actor collaboration networks.

2.2.3 Scalability

Above we state that the method should be able to consistently distinguish networks of different types—this language is intentionally imprecise. We use "types" loosely such that it could refer to distinguishing networks of mutualistic interactions from those depicting antagonisms, or it could refer to distinguishing hummingbird pollination networks from those in which the pollinators are bees, or even just distinguishing ecological networks from anthropogenic ones. This ambiguity is intentional in order to remain agnostic with respect to what the appropriate level of differentiation should be, or even whether there should be one primary level. Analogous to the issue of classifying levels of taxonomy, perhaps there are many different levels of differentiation that can be distinguished. For this reason, a successful method should be able to identify levels of segregation within the data according to some hierarchy of network similarity.

We demonstrate this property in fig. 2.3, where we look more closely at the cluster of crime networks in the top panel of fig. 2.2. Although fig. 2.2 shows that crime networks cluster more closely to one another than to networks of actor collaboration, legislature voting records, or human microbiome networks, fig. 2.3, shows that substructure within this cloud corresponding to networks collected from the same city. The fact that there is much more overlap at this level suggests that perhaps the higher level displayed in fig. 2.2 is a stronger signal, but this was not obvious *a priori*—it had to be revealed through the analysis.

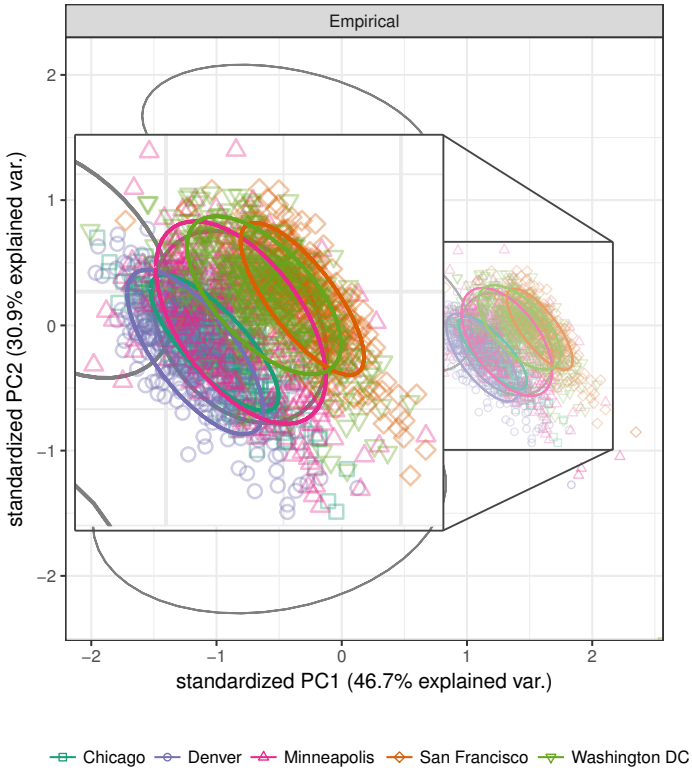


Figure 2.3: Subsetting the points from the top panel of fig. 2.2 to only include the networks of crime locations, we now color the points according to the city from which each crime network was collected. Note that, though all crime networks cluster together fig. 2.2, those from the same city cluster more tightly, albeit with significant overlap between cities. This indicates that there is additional structure beyond that used to disambiguate crime networks from those of authorship, legislature, *etc.* and by looking at these finer differences, more subtle distinctions can be made.

2.3 Ecological Interaction Networks

Having demonstrated some level of success in achieving generality, specificity, and scalability in our method, we now turn to the focus of our challenge: applying the method to ecological interaction networks. For this, we have a collection of networks depicting various forms of antagonistic or mutualistic interactions. We project these networks into the principal component space constructed using the non-ecological networks as the input (fig. 2.4). That is, we do not fit the PCA to the ecological data themselves, but see whether or not the "map" we have constructed also works to distinguish ecological networks from one another. As can

be seen, there is substantial overlap in these ecological networks. This persists even if only highly replicated networks are used, if ecological networks are utilized in the construction of the PCA space, and in all alternative metric sets tested (Supplementary Information).

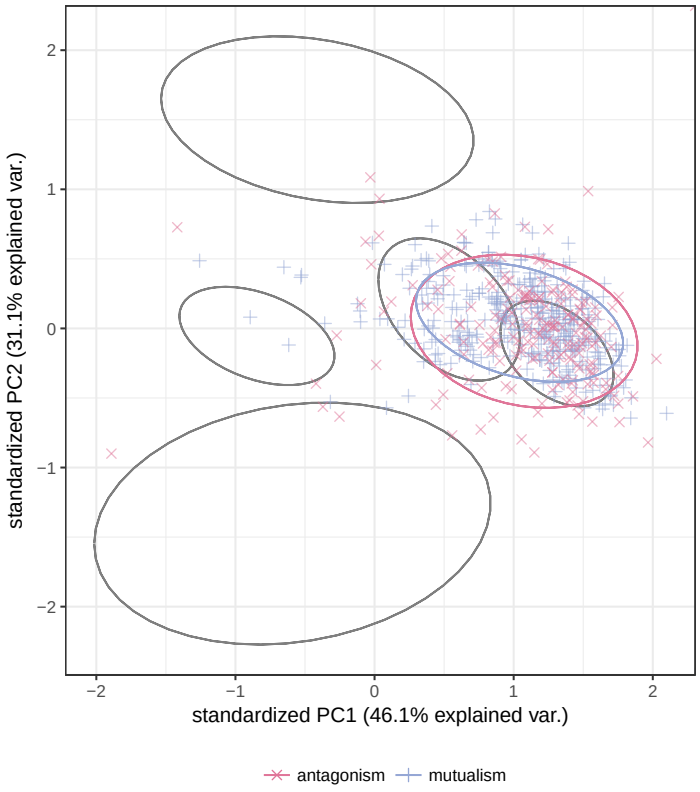


Figure 2.4: Overlaying more than 400 ecological networks of either mutualistic or antagonistic interactions. These networks do not cluster as nicely as the non-ecological webs did, suggesting that there is much more variation within ecological network categories than between them.

Since the non-ecological webs are more consistent in structure and more highly replicated, using just these to fit the PCA is a best case scenario for the PCA to be able to distinguish between the different network types and map out a wide range of possible bipartite networks structures. By just overlaying the ecological webs into this space, we can evaluate clustering in the ecological data without concern that our mapping is particular to the specific ecological datasets we have assembled. The risk in this approach is that the structural differences between ecological networks might differ from those of the data used to generate the principal

component space. This does not appear to be the source of the lack of disambiguation in ecological interaction networks in our case (Supplementary Information).

2.4 Data

All networks analyzed in this work are bipartite. To develop our method, producing the results and figures reported in sections 2.2.1 to 2.2.3, we utilize a large number of highly replicated non-ecological networks assembled from publicly accessible websites. These include:

crimes Daily networks of crime occurrence. Nodes indicate either a crime or neighborhood, and links indicating which crimes took place in which neighborhood on the given day of the year.

legislature Annual networks of legislature voting rolls. Nodes indicate either a legislator or a bill and links indicate positive votes of a given legislator for a given bill.

microbiome Individual microbiome biogeographic networks. Nodes are either location on the body of an individual or operational taxonomic units (OTU) of bacteria living on the individual. Links indicate the presence of a given OTU at a particular bodily location.

authorship Annual co-authorship at the level of academic journals, with each network considering all authors who have published in an indicated journal in a specified year. Nodes indicate either authors or academic journals and links indicate that a given author has published at least one paper in that journal in that year. Note, the journal used to subset the authors will necessarily be connected to all of the authors in the network.

actor collaboration Networks of actor/actress co-participation in films of a given genre in a specified time range. Nodes indicate actors/actresses or films, and links indicate that the actor/actress acted in the specified film.

The majority ($\approx 70\%$) of the collected webs are crime location networks, but each category contains at least 100 webs (table 2.2).

Additionally, to test the method, we have assembled a large number of ecological networks from various online databases. In order to ensure applicability of our method to more than a particular subset of, for example, pollination networks, we assembled a diverse collection of networks for each type of ecological interaction including seed dispersal, ant-plant mutualisms, pollination, parasitism, herbivory, bacteria-phage, *etc.* Altogether, we have amassed more than 500 ecological interaction networks.

Moreover, for each network (ecological or otherwise), we have cataloged some meta-information, including more specific classifications (*e.g.* which city the crime networks are from, which location and treatment corresponds to each ecological interaction web, what sub-type of mutualistic interaction is being described, *etc.*)

Type	Metric	Minimum	Maximum	Mean	Median
Actor Collaboration (172)	Connectance	0.0002	0.072	0.005	0.0029
	Links	258	78,145	11,808.186	6,086.5
	Rows	15	6,675	1,037.75	508
	Cols	239	43,872	7,354.5	4,355.5
Antagonism (197)	Connectance	0.0078	0.9301	0.27	0.2361
	Links	14	5,203	165.472	82
	Rows	5	749	38.67	18
	Cols	5	888	34.274	21
Authorship (109)	Connectance	0.0303	0.0588	0.04	0.0388
	Links	2,186	16,611	9,479.532	9,939
	Rows	86	712	389.349	396
	Cols	438	681	623.138	639
Biogeography (33)	Connectance	0.0154	0.395	0.107	0.054
	Links	78	1,242	260.424	206
	Rows	16	492	172.909	142
	Cols	5	89	30.545	25
Crimes (1815)	Connectance	0.0457	0.6053	0.25	0.2406
	Links	12	470	153.233	120
	Rows	7	77	39.85	33
	Cols	5	64	22.272	21
Legislature (245)	Connectance	0.1715	0.8455	0.593	0.6056
	Links	277	367,956	19,162.89	7,092
	Rows	16	972	167.898	138
	Cols	10	1,427	151.698	87
Microbiome (203)	Connectance	0.0621	0.1891	0.103	0.0982
	Links	7,585	46,936	19,396.808	18,500
	Rows	4,559	16,163	8,625.236	8,442
	Cols	9	44	22.616	23
Mutualism (363)	Connectance	0.0173	0.6875	0.201	0.1905
	Links	12	15,255	183.325	58
	Rows	5	1,044	47.149	27
	Cols	5	647	27.248	15

Table 2.2: Summary statistics on the size and fill of collected networks.

2.5 Discussion

2.5.1 Multiple-interaction Networks

Recent work in networks of ecological interactions has begun to address the reality of species interacting with one another in multiple ways through the consideration of interaction networks containing both mutualistic and antagonistic interactions. While the simple classification problem outlined here does not apply as cleanly to these types of networks, the insights into structural signatures of interaction types in isolation have clear value when generalizing to these more complicated networks. For instance, with the rise of large-scale data collection techniques such as environmental DNA and remote sensing, we are entering an age of data availability that will allow rapid construction of co-occurrence networks. Yet, though we can often infer some sort of interaction between species based on their frequent co-occurrence, discerning the type of interaction can be more difficult. If we are able to determine which structures are typically associated with mutualistic *versus* antagonistic interactions, we might be able to leverage these patterns to infer interaction types between members of co-occurrence networks.

2.5.2 Choice of Metrics

In addition to the choice of underlying model framework (PCA in the provided example), one of the key decisions is which (and how many) metrics to include in the analysis. We have attempted to keep the number of metrics to a minimum for clarity of presentation, however there are an ever-increasing number to choose from. The choice is not inconsequential: including too few metrics will provide insufficient power to distinguish between more similar networks, while too many can lead (in some modeling frameworks) to noise or multicollinearity that can obscure meaningful differences. There are also logistical concerns, as some metrics are much more computationally intensive than others to compute. Certain

estimates of modularity, for instance, require first identifying the appropriate clustering of the species involved—an NP-hard problem [26]. Finally, it might be that certain metrics are more appropriate when discerning between particular pairs of network types. Put another way, perhaps there is no general solution for disambiguating networks *per se*, but rather only techniques for telling one particular type from another.

2.5.3 Reconsidering Old Results

The results presented here are somewhat surprising—why are ecological networks so hard to classify when there have been previous publications claiming success? Using our newer, larger dataset, we can re-investigate the proposed differences between networks of antagonistic and mutualistic interactions. Using a smaller dataset, **(author?)** [142] found significant differences in NODF and modularity between networks depicting mutualisms and antagonisms with respect to a configuration null model. We attempt to repeat and extend this analysis, by looking at additional measures of nestedness, constraining the null model to ensure connectedness, and adding a second null model. We find conflicting support for the use of these measures in the classification of ecological networks. Indeed, both the size and direction of these differences can be influenced by metric type. Moreover, even in the case of significant difference, there is substantial overlap in the metric distributions for different network types. This suggests that this method cannot reliably distinguish network types.

2.5.4 Alternative Methods

There are, of course, a plethora of methods that could be applied to this problem. While we used PCA in this example for ease of presentation, a number of techniques, in particular those in the rapidly growing field of machine learning, are arguably better suited to this task.

Nevertheless, there are features of PCA which make it amenable to this type of prob-

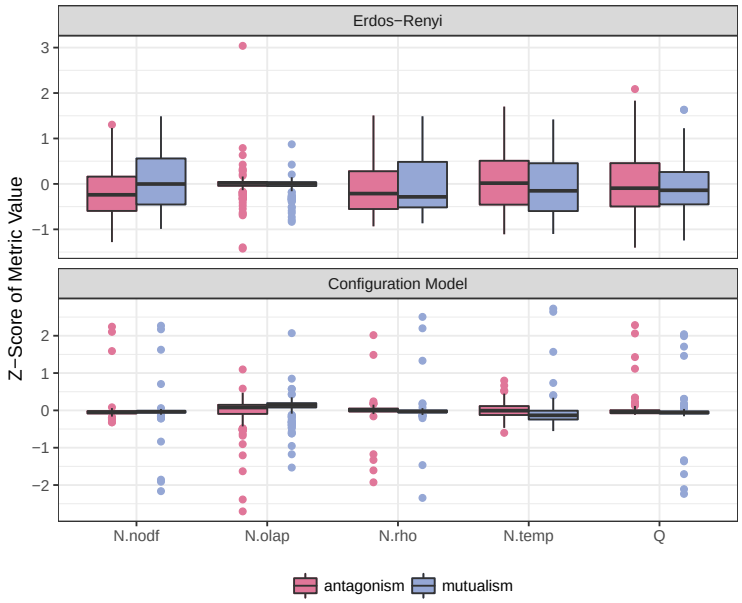


Figure 2.5: We compared networks of ecological interactions based on their value of nestedness and modularity as was done in [142]. Our methodology differs in several respects from that in [142]. We took each empirical web and randomized it 1000 times according to a modified Erdős Rényi (top) or configuration model (bottom) scheme, each modified slightly to guarantee only connected graphs. For each of these randomizations, we calculated nestedness (in a variety of ways) and modularity. We plot here the Z-score of the empirical value for each measure with respect to the 1000 randomizations. For the Erdős Rényi randomizations, the only significant (using a Wilcoxon signed-rank test) difference is that for NODF (N.nodf; 8), showing mutualistic networks to be more nested than antagonistic ones. For the configuration null model, all differences are significant, but the direction depends upon the metric considered, with NODF and overlap nestedness (N.olap; 139) showing mutualisms to be more nested, and nestedness temperature (N.temp; 12) and spectral radius (N.rho; 136) showing antagonisms to be more nested. The configuration model randomization also shows that networks of antagonisms are more modular than those depicting mutualisms.

lem. Perhaps most prominently, as an ordination technique, it attempts to project a high-dimensional space into fewer dimensions. This means we can make figures like those in this work, which serve as a sort of map onto which we can place new networks. Yet, PCA has a number of shortcomings as well. Foremost is that the space constructed varies with the data used. If we were to include more, or different non-ecological networks in the generation phase, we would have a different map. Additionally, we have not run statistics on the clustering we observe in PCA space, introducing an element of subjectivity into the analysis as

stated. Alternative or additional methods could be implemented that incorporate objective grouping to enhance the robustness of any results.

2.5.5 The Meaning of Failure

What if, in spite of our best efforts, we are never able to distinguish ecological interaction networks of differing types from one another? Despite being a "negative" result, this type of failure would still inform our understanding of ecological networks in a positive way. It would indicate that binary network structure is not the level upon which the known differences in ecological and evolutionary pressures act, and would suggest that further studies comparing interaction networks should look to alternative sources of variation, such as interaction strengths or temporal variation. As more networks including interaction strength are published, new measures of structure might be revealed as key distinguishing characteristics. Importantly, several of the metrics we use in this demonstration, including both modularity and nestedness, have only recently begun being expanded to account for non-binary link strength [for example 136]. Other metrics (such as subgraph distributions) are still completely restricted to the binary case.

Alternatively, perhaps these interaction networks, which are select subsets of the larger network enmeshing ecological communities, are not distinct *per se*, but rather in the ways in which they combine to form the larger web. Further exploration into the properties of these multi-layer networks will be key to clarifying any biases introduced over the past decades by treating different types of interactions as independent of one another.

We hope that the publication of this work and accompanying datasets and code-base will spur further investigation into this interesting computational challenge. The solution to this problem would serve as a substantial step forward in our understanding of ecological networks as well as a confirmation of the value of network science in ecology. Discovery of which structural aspects differ between ecological networks of different interaction types

would provide valuable insights into both the underlying evolutionary processes of network formation as well as into the relevant ecological variables that differentially affect these systems. Aspects which are found to be similar between networks is also valuable information, indicating convergent forces or, at the very least, indicating areas in which empiricist need not direct their efforts.

2.6 Acknowledgements

This work benefited from the feedback of Mercedes Pascual, Trevor Price, Sarah Cobey, and Amos Maritan. We would also like to thank Jacopo Grilli, Phillip Staniczenko, Gyuri Barabás, Samraat Pawar, and Liz Sander for conversations regarding the analysis. This work utilized the computational resources of the Center for Research Informatics' Gardner HPC cluster at the University of Chicago (http://cri.uchicago.edu). Funding includes U.S. Department of Education grant P200A150101 and NSF DEB-1148867.

Supplemental Information

2.7 Randomizations

For each network collected, we created two randomized versions. The first was a network containing exactly the same number of nodes and links, but having the links rearranged such that any two nodes had the same probability of being connected. This type of randomization is termed Erds-Rényi. We made one change to the traditional formuation of Erds-Rényi randomizations, however, in that we ensured the network would be fully connected (*i.e.* that there is at least one path between any two pairs of nodes). We did this by first drawing a random spanning tree through the network and then randomly assigning the remaining links (to bring the total number up to that of the original network) as described above. This should not bias our randomizations, as this spanning tree is a necessary backbone of any connected network.

For the configuration model, we utilized a fast, unbiased randomization algorithm termed the "curveball" algoritm[138]. This algorithm rearranges the links in a network in such a way as to preserve the degree distribution (*i.e.* the number of links connected to each node).

It is more compuationally taxing to ensure connectedness in the configuration model randomized networks, so for this case, we simply discarded disconnected graphs and re-randomized from the original network again until we found a connected one. For several networks, despite more than 1000 attempts, we were still unable to find a connected network. This is often due to very low connectance, which reduces the number of possible connected networks with the given degree distribution. For these cases, we simply omitted those networks in the configuration model panel of figures. Due to the unbiased nature of the curveball algorithm, this process of discarding and re-randomizing is not expected to introduce bias into our analysis.

2.8 Metrics Collected

For each network, including the two randomized versions of each empirical network, we measured a number of network structural properties. These included (in addition to network size (number of rows and columns in the matrix representation of the bipartite network) and fill (absolute number of links as well as the ratio of the number of links to the number of possible links, termed "connectance"):

1. eigenvalues
   (a) the three right-most (maximum real part) eigenvalues of the matrix representation of the network)
   (b) the three right-most (maximum real part) eigenvalues of the laplacian matrix associated with the network)
   (c) an analytical estimate of the leading eigenvalue of a random network with the same size and fill

$$\lambda_1^{er} = \frac{1+\frac{(n_{cols}-1)(n_{links}-n_{rows})}{(n_{cols}-1)n_{rows}}\left(3+\frac{(n_{cols}-2)(n_{links}-n_{rows})}{(n_{cols}-1)n_{rows}}\right)}{1+\frac{(n_{cols}-1)(n_{links}-n_{rows})}{(n_{cols}-1)n_{rows}}} * \frac{1+\frac{(n_{rows}-1)(n_{links}-n_{cols})}{(n_{rows}-1)n_{cols}}\left(3+\frac{(n_{rows}-2)(n_{links}-n_{cols})}{(n_{rows}-1)n_{cols}}\right)}{1+\frac{(n_{rows}-1)(n_{links}-n_{cols})}{(n_{rows}-1)n_{cols}}}$$

   (d) an analytical estimate of the leading eigenvalue of a random network with the same degree distribution

$$\lambda_1^{cm} = \sqrt{\frac{\overline{\boldsymbol{d}_{col}^2}}{\overline{\boldsymbol{d}}_{col}} \times \frac{\overline{\boldsymbol{d}_{row}^2}}{\overline{\boldsymbol{d}}_{row}}}$$

   (e) an analytical estimate of the leading eigenvalue of a random regular (i.e. all nodes of the same class have the same degree) network with the same size and fill

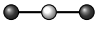
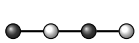
$$\lambda_1^{reg} = \frac{\overline{\boldsymbol{d}_{col}} n_{col} + \overline{\boldsymbol{d}_{row}} n_{row}}{n_{row} n_{col}}$$

   (f) analytical estimates of the second and third leading eigenvalues expected from

the Marchenko-Pastur distribution

$$\lambda_2^{mp} = \frac{n_{row} - (\lambda_1^{cm})^2}{n_{col}} \left(1 + \sqrt{\frac{n_{col}}{n_{row}}}\right)^2$$

$$\lambda_3^{mp} = \lambda_2^{mp} - \frac{\sqrt[3]{n_{row}\pi^2} * C * (1-C) * \left(1 + \sqrt{\frac{n_{col}}{n_{row}}}\right)^{4/3}}{(n_{col}/n_{row})^{1/6}}$$

2. the modularity[111] of the network (where computationally feasible)
3. the nestedness of the network (where computationally feasible)
   (a) using nestedness temperature[12]
   (b) using NODF[8]
   (c) using the overlap measure of nestedness[139]
4. counts of three-node subgraphs present in the network
   (a) H2: $\sum_i \binom{d_i}{2}$ 
   (b) H3: $(\boldsymbol{d}_{row} - 1) \times \mathbf{B} \times (\boldsymbol{d}_{col} - 1)$ 
   (c) H4: $\sum_i \binom{d_i}{3}$ 
   (d) H17: $\sum_i \binom{d_i}{4}$ 
5. measures of network centralization using
   (a) betweenness centrality
   (b) closeness centrality
   (c) eigenvector centrality
6. network diameter
7. average path length between nodes of the network
8. degree heterogeneity in each of the row/column nodes
9. algebraic connectance
10. inverse participation ratio for each of the three leading eigenvalues

2.9 Alternative Metrics

For the figures in the main text, we use a subset of these metrics for clarity of demonstration.

In this section, we provide analogous figures to those in the main text including all metrics

for comparison. Note that, because several of these metrics have strong correlations with size and fill, we do not see the desired collapsing behavior in the randomized networks (figs. 2.6 to 2.9).

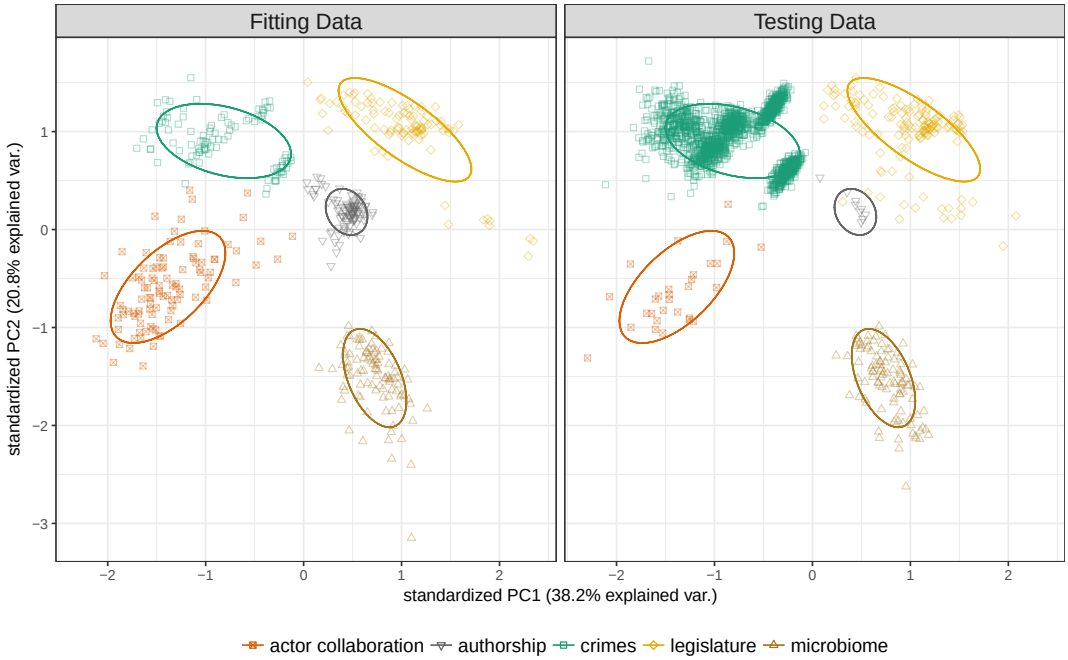


Figure 2.6: As fig. 2.1, but utilizing all collected structural measures as input for the PCA.

Note that, while we obtain even better separation for the non-ecological networks, the mutualistic and antagonistic networks still fall nearly on top of one another.

As an alternative to Figure 5 in the main text, we could investigate the role of Nestedness and Modularity in distinguishing ecological networks using PCA. Here, we ran a pca just on the ecological data using just measures of nestedness and modularity as input values. As before, there is no separation between interaction type (fig. 2.10).

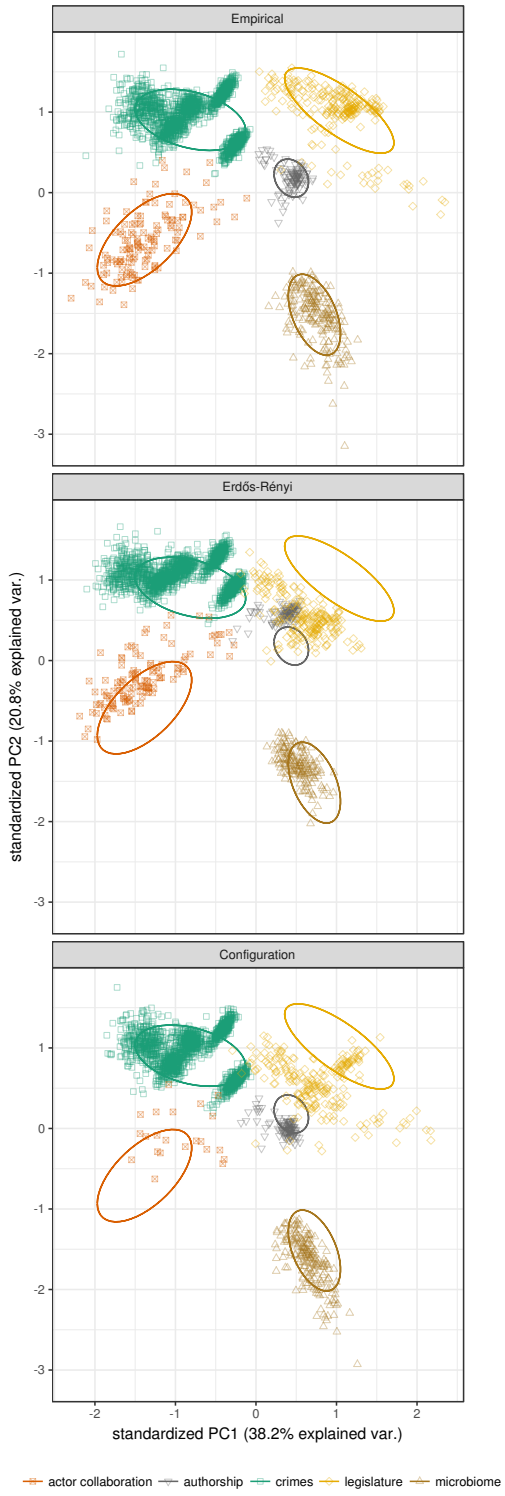


Figure 2.7: As fig. 2.2, but utilizing all collected structural measures as input for the PCA.

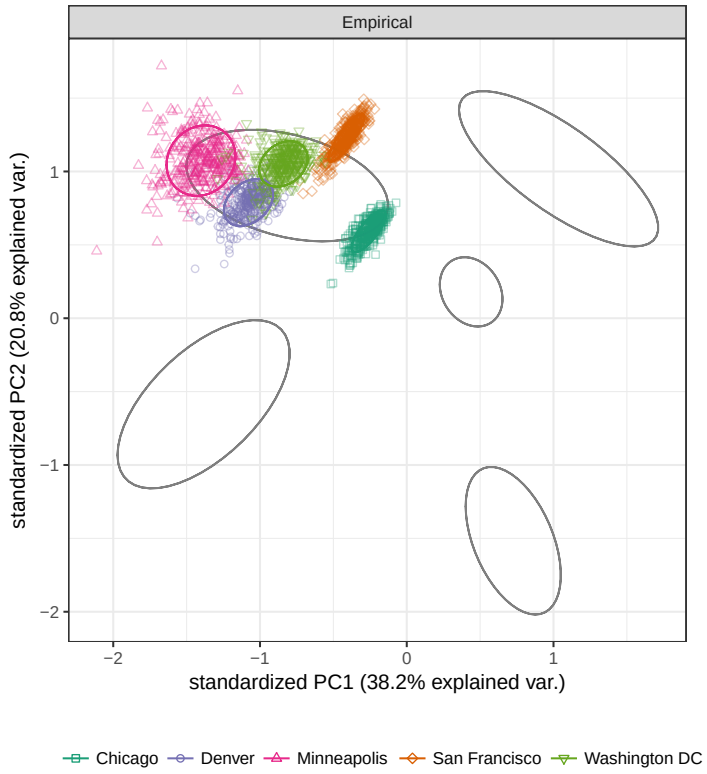


Figure 2.8: As fig. 2.3, but utilizing all collected structural measures as input for the PCA.

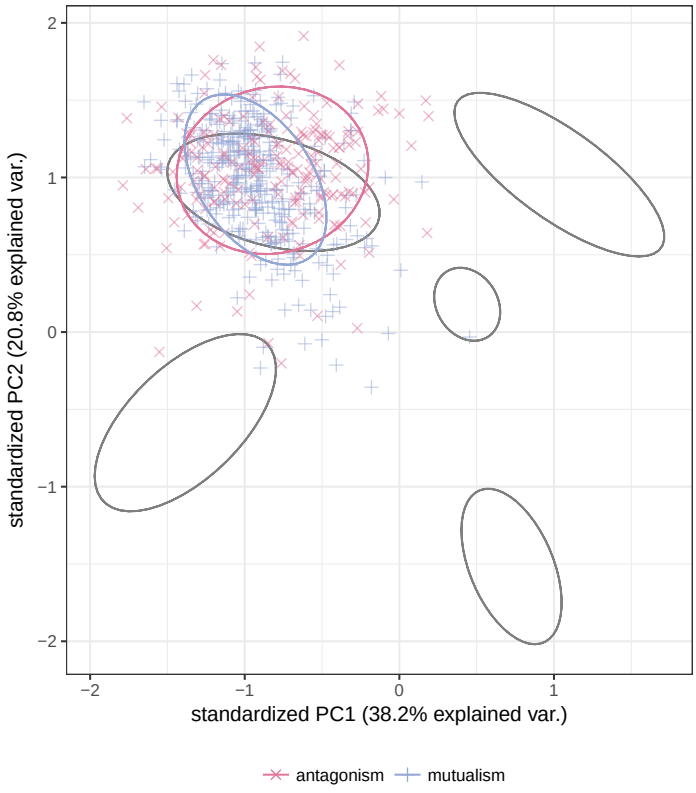


Figure 2.9: As fig. 2.4, but utilizing all collected structural measures as input for the PCA.

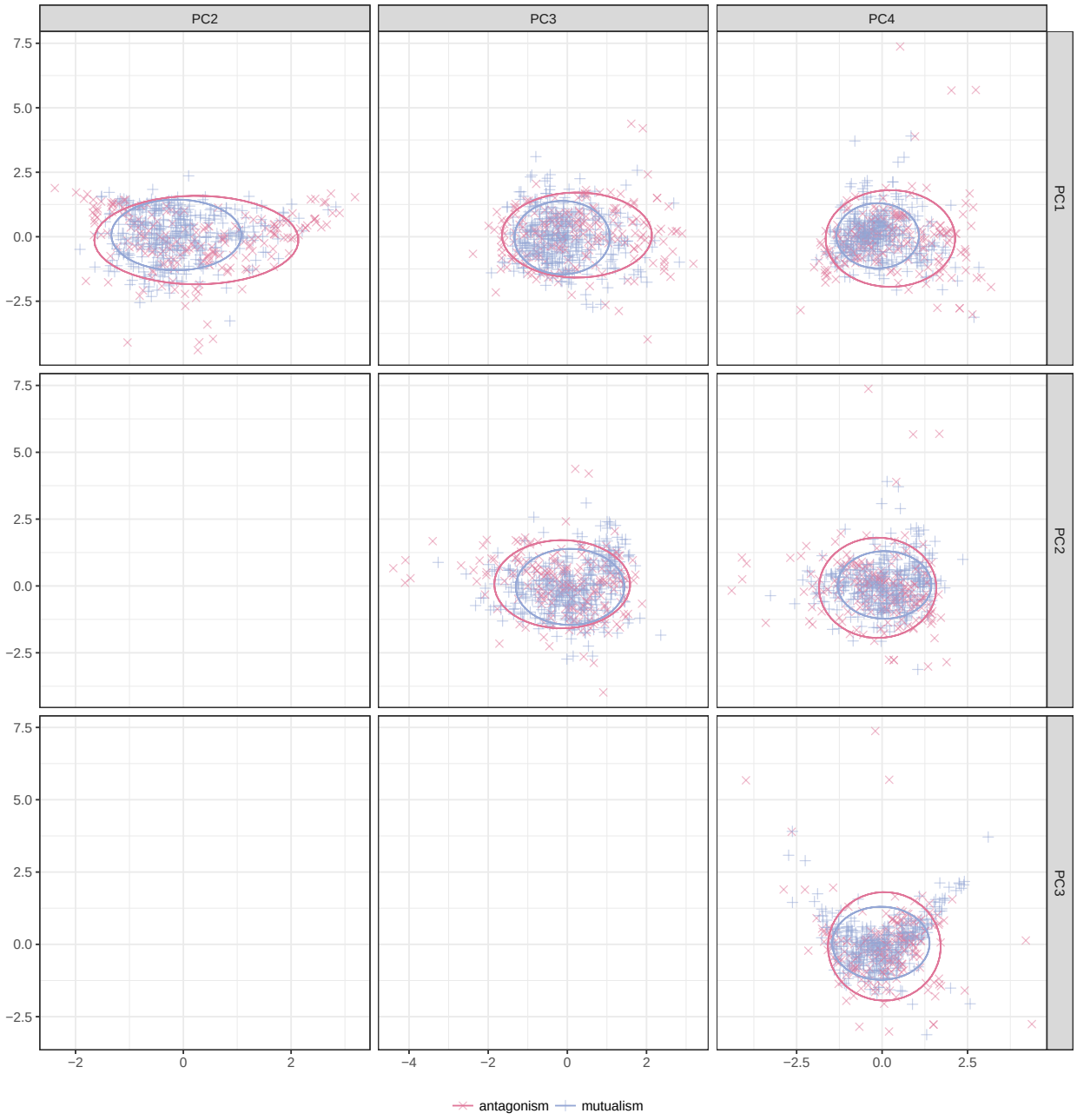


Figure 2.10: All pairwise combinations of principal components from a PCA run on the ecological data, but only using nestedness and modularity as inputs.

2.10 Alternative Ecological Data

One possible explanation for this failure to disambiguate ecological networks is that of data quality. Here we are using hundreds of ecological networks from a variety of sources. These networks were assembled using a variety of methods and span a large geographic and temporal range.

Here, we re-plot the ecological networks using only cases where there are multiple networks published by the same group. Often, these are networks in similar environmental conditions and were collected using similar methods over a relatively constrained amount of time. Moreover, because networks that meet these criteria also tend to be more recently published, they also tend to be larger, on average, than older, lower-quality networks (fig. 2.11).

Though there is better separation, we still see significant overlap and thus much progress to be made.

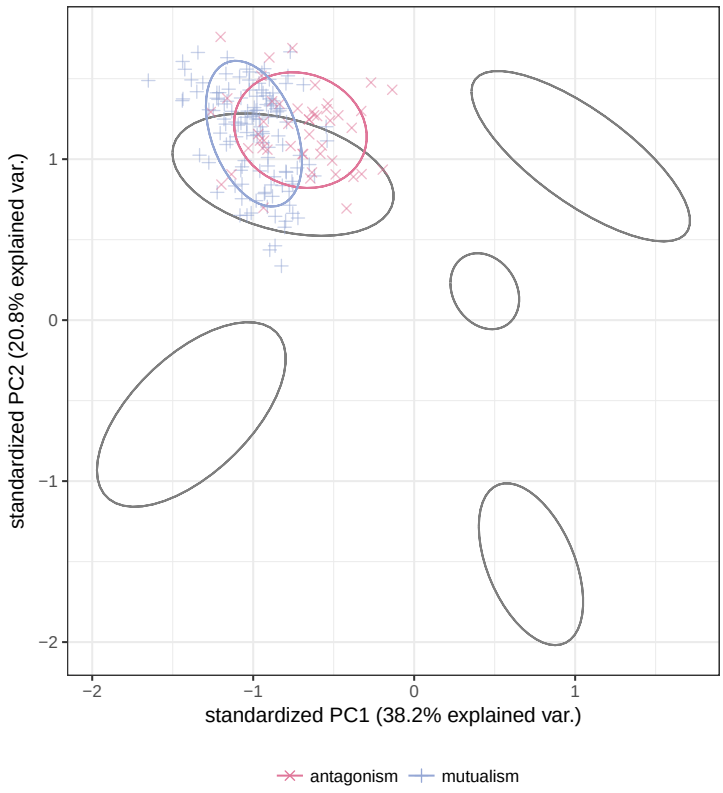


Figure 2.11: Overlaying ecological data onto the full PCA as in fig. 2.9, but restricting datasets to those with clear replication.

2.11 More Dimensions

Perhaps there is separation, just not within the first two principal component dimensions. Here we plot each pairwise combination of dimensions for the PCA from the main text (fig. 2.12; note that here we are only plotting the ecological networks, but we are still projecting them into the space defined by the PCA run on the non-ecological networks), and the same for each pairwise combination of the first five dimensions of the full PCA run here in the supplement (fig. 2.13). For all cases, the level of overlap is substantially greater than that seen between the non-ecological networks.

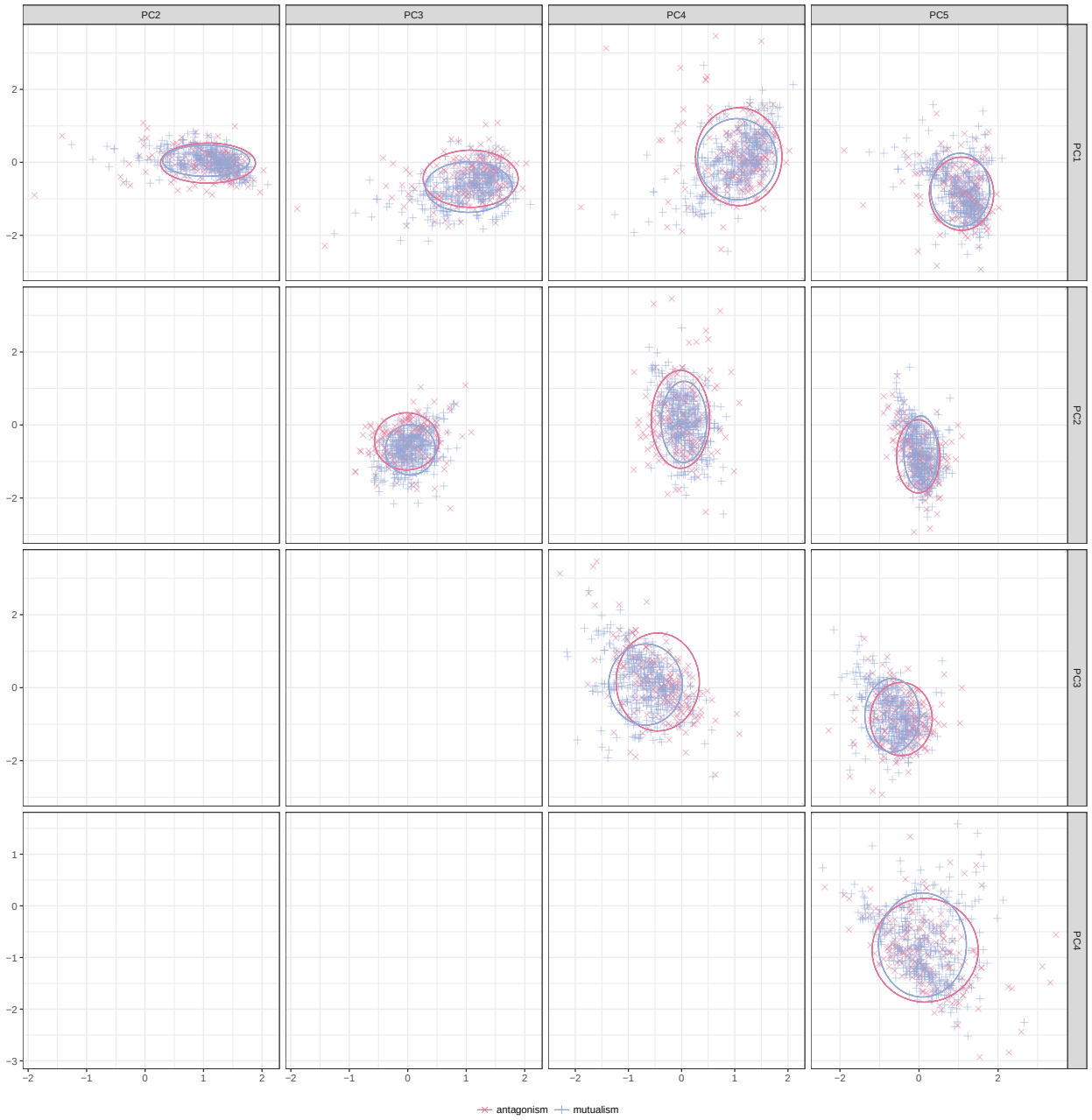


Figure 2.12: All pairwise combinations of principal components as in fig. 2.10, but using the PCA definined in the main text.

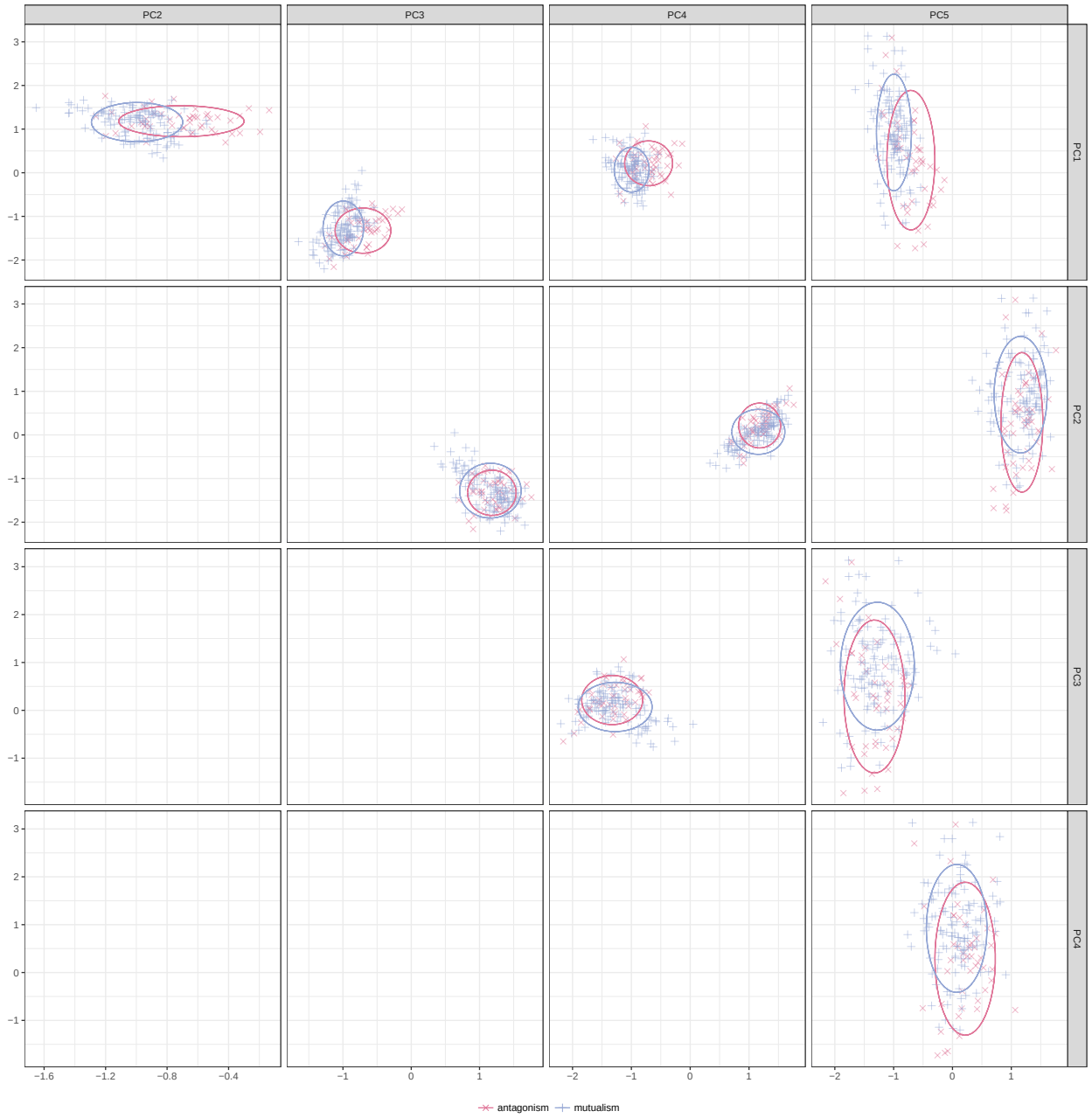


Figure 2.13: All pairwise combinations of principal components as in fig. 2.10, but using the full PCA definined in fig. 2.6.

2.12 Including Ecological Data in the Fitting

Perhaps the lack of separation is due to our process of not including the ecological data in the model construction. We did this in order to both make our model more general, but also because including the ecological data in the PCA makes the PCA dependent upon which ecological data is included. In our view, this reduces the robustness of any results, but we here include those results for consideration (again mirroring figures from the main text; figs. 2.14 and 2.15).

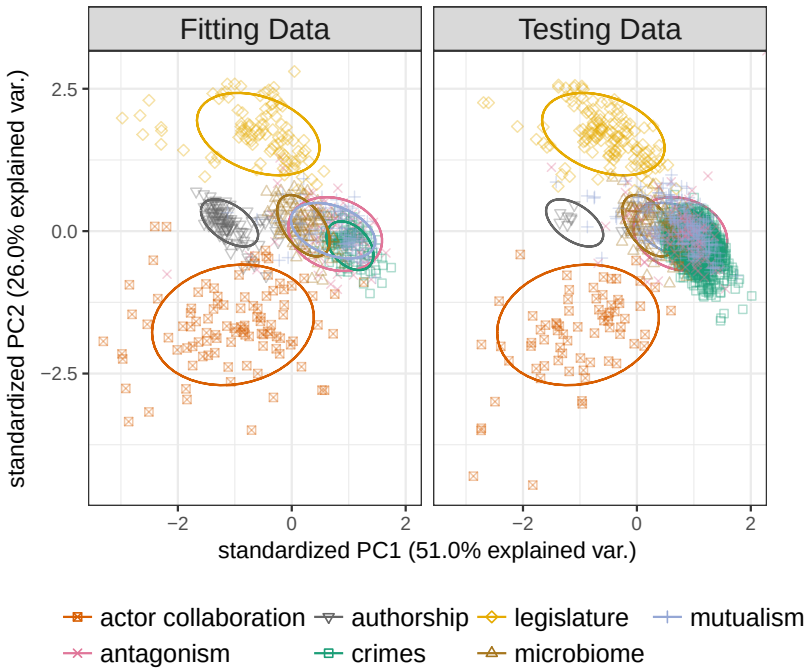


Figure 2.14: As fig. 2.1, but including ecological data as some of the input networks for the PCA.

As before, there is no separation between mutualistic and antagonisitic networks, which fall over the microbiome and crime networks.

Similar results are obtained when restricting to the "high quality" ecological datasets (figs. 2.16 and 2.17) and when using the full set of metrics collected figs. 2.18 and 2.19).

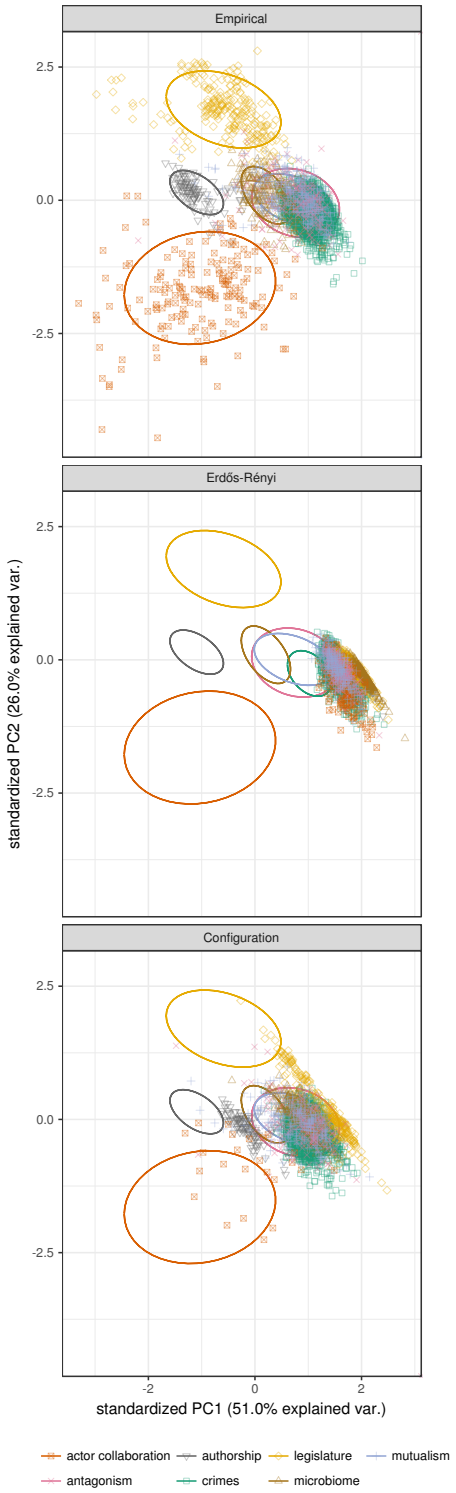


Figure 2.15: As fig. 2.2, but including ecological data as some of the input networks for the PCA.

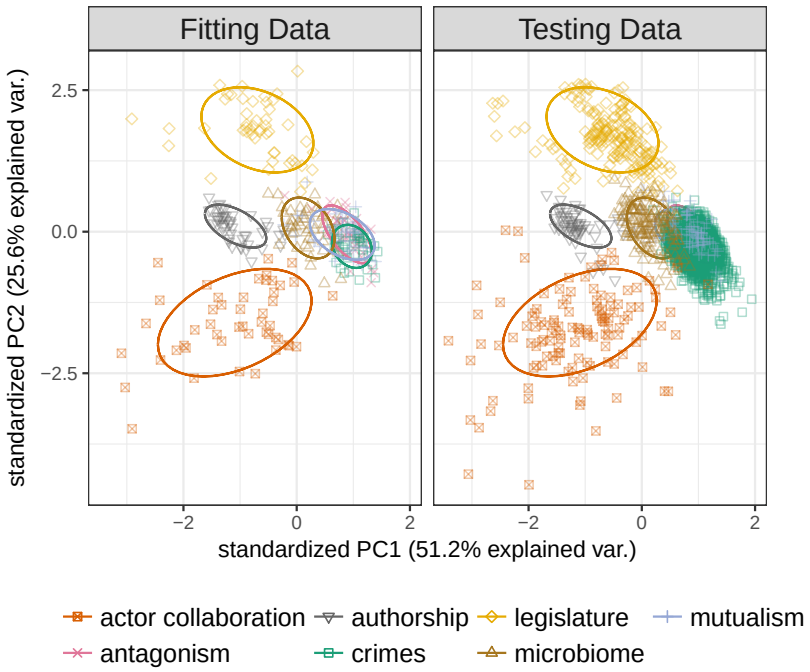


Figure 2.16: As fig. 2.14 but restricted to ecological data to those datasets with clear replication.

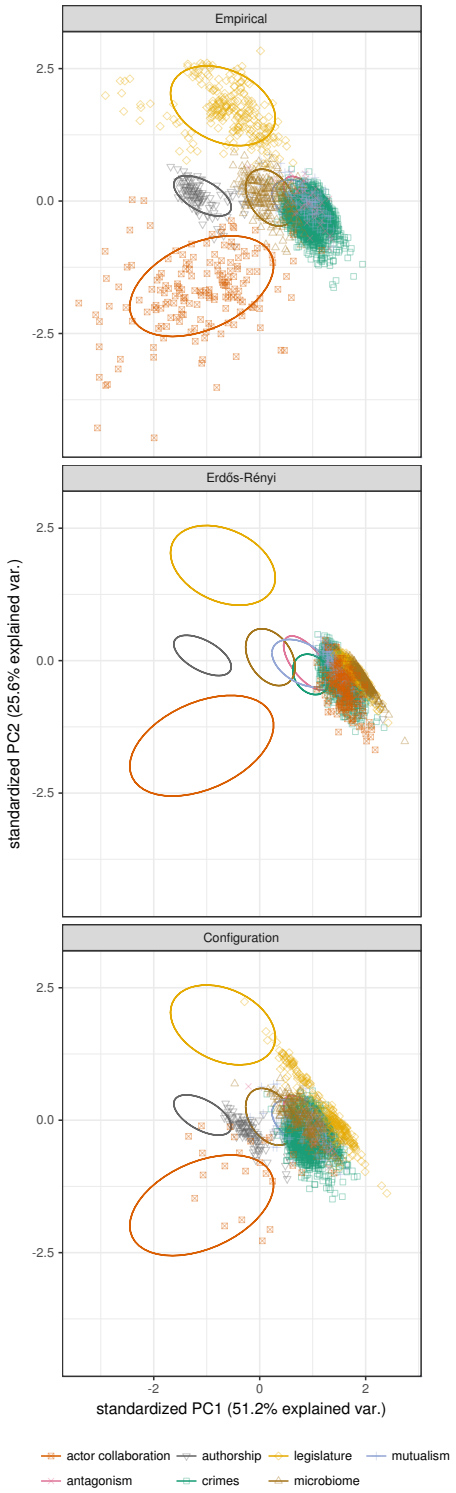


Figure 2.17: As fig. 2.15 but restricted to ecological data to those datasets with clear replication.

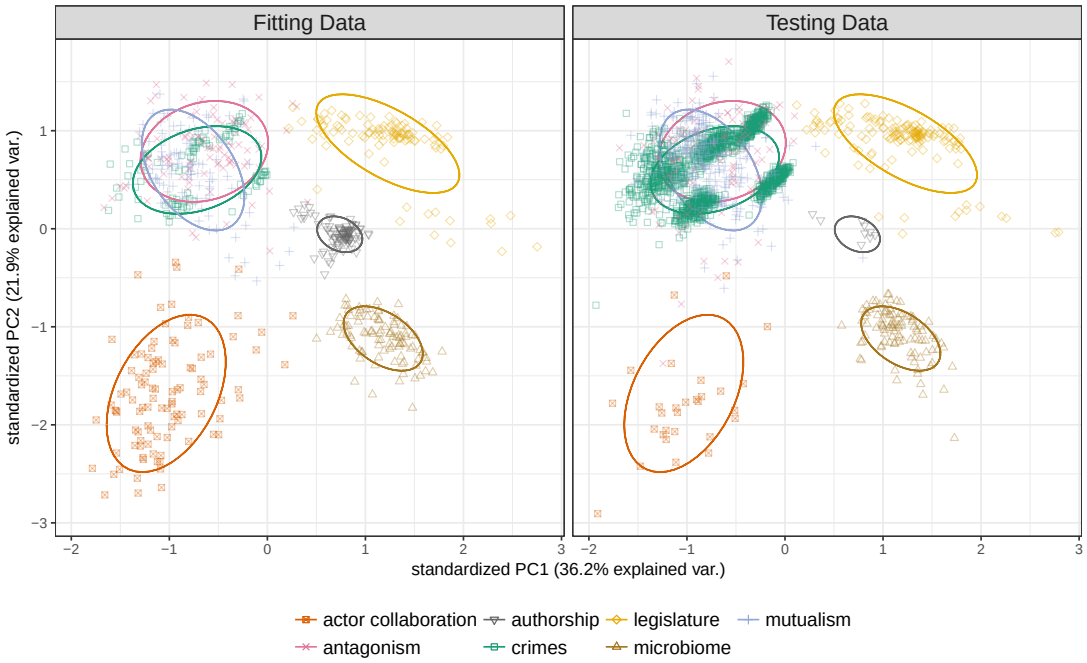


Figure 2.18: As fig. 2.14 but utilizing all structural metrics as input to the PCA.

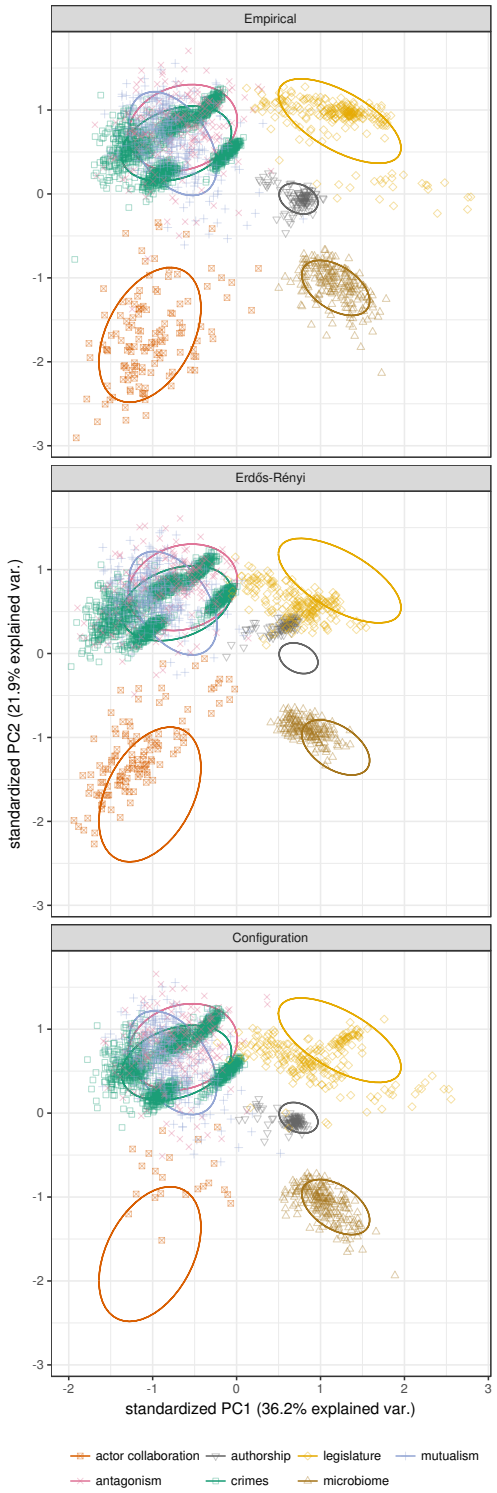


Figure 2.19: As fig. 2.15 but utilizing all structural metrics as input to the PCA.

CHAPTER 3

UNDERSTANDING THE ROLE OF PARASITES IN FOOD WEBS USING THE GROUP MODEL

Matthew J. Michalska-Smith[1], Elizabeth L. Sander[1], Mercedes Pascual & Stefano Allesina

The following was originally published in *The Journal of Animal Ecology* in April 2018.

MichalskaSmith, Matthew J., Elizabeth L. Sander, Mercedes Pascual, and Stefano Allesina. "Understanding the role of parasites in food webs using the group model." Journal of Animal Ecology 87, no. 3 (2018): 790-800.

1. **Parasites are ubiquitous and have been shown to influence macroscopic measures of ecological network structure, such as connectance and robustness, as well as local structure, such as subgraph frequencies. Nevertheless, they are often underrepresented in ecological studies due to their small size and often complex life cycles.**

2. **We consider whether or not parasites play structurally unique roles in ecological networks; that is, can we distinguish parasites from other species using network structure alone?**

3. **We partition the species in a community statistically using the group model, and we test whether or not parasites tend to cluster in their own groups, using a measure of "imbalance."**

1. These authors contributed equally to this work.

4. **We find that parasites form highly imbalanced groups, and that concomitant predation, in which a predator consumes a prey and its parasites, but not the number of interactions, improves the group model's ability to distinguish parasites from non-parasites.**

5. **This work demonstrates that parasites and non-parasites interact in networks in statistically distinct ways, and that these differences are partly, but not entirely, due to the existence of concomitant predation.**

3.1 Introduction

Parasites are ecologically significant players in many communities, and several authors have urged the incorporation of these species into ecological networks [81, 96, 97]. While many modern networks are well resolved with respect to most free living species, parasites are often excluded entirely. In networks which do incorporate parasites, these species affect several general aspects of food web structure; for example, increasing the proportion of possible links that are observed (connectance), the number of consumptive links between the highest and lowest trophic levels (trophic chain length), and, trivially, the number of species (richness) [11, 46, 73, 82, 146]. Conversely, many parasite species may decrease network robustness—usually quantified as the proportion of species lost following a given number of primary extinctions [48]—because highly complex and specialized life cycles may make them prone to secondary extinction in response to host removal [81].

Parasites also affect local network structure. In a comprehensive analysis, Dunne *et al.* [46] show that parasites change the relative frequency of certain network subgraphs, and tend to have niches which are broader, but contain more gaps, than predators in aquatic food webs. The intimate connection between parasites and their hosts results in another major effect. Concomitant predation, wherein a predator consumes both the prey and its parasites, is sometimes a necessary part of parasitic life cycles, and may also be a relevant source of

biomass for predators, given the high biomass of parasites in ecological communities [80]. These interactions increase the connectance of the network and affect the degree, *i.e.* the number of consumptive interactions a given species is involved in, of both parasites and their incidental predators.

There is some disagreement in the literature about how exactly parasite degree differs from that of free-living species. Parasites are often highly specialized (*e.g.* 9, 51), suggesting that parasites might have a lower in-degree (number of prey/hosts) than free-living predators. However, parasites have been found to increase overall connectance, depending on how the calculation is done [82]. This would suggest that parasites have, on average, more interactions than their free-living counterparts. Concomitant predation will also affect degree, increasing the in-degree of free-living predators and the out-degree (number of predators) of parasites. Taken together, these observations point toward degree as a structurally distinguishing feature of parasitic species, especially when in-degree and out-degree are considered separately.

Ecological networks often contain hundreds of species and thousands of consumer-resource interactions. To study these complex networks, it is useful to understand the general roles species play in the community. Species roles are sometimes classified based on phylogenetic (*e.g.* a terrestrial ungulate can be assumed to be herbivorous) or *a priori* trophic strategies (e.g. an herbivore consumes exclusively primary producers by definition), but they may also be identified statistically. *Ecologically equivalent* species (also known as trophic species) have the same set of predators and prey, and therefore play identical roles in the network structure [47, 91]. This concept can be relaxed and generalized using the group model [7], which organizes species into groups, such that species in a group tend to eat and be eaten by members of the same other groups of species (fig. 3.1). Equivalent to the stochastic blockmodel from the social science literature [76, 131], the group model uses network structure to form groups that often have straightforward ecological interpretations [127]. Indeed the

species roles defined by the group model are essentially functional groups, in that species within a group tend to interact with the same sets of species in the same way.

Using the group model, we consider the structural distinction between parasites and free-living predators. Whether or not parasites alter general network metrics, if the patterns of their interactions are structurally unique within the network [40], then the groups identified by the model should reflect this distinction. For our study, we consider a set of large food webs that include information on both parasites and free-living species (66, 72, 108, 143, 156; table 3.2), and we quantify how well the network's group structure matches broad trophic strategies. In general, it is difficult to identify the specific ecological drivers that contribute to the group structure. Here, we are able to isolate the effects of two ecological factors, degree and concomitant predation, which may influence group structure and how well it corresponds to the trophic strategies we expect. To examine the effect of degree, we compare the groupings found using a standard group model, and a variant of the model that removes the effect of degree on the group structure. To study concomitant predation, we compare groupings found when concomitant links are included and excluded.

We find that parasites perform unique roles in ecological communities, whether or not concomitant links are included and whether or not the model is corrected for degree. However, although the presence of concomitant links improves the model's ability to distinguish parasites from free-living species, degree heterogeneity does not.

3.2 Materials and methods

3.2.1 Data

We analyzed the seven well-resolved marine and estuarine food webs described in [45, 66, 72, 108, 143, 156]. We analyzed two versions of each network: one which includes concomitant links, and one which excludes them. Concomitant interactions are inferred links based on

the assumption that predators eat all parasite species of their prey Dunne *et al.* [46]. For all webs, parasites with complex life-cycles had their various life-stages aggregated into a single node.

Species were classified into four trophic strategies: primary producer, herbivore, predator, or parasite. Primary producers were identified as any species with no prey. Herbivores were identified as species which consumed only primary producers. Parasites were identified based on Dunne *et al.* [46]. All other species were labeled as predators; therefore, this group contains both carnivores and omnivores.

3.2.2 Group Model

Metrics of categorizing network structure are common in analyses of ecological networks. One of the most popular of such metrics is modularity [111], which evaluates the presence of compartments within a community. These compartments contain individuals/species which interact more strongly with fellow members of their compartment than they do with members of other compartments (*e.g.* benthic versus pelagic species or flowers which bloom in the early versus late summer). This results in a structure of dense blocks along the diagonal of a matrix when properly ordered (fig. 3.1). The group model can be thought of as a generalization of modularity, in which compartments are not defined exclusively by strong within-compartment connections, but rather by patterns of strong connections between compartments as well (*e.g.* between herbivores and primary producers). Note that the group model does not exclude the possibility of strong connections with one's own compartment, such that modularity is a subset of the possible groupings identified by the group model.

The group model provides a likelihood-based framework to calculate how well a specific grouping fits the observed network structure. High-likelihood groupings will tend to have groups which act as functional groups, that is, species within a group tend to eat and be

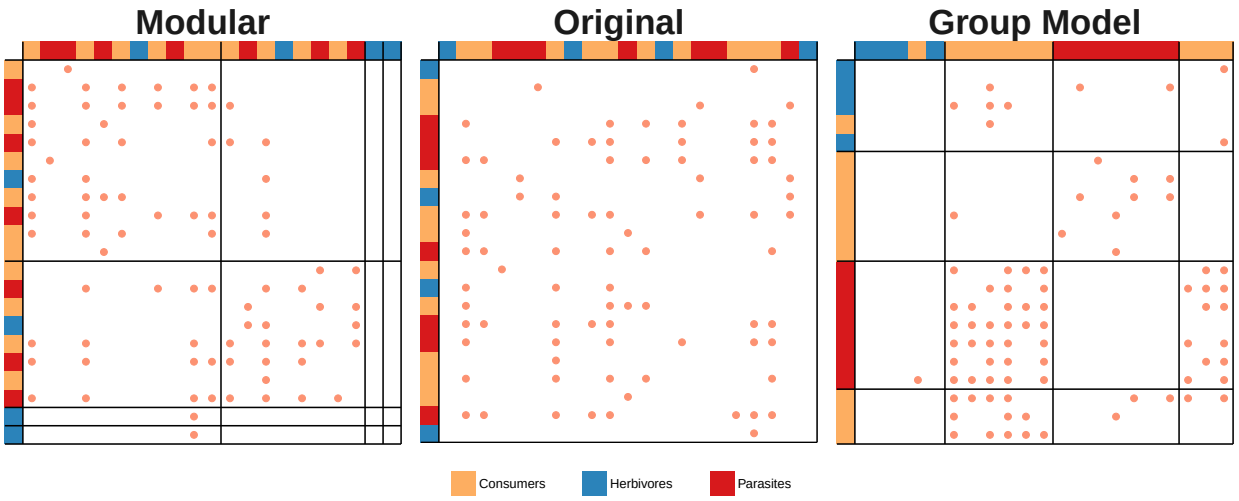


Figure 3.1: Example of a group-model-produced grouping of an empirical adjacency matrix. The center matrix represents a subset of one of the empirical webs used in this analysis. To the right and left are the same subset, but with the rows/columns re-ordered to maximize the modularity [111] (left) or the group model (right). Both modularity and the group model attempt to condense the links into groups of species that are strongly connected, producing a pattern in which the matrix is divided into areas with either very high or very low connectance. Note that, though the links appear randomly distributed before sorting, applying a walk-trap algorithm to find modules partitions the matrix into four groups, with links concentrated within modules (in blocks along the diagonal). Applying the group model (unrestricted for number of groups) also finds four groups, but partitions the matrix differently, creating more strongly connected blocks that are often (though not necessarily) off the diagonal. In all cases, links are indicated by orange dots and groupings by black lines. The trophic species of each node is indicated with a colored box along the margins.

eaten by the same other groups. Consider a food web with S species and L links, represented by directed adjacency matrix A. Modelling the network as an Erdős Rényi random graph with connectance (the proportion of possible links that are realized) c, the likelihood of obtaining A can be given by:

$$Pr(A(S,L)|c) = c^{L}(1-c)^{S^2-L} \tag{3.1}$$

The likelihood is maximized when $\hat{c} = \frac{L}{S^2}$, the observed connectance. Using a partition (grouping) G containing g groups, we can split the network into a series of blocks, where each block represents all of the interactions from group r to group s, and where the groups contain S_r and S_s species, respectively. A block has L_{rs} links and connectance c_{rs} (note that

because the network is directed, block rs is distinct from block sr). Then the full likelihood can be calculated as the product of the likelihoods of each individual block, as follows:

$$Pr(A(S,L)|c_{rs}, r, s \in 1:g) = \prod_{r=1}^{g}\prod_{s=1}^{g} c_{rs}^{L_{rs}}(1-c_{rs})^{S_r S_s - L_{rs}} \tag{3.2}$$

which is maximized when $\hat{c}_{rs} = \frac{L_{rs}}{S_r S_s}$ for every r, s.

Model selection can be performed by calculating the Bayes factor, or, equivalently, by choosing the partition with the highest marginal likelihood, which can be calculated as:

$$Pr(A|G) = \prod_{r=1}^{g}\prod_{s=1}^{g} \frac{L_{rs}!(S_r S_s - L_{rs})!}{(1+L_{rs})(1+S_r S_s)!} \tag{3.3}$$

For a full derivation of the group model and the Bayes factor, see Eklöf *et al.* [54] and Sander *et al.* [127].

The group model may be extended to correct for degree. For this version of the model, the marginal likelihood may be calculated as:

$$Pr(A|G) = \left[\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{g^2} \prod_{r=1}^{g}\prod_{s=1}^{g} (1+\beta)^{-(\alpha - L_{rs})}\Gamma(\alpha + L_{rs})\right] \times \left[\prod_{r=1}^{g} \frac{\prod_{i=1}^{S_r}\left(k_i^{\text{in}}\right)!\left(k_i^{\text{out}}!\right)}{\Gamma(S_r)^2\Gamma\left(S_r + \sum_{i=1}^{S_r} k_i^{\text{in}}\right)\Gamma\left(S_r + \sum_{i=1}^{S_r} k_i^{\text{out}}\right)}\right] \tag{3.4}$$

where α and β are parameters on a Gamma prior ($\alpha = 1$ and $\beta = 1$ used in our analyses), and k_i^{in} and k_i^{out} are the in-degree (number of prey/hosts) and out-degree (number of predators/parasites) for species i, respectively. For a derivation of the likelihoods and Bayes factor for the degree-corrected model, see the Supplemental Information and Karrer & Newman [76].

We searched for partitions that best fit the group model, one for each combination of the following variables: including/excluding concomitant predation, standard group model/degree

corrected model, and maximum number of groups (2, 3, 5, 10, or 100). When up to 100 groups were allowed, the groupings collapsed down to a number that was more statistically parsimonious. Allowing for 100 groups gives the model the flexibility to find a truly optimal grouping, but constraining the number of groups makes the structure easier to visualize, understand, and interpret. In addition, it allows for a clear comparison between partitions with the same number of groups. Therefore, although we present results for all groupings, we focus on the 10 group case, which gives the model some flexibility, but is feasible to visualize and compare.

We used Metropolis-Coupled Markov Chain Monte Carlo (MC^3) with a Gibbs sampler to search for the partition of species into groups that maximizes the marginal likelihood (for details, see Supplemental Information and 127). Since exhaustively searching all possible groupings is computationally infeasible, we performed 200 independent MC^3 runs for each grouping reported, with 10 chains and 200,000 steps. Differences from the true optimum are likely to be small, so finding the true optimum is unlikely to have a large effect on the results. For convenience, we refer to the best partitions found as "best groupings", although they are not guaranteed to be optimal.

We studied the effect of degree correction on the groupings by calculating the mutual information between degree-corrected and degree-uncorrected partitions. Mutual information is described in more detail in the section on Taxonomic Comparison.

3.2.3 Imbalance

Once the group structure was inferred, we evaluated how well these statistically defined groups correspond to ecologically relevant *a priori* partitions, such as those specifying general trophic strategies (*e.g.*, herbivores or parasites). We did this by characterizing the "imbalance" of the distribution of species employing a given strategy across the various groups specified by the group model. For instance, considering the distribution of parasites across

the group structure, we can measure the imbalance by calculating the proportion of species belonging to the dominant trophic strategy (*e.g.* parasites or non-parasites) in any given group:

$$\psi_i^{parasites} = \frac{max(\pi_i, \phi_i)}{\pi_i + \phi_i} \tag{3.5}$$

where π_i is the number of parasites in group i and ϕ_i is the number of free-living, *i.e.* non-parasitic, species. This index can range from $\frac{1}{2}$ in the case where both trophic strategies are present in equal numbers, to 1 when all species in the group employ the same strategy. For a given network and partition, we can calculate the full imbalance for a given trophic strategy by taking the product across all groups in the partition:

$$\Psi^{parasites} = \prod_{i=1}^{g} \psi_i = \prod_{i=1}^{g} \frac{max(\pi_i, \phi_i)}{\pi_i + \phi_i} \tag{3.6}$$

We calculated imbalance in this way for all trophic strategies we considered: primary producers, herbivores, free-living predators, and parasites. We used a generalization of this measure to calculate the imbalance value for the full network (incorporating all strategies and groups simultaneously)[2]:

$$\Psi^{All} = \prod_{i=1}^{g} \prod_{k=1}^{c} \psi_i^k = \prod_{i=1}^{g} \prod_{k=1}^{c} \frac{max(\chi_i^1, \chi_i^2, \ldots, \chi_i^c)}{\sum_{k=1}^{c} \chi_i^k} \tag{3.7}$$

where g is the number of groups in the partition, c is the number of unique trophic strategies, and χ_i^k is the number of species with strategy k in group i. Note that $\left(\frac{1}{c}\right)^g \leq \Psi^{All} \leq 1$. To determine whether a given value for Ψ is higher than expected by chance, and therefore whether parasites significantly aggregate with other parasites in the partition, we wanted to associate Ψ with a p-value, measuring the probability of obtaining an equal or greater

2. Note that in the case of just two strategies (*e.g.* parasites and non-parasites, this equation collapses into Eqn. eq. (3.6) and is the same for both strategies.

imbalance at random.

To get at this, we consider the following example. Suppose we partition S species into two groups ($g = 2$). The first group contains ϕ_1 free-living species and π_1 parasites, while the second group has ϕ_2 free-living species and π_2 parasites. Thus, the total number of parasites in the network is $P = \pi_1 + \pi_2 = \sum_{i=1}^{g} \pi_i$ and the total number of free-living species is $F = \phi_1 + \phi_2 = \sum_{i=1}^{g} \phi_i$. Clearly, $S = P + F$. The probability of obtaining exactly ϕ_1, ϕ_2, π_1, π_2 at random can be computed using the hypergeometric distribution:

$$
\begin{aligned}
Pr(\phi_1, \phi_2, \pi_1, \pi_2 | \vec{\mathcal{P}}, P, F) = \\
&= \frac{\binom{P}{\pi_1}\binom{F}{\phi_1}}{\binom{P+F}{\pi_1+\phi_1}} \frac{\binom{P-\pi_1}{\pi_2}\binom{F-\phi_1}{\phi_2}}{\binom{P+F-\pi_1-\phi_1}{\pi_2+\phi_2}} = \frac{\binom{P}{\pi_1}\binom{F}{\phi_1}}{\binom{P+F}{\pi_1+\phi_1}} \cdot 1 = \frac{\binom{P}{\pi_1}\binom{F}{\phi_1}}{\binom{P+F}{\pi_1+\phi_1}} \\
&= \frac{\binom{P}{\pi_2}\binom{F}{\phi_2}}{\binom{P+F}{\pi_2+\phi_2}} \frac{\binom{P-\pi_2}{\pi_1}\binom{F-\phi_2}{\phi_1}}{\binom{P+F-\pi_2-\phi_2}{\pi_1+\phi_1}} = \frac{\binom{P}{\pi_2}\binom{F}{\phi_2}}{\binom{P+F}{\pi_2+\phi_2}} \cdot 1 = \frac{\binom{P}{\pi_2}\binom{F}{\phi_2}}{\binom{P+F}{\pi_2+\phi_2}}
\end{aligned}
\tag{3.8}
$$

where $\vec{\mathcal{P}}$ is the partition structure (in this case there are only two groups, *i.e.*, $|\vec{\mathcal{P}}| = 2$) provided by the group model. Note that the probability is the same regardless of how we label the groups. Therefore, we can associate a probability of obtaining this result at random to each possible partition encompassing a given number of parasites and free-living species. The formula above can be generalized to an arbitrary number of groups $|\vec{\mathcal{P}}| = g$:

$$
Pr(\vec{\phi}, \vec{\pi} | \vec{\mathcal{P}}, P, F) = \prod_{i=1}^{g} \frac{\binom{P-\sum_{j=0}^{j<i} \pi_j}{\pi_i}\binom{F-\sum_{j=0}^{j<i} \phi_j}{\phi_i}}{\binom{P+F-\sum_{j=0}^{j<i}(\pi_j+\phi_i)}{\pi_i+\phi_i}} \tag{3.9}
$$

and trophic strategies c:

$$
Pr(\vec{\chi^1}, \vec{\chi^2}, \ldots, \vec{\chi^c} | \vec{\mathcal{P}}, X^1, X^2, \ldots, X^c) = \prod_{i=1}^{g} \frac{\prod_{k=1}^{c} \binom{X^k - \sum_{j=0}^{j<i} \chi_j^k}{\chi_i^k}}{\binom{S-\sum_{j=0}^{j<i}\sum_{k=1}^{c} \chi_j^k}{\sum_{k=1}^{c} \chi_i^k}} \tag{3.10}
$$

where by definition $\phi_0 = \pi_0 = \chi_0 = 0$, $X^k = \sum_{i=1}^{g} \chi_i^k$ is the total number of species with strategy k, and, as above, $S = \sum_{k=1}^{c} X^k$ is the total number of species in the network. For a description of how to calculate p-values based on these probabilities, see Box 1.

All data and code needed to run the search algorithm and perform all analyses may be found at `https://git.io/vXciH`.

Box 1. Calculation of partition imbalance p-values. We are interested in the probability of observing an equal or larger value of imbalance at random. For networks with few groups and/or trophic strategies, we can compute this probability analytically by enumerating all possible cases and adding the probabilities of observing each imbalance value greater or equal to that observed. For example, take a network that is composed of 8 species ($S = 8$), of which 3 are parasites ($P = 3$) and 5 are free-living ($F = 5$). Suppose that when we use the group model to find the optimal partitioning into 3 groups we find that $\vec{\pi} = [1, 0, 2]$ and $\vec{\phi} = [0, 4, 1]$. The imbalance is $\Psi = \frac{2}{3}$. We can compute all the possible cases in which we arrange the P parasites and F free-living species into 3 boxes of sizes $\vec{\mathcal{P}} = [1, 4, 3]$, with each configuration having an associated imbalance value and probability of obtaining this configuration at random. We can then compute a p-value for the empirical distribution of trophic strategies across the partition produced by the group model by summing the probabilities associated with the configurations yielding imbalance equal or higher than that found in the partition produced by the group model (fig. 3.2). Though this brute-force method becomes infeasible for networks with many groups or trophic strategies, we can still calculate the p-value numerically by comparing the observed imbalance to a large number of randomized species strategy distributions sampled uniformly across the provided group structure, with fairly rapid convergence (fig. 3.22).

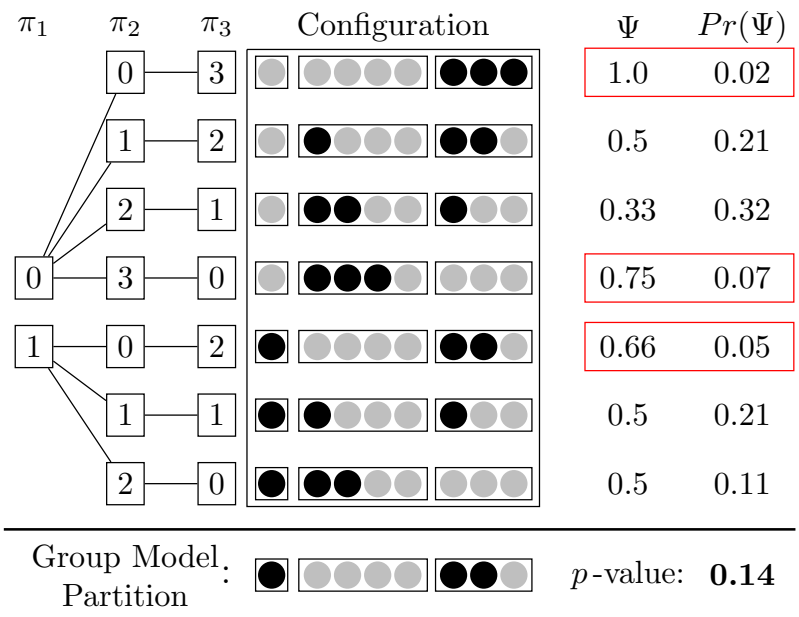


Figure 3.2: Process for calculating the p-value for a hypothetical group model partition. Under "Configuration", we list all seven unique configurations for the three parasites (black circles) and five free-living species (grey circles) into three groups (boxes) whose sizes have been determined by the partition produced by the group model. π_1, π_2, and π_3 show the combinatorial tree to obtain these configurations. Using equation eq. (3.6), we associate each configuration with an imbalance value ψ. Next, using equation eq. (3.9), we compute the probability of obtaining each configuration at random. Finally we sum the probabilities for all configurations with equal or greater imbalance than the empirical partititon (those in red boxes) to compute a p-value.

3.2.4 Taxonomic Comparison

In order to examine the biological relevance of the groupings found by the group model, we compared the group model results to a natural biological grouping by taxonomy. For this analysis, we utilized the Ythan network, which contained the most taxonomically detailed information about parasitic species, as a representative case-study. Taxonomic information to the family and order levels were gathered from the WoRMS database [70], and compared the resulting partitions to the 10-group partitions identified by the group model. The 10-group partitions were chosen because the number of groups was most similar to the number of families and orders in the taxonomic grouping, and were therefore the most analagous partitions.

The similarity of the two partitions was determined using the information theoretic measure of mutual information (MI). MI measures the reduction in entropy of partition B when partition A is known, such that if the MI is 1, the two partitions contain identical information. It can be thought of as the intersection between the two entropies, and is calculated as follows:

$$MI_{AB} = H(A) + H(B) - H(A,B) = \sum_{a \in A} \sum_{b \in B} p(a,b) \ln \frac{p(a,b)}{p(a)p(b)} \tag{3.11}$$

where $H(A)$ is the entropy of partition A, and $H(A,B)$ is the joint entropy of partitions A and B. The significance of MI_{AB} was calculated using a randomization test. For more details, see [127].

3.2.5 Subgraph-Roles

We furthermore considered the usefulness of meso-scale network structures, such as subgraph participation, in distinguishing trophic roles. We tabulated the number of times each species participated in each of the thirty potential roles within the thirteen three-node subgraphs

(Supplemental Information) and formed groups based on similarity of participation using a k-means algorithm. We then repeated the imbalance analysis described above using these alternative groupings and compared the results to those of the group model.

3.2.6 Life-Stages

Species, such as many parasites, which undergo complex life-cycles composed of stages with potentially distinct sets of interactions, face a dilemma in data organizations where ecological networks condense all interactions of a given species into the same node. We investigated the importance of this type of node resolution by defining nodes according to life-stages rather than species as in the previous analyses and broadening our trophic-strategy labels to incorporate this new information. We then repeated the grouping (using the group model) and imbalance analyses and compared the results to those found previously. Finally, using these same expanded networks, with separate nodes for each life-stage of the species involved, we repeated the imbalance analysis using simplified labels matching those used in the species-node case (*i.e.* the various life-stages of a given parasite species were still kept as separate nodes, but each stage had the same label of "parasite"). These results were then compared back to the original analysis as well as to the results of the life-stage-labeled imbalance values.

3.3 Results

Results were similar across networks. We report statistical results across all networks, but for simplicity we display figures and imbalance scores only for the largest network (Punta Banda) in the main text. Figures for the other six networks may be found in the Supplemental Information.

Partitions were significantly imbalanced in almost all cases, across different numbers of groups and all networks (table 3.1, section 3.15). Hence, the groupings maximizing the

marginal likelihoods separated parasites (and other trophic strategies) from other strategies more than expected by chance alone. Results were significant whether or not the model corrected for degree, and whether or not concomitant predation was considered. The only exception was that producers and herbivores were often highly imbalanced, but not significantly so. Since these networks generally had few producers and herbivores, this could be due to the relatively low statistical power. It is fairly easy for trophic strategies with few species to appear in the same groups simply by chance; as a result, trophic strategies with few species tend to have higher imbalance, but lower significance. This is a common problem for permutation tests and other procedures involving discrete outcomes, such as in Fisher's exact test. Since raw imbalance scores depend both on the number of groups and the number of species in each trophic strategy, the scores alone can be misleading; for this reason, we focus our interpretation on the significance rather than the scores themselves. Producers did tend to group together, but these groups often contained non-producers as well. This could be because they were being consumed by a similar group of predators, *e.g.* by a group of omnivores. Parasites and predators in particular tended to form groups which were distinct from all other strategies (fig. 3.3). Taking the network with the most resolved taxonomic information (Ythan), we found that the group model partitions shared a highly significant amount of information with the taxonomic partitions ($p \approx 0$ for all cases, see table 3.22), demonstrating a strong overlap between the group model and a natural biological partition.

Average in- and out-degree varied across trophic strategies, whether or not concomitant predation was included (one-way ANOVA, $p < .0001$ for all four tests: in-degree without concomitant, $F = 31.01$; in-degree with conomitant, 126.2; out-degree without concomitant, $F = 84.93$; out-degree with concomitant, $F = 174.5$) (fig. 3.4). Mean in-degree for predators was significantly higher than other trophic strategies (mean in-degree $\mu_{in} = 15.64$ without concomitant predation, $\mu_{in} = 30.54$ with concomitant predation), followed by parasites ($\mu_{in} = 12.80$ without concomitant, $\mu_{in} = 13.16$ with concomitant), followed by pro-

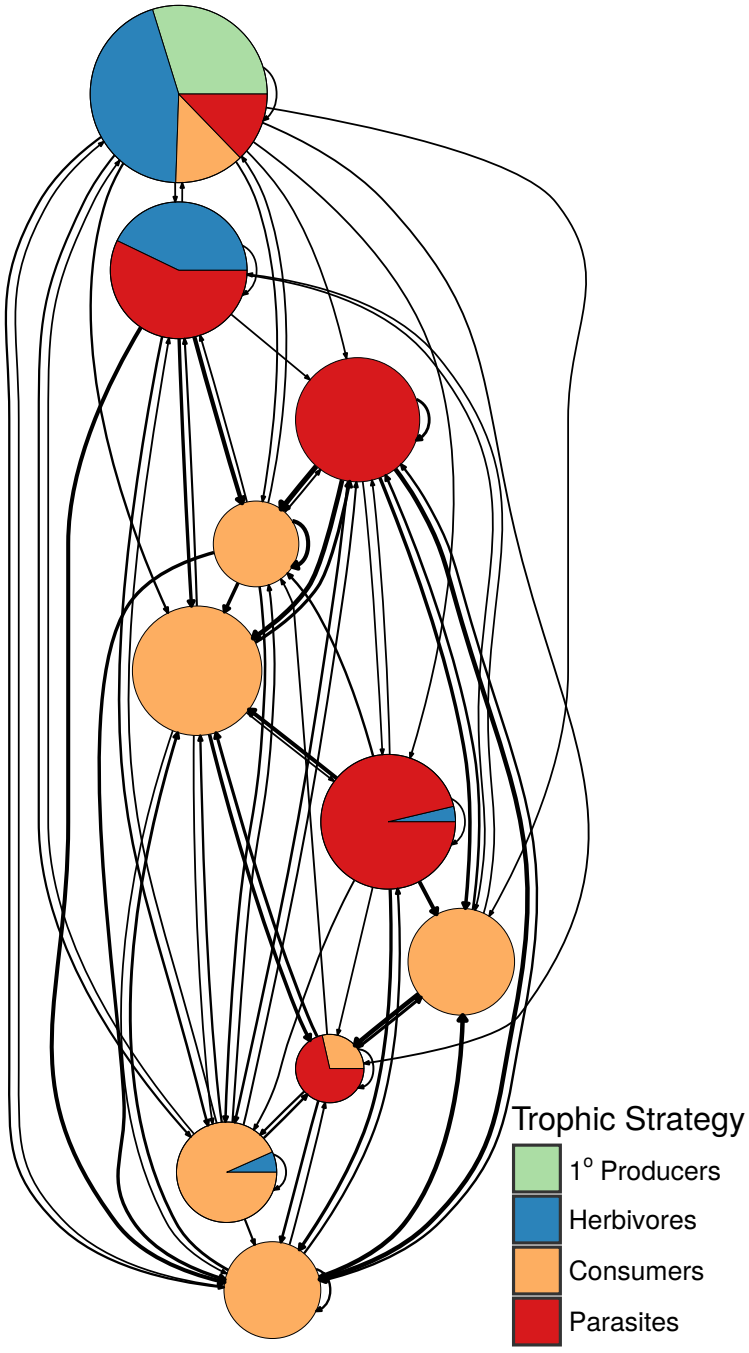


Figure 3.3: Condensed graph representation of the Punta Banda network using the best partitioning found by the group model. Each group is depicted by a pie-chart in which the fraction of nodes of each trophic strategy are indicated by the colored slices and the overall size is proportional to the number of nodes in the group. Number of links between (or within) groups is given by the thickness of each arrow and arrow direction of represents energy flow in the network; that is, arrows point to the consumer.

ducers and herbivores, which were not statistically distinct ($\mu_{in} = 0$ and 1.65 for producers and herbivores, respectively, both with and without concomitant predation). Mean out-degree was highest for producers, herbivores, and predators without concomitant predation ($\mu_{out} = 14.71$, 13.73, and 13.25, respectively), with lower out degree for parasites ($\mu_{out} = 6.79$). When concomitant predation was added, the pattern flipped: out-degree for parasites was highest ($\mu_{out} = 31.00$), with the other trophic strategies significantly lower ($\mu_{out} = 14.71$, 13.88, and 13.25 for producers, herbivores, and predators, respectively). In all four cases, parasites were significantly different from free-living predators.

Despite the differences in degree between trophic strategies, the degree-corrected group model produced significantly more imbalanced groups overall (across webs, trophic strategies, and number of groups), both when concomitant predation was included (paired t-test, estimated difference: .12, $p < .001$) and excluded (estimated difference: .067, $p < .001$). Including concomitant links also produced more imbalanced groups, under both the degree-corrected (paired t-test, estimated difference: .034, $p = .0049$) and uncorrected (estimated difference: .085, $p < .001$) models.

In general, degree-corrected and uncorrected partitions contained similar information when allowed to form more than 3 groups (table 3.21). The mutual information between corrected and uncorrected partitions increases as the number of groups increases. As expected, degree-corrected partitions tended to have fairly evenly sized groups [76], whereas group size was significantly less even for uncorrected partitions, both with concomitant predation (paired t test, estimated difference in Pielou's evenness [117]: $-.030$, $p = .031$) and without (estimated difference: $-.025$, $p = .001$).

Considering the distribution of subgraph-roles across trophic strategies, we see little consistency without the inclusion of concomitant links (fig. 3.6). Though including concomitant links does reveal some trends (fig. 3.7), taking the additional step of applying our imbalance analysis to groupings of species informed by subgraph-role participation yielded less consis-

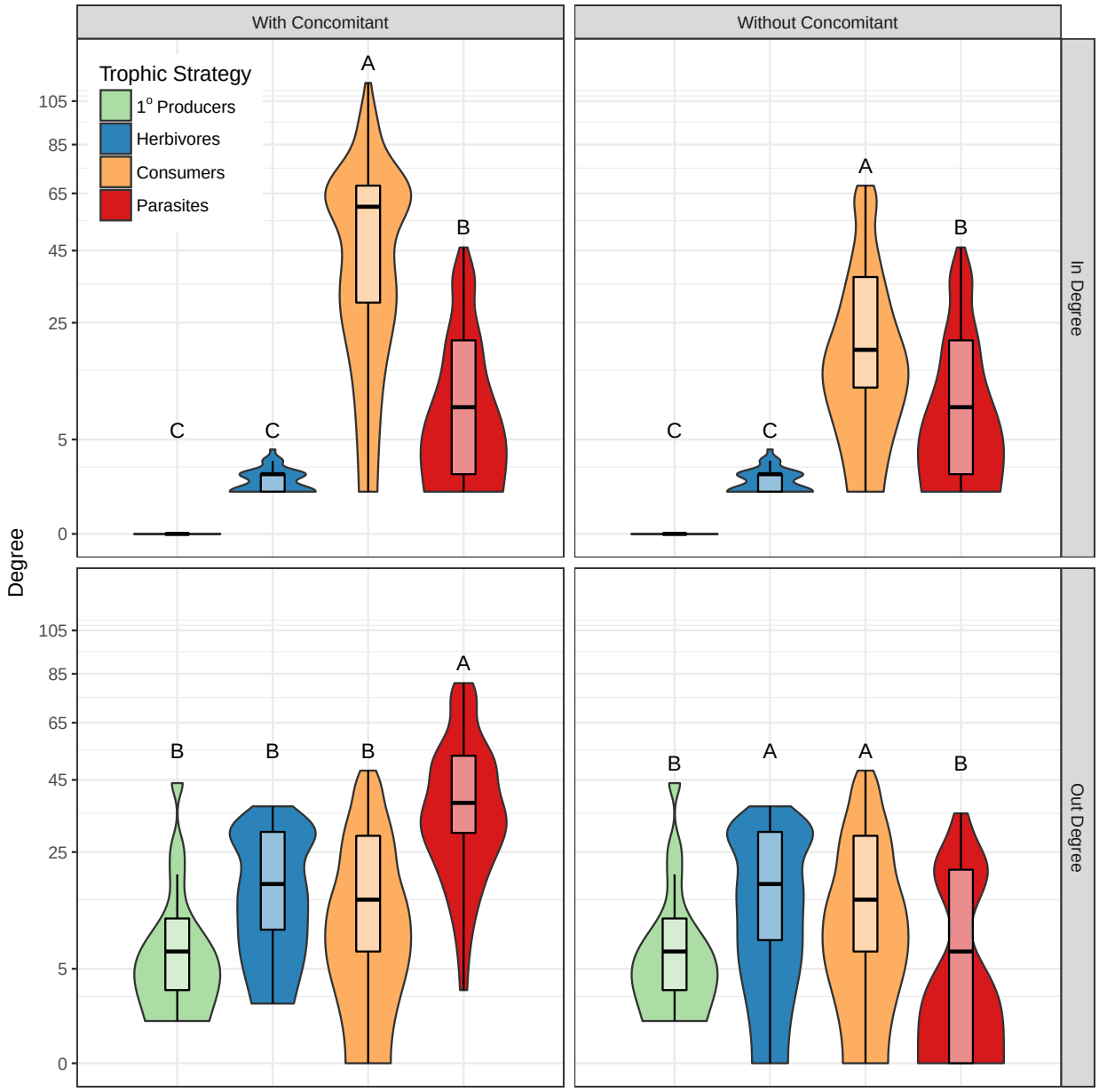


Figure 3.4: Violin and boxplots of in-degree (number of prey) and out-degree (number of predators) for different trophic strategies in the Punta Banda network. Degree is plotted on a square root scale. Boxes indicate the traditional 25^{th}, 50^{th} and 75^{th} quartiles, with wiskers extending to 1.5 times the inter-quartile range. Above each violin are grouping letters as indicated by a Tukey's HSD (honest significant difference) test.

tently significant imbalanced groups (compared to using the overall network structure as in the group model; section 3.16), as did repeating the analysis for food webs in which nodes are defined by life-stage rather than by species for parasites (section 3.17). In this latter case,

however, we still found significantly imbalanced groups with respect to particular life-stages in about half of the cases (*i.e.* combinations of concomitance, degree-correction, and number of groups) we considered. Re-labeling the life-stages with species-specific trophic strategies (*i.e.* labeling all life-stages of a parasite as just "parasite") yielded significant imbalances for these same groupings (section 3.18).

Concomitant Links	Degree Corrected	g/g_{max}	1° Producers	Herbivores	Consumers	Parasites	All
No	No	2/2	0.895***	0.737***	0.491***	0.421	0.246***
		3/3	0.826**	0.589	0.331***	0.260	0.120***
		5/5	0.757***	0.551***	0.628***	0.430***	0.307***
		10/10	0.585*	0.448***	0.439***	0.445***	0.214***
		18/100	0.481**	0.441***	0.494***	0.292***	0.103***
No	Yes	2/2	0.874	0.692	0.660***	0.449***	0.397***
		3/3	0.815	0.565	0.381***	0.521***	0.250***
		5/5	0.689	0.350	0.289***	0.559***	0.109***
		10/10	0.487	0.145	0.244***	0.138***	0.027***
		43/100	0.162***	0.036***	0.003***	0.005***	0.000***
Yes	No	2/2	0.898**	0.745***	0.912***	0.555***	0.555***
		3/3	0.839**	0.598	0.835***	0.603***	0.370***
		5/5	0.731	0.440*	0.702***	0.710***	0.327***
		10/10	0.702***	0.284***	0.582***	0.343***	0.164***
		23/100	0.411***	0.253***	0.590***	0.504***	0.169***
Yes	Yes	2/2	0.874	0.709***	0.812***	0.609***	0.537***
		3/3	0.833**	0.609***	0.564***	0.515***	0.282***
		5/5	0.716	0.360	0.365***	0.506***	0.173***
		10/10	0.451	0.198***	0.364***	0.445***	0.114***
		54/100	0.148***	0.006***	0.000***	0.024***	0.000***

Table 3.1: Imbalance values for the Punta Banda network (with or without concomitant predation) found via uniform sampling of 10^6 possible distributions of trophic strategies across the partitionings found by the (degree-corrected or otherwise) group model with g groups. Significance is indicated by the trailing asterisks, with $p < 0.05$, $p < 0.01$, and $p < 0.001$, corresponding to $*$, $**$, and $***$, respectively. Note that the lower bound for the imbalance varies with the number of groups, so it is only appropriate to compare raw imbalance scores when the number of groups is the same.

3.4 Discussion

The group model can be used to find the coarse-grained ecological roles, similar to functional groups, that are present in a community. Here, we use the group model to identify general patterns in groupings across networks, to determine if parasites are structurally unique. Since groupings are based on the entire network structure, and the quality of a group depends on the quality of all other groups in the network, it is generally difficult to study how an ecologically relevant trait affects the group structure. In this study, we are able to consider the effect of two ecologically distinguishing features of parasites: concomitant predation and degree, by including or excluding concomitant links and by using a degree-corrected variant of the group model, respectively. We find that parasites are, in general, structurally distinct from free-living species, regardless of number of groups in the model, the inclusion or exclusion of concomitant predation, and whether or not the model corrects for degree. These findings contrast with those of [46] on changes in local structure with the inclusion of parasites resulting mostly from the general increases in diversity and complexity. This contrast is only apparent as it reflects a difference in focus; while [46] addressed changes in the overall structure of the networks, we address here the position of parasites within this structure.

Concomitant predation tends to increase the distinction between parasites and free living species. This is not very surprising; concomitant interactions can create loops in the network, causing it to look less "cascade-like", that is, less like a network where species only consume species which are below them along some niche axis (section 3.21). In the absence of parasites, food webs tend to follow a largely cascade-like structure [42], so the group model can easily use these interactions to distinguish parasitic species from free-living ones. Without concomitant links, parasites which have similar prey to free-living predators might end up in mixed groups of free-living and parasitic species; however, by including these links, parasitic species have additional predators that distinguish them from otherwise similar free-living

species.

More surprising is the relationship we find with degree. Since parasites have different average in- and out-degree than free-living species, we might expect that degree would help the group model cluster parasites together. We found that the opposite was true: compensating for degree heterogeneity with the degree-corrected group model still produced more imbalanced groups. This suggests that degree is not the strongest structural signal that separates parasites from non-parasites. Indeed, although the mean degree is significantly different, the degree distribution of parasitic species overlaps considerably with the degree distributions of free-living categories. When uncorrected for degree, the group model tends to form a few groups with a small number of high-degree species. If the number of groups is constrained, as we have done here, this results in a few small groups and several larger groups. Thus, while the small groups may be highly imbalanced, the larger groups are often less imbalanced. The degree-corrected model counteracts this effect, producing groupings are significantly more even in size and even more strongly imbalanced. Relaxing the constraint on the number of groups improves the fit of the uncorrected model, and indeed, we see that the two models form more similar groupings as the number of groups increases (table 3.21). This pattern suggests that the corrected and uncorrected models are identifying similar underlying structures, but that the uncorrected model "prioritizes" grouping of high-degree species over grouping species which are structurally similar. Put another way, the uncorrected model can be affected by high-degree outliers, especially when the number of groups is heavily constrained.

Finding little consolation in degree, one might think that the key to parasite structural uniqueness could be found in a slightly higher-order form of network structure, such as the local patterns of connectance termed "motifs" or subgraphs [46, 137]. Unfortunately, though some subgraph-roles are more associated with particular trophic strategies (40; Supplemental Information), we found these trends to be less consistent than the group model

at distinguishing strategies, such that groupings formed from species showing similarity in their subgraph-role participation were found to be less consistently significantly imbalanced and less strongly strongly significant on average when compared to those groupings found by the group model.

These results provide evidence that parasites are structurally distinguished, not by how many predators and hosts they have, but by who those predators and hosts are at a global scale. What are the ecological drivers of this difference? The groups identified by the group model do not lend themselves to easy interpretation in terms of one, or even a combination of several, node-specific properties of the network (*e.g.* degree, subgraph-role, taxonomy, body-size, trophic-level, centrality, *etc.*). Instead, the group model coalesces nodes that share similar roles within the network, *i.e.* species which interact with similar sets of other species in similar ways. Put another way, the group model acts upon the links between species, grouping links that go from and to similar species. Previous results have suggested that the group model is able to find more informative groupings than any one property alone [135], but perhaps some combination of properties not investigated here could lend a more ecological explanation for the similarity of species within groups. In the Ythan network, we find that the group model partitions are closely related to an important biological property: evolutionary history, as captured by taxonomic partitions. The group model is able to capture taxonomic information using the results of the community's evolutionary history: the patterns of interaction between species.

One aspect of food web reporting that disproportionately affects parasites is the common practice of combining all life-stages of a species into a single node in the network. Parasites often have complex life-cycles, distinct stages of which might show wildly different patterns of interactions. The bulk of the results presented thus far consider food webs arranged in this way. To evaluate the effect of this data limitation on our results, for a subset of these networks we also looked at versions with these life-stages disaggregated. While we found the group

model to be less consistent at distinguishing the various parasite life-stages from one another than it was at distinguishing species-aggregated trophic strategies, surprisingly, it still does fairly well at separating life-stages into different groups (section 3.17). More remarkably, if each life-stage is relabeled with the species-defined trophic strategies from before, we find even more consistently significantly imbalanced groupings than in the aggregated webs. This suggests that, when unable to distinguish between parasite life-stages, the group model tends to groups these various parasitic life-stages together, rather than lumping such stages in with another trophic strategy, and that the addition of this information on life-stages actually improves the group model's efficacy at grouping similar trophic strategies together at the species level.

Parasites constitute a very broad set of organisms. They can vary in many ways: size (from microscopic viruses to parasitic worms reaching a meter or longer; 121); life cycle complexity; level of specialization; presence of free living stages; and whether they live in or on their hosts. They are also extraordinarily phylogenetically diverse. Given these major differences, it is encouraging to see that a human-chosen categorization as parasite is indeed structurally relevant and statistically robust in food webs.

3.5 Conclusion

Network structure has been found to influence many important features of ecological systems, including robustness [50], stability [6], and resilience [77]. General patterns of network structure are also used to develop structural [42, 116, 153] and dynamic [19, 64] models. However, many of these models were developed from data that excluded parasites, and parasites violate many of the patterns that they are based on. For example, concomitant predation creates loops that violate the cascade model, and allometric patterns which hold for free-living species (*e.g.*, 27), such as predator:prey body mass ratios, are inverted for parasitic interactions [123]. Models such as Allometric Diet Breadth Model [116] and the

Allometric Trophic Network [19], which are based on body size data, are unlikely to capture parasites successfully.

Our finding that parasites have unique structural roles – in essence, form unique functional groups – suggests that existing food web models should be reevaluated to better fit these distinct structural patterns. This stands in contrast to previous work suggesting that parasites' effect on network structure is mainly due to changes in connectance and diversity [46]. Using the same set of networks, we instead find that parasites perform statistically distinct roles in networks, even when correcting for degree, and even when concomitant links are excluded. These results add to the growing evidence that parasites must be considered as we continue to study and model ecological networks.

3.6 Acknowledgements

Thanks to G Barabás, A Dobson, J Dunne, J Grilli, K Lafferty, and N Martinez for helpful discussions. ELS is supported by the NSF GRFP. MJM is supported by the U.S. Department of Education grant P200A150101. SA is supported by NSF DEB-1148867. This work was inspired by the Parasites and Food Webs Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (DEB-0553768), the University of California, Santa Barbara and the State of California.

3.7 Authors' Contributions

All authors conceived the ideas and designed methodology; MJM and ELS analyzed the data, generated figures, and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

3.8 Data accessiblity

All food web data used in this project can be found in the Dryad Digital Repository: `http://dx.doi.org/10.5061/dryad.b8r5c` [45].

Supplemental Information

3.9 Group Model

The group model organizes species into groups such that species within a group tend to interact with other groups in the same way. Here, and throughout, we consider the directed case. The undirected case is mathematically very similar and is given a fuller treatment in [76]. In the most basic version of this model, it is possible to calculate the likelihood of obtaining the observed network A given the block structure or grouping G, where each block is treated as a separate Erdős-Rényi random graph. This likelihood is given as:

$$\mathcal{L}(A|G, c_{rs}; r, s \in 1, \ldots, g) = \prod_{r=1}^{g} \prod_{s=1}^{g} c_{rs}^{L_{rs}} (1 - c_{rs})^{S_r S_s - L_{rs}} \tag{3.12}$$

where c_{rs} is the connectance between groups r and s (note that, since the graph is directed, c_{rs} is not necessarily equal to c_{sr}), g is the number of groups, L_{rs} is the number of edges going from group r to group s, and S_r is the number of species in group r [127]. Using a uniform prior, the Bayes factor can be calculated for model selection. For two groupings G_1 and G_2, the Bayes factor is given by

$$B = \frac{P(A|G_1)}{P(A|G_2)} \tag{3.13}$$

where $P(A|G_1)$ is the marginal likelihood

$$\int_0^1 \cdots \int_0^1 P(c_{rs}; r, s \in 1, \ldots, g)\mathcal{L}(A|G, c_{rs}, r, s \in 1:g)\mathrm{d}c_{11} \ldots \mathrm{d}c_{gg} \tag{3.14}$$

which can be integrated to give

$$\prod_{r=1}^{g}\prod_{s=1}^{g}\frac{L_{rs}!(S_rS_s - L_{rs})!}{(1+L_{rs})(1+S_rS_s)!} \tag{3.15}$$

3.10 Group Model for Multigraphs

Here we consider the group model for multigraphs, which is consistent with the degree-corrected case. For sparse networks such as ecological networks, the possibility of multiple edges between nodes does not significantly affect the results [76]. Using the multigraph group model, the number of interactions between a species from group r and a species from group s are drawn from a Poisson distribution with rate parameter ω_{rs}, such that $E[A_{ij}] = \omega_{rs}$ when $i \in r, j \in s$. Then the likelihood of the network A given the block structure G is:

$$\mathcal{L}(A|G, \omega_{rs}; r, s \in 1, \ldots, g) = C\prod_{i=1}^{S}\prod_{j=1}^{S}\left(\omega_{g_ig_j}\right)^{A_{ij}}exp\left(-\omega_{g_ig_j}\right) \tag{3.16}$$

where

$$C = \prod_{i=1}^{S}\prod_{j=1}^{S}\left(A_{ij}!\right)^{-1} \tag{3.17}$$

The likelihood of a single block rs is then

$$\prod_{i\in r}\prod_{j\in s}\left(\omega_{g_ig_j}\right)^{A_{ij}}exp\left(-\omega_{g_ig_j}\right)$$

$$= (\omega_{rs})^{\sum_{i\in r}\sum_{j\in s}A_{ij}}exp(-\omega_{rs}) \tag{3.18}$$

$$= \omega_{rs}^{L_{rs}} exp(-\omega_{rs})$$

Substituting this back into the likelihood for the full network, we get

$$\mathcal{L}(A|G, \omega_{rs}; r, s \in 1, \ldots, g) = C \prod_{r=1}^{g} \prod_{s=1}^{g} \omega_{rs}^{L_{rs}} exp(-\omega_{rs}) \tag{3.19}$$

Note that the fractional term C is constant with respect to the network structure, and therefore can be ignored when using Bayes factors for model selection on the same network. Note that $C = 1$ when the network is not a multigraph. Using equation 2 for the Bayes factor, we need to calculate the marginal likelihood. We can then use a Gamma distribution as a (conjugate) prior for the ωs, for the marginal likelihood

$$P(A|G_1) = C \int_0^\infty \cdots \int_0^\infty \prod_{r=1}^{g} \prod_{s=1}^{g} \frac{\beta^\alpha}{\Gamma(\alpha)} \omega_{rs}^{L_{rs}+\alpha-1} exp(-\omega_{rs}(1+\beta)) \mathrm{d}\omega_{11} \ldots \mathrm{d}\omega_{gg} \tag{3.20}$$

Since the likelihood is a function, and the prior is the probability of that function, the quantity inside the integrals and products is an expectation. Since the expectation of the product is the product of expectations, this may be rewritten as

$$P(A|G_1) = C \prod_{r=1}^{g} \prod_{s=1}^{g} \int_0^\infty \cdots \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \omega_{rs}^{L_{rs}+\alpha-1} exp(-\omega_{rs}(1+\beta)) \mathrm{d}\omega_{11} \ldots \mathrm{d}\omega_{gg} \tag{3.21}$$

which can be integrated to give

$$C \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^{g^2} (1+\beta)^{-\alpha g^2 + L} \prod_{r=1}^{g} \prod_{s=1}^{g} \Gamma(\alpha + L_{rs}) \tag{3.22}$$

where L is the total number of links in the network.

3.11 Degree-Corrected Directed Group Model

Now to incorporate degree correction into the group model for multigraphs, consider parameters θ_i and ϕ_i, $i \in 1:S$, which control the expected in- and out-degree of vertex i, respectively. Then the expected number of edges going from species j to i is given by

$$E\Big[A_{ij}\Big] = \theta_i \phi_j \omega_{g_i g_j} \tag{3.23}$$

Then the likelihood may be written as

$$\mathcal{L}(A|G,\theta,\phi,\omega) = C \prod_{i=1}^{S} \prod_{j=1}^{S} \Big(\theta_i \phi_j \omega_{g_i g_j}\Big)^{A_{ij}} exp\Big(-\theta_i \phi_j \omega_{g_i g_j}\Big) \tag{3.24}$$

To normalize the θs and ϕs, we impose the following constraint for all groups r:

$$\sum_{i=1}^{S} \theta_i \delta_{g_i,r} = \sum_{i=1}^{S} \phi_i \delta_{g_i,r} = 1 \tag{3.25}$$

where $\delta_{g_i,r}$ is the Kronecker delta. The likelihood can then be simplified, again by considering a single block rs:

$$\omega_{rs}^{L_{rs}} \prod_{i \in r} \prod_{j \in s} \Big(\theta_i \phi_j\Big)^{A_{ij}} exp\Big(-\omega_{rs} \theta_i \phi_j\Big) \tag{3.26}$$

$$= \omega_{rs}^{L_{rs}} exp\left[-\omega_{rs} \sum_{i \in r} \sum_{j \in s} \theta_i \phi_j\right] \prod_{i \in r} \prod_{j \in s} \theta_i^{A_{ij}} \phi_j^{A_{ij}} \tag{3.27}$$

$$= \omega_{rs}^{L_{rs}} exp\left[-\omega_{rs} \sum_{i \in r} \theta_i \sum_{j \in s} \phi_j\right] \prod_{i \in r} \theta_i^{\sum_{j \in s} A_{ij}} \prod_{j \in s} \phi_j^{\sum_{i \in r} A_{ij}} \tag{3.28}$$

then we can use the normalization constraints and plug this back into the likelihood for the full network to get

$$\mathcal{L}(A|G,\theta,\phi,\omega) = C \prod_{i=1}^{S} \theta_i^{k_i^{\text{out}}} \prod_{i=1}^{S} \phi_i^{k_i^{\text{in}}} \prod_{r=1}^{g} \prod_{s=1}^{g} \omega_{rs}^{L_{rs}} exp(-\omega_{rs}) \tag{3.29}$$

where k_i^{in} and k_i^{out} are the observed in- and out-degree of species i, respectively. Within each groups, the θs and ϕs are on a simplex (using the constraints in equation 14). Thus for each group r, we can set a prior of $Dirichlet(\vec{1})$ over the θs and ϕs in each group to get a flat prior over the simplex. As before, we use a Gamma prior over each ω_{rs}, for the marginal likelihood

$$P(A|G_1) = C \int_0^\infty \cdots \int_0^\infty \prod_{r=1}^{g} \prod_{s=1}^{g} \frac{\beta^\alpha}{\Gamma(\alpha)} \omega_{rs}^{L_{rs}+\alpha-1} exp(-\omega_{rs}(1+\beta)) \mathrm{d}\omega_{11} \ldots \mathrm{d}\omega_{gg} \times$$
$$\int \cdots \int_{\Delta_r, r\in 1:g} \prod_{i\in r} \theta_i^{k_i^{\text{out}}} \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_S \int \cdots \int_{\Delta_s, s\in 1:g} \prod_{j\in s} \phi_j^{k_j^{\text{in}}} \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_S \tag{3.30}$$

Using the same technique as before, the products may be moved outside of the integrals, and the result may be integrated to give:

$$P(A|G_1) = C \left[\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^{g^2} \prod_{r=1}^{g} \prod_{s=1}^{g} (1+\beta)^{-(\alpha - L_{rs})} \Gamma(\alpha + L_{rs}) \right] \times$$
$$\left[\prod_{r=1}^{g} \frac{\prod_{i=1}^{S_r} \left(k_i^{\text{in}}\right)! \left(k_i^{\text{out}}!\right)}{\Gamma(S_r)^2 \Gamma\left(S_r + \sum_{i=1}^{S_r} k_i^{\text{in}}\right) \Gamma\left(S_r + \sum_{i=1}^{S_r} k_i^{\text{out}}\right)} \right] \tag{3.31}$$

3.12 Search Algorithm

High-quality partitions were searched for using Metropolis-coupled Markov Chain Monte Carlo (MC^3). This algorithm uses multiple MCMC chains, run in parallel at different temperatures, with occasional opportunities for chains to swap temperatures. The temperature parameter tunes the probability of accepting a "bad" move, that is, accepting a move that

reduces the marginal likelihood. At low temperatures, the chain acts as a local search, only accepting steps which improve the marginal likelihood. At high temperatures, the chain acts more like a random walk, accepting many "bad" steps, in hopes of escaping local optima to find the globally optimal solution.

The algorithm was given a maximum number of groups g. Solutions were initialized by randomly assigning each species a group assignment between 1 and g. Throughout the search, partitions were allowed to collapse down to fewer than g groups (that is, some groups were allowed to be "empty"), but were never permitted to have more than g groups.

For further implementation details, see the main text and supplement of [127]. The structure of the search algorithm used here is identical, but with the added constraint of a maximum number of groups.

3.13 Subgraph Role Analysis

Here we investigate the usefulness of a species' subgraph contributions in classifying trophic strategies. In performing the following analyses, we utilize a subgraph-role naming convention which assigns a number to each of the thirteen non-isomorphic, three-node subgraphs [105] and then distinguishes the roles within each subgraph by each node's degree distribution. For example, 1.2.0 corresponds to a node in subgraph 1 (apparent competition, see fig. 3.5), with an in-degree (number of prey) of 2 and an out-degree (number of predators) of 0 (*i.e.* the green node at the top). Likewise, the red nodes at the bottom share the same subgraph-role of 1.0.1 (subgraph 1, in-degree of 0, out-degree of 1). We start by enumerating how many of each of the thirty possible subgraph-role combinations each node of the network participates in.

Previous studies have indicated that the distribution of these subgraph-roles varies across trophic strategies. We see this as well (figs. 3.6 and 3.7), but without the inclusion of concomitant predation, no subgraph-role shows a consistent trend across all of the webs we

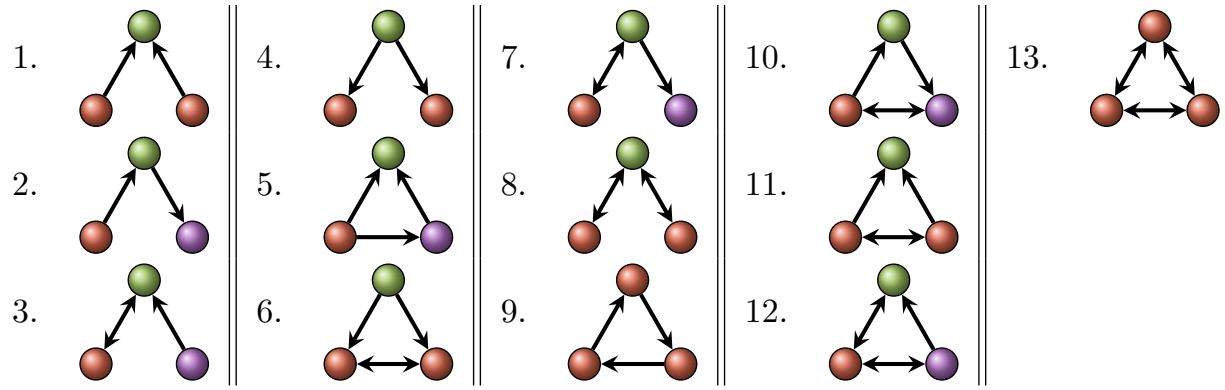


Figure 3.5: A graphical portrayal of the thirteen unique, connected, three-node subgraph structures. The direction of the arrows represents energy flow in the network; that is, arrows point to the consumer. Some of these structures have been given names in the ecological literature, for instance subgraph 1 is often termed apparent competition while subgraph 5 can depict omnivory. Unique "roles" (*i.e.* unique degree distributions within the subgraph) are depicted by differing colors. There are thirty unique subgraph-role combinations in total across these subgraphs: two in the first subgraph (in-degree of 0, out-degree of 1 (red) and in-degree of 2, out-degree of 0 (green)), three in the second, *etc.*

looked at. With the inclusion of links depicting concomitant predation, however, several consistent trends emerge. For instance, Parasitism is consistently found to be enriched in subgraph-roles 3.1.1, 7.1.2, and 10.1.2 compared to other trophic strategies. These are all roles involving a bidirectional relationship with another species (*i.e.* the parasite both feeds on and is eaten by the same species). Similarly, predators are could to be consistently enriched in subgraph-roles 1.2.0, 4.1.0, 5.2.0, and 7.1.0. These are all roles without any out-degree (*i.e.* species which are not consumed by another species). Taken together, we can conclude that the addition of concomitant links better allows the separation of parasites from predators insofar as this inclusion adds loops (see discussion in main text) to the network while increasing the out-degree of parasites and the in-degree of predators (as was also suggested by examination of node degree; section 3.24).

To investigate whether subgraph-role contribution could be used to distinguish trophic strategies more generally, these distributions (a vector of length 30 for each node in the network) were run through a principal component analysis to remove co-linearities and the resulting principal component coordinates were then clustered using a k-means algorithm

into 2, 3, 5, or 10 groups in `R` to correspond to the number of groups found using the group model in other analyses. The k-means algorithm divides the data into a set number of groups (without the possibility of empty groups), thus it did not make sense to also repeat the $g = 100$ case in which we were looking for a natural upper-bound on the number of groups found by the group model.

The groupings found by the k-means algorithm were then evaluated for imbalance with respect to trophic strategy as was done for the group model groupings. The results of this analysis are depicted in section 3.16, in tables analogous to those for the group-model groupings in section 3.15. Because this analysis is computationally intensive, we omitted the largest network (Punta Banda). In summary, we find these groupings to be less significantly imbalanced than those found by the group model. This is despite the trends observed in the subgraph distributions depicted in figs. 3.6 and 3.7 likely because of the substantial overlap in the distributions, even in the case of statistically distinct means.

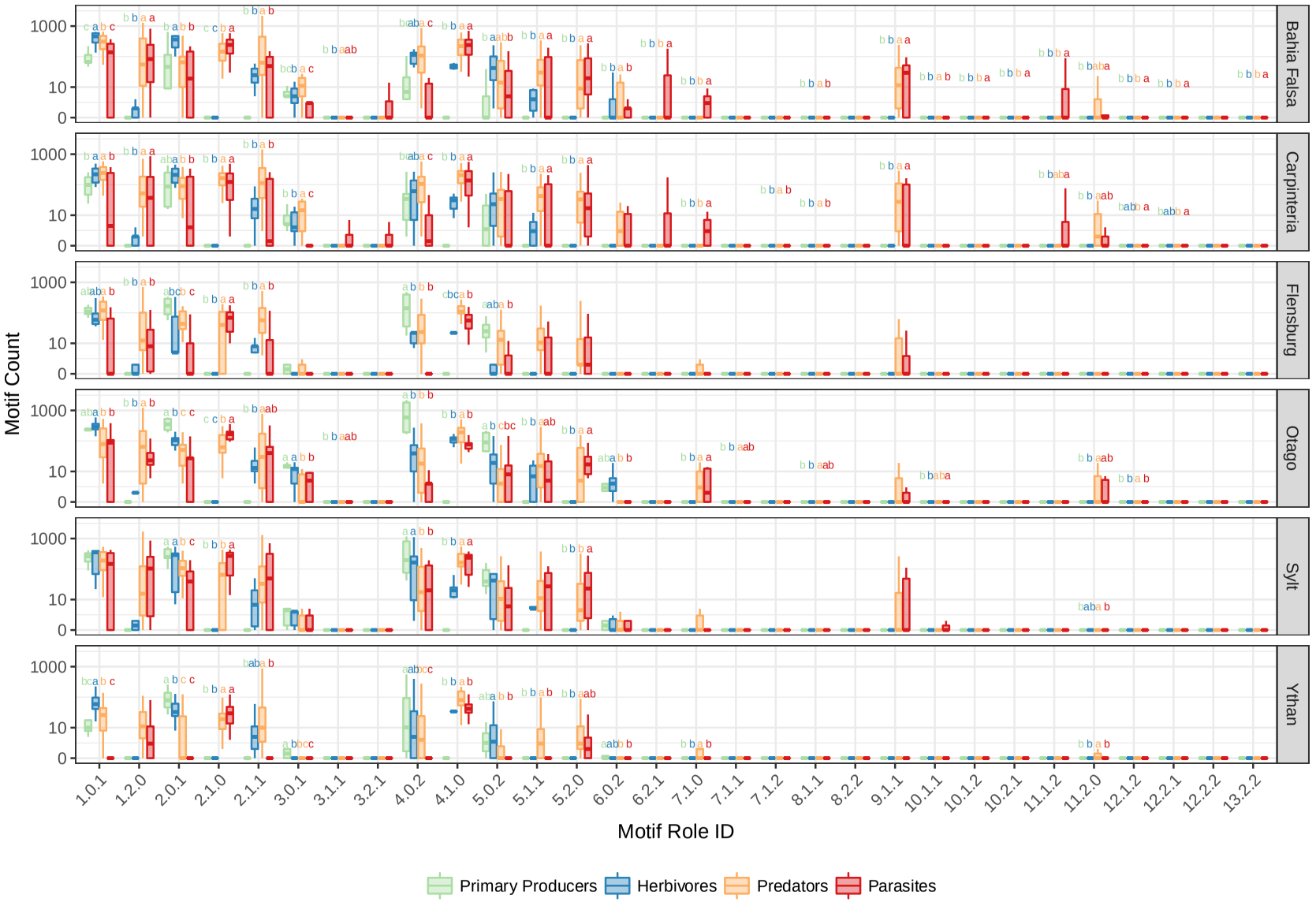


Figure 3.6: Distribution of subgraph-roles across trophic strategies. Subgraph counts are plotted on a square root scale. Boxes indicate the traditional 25^{th} , 50^{th}, and 75^{th} quartiles, with whiskers extending to 1.5 times the inter-quartile range. Above each boxplot are grouping letters as indicated by a Tukeys HSD (honest significant difference) test. Note that, though some panels show a difference in subgraph role counts by trophic strategy, none of these trends are consistent across all webs listed.

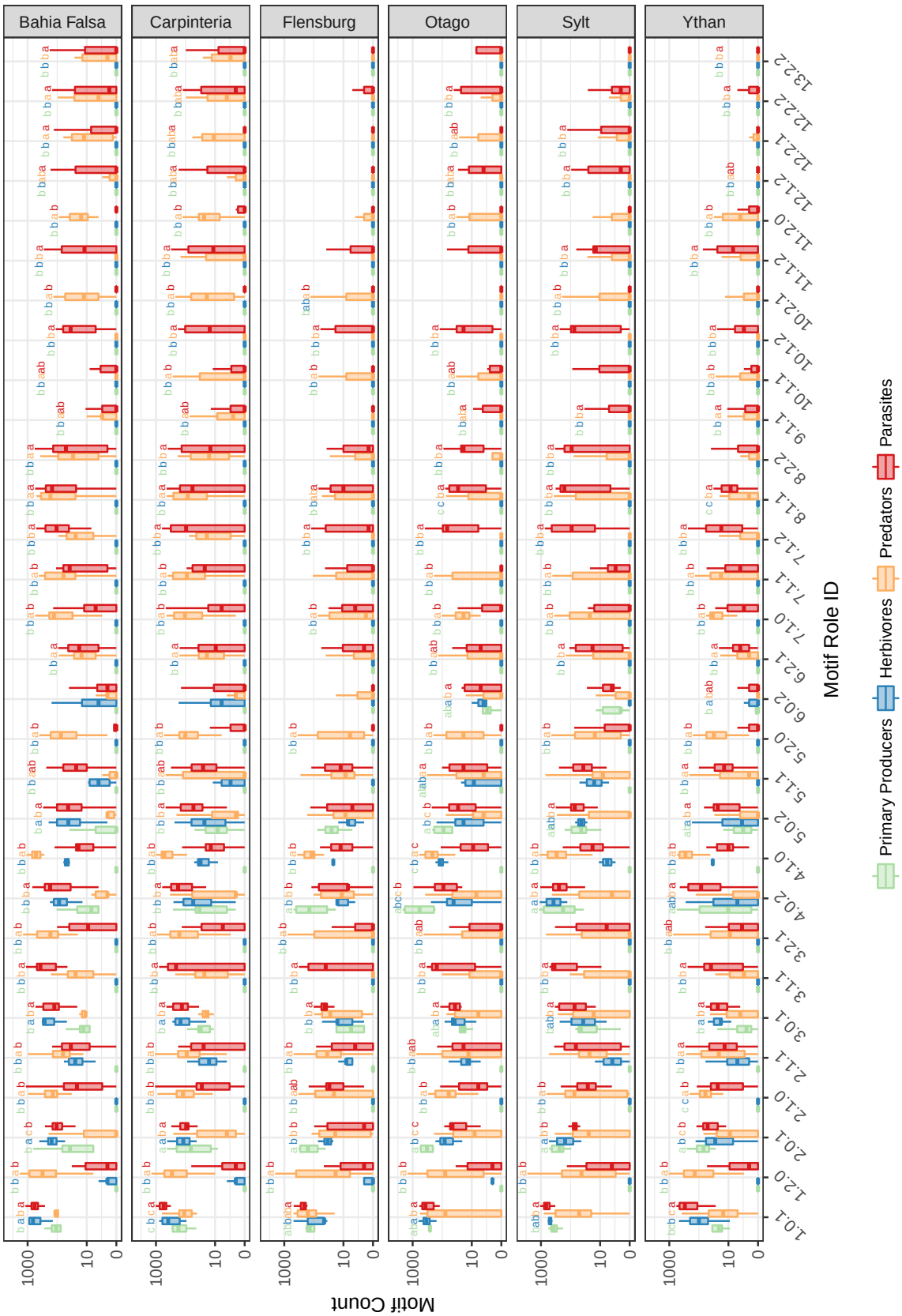


Figure 3.7: As fig. 3.6, but for webs including links corresponding to concomitant predation. In this case, we observe several trends that are consistent across webs (see text for details).

3.14 Data

Name	Species	Links (with concomitant predation)	Reference
Bahia Falsa	171	2234 (3720)	[66]
Carpinteria Salt Marsh	165	2187 (3708)	[66]
Flensburg Fjord	123	968 (1406)	[156]
Otago Harbor	142	1487 (1844)	[108]
Punta Banda	214	3334 (5653)	[66]
Sylt Tidal Basin	161	1950 (3005)	[143]
Ythan Estuary	133	597 (1391)	[72]

Table 3.2: Empirical food web data used in this paper. Each row corresponds to a different food web. Columns indicate, respectively, the name of the web, its number of species, number of links without (and with) including concomitant links, and a reference for the source of the original data.

3.15 Full Imbalance Results

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2/2	0.191**	0.636***	0.423	0.382	0.841**
		3/3	0.218***	0.534***	0.406***	0.446***	0.809***
		5/5	0.313***	0.717***	0.575***	0.740***	0.644*
		10/10	0.202***	0.480***	0.459***	0.655***	0.439*
		15/100	0.132***	0.329***	0.436***	0.427***	0.439***
No	Yes	2/2	0.309***	0.597*	0.494*	0.679***	0.817
		3/3	0.326***	0.452	0.576***	0.646***	0.737
		5/5	0.221***	0.544***	0.403***	0.568***	0.587
		10/10	0.115***	0.516***	0.264***	0.246***	0.450***
		38/100	0.001***	0.015***	0.021***	0.012***	0.131***
Yes	No	2/2	0.448***	0.649***	0.544***	0.931***	0.845**
		3/3	0.541***	0.580***	0.798***	0.906***	0.754***
		5/5	0.588***	0.625***	0.897***	0.882***	0.734***
		10/10	0.325***	0.645***	0.535***	0.573***	0.567***
		19/100	0.358***	0.704***	0.425***	1.000	0.842***
Yes	Yes	2/2	0.422***	0.612***	0.497*	0.820***	0.819
		3/3	0.273***	0.464*	0.473***	0.698***	0.724
		5/5	0.312***	0.425***	0.565***	0.690***	0.653***
		10/10	0.091***	0.241***	0.553***	0.325***	0.321
		43/100	0.000***	0.013***	0.067***	1.000	0.100***

Table 3.3: As table 3.1, but for the Bahia Falsa network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2/2	0.198	0.735	0.418	0.438***	0.860
		3/3	0.195***	0.685***	0.353***	0.465***	0.836***
		5/5	0.346***	0.539***	0.576***	0.625***	0.685
		10/10	0.283***	0.587***	0.468***	0.469***	0.656***
		14/100	0.234***	0.730***	0.312***	0.494***	0.510*
No	Yes	2/2	0.205	0.750***	0.420	0.403***	0.865
		3/3	0.225***	0.601	0.498***	0.321***	0.776
		5/5	0.089***	0.405	0.289***	0.150***	0.694
		10/10	0.041***	0.282***	0.146***	0.147***	0.367
		39/100	0.001***	0.073***	0.005***	0.011***	0.234***
Yes	No	2/2	0.236**	0.693	0.420	0.542***	0.840
		3/3	0.358***	0.623	0.639***	0.857***	0.803
		5/5	0.344***	0.623***	0.614***	0.823***	0.803***
		10/10	0.259***	0.529***	0.500***	0.552***	0.769***
		17/100	0.103***	0.264***	0.239***	0.517***	0.586***
Yes	Yes	2/2	0.535***	0.743	0.608***	0.788***	0.862
		3/3	0.234***	0.647	0.352***	0.572***	0.785
		5/5	0.160***	0.401	0.502***	0.318***	0.632
		10/10	0.061***	0.207	0.369***	0.194***	0.457
		47/100	0.002***	0.055***	0.072***	0.010***	0.315***

Table 3.4: As table 3.1, but for the Carpinteria network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2/2	0.425***	0.779	0.496***	0.567***	0.910
		3/3	0.281***	0.730	0.404***	0.358***	0.818
		5/5	0.312***	0.642	0.461***	0.470***	0.750
		10/10	0.365***	0.440	0.970***	0.426***	0.750
		13/100	0.478***	0.674***	0.827***	0.557***	0.750**
No	Yes	2/2	0.216	0.797	0.392	0.333	0.902
		3/3	0.433***	0.622	0.728***	0.492***	0.888*
		5/5	0.265***	0.549	0.789***	0.265***	0.647
		10/10	0.117***	0.347	0.404***	0.268***	0.500
		27/100	0.006***	0.200***	0.051***	0.023***	0.200
Yes	No	2/2	0.438***	0.876	0.562***	0.619***	0.943
		3/3	0.291***	0.829**	0.470***	0.423***	0.921
		5/5	0.300***	0.690**	0.525***	0.315***	0.857**
		10/10	0.426***	0.680***	0.857***	0.533***	0.615
		14/100	0.486***	0.700***	0.814***	0.567***	0.750**
Yes	Yes	2/2	0.356***	0.742	0.561***	0.387*	0.854
		3/3	0.189***	0.680	0.347***	0.208	0.835
		5/5	0.177***	0.541	0.655***	0.207***	0.538
		10/10	0.237***	0.477***	0.909***	0.276***	0.500
		32/100	0.016***	0.200***	0.240***	0.024***	0.250

Table 3.5: As table 3.1, but for the Flensburg network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2/2	0.380***	0.517***	0.756	0.504***	0.931
		3/3	0.326***	0.588***	0.632	0.343***	0.902
		5/5	0.191***	0.603***	0.442	0.263***	0.667
		10/10	0.281***	0.527***	0.533***	0.351***	0.800
		15/100	0.113***	0.277***	0.433***	0.169***	0.667
No	Yes	2/2	0.278	0.464	0.748	0.278	0.941
		3/3	0.182**	0.288	0.619	0.221*	0.911
		5/5	0.083***	0.162	0.546***	0.145***	0.822
		10/10	0.024***	0.105***	0.412***	0.063***	0.577
		34/100	0.005***	0.055***	0.321***	0.007***	0.211
Yes	No	2/2	0.452***	0.538***	0.786**	0.690***	0.956
		3/3	0.371***	0.569***	0.589	0.543***	0.833
		5/5	0.165***	0.606***	0.358	0.297***	0.500
		10/10	0.373***	0.401***	0.794***	0.465***	1.000***
		17/100	0.217***	0.257***	0.755***	0.317***	0.831*
Yes	Yes	2/2	0.294	0.491	0.740	0.294	0.945
		3/3	0.178**	0.354***	0.613	0.178	0.920*
		5/5	0.086***	0.256***	0.490	0.101*	0.789
		10/10	0.052***	0.193***	0.572***	0.052***	0.562
		36/100	0.008***	0.072***	0.500***	0.008***	0.225

Table 3.6: As table 3.1, but for the Otago network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2/2	0.491	0.933*	0.579	0.491	0.933*
		3/3	0.347	0.871	0.493	0.347	0.842
		5/5	0.487***	0.879*	0.660***	0.487***	0.870
		10/10	0.290***	0.756	0.543***	0.507***	0.473
		15/100	0.191***	0.633	0.514***	0.280***	0.510
No	Yes	2/2	0.494	0.927	0.605	0.494	0.925
		3/3	0.603***	0.896	0.688***	0.692***	0.871
		5/5	0.335***	0.809	0.470***	0.396***	0.820
		10/10	0.129***	0.632	0.196***	0.338***	0.635
		36/100	0.007***	0.268	0.064***	0.018***	0.169
Yes	No	2/2	0.476	0.972	0.551	0.476	0.927
		3/3	0.767***	0.959	0.812***	0.849***	0.899
		5/5	0.554***	0.947	0.718***	0.554***	0.829
		10/10	0.351***	0.778	0.675***	0.468***	0.540
		15/100	0.231***	0.771	0.452***	0.309***	0.540
Yes	Yes	2/2	0.484	0.975	0.547	0.484	0.929
		3/3	0.707***	0.964	0.728***	0.831***	0.878
		5/5	0.468***	0.915	0.644***	0.504***	0.758
		10/10	0.207***	0.859	0.476***	0.207***	0.559
		43/100	0.003***	0.500	0.042***	0.007***	0.213

Table 3.7: As table 3.1, but for the Sylt network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2/2	0.248***	0.479	0.562***	0.390	0.926
		3/3	0.373***	0.533***	0.538***	0.611***	0.899
		5/5	0.371***	0.476***	0.778***	0.412***	0.912
		8/10	0.212***	0.303***	0.713***	0.255***	0.856
		8/100	0.212***	0.303***	0.713***	0.255***	0.856
No	Yes	2/2	0.269***	0.514	0.470	0.479***	0.935
		3/3	0.550***	0.754***	0.733***	0.611***	0.914
		5/5	0.344***	0.740***	0.452***	0.426***	0.846
		10/10	0.035***	0.398***	0.079***	0.063***	0.636
		27/100	0.000***	0.007***	0.001*	0.004***	0.480
Yes	No	2/2	0.457***	0.609***	0.543***	0.891***	0.957
		3/3	0.426***	0.483***	0.604***	0.753***	0.929*
		5/5	0.312***	0.476***	0.374***	0.745***	0.819
		10/10	0.426***	0.598***	0.519***	0.643***	0.927**
		13/100	0.276***	0.326***	0.455***	0.651***	0.874**
Yes	Yes	2/2	0.348***	0.497	0.448	0.706***	0.940
		3/3	0.197***	0.394***	0.304	0.593***	0.909
		5/5	0.350***	0.492***	0.445***	0.699***	0.853
		10/10	0.198***	0.362***	0.397***	0.386***	0.764
		34/100	0.001***	0.013***	0.064***	0.004***	0.338

Table 3.8: As table 3.1, but for the Ythan network.

3.16 Subgraph Role Imbalance Results

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2	0.243*	0.748*	0.482	0.423	0.893
		3	0.253***	0.730**	0.621***	0.497**	0.786
		5	0.092***	0.427**	0.399***	0.218**	0.706
		10	0.015***	0.407***	0.071***	0.028**	0.394
Yes	No	2	0.128	0.560	0.517*	0.440**	0.713
		3	0.201***	0.442	0.441	0.407***	0.809**
		5	0.096***	0.362*	0.208	0.330***	0.590
		10	0.055***	0.235**	0.237***	0.298***	0.568

Table 3.9: As table 3.1, but for groupings based on k-means clustering of the subgraph-role contributions of each node of the Bahia Falsa network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2	0.224	0.819**	0.369	0.303	0.906*
		3	0.156**	0.744***	0.288	0.321**	0.867***
		5	0.124***	0.630*	0.207*	0.180**	0.800
		10	0.008***	0.276	0.040**	0.029**	0.648***
Yes	No	2	0.278***	0.767	0.451*	0.469***	0.892**
		3	0.262**	0.847	0.418	0.404*	0.920
		5	0.205***	0.441	0.350***	0.586***	0.797*
		10	0.049***	0.362	0.159***	0.313***	0.554

Table 3.10: As table 3.1, but for groupings based on k-means clustering of the subgraph-role contributions of each node of the Carpinteria network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2	0.475	0.893	0.631	0.525	0.951
		3	0.146	0.821*	0.226	0.208	0.813
		5	0.078**	0.739***	0.209**	0.152**	0.652
		10	0.094***	0.548	0.286***	0.172***	0.364
Yes	No	2	0.280	0.699	0.582**	0.336	0.815
		3	0.298*	0.670	0.513*	0.459**	0.730
		5	0.108*	0.616	0.361***	0.243**	0.809
		10	0.134***	0.528*	0.324***	0.202***	0.604

Table 3.11: As table 3.1, but for groupings based on k-means clustering of the subgraph-role contributions of each node of the Flensburg network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2	0.525	0.688	0.865	0.525	0.972
		3	0.139	0.392**	0.469	0.249*	0.924
		5	0.172***	0.392***	0.459	0.244***	0.934
		10	0.030*	0.133*	0.405	0.047*	0.432
Yes	No	2	0.272	0.691	0.583	0.272	0.815
		3	0.446**	0.679*	0.727	0.446*	0.969
		5	0.228***	0.478***	0.558	0.241**	0.963
		10	0.051***	0.162**	0.243	0.116***	0.750

Table 3.12: As table 3.1, but for groupings based on k-means clustering of the subgraph-role contributions of each node of the Otago network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2	0.477	0.923	0.600	0.477	0.917
		3	0.398***	0.897	0.534***	0.398***	0.880
		5	0.149	0.897*	0.311	0.186	0.458
		10	0.054	0.798	0.148	0.054	0.495
Yes	No	2	0.599*	0.964	0.682*	0.599*	0.933
		3	0.543**	0.979	0.613	0.543**	0.936*
		5	0.171	0.873	0.376	0.214	0.632
		10	0.197***	0.881	0.241**	0.354***	0.612

Table 3.13: As table 3.1, but for groupings based on k-means clustering of the subgraph-role contributions of each node of the Sylt network.

Concomitant Links	Degree Corrected	g	all	Herbivores	Parasites	Predators	1° Producers
No	No	2	0.379	0.727	0.682	0.621	0.970
		3	0.232***		0.382***	0.290**	0.909
		5	0.144***	0.387***	0.410***	0.204*	0.789
		10	0.076***	0.326***	0.165***	0.213***	0.780
Yes	No	2	0.278***	0.582***	0.506*	0.572***	0.955
		3	0.129**	0.576***	0.226		0.953
		5	0.147***	0.600***	0.273	0.257**	0.956
		10	0.050***	0.735***	0.074*	0.062**	

Table 3.14: As table 3.1, but for groupings based on k-means clustering of the subgraph-role contributions of each node of the Ythan network.

3.17 Disaggregated Life-cycles—Node-specific Trophic Strategies

Concomitant Links	Degree Corrected	g	all	Herbivore detritivore	Parasite	Parasite castrator	Parasite nonfeeding	Parasite pathogen	Parasite trophic-transmission	Predator	1°Producer	1°Producer other
Yes	No	2/2	0.312***	0.994	0.775**	0.912	0.925	0.975	0.769**	0.688***	0.975	0.988
		3/3	0.113***	0.991	0.656***	0.725	0.890**	0.953	0.572*	0.449***	0.963	0.982
		5/5	0.124***	0.982	0.568***	0.731	0.786	0.924	0.439**	0.351***	0.929**	0.964
		10/10	0.068***	0.964	0.465***	0.562	0.571	0.894**	0.235***	0.502***	0.857	0.929
		14/100	0.042***	0.800	0.453***	0.562	0.556	0.896*	0.216***	0.346***	0.578	0.770
	Yes	2/2	0.194***	0.982	0.712***	0.888**	0.782	0.968	0.704***	0.477***	0.927	0.964
		3/3	0.061	0.979	0.563***	0.702	0.745	0.945	0.514	0.259**	0.915	0.957
		5/5	0.092***	0.960	0.440***	0.576	0.520	0.917	0.335	0.595***	0.840	0.920
		10/10	0.024***	0.957	0.211***	0.331	0.522	0.797	0.128*	0.267***	0.826	0.913
		34/100	0.000***	0.750	0.095***	0.533***	0.300***	0.555	0.046***	0.033***	0.417	0.500
No	No	2/2	0.321***	0.994	0.778*	0.914	0.926	0.975	0.772*	0.679***	0.975	0.988
		3/3	0.198***	0.987	0.566	0.823	0.854	0.952	0.593*	0.439***	0.949	0.975
		5/5	0.127***	0.982	0.554***	0.659	0.786	0.931	0.461**	0.304***	0.929	0.964
		10/10	0.044***	0.964	0.483***	0.583**	0.571	0.885*	0.126	0.247***	0.857	0.929
		14/100	0.034***	0.800	0.405***	0.700***	0.556	0.877*	0.169***	0.193***	0.578	0.770
	Yes	2/2	0.180***	0.983	0.700***	0.877*	0.800	0.967	0.677***	0.430***	0.933	0.967
		3/3	0.233***	0.989	0.519	0.839***	0.862***	0.943	0.542	0.499***	0.954	0.977
		5/5	0.085***	0.960	0.418***	0.576	0.520	0.916	0.331	0.538***	0.840	0.920
		10/10	0.024***	0.952	0.268***	0.389	0.571	0.877*	0.177***	0.132***	0.714	0.905
		34/100	0.000***	0.800	0.050***	0.500***	0.333***	0.476	0.026***	0.023***	0.420	0.400

Table 3.15: As table 3.1, but for the Flensburg network with parasite life-stages disaggregated.

Table 3.16: As table 3.15, but for the Otago network.

Concomitant Links	Degree Corrected	g	all	Herbivore detritivore	Parasite	Parasite castrator	Parasite nonfeeding	Parasite trophic-transmission	Predator	1°Producer
Yes	No	2/2	0.505***	0.982	0.863**	0.895**	0.903**	0.879**	0.505***	0.966
		3/3	0.297	0.988	0.710	0.845***	0.791	0.818***	0.331***	0.965
		5/5	0.224***	0.974	0.644	0.802***	0.814***	0.750**	0.277***	0.923
		10/10	0.072***	0.833	0.551**	0.626*	0.620	0.472	0.371***	0.500
		16/100	0.024***	0.500	0.577***	0.612***	0.375	0.273	0.156***	0.250
Yes	Yes	2/2	0.433	0.988	0.819	0.862	0.860	0.842*	0.433	0.965
		3/3	0.262	0.985	0.707	0.783	0.818*	0.757	0.313***	0.955*
		5/5	0.148**	0.966	0.499	0.695	0.586	0.676	0.536***	0.897
		10/10	0.046***	0.875	0.750***	0.400	0.602***	0.358	0.166***	0.625
		38/100	0.001***	0.800	0.238***	0.306***	0.250***	0.152***	0.003***	0.188
No	No	2/2	0.509***	0.982	0.864**	0.896**	0.904*	0.880**	0.509***	0.966
		3/3	0.288	0.988	0.720	0.832**	0.804	0.809**	0.366***	0.963
		5/5	0.215***	0.973	0.706**	0.768*	0.785*	0.733*	0.257***	0.919
		10/10	0.088***	0.800	0.268***	1.000	1.000	1.000	0.335***	0.600
		16/100	0.056***	0.800	0.577***	0.627***	0.417	0.352	0.166***	0.600
No	Yes	2/2	0.435	0.988	0.821	0.863**	0.859	0.843*	0.435	0.965
		3/3	0.254	0.985	0.693	0.783	0.824**	0.745	0.307***	0.956
		5/5	0.134	0.967	0.492	0.666	0.600	0.639	0.574***	0.900
		10/10	0.045***	0.875	0.761***	0.459	0.602***	0.367	0.189***	0.625
		38/100	0.000***	0.800	0.190***	0.260***	0.250***	0.141***	0.002***	0.281

Table 3.17: As table 3.15, but for the Sylt network.

Concomitant Links	Degree Corrected	g	all	Herbivore detritivore	Parasite	Parasite castrator	Parasite nonfeeding	Parasite trophic-transmission	Predator	1°Producer
Yes	No	2/2	0.376***	0.994	0.826***	0.848***	0.893**	0.837***	0.624***	0.978
		3/3	0.166**	0.991	0.712***	0.697	0.821***	0.688	0.362***	0.959
		5/5	0.106***	0.962	0.455	0.671***	0.684	0.613***	0.349***	0.846
		10/10	0.082***	0.955	0.490***	0.442***	0.797***	0.278	0.446***	0.862
		18/100	0.041***	0.800	0.502***	0.450***	0.776***	0.216**	0.282***	0.564
Yes	Yes	2/2	0.277	0.989	0.777***	0.806***	0.791	0.788***	0.449***	0.960
		3/3	0.140	0.982	0.648	0.713**	0.670	0.695***	0.259***	0.940
		5/5	0.180***	0.978	0.430	0.539	0.597	0.509	0.628***	0.919
		10/10	0.057***	0.938	0.614***	0.540***	0.792***	0.239	0.182***	0.782
		44/100	0.000***	0.667	0.141***	0.050***	0.370***	0.016***	0.005***	0.400
No	No	2/2	0.373***	0.994	0.825***	0.847***	0.893**	0.836***	0.627***	0.977
		3/3	0.167**	0.991	0.716***	0.695	0.824***	0.686	0.352***	0.959
		5/5	0.100***	0.962	0.469	0.665***	0.684	0.611***	0.348***	0.846
		10/10	0.090***	0.955	0.460***	0.493***	0.800***	0.311	0.309***	0.869
		18/100	0.026***	0.800	0.413***	0.375***	0.770***	0.220***	0.151***	0.556
No	Yes	2/2	0.278*	0.989	0.777***	0.800***	0.796	0.788***	0.425***	0.960
		3/3	0.139	0.982	0.660	0.708**	0.671	0.677	0.221**	0.942
		5/5	0.184***	0.978	0.430	0.539	0.597	0.509	0.655***	0.919
		10/10	0.110***	1.000	0.229***	1.000	0.791***	1.000	0.239***	0.864**
		44/100	0.000***	1.000	0.007***	1.000	0.476***	1.000	0.016***	0.360

3.18 Disaggregated Life-cycles—Species-specific Trophic Strategies

Concomitant Links	Degree Corrected	g	all	Herbivore	Parasite	Predator	1°Producer
		2/2	0.644***	0.994	0.644***	0.688***	0.962
		3/3	0.409***	0.991	0.409***	0.449***	0.945*
Yes	No	5/5	0.351***	0.982	0.431***	0.351***	0.893
		10/10	0.317***	0.964	0.422***	0.502***	0.786
		14/100	0.257***	0.800	0.370***	0.346***	0.741
		2/2	0.477***	0.982	0.569***	0.477***	0.891
		3/3	0.259***	0.979	0.328***	0.259**	0.872
Yes	Yes	5/5	0.375***	0.960	0.407***	0.595***	0.760
		10/10	0.169***	0.957	0.169***	0.267***	0.739
		34/100	0.022***	0.750	0.133***	0.033***	0.185
		2/2	0.636***	0.994	0.636***	0.679***	0.963
		3/3	0.368***	0.987	0.439***	0.439***	0.924
No	No	5/5	0.304***	0.982	0.373***	0.304***	0.893
		10/10	0.156***	0.964	0.208***	0.247***	0.786
		14/100	0.144***	0.800	0.207***	0.193***	0.741
		2/2	0.430***	0.983	0.514***	0.430***	0.900
		3/3	0.426***	0.989	0.478***	0.499***	0.931*
No	Yes	5/5	0.340***	0.960	0.368***	0.538***	0.760
		10/10	0.088***	0.952	0.110***	0.132***	0.635
		34/100	0.005***	0.800	0.032***	0.023***	0.158

Table 3.18: As table 3.16 (*i.e.* using the disaggregated food web), but labeling nodes with the trophic strategy of their aggregated node from the prior analyses.

Concomitant Links	Degree Corrected	g	all	Herbivore	Parasite	Predator	1°Producer
Yes	No	2/2	0.505***	0.982	0.540***	0.505***	0.966
		3/3	0.331***	0.988	0.356**	0.331***	0.965
		5/5	0.277***	0.974	0.313***	0.277***	0.923
		10/10	0.186***	0.833	0.247***	0.371***	0.500
		16/100	0.039***	0.500	0.078***	0.156***	0.250
Yes	Yes	2/2	0.433	0.988	0.459	0.433	0.965
		3/3	0.313***	0.985	0.339***	0.313***	0.955*
		5/5	0.462***	0.966	0.462***	0.536***	0.897
		10/10	0.083***	0.875	0.166***	0.166***	0.625
		38/100	0.003***	0.800	0.021***	0.003***	0.188
No	No	2/2	0.509***	0.982	0.544***	0.509***	0.966
		3/3	0.366***	0.988	0.394***	0.366***	0.963
		5/5	0.257***	0.973	0.291***	0.257***	0.919
		10/10	0.201***	0.800	0.268***	0.335***	0.600
		16/100	0.099***	0.800	0.132***	0.166***	0.600
No	Yes	2/2	0.435	0.988	0.461	0.435	0.965
		3/3	0.307***	0.985	0.333***	0.307***	0.956
		5/5	0.497***	0.967	0.492	0.574***	0.900
		10/10	0.095***	0.875	0.189***	0.189***	0.625
		38/100	0.002***	0.800	0.011***	0.002***	0.281

Table 3.19: As table 3.16 (*i.e.* using the disaggregated food web), but labeling nodes with the trophic strategy of their aggregated node from the prior analyses.

Concomitant Links	Degree Corrected	g	all	Herbivore	Parasite	Predator	1°Producer
Yes	No	2/2	0.596***	0.994	0.596***	0.624***	0.978
		3/3	0.350***	0.991	0.369***	0.362***	0.959
		5/5	0.280***	0.962	0.280***	0.349***	0.846
		10/10	0.382***	0.955	0.408***	0.446***	0.862
		18/100	0.141***	0.800	0.306***	0.282***	0.564
Yes	Yes	2/2	0.444***	0.989	0.470***	0.449***	0.960
		3/3	0.259***	0.982	0.292***	0.259***	0.940
		5/5	0.585***	0.978	0.609***	0.628***	0.919
		10/10	0.136***	0.938	0.214***	0.182***	0.782
		44/100	0.001***	0.667	0.005***	0.005***	0.400
No	No	2/2	0.599***	0.994	0.599***	0.627***	0.977
		3/3	0.340***	0.991	0.358***	0.352***	0.959
		5/5	0.279***	0.962	0.279***	0.348***	0.846
		10/10	0.265***	0.955	0.285***	0.309***	0.869
		18/100	0.076***	0.800	0.173***	0.151***	0.556
No	Yes	2/2	0.420***	0.989	0.446***	0.425***	0.960
		3/3	0.221**	0.982	0.250***	0.221**	0.942
		5/5	0.610***	0.978	0.635***	0.655***	0.919
		10/10	0.212***	0.963	0.229***	0.239***	0.864**
		44/100	0.002***	0.667	0.007***	0.016***	0.360

Table 3.20: As table 3.16 (*i.e.* using the disaggregated food web), but labeling nodes with the trophic strategy of their aggregated node from the prior analyses.

3.19 Mutual Information Between Group Models

Table 3.21: Overlap between partitions that are corrected for degree and those that are not. Columns list the food web (Web), whether or not concomitant predation is included (Concomitant), the maximum number of groups the network is split into (G), the entropy of the degree-corrected and non-degree-corrected partitions ($H(Corrected)$ and $H(Uncorrected)$, the mutual information shared by the two partitions (MI), and the overlap represented as a Venn diagram. The left (purple) circle corresponds to the partition found by the group model without degree correction, and the right (green) circle corresponds to the partition found by the degree-corrected model. The area of each circle is proportional to the corresponding entropy, and the area of overlap between the circles is proportional to the mutual information. Stars next to the mutual information values correspond to the level of significance ($< .05, < .01, < .001$), as calculated by a randomization test [127]. As the number of groups increases, the entropy also increases, but the two partitions become increasingly similar. Corrected and uncorrected partitions become very similar when the network is partitioned into 10 groups, although there is always some distinct information in each.

Web	Concomitant	G	$H(Uncorrected)$	$H(Corrected)$	MI	
Flensburg	Yes	2	0.42	0.64	0.19^{***}	
Flensburg	Yes	3	0.92	1.09	0.36^{***}	
Flensburg	Yes	5	1.43	1.58	0.78^{***}	
Flensburg	Yes	10	2.14	2.26	1.71^{***}	
Flensburg	No	2	0.69	0.69	0.012	
Flensburg	No	3	1.08	1.06	0.23^{***}	
Flensburg	No	5	1.45	1.58	0.85^{***}	

Continued on next page

Table 3.21 – *Continued from previous page*

Web	Concomitant	G	$H(Uncorrected)$	$H(Corrected)$	MI	
Flensburg	No	10	2.02	2.26	1.50^{***}	
Carpinteria	Yes	2	0.69	0.69	0.073^{***}	
Carpinteria	Yes	3	1.08	1.09	0.39^{***}	
Carpinteria	Yes	5	1.49	1.58	0.93^{***}	
Carpinteria	Yes	10	2.06	2.27	1.49^{***}	
Carpinteria	No	2	0.69	0.69	0.03^{**}	
Carpinteria	No	3	1.07	1.08	0.37^{***}	
Carpinteria	No	5	1.58	1.60	0.83^{***}	
Carpinteria	No	10	2.11	2.28	1.30^{***}	
Otago	Yes	2	0.65	0.69	0.0021	

Continued on next page

Table 3.21 – *Continued from previous page*

Web	Concomitant	G	$H(Uncorrected)$	$H(Corrected)$	MI	
Otago	Yes	3	1.02	1.10	0.14***	
Otago	Yes	5	1.42	1.59	0.72***	
Otago	Yes	10	2.07	2.25	1.48***	
Otago	No	2	0.65	0.69	0.0055	
Otago	No	3	1.06	1.10	0.21***	
Otago	No	5	1.40	1.59	0.59***	
Otago	No	10	2.04	2.26	1.34***	
BahiaFalsa	Yes	2	0.64	0.67	0.46***	
BahiaFalsa	Yes	3	1.08	1.08	0.43***	
BahiaFalsa	Yes	5	1.50	1.57	1.07***	
BahiaFalsa	Yes	10	2.16	2.25	1.51***	

Continued on next page

Table 3.21 – *Continued from previous page*

Web	Concomitant	G	$H(Uncorrected)$	$H(Corrected)$	MI	
BahiaFalsa	No	2	0.65	0.68	0.024^{**}	
BahiaFalsa	No	3	1.00	1.09	0.27^{***}	
BahiaFalsa	No	5	1.53	1.58	0.79^{***}	
BahiaFalsa	No	10	2.12	2.27	1.54^{***}	
Ythan	Yes	2	0.62	0.69	0.30^{***}	
Ythan	Yes	3	1.08	1.10	0.25^{***}	
Ythan	Yes	5	1.52	1.61	0.86^{***}	
Ythan	Yes	10	1.93	2.28	1.22^{***}	
Ythan	No	2	0.59	0.69	0.0017	
Ythan	No	3	0.96	1.10	0.46^{***}	

Continued on next page

Table 3.21 – *Continued from previous page*

Web	Concomitant	G	$H(Uncorrected)$	$H(Corrected)$	MI	
Ythan	No	5	1.33	1.58	0.68***	
Ythan	No	10	1.72	2.28	1.02***	
PuntaBanda	Yes	2	0.65	0.69	0.44***	
PuntaBanda	Yes	3	1.07	1.08	0.44***	
PuntaBanda	Yes	5	1.57	1.59	0.82***	
PuntaBanda	Yes	10	2.18	2.28	1.62***	
PuntaBanda	No	2	0.66	0.69	0.059***	
PuntaBanda	No	3	1.09	1.10	0.066***	
PuntaBanda	No	5	1.58	1.59	0.67***	
PuntaBanda	No	10	2.18	2.28	1.41***	
Sylt	Yes	2	0.64	0.69	0.018*	

Continued on next page

Table 3.21 – *Continued from previous page*

Web	Concomitant	G	$H(Uncorrected)$	$H(Corrected)$	MI	
Sylt	Yes	3	0.95	1.08	0.34^{***}	
Sylt	Yes	5	1.59	1.57	0.74^{***}	
Sylt	Yes	10	2.12	2.28	1.46^{***}	
Sylt	No	2	0.69	0.69	0.01	
Sylt	No	3	1.06	1.06	0.16^{***}	
Sylt	No	5	1.51	1.59	0.74^{***}	
Sylt	No	10	2.02	2.28	1.30^{***}	

3.20 Mutual Information between Ythan Taxonomy and Group Model

Concomitant Links	Degree Correction	Rank	$H(Group)$	$H(Taxonomic)$	MI
No	No	Family	1.75	4.02	1.57^{***}
		Order	1.76	3.11	1.39^{***}
No	Yes	Family	2.27	4.02	2.00^{***}
		Order	2.27	3.11	1.58^{***}
Yes	No	Family	1.98	4.02	1.64^{***}
		Order	1.98	3.11	1.30^{***}
Yes	Yes	Family	2.28	4.02	2.08^{***}
		Order	2.29	3.11	1.67^{***}

Table 3.22: Overlap between taxonomic partition and 10-group group model partition for the Ythan network. Columns list whether or not concomitant predation is included (Concomitant Links), whether or not the model is corrected for degree (Degree Correction), taxonomic rank (Rank), entropy of the group model partition ($H(Group)$), entropy of the taxonomic partition ($H(Taxonomic)$), and mutual information (MI), with the significance from a randomization test marked in asterisks.

3.21 Empirical Network Adjacency Matrices – Grouped by Trophic Strategy

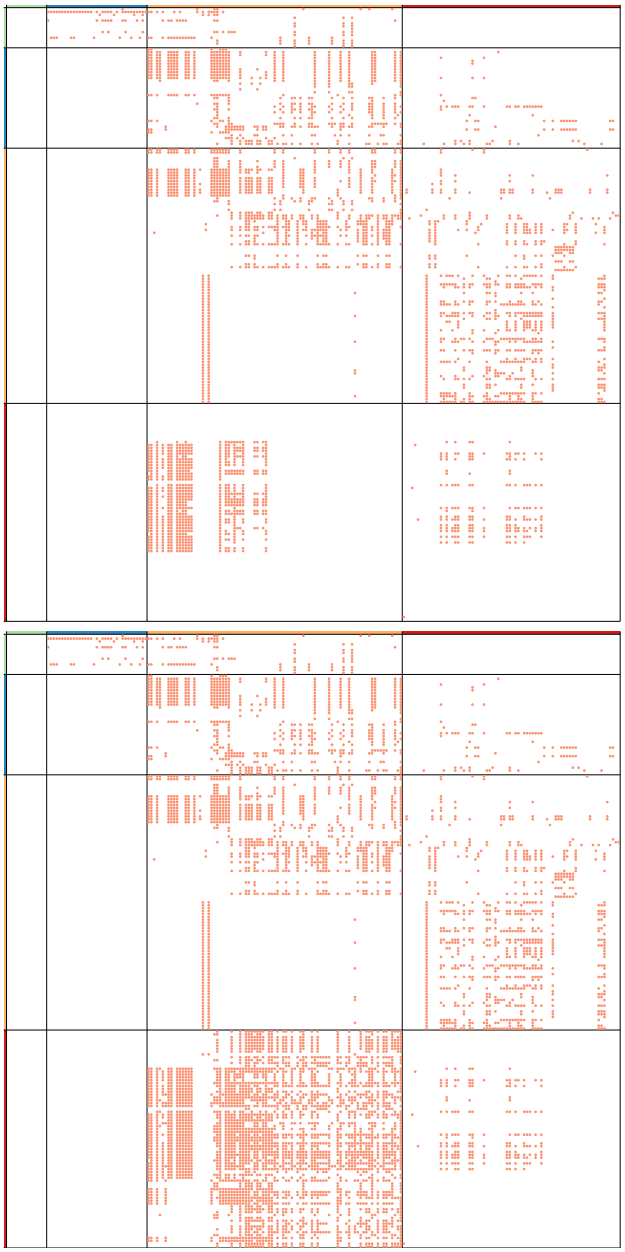


Figure 3.8: Punta Banda network structure with (top) and without (bottom) concomitant predation and species grouped by trophic strategy. Colored squares represent trophic strategy (green for primary producers, blue for herbivores, red for parasites, and yellow for other predators), and dots represent feeding interactions wherein the column species consumes the row species. Note the addition of concomitant links increases the number of consumer-resource interactions from predators to parasites and decreases the cascade-like structure seen in the top matrix.

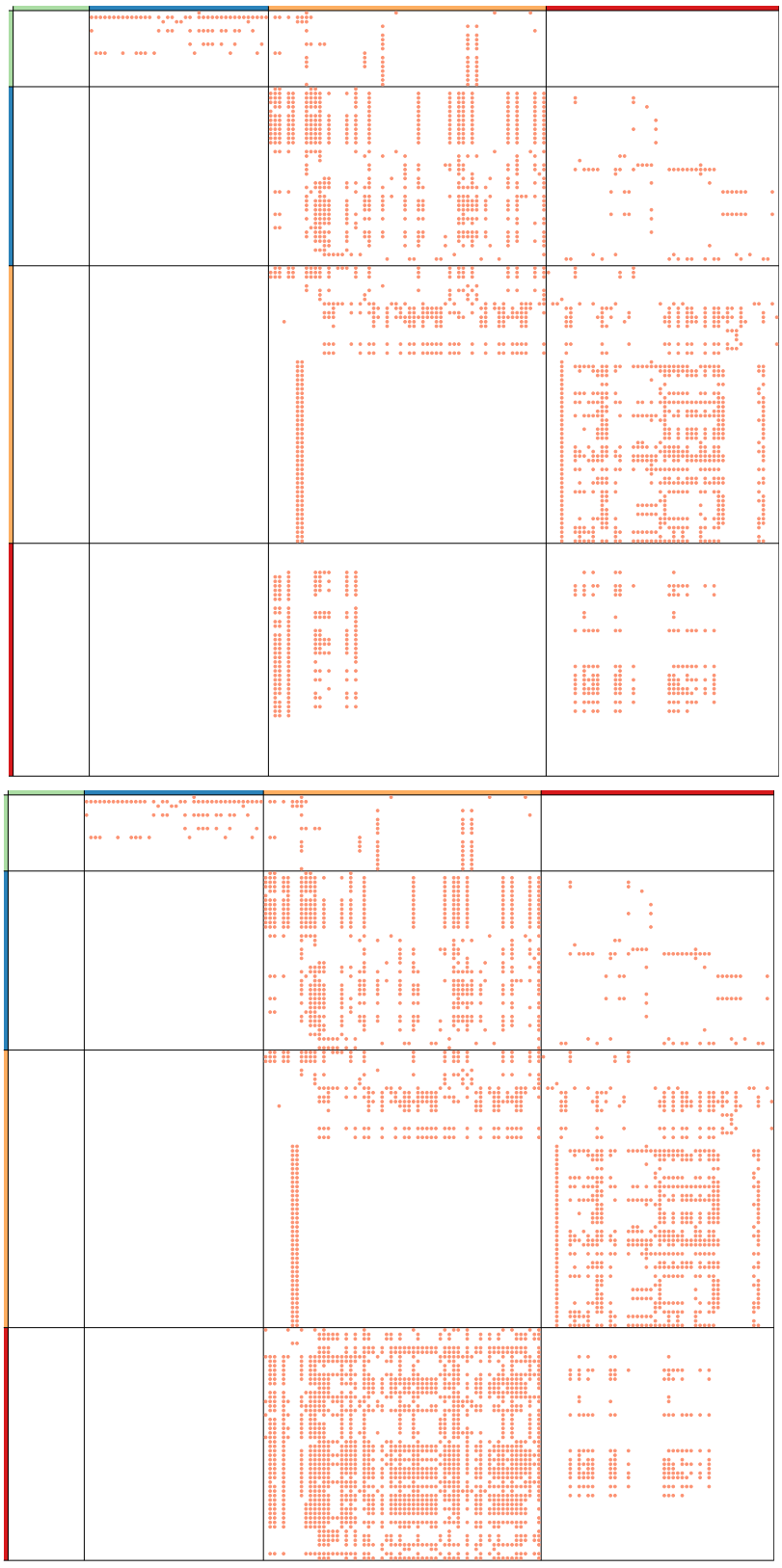


Figure 3.9: As fig. 3.8, but showing the Bahia Falsa network.

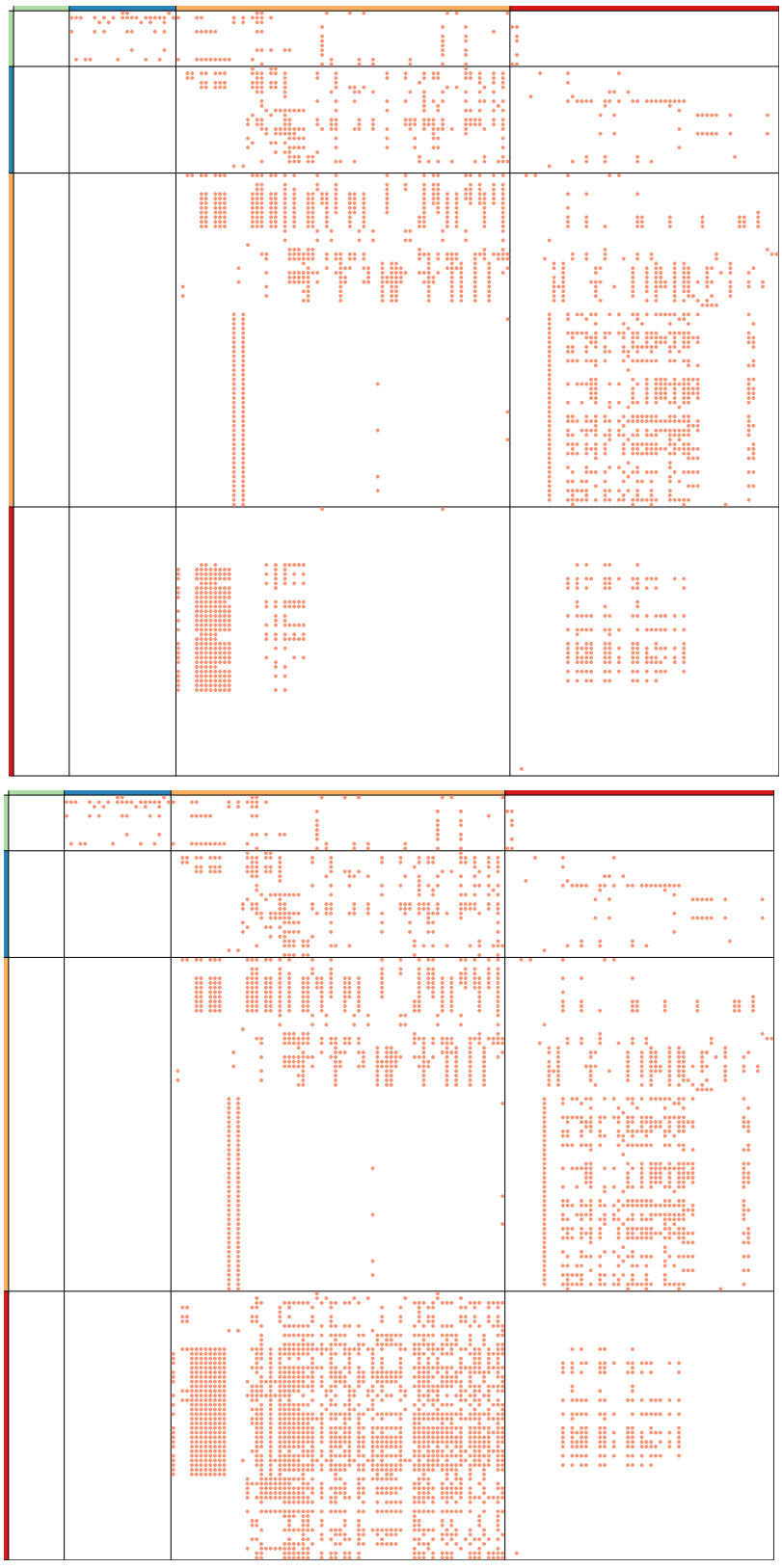


Figure 3.10: As fig. 3.8, but showing the Carpinteria network.

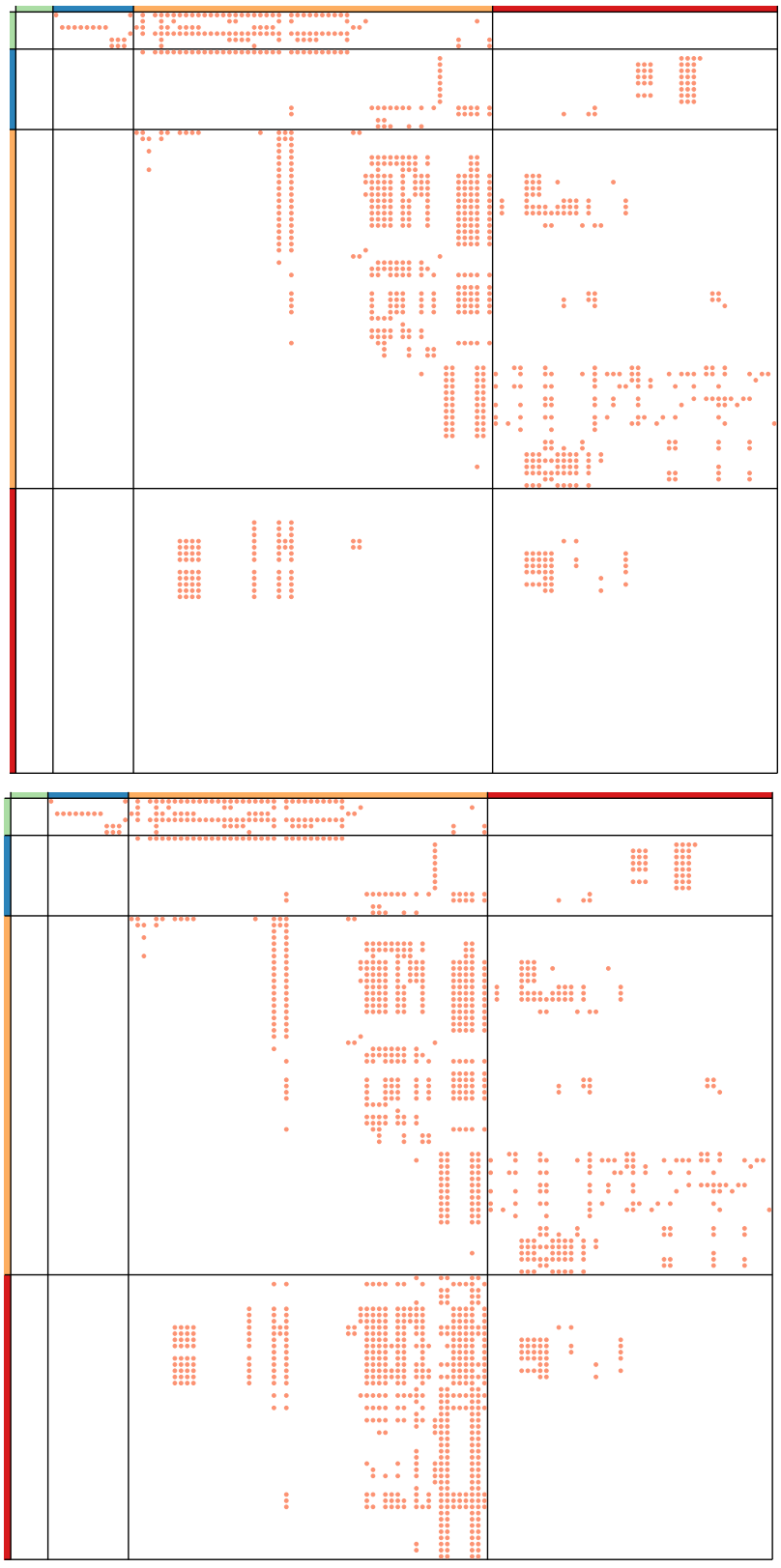


Figure 3.11: As fig. 3.8, but showing the Flensburg network.

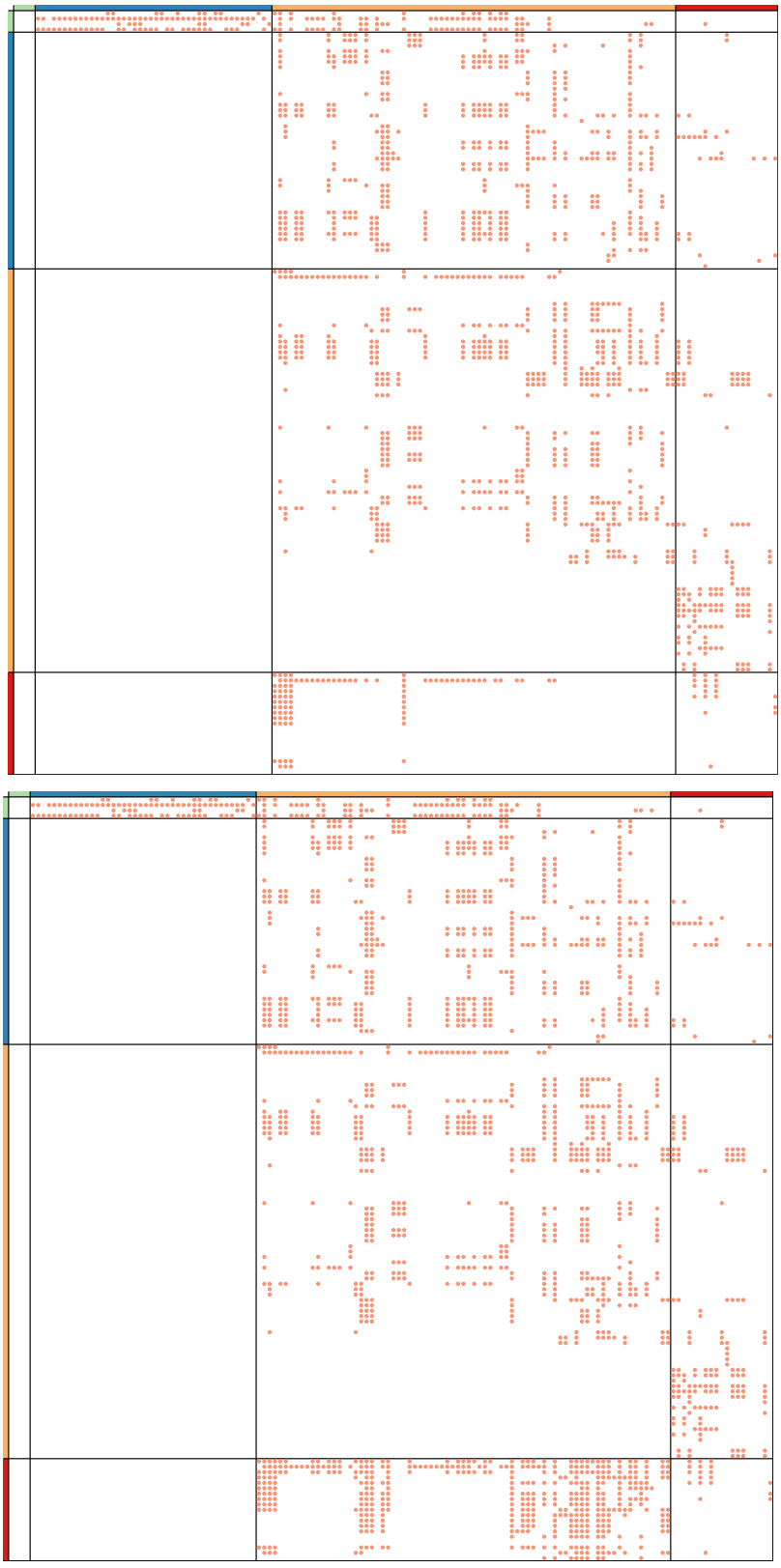


Figure 3.12: As fig. 3.8, but showing the Otago network.

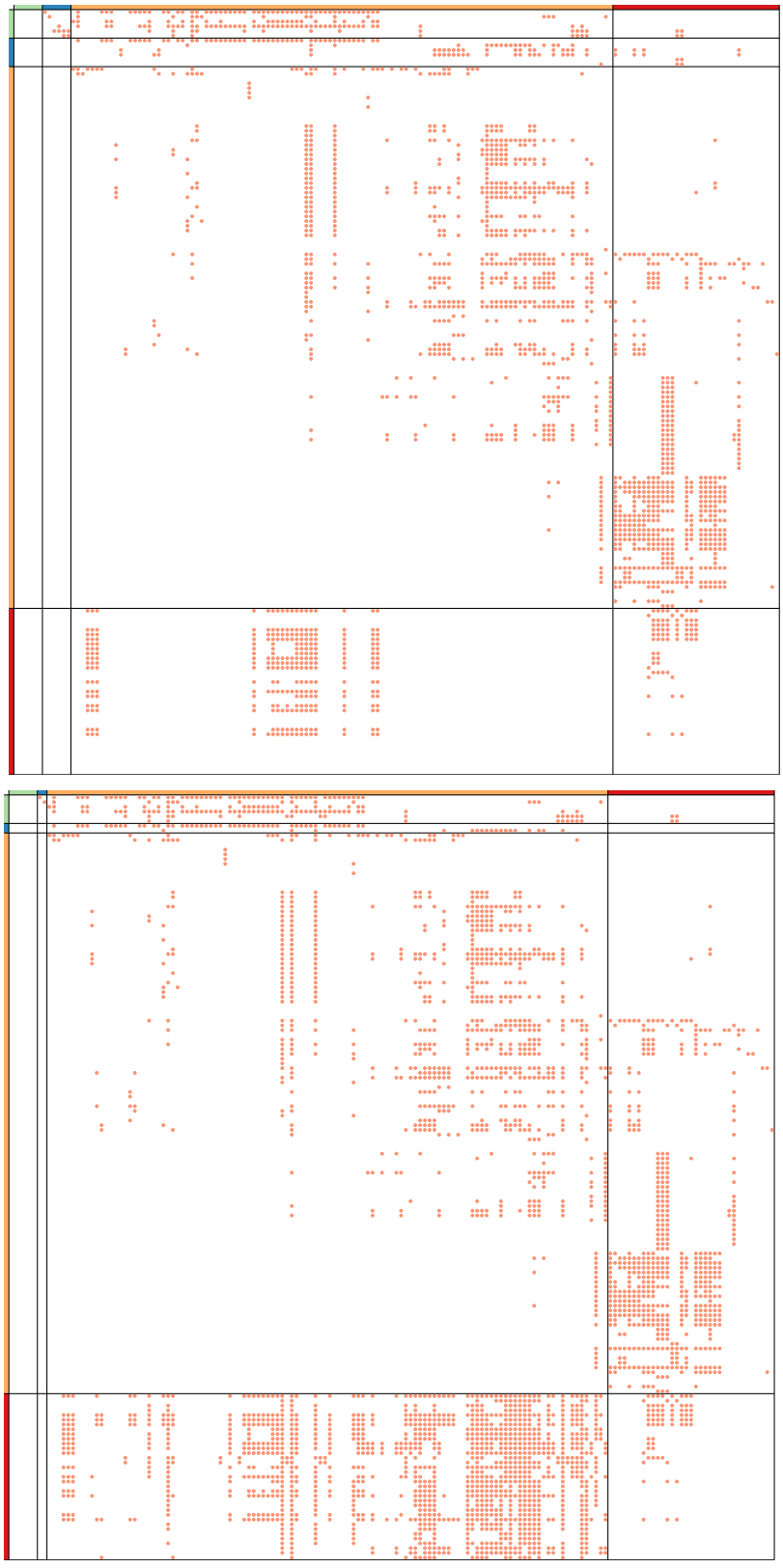


Figure 3.13: As fig. 3.8, but showing the Sylt network.

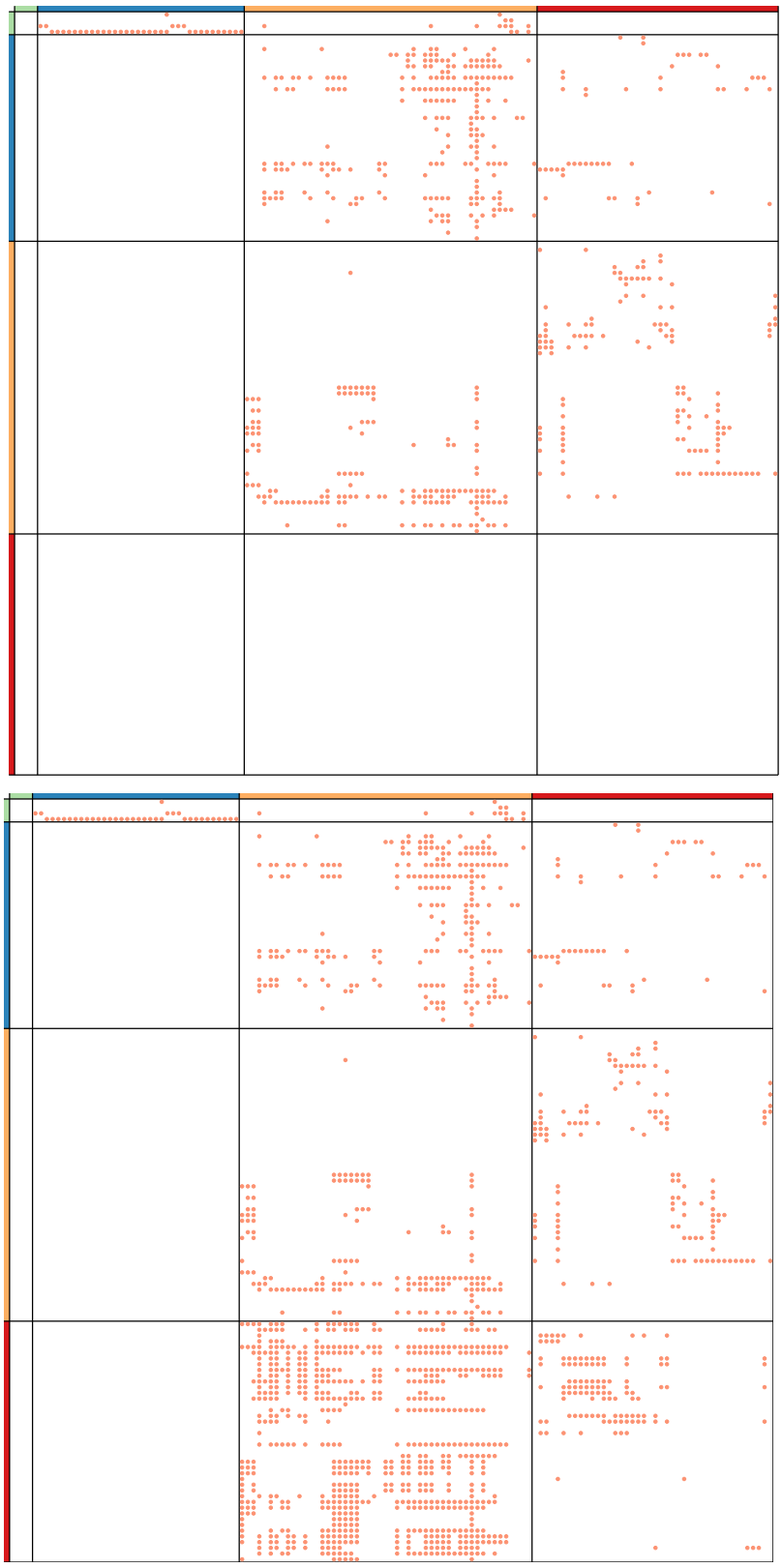


Figure 3.14: As fig. 3.8, but showing the Ythan network.

3.22 Empirical Network Adjacency Matrices – Grouped by Model

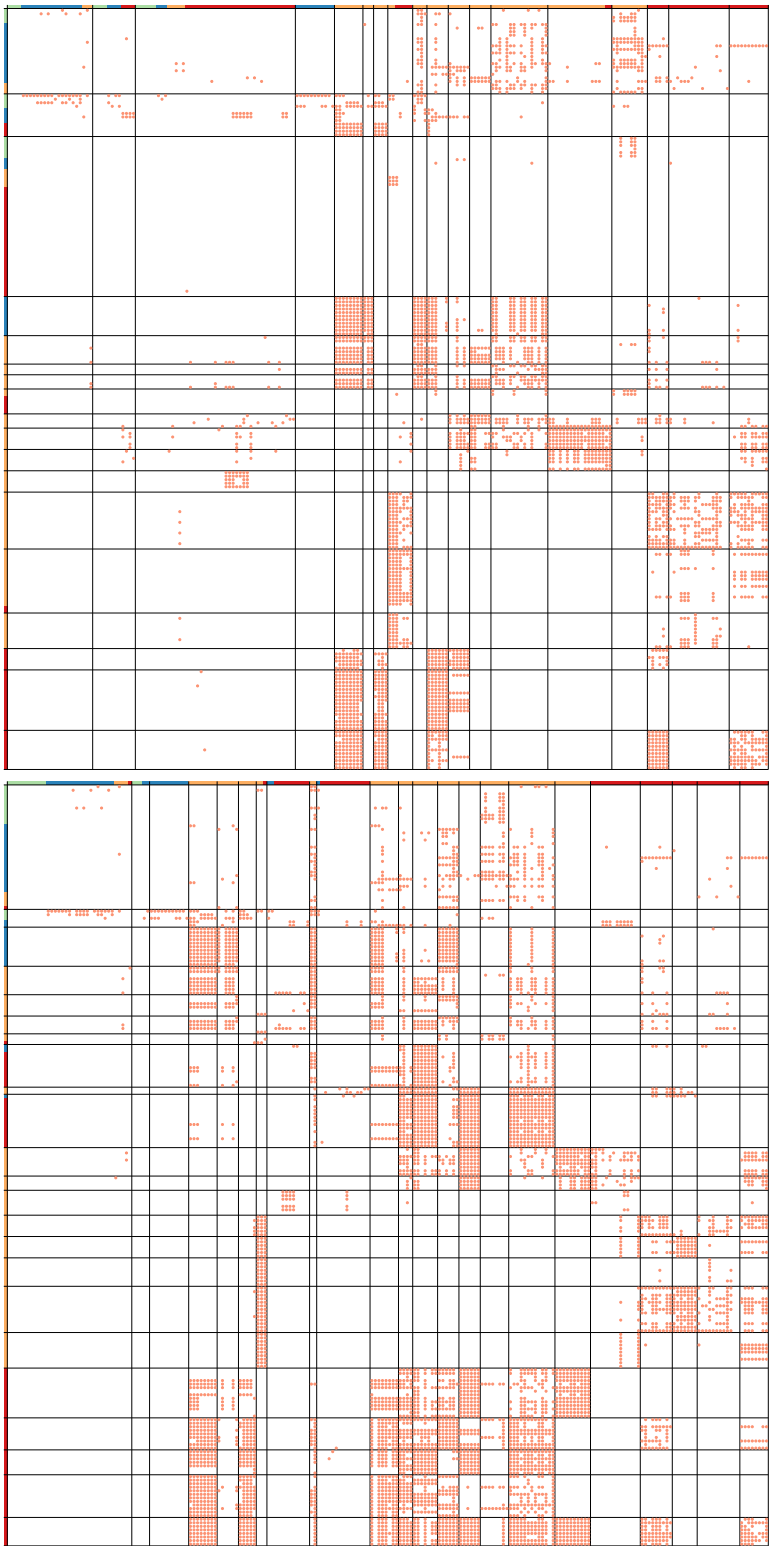


Figure 3.15: Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Punta Banda network.

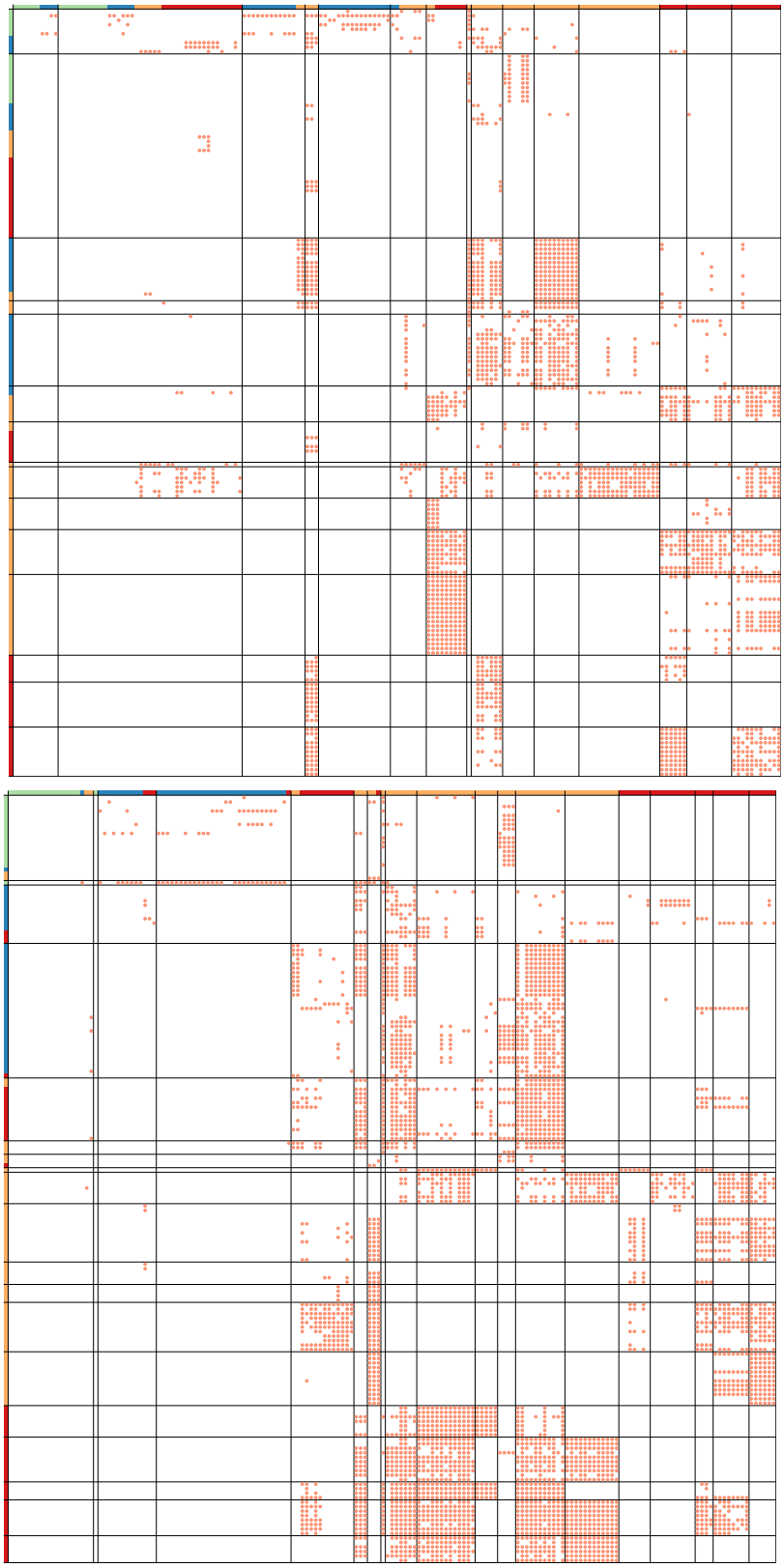


Figure 3.16: As fig. 3.1 in main text. Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Bahia Falsa network.

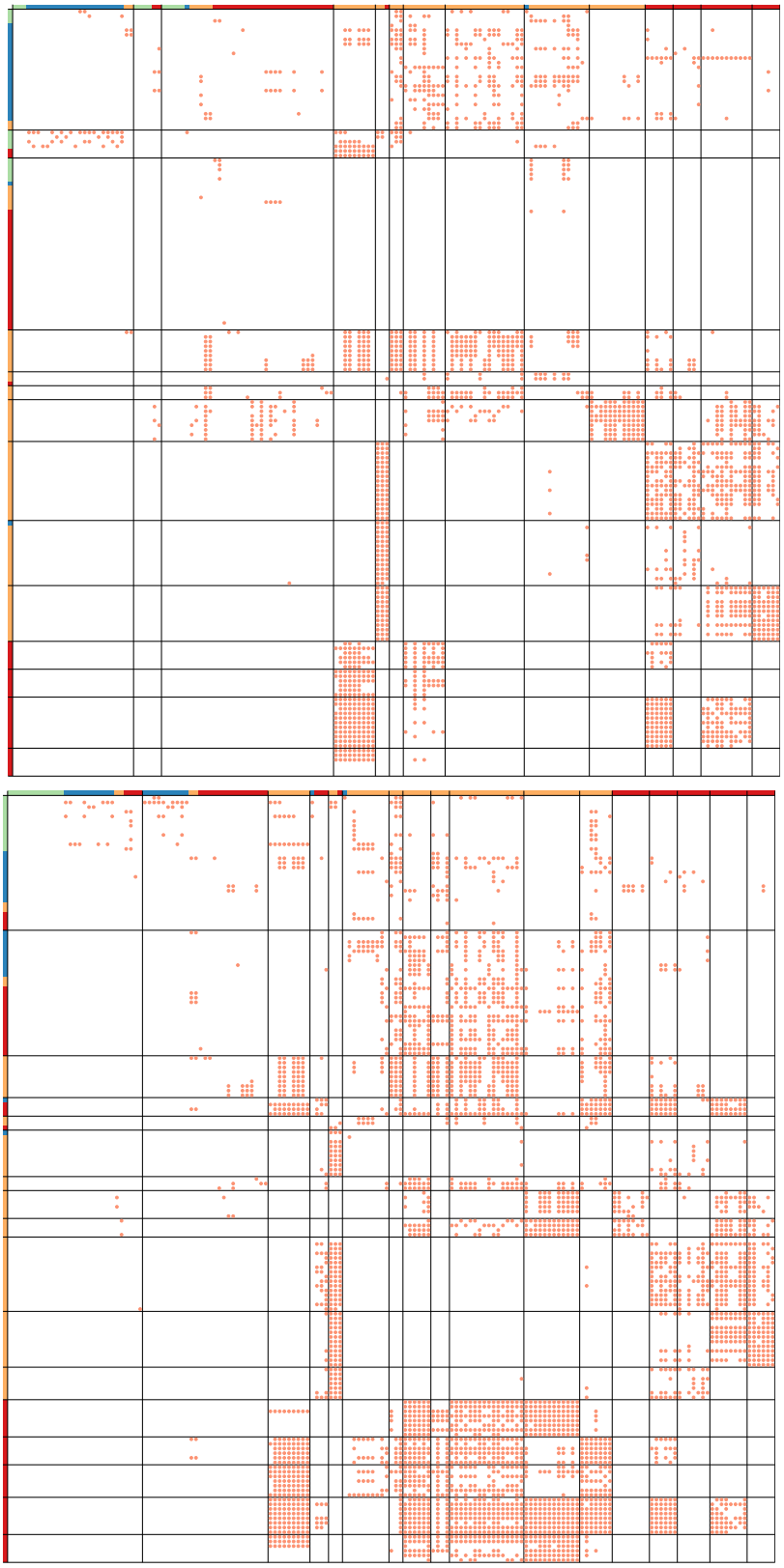


Figure 3.17: As fig. 3.1 in main text. Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Carpinteria network.

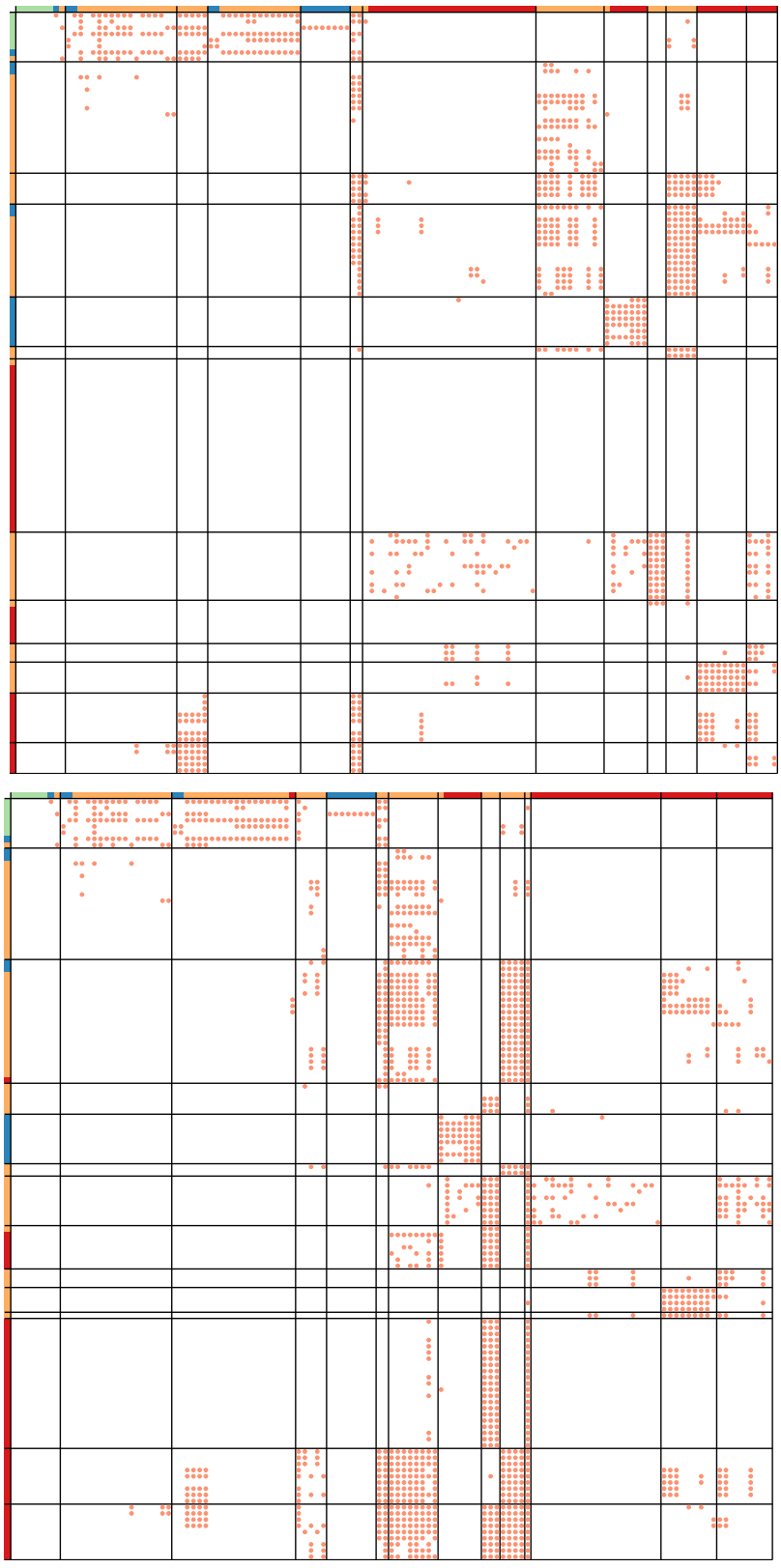


Figure 3.18: As fig. 3.1 in main text. Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Flensburg network.

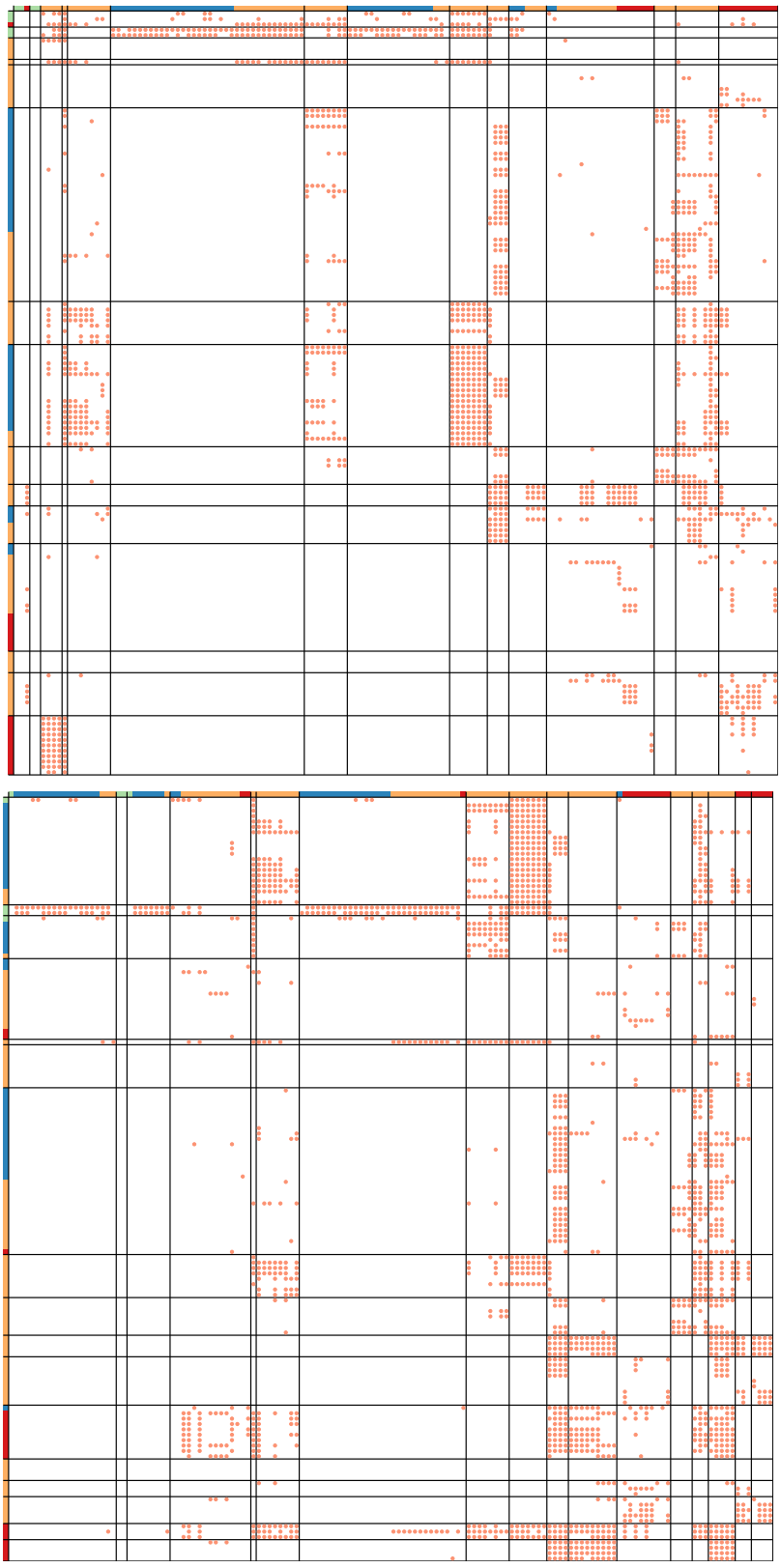


Figure 3.19: As fig. 3.1 in main text. Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Otago network.

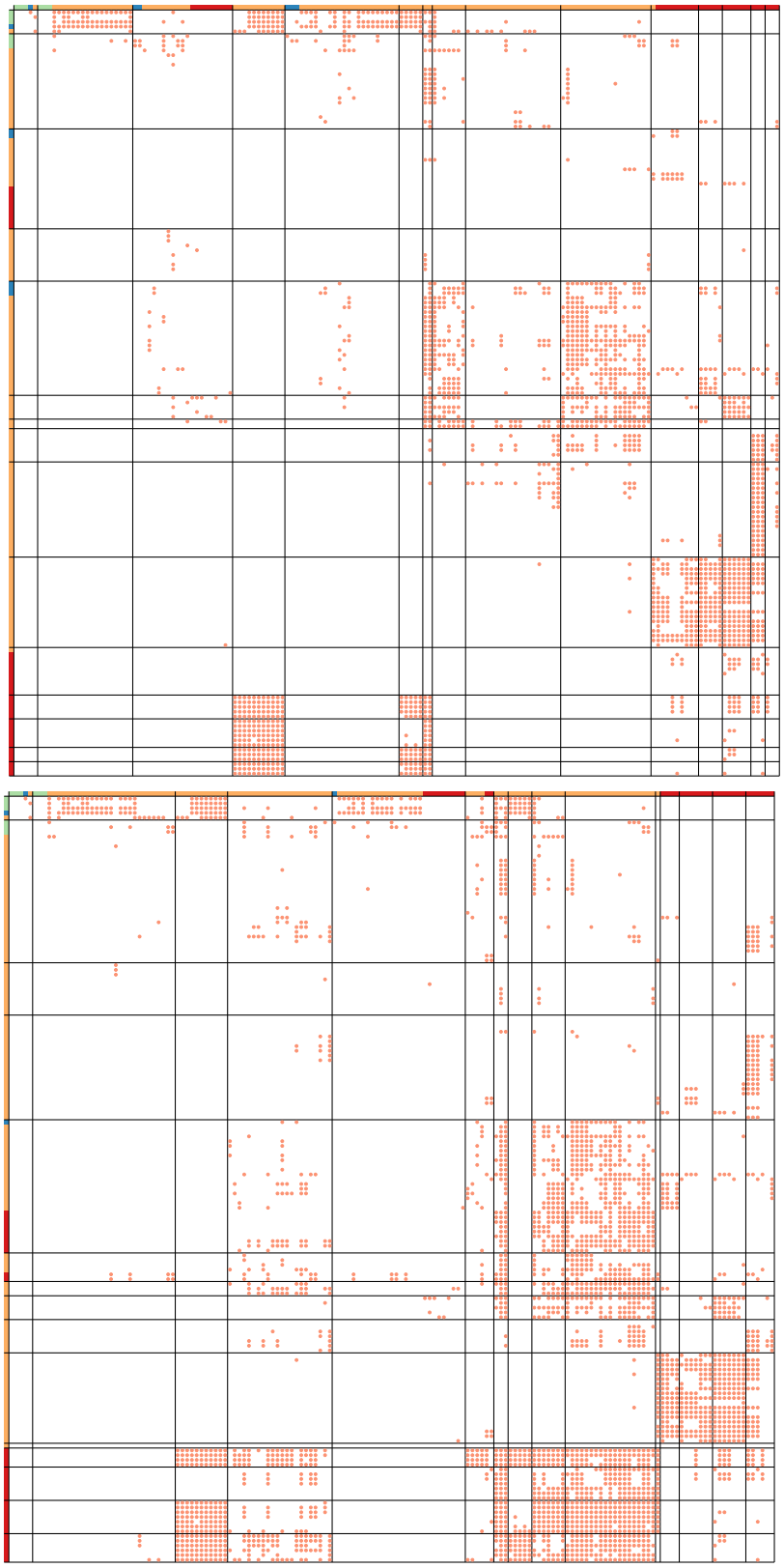


Figure 3.20: As fig. 3.1 in main text. Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Sylt network.

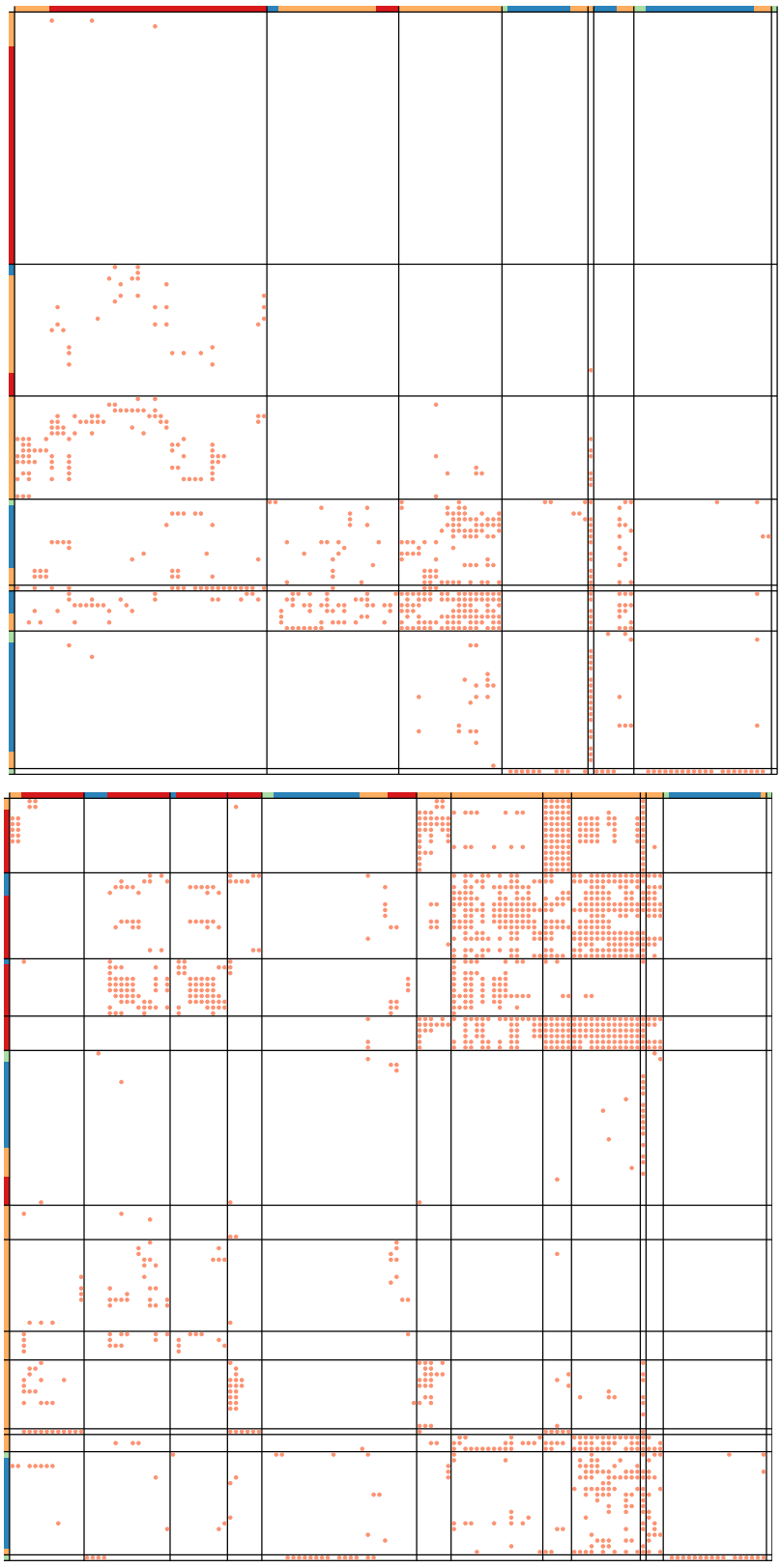
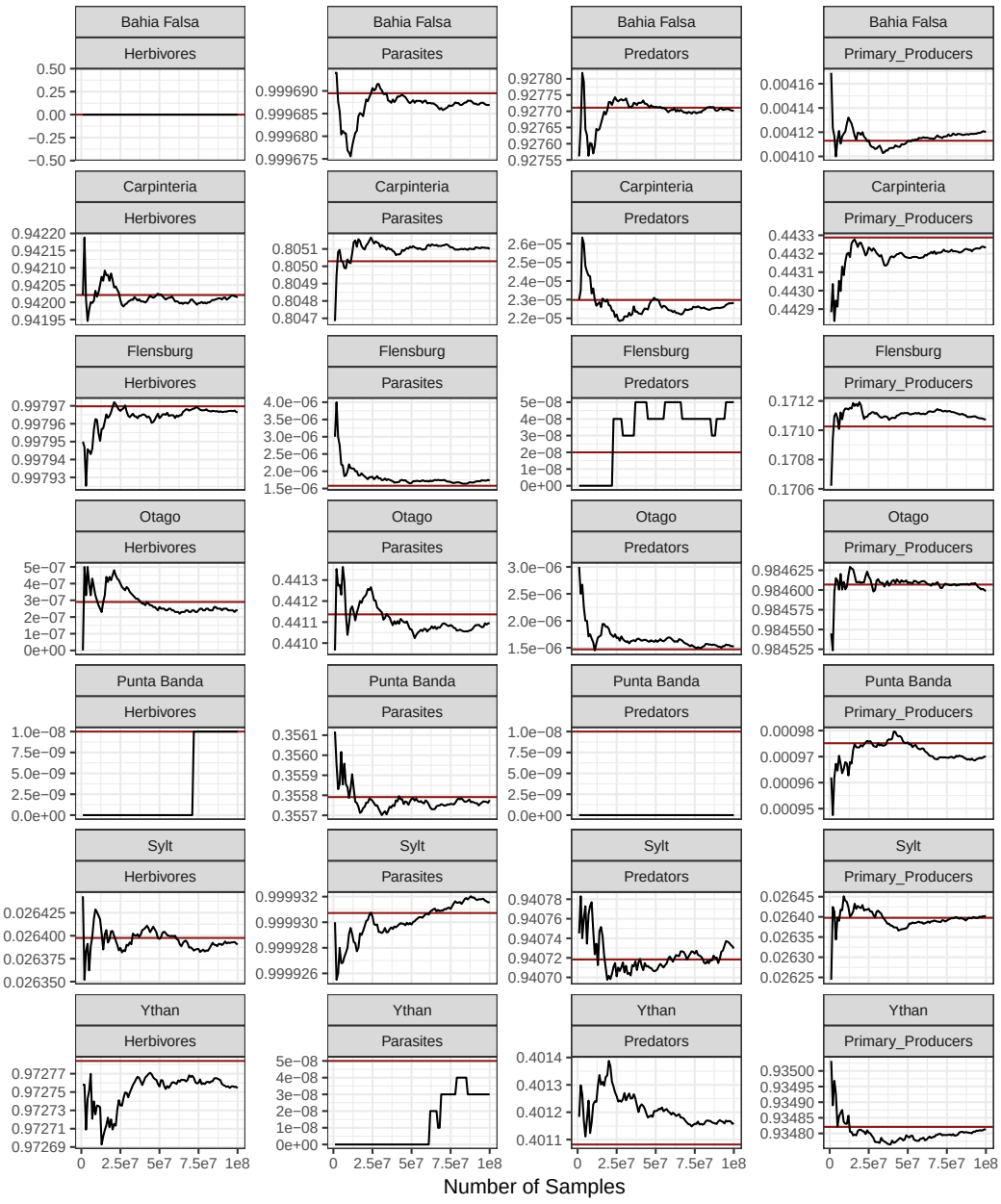
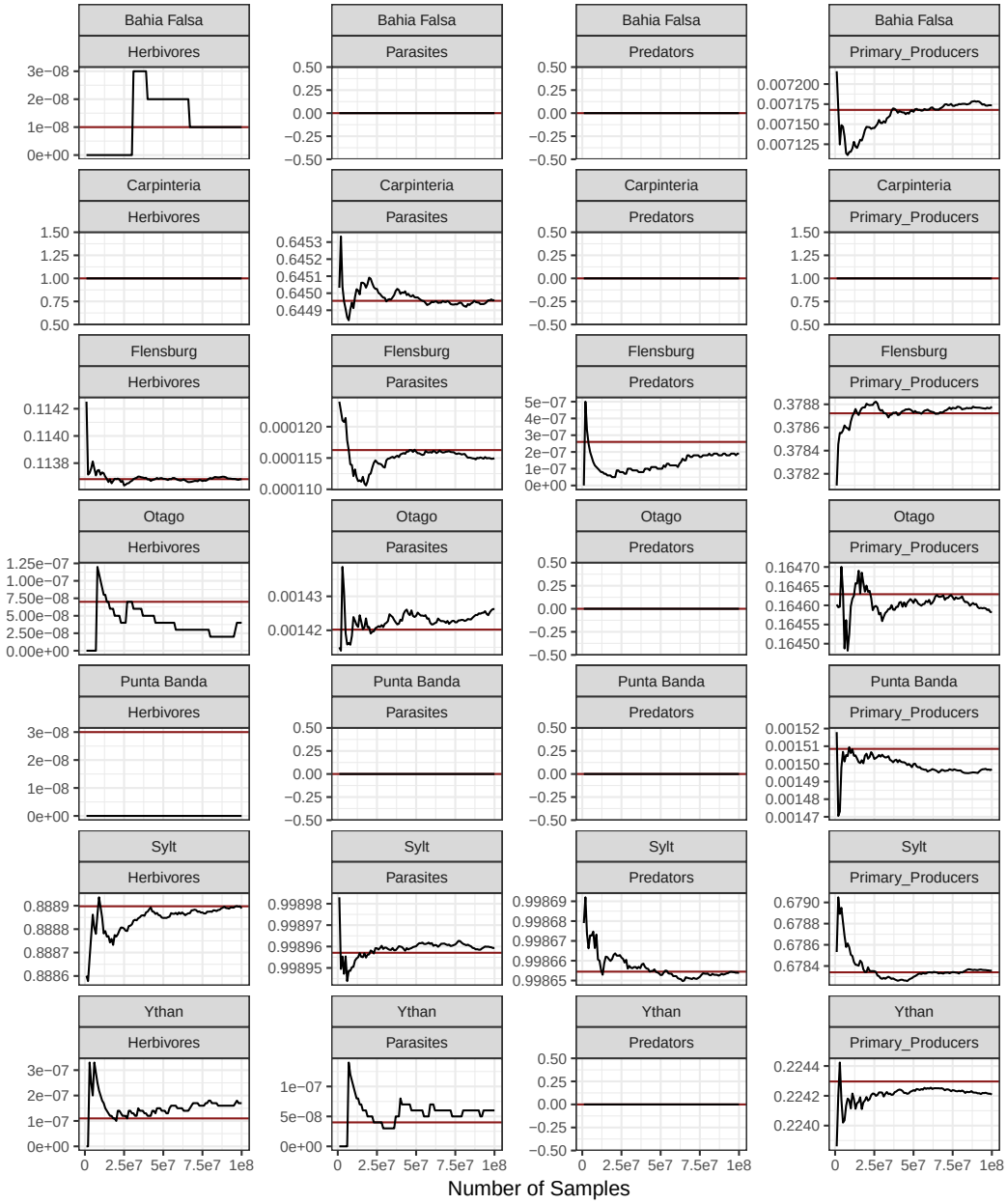


Figure 3.21: As fig. 3.1 in main text. Group model results for uncorrected group model and either including concomitant links (bottom) or not (top) for the Ythan network.

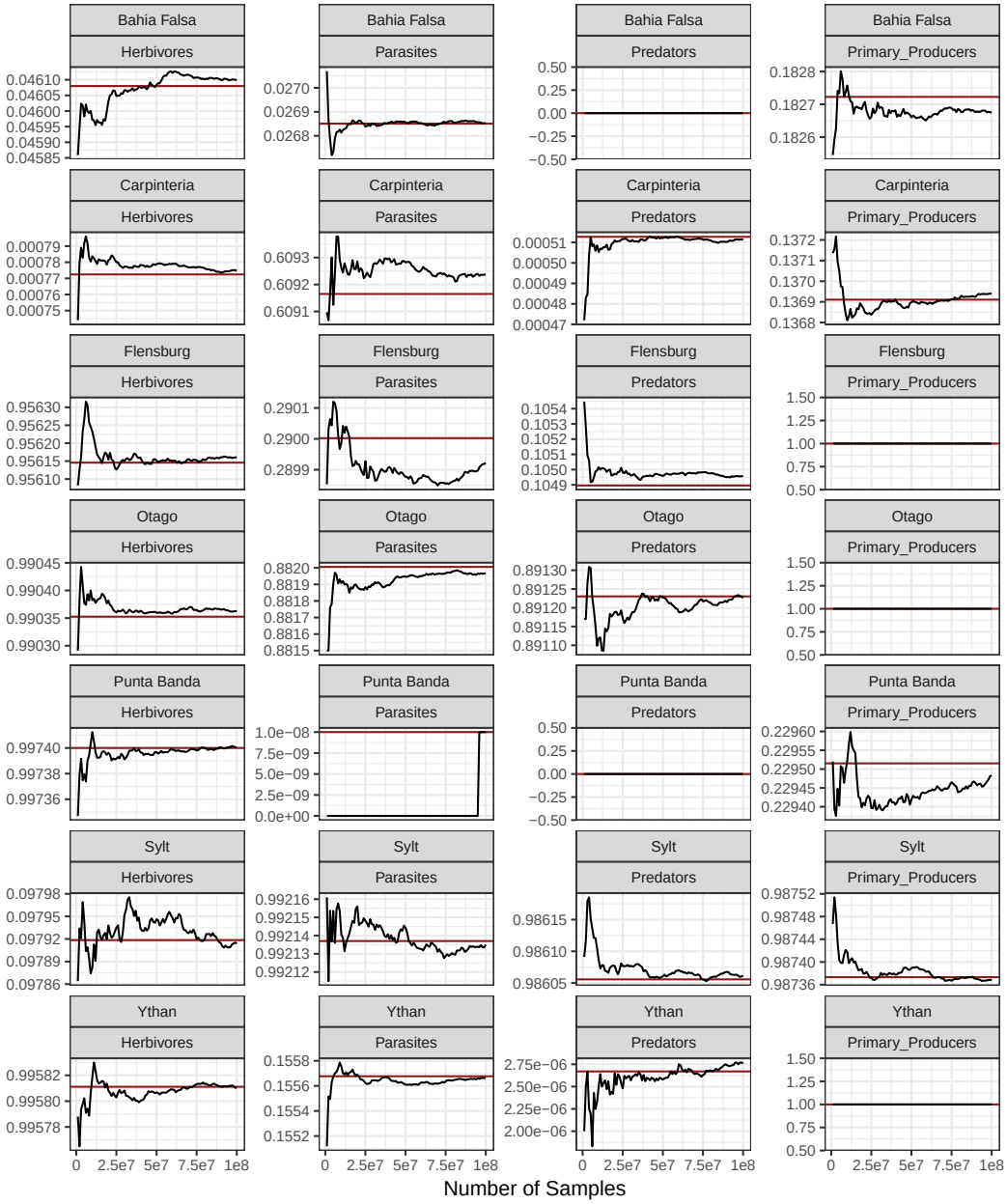
3.23 Imbalance Sampling Convergence



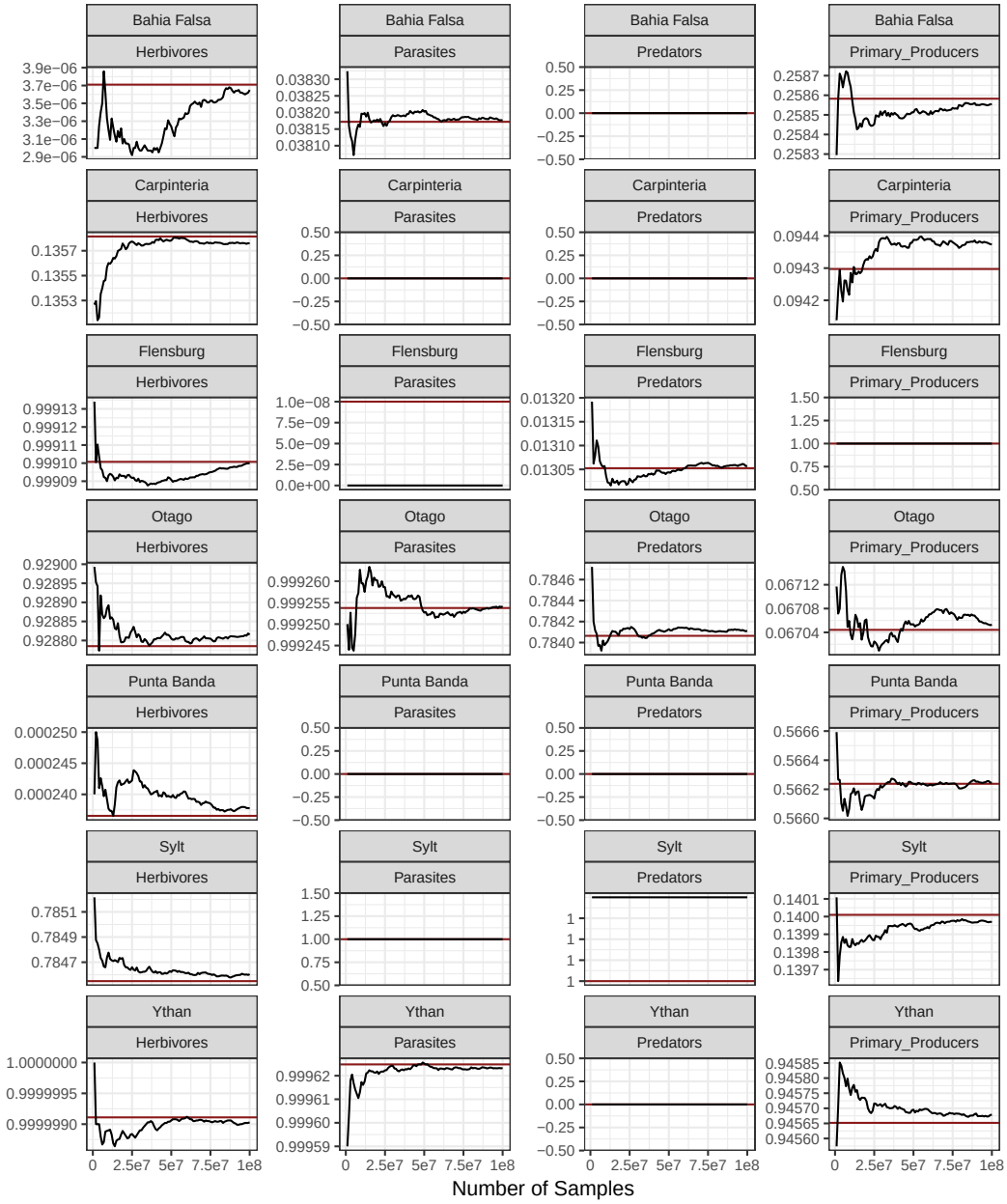
Continued on following page.



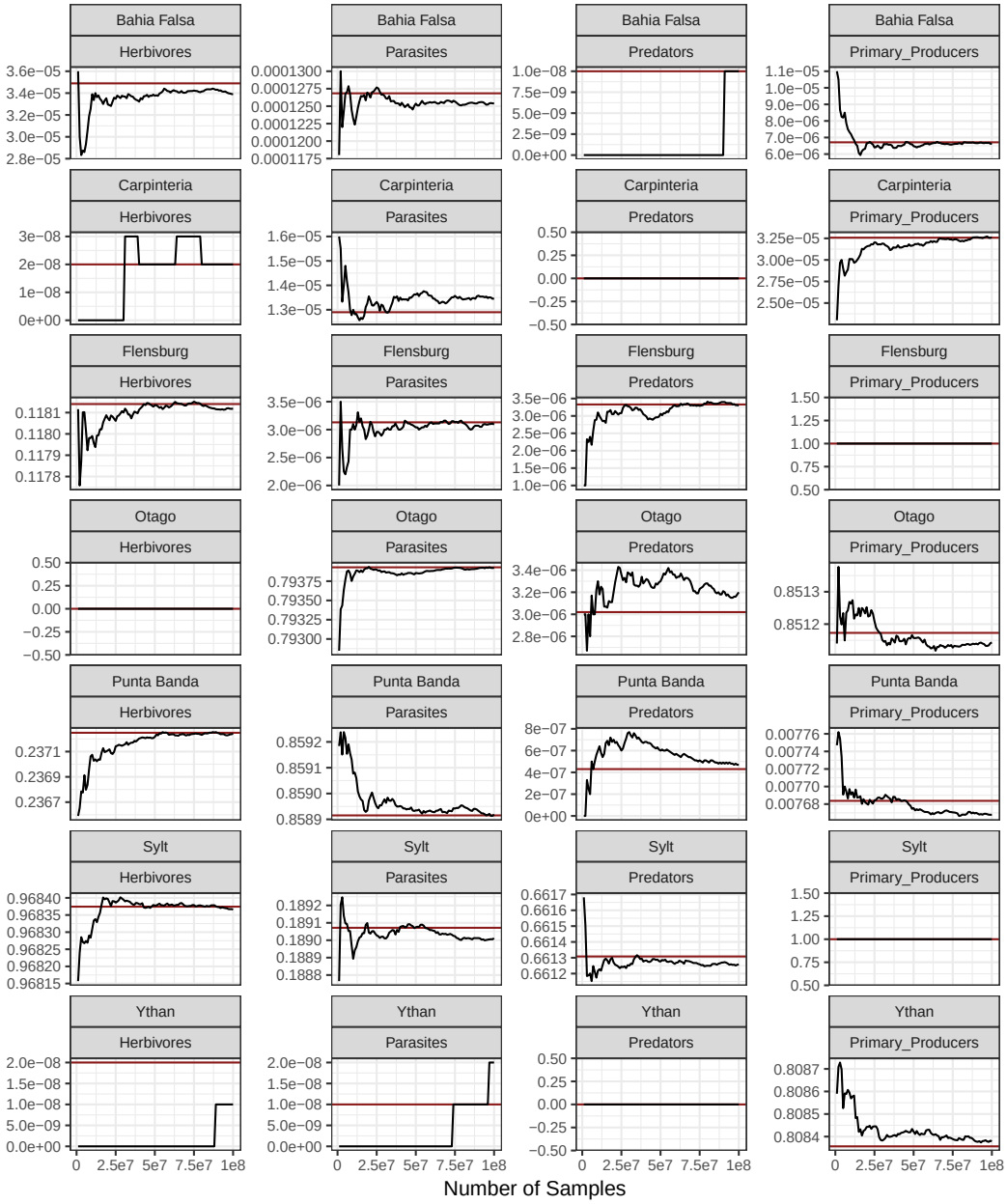
Continued on following page.



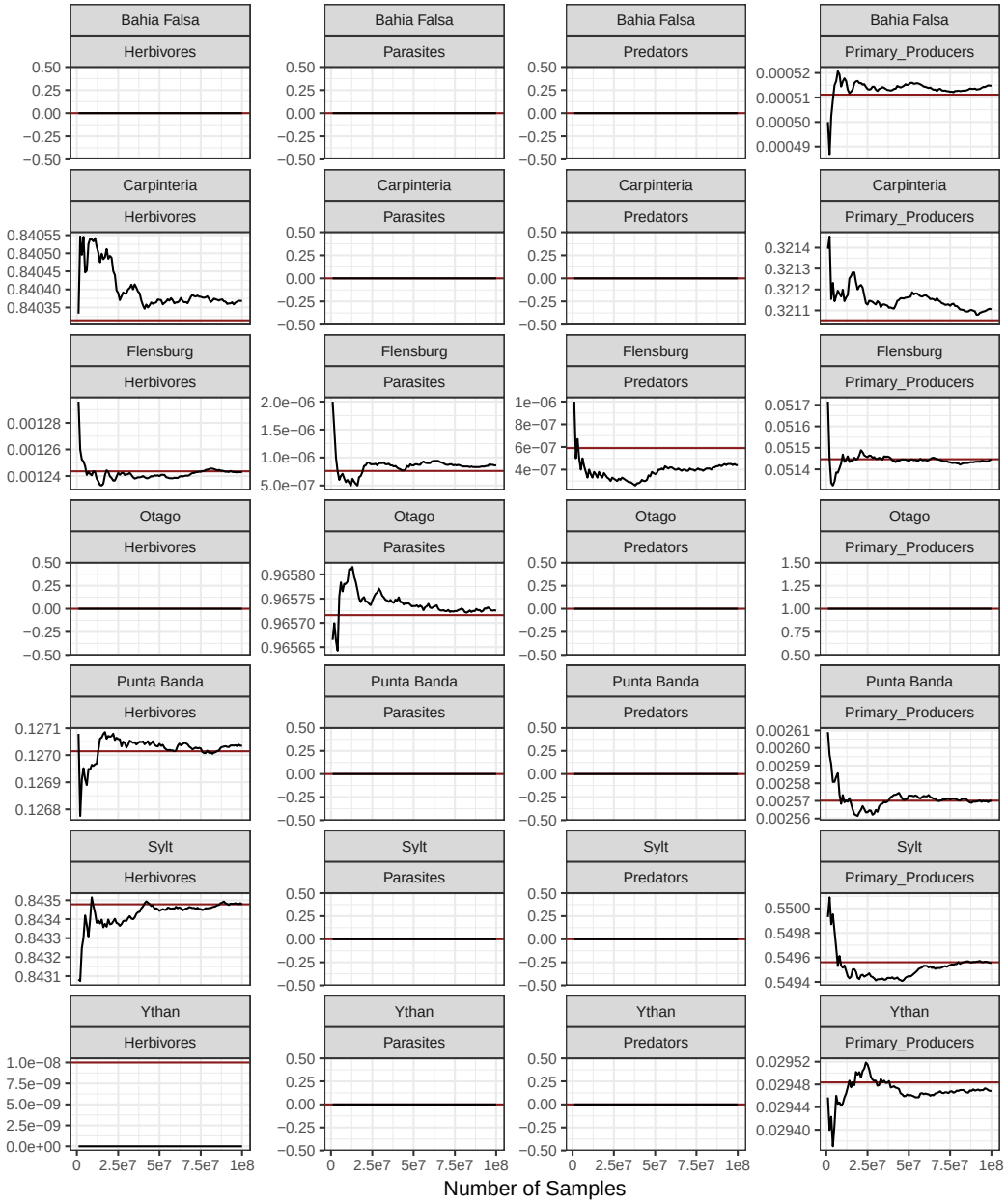
Continued on following page.



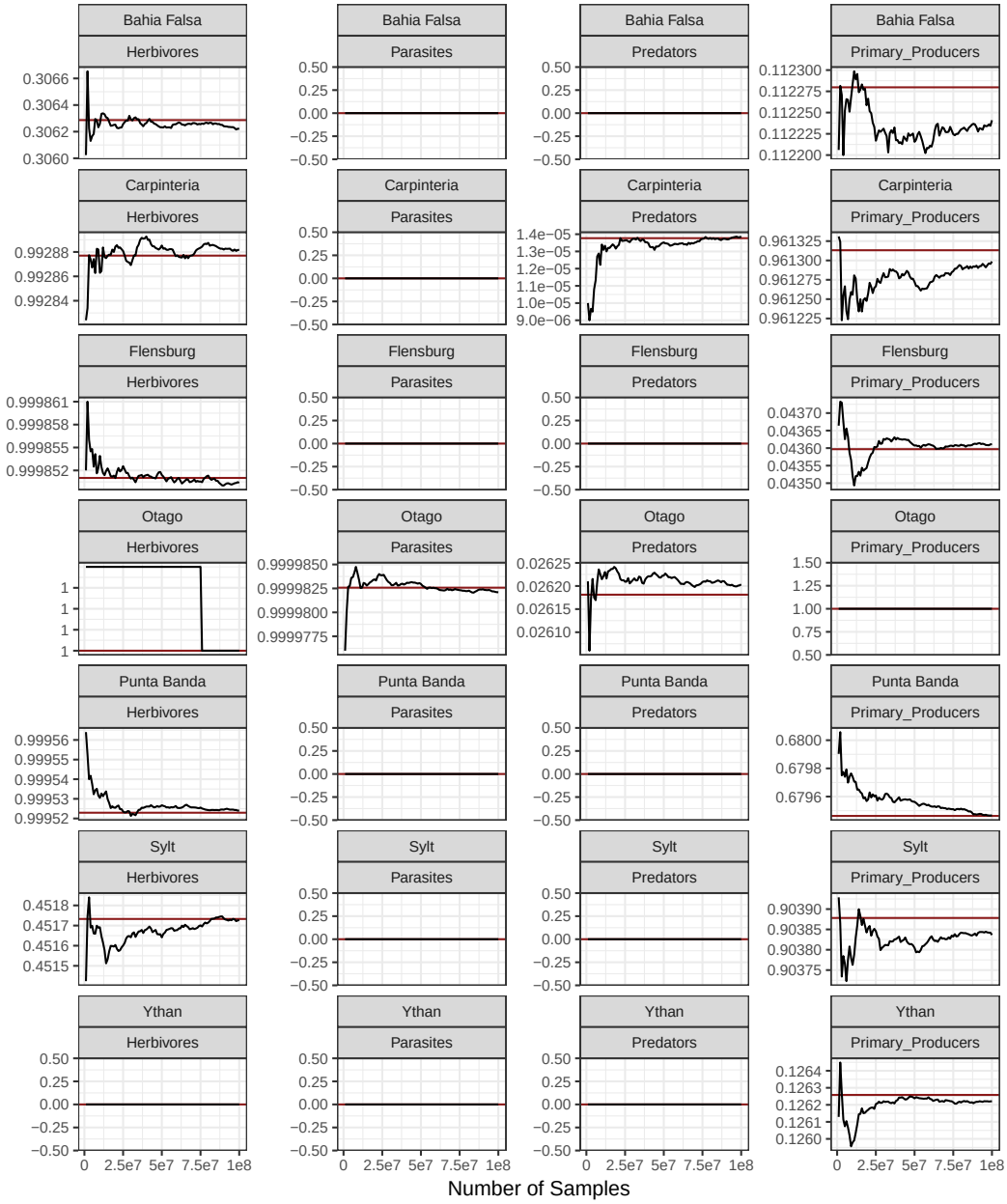
Continued on following page.



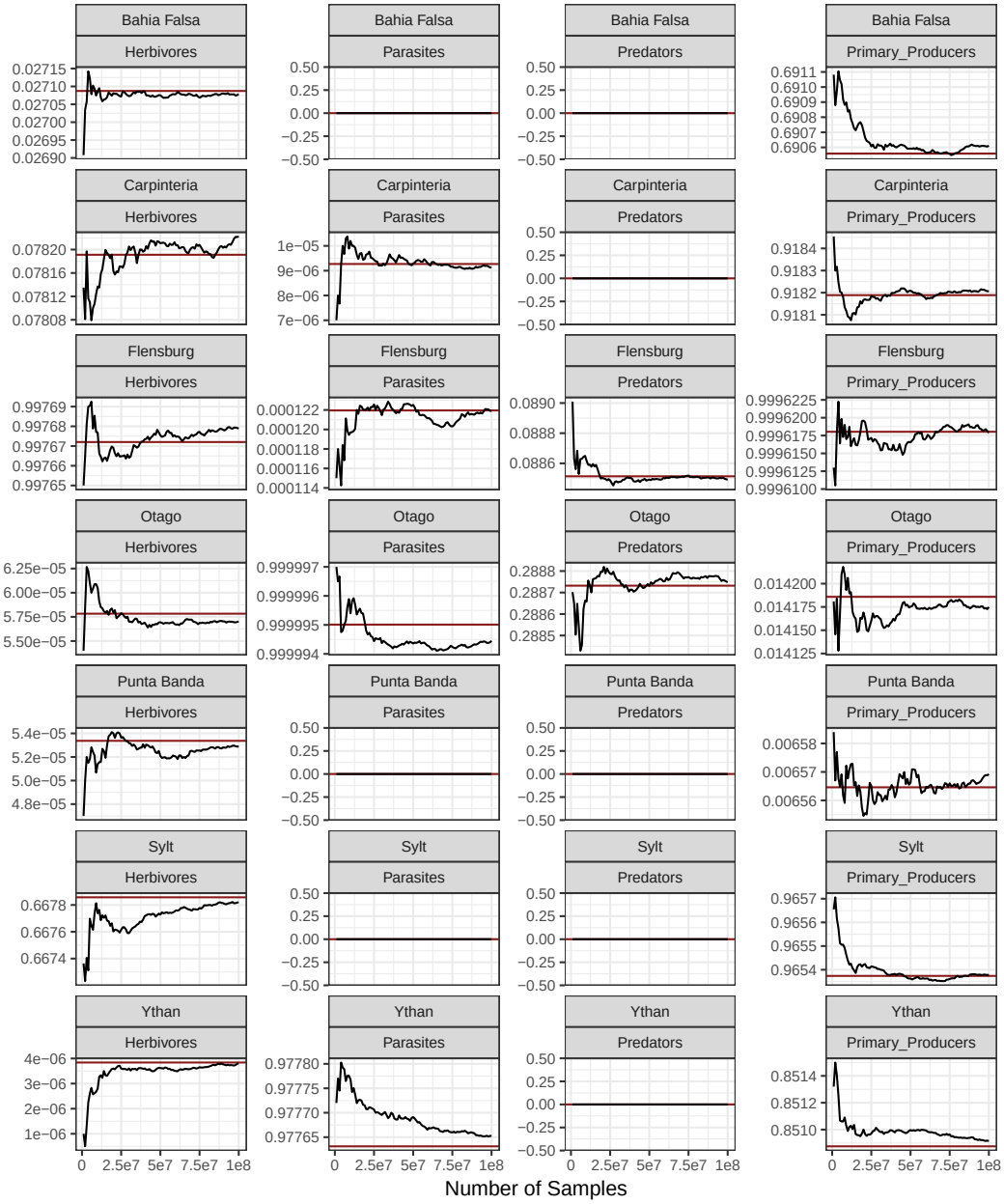
Continued on following page.



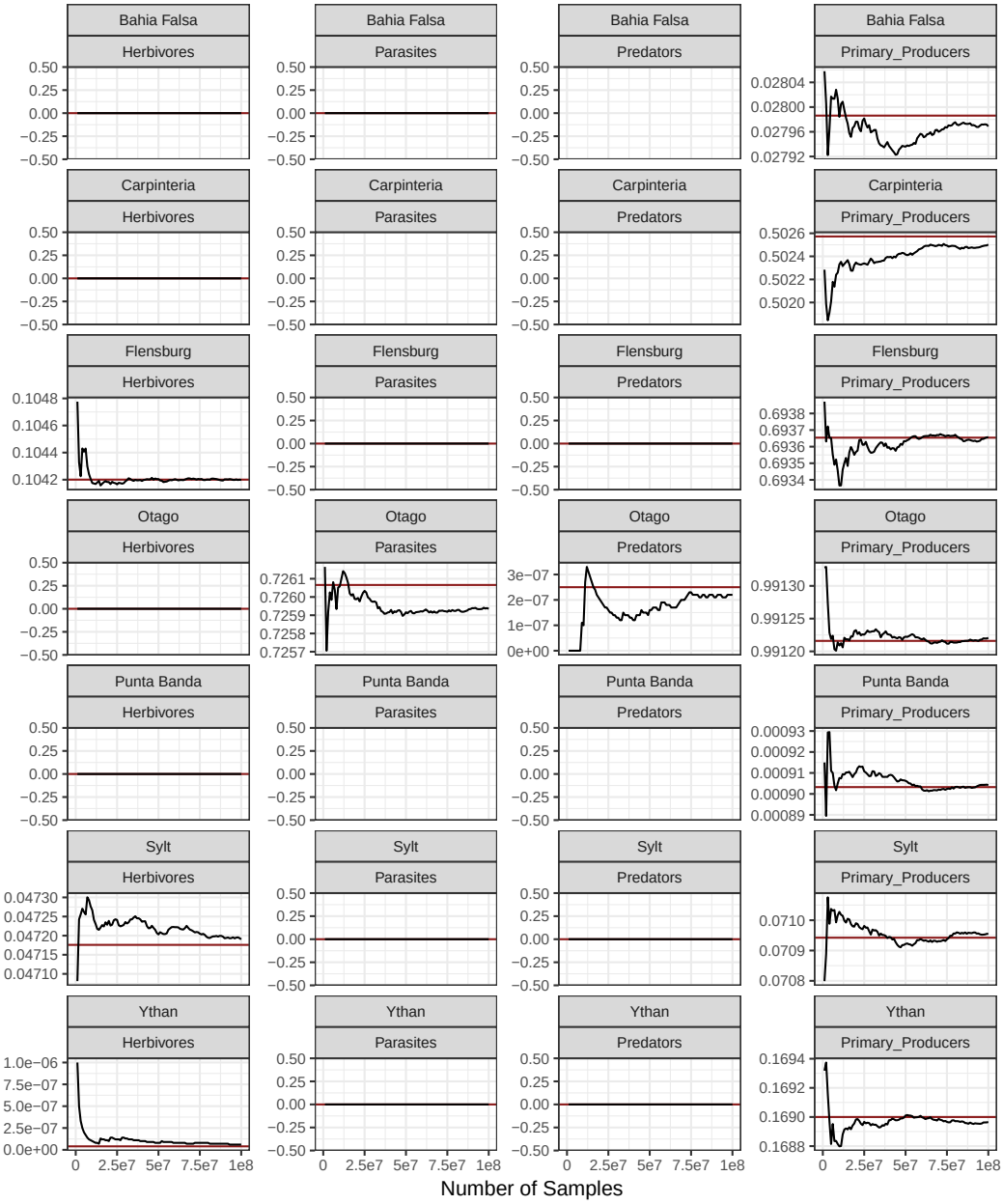
Continued on following page.



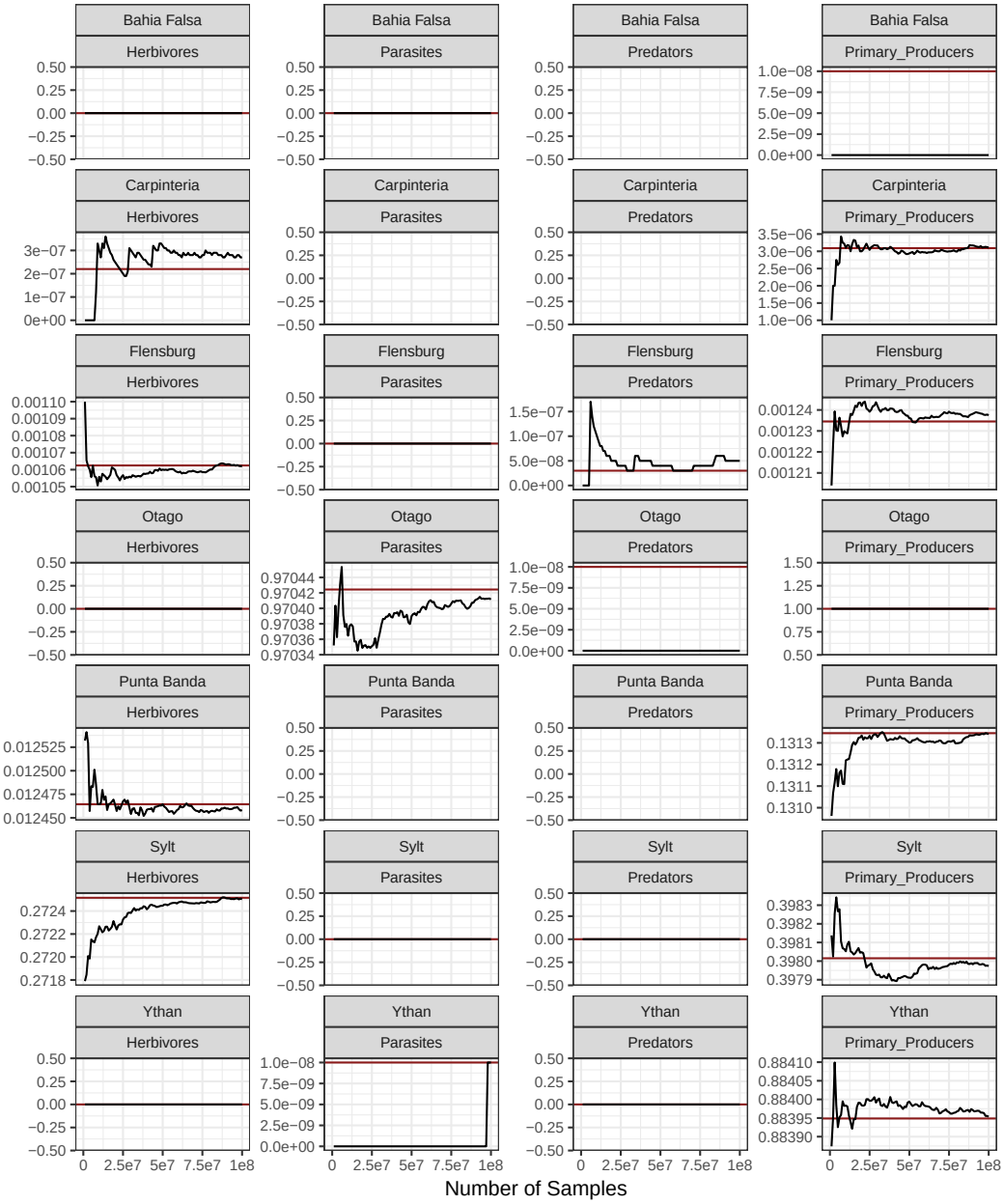
Continued on following page.



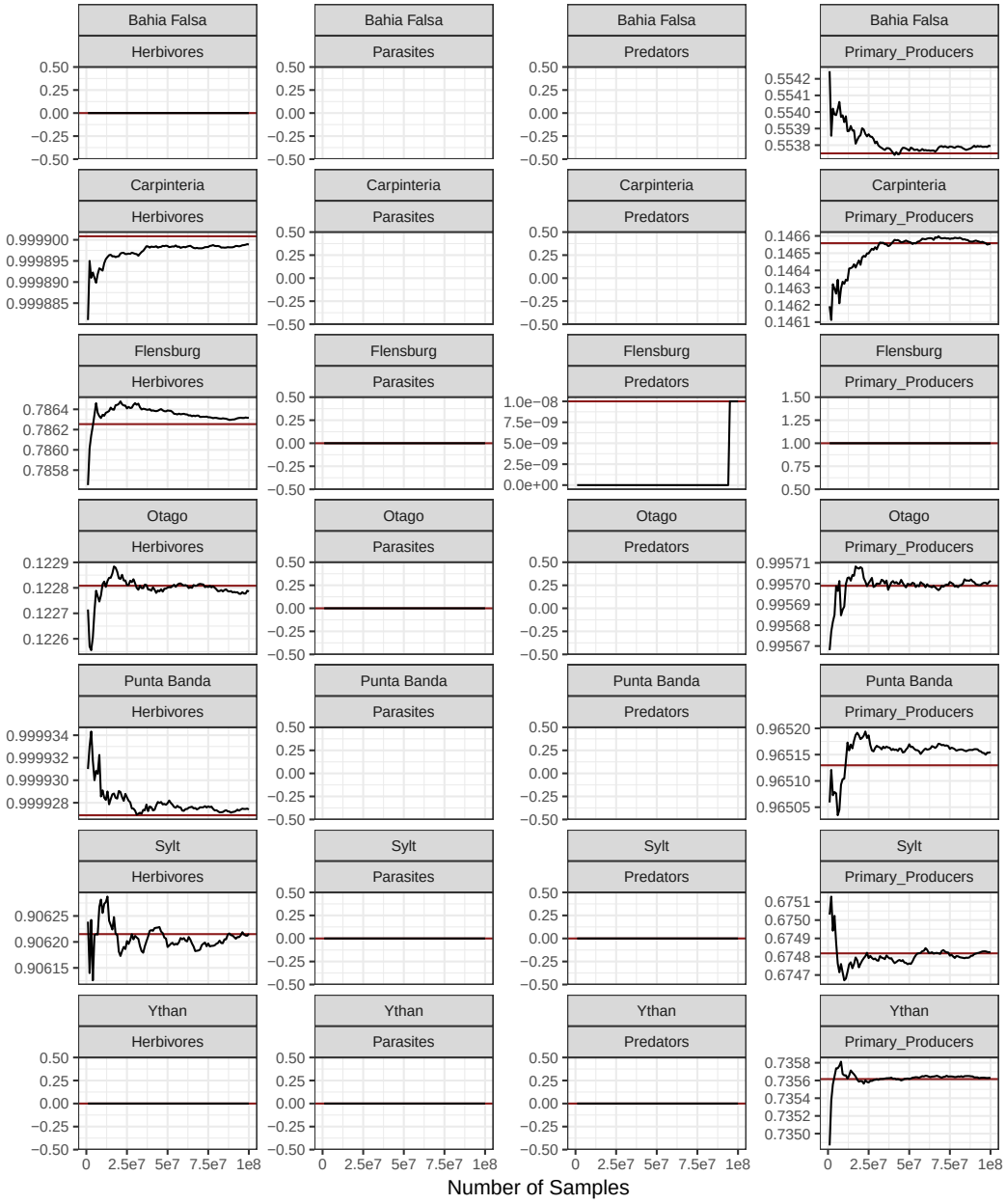
Continued on following page.



Continued on following page.



Continued on following page.



Continued on following page.

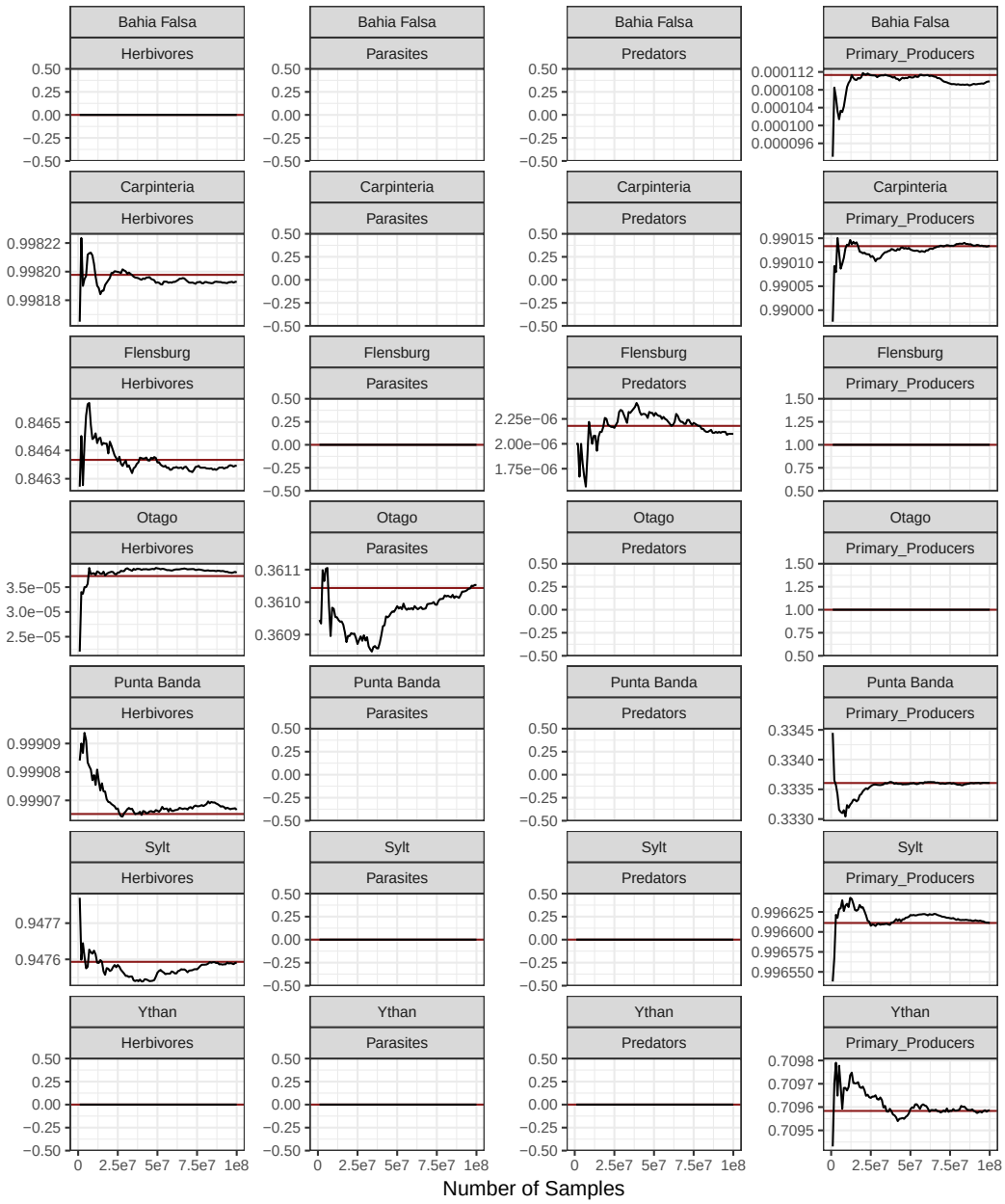


Figure 3.22: Convergence of sampling routine to analytically calculated p-values. Horizontal red line is the analytical value, while the black lines are the sampled value as the number of samples increases. Blank plots indicate computationally infeasible analytical values.

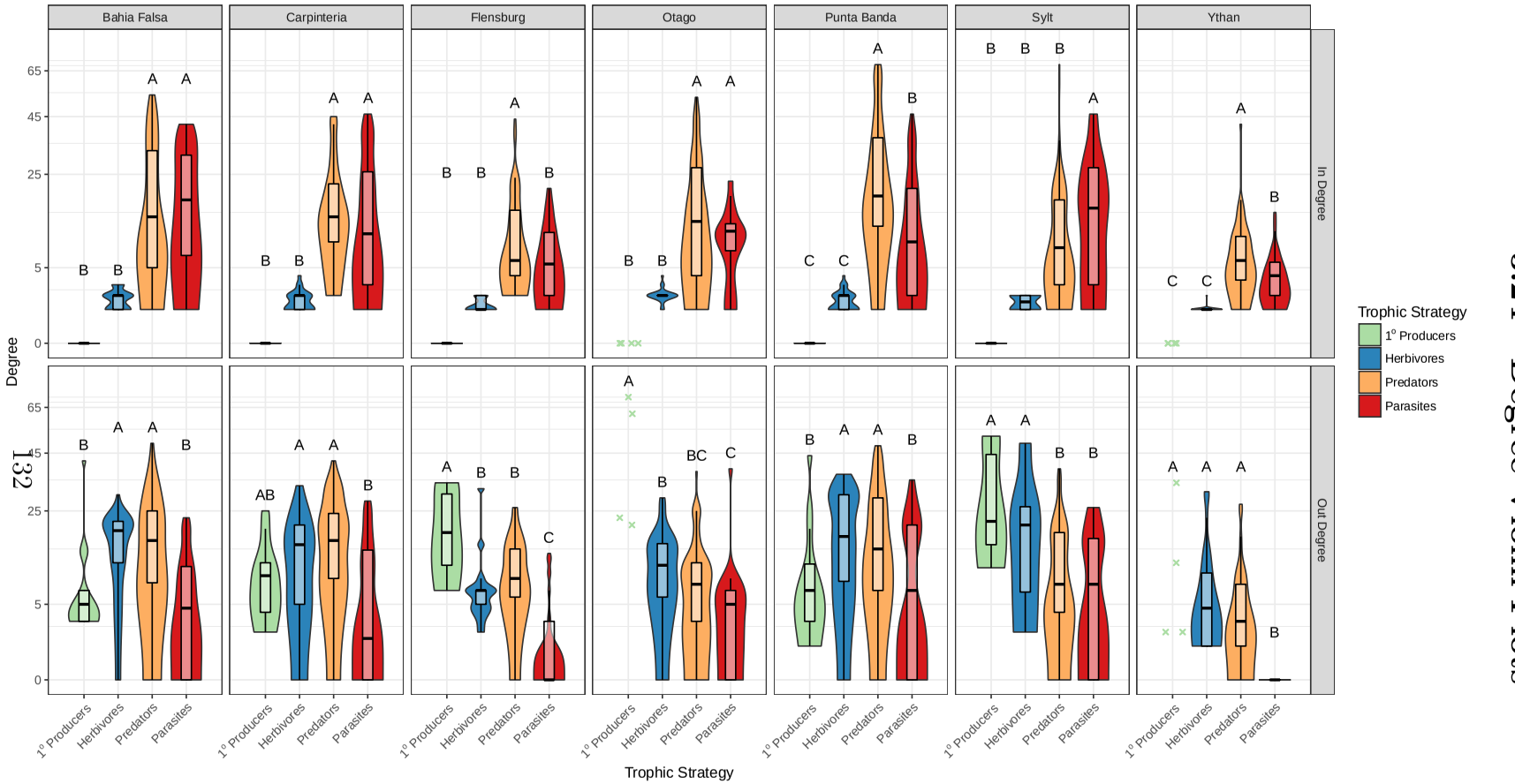


Figure 3.23: Violin and boxplots of in-degree (number of prey) and out-degree (number of predators) for different trophic strategies in all networks *excluding* concomitant predation. Degree is plotted on a square root scale. Boxes indicate the traditional 25^{th} , 50^{th}, and 75^{th} quartiles, with whiskers extending to 1.5 times the inter-quartile range. Above each violin are grouping letters as indicated by a Tukeys HSD (honest significant difference) test.

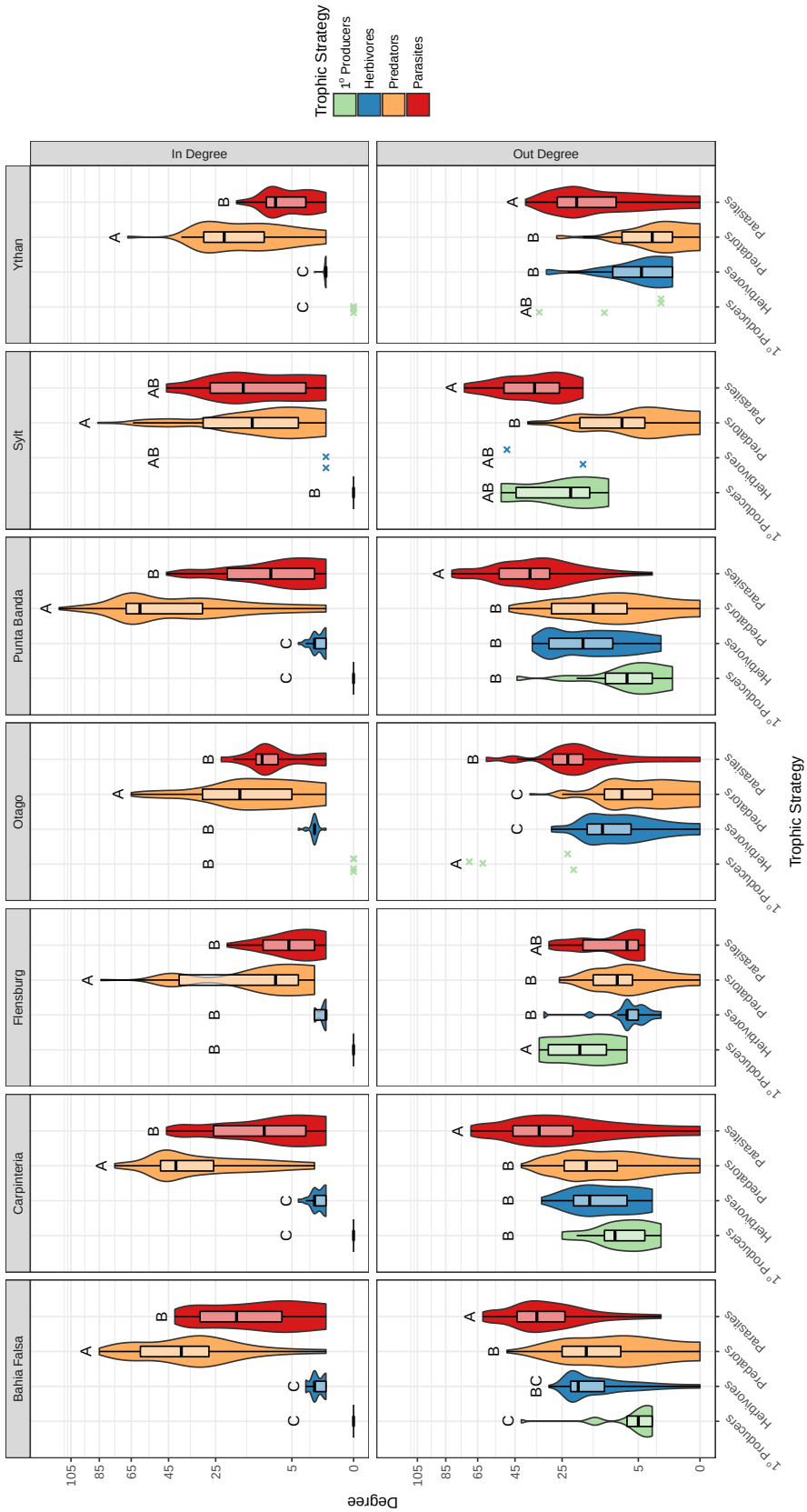


Figure 3.24: As fig. 3.23 but for networks *including* concomitant predation.

3.25 Condensed Network Diagrams

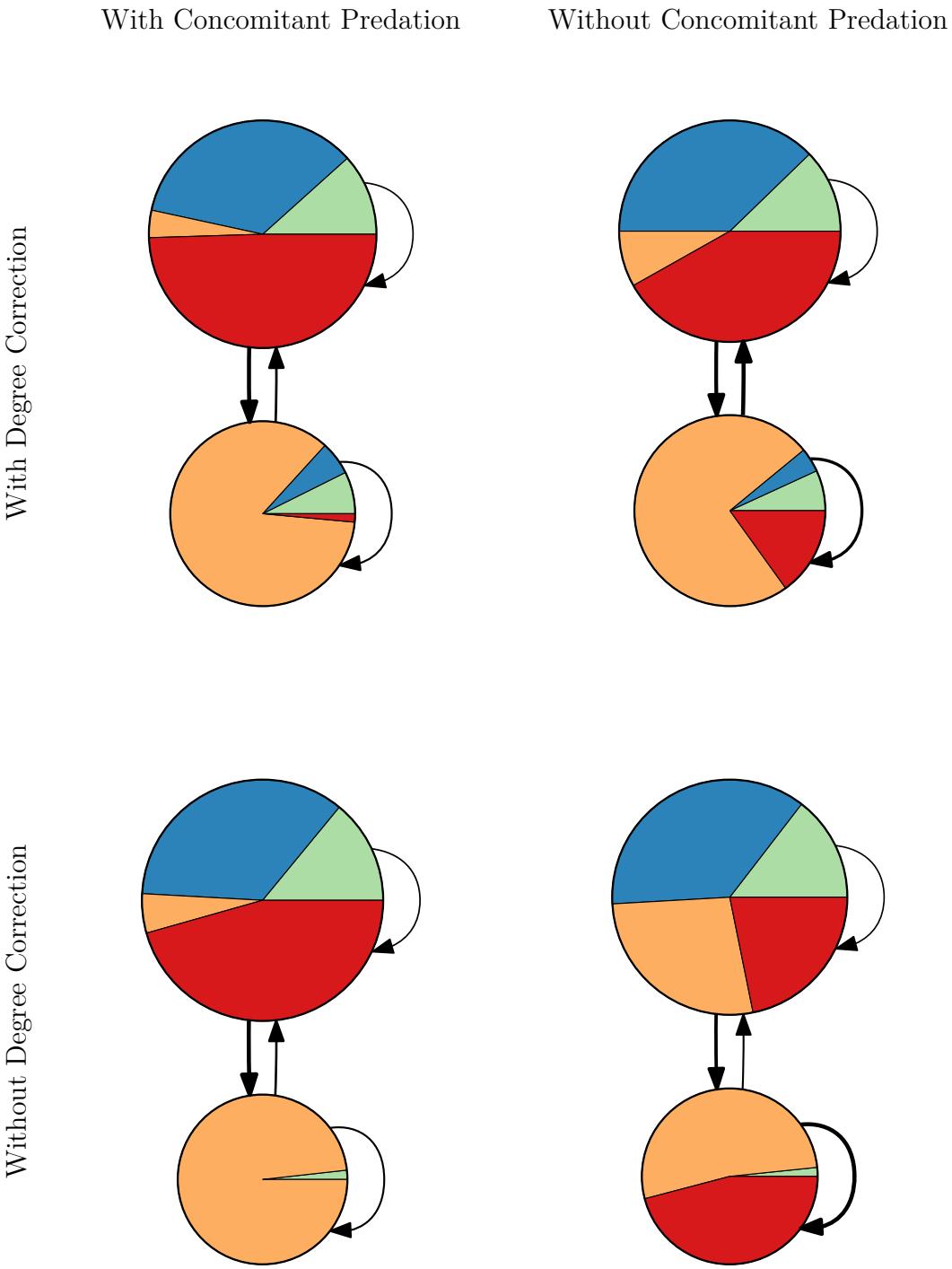


Figure 3.25: As fig. 3.3 in main text, but for the Bahia Falsa network and $g = 2$.

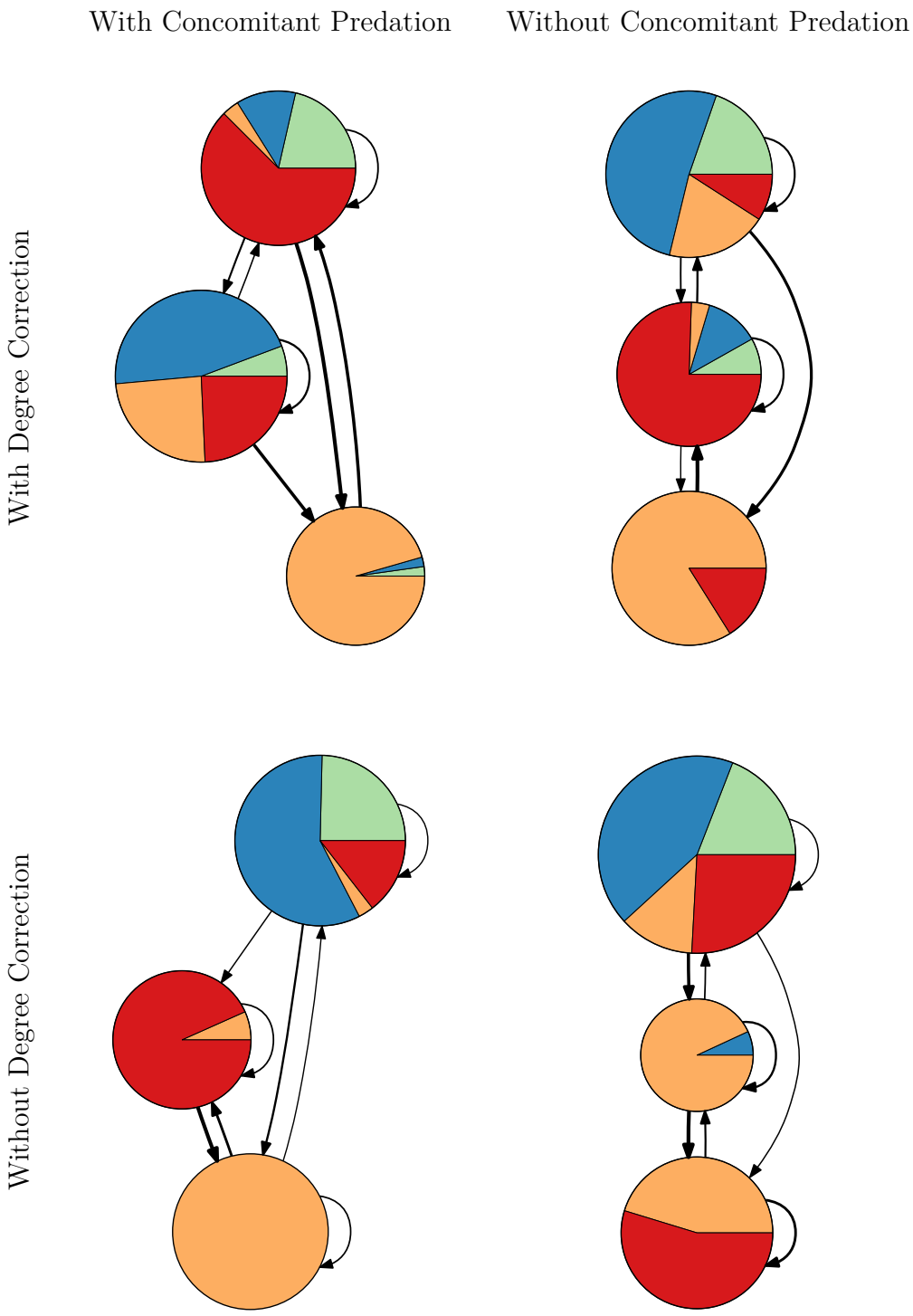


Figure 3.26: As fig. 3.3 in main text, but for the Bahia Falsa network and $g = 3$.

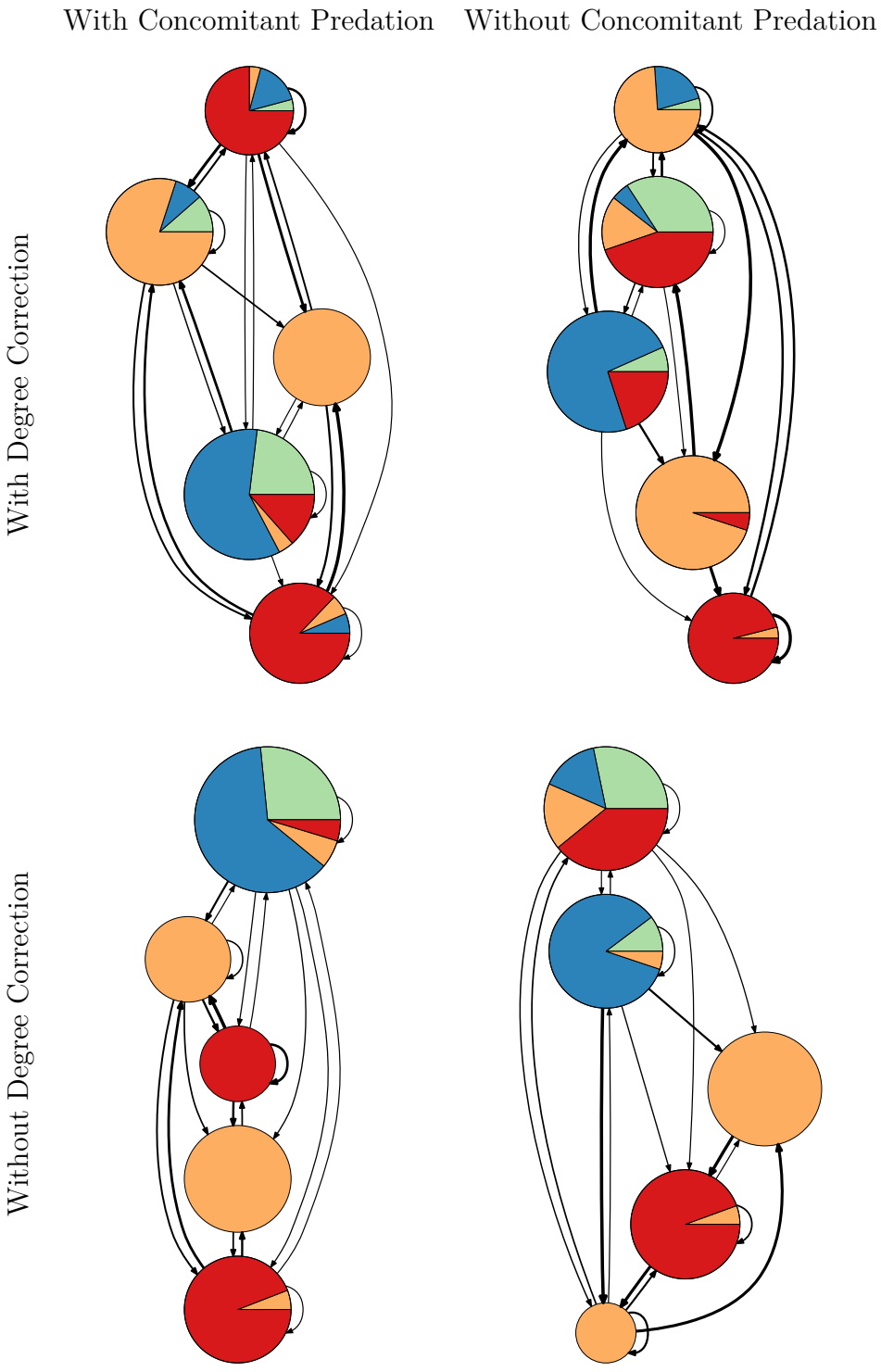


Figure 3.27: As fig. 3.3 in main text, but for the Bahia Falsa network and $g = 5$.

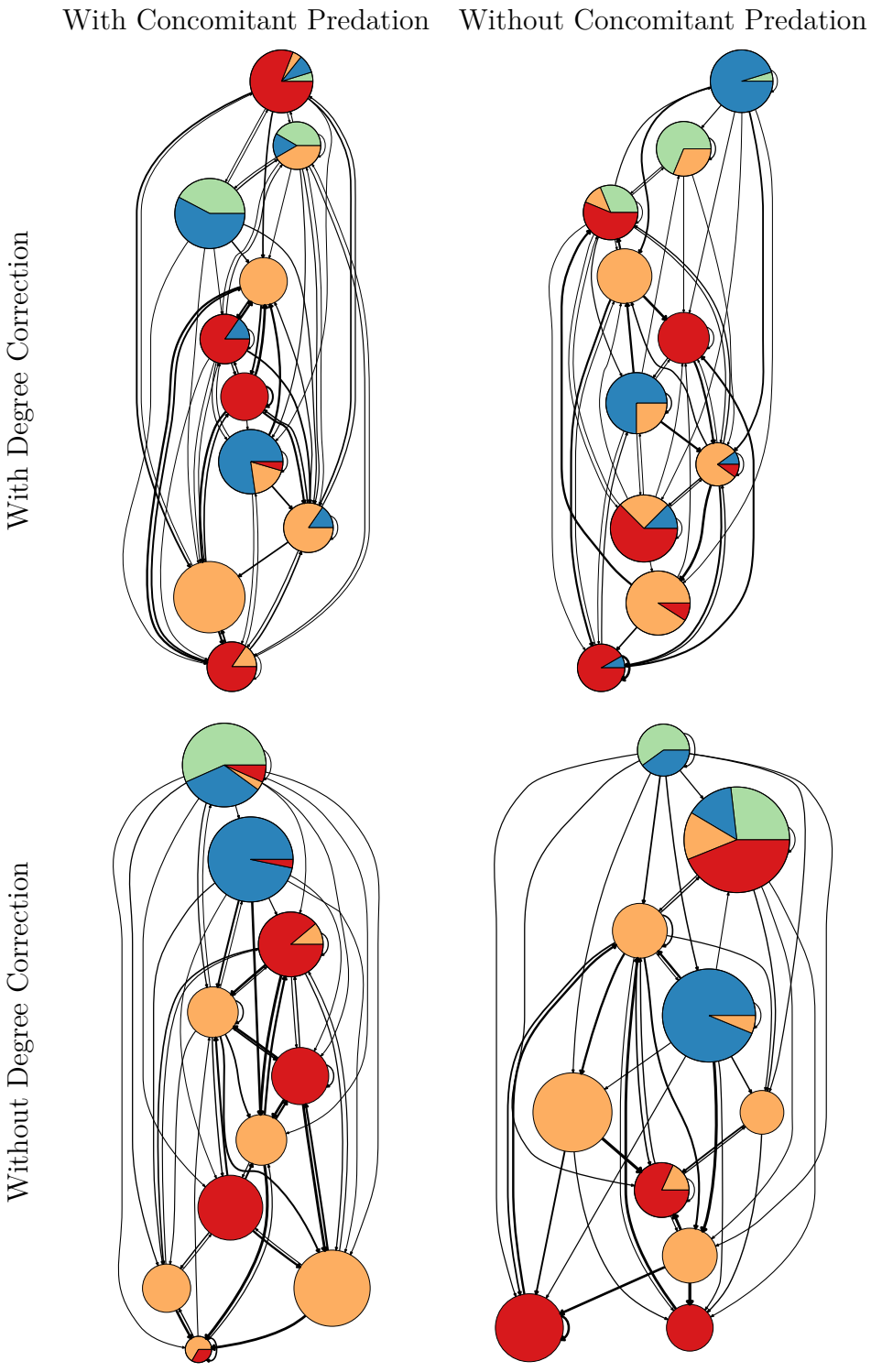


Figure 3.28: As fig. 3.3 in main text, but for the Bahia Falsa network and $g = 10$.

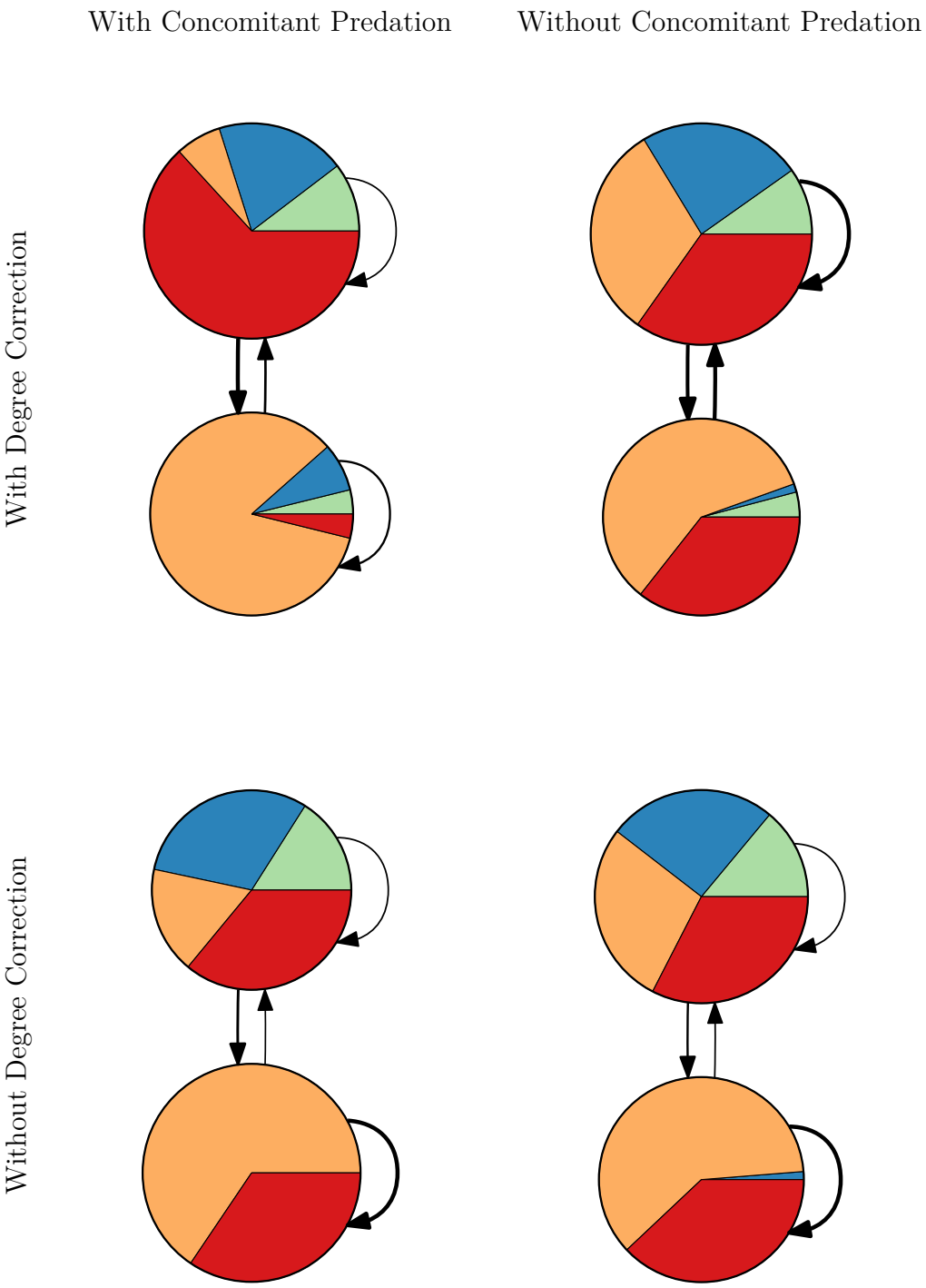


Figure 3.29: As fig. 3.3 in main text, but for the Carpinteria network and $g = 2$.

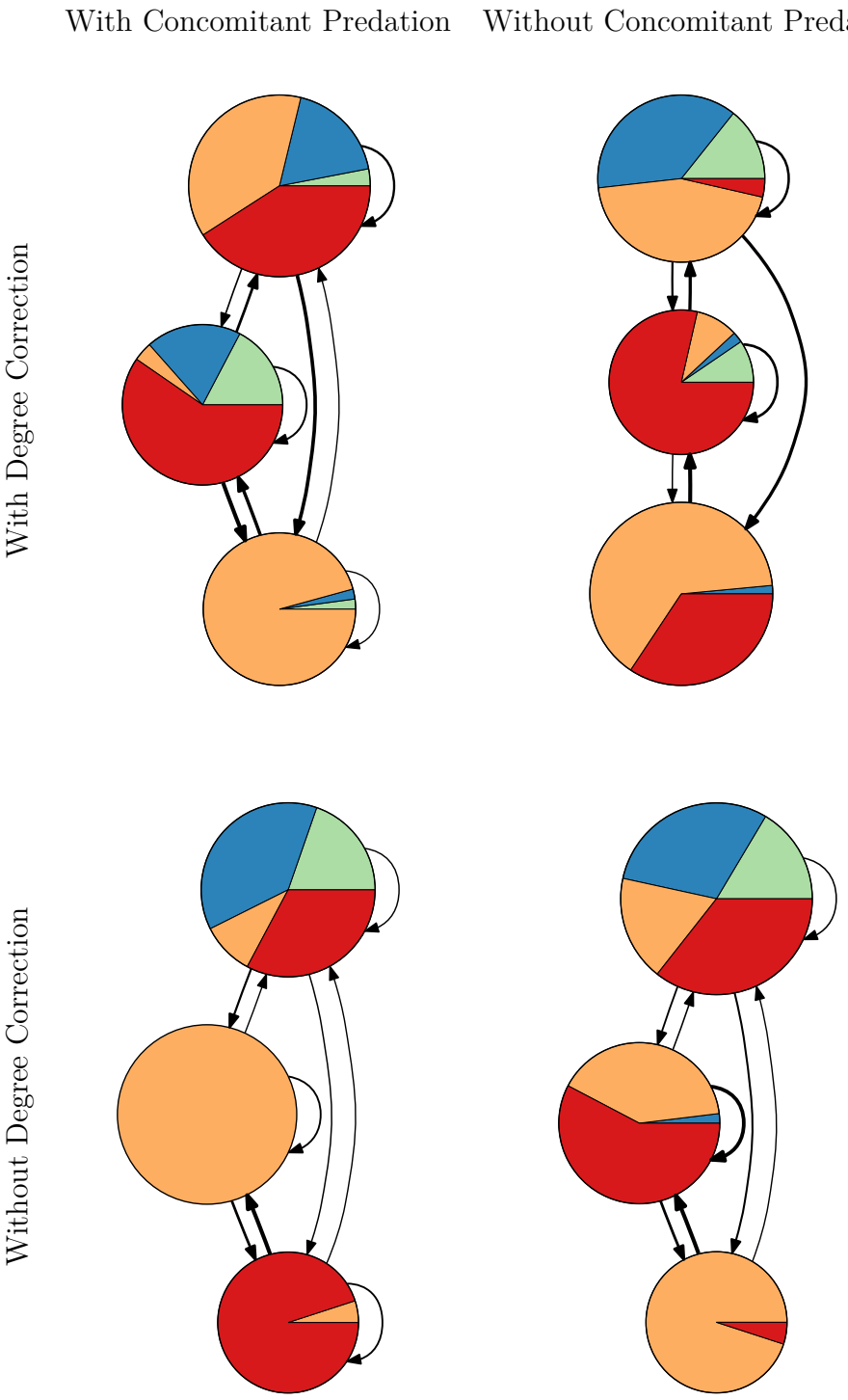


Figure 3.30: As fig. 3.3 in main text, but for the Carpinteria network and $g = 3$.

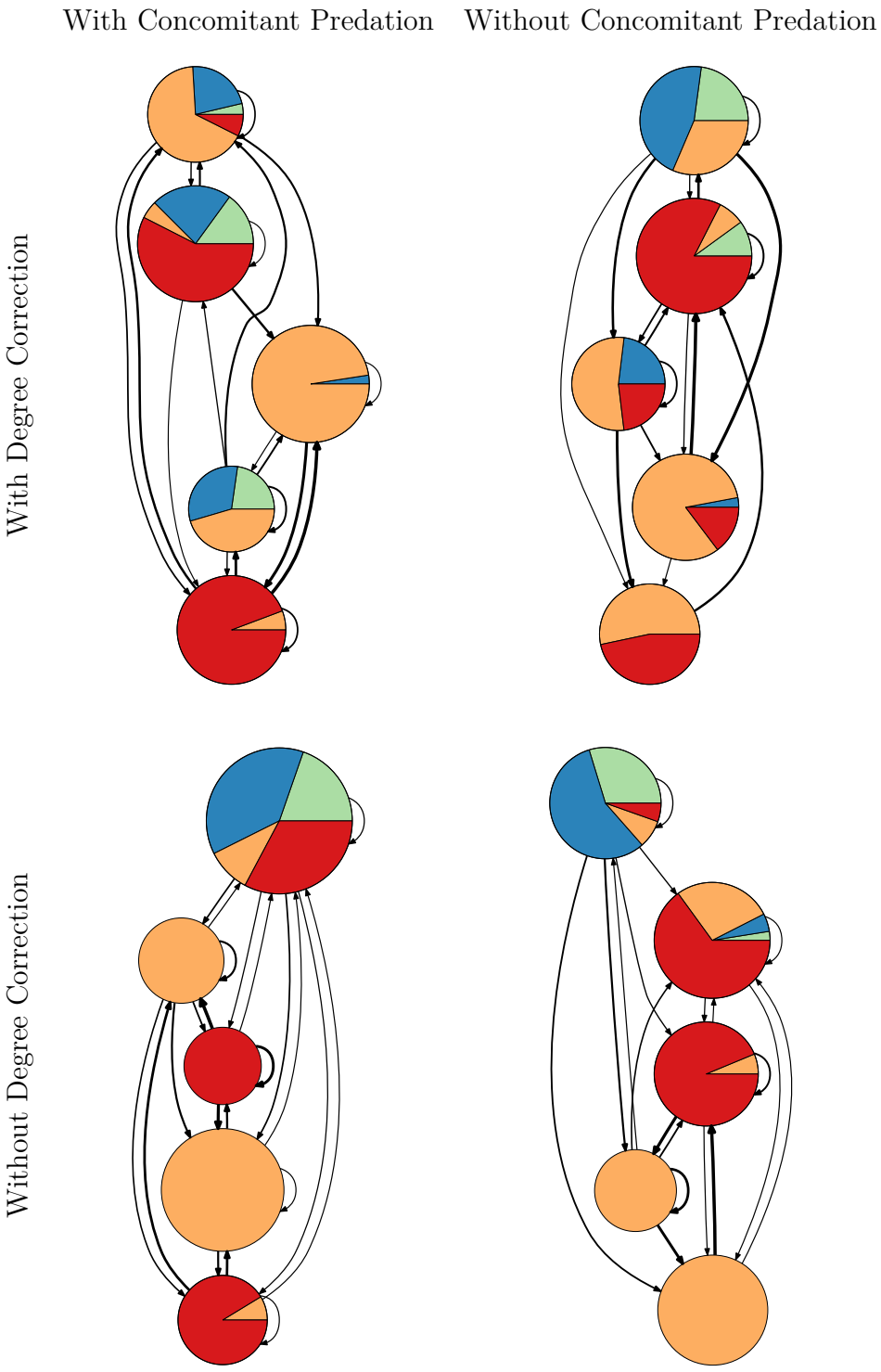


Figure 3.31: As fig. 3.3 in main text, but for the Carpinteria network and $g = 5$.

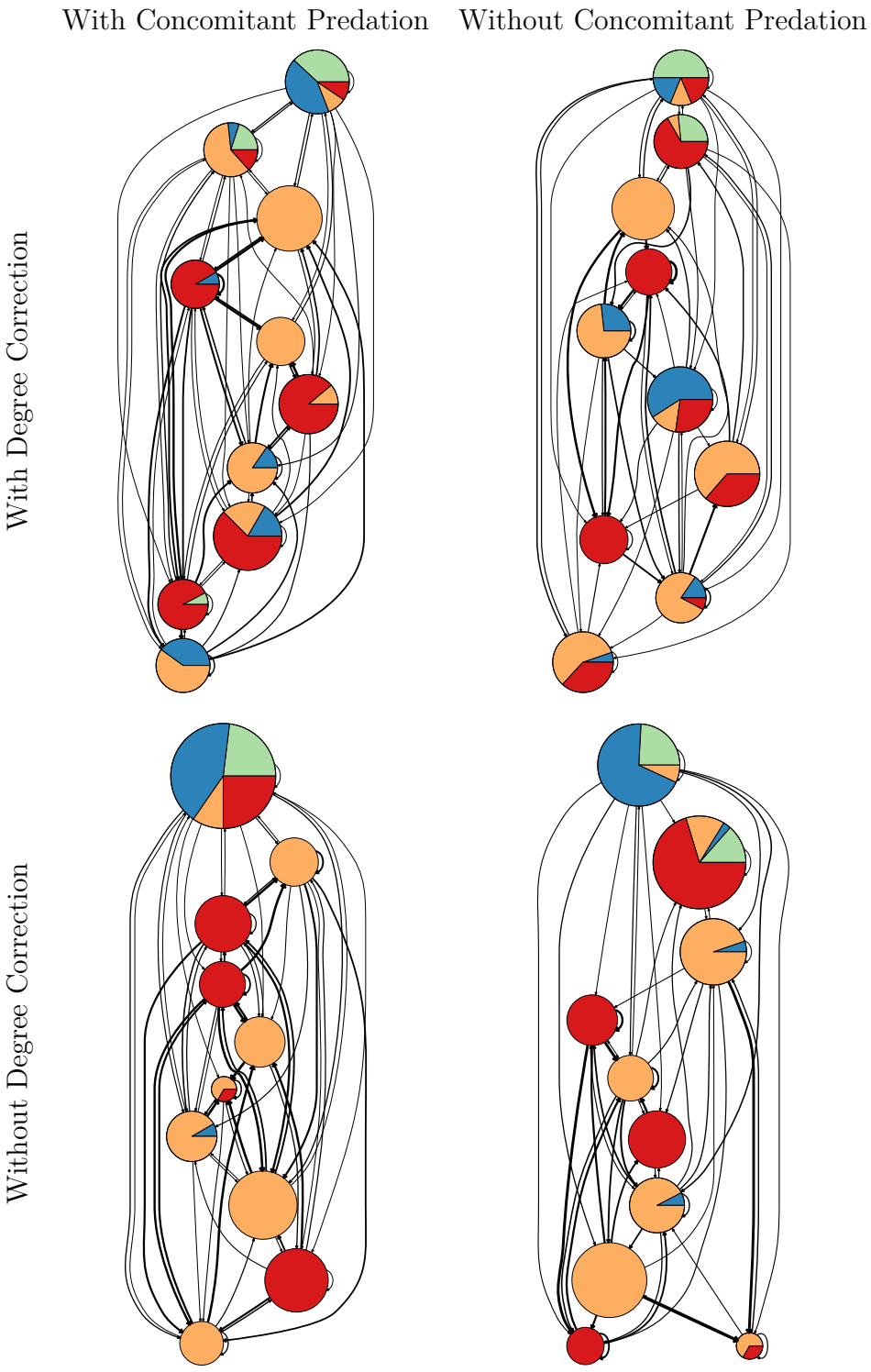


Figure 3.32: As fig. 3.3 in main text, but for the Carpinteria network and $g = 10$.

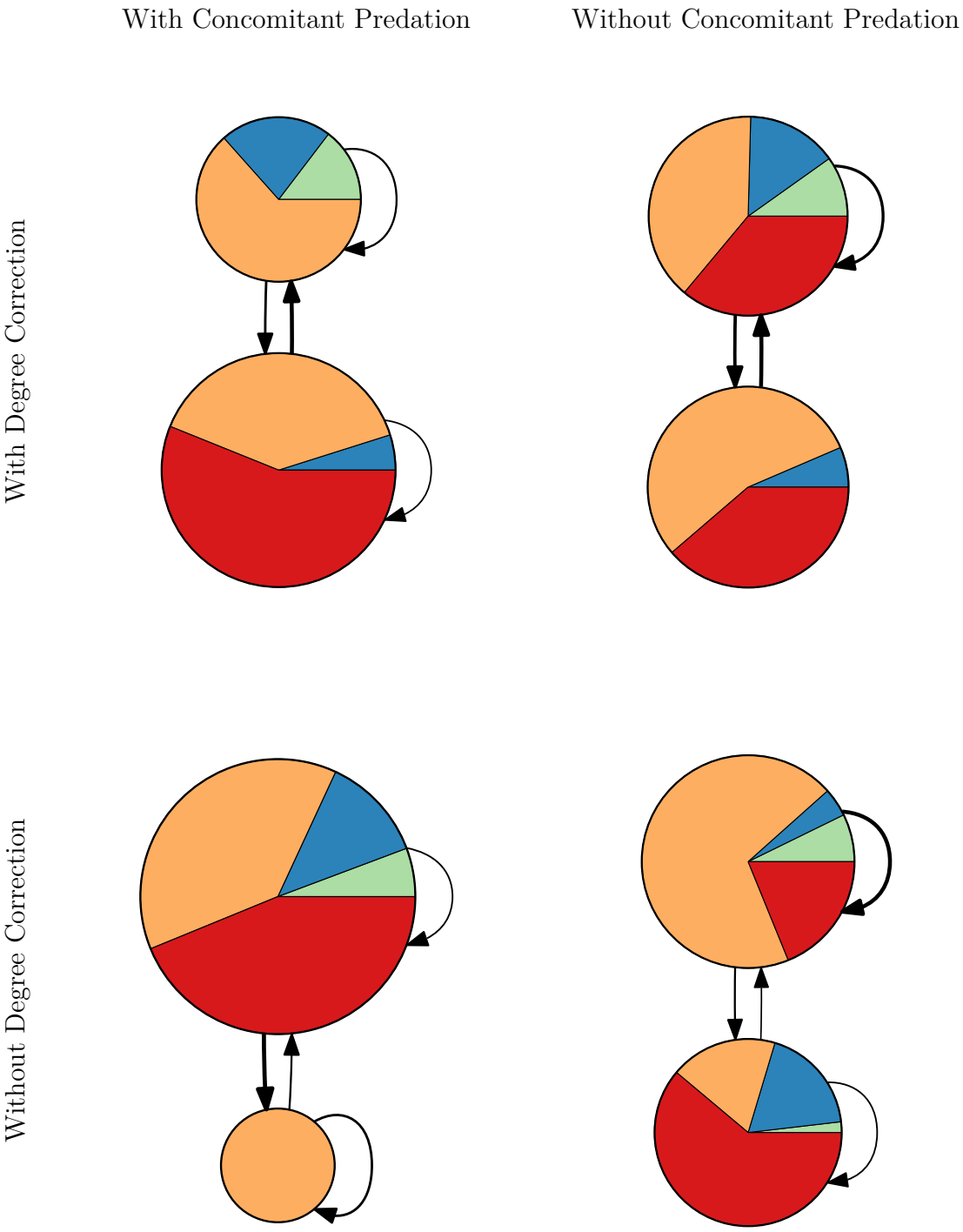


Figure 3.33: As fig. 3.3 in main text, but for the Flensburg network and $g = 2$.

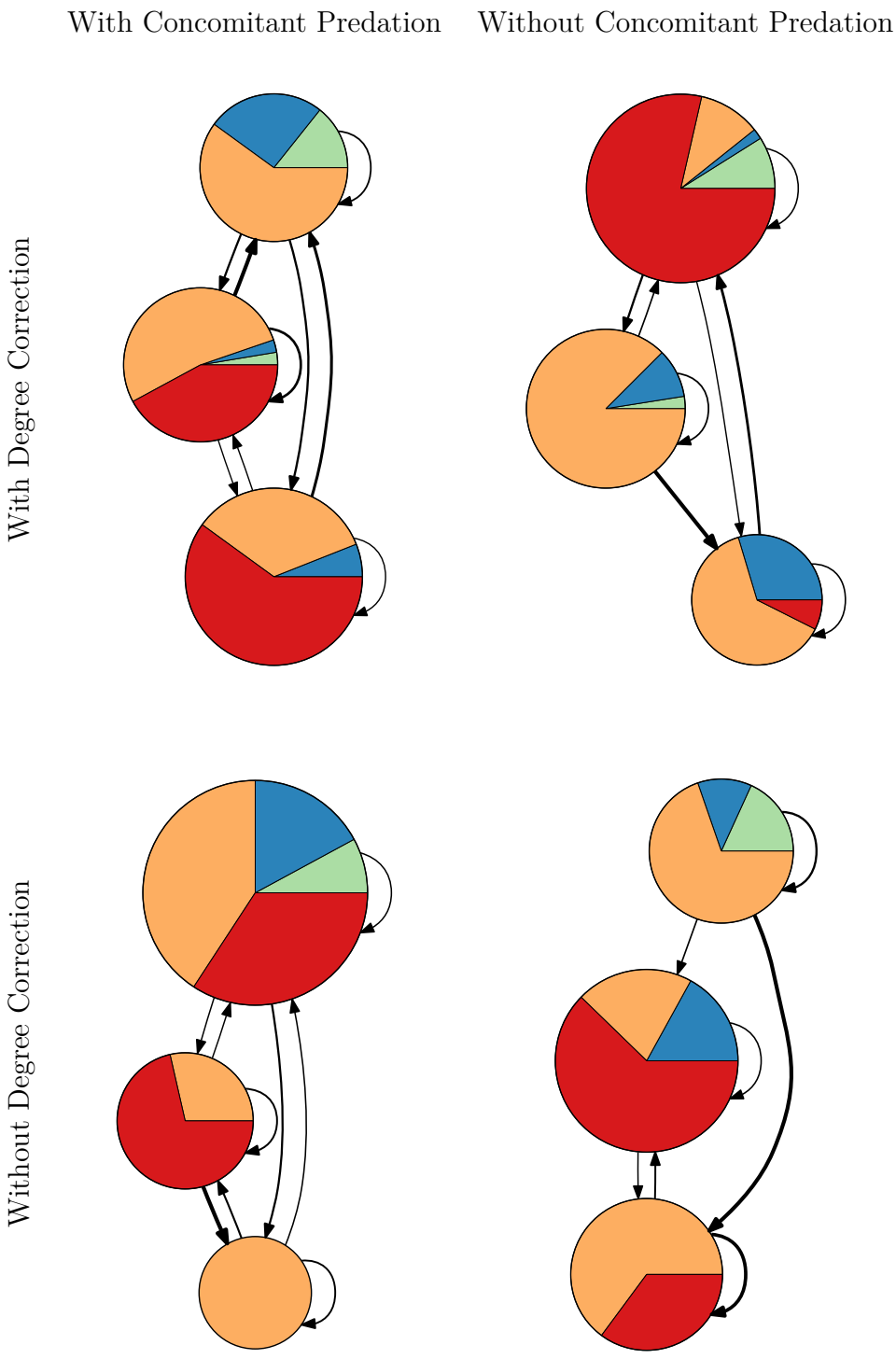


Figure 3.34: As fig. 3.3 in main text, but for the Flensburg network and $g = 3$.

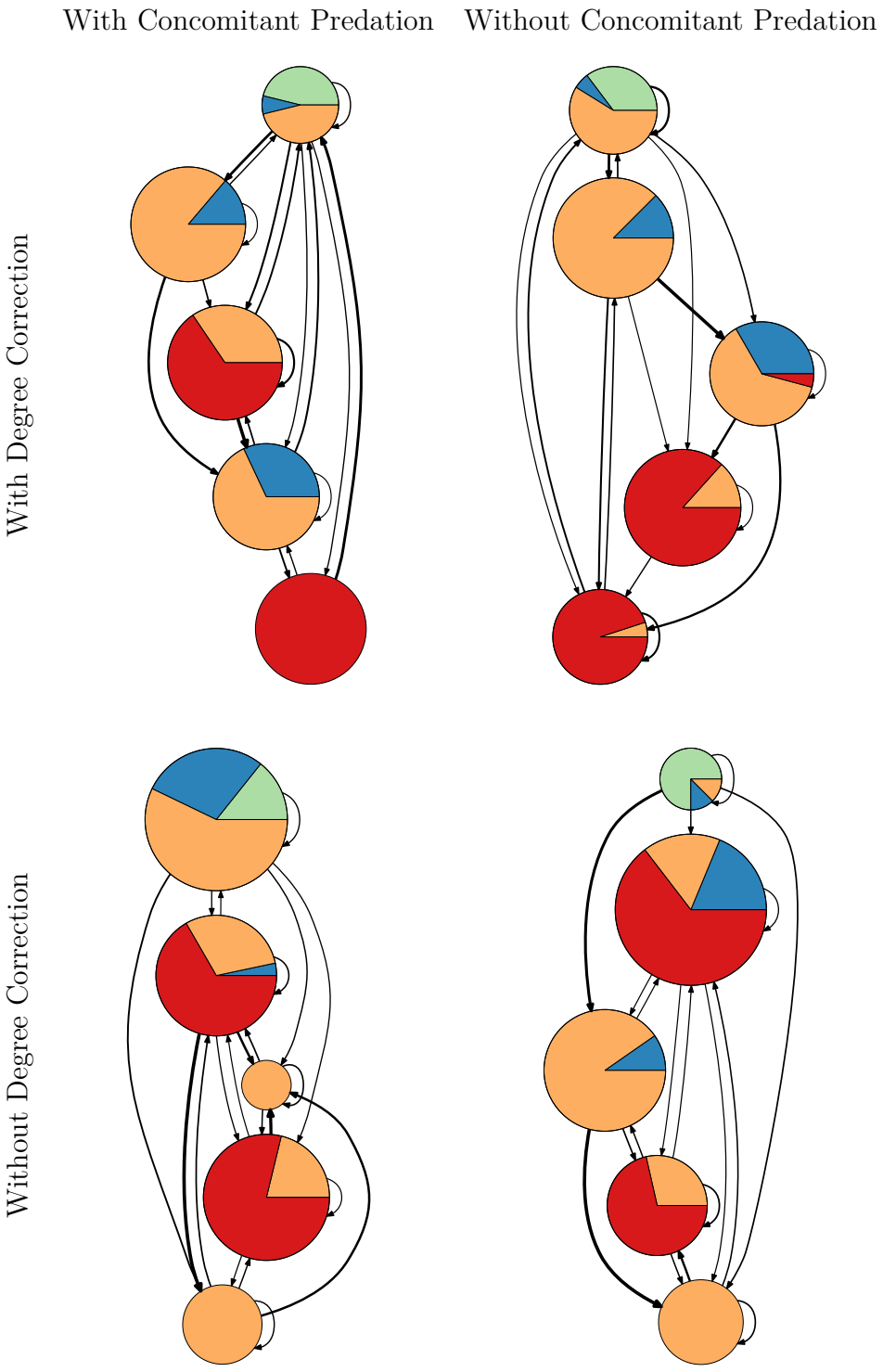


Figure 3.35: As fig. 3.3 in main text, but for the Flensburg network and $g = 5$.

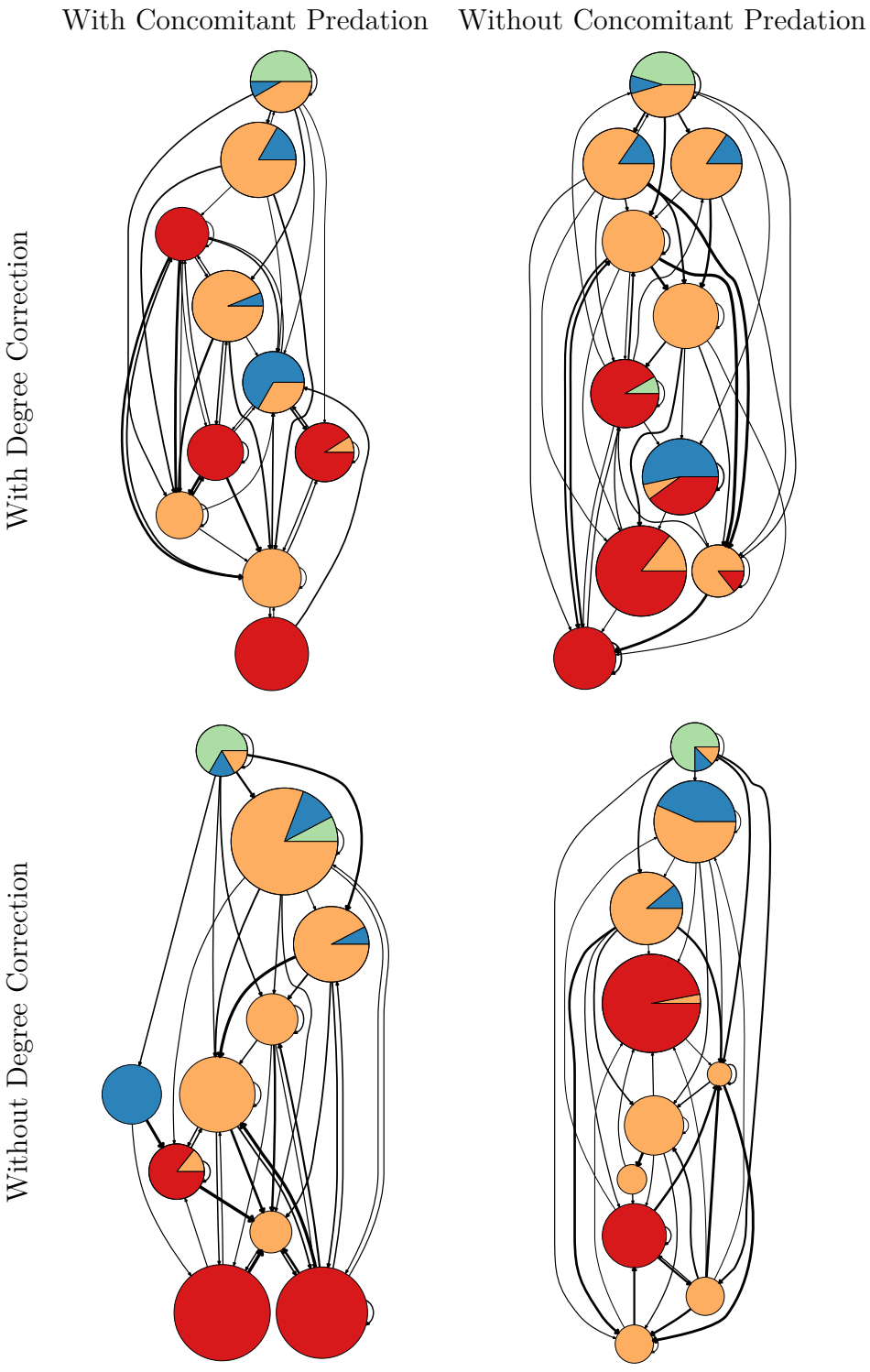


Figure 3.36: As fig. 3.3 in main text, but for the Flensburg network and $g = 10$.

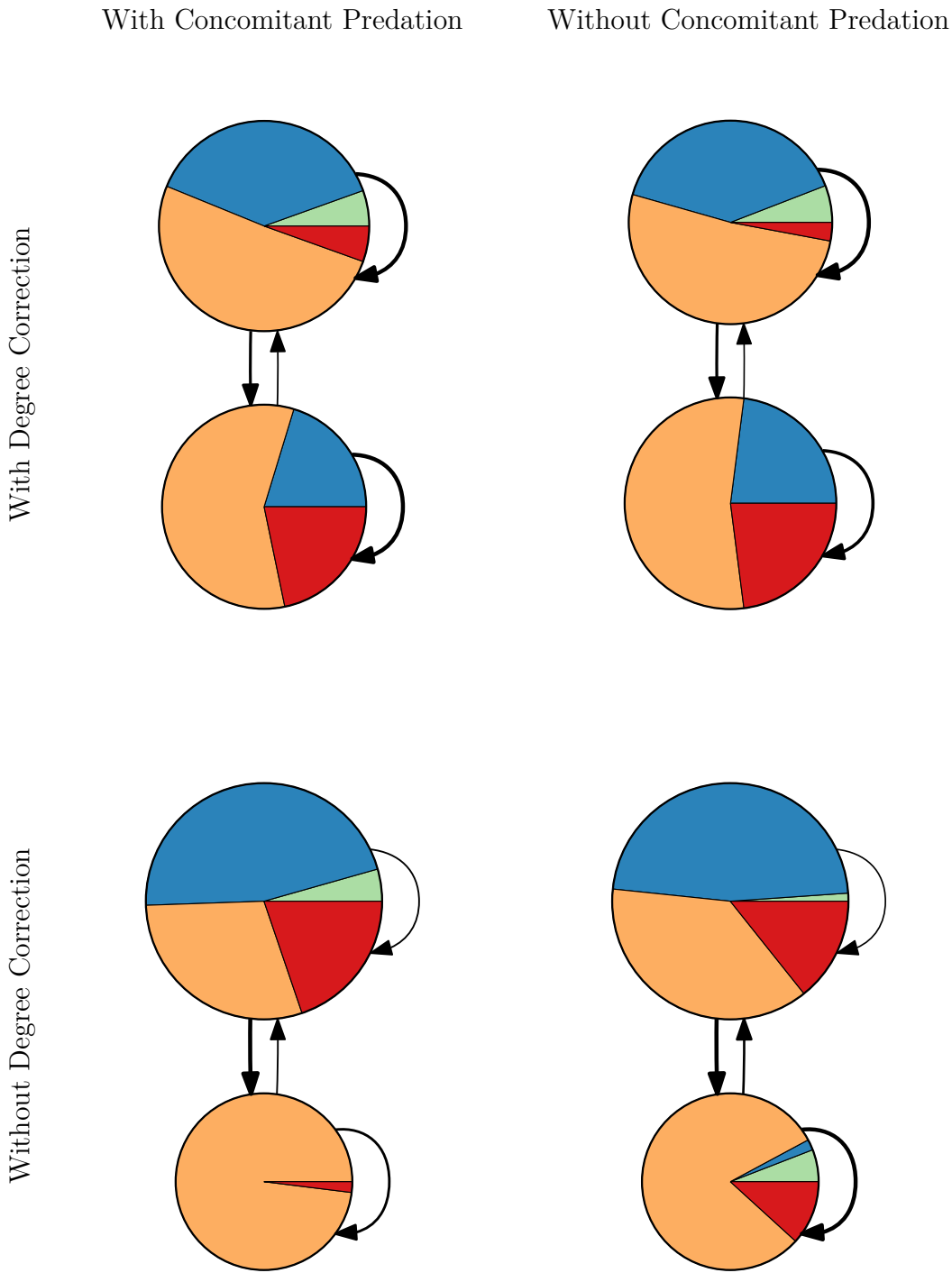


Figure 3.37: As fig. 3.3 in main text, but for the Otago network and $g = 2$.

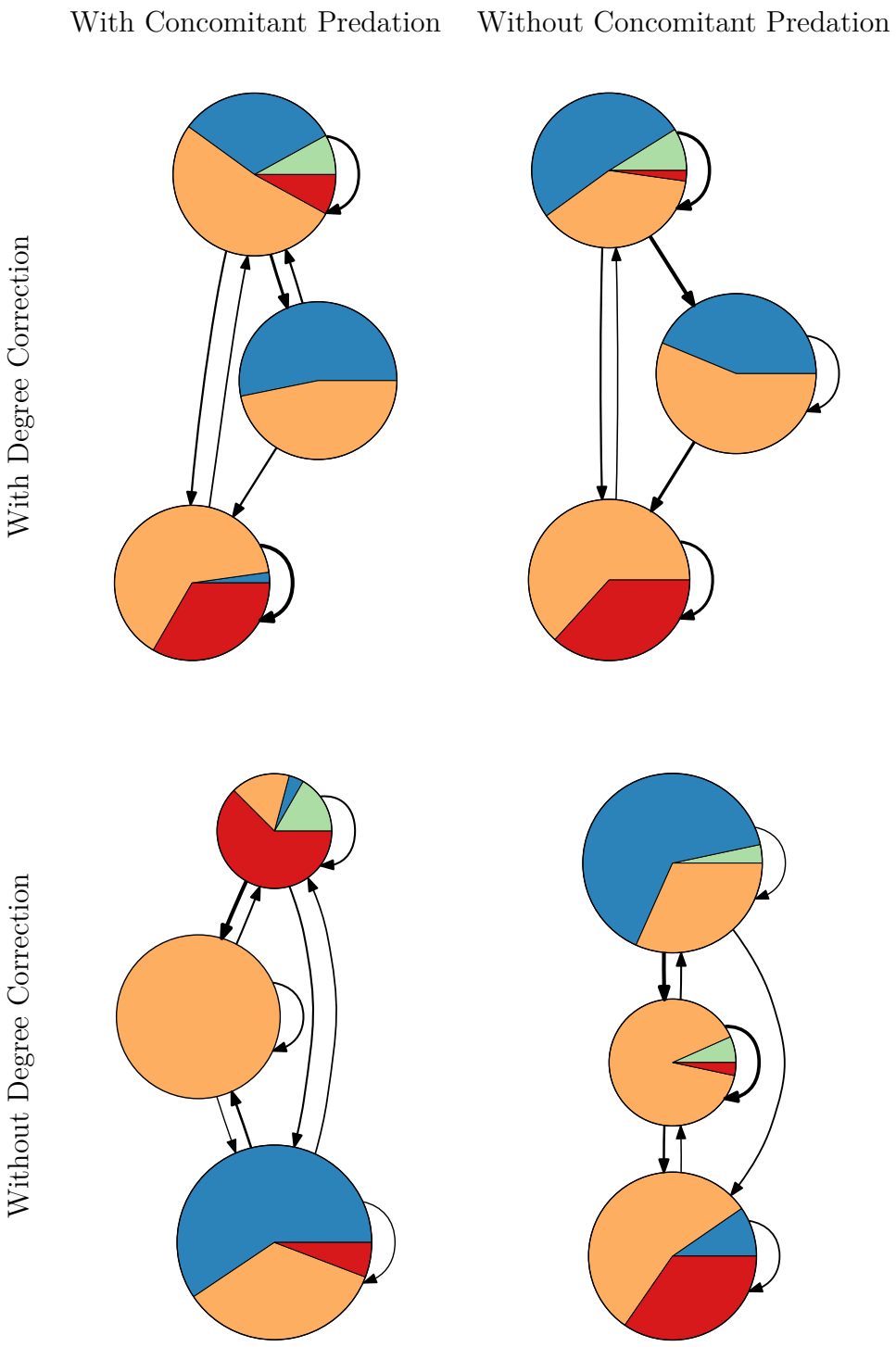


Figure 3.38: As fig. 3.3 in main text, but for the Otago network and $g = 3$.

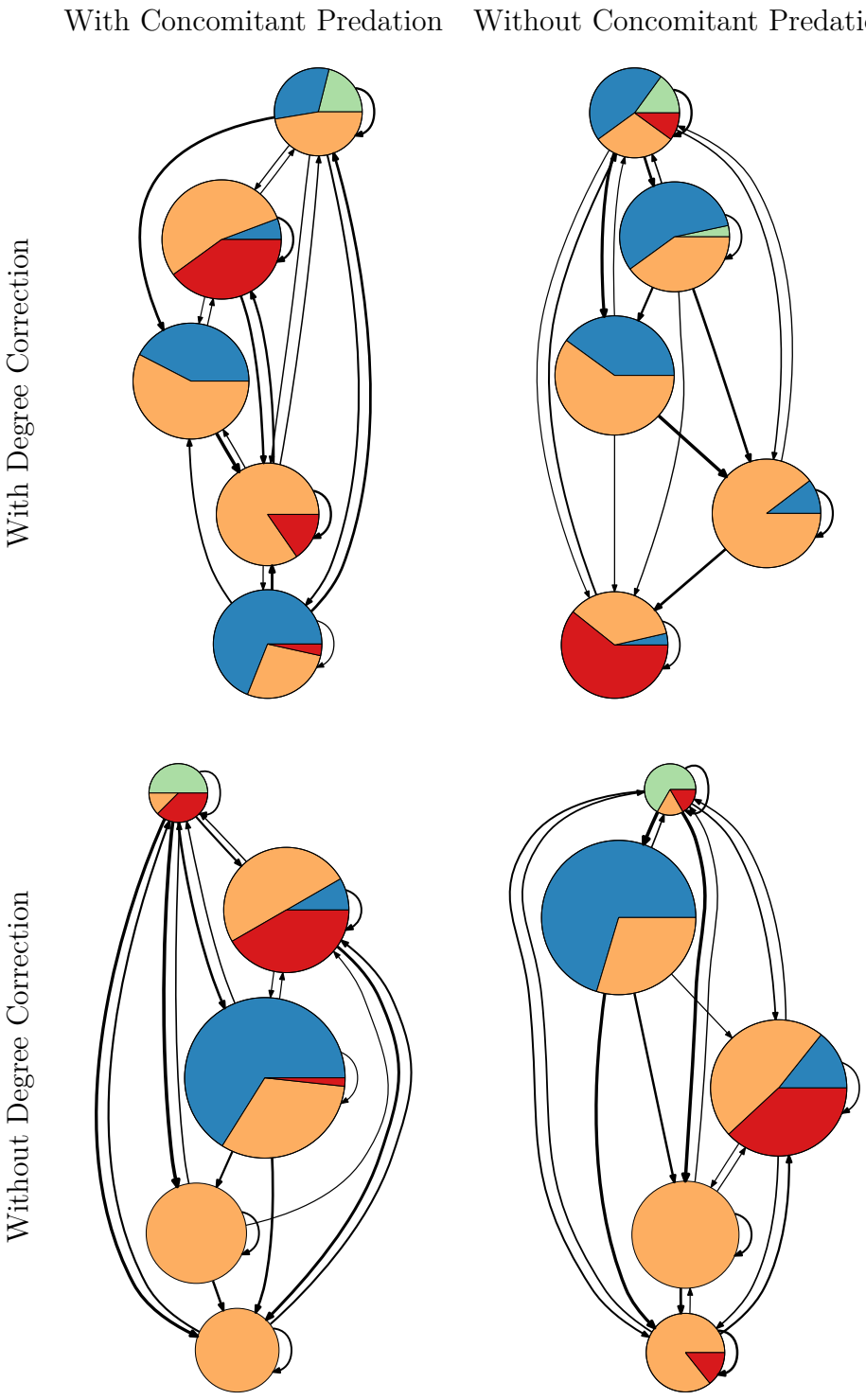


Figure 3.39: As fig. 3.3 in main text, but for the Otago network and $g = 5$.

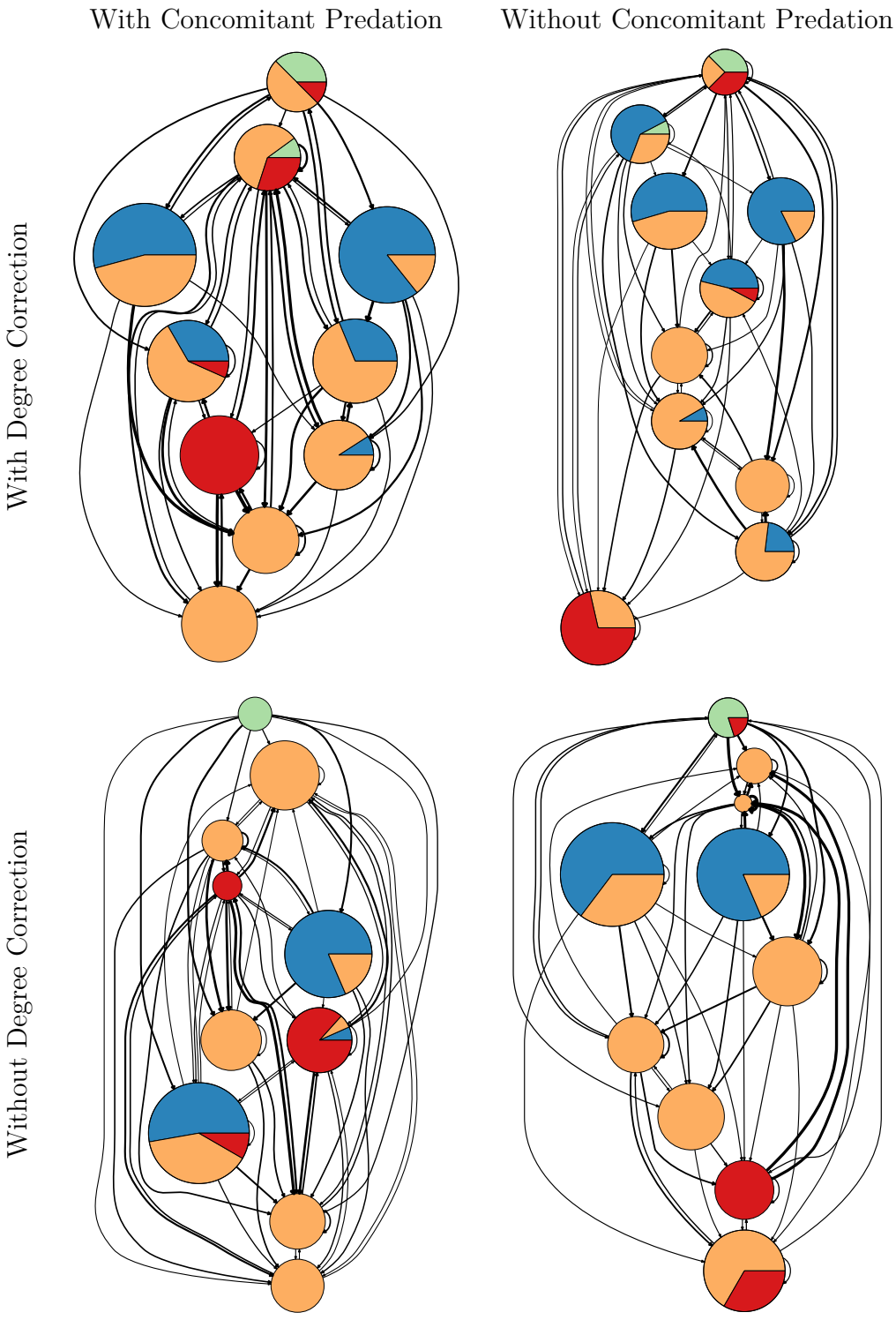


Figure 3.40: As fig. 3.3 in main text, but for the Otago network and $g = 10$.

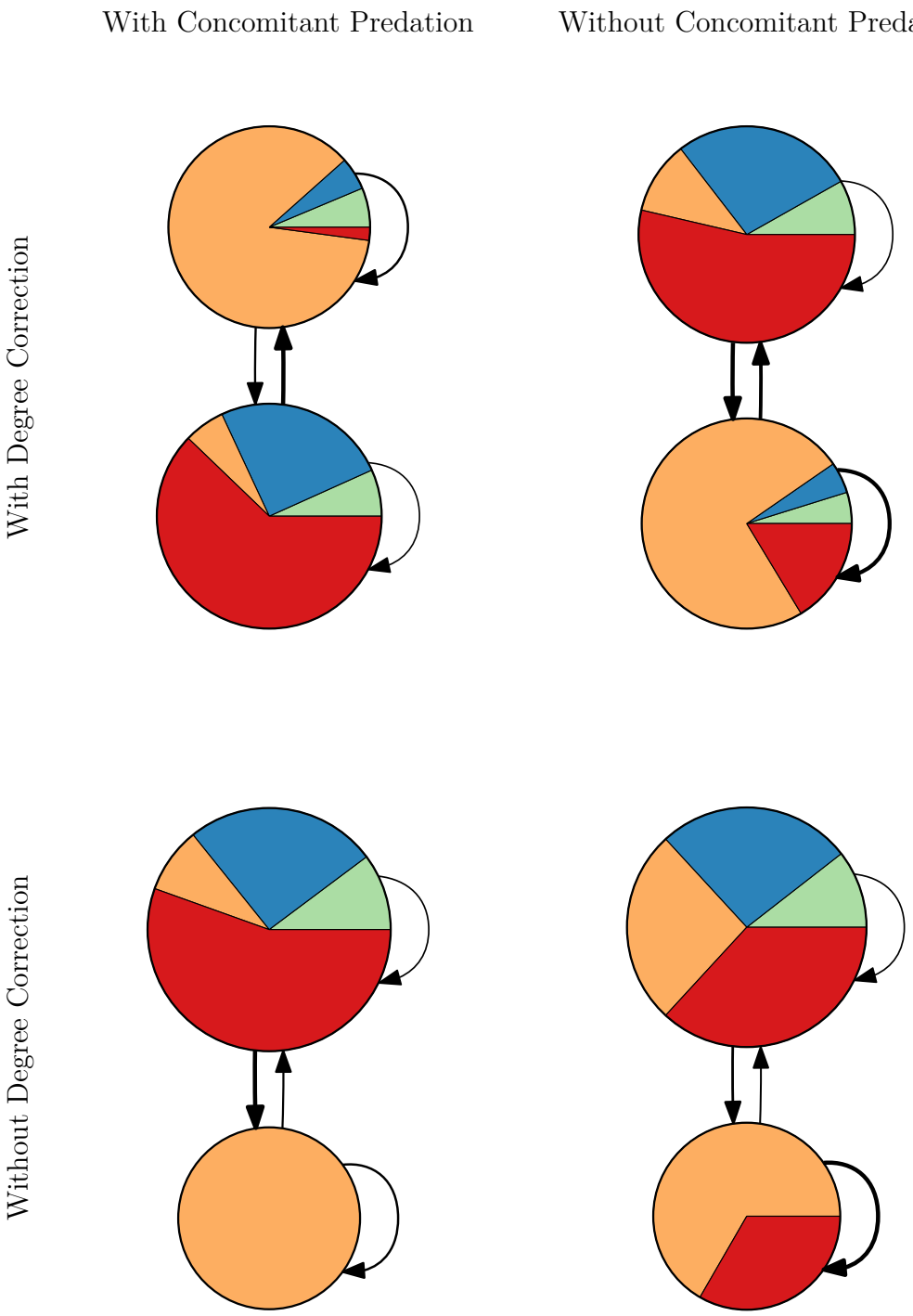


Figure 3.41: As fig. 3.3 in main text, but for the Punta Banda network and $g = 2$.

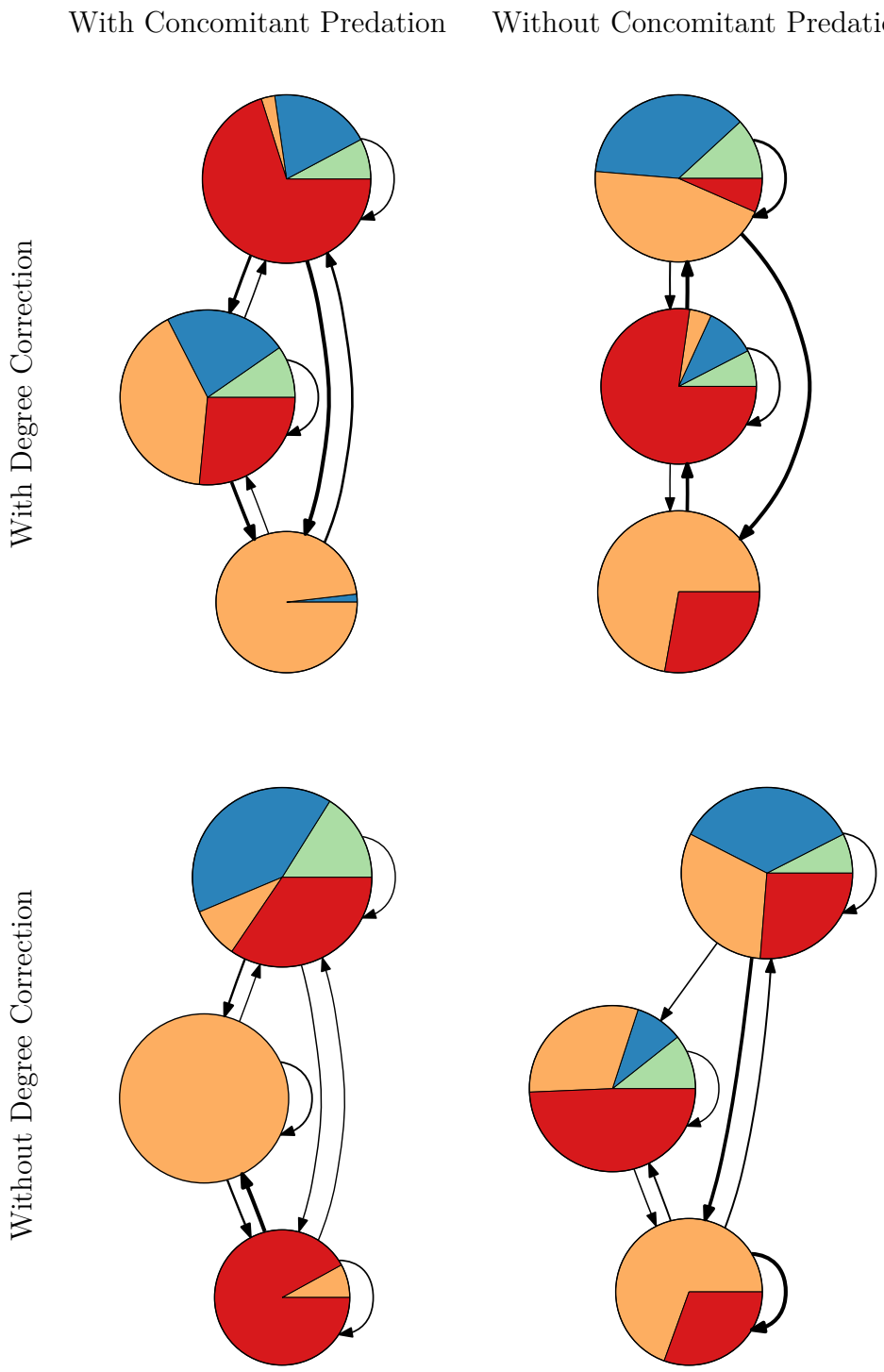


Figure 3.42: As fig. 3.3 in main text, but for the Punta Banda network and $g = 3$.

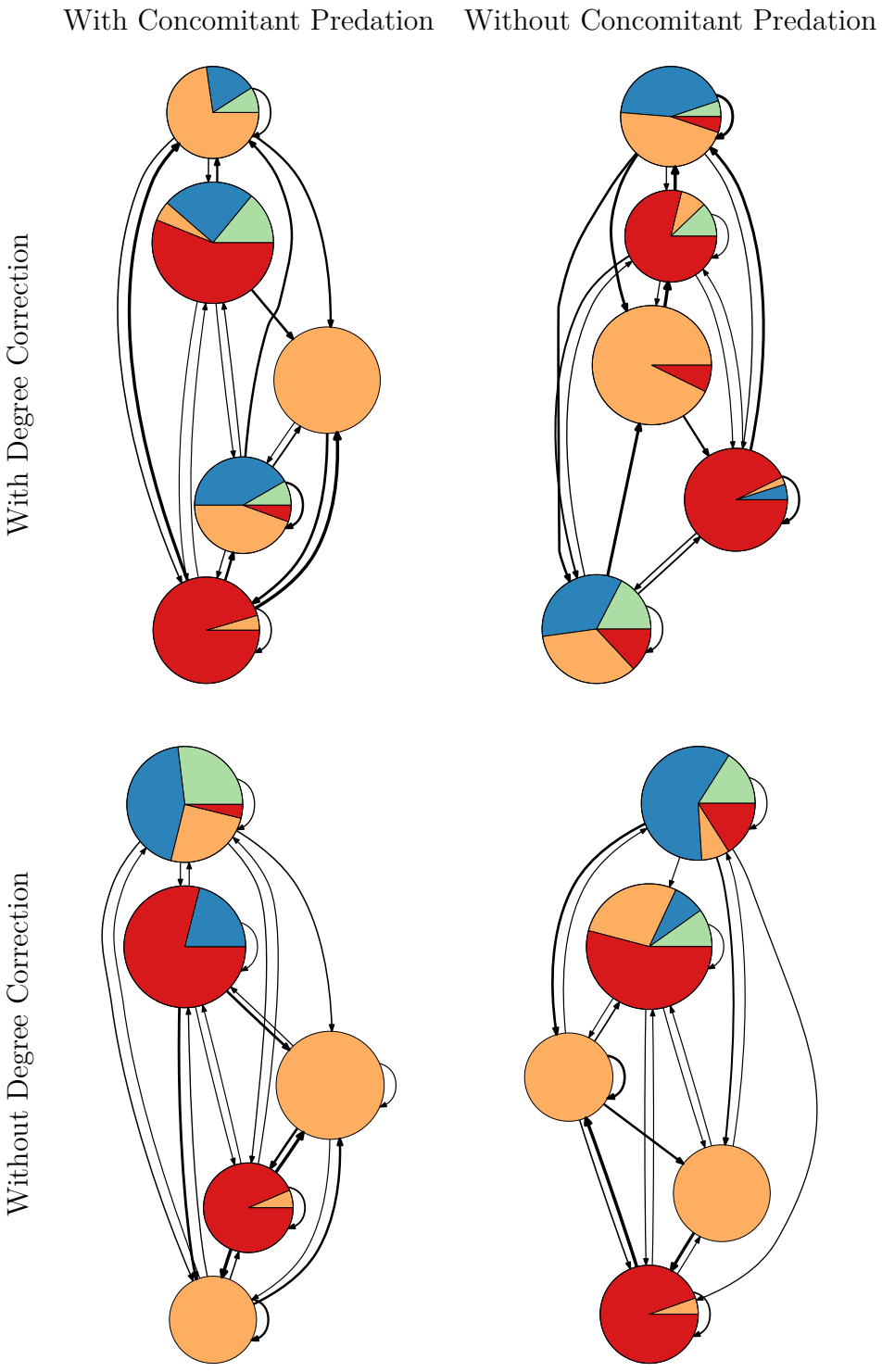


Figure 3.43: As fig. 3.3 in main text, but for the Punta Banda network and $g = 5$.

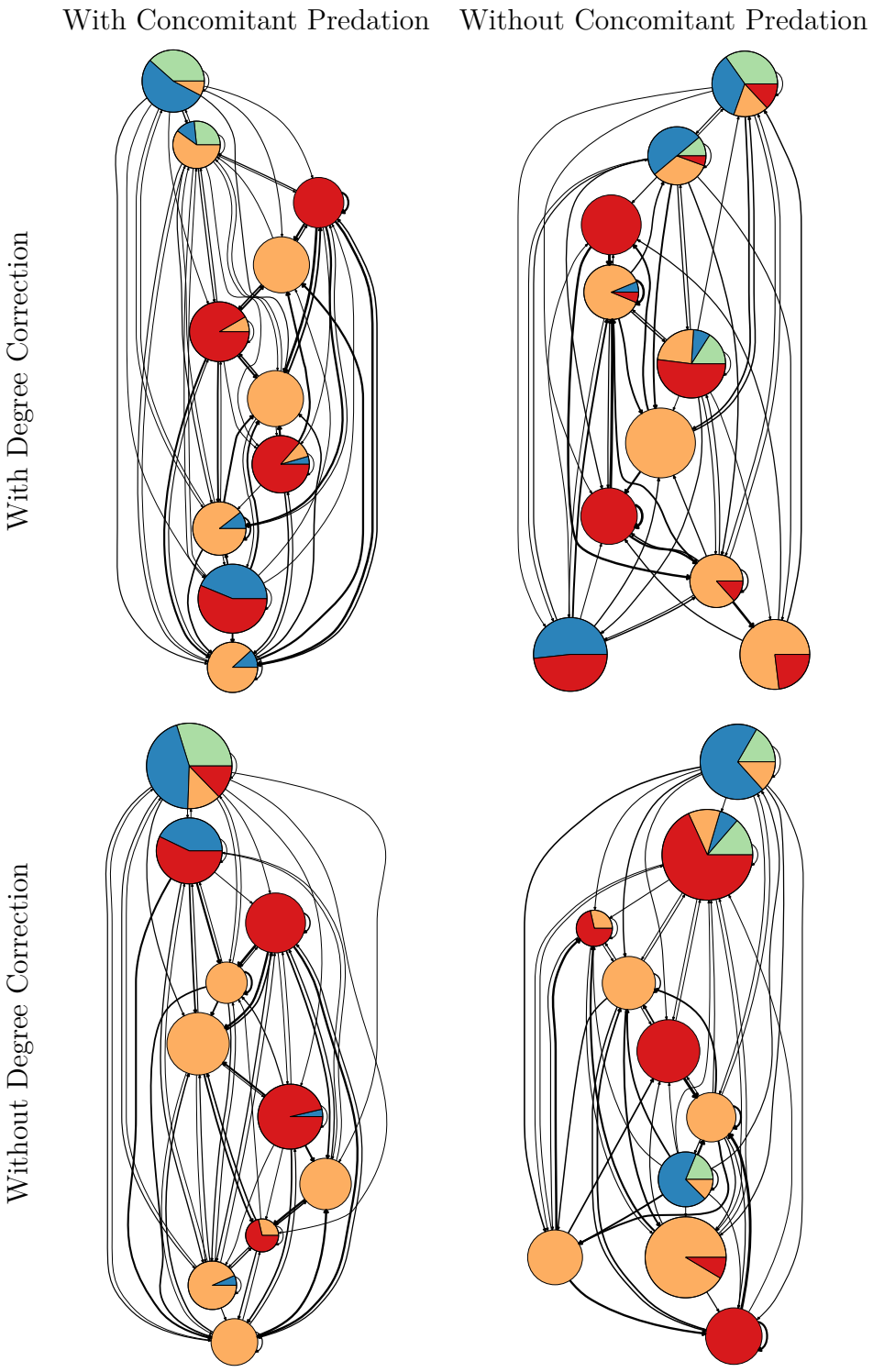


Figure 3.44: As fig. 3.3 in main text, but for the Punta Banda network and $g = 10$.

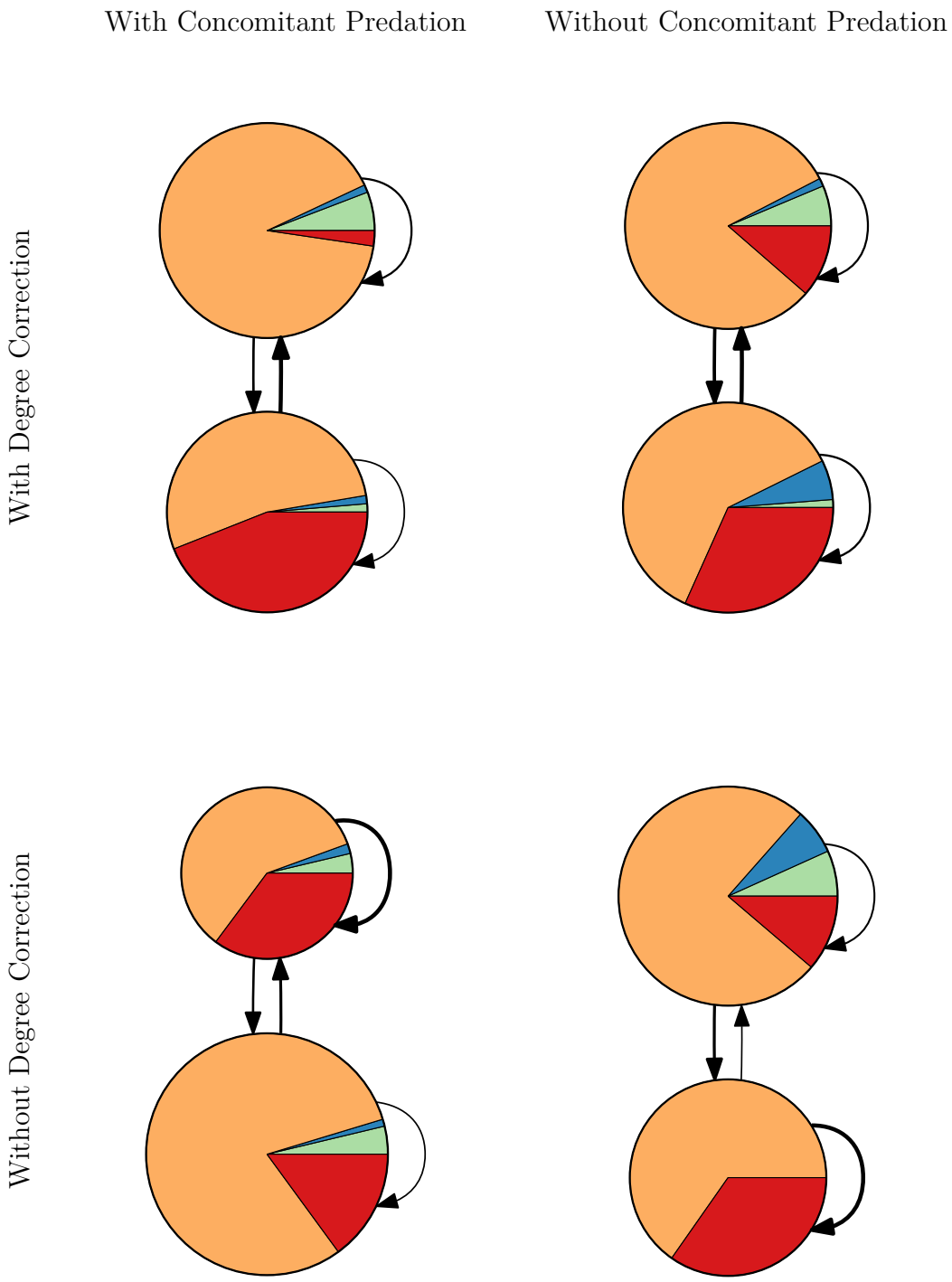


Figure 3.45: As fig. 3.3 in main text, but for the Sylt network and $g = 2$.

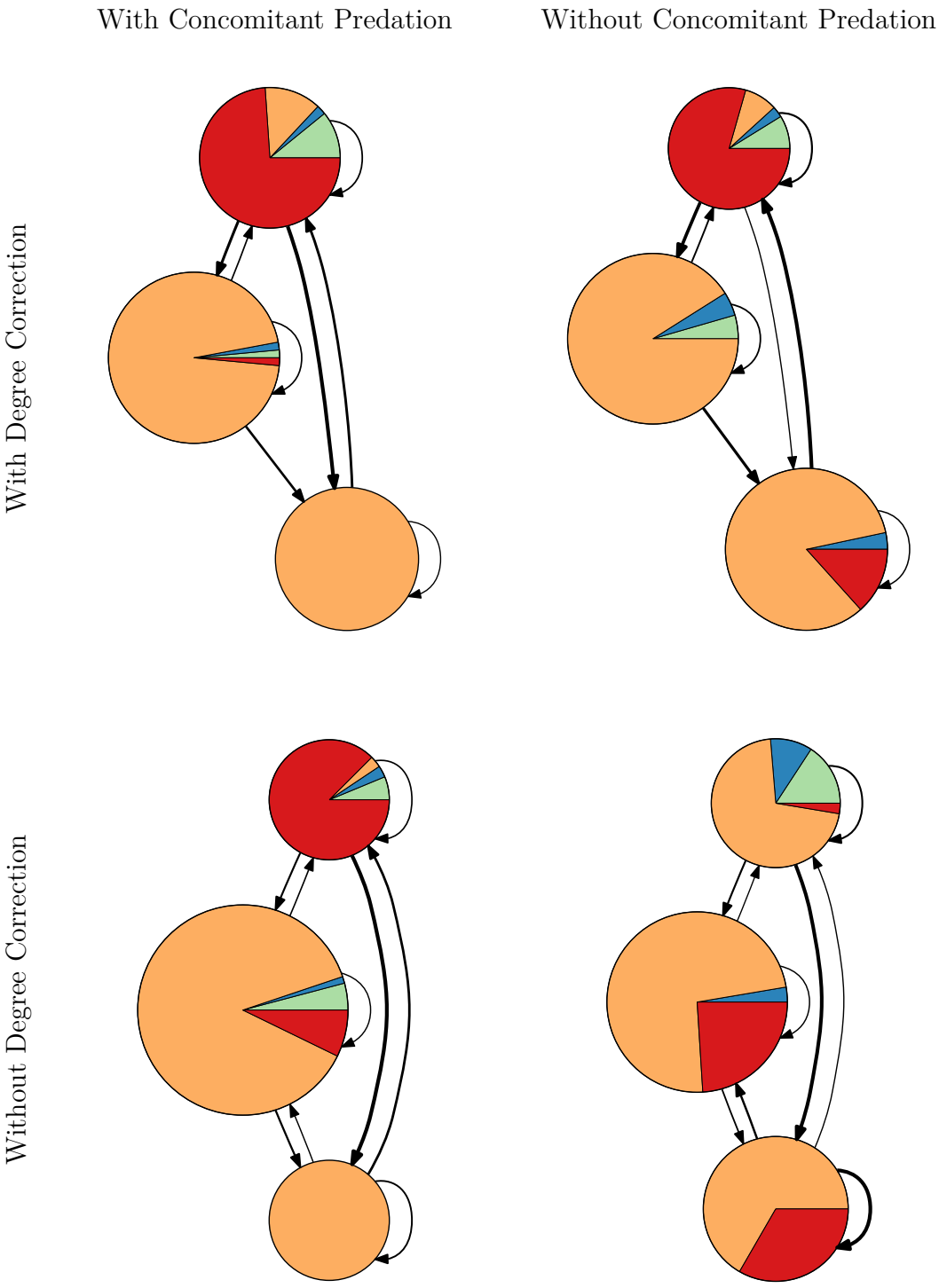


Figure 3.46: As fig. 3.3 in main text, but for the Sylt network and $g = 3$.

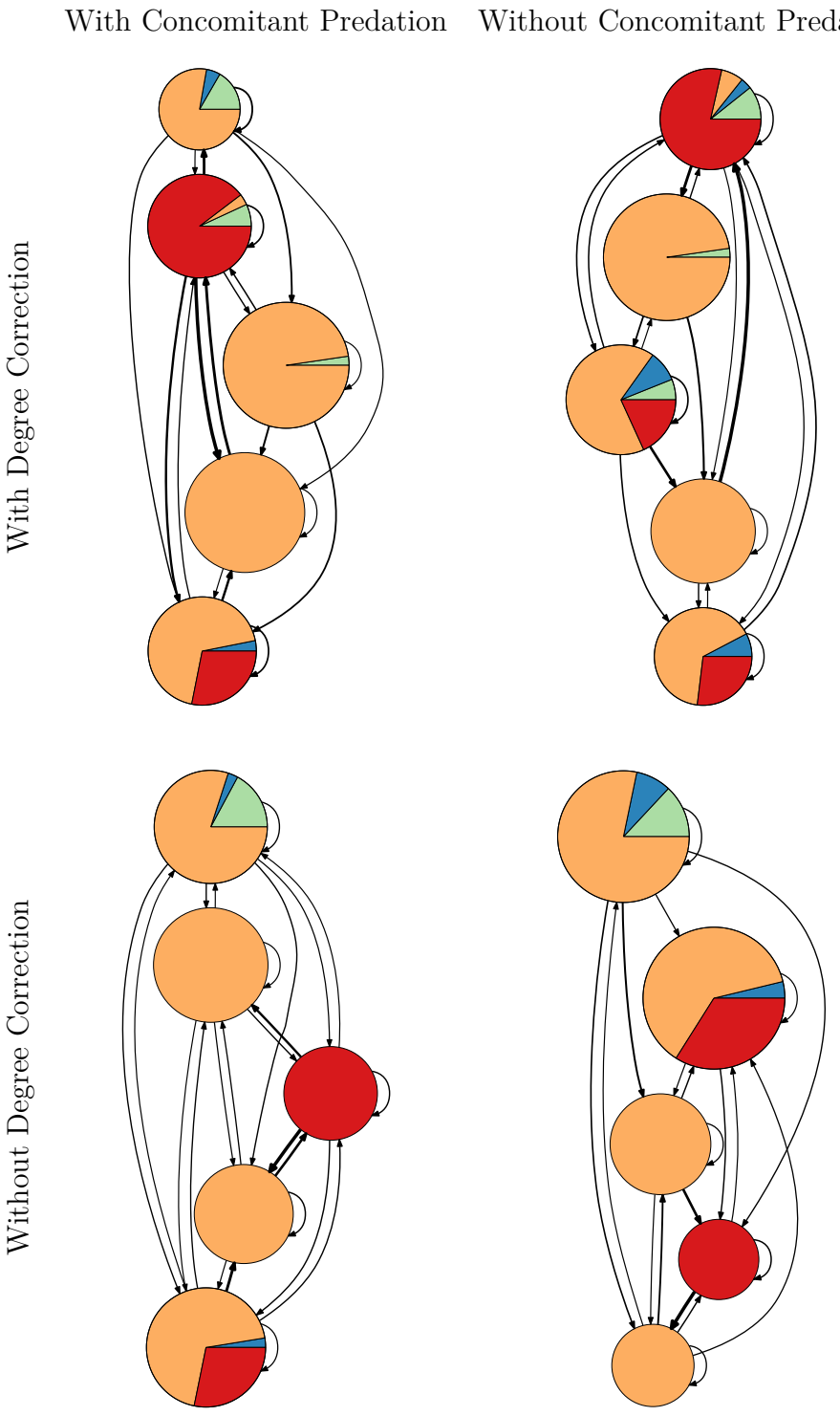


Figure 3.47: As fig. 3.3 in main text, but for the Sylt network and $g = 5$.

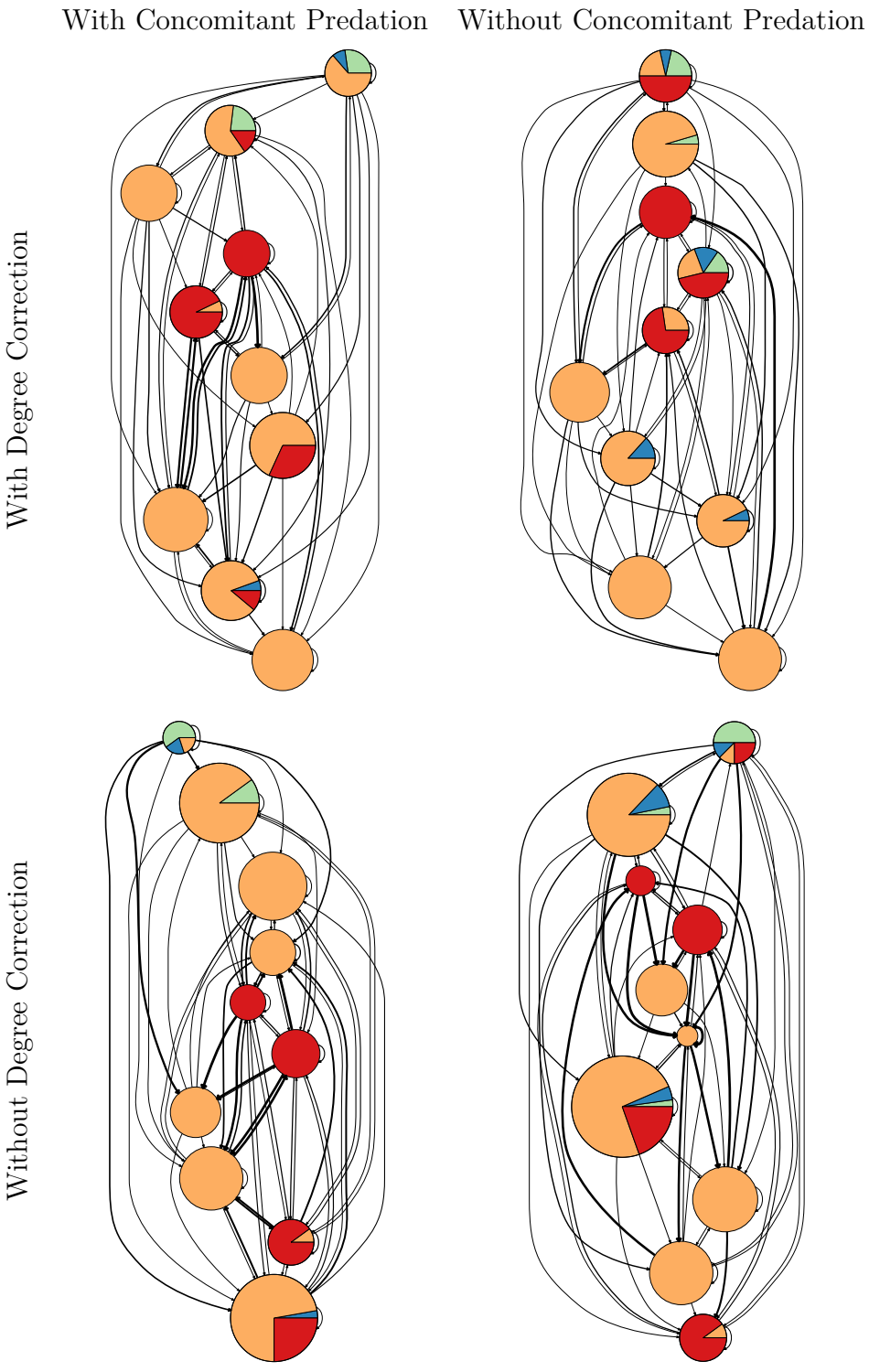


Figure 3.48: As fig. 3.3 in main text, but for the Sylt network and $g = 10$.

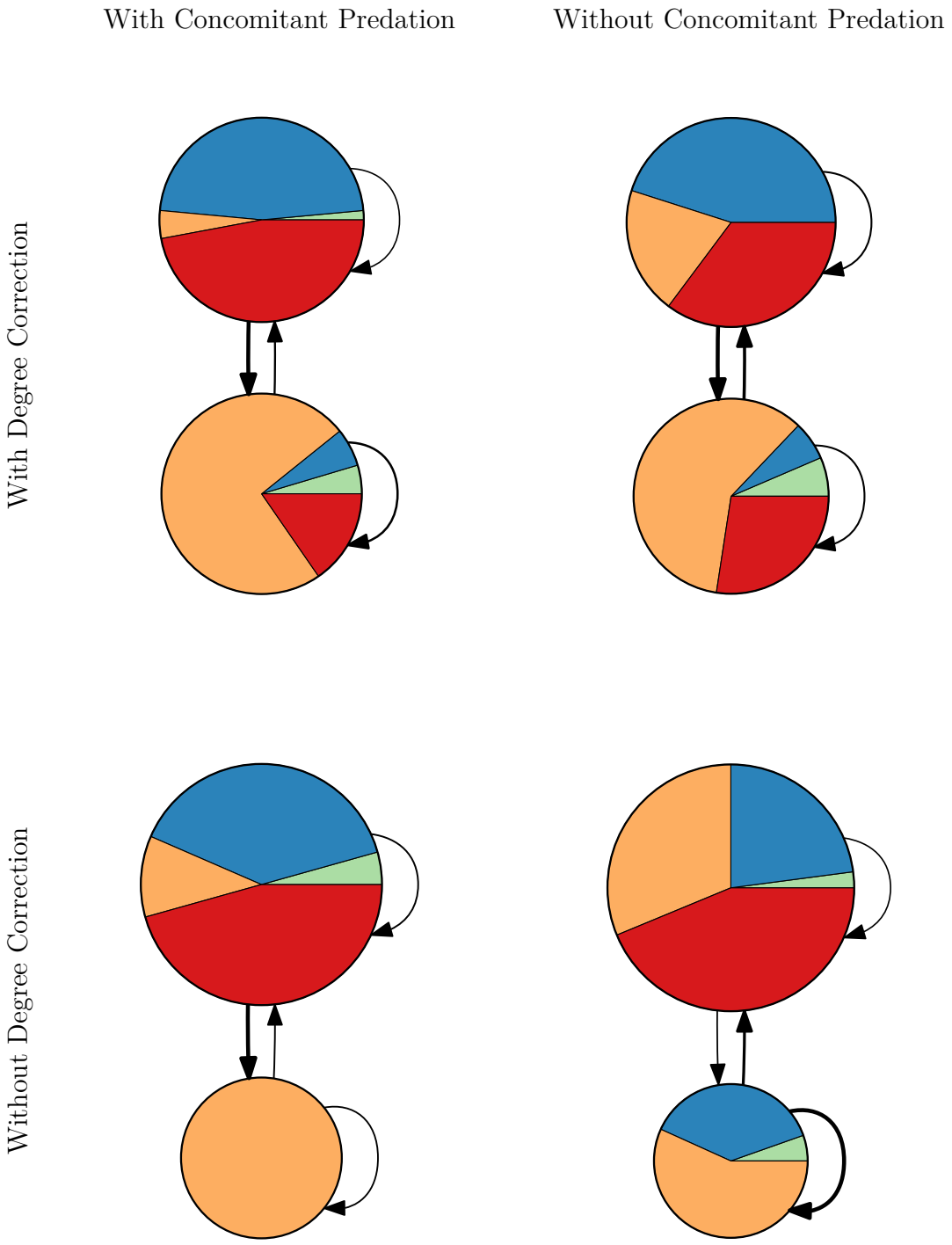


Figure 3.49: As fig. 3.3 in main text, but for the Ythan network and $g = 2$.

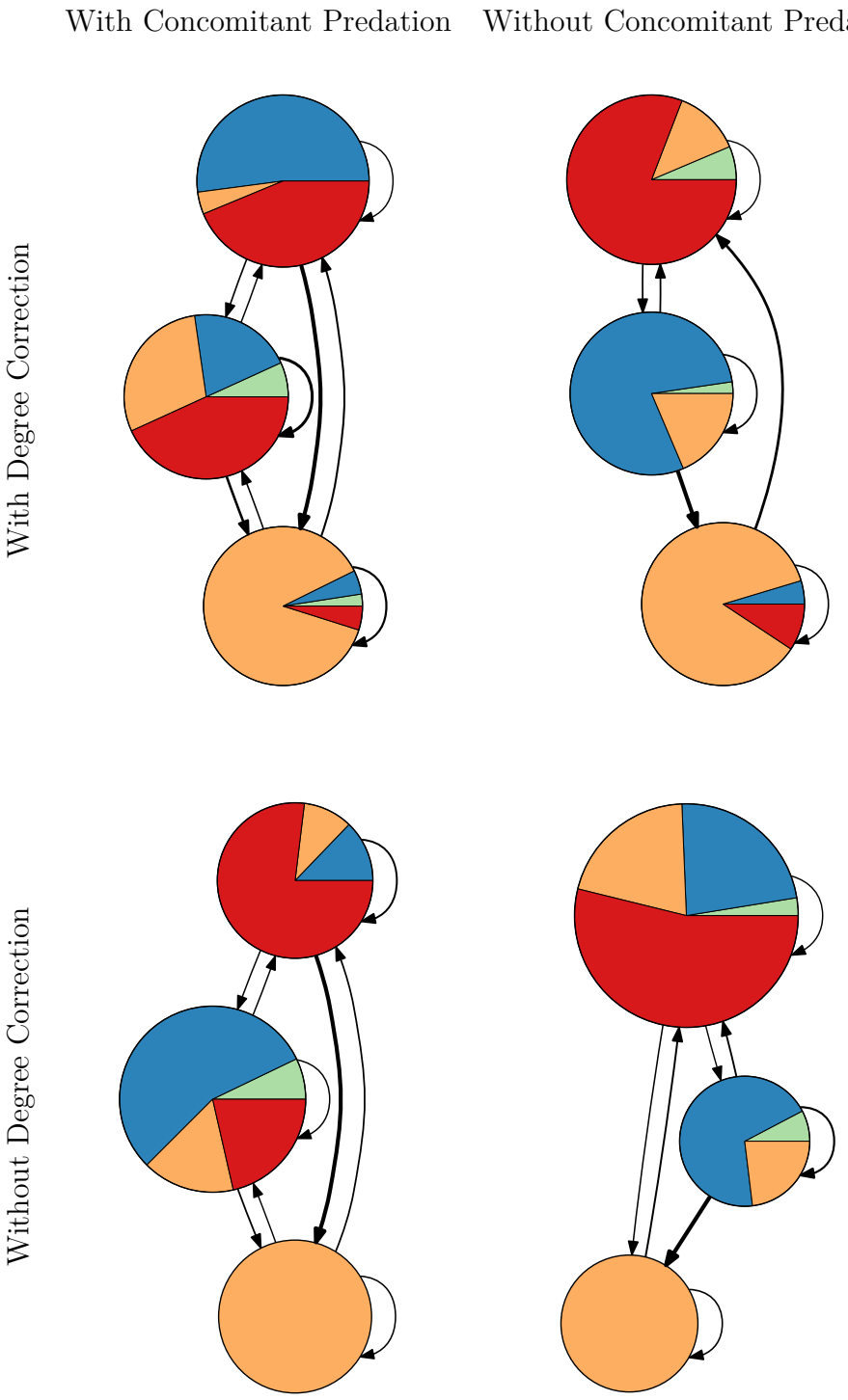


Figure 3.50: As fig. 3.3 in main text, but for the Ythan network and $g = 3$.

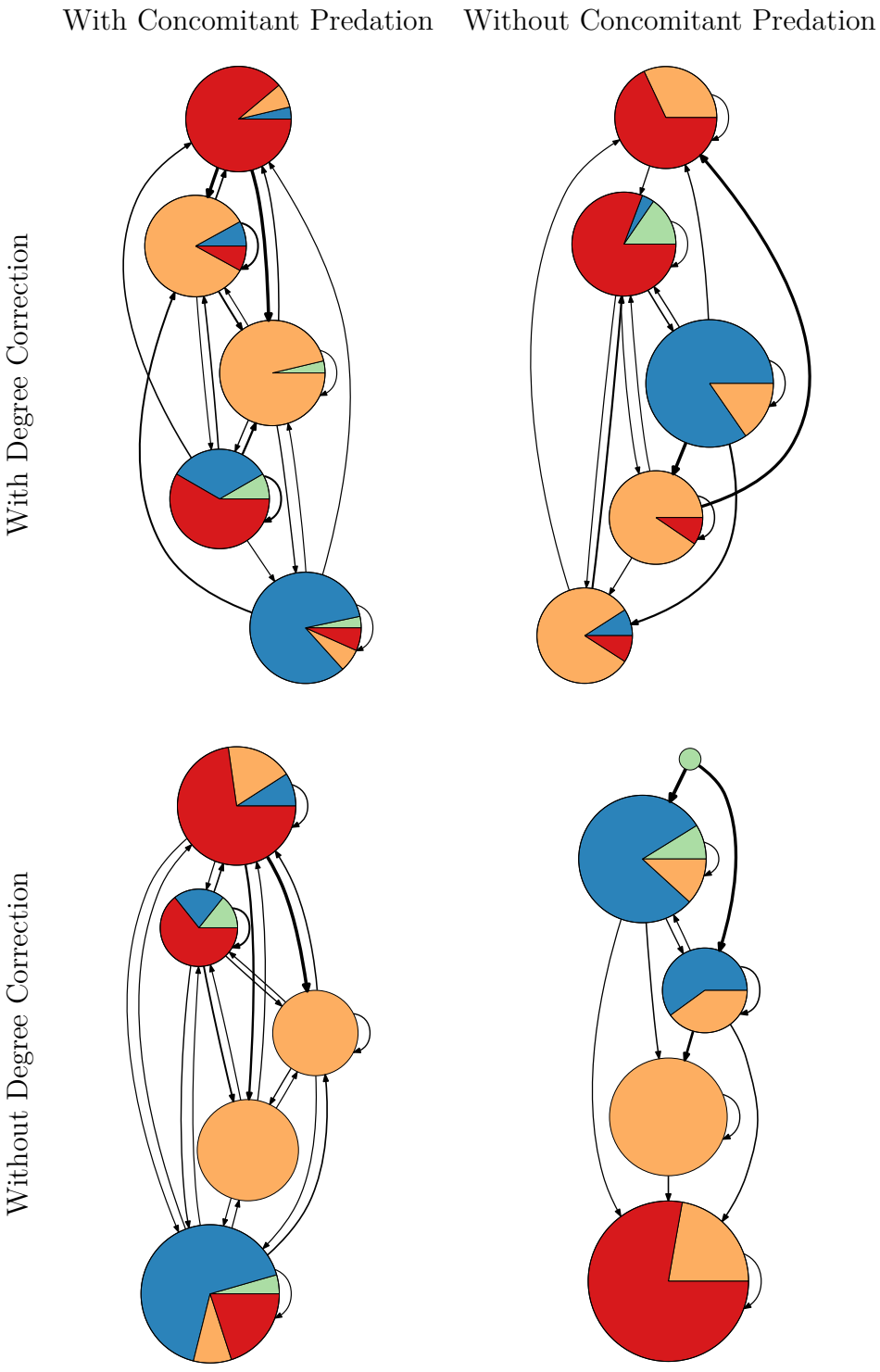


Figure 3.51: As fig. 3.3 in main text, but for the Ythan network and $g = 5$.

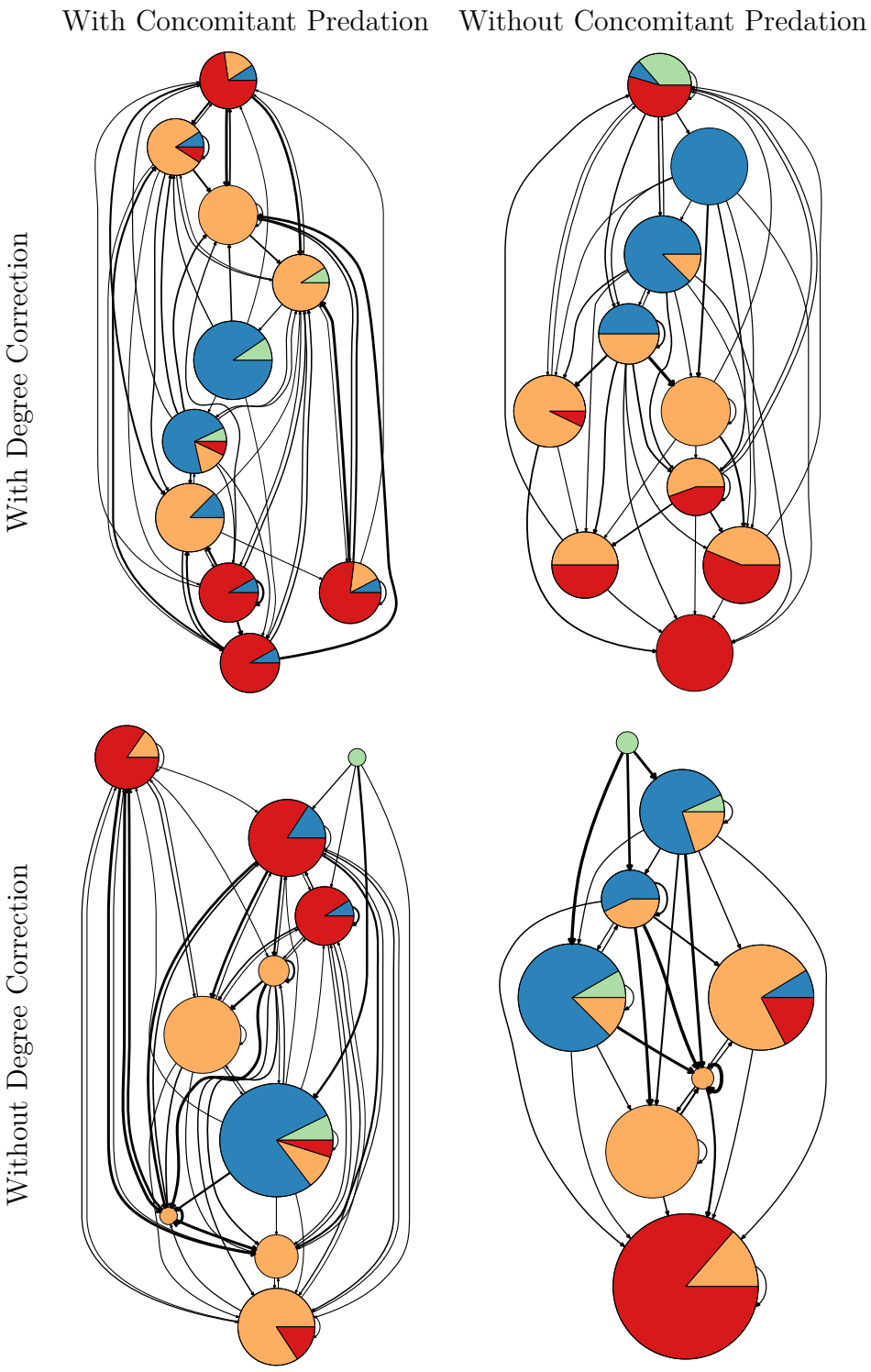


Figure 3.52: As fig. 3.3 in main text, but for the Ythan network and $g = 10$. Note that the group model without degree correction found that 8 groups outperforms 10 for the Ythan web without concomitant predation. This has the effect of making the nodes for this network disproportionately large and thus uncomparable to the condensed graphs with 10 nodes.

CHAPTER 4

THE EFFECT OF INTRA- AND INTERSPECIFIC COMPETITION ON COEXISTENCE IN MULTISPECIES COMMUNITIES

György Barabás[1], Matthew J. Michalska-Smith[1] & Stefano Allesina

The following was originally published in *The American Naturalist* in July 2016.

Barabás, György, Matthew J. Michalska-Smith, and Stefano Allesina. “”“The effect of intra-and interspecific competition on coexistence in multispecies communities.” The American Naturalist 188, no. 1 (2016): E1-E12.

Reprinted with permission.

For two competing species, intraspecific competition must exceed interspecific competition for coexistence. To generalize this well-known criterion to multiple competing species, one must take into account both the distribution of interaction strengths and community structure. Here we derive a multispecies generalization of the two-species rule in the context of symmetric Lotka–Volterra competition, and obtain explicit stability conditions for random competitive communities. We then explore the influence of community structure on coexistence. Results show that both the most and least stabilized cases have striking global structures, with a nested pattern emerging in both cases. The distribution of intraspecific coefficients leading to the most and least stabilized communities also follows a predictable pattern that can be justified analytically. In addition, we show that the size of the parameter space allowing for feasible communities always increases with the strength of intraspecific effects in a characteristic way

1. These authors contributed equally to this work.

that is independent of the interspecific interaction structure. We conclude by discussing possible extensions of our results to nonsymmetric competition.

4.1 Introduction

The familiar textbook statement, based on the analysis of two-species Lotka–Volterra competition, is that intraspecific competition must be greater than interspecific competition for two species to coexist (31, p. 331, 61, p. 104, 106, p. 130, 151, p. 208). This mathematical result formalizes the attractively simple intuition that species must limit themselves more than their competitors for coexistence [35]: if any one of the species reaches high abundance, it must hinder itself more than its competitor, ensuring that neither can completely take over.

One important question is how this simple and well-known coexistence rule generalizes to multispecies communities. When more than two species compete, things quickly get very complicated, even in the context of the classic competitive Lotka–Volterra model (99, 31, pp. 333-341). The difficulty is that community structure influences coexistence, but whereas two-species communities have a very simple structure, the same is not true of multispecies communities. Importantly, the two-species coexistence rule no longer holds. Indeed, it is possible to formulate even three-species scenarios where 1) every pair satisfies the rule, yet the three species cannot coexist; and 2) the pairs violate the rule, yet the three species do coexist at a stable equilibrium (Appendix B, Section 4.8.3). Therefore, in larger systems the two-species condition is neither necessary nor sufficient: it is, in general, no condition for coexistence at all. This means that the original, straightforward two-species intuition is not quite so straightforward in multispecies communities. One cannot simply think of such systems as collections of pairwise interactions, so that every pair fulfilling the two-species coexistence condition would guarantee the dynamical stability of feasible equilibria (a "feasible equilibrium" is one where all species have positive abundances).

Techniques for evaluating the stability of multispecies communities have been available for many decades, and include invasion analysis, loop analysis, and qualitative modeling. Invasion analysis [4, 34, 36, 38, 129] attempts to reduce the single, complicated problem of S competing species to S simple problems, where one species invades, from low abundance, the community formed by the other $S-1$ species at the stationary state corresponding to the absence of the invader. The species are considered coexisting if each has positive growth when invading. This approach has generated great advances, both on the theoretical [4, 34, 35, 129] and empirical [3, 85, 86, 128] side of community ecology. Unfortunately however, like any method, invasion analysis has its limitations. The outcome of invasion depends on the steady state of the $(S-1)$-species community where the invader is absent—however, obtaining this is often every bit as difficult as the original S-species problem, so nothing is gained. Also, there is no guarantee that in the absence of a single invader the rest of the community will not experience further extinctions, resulting in a state where the originally removed species cannot invade alone, but would be able to invade in concert with some other species [30, 53, 83]. Invasibility can therefore be too strict a criterion, which might erroneously classify cases of true coexistence as mere co-occurrence, in the sense of Leibold & McPeek [84].

Loop analysis and qualitative modeling [75, 87–89, 98] focus on evaluating the local stability of coexistence equilibria. Indeed, the two-species coexistence rule can be derived using loop analysis. The main limitation of these approaches is that results can typically be obtained only in cases where either the number of species is low, or the general structure of interactions especially simple [75]. The reason is that loop analysis is a "graphical translation" of the Routh-Hurwitz stability criteria [52]: as the number of species increases, both the number and complexity of these criteria increase, and even if one were to evaluate them somehow, the results would lack the clear biological interpretation that is available in e.g. the two-species case in terms of the intra- and interspecific competition coefficients.

Here we ask what can be said about the relationship between intra- and interspecific competition and coexistence in multispecies competitive communities, and how community structure affects this relationship in the context of Lotka–Volterra competition. Due to the aforementioned difficulties with classical methods of analysis, our strategy is instead to decompose the interaction matrix as the sum of an intra- and an interspecific part, and to study the dynamical stability of coexistence via the eigenvalues of these parts. We restrict most of our analyses to symmetric competition, commenting on the more general case in Appendix A ("The case of nonsymmetric interaction matrices") and the Discussion. Symmetric interaction matrices naturally arise in many competition models, e.g., in those based on niche overlap [31, 67, 87, 93, 94, 124]. Also, symmetry allows for a clear disentangling of the problems of stability and feasibility, makes local stability properties reflect global stability, and offers a way of relating intra- to interspecific competition not available in the nonsymmetric case.

This work is structured as follows. After reviewing our methods, we present simple analytical results extending the two-species coexistence rule to the multispecies case. Next, we relate the effect of community structure on intra- and interspecific competition and community stability, using random communities as our baseline (i.e., with competition coefficients randomly and independently sampled). We find that both the most and least stabilized communities possess simple, characteristic structures that can be interpreted biologically. Additionally, since an equilibrium needs to be both stable and feasible to describe coexistence [14, 62, 104, 122], we explore the influence of community structure on feasibility as well as stability. We conclude by discussing possible extensions of our results to the case of nonsymmetric competition.

4.2 Methods

We start from the generalized Lotka–Volterra equations for S species:

$$\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}t} = \mathrm{diag}(\mathbf{n})(\mathbf{b} + \mathbf{A}\mathbf{n}) \tag{4.1}$$

(Appendix B, Section 4.8.1), where t is time, $\mathbf{n}$ is the vector of species densities, $\mathrm{diag}(\mathbf{n})$ is the $S \times S$ matrix with the entries of $\mathbf{n}$ along its diagonal and zeros elsewhere, $\mathbf{b}$ is the vector of intrinsic growth rates, and $\mathbf{A}$ is the interaction matrix; its (i,j)th entry is the amount of change in species i's per capita growth rate caused by a unit increase in species j's density.

We wish to study the stability and feasibility of coexistence generated by eq. (4.1) in light of the distribution of intra- and interspecific interaction strengths. To disentangle the effects of the two, we write $\mathbf{A} = \mathbf{B} + \mathbf{C}$, where $\mathbf{B}$ only contains interspecific and $\mathbf{C}$ only intraspecific effects; therefore $\mathbf{B}$ is equal to $\mathbf{A}$ except it has zeros along its diagonal, and $\mathbf{C}$ contains only the diagonal entries of $\mathbf{A}$ with all offdiagonal entries being zero. We denote the (i,j)th entries of these matrices by A_{ij}, B_{ij}, and C_{ij}, respectively; since we are concerned with competition, all coefficients are nonpositive.

For $\mathbf{A}$ symmetric, its rightmost eigenvalue being less than zero guarantees the global stability of coexistence in the Lotka–Volterra model, provided the coexistence equilibrium is feasible (Appendix B, Section 4.8.1). This greatly simplifies stability analysis by obviating the need to evaluate the Jacobian at equilibrium. Also, since $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are all symmetric, their eigenvalues are real, so they can be numbered in decreasing order. Let the eigenvalues of $\mathbf{A}$ be $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_S$, those of $\mathbf{B}$ $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_S$, and those of $\mathbf{C}$ $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_S$. Note that, since $\mathbf{C}$ is a diagonal matrix, its eigenvalues are the diagonal entries themselves. Therefore $\gamma_1 = \max A_{ii} = \max C_{ii}$ is the weakest and $\gamma_S = \min A_{ii} = \min C_{ii}$ is the strongest intraspecific competition coefficient.

Moreover, the symmetry of $\mathbf{B}$ and $\mathbf{C}$ allows one to derive further conditions on $\mathbf{A}$'s stability. From Weyl's inequality [59], we have, respectively, the necessary and sufficient conditions

$$\beta_1 + \gamma_S < 0, \qquad \beta_1 + \gamma_1 < 0 \tag{4.2}$$

for the stability of $\mathbf{A}$ (Appendix B, Section 4.9). In words, the strongest intraspecific interaction being able to offset the interspecific effects is a necessary condition, while the weakest intraspecific interaction being able to offset the interspecific effects is a sufficient condition for the stability of $\mathbf{A}$. If all intraspecific coefficients are equal, then $\gamma_1 = \gamma_S$ and the two conditions are equivalent, providing a necessary and sufficient condition for stability. In case the intraspecific coefficients are not equal (the most general situation), fulfilling the sufficient condition $\beta_1 + \gamma_1 < 0$ guarantees stability.

Since the rightmost eigenvalue β_1 of $\mathbf{B}$ appears in both conditions, it is obviously important for the stability of coexistence. We are therefore interested in how community structure influences β_1. To study this question, we first generate $\mathbf{B}$ by sampling its entries randomly—this ensures that any resulting community structure is coincidental. Then, using a genetic algorithm (Appendix B, Section 4.11), we rearrange $\mathbf{B}$'s entries to minimize/maximize β_1, and analyze the structure of the resulting matrices. In other words, we keep the same set of entries but, preserving the symmetry of $\mathbf{B}$, assign them to different species pairs such that the community composition becomes easier/more difficult to stabilize than if we were to shuffle the interaction strengths at random. The minimum possible β_1 corresponds to the most stabilized, while the maximum β_1 to the least stabilized community.

Turning to the matrix $\mathbf{C}$ of intraspecific coefficients, we ask the same question: how should they be distributed across species to minimize/maximize the rightmost eigenvalue of $\mathbf{A}$; i.e., to lead to the most and least stabilized communities? Since the answer depends on the arrangement of the coefficients of $\mathbf{B}$, we perform the optimization of $\mathbf{C}$ (using the same

method as for $\mathbf{B}$) for all three cases: $\mathbf{B}$ arranged to minimize β_1, $\mathbf{B}$ random, and $\mathbf{B}$ arranged to maximize β_1.

Finally, we investigate the feasibility of the coexistence, i.e., determine when are all equilibrium densities strictly positive. The only equilibrium solution of eq. (4.1) where all species may coexist is $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{b}$, so for a given matrix $\mathbf{A}$ we can determine which vectors $\mathbf{b}$ make this product all-positive. Note that if $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{b}$ is feasible for some $\mathbf{b}$, then it is also feasible for $\eta\mathbf{b}$ with η an arbitrary positive constant, since this will lead to $-\mathbf{A}^{-1}(\eta\mathbf{b}) = \eta\mathbf{n}$ as the new equilibrium solution. Therefore, the magnitude of $\mathbf{b}$ is inconsequential for feasibility; only its direction matters. Using an appropriate integral formula (Appendix B, Section 4.13), we numerically determine the ratio of feasible to nonfeasible directions of the vector $\mathbf{b}$.

4.3 Interspecific competition and community stability

4.3.1 A simple multispecies generalization of the two-species coexistence rule

For two competing species, stability of the coexistence equilibrium $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{b}$ of eq. (4.1) is achieved if $A_{11}A_{22} > A_{12}A_{21}$, meaning that the geometric mean of the intraspecific competition coefficients has to exceed the geometric mean of the interspecific ones (Appendix B, Section 4.8.2; note that this two-species coexistence rule only guarantees stability, but not feasibility). What happens when the number of species is greater than two? We have seen that Weyl's inequality provides necessary and sufficient stability criteria if $\mathbf{A}$ is symmetric. In that case however, there also exists a more direct multispecies generalization. The result, based on a simple corollary of Sylvester's criterion (Appendix B, Section 4.8.5), is that for stability, all species pairs must individually satisfy the two-species coexistence rule. This condition is necessary but not sufficient, i.e., violating it guarantees instability, but fulfilling it does not guarantee stability.

4.3.2 Random interspecific interaction matrices

To study the effect of community structure on stability, let us start from a randomly assembled $S \times S$ interaction matrix. In particular: the matrix $\mathbf{B}$ of interspecific coefficients has zeros along the diagonal, every entry in the upper triangle is drawn independently from some distribution with mean $\mu < 0$ and variance V, and the lower triangle is filled out to make $\mathbf{B}$ symmetric. A similar approach has been followed for studying the stability of replicator dynamics; see Diederich & Opper [44] in the context of symmetric payoff matrices, and Opper & Diederich [114] and Galla [60] in the context of nonsymmetric ones. Importantly, the precise shape of the distribution from which the matrix entries are drawn does not matter; the mean and the variance fix all relevant properties of the matrix (Appendix B, Section 4.10). The matrix $\mathbf{C}$ of intraspecific coefficients has zeros on the offdiagonal and each diagonal entry is sampled from some distribution with mean $\mu_d < 0$ and variance V_d. The full interaction matrix is then given by $\mathbf{A} = \mathbf{B} + \mathbf{C}$. In constructing the matrices, we keep all entries of $\mathbf{A}$ nonpositive.

For random matrices, one can immediately obtain analytical stability criteria: for S large, the rightmost eigenvalue β_1 of $\mathbf{B}$ depends not on any of the fine details of the interactions within the community, but only on the number of species S, the mean μ, and the variance V via

$$\beta_1 \approx 2\sqrt{SV} - \mu \tag{4.3}$$

(Appendix B, Section 4.10), where the approximation improves with increasing S. Combining with Weyl's inequalities (eq. (4.2)), we obtain $\gamma_S < \mu - 2\sqrt{SV}$ and $\gamma_1 < \mu - 2\sqrt{SV}$ as approximations to the necessary and sufficient conditions for stability, respectively in the case of large S. Of course, for small communities, the approximation of β_1 is less accurate and correspondingly the inequalities may not hold absolutely.

There are three implications of this result. First, note that a matrix $\mathbf{A}$ can always be

stabilized by making its diagonal entries larger in magnitude than the sum of the magnitudes of the other entries in the same row ("diagonal dominance"; the stability of such a matrix follows directly from Gershgorin's circle theorem). If the offdiagonal entries have an average value of μ, then for S large, one would expect γ_1 to have to scale with S to achieve stability. Instead, from eq. (4.3) it is clear that γ_1 only needs to scale with $\sqrt{S}$, meaning that much weaker intraspecific coefficients are sufficient to stabilize coexistence. For instance, in a community of 100 species, γ_1 would need to be proportional to $\sim 10\mu$ instead of $\sim 100\mu$.

Second, β_1 scales with $\sqrt{V}$ as well. The variance of the interspecific interaction strengths thus has a very important effect on stability: all other things being equal, communities with a smaller variance are easier to stabilize than communities with larger variances.

Third, μ must of course be a negative number in competitive systems, measuring the average competitive effect of one species on another across the community. It therefore impacts the amount of intraspecific stabilization required for coexistence via eq. (4.3). However, for S large, the effect of μ in the expression $\gamma_1 < \mu - 2\sqrt{SV}$ is going to be dominated by the square root term. For large communities, we expect the mean interaction strength μ to be much less important in determining stability than the variance V.

This means that, for species-rich assemblages, intraspecific competition has to be not just stronger, but substantially stronger than average interspecific competition. Using the condition $\gamma_1 < \mu - 2\sqrt{SV}$, if the interspecific coefficients are uniformly sampled from $[-1, 0]$ ($\mu = -1/2,\ V = 1/12$), then for $S = 12$ we get $\gamma_1 < -2.5 = 5\mu$, while $S = 300$ leads to $\gamma_1 < -10.5 = 21\mu$, with intraspecific effects having to be at least 21 times stronger than interspecific ones to achieve stability.

4.3.3 Nonrandom interspecific interaction matrices

What is the effect of a nonrandom community structure on competitive coexistence? We now take the original, randomly assembled matrix $\mathbf{B}$ and rearrange its entries to obtain the

minimum/maximum possible values for β_1 (fig. 4.1; note that, for β_1 minimal, species are sorted in increasing order of the leftmost eigenvector's entries, while for for β_1 random and β_1 maximal, they are sorted in increasing order of the rightmost eigenvector's entries). Both the minimum and maximum β_1 cases possess a characteristic structure. The minimum β_1 case is perfectly hierarchical [134]: interaction strengths always decrease moving downwards and right when starting from the top left entry of the matrix. The maximum β_1 case is also hierarchical (here competitive effects always increase going upwards and right—or downwards and left—in the matrix, starting anywhere along the main diagonal), but in addition, one can also classify the species into two roughly distinct groups, with competitive interactions significantly weaker within than between those groups.

Both these scenarios have ecological interpretations. For the minimum β_1 case, consider a resource continuum. Now imagine that Species 1 consumes some range of these resources. Species 2 consumes a proper subset of the resources Species 1 consumes, Species 3 a proper subset of the resources Species 2 consumes, and so on, with the resource spectrum of Species S being the narrowest and a proper subset of the resource spectra of all the other species. Assume now that the competition coefficients are proportional to the overlap in resource use, and we end up with the structure in the top left of fig. 4.1. Since this arrangement leads to the smallest β_1 possible, this biological scenario is the easiest to stabilize from a dynamical perspective.

For the maximum β_1 case, imagine a unidimensional trait axis along which the intensity of competition is an *increasing* function of trait difference, instead of the usual assumption of decreasing competition. For instance, the axis could describe the quality of toxin produced by species of allelopathic plants, where each plant species is more resistant to toxins that are similar to its own. If we now assume that the species are sorted into two groups and that between-group distances along the axis are much larger than within-group distances, we get the structure in the bottom left of fig. 4.1. This scenario is also the most difficult to stabilize,

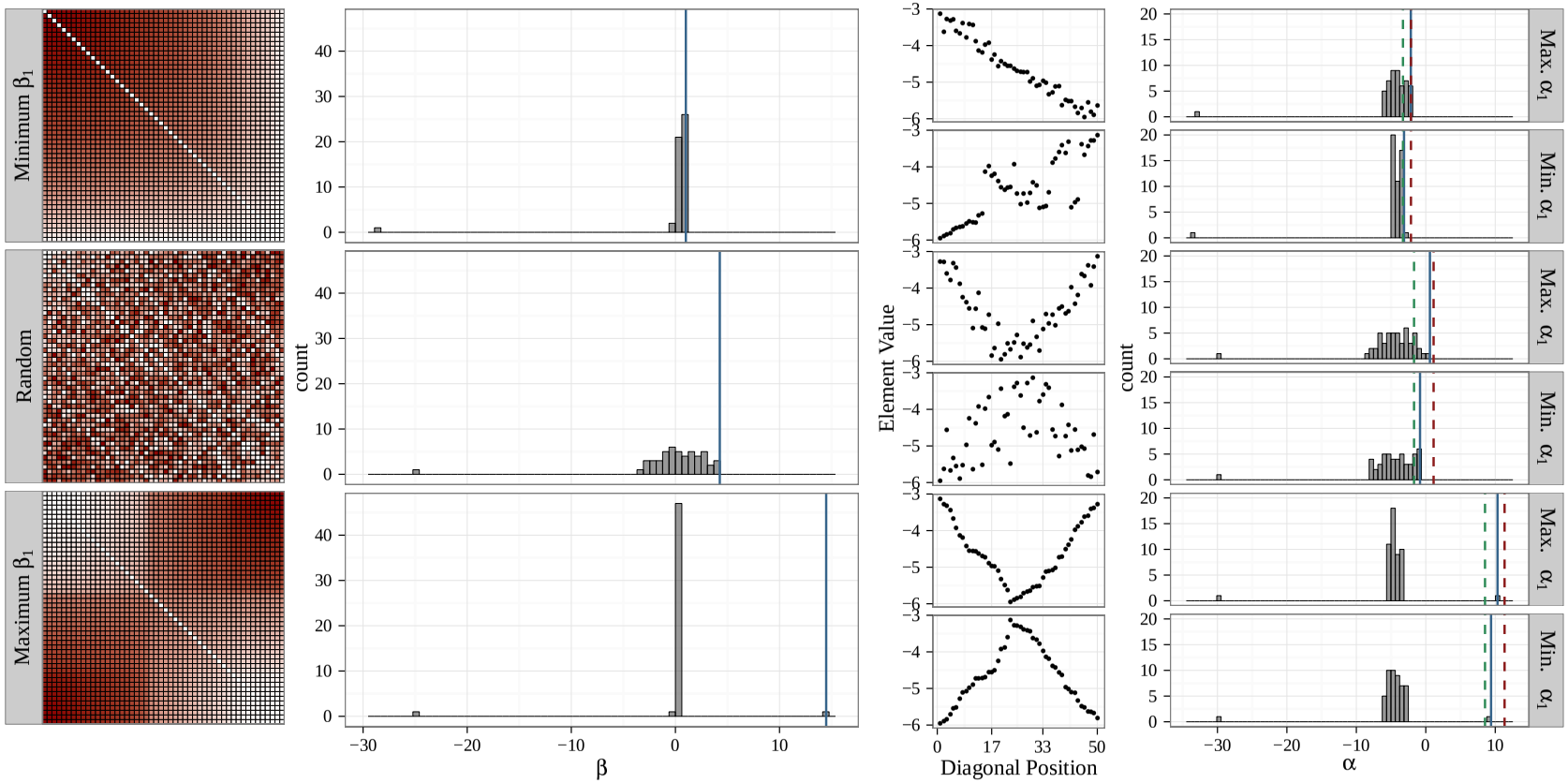


Figure 4.1: Columns from left to right: first, matrices of interspecific competition coefficients $\mathbf{B}$ in a 50-species community with the same set of interaction strengths (uniformly sampled from $[-1, 0]$), but arranged in three different ways: minimizing the rightmost eigenvalue β_1 (top), random arrangement (middle), and maximizing β_1 (bottom). Darker colors correspond to stronger coefficients. While the minimizing case (top) is sorted in increasing order of the components of the eigenvector corresponding to the leftmost eigenvalue, the other two are sorted in increasing order of the components of the eigenvector associated with the rightmost eigenvalue. Second, histograms of the eigenvalue distributions of matrices depicted in the first column. The solid blue line marks the precise position of the rightmost eigenvalue β_1. Third, for each matrix $\mathbf{B}$ in column one, the coefficients of matrix $\mathbf{C}$ (uniformly sampled from $[-6, -3]$) are rearranged to provide the maximum and minimum α_1. The arrangement of these coefficients is plotted, sorted according to their position in the matrices in column one. Fourth, histograms of the eigenvalue distributions of $\mathbf{A} = \mathbf{B} + \mathbf{C}$, where $\mathbf{B}$ is given in the first column, and $\mathbf{C}$ is a matrix of intraspecific effects, with diagonal entries indicated by column three. The solid blue line again marks the precise position of the rightmost eigenvalue α_1, and the dashed green/red lines are the bounds on the rightmost eigenvalue calculated from eq. (4.2). (See (figs. 4.10 to 4.29) for replicates with different parameterizations, showing the robustness of the results.)

since it produces the largest leading eigenvalue β_1 possible with the given set of coefficients.

One can look at the structure of interactions in a different, pairwise way (fig. 4.2). We line up all 50 species of the community horizontally, and connect two species by a blue arc if they would coexist in isolation (intraspecific competition greater than interspecific competition between them), and with a red arc if they would not. In the minimum β_1 case, red arcs disappear in a pattern from right to left as intraspecific competition is increased, reflecting species' positions in the hierarchy. In the maximum β_1 case, two clear groups emerge as intraspecific competition is increased: those within the groups can coexist pairwise, but those across the groups cannot, in line with the biological interpretation given above. The size of the groups is dictated by the relative strength of intra- to interspecific competition. In all cases, there can be no coexistence in the limit of no intraspecific competition, while in the case of strong intraspecific competition all species pairs would coexist in isolation—though, as discussed before, this does not guarantee community-wide coexistence.

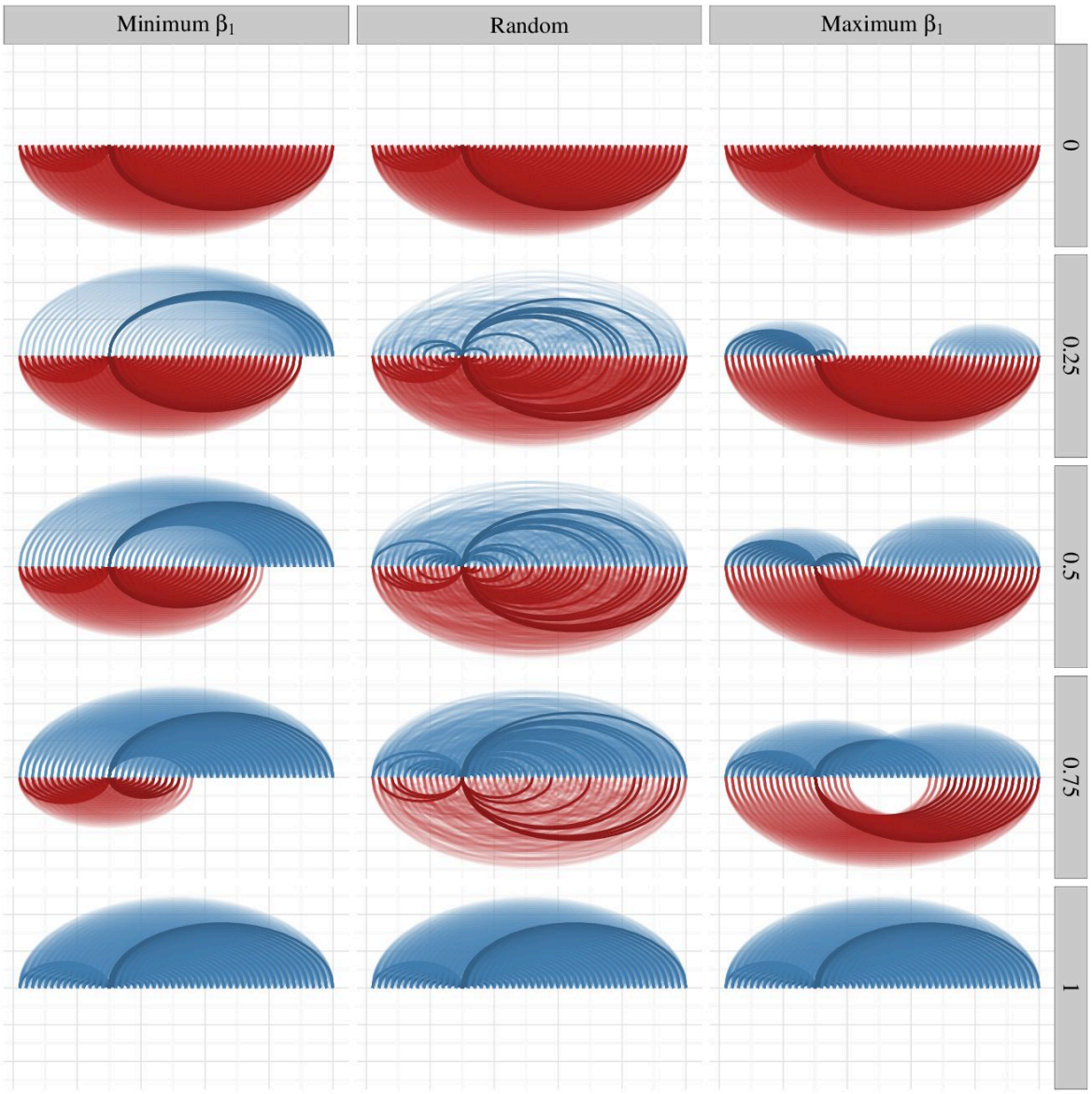


Figure 4.2: Pairwise interactions in a 50-species competitive community. Columns correspond to different arrangements of the interspecific interaction matrix **B**: minimizing its rightmost eigenvalue β_1 (left), random (middle), and maximizing β_1 (right). Rows correspond to different values of the intraspecific competition coefficient, here assumed to be the same for all species. In each panel, a blue/red connection between two species means that those two species could/could not coexist in a pairwise manner. Highlighted arcs are interactions of a particular species (arbitrarily chosen to be the 15th species from the left in each ordering). Note the regular patterns of blue and red arcs emerging in the minimum and maximum β_1 columns, reflecting the community structures seen on fig. 4.1.

4.4 Intraspecific competition and community stability

Once we arrange the matrix **B** of interspecific coefficients to minimize/maximize β_1, we can ask how one should distribute a given set of intraspecific coefficients **C** across species to obtain the smallest/largest value of α_1, the rightmost eigenvalue of the full interaction matrix **A** determining overall community stability.

Figure 4.1 shows how the eigenvalue distribution changes according to the arrangement of intra- and interspecific coefficients. For each arrangement of interspecific competition coefficients, we found the arrangement of intraspecific ones minimizing and maximizing the rightmost eigenvalue α_1 of **A**. While each of the six cases exhibits a different pattern, they are all consistent with a simple rule of thumb: to maximize α_1, stronger intraspecific entries should be associated with species whose interspecific interaction strengths have smaller variance, and vice versa. Conversely, to minimize α_1, stronger intraspecific coefficients should be paired with larger variance in interspecific interaction strengths. For example, to maximize α_1 in the minimum β_1 case in fig. 4.1, the first species should be assigned the weakest intraspecific interaction, the second species the second weakest, and so on, with the Sth species (smallest variance in interspecific interaction strengths) having the strongest one.

These patterns are only approximate, and are noticeably fuzzy in the case where both β_1 and α_1 are minimized, and in both of the random cases[2]. We do not know whether our rule of thumb is indeed a general rule. However, It can be shown that the observed patterns are always the theoretically expected ones in our particular scenarios (Appendix B, Section 4.12). A more intuitive explanation can also be given. In general, a higher variance in the interspecific effects also implies a larger absolute magnitude of the strongest intraspecific effect experienced by a species (fig. 4.1). Assigning the weakest interspecific

2. One might wonder why there is a pattern to the random cases at all. The answer is that species are ordered by the leading eigenvector, which is nonrandom with respect to the (otherwise randomly determined) variance of the rows' entries. See the random case in fig. 4.1: there is a visible trend for stronger coefficients towards the upper right and lower left corners of the matrix.

coefficients to species with the highest variance is therefore the most likely way of violating Sylvester's necessary criterion that all species pairs follow the two-species rule for coexistence, destabilizing the system.

4.5 Feasibility

Dynamical instability precludes, but dynamical stability does not guarantee coexistence: the equilibrium solution also needs to be feasible, i.e., $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{b}$ needs to be strictly positive. For two species, writing $\mathbf{b} = (\cos\theta, \sin\theta)$ where θ defines the direction of $\mathbf{b}$, this translates into the the feasibility condition

$$\arctan\left(\frac{A_{21}}{A_{11}}\right) < \theta < \arctan\left(\frac{A_{22}}{A_{12}}\right) \tag{4.4}$$

(Appendix B, Section 4.8.2). That is, for any given interaction matrix $\mathbf{A}$, one can determine all directions of $\mathbf{b}$ leading to a feasible equilibrium. This is best expressed via the concept of the *feasibility domain* Ξ, the proportion of feasible directions to all possible directions [62, 122, 140]. In the two-species case,

$$\Xi = \frac{2}{\pi}\max\left\{\arctan\left(\frac{A_{22}}{A_{12}}\right) - \arctan\left(\frac{A_{21}}{A_{11}}\right), 0\right\}. \tag{4.5}$$

Below, instead of Ξ, we use the (geometric) average feasible domain *per dimension* $\sqrt[S]{\Xi}$, a more meaningful metric which makes feasilibity comparable across systems with different numbers of species (Appendix B, Section 4.13).

To explore the effect of intra- and interspecific interaction strengths on the feasibility of competitive communities, we obtained $\sqrt[S]{\Xi}$ via numerical integration (Appendix B, Section 4.13) for all three cases of distributing the interspecific interaction strengths: minimum β_1, random, and maximum β_1. For each of the three cases we gradually increased the strength of overall intraspecific competition, assumed to be the same for all species. We only calculated

the feasibility domain for stable matrices, since a feasible unstable equilibrium does not describe coexistence. We know a priori that $\sqrt[S]{\Xi}$ must be zero at the boundary of stability and instability where the rightmost eigenvalue crosses the origin of the complex plane, because at this point the model is structurally unstable. As intraspecific competition is increased from this point on, the matrices invariably obtain larger domains of feasibility $\sqrt[S]{\Xi}$ (fig. 4.3). Surprisingly, after an initial, very steep increase of feasibility from zero up, $\sqrt[S]{\Xi}$ settles down to the same characteristic curve independent of the structure of interspecific interactions. In the big picture, intraspecific competition has the most influence on feasibility, whereas interspecific competition has the role of determining the point at which the community becomes stable in the first place.

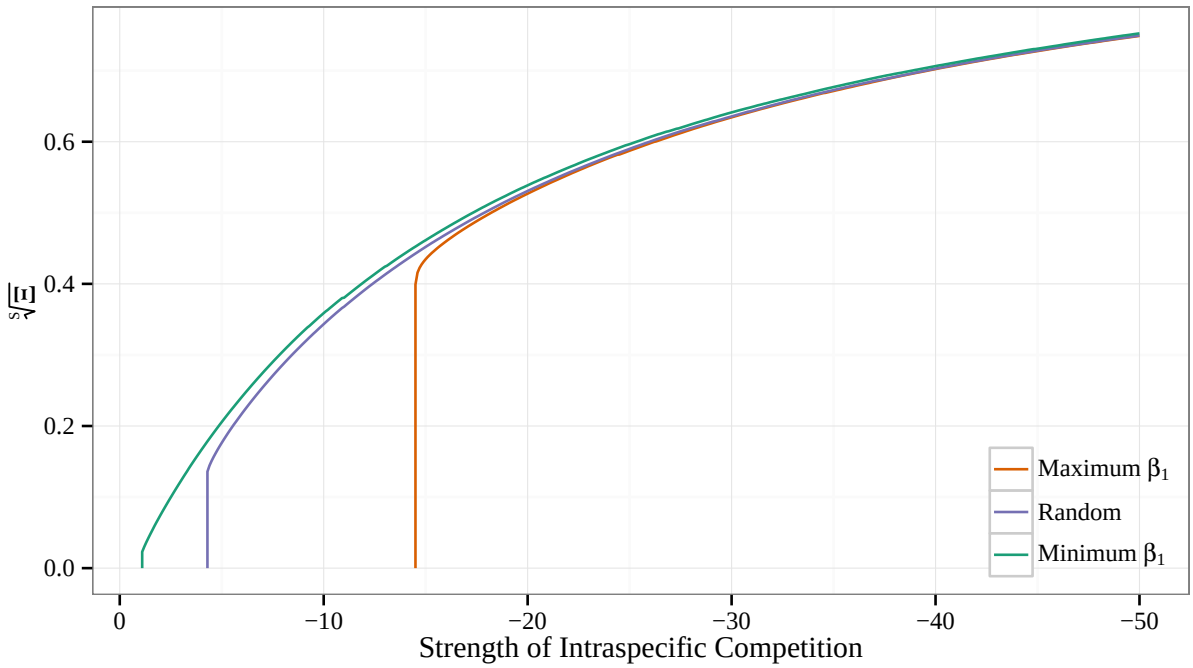


Figure 4.3: The feasibility domain $\sqrt[S]{\Xi}$ as a function of intraspecific interaction strength, assumed here to be constant across $S = 50$ species. The curves correspond to β_1 maximized (orange), β_1 corresponding to a random competition matrix (purple), and β_1 minimized (green), where β_1 is the rightmost eigenvalue of the matrix $\mathbf{B}$ of interspecific competition coefficients. Note that the abscissa is reversed in order to better indicate increasing strength of intraspecific competition. In each case, $\sqrt[S]{\Xi}$ is plotted only for stable matrices. Except for the discrepancy near the boundary between stability and instability, all three cases eventually converge on the same characteristic feasibility curve for any strength of intraspecific competition.

4.6 Discussion

In this work we explored the relationship between intra- and interspecific competition and its effect on coexistence in multispecies competitive communities. Our strategy was to decompose the interaction matrix into intra- and interspecific parts, allowing us to bound the stability of the community in terms of their eigenvalues via Weyl's inequality. We could then obtain simple stability criteria for random competitive communities. Using the random case as a springboard, we searched for the community structures that minimize/maximize the rightmost eigenvalue of the interaction matrix, both in terms of intra- and interspecific

effects. Both the most and the least stabilized case exhibited a characteristic interspecific structure, with a hierarchical arrangement of coefficients, and a quasi-bipartite structure in the least stabilized configuration. Intraspecific effects were most stabilizing when species with the largest variance in interspecific interactions were assigned strong intraspecific coefficients, and most destabilizing when they were assigned weak ones. The feasible fraction of parameter space invariably increased with a decreasing rightmost eigenvalue, showing that systems that are more likely to be stable are also more robust against environmental perturbations of the intrinsic growth rates.

Conclusions about the relationship between intra- and interspecific competition have largely stemmed from experiments and analyses which considered very few (usually two) species (e.g., 1, 4, 21, 33, 58, 65, 109). In this case, the conditions for the stability of the interaction matrix translate into simple inequalities containing the competition coefficients, with the direct biological interpretation that species must limit themselves more than their competitors for coexistence. As the number of species increases, the Routh-Hurwitz stability criteria [52] grow in number and complexity; each must be satisfied for stability, yet their biological implications are no longer easily derived. Therefore, in general, there is no simpler way to answer the question of stability than to state that "the rightmost eigenvalue has to be in the left half plane". The necessary condition for symmetric matrices that all species pairs must individually satisfy the familiar two-species coexistence condition is a convenient exception to this rule.

Two-species competitive communities have a simple and transparent structure. For multiple species, many different structures are possible, with large effects on community stability. We have shown that the same set of coefficients can lead to vastly different stability properties: communities that were most stabilized (hierarchical arrangement) or least stabilized (hierarcy, but with two groups and weak interactions within, strong interaction between group members) both exhibited a characteristic structure. The wide range of possible right-

most eigenvalues for the exact same set of coefficients, with only the community structure varying, is one of the main reasons why there is no trivial generalization of the two-species coexistence rule to multispecies communities. Nevertheless, it is often implied, without being explicitly stated, that the two-species criterion holds for an arbitrary number of species [4, 10, 15, 37, 39, 71]. It does not—except as a necessary but insufficient stability condition for symmetric interaction matrices only.

Of course, it must be true in a broader sense that, in order to coexist, species must limit themselves more strongly than they limit their competitors: if we add individuals to a species at a stable attractor, that same number of individuals will eventually die and the community will return to its state prior to the perturbation. Attributing this return exclusively to the relative strength of intra- and interspecific competition between pairs of species, however, potentially ignores the myriad indirect higher-order feedbacks influencing species trajectories, which can override the simple two-species coexistence rule. For instance, in reactive systems a perturbation to species abundances may initially be amplified before returning to equilibrium [110, 141]. In such cases, the overall requirement for stability does not translate directly into a straightforward rule for the competition coefficients themselves. Interestingly, symmetric matrices cannot be reactive—the same situation for which there *is* a reasonably straightforward generalization of the two-species rule of coexistence.

What can be said about nonsymmetric interaction matrices? Even in cases where symmetric competition is a consequence rather than an assumption of the model in question [31, 67, 93, 124], other confounding factors not included in the model will likely distort perfect symmetry. In other cases, such as size-structured competition for light in forest trees [2, 78, 79], asymmetry is an essential part of the system's ecology.

We do not expect our results to change substantially for mild degrees of asymmetry, since that can be accounted for as a small perturbation to the symmetric model. For substantial asymmetry however, several difficulties emerge, rendering most of the techniques applied

in this work inapplicable (see Appendix A). Some conclusion can still be made though for random competitive communities, as long as the variance of the intraspecific effects around a mean γ is not very large. Then one can derive the stability condition $\gamma > (1+\rho)\sqrt{SV} - \mu$, with ρ being the average correlation between the effect of species i on j and the effect of j on i. For strongly negative values of ρ, this implies that interspecific effects can be much larger than intraspecific ones and coexistence can still be stable, at least locally. In fact, the aforementioned case of size-structured competition for light in forests, where taller trees have a much larger effect on shorter ones than vice versa, presents an example where negative values of ρ are expected. Such systems could then be stabilized by much weaker intraspecific effects than corresponding symmetric cases. It is interesting to wonder whether and how much this structural property is responsible for explaining the general prevalence of such hierarchical competitive systems. Whether their stability is in fact explained in this way awaits further investigations.

Acknowledgements

We thank R. D'Andrea, J. Grilli, G. Kylafis, A. Ostling, L. Pásztor, R. Rael, and E. Sander for discussions. Comments by M. Leibold, D. Stouffer, and an anonymous reviewer all helped improve the clarity of the manuscript. This work was supported by NSF #1148867.

Appendix A: The case of nonsymmetric interaction matrices

When the interaction matrix $\mathbf{A}$ is nonsymmetric, substantial complications emerge which make it difficult to say much about the relationship between intra- and interspecific competition and coexistence. First, Sylvester's criterion no longer applies, therefore there is

no guarantee that a species pair violating the two-species coexistence rule will lead to the loss of coexistence (see Appendix B, Section 4.8.3 for a three-species, and fig. 4.4 for a 50-species example). Second, local stability no longer guarantees global stability of coexistence, and local instability does not mean the lack of coexistence, e.g., along some periodic or chaotic orbit (Appendix B, Section 4.8.3 and section 4.8.4). Third, $\mathbf{A}$ no longer determines stability; instead, the Jacobian evaluated at the equilibrium $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{b}$ does. This reads $\mathbf{J} = \mathrm{diag}(\mathbf{n})\mathbf{A} = -\mathrm{diag}(\mathbf{A}^{-1}\mathbf{b})\mathbf{A}$ (Appendix B, Section 4.8.1). Since $\mathbf{J}$ now depends on the intrinsic rates, it is impossible to give a formal coexistence criterion independent of $\mathbf{b}$. Fourth, Weyl's inequality no longer holds, making the decomposition of $\mathbf{J}$ into intra- and interspecific parts ineffective.

Despite these difficulties, some conclusions can still be made for random asymmetric competitive communities. As long as the variance of the equilibrium densities is small, we have $\mathrm{diag}(\mathbf{n}) \approx \bar{n}\,\mathbf{I}$, where $\bar{n}$ is the average abundance and $\mathbf{I}$ is the identity matrix. Then, $\mathbf{J} = \mathrm{diag}(\mathbf{n})\mathbf{A} \approx \bar{n}\,\mathbf{A}$, therefore $\mathbf{A}$ still determines stability, albeit only locally. If we now perform the $\mathbf{A} = \mathbf{B} + \mathbf{C}$ decomposition where $\mathbf{B}$ contains the interspecific and $\mathbf{C}$ the intraspecific coefficients, one can apply the elliptic law of random matrix theory (115, 133; Appendix B, Section 4.10.2) to find the leading eigenvalue β_1 of $\mathbf{B}$:

$$\beta_1 \approx (1+\rho)\sqrt{SV} - \mu, \tag{4.6}$$

where ρ is the average correlation between B_{ij} and B_{ji}. Then, as long as the variance of the intraspecific effects around a mean γ is not too large (i.e., $\mathbf{C} \approx \gamma\,\mathbf{I}$), we have the stability condition

$$\gamma > (1+\rho)\sqrt{SV} - \mu \tag{4.7}$$

(Appendix B, Section 4.10).

Due to the similarity of eqs. (4.3) and (4.6), much same conclusions apply as in the symmetric case—but only as long as ρ is positive or not too strongly negative. Strong negative correlations on the other hand can substantially reduce the amount of intraspecific competition required for stability. As an extreme example, if $\rho = -1$, then $\beta_1 = -\mu$, and therefore $\gamma < \mu$ will lead to stability regardless of the number of species S or the variance V in interspecific effects. Even if the variance is much larger than the mean, stability is ensured. In other words, for strongly negative values of ρ, interspecific competitive effects may dominate over intraspecific ones, and the system can still be stable. See fig. 4.4 for an example with $\gamma = -2.2$, $\mu = -2$, $V = 0.3$, and $\rho = -0.95$.

The most important restrictive assumption above is that the variance of the species abundances around $\bar{n}$ is small. Real communities, on the other hand, possess a characteristic species abundance distribution close to lognormal, with a large variance [102]. Moreover, theoretical investigations of the species abundance distribution under the replicator dynamics with both symmetric [148] and nonsymmetric [155] payoff matrices have arrived at the exact same conclusion. The species abundance distributions produced by the model are therefore consistent with empirically observed ones, rendering the assumption of nearly equal equilibrium abundances implausible.

However, preliminary explorations [5] show that when the equilibrium abundances $\mathbf{n}$ follow realistic species abundance distributions, the elliptic law consistently overestimates the leading eigenvalue. That is, $\bar{n}\,\mathbf{A}$ always has a larger leading eigenvalue than $\mathrm{diag}(\mathbf{n})\mathbf{A}$, and sometimes substantially so. Equation (4.6) can therefore be seen as a conservative estimate of stability. We therefore expect this criterion, based on the elliptic law, to provide a sufficient stability condition for realistic communities.

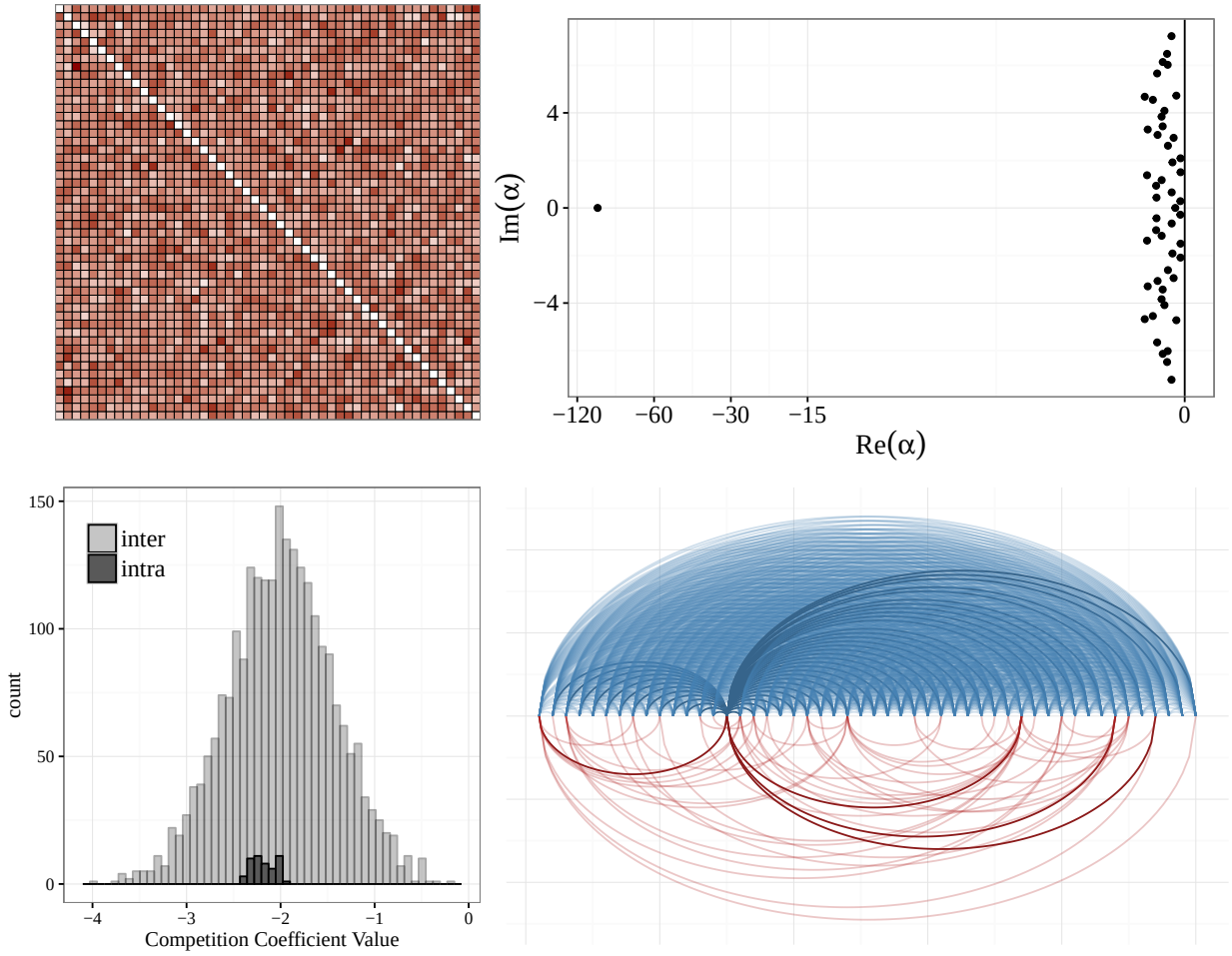


Figure 4.4: A highly asymmetric, stable competitive community of 50 species. Interspecific coefficient pairs (B_{ij}, B_{ji}) were randomly drawn from a bivariate normal distribution with marginal means $\mu = -2$, marginal variances $V = 0.3$, and correlation $\rho = -0.95$. Intraspecific coefficients C_{ii} were uniformly sampled from $[-2.4, -2]$. Top left: the matrix $\mathbf{B}$ of interspecific coefficients. Top right: eigenvalue distribution of $\mathbf{A} = \mathbf{B} + \mathbf{C}$ in the complex plane. Since all eigenvalues have negative real parts, the system is locally stable. Bottom left: distribution of intra- and interspecific interaction strengths; note the substantially larger variance of the interspecific coefficients. Bottom right: species (nodes) connected by blue/red arcs if they can/cannot coexist as pairs. Highlighted arcs are connections of species no. 15. Note that several species pairs violate the pariwise coexistence rule, yet the community as a whole is locally stable.

Appendix B: Supporting Information

4.7 Dynamical stability, structural stability, and feasibility

Consider an autonomous dynamical system with S variables $n_1, \ldots, n_S$ (representing species abundances or biomasses) and parameters $\mu_1, \ldots, \mu_k$:

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = f_i(n_1, \ldots, n_S, \mu_1, \ldots, \mu_k), \tag{4.8}$$

where t is time, and the f_i are functions of the n_i and the model parameters (the time-dependence of the n_i is suppressed for notational convenience). An equilibrium of the system is a set of values $\hat{n}_1, \ldots, \hat{n}_S$ such that all rates of change are simultaneously zero:

$$f_i(\hat{n}_1, \ldots, \hat{n}_S, \mu_1, \ldots, \mu_k) = 0 \tag{4.9}$$

for all $i = 1, \ldots, S$. The *local dynamical stability* of an equilibrium can be determined by inspecting the eigenvalues of the Jacobian matrix $J_{ij} = \partial f_i / \partial n_j$ evaluated at the equilibrium: stability requires that all eigenvalues have negative real parts. An equilibrium that is not locally dynamically stable is called (locally, dynamically) unstable. A "stable matrix" is one with all its eigenvalues having negative real parts.

The dynamics is said to be *structurally stable* if sufficiently small perturbations of the parameters $\mu_1, \ldots, \mu_k$ cannot qualitatively alter the behavior of the system's trajectories. Equilibria with eigenvalues sitting exactly on the imaginary axis present an example of structural instability, as an arbitrarily small perturbation could in principle tip those eigenvalues either in the stable or the unstable direction in the complex plane, qualitatively changing system behavior.

The values of the abundances n_i at an equilibrium point are in principle arbitrary. However, in ecological systems we are usually interested in equilibria that are strictly positive, i.e., $n_i > 0$ for all i. These are called *feasible* equilibrium points.

4.8 The Lotka–Volterra model

4.8.1 General properties

The generalized Lotka–Volterra model for S species with abundances $n_1, \ldots, n_S$ can be written

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = n_i \left(b_i + \sum_{j=1}^{S} A_{ij} n_j \right), \tag{4.10}$$

where b_i is the intrinsic growth rate of species i, and the A_{ij} are interaction coefficients: A_{ij} measures the amount of change in species i's per capita growth rate caused by a unit change in species j's density.

The equation can be written in vector form. Collecting the coefficients b_i into the vector $\mathbf{b}$, the abundances n_i into the vector $\mathbf{n}$, the interaction coefficients A_{ij} into the matrix $\mathbf{A}$, and letting $\mathrm{diag}(\mathbf{n})$ be the diagonal matrix with the entries of the vector $\mathbf{n}$ along its diagonal and zeros elsewhere, eq. (4.10) reads

$$\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}t} = \mathrm{diag}(\mathbf{n})(\mathbf{b} + \mathbf{A}\mathbf{n}). \tag{4.11}$$

The Lotka–Volterra model with S species may have up to 2^S equilibria, out of which at most one is feasible. We call this the *coexistence equilibrium* (even when it does not describe actual coexistence). From eq. (4.11), this equilibrium is given when all the $\mathbf{n}$ are assumed nonzero:

$$\mathbf{b} + \mathbf{A}\hat{\mathbf{n}} = 0, \tag{4.12}$$

where the hat denotes equilibrium values. Therefore,

$$\hat{\mathbf{n}} = -\mathbf{A}^{-1}\mathbf{b}, \tag{4.13}$$

with $\mathbf{A}^{-1}$ being the inverse of $\mathbf{A}$. This equilibrium is feasible if $\hat{\mathbf{n}}$ is strictly positive. Note that if a vector $\mathbf{b}$ leads to a feasible solution, so does the vector $\eta\mathbf{b}$ for an arbitrary positive constant η. Indeed, for $\mathbf{b}^* = \eta\mathbf{b}$ we get the solution $\hat{\mathbf{n}}^* = -\mathbf{A}^{-1}\mathbf{b}^* = -\mathbf{A}^{-1}\eta\mathbf{b} = \eta\hat{\mathbf{n}}$, which is still feasible if $\eta > 0$. For a given matrix $\mathbf{A}$ therefore, only the direction of $\mathbf{b}$ matters for feasibility, not its absolute magnitude. This opens the way for describing feasibility via the fraction of directions in which $\mathbf{b}$ is feasible, out of all possible directions.

To determine the stability of the equilibrium given by eq. (4.13), we calculate the Jacobian $\mathbf{J}$ with entries $J_{ij} = \partial(\mathrm{d}n_i/\mathrm{d}t)/\partial n_j$, using eq. (4.11). The result is $\mathbf{J} = \mathrm{diag}(\mathbf{b} + \mathbf{A}\mathbf{n}) + \mathrm{diag}(\mathbf{n})\mathbf{A}$. Evaluating $\mathbf{J}$ at equilibrium, the first term cancels because $\hat{\mathbf{n}} = -\mathbf{A}^{-1}\mathbf{b}$, therefore we get $\mathbf{J}|_{\hat{\mathbf{n}}} = \mathrm{diag}(\hat{\mathbf{n}})\mathbf{A} = -\mathrm{diag}(\mathbf{A}^{-1}\mathbf{b})\mathbf{A}$. Stability of this matrix guarantees the local stability of the coexistence fixed point, though not its feasibility.

Important results follow if $\mathbf{A}$ is a symmetric matrix with all negative entries; i.e., when interactions are purely competitive. We then have the following [67, 68, 92]. First, the system cannot exhibit cycles, chaos, or any other behavior than simple convergence to a fixed point. Second, if $\hat{\mathbf{n}}$ is feasible and locally stable, it is globally stable as well. Third, stability of the coexistence equilibrium depends only on the stability of $\mathbf{A}$ instead of the full Jacobian evaluated at equilibrium, $\mathbf{J}|_{\hat{\mathbf{n}}} = \mathrm{diag}(\hat{\mathbf{n}})\mathbf{A}$. Since $\mathbf{A}$ is symmetric, all its eigenvalues are real; as long as all of them are negative, the coexistence equilibrium is (globally) stable.

4.8.2 Two species

In the well-known textbook example of the competitive two-species Lotka–Volterra equations, one has the great simplification that the coexistence equilibrium is globally stable or unstable even if $\mathbf{A}$ is not symmetric (the easiest way to prove this is by constructing appropriate Lyapunov functions; see, e.g., 157). Therefore, if coexistence is locally stable, it is also globally stable. Based on this, four outcomes are possible:

1. No feasible equilibrium point exists; species 1 wins from any positive initial condition.

2. No feasible equilibrium point exists; species 2 wins from any positive initial condition.

3. There is a feasible equilibrium point but it is unstable; species 1 or 2 wins depending on initial conditions.

4. There is a feasible equilibrium point and it is stable; the two species coexist from any positive initial condition.

It should be mentioned that there are other, degenerate outcomes as well (for instance, the two species could be exactly identical in all their parameters, leading to a one-dimensional, neutrally stable manifold), but these examples are structurally unstable.

The coexistence equilibrium will be feasible if it is strictly positive: $\hat{\mathbf{n}} > \mathbf{0}$. Using eq. (4.13), this requirement can be written as $-\mathbf{A}^{-1}\mathbf{b} > \mathbf{0}$. Writing this inequality out in components for two species and simplifying, one arrives at

$$\frac{A_{12}}{A_{22}} < \frac{b_1}{b_2} < \frac{A_{11}}{A_{21}} \tag{4.14}$$

(e.g., 95, 150). This same set of conditions may be written in a slightly different form. As discussed before, only the direction of $\mathbf{b}$ matters for feasibility, not its magnitude. We therefore parameterize the intrinsic growth rates as $\mathbf{b} = (\cos\theta, \sin\theta)$, i.e., $\mathbf{b}$ has length 1, and $0 < \theta < \pi/2$ is the angle measured counterclockwise from the abscissa of the two-dimensional plane of possible $\mathbf{b}$ vectors. Substituting this into eq. (4.14) and simplifying, we get

$$\arctan\left(\frac{A_{21}}{A_{11}}\right) < \theta < \arctan\left(\frac{A_{22}}{A_{12}}\right). \tag{4.15}$$

Expressing this as the proportion of feasible θ values out of all possible values $0 < \theta < \pi/2$, and calling this fraction Ξ, we get

$$\Xi = \frac{2}{\pi} \max\left\{\arctan\left(\frac{A_{22}}{A_{12}}\right) - \arctan\left(\frac{A_{21}}{A_{11}}\right),\, 0\right\}, \tag{4.16}$$

where $\max(\cdot, \cdot)$ picks the larger of the two arguments.

For dynamical stability, all eigenvalues of $\mathbf{J}|_{\hat{\mathbf{n}}} = \mathrm{diag}(\hat{\mathbf{n}})\mathbf{A}$ (the Jacobian evaluated at the equilibrium point) must have negative real parts. For two species, and two species only, there is a shortcut: the coexistence equilibrium for competitive dynamics is stable if and only if $\mathbf{A}$ has a positive determinant [157]. Writing this condition out, we get

$$\det(\mathbf{A}) = A_{11}A_{22} - A_{12}A_{21} > 0, \tag{4.17}$$

or, equivalently,

$$\sqrt{A_{11}A_{22}} > \sqrt{A_{12}A_{21}}\,. \tag{4.18}$$

The left hand side is the geometric mean of the two intraspecific competition coefficients, while the right hand side is the geometric mean of the interspecific coefficients.

This condition is often interpreted as intraspecific competition having to be stronger than interspecific competition for coexistence. Note that this is a necessary but not a sufficient coexistence condition, as eq. (4.14) also needs to be satisfied. While true in the form just presented, one should be careful not to overinterpret this simple result. For instance, consider the following parameterization of the two-species Lotka–Volterra model:

$$\mathbf{b} = \begin{pmatrix} 6 \\ 10 \end{pmatrix}, \quad \mathbf{A} = -\begin{pmatrix} 9 & 1 \\ 13 & 2 \end{pmatrix}. \tag{4.19}$$

This parameterization satisfies eqs. (4.14) and (4.18), therefore this system has a globally stable, feasible coexistence equilibrium. Yet the interspecific coefficient A_{21} is larger in magnitude than the sum of all the other coefficients taken together. The intra- versus interspecific competition argument is valid only for the geometric averages as in eq. (4.18), not for the individual coefficients themselves.

Finally, note that an alternative form of the Lotka–Volterra equations, preferred by Ches-

son [35], reads

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = n_i b_i \left(1 + \sum_{j=1}^{S} \alpha_{ij} n_j\right), \tag{4.20}$$

where α_{ij} measures the effect of species j on species i's intrinsic growth rate b_i. eqs. (4.10) and (4.20) are both perfectly legitimate parameterizations, and can be transformed into one another by $A_{ij} = b_i \alpha_{ij}$. For $S = 2$, the feasibility condition eq. (4.14) reduces to $\alpha_{12}/\alpha_{22} < 1 < \alpha_{11}/\alpha_{21}$, implying $\alpha_{12} < \alpha_{22}$ and $\alpha_{21} < \alpha_{11}$. This yields a much more direct version of the principle that two species can only coexist if intraspecific effects exceed interspecific ones [35]. Unfortunately, this more direct interpretation of the two-species coexistence rule via eq. (4.20) is only available in the two-species case. For $S > 3$, violating the condition does not preclude, and vice versa: fulfilling it does not ensure coexistence. Below we construct examples to show this. To make sure that our conclusions will hold in both the parameterizations of eq. (4.10) and eq. (4.20), we will deliberately set $b_i = 1$ for all i, making the two parameterizations equivalent.

4.8.3 Three species

The only possible structurally stable dynamical behavior of the two-species competitive Lotka–Volterra model is the convergence to a stable fixed point. With three species, and still assuming purely competitive dynamics, limit cycles are also possible [99], but not chaos. Importantly, eq. (4.18), the familiar two-species condition between intra- and interspecific competition, no longer holds. In fact, one can construct counterexamples going both ways: competitive matrices where every pair of species satisfies the criterion and yet the system as a whole is unstable, and matrices where the criterion is violated yet coexistence is stable.

An example of the former is

$$\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{A} = -\begin{pmatrix} 10 & 9 & 5 \\ 9 & 10 & 9 \\ 5 & 9 & 10 \end{pmatrix}. \tag{4.21}$$

Since $\mathbf{A}$ is symmetric, its eigenvalues determine the global stability of coexistence [92], as discussed before. The eigenvalues are -5 and $(-25 \pm \sqrt{673})/2$, the largest of which is positive. Coexistence is therefore globally unstable: regardless of feasibility, the inevitable outcome is the extinction of at least one species (fig. 4.5, left panel). This is despite the fact that all species pairs individually satisfy the coexistence criteria, both in terms of eq. (4.10) and eq. (4.20), and so would stably coexist in the absence of the third species.

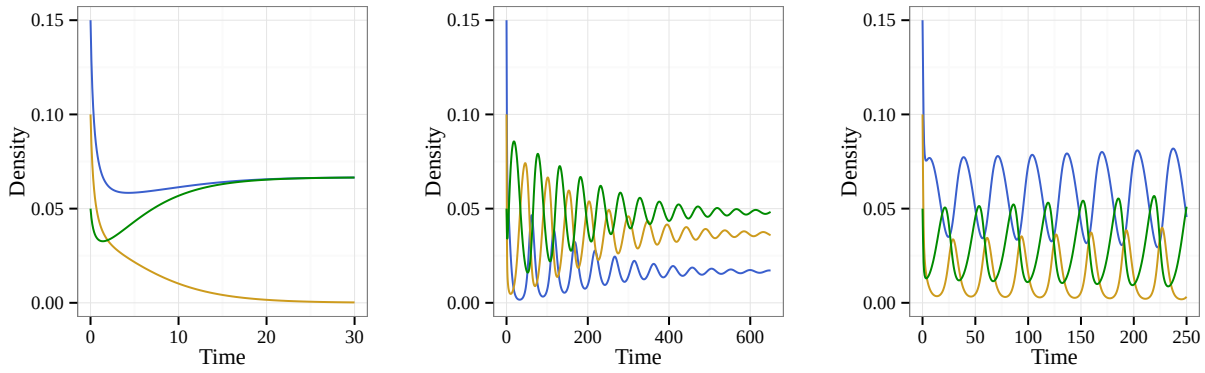


Figure 4.5: Dynamics of three-species competitive Lotka–Volterra equations. Left: intraspecific competition exceeds interspecific competition for all species pairs, yet there is no coexistence. Center: intraspecific competition does not exceed interspecific competition for all species pairs, yet coexistence is possible. Right: the coexistence equilibrium is unstable, but the species still coexist on a stable limit cycle.

Similarly, let us consider the following parameterization:

$$\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{A} = -\begin{pmatrix} 10 & 7 & 12 \\ 15 & 10 & 8 \\ 7 & 11 & 10 \end{pmatrix}. \tag{4.22}$$

Taking just species 1 and 2, the geometric mean of their interspecific competition coefficients exceeds that of the intraspecific ones. Also, $(b_2 A_{21})/(b_1 A_{11}) = \alpha_{21}/\alpha_{11} > 1$, violating two-species coexistence in the parameterization of eq. (4.20) as well. Yet the equilibrium densities are feasible: $\hat{\mathbf{n}} = -\mathbf{A}^{-1}\mathbf{b} = (10,\, 22,\, 29)/602$, and the eigenvalues of $\mathbf{J}|_{\hat{\mathbf{n}}} = \mathrm{diag}(\hat{\mathbf{n}})\mathbf{A}$ are -1 and $(-2 \pm \sqrt{-1}\sqrt{1591})/301$, all of which have negative real parts. The coexistence equilibrium is therefore feasible and at least locally stable (fig. 4.5, center panel).

Finally, an example of a stable limit cycle solution is given by

$$\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{A} = -\begin{pmatrix} 10 & 6 & 12 \\ 14 & 10 & 2 \\ 8 & 18 & 10 \end{pmatrix} \tag{4.23}$$

(fig. 4.5, right panel).

4.8.4 Four or more species

For four species, even purely competitive Lotka–Volterra dynamics may produce chaotic solutions. One example is given by

$$\mathbf{b} = \begin{pmatrix} 1.00 \\ 0.72 \\ 1.53 \\ 1.27 \end{pmatrix}, \quad \mathbf{A} = -\begin{pmatrix} 1.0000 & 1.0900 & 1.5200 & 0.0000 \\ 0.0000 & 0.7200 & 0.3168 & 0.9792 \\ 3.5649 & 0.0000 & 1.5300 & 0.7191 \\ 1.5367 & 0.6477 & 0.4445 & 1.2700 \end{pmatrix} \tag{4.24}$$

(after 152); the dynamics is shown on fig. 4.6. For five or more species, a classic theorem by Smale [130] establishes that one can construct the parameters of the competitive Lotka–Volterra equations to be compatible with *any* asymptotic dynamical behavior, i.e., the range of dynamics exhibited by these models is unrestricted.

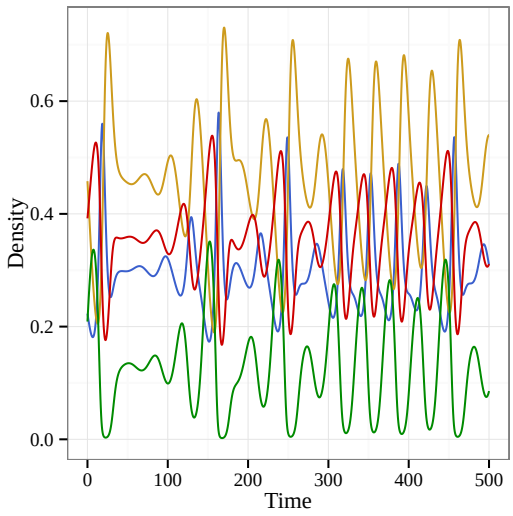


Figure 4.6: Chaos in a four-species competitive Lotka–Volterra system. The coexistence equilibrium is unstable, but the species still coexist on the stable chaotic attractor.

4.8.5 Multispecies generalization of the two-species coexistence rule

Although the two-species coexistence rule does not apply in communities of more than two species, a generalization can be given for competitive matrices $\mathbf{A}$ that are symmetric. This generalization is based on a simple corollary of Sylvester's criterion.

Sylvester's criterion for positive definite matrices [69] states that a real symmetric matrix $\mathbf{A}$ has positive eigenvalues if and only if all its principal minors are positive (a "principal minor" of a square matrix $\mathbf{A}$ is the determinant of a matrix obtained by deleting rows and columns of the same index from $\mathbf{A}$). Since $\mathbf{A}$ has positive eigenvalues if and only if $-\mathbf{A}$ has negative ones, an immediate consequence of the above is Sylvester's criterion for negative

definite matrices: a real symmetric matrix has negative eigenvalues if and only if all its odd-sized principal minors are negative and all its even-sized principal minors are positive.

The set of all principal minors includes every 2×2 principal minor. Therefore a weaker but simpler version of Sylvester's criterion for negative definite matrices states that all 2×2 principal minors must be positive for $\mathbf{A}$ to have negative eigenvalues; i.e., failing to fulfill this criterion guarantees some nonnegative eigenvalues, but fulfilling it does not guarantee all negative ones.

In ecological terms, a 2×2 principal minor is simply the determinant of the interaction matrix of a two-species subsystem, with all other species absent. Requiring all 2×2 principal minors to be positive is therefore equivalent to requiring that all species pairs independently satisfy eq. (4.18). In this way, Sylvester's criterion generalizes the two-species rule to multiple species: a community of S symmetrically competing species cannot be stable if there is even a single species pair violating the two-species coexistence rule.

4.9 Decomposition of the interaction matrix

We decompose the interaction matrix $\mathbf{A}$, with coefficients $A_{ij} \leq 0$, as the sum of two matrices $\mathbf{B}$ and $\mathbf{C}$, defined as

$$B_{ij} = \begin{cases} A_{ij} & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \tag{4.25}$$

$$C_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ A_{ij} & \text{if } i = j. \end{cases} \tag{4.26}$$

Thus, $\mathbf{B}$ contains the interspecific effects, $\mathbf{C}$ the intraspecific effects, and $\mathbf{A} = \mathbf{B} + \mathbf{C}$.

When $\mathbf{A}$ is symmetric (and therefore both $\mathbf{B}$ and $\mathbf{C}$ are symmetric), it is simple to write bounds for the eigenvalues of $\mathbf{A}$ in terms of the eigenvalues of $\mathbf{B}$ and $\mathbf{C}$, and thus illustrate the effects on the stability of the system using the familiar language of intra- and interspecific

competition.

The mathematical setting is as follows: take $\mathbf{B}$ and $\mathbf{C}$ symmetric, $S \times S$ matrices, and define $\mathbf{A} = \mathbf{B} + \mathbf{C}$. Because the matrices are symmetric, all eigenvalues are real. We number the eigenvalues in order: the eigenvalues of $\mathbf{A}$ are $\alpha_1 \geq \alpha_2 \geq \ldots \alpha_S$, those of $\mathbf{B}$ are $\beta_1 \geq \beta_2 \geq \ldots \beta_S$, and those of $\mathbf{C}$ are $\gamma_1 \geq \gamma_2 \geq \ldots \gamma_S$. Since every coefficient A_{ij} is negative or zero, it follows that all $\gamma_i \leq 0$, because the eigenvalues of $\mathbf{C}$ are simply the coefficients on the diagonal of $\mathbf{A}$ sorted in decreasing order. Also, since $\mathbf{B}$ has all zeros along its diagonal, $\beta_1 \geq 0$ and $\beta_S \leq 0$. This is because the sum of the diagonal entries (zero in our case) is also the sum of all eigenvalues, so either every eigenvalue is zero, or some are positive and some are negative. For stability, we need $\alpha_1 < 0$.

When the matrices are symmetric, we can write the following inequalities [69]:

$$\max_{i+j=S+k} \beta_i + \gamma_j \leq \alpha_k \leq \min_{i+j=k+1} \beta_i + \gamma_j, \tag{4.27}$$

where $k = 1, \ldots, S$. Setting $k = 1$ (i.e., bounding α_1, the rightmost eigenvalue of $\mathbf{A}$, determining stability), we can simplify the inequalities:

$$\max_j \beta_j + \gamma_{S-j+1} \leq \alpha_1 \leq \beta_1 + \gamma_1, \tag{4.28}$$

which implies

$$\beta_1 + \gamma_S \leq \alpha_1 \leq \beta_1 + \gamma_1. \tag{4.29}$$

These relationships allow us to write a *necessary* condition for stability:

$$\beta_1 + \gamma_S < 0, \tag{4.30}$$

meaning that the strongest self-regulation $\gamma_S = \min A_{ii}$ needs to be strong enough to offset the rightmost eigenvalue β_1 stemming from interspecific interactions. Similarly, we can write

a *sufficient* condition for stability:

$$\beta_1 + \gamma_1 < 0, \tag{4.31}$$

meaning that the weakest self-regulation $\gamma_1 = \max A_{ii}$ needs to be strong enough to offset the rightmost eigenvalue β_1 stemming from interspecific interactions. For instance, in the two-species case, β_1 is simply equal to $-A_{12} = -A_{21}$. Assuming $A_{11} < A_{22}$ without loss of generality, $A_{11} < A_{12}$ is a necessary, while $A_{22} < A_{12}$ is a sufficient condition for two-species coexistence—consistently with the precise condition $A_{11}A_{22} > A_{12}^2$ coming from eq. (4.18).

Clearly, when the diagonal entries of $\mathbf{A}$ are all equal, the two conditions are the same, providing a *necessary and sufficient* condition for stability.

How sharp are these bounds? Because we are foregoing the exact matrices $\mathbf{B}$ and $\mathbf{C}$ and only consider their eigenvalues, the bounds cannot be sharp in all cases. In fact, there are very many possible matrices having exactly the same eigenvalues (isospectral matrices). In particular, when we consider only the eigenvalues of $\mathbf{C}$, we forego the precise identity of the species, so that the $S!$ different ways of arranging the coefficients C_{ii} on the diagonal would yield exactly the same eigenvalues γ_i, but would alter the eigenvalues of $\mathbf{A}$. This subtle point is better understood through a simple example.

Take a given matrix $\mathbf{B}$, and suppose that the diagonal matrix $\mathbf{C}$ has eigenvalues $\gamma_1 = -2$, $\gamma_2 = -9$, and $\gamma_3 = -10$. There are $3! = 6$ matrices $\mathbf{C}$ with the same exact eigenvalues. However, each possible $\mathbf{C}$ will yield a different $\mathbf{A} = \mathbf{B} + \mathbf{C}$ matrix, and thus different eigenvalues α_i.

Table 4.1 shows that for the very same eigenvalues of $\mathbf{B}$ and $\mathbf{C}$, we can obtain α_1 as low as -0.928, or as high as 1.057. This is reflected in the bounds introduced above:

$$\beta_1 + \gamma_S = 5.429 - 10 = -4.571 \leq \alpha_1 \leq \beta_1 + \gamma_1 = 5.429 - 2 = 3.429. \tag{4.32}$$

C, **B**, and **A**										Eigenvalues	$\alpha_1 < 0$
$\begin{bmatrix} -2 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & -10 \end{bmatrix}$			$\begin{bmatrix} 0 & -3 & -1 \\ -3 & 0 & -5 \\ -1 & -5 & 0 \end{bmatrix}$			$\begin{bmatrix} -2 & -3 & -1 \\ -3 & -9 & -5 \\ -1 & -5 & -10 \end{bmatrix}$			$\gamma_1 = -2, \gamma_2 = -9, \gamma_3 = -10$ $\beta_1 = 5.429, \beta_2 = 0.876, \beta_3 = -6.305$ $\alpha_1 = -0.792, \alpha_2 = -5.094, \alpha_3 = -15.114$	Yes
$\begin{bmatrix} -2 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0 & -9 \end{bmatrix}$			$\begin{bmatrix} 0 & -3 & -1 \\ -3 & 0 & -5 \\ -1 & -5 & 0 \end{bmatrix}$			$\begin{bmatrix} -2 & -3 & -1 \\ -3 & -10 & -5 \\ -1 & -5 & -9 \end{bmatrix}$			$\gamma_1 = -2, \gamma_2 = -9, \gamma_3 = -10$ $\beta_1 = 5.429, \beta_2 = 0.876, \beta_3 = -6.305$ $\alpha_1 = -0.928, \alpha_2 = -4.903, \alpha_3 = -15.169$	Yes
$\begin{bmatrix} -9 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -10 \end{bmatrix}$			$\begin{bmatrix} 0 & -3 & -1 \\ -3 & 0 & -5 \\ -1 & -5 & 0 \end{bmatrix}$			$\begin{bmatrix} -9 & -3 & -1 \\ -3 & -2 & -5 \\ -1 & -5 & -10 \end{bmatrix}$			$\gamma_1 = -2, \gamma_2 = -9, \gamma_3 = -10$ $\beta_1 = 5.429, \beta_2 = 0.876, \beta_3 = -6.305$ $\alpha_1 = 0.941, \alpha_2 = -8.382, \alpha_3 = -13.559$	No
$\begin{bmatrix} -9 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0 & -2 \end{bmatrix}$			$\begin{bmatrix} 0 & -3 & -1 \\ -3 & 0 & -5 \\ -1 & -5 & 0 \end{bmatrix}$			$\begin{bmatrix} -9 & -3 & -1 \\ -3 & -10 & -5 \\ -1 & -5 & -2 \end{bmatrix}$			$\gamma_1 = -2, \gamma_2 = -9, \gamma_3 = -10$ $\beta_1 = 5.429, \beta_2 = 0.876, \beta_3 = -6.305$ $\alpha_1 = 0.421, \alpha_2 = -7.148, \alpha_3 = -14.273$	No
$\begin{bmatrix} -10 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -9 \end{bmatrix}$			$\begin{bmatrix} 0 & -3 & -1 \\ -3 & 0 & -5 \\ -1 & -5 & 0 \end{bmatrix}$			$\begin{bmatrix} -10 & -3 & -1 \\ -3 & -2 & -5 \\ -1 & -5 & -9 \end{bmatrix}$			$\gamma_1 = -2, \gamma_2 = -9, \gamma_3 = -10$ $\beta_1 = 5.429, \beta_2 = 0.876, \beta_3 = -6.305$ $\alpha_1 = 1.057, \alpha_2 = -8.725, \alpha_3 = -13.333$	No
$\begin{bmatrix} -10 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & -2 \end{bmatrix}$			$\begin{bmatrix} 0 & -3 & -1 \\ -3 & 0 & -5 \\ -1 & -5 & 0 \end{bmatrix}$			$\begin{bmatrix} -10 & -3 & -1 \\ -3 & -9 & -5 \\ -1 & -5 & -2 \end{bmatrix}$			$\gamma_1 = -2, \gamma_2 = -9, \gamma_3 = -10$ $\beta_1 = 5.429, \beta_2 = 0.876, \beta_3 = -6.305$ $\alpha_1 = 0.629, \alpha_2 = -7.597, \alpha_3 = -14.032$	No

Table 4.1: Eigenvalues of $\mathbf{C} + \mathbf{B} = \mathbf{A}$, as a function of the ordering of the diagonal entries of $\mathbf{C}$. The same set of coefficients produce different eigenvalues and stability properties for $\mathbf{A}$ depending on the ordering, despite the fact that the eigenvalues of the two constituent matrices $\mathbf{C}$ and $\mathbf{B}$ are unchanged by this rearrangement.

4.10 Random matrices

4.10.1 Symmetric matrices

Let $\mathbf{A}$ be an $S \times S$ matrix, constructed as follows. Each diagonal entry is set to zero. Then each upper triangular entry is drawn from some probability distribution with mean zero, variance V, and all moments finite. Finally, each lower triangular entry A_{ij} is set equal to A_{ji}, making $\mathbf{A}$ symmetric.

Matrices constructed this way are called *Wigner matrices*. Consider now the matrix $\mathbf{A}/\sqrt{SV}$. Since it is symmetric, all eigenvalues are real. Their empirical spectral distribution, in the limit of large S, follows the *Wigner semicircle distribution*:

$$\tilde{f}(x) = \frac{\sqrt{4 - x^2}}{2\pi}, \quad -2 \leq x \leq 2 \tag{4.33}$$

and 0 otherwise, where x represents the position along the real line where the eigenvalues may fall [5, 13]. The most important property of this distribution is its *universality*: any

underlying distribution from which the matrix entries are drawn will lead to $\tilde{f}(x)$ as $S \to \infty$; the only constraint is a zero mean and finite higher moments. For finite S, one can approximate the above distribution for $\mathbf{A}$ instead of $\mathbf{A}/\sqrt{SV}$ as

$$f(x) \approx \frac{\sqrt{4SV - x^2}}{2\pi SV}, \quad -2\sqrt{SV} \leq x \leq 2\sqrt{SV} \tag{4.34}$$

and 0 otherwise, where the approximation improves with increasing S. See the left panel of fig. 4.7 for an example.

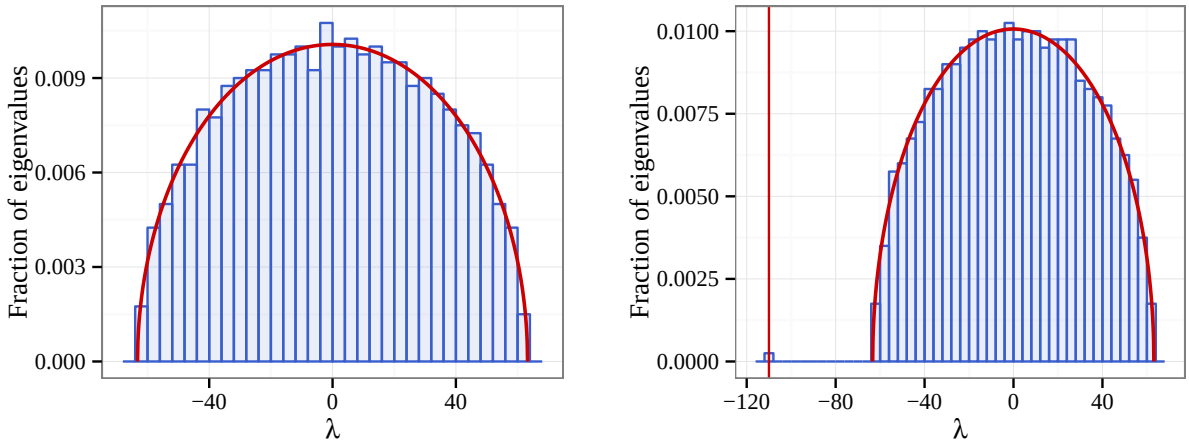


Figure 4.7: Left panel: eigenvalue distribution (blue histogram) of a Wigner random matrix with $S = 1000$, $V = 1$, and the entries drawn from a normal distribution. The red curve is the analytical prediction of eq. (4.34). Right panel: eigenvalue distribution (blue histogram) of a symmetric random matrix with $S = 1000$, $\mu = -1/10$, $V = 1$, $\mu_d = 0$, $V_d = 1$, and all entries drawn from a normal distribution. The red curve is the analytical prediction based on eq. (4.39); the single outlier eigenvalue is predicted to be at $S\mu + V/\mu = -110$. Notice also that the variance of the diagonal entries is as large as the offdiagonal variance, yet this has no discernible effect on the eigenvalue distribution (the red semicircle's prediction is still almost perfect).

The requirement that all diagonal entries are zero can be relaxed. Assume the diagonal entries have mean zero and variance V_d. By definition,

$$V_d = \frac{1}{S}\sum_{i=1}^{S}(A^2)_{ii}, \tag{4.35}$$

where $(A^2)_{ij}$ is the (i,j)th entry of $\mathbf{AA} = \mathbf{A}^2$. The variance of all matrix entries may then be written

$$\frac{1}{S^2}\sum_{i=1}^{S}\sum_{j=1}^{S}(A^2)_{ij} = \frac{1}{S^2}\sum_{i=1}^{S}(A^2)_{ii} + \frac{1}{S(S-1)}\sum_{i=1}^{S}\sum_{j\neq i}^{S}(A^2)_{ij} = \frac{V_d}{S} + V. \tag{4.36}$$

In other words, as long as V_d does not scale with S for some reason, the contribution of the diagonal entries' variance to the total variance will be negligible in the large S limit, and only the offdiagonal variance V matters. Adding diagonal variance to a large Wigner matrix therefore leaves its spectral properties unchanged.

What happens if the diagonal entries have a nonzero mean, say, μ_d? (The mean of the offdiagonal entries is still assumed to be zero.) Let $\mathbf{I}$ be the $S \times S$ identity matrix; $\mathbf{A}$ can then be written as $\mathbf{A} = (\mathbf{A} - \mu_d\mathbf{I}) + \mu_d\mathbf{I}$. The matrix $(\mathbf{A} - \mu_d\mathbf{I})$ has zero mean diagonal, therefore it is a Wigner matrix for which our previous results hold. When adding the diagonal matrix $\mu_d\mathbf{I}$ to this, μ_d is simply added to each eigenvalue.[3] The eigenvalue distribution in the large S limit therefore reads

$$f(x) = \frac{\sqrt{4SV - (x-\mu_d)^2}}{2\pi SV}, \quad -2\sqrt{SV} \leq x - \mu_d \leq 2\sqrt{SV} \tag{4.37}$$

and 0 otherwise.

Finally, one can consider the case when the offdiagonal entries also have a nonzero mean μ. Let $\mathbf{E}$ be the $S \times S$ matrix with all entries equal to one; then we can write

$$\mathbf{A} = (\mathbf{A} - \mu_d\mathbf{I} - \mu\mathbf{E} + \mu\mathbf{I}) + \mu_d\mathbf{I} + \mu\mathbf{E} - \mu\mathbf{I}. \tag{4.38}$$

In the parentheses, we first shift the diagonal to have zero mean, then we shift the matrix

3. The eigenvalues of $\mathbf{A}$ are, by definition, numbers λ satisfying $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. The eigenvalues of $\mathbf{A} + \mu_d\mathbf{I}$ are numbers λ' satisfying $\det(\mathbf{A} + \mu_d\mathbf{I} - \lambda'\mathbf{I}) = \det(\mathbf{A} - (\lambda' - \mu_d)\mathbf{I}) = 0$. Therefore $\lambda = \lambda' - \mu_d$, or $\lambda' = \lambda + \mu_d$.

so that its offdiagonal entries have zero mean, but in the process also affect the diagonal—the final term restores it back to have zero mean. The matrix in parentheses is therefore a Wigner matrix, its eigenvalue distribution given by the Wigner semicircle distribution. The matrix $\mu_d\mathbf{I} - \mu\mathbf{I} = (\mu_d - \mu)\mathbf{I}$ simply adds $\mu_d - \mu$ to each eigenvalue. Finally, note that $\mathbf{E}$ can be written as the outer product of the vector (1, 1, ..., 1) (a vector of all ones) with itself—that is, $\mathbf{E}$ is a rank-one matrix. According to the theory of low-rank perturbations to random matrices [115], the effect of adding $\mu\mathbf{E}$ is to make one single eigenvalue assume the value $S\mu + V/\mu$ (for S large, the second term is negligible).[4] The shape of the rest of the eigenvalue distribution remains unchanged. The final form of the empirical spectral distribution for a symmetric random matrix with offdiagonal mean and variance μ, V and diagonal mean and variance μ_d, V_d therefore reads

$$f(x) = \frac{\sqrt{4SV - (x - \mu_d + \mu)^2}}{2\pi SV}, \quad -2\sqrt{SV} \leq x - \mu_d + \mu \leq 2\sqrt{SV} \tag{4.39}$$

and 0 otherwise, except for a single outlier eigenvalue equal to $S\mu + V/\mu$ (with the V/μ term missing if μ is too small). See the right panel of fig. 4.7 for an example.

4.10.2 Nonsymmetric matrices

The Wigner semicircle distribution holds only for symmetric random matrices. More generally, one may consider the nonsymmetric random matrix $\mathbf{A}$, constructed as follows. First, each diagonal entry is set to zero. Then each pair of entries (A_{ij}, A_{ji}) are sampled from a bivariate distribution with zero marginal means, marginal variances $V < \infty$, and pairwise correlation ρ. All moments of the bivariate distribution are assumed finite.

Matrices constructed this way are called *elliptic matrices*. According to the *elliptic law*

4. Technical conditions hold on whether the V/μ term is actually present (see 115); in essence, this term will be there as long as μ is large enough to make the single modified eigenvalue isolated from the rest of the distribution. See the right panel of fig. 4.7 for an example.

[115, 133], the eigenvalues of $\mathbf{A}/\sqrt{SV}$ are uniformly distributed in an ellipse in the complex plane, with horizontal and vertical semiaxes $1 \pm \rho$. For finite but large S, the eigenvalues of $\mathbf{A}$ are distributed uniformly in an ellipse with horizontal/vertical semiaxes $\sqrt{SV}(1 \pm \rho)$ (fig. 4.8, left panel). Just as in the symmetric case, the elliptic law is universal: *any* bivariate distribution with the same marginal variances and correlation will result in the eigenvalues being distributed in the same ellipse.

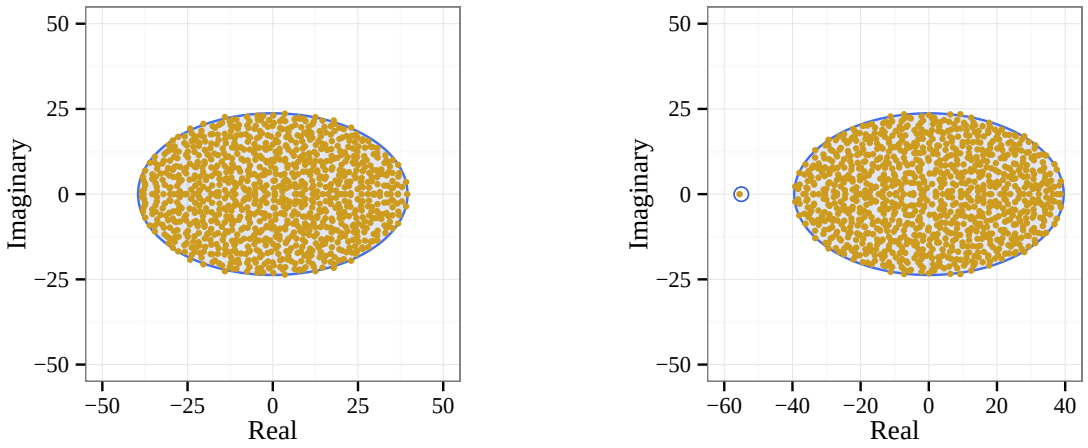


Figure 4.8: Left panel: eigenvalues (yellow points) of an elliptic random matrix with $S = 1000$, and all pairs (A_{ij}, A_{ji}) sampled from a bivariate normal distribution with $\mu = 0$, $V = 1$, and $\rho = 1/4$. The blue ellipse is the analytical prediction for the region where the eigenvalues are uniformly distributed. Right panel: eigenvalues (yellow points) of an elliptic random matrix with $S = 1000$, and all pairs (A_{ij}, A_{ji}) sampled from a bivariate normal distribution with $\mu = -1/20$, $V = 1$, and $\rho = 1/4$. The blue ellipse and circle are the analytical prediction; the single outlier eigenvalue is predicted to be at $S\mu + \rho V/\mu = -55$.

The generalizations that we performed in the symmetric case carry over to the nonsymmetric case as well, as there is nothing in these generalizations that is specific to Wigner matrices. Therefore, if the bivariate distribution has mean μ, and the diagonal entries have mean μ_d and variance V_d, the eigenvalues will be uniformly distributed in an ellipse with horizontal/vertical semiaxes $\sqrt{SV}(1 \pm \rho)$ centered at $(\mu_d - \mu, 0)$ in the complex plane, and there will be a single outlying eigenvalue at $S\mu + \rho V/\mu$ (the second term is zero if μ is too small). See the right panel of fig. 4.8 for an example. Finally, note that Wigner matrices are

really just a special case of elliptic matrices: a Wigner matrix is an elliptic one with $\rho = 1$.

4.11 Maximizing/minimizing the rightmost eigenvalue

Here we describe the algorithm we used to find the arrangement of interspecific interaction strengths that optimized (either maximized or minimized) the largest (rightmost) eigenvalue of $\mathbf{B}$. (Since all eigenvalues are real, the rightmost one is necessarily the largest in value, though not necessarily in magnitude.) We used the same algorithm to find the arrangement of intraspecific effects in $\mathbf{C}$ maximizing/minimizing the rightmost eigenvalue of $\mathbf{A} = \mathbf{B} + \mathbf{C}$ for a fixed $\mathbf{B}$, with some minor differences in parameterization described below.

For optimizing the rightmost eigenvalue of $\mathbf{B}$, we used a population of 5000 matrices in any one generation of a genetic algorithm. For reproduction, first 4900 pairs of matrices were drawn randomly with replacement. From each pair we picked the matrix that had the more optimal rightmost eigenvalue, and put that matrix into the next generation. The remaining 100 matrices were direct copies of the most optimal matrix of the previous generation (elite reproduction). After reproduction, each matrix in the next generation was mutated by swapping two entries in the upper triangular part, and the corresponding two in the lower triangle to keep the matrix symmetric. The algorithm was run for $500,000$ generations, after which the matrix with the most optimal largest eigenvalue was taken to be the candidate for the final solution. We then ran a hill-climbing algorithm starting from this matrix to explore the matrices in the direct vicinity of our candidate solution. We generated 5000 mutant matrices, and only the most optimal one of the 5000 was kept at every generation in the algorithm. We repeated this until no better solutions were found in 5000 further generations. The most optimal matrix we found via the hill-climber was then taken as our final solution.

We did the same when finding the optimal largest eigenvalue of $\mathbf{A}$ as a function of the arrangement of its diagonal part $\mathbf{C}$ (with $\mathbf{B}$ fixed), except that we ran the genetic algorithm only for $50,000$ generations, and that a mutation event was defined by swapping two randomly

chosen diagonal entries. Just as before, we ran a hill-climbing algorithm on the best solution found by the genetic algorithm to try to further improve on that solution. We did this in the same way as in the offdiagonal case, except, instead of generating only 5000 mutants, we generated all possible mutations in every generation of the hill-climber. Because of this, the algorithm could be stopped the moment no better solutions were found in a generation, instead of waiting for 5000 further steps without improvement.

For each optimization, both diagonal and offdiagonal, we performed 10 separate optimization runs, each with a different random seed, to make sure they ended up with the same results.

The source code and documentation for an implementation of this algorithm, written in C, can be downloaded from `https://github.com/dysordys/intra-inter`.

4.12 Explaining the pattern of intraspecific interactions leading to the largest and smallest rightmost eigenvalue

Here we show that the patterns of intraspecific coefficients on fig. 4.1 of the main text are generic and expected to hold in all cases, except when the variance V_d of the diagonal entries is large. As explained in section 4.9, $\mathbf{A}$ is the interaction matrix, decomposed as $\mathbf{A} = \mathbf{B} + \mathbf{C}$, where $\mathbf{B}$ is the matrix of interspecific and $\mathbf{C}$ the diagonal matrix of intraspecific effects. The rightmost eigenvalue of $\mathbf{A}$ is α_1, and that of $\mathbf{B}$ is β_1. Summarizing the patterns of $\mathbf{C}$'s diagonal entries to be explained (fig. 4.1 in main text):

1. ($\max \beta_1$, $\max \alpha_1$): The strength (absolute value) of the diagonal entries first increases, then decreases, approximately forming a V-shape.

2. ($\max \beta_1$, $\min \alpha_1$): The strength of the diagonal entries first decreases, then increases, approximately forming an upside-down V-shape.

3. (random β_1, $\max \alpha_1$): The strength of the diagonal entries first increases, then de-

creases, approximately forming a V-shape. This V-shape is less pronounced, "fuzzier", than in the maximum β_1 case.

4. (random β_1, min α_1): The strength of the diagonal entries first decreases, then increases, approximately forming an upside-down V-shape. This upside-down V-shape is less pronounced than in the maximum β_1 case.

5. (min β_1, max α_1): The strength of the diagonal entries increases monotonically with the variance of the corresponding species' interspecific coefficients. That is, species 1 will have the strongest and species S the weakest intraspecific coefficient, provided that the variance of the offdiagonal entries is largest in the first row and smallest in the Sth.

6. (min β_1, min α_1): The strength of the diagonal entries roughly decreases with the variance of the corresponding species' interspecific coefficients, but this decrease is not necessarily strict or monotonic.

These patterns can be explained via standard eigenvalue sensitivity theory. We have $\mathbf{A} = \mathbf{B} + \mathbf{C}$. Let us write $\mathbf{C}$ as the mean intraspecific interaction μ_d plus the deviation from the mean: $\mathbf{C} = (\mathbf{C} - \mu_d\mathbf{I}) + \mu_d\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. With the notation $\Delta\mathbf{C} = \mathbf{C} - \mu_d\mathbf{I}$, we have $\mathbf{A} = \mathbf{B} + \mu_d\mathbf{I} + \Delta\mathbf{C}$. The eigenvalues of $\mathbf{B} + \mu_d\mathbf{I}$ are the eigenvalues of $\mathbf{B}$ plus μ_d (see section 4.10), and the eigenvectors remain unchanged. We therefore only need to determine the effect of $\Delta\mathbf{C}$ to obtain α_1. This we can do, to a first-order approximation, by writing the eigenvalues of $\mathbf{A}$ as those of $\mathbf{B} + \mu_d\mathbf{I}$ plus a correction imposed by the perturbing matrix $\Delta\mathbf{C}$:

$$\alpha_1 \approx (\beta_1 + \mu_d) + \mathbf{w}\,\Delta\mathbf{C}\,\mathbf{w}, \tag{4.40}$$

where $\mathbf{w}$ is the eigenvector corresponding to α_1. Here we used the general sensitivity formula for eigenvalues (e.g., 32, chapter 9); note that, because $\mathbf{B}$ is symmetric, the left and right eigenvectors are identical. Since $\mathbf{C}$ is diagonal, the product $\mathbf{w}\,\Delta\mathbf{C}\,\mathbf{w}$ can be written $w_1^2\Delta C_{11} +$

$w_2^2 \Delta C_{22} + \cdots + w_S^2 \Delta C_{SS}$, so we have

$$\alpha_1 \approx (\beta_1 + \mu_d) + \sum_{i=1}^{S} w_i^2 \Delta C_{ii}. \tag{4.41}$$

Therefore, maximizing/minimizing α_1 is equivalent to maximizing/minimizing the sum on the right hand side. The coefficients w_i^2 are always nonnegative while the C_{ii} are both negative and positive (since they are the deviations from the mean intraspecific interaction strength). Maximization (minimization) of the sum is therefore achieved by pairing up large values of w_i^2 with large (small) entries of ΔC_{ii}, and small values of w_i^2 with small (large) entries of ΔC_{ii}.

In the maximum β_1 case (1. and 2. above), species are sorted in increasing order of the components of the eigenvector $\mathbf{w}$ corresponding to the rightmost eigenvalue β_1 of $\mathbf{B}$. It is therefore both the vector corresponding to the eigenvalue to be perturbed and also the one based on which the species are ordered. Due to this fact, we have $w_1 \leq w_2 \leq \ldots \leq w_S$. Note that $\mathbf{w}$ is the eigenvector belonging to the rightmost and not the dominant eigenvalue; and since only the dominant eigenvector can have components with all the same sign, $\mathbf{w}$ will necessarily have both positive and negative components (a simple corollary of the Perron–Frobenius theorem). Because of this, it is *not* true that $w_1^2 \leq w_2^2 \leq \ldots \leq w_S^2$. Instead, the w_i^2 first decrease, then increase as i runs from 1 to S, forming a V-shape (fig. 4.9, top row). Then, to maximize α_1, the largest ΔC_{ii} should be paired with the largest w_i^2, leading to a diagonal that also follows a V-shape—which is precisely what we see in the cases of maximizing α_1 with either a random or maximal β_1. In turn, α_1 is minimized when the smallest ΔC_{ii} are paired up with the largest w_i^2, leading to the upside-down V-shape.

We still see these arrangements in the random cases (3. and 4. above), though they are less pronounced than in the maximum β_1 case. Since the matrix $\mathbf{B}$ is random, one might wonder why we see a pattern at all. The reason is that species are still ordered based on the leading eigenvector. Therefore, despite the randomness, matrix entries tend to increase

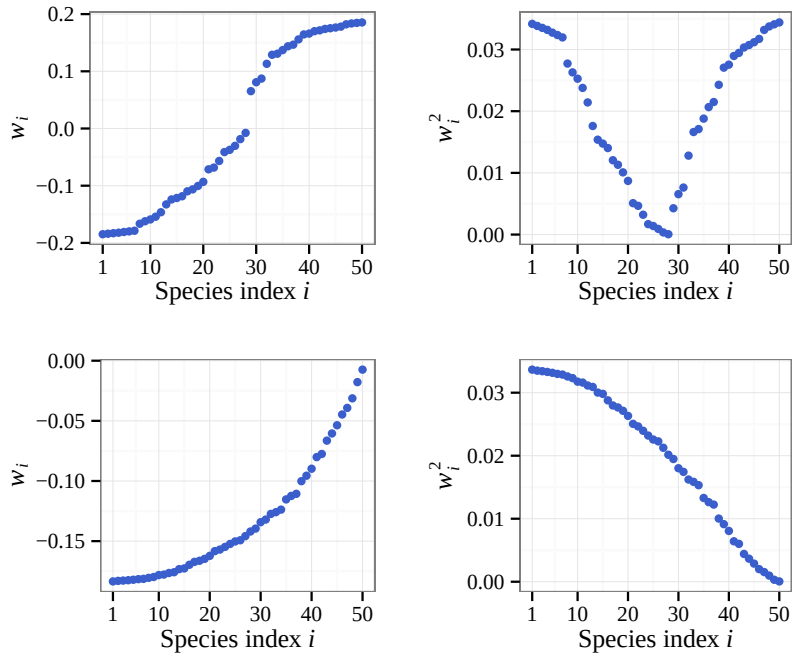


Figure 4.9: Eigenvectors' components sorted in increasing order, and the squared components in that same order. Top: the leading eigenvector has both negative and positive entries. The squared components will therefore first decrease, then increase. Bottom: the dominant eigenvector can be chosen all-negative by the Perron–Frobenius theorem. Its squared components therefore monotonically decrease.

towards the upper-right and lower-left corners of the matrices, though this increase is no longer strictly monotonic. This means that the broad pattern is retained, but it is expected to be fuzzier than before, which is exactly what we see.

Finally, in the minimum β_1 cases (5. and 6. above), species are ordered not by the leading but by the dominant eigenvector (i.e., the one belonging to the eigenvalue that is largest in magnitude, not in value). Since the matrix $\mathbf{A}$ is negative, the Perron–Frobenius theorem applies to $-\mathbf{A}$, and so the leading eigenvector can be chose to have all negative components. When ordering species based on it, we have $w_1 \leq w_2 \leq \ldots \leq w_S$, and since all w_i are negative, it follows that $w_1^2 \geq w_2^2 \geq \ldots \geq w_S^2$ (fig. 4.9, bottom row). Therefore, were our goal to maximize (minimize) this eigenvalue instead of the leading one, ΔC_{ii} would have to be arranged in decreasing (increasing) order. When maximizing/minimizing the dominant

eigenvalue, keep in mind that the mean and the variance of $\mathbf{A}$'s eigenvalue distribution does not change by swapping diagonal entries. Therefore, maximizing (minimizing) this eigenvalue will cause all the other $S-1$ eigenvalues to a) shift very slightly in the negative (positive) direction to retain the mean, and b) to reduce (increase) their spread to retain the variance. The shift in the mean is going to be negligible for S sufficiently large, therefore the change in the spread of the $S-1$ other eigenvalues will determine whether the leading (not the dominant!) eigenvalue will increase or decrease. It decreases when their spread decreases, and vice versa. Therefore, minimizing the dominant eigenvalue is expected to also minimize the leading eigenvalue, and conversely, maximizing the leading eigenvalue will maximize the dominant one, leading to the observed patterns.

In the random and minimum β_1 cases, there is an extra complication when minimizing α_1. Since many eigenvalues have very similar values, minimizing the leading eigenvalue may cause it to cross other eigenvalues. It therefore ceases to be the leading eigenvalue. To truly minimize the rightmost eigenvalue, an appropriate linear combination of eigenvectors must be considered, which may confound the simple patterns described above. This is seen in fig. 4.1 of the main text: for minimum β_1 and α_1, the decrease in the magnitude of interaction strengths is not quite monotonic.

As a final note, the argument of this section relied on the linear eigenvalue perturbation formula, eq. (4.40). Thus, one expects it to hold if the entries of the perturbing matrix $\mathbf{\Delta C}$ are not too large in magnitude compared to those of $\mathbf{B}+\mu_d\mathbf{I}$. Deviations from the expected pattern may happen if $\mathbf{\Delta C}$ has very strong entries, i.e., if the diagonal variance V_d is too large. However, even in those cases, the general shapes predicted by our argument hold up (see (figs. 4.10 to 4.29).

4.13 Calculating the domain of feasibility

The coexistence equilibrium $\hat{\mathbf{n}} = -\mathbf{A}^{-1}\mathbf{b}$ of the Lotka–Volterra model is feasible if all equilibrium densities are positive (section 4.7). For a given interaction matrix $\mathbf{A}$, the direction of $\mathbf{b}$ determines feasibility but not its magnitude, because if $\hat{\mathbf{n}} = -\mathbf{A}^{-1}\mathbf{b} > \mathbf{0}$, then so is $\eta\hat{\mathbf{n}} = -\mathbf{A}^{-1}(\eta\mathbf{b})$ for an arbitrary positive constant η (section 4.8). To evaluate feasibility, we can therefore determine the number of directions of $\mathbf{b}$ that lead to a feasible solution, out of all possible directions. We call this quantity Ξ:

$$\Xi = \frac{(\text{number of directions of } \mathbf{b} \text{ leading to a feasible coexistence equilibrium})}{(\text{all possible directions of } \mathbf{b} \text{ in the positive orthant})} \,. \qquad (4.42)$$

Restricting the possible directions of $\mathbf{b}$ to the positive orthant is equivalent to assuming that all intrinsic growth rates are positive. Since the system is competitive, no coexistence is possible without this assumption in the first place. One can evaluate Ξ analytically for two (section 4.8.2) and three [140, pp. 203-204] species. For more than three species, no known closed-form solutions exist. In lieu of an analytical formula, Grilli *et al.* [62] provide an accurate and efficient method for calculating Ξ by numerically evaluating the integral

$$\Xi = \frac{2^S\ \Gamma(S/2)\ |\det(\mathbf{A})|}{2\pi^{S/2}} \int_{\Omega^+} \left(\sum_{i=1}^{S} \sum_{j=1}^{S} \sum_{k=1}^{S} b_i A_{ki} A_{kj} b_j \right)^{-S/2} \mathrm{d}b_1\, \mathrm{d}b_2 \cdots \mathrm{d}b_S. \qquad (4.43)$$

Here S is the number of species, $\Gamma(\cdot)$ is the Gamma function, $\det(\mathbf{A})$ is $\mathbf{A}$'s determinant, Ω^+ is the part of the S-dimensional unit sphere's surface that falls in the positive orthant, $b_1, \ldots, b_S$ are the components of $\mathbf{b}$ normalized so that the vector has length 1 (i.e., $\mathbf{b}$ always lies on the unit sphere's surface within the positive orthant), and A_{ij} is the (i,j)th entry of $\mathbf{A}$. To obtain our feasibility results, we numerically evaluated this integral via the Monte

Carlo method.

An increasing number of species S has a trivial dimensional effect on the fraction of feasible directions Ξ. If half the total angle per direction in an S-species system is feasible, then by eq. (4.42) we get $\Xi = 2^{-S}$. But, although the total fraction of feasible directions is decreasing, the fraction of feasible directions *per dimension* stays the same, 1/2. For this reason, we use $\sqrt[S]{\Xi}$ instead of Ξ to quantify feasibility, which is a measure of the (geometric) average fraction of feasible directions per dimension. It also has the advantage of allowing for the comparison of systems with different numbers of species S.

Supplementary Figures

The figures below (figs. 4.10 to 4.29) show that the results reported in the main text are robust across various parameterizations. The first three of each parameterization read exactly like (figs. 4.1 to 4.3) in the main text, except for the set of intra- and interspecific coefficients, which we report in the figure captions. The additional two figures of each parameterization show the relationship of the leading eigenvalue of the most/least stabilized configuration of coefficients with respect to a distribution generated by randomly assorting those coefficients.

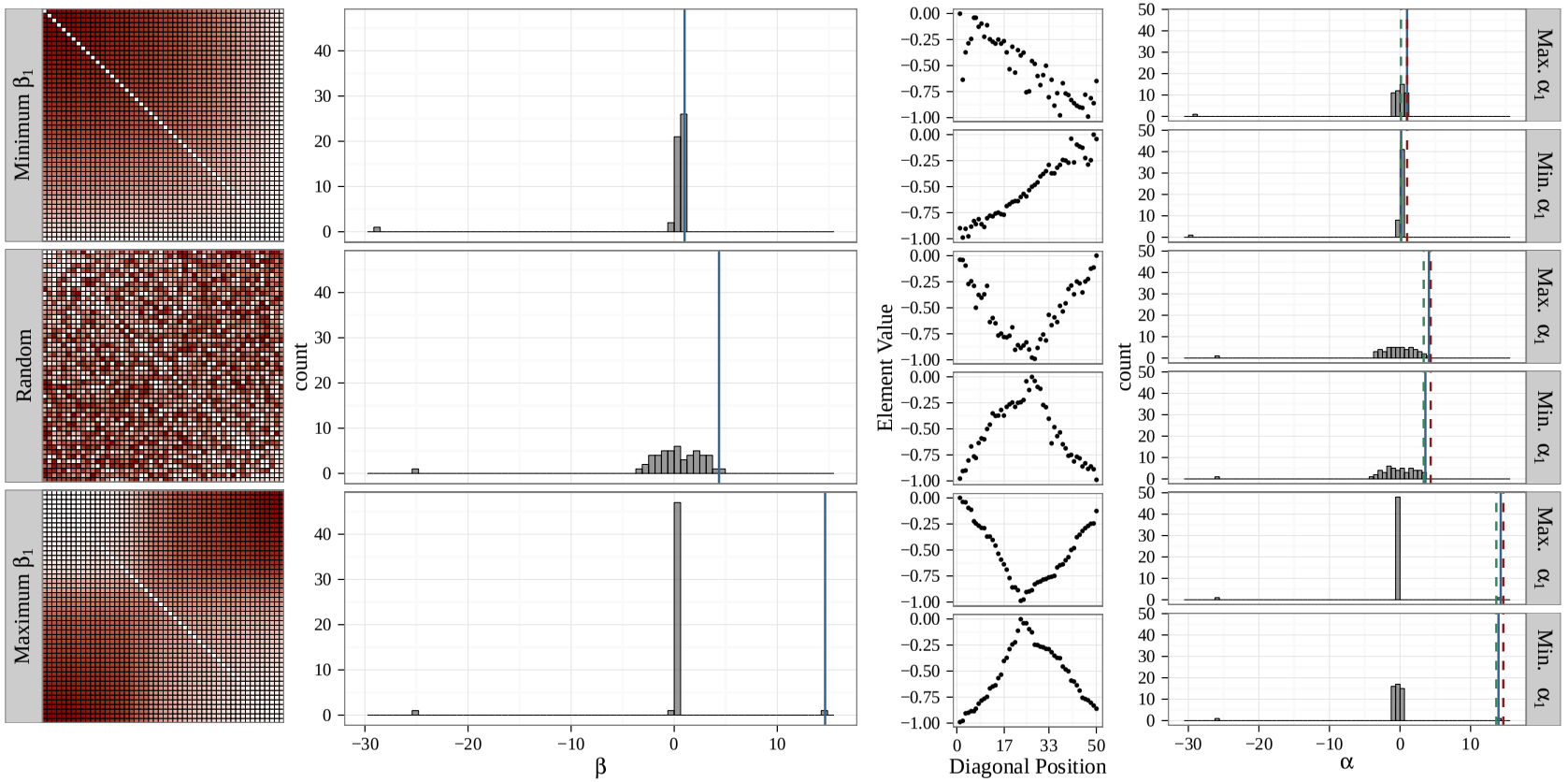


Figure 4.10: As fig. 4.1 in the main text, except the interspecific competition coefficients are uniformly sampled from $[-1, 0]$, and the intraspecific coefficients are also uniformly sampled from $[-1, 0]$.

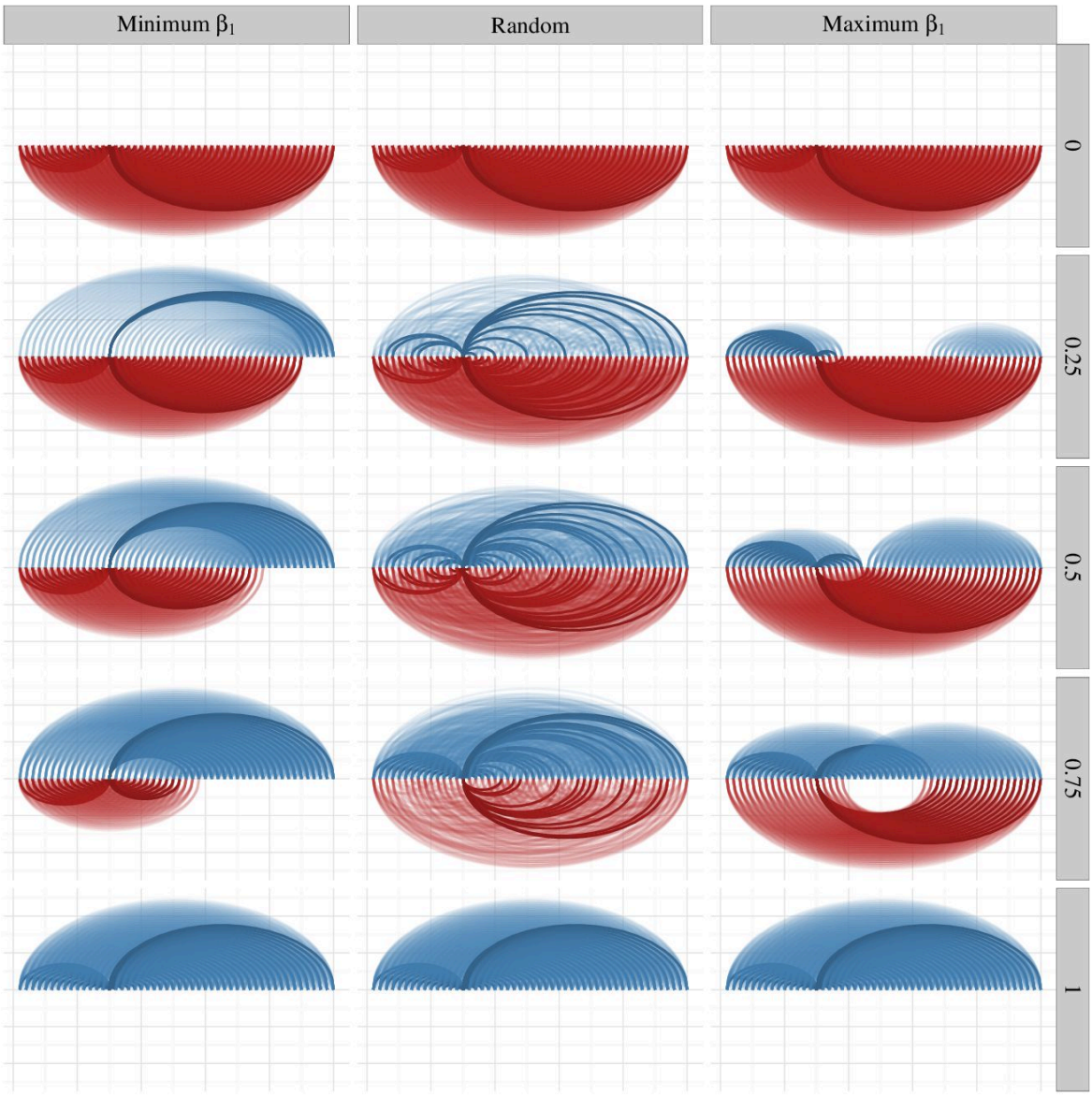


Figure 4.11: As fig. 4.2 in the main text, except with the three interspecific interaction matrices in fig. 4.10.

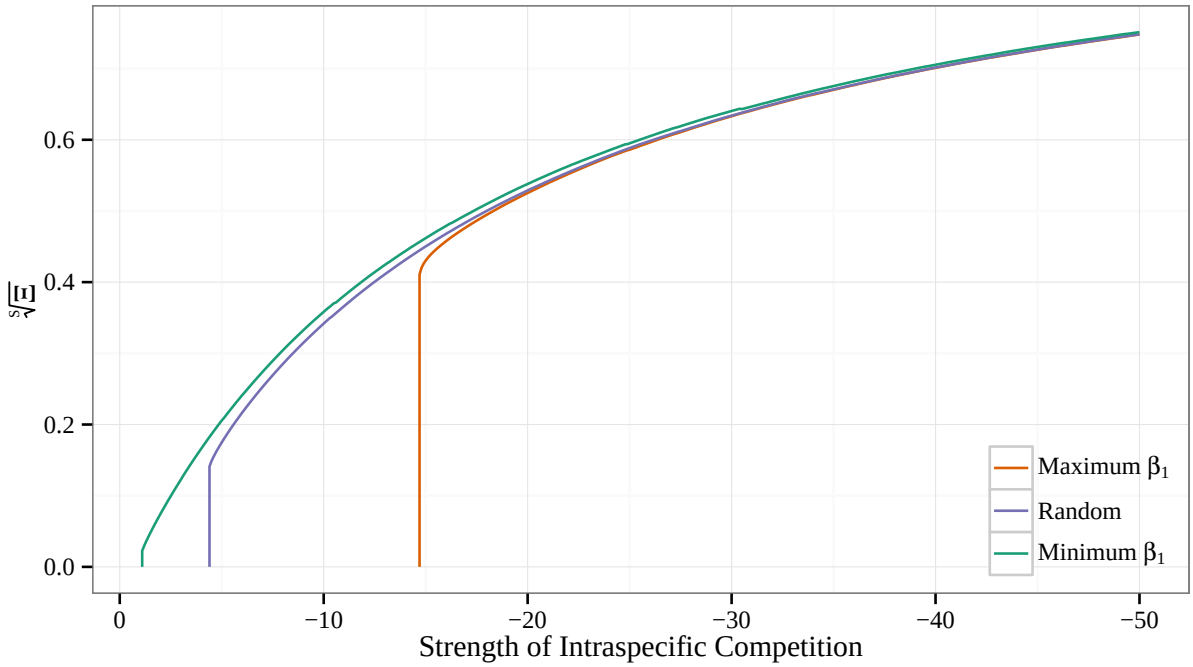


Figure 4.12: As fig. 4.3 in the main text, except the interspecific competition coefficients are uniformly sampled from $[-1, 0]$.

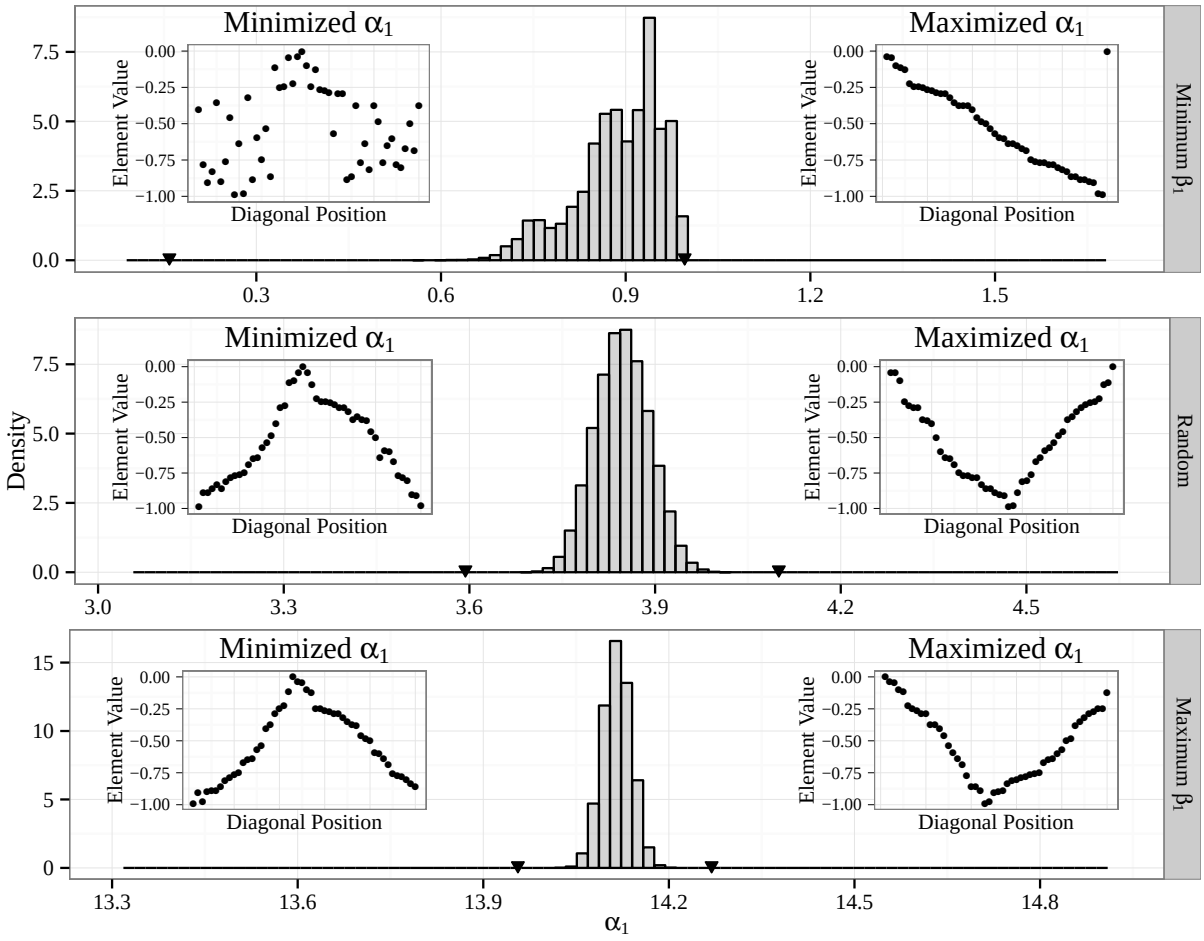


Figure 4.13: Distributions of the rightmost eigenvalue α_1 of $\mathbf{A} = \mathbf{B} + \mathbf{C}$ ($\mathbf{B}$ contains only interspecific, $\mathbf{C}$ only intraspecific effects) in a 50-species community, when the rightmost eigenvalue β_1 of $\mathbf{B}$ is minimized (top), corresponds to a random matrix (middle), or maximized (bottom). The histogram in each row is generated by 100,000 random permutations of $\mathbf{C}$'s diagonal coefficients. Species in the insets are ordered as in fig. 4.10. In each of the three cases, we also looked for the particular arrangement which minimizes/maximizes α_1, using a genetic optimization algorithm (left/right insets showing the magnitudes of $\mathbf{C}$'s diagonal entries in order; their corresponding values of α_1 are marked by the arrows). Interspecific competition coefficients are uniformly sampled from $[-1, 0]$, and the intraspecific coefficients are also uniformly sampled from $[-1, 0]$.

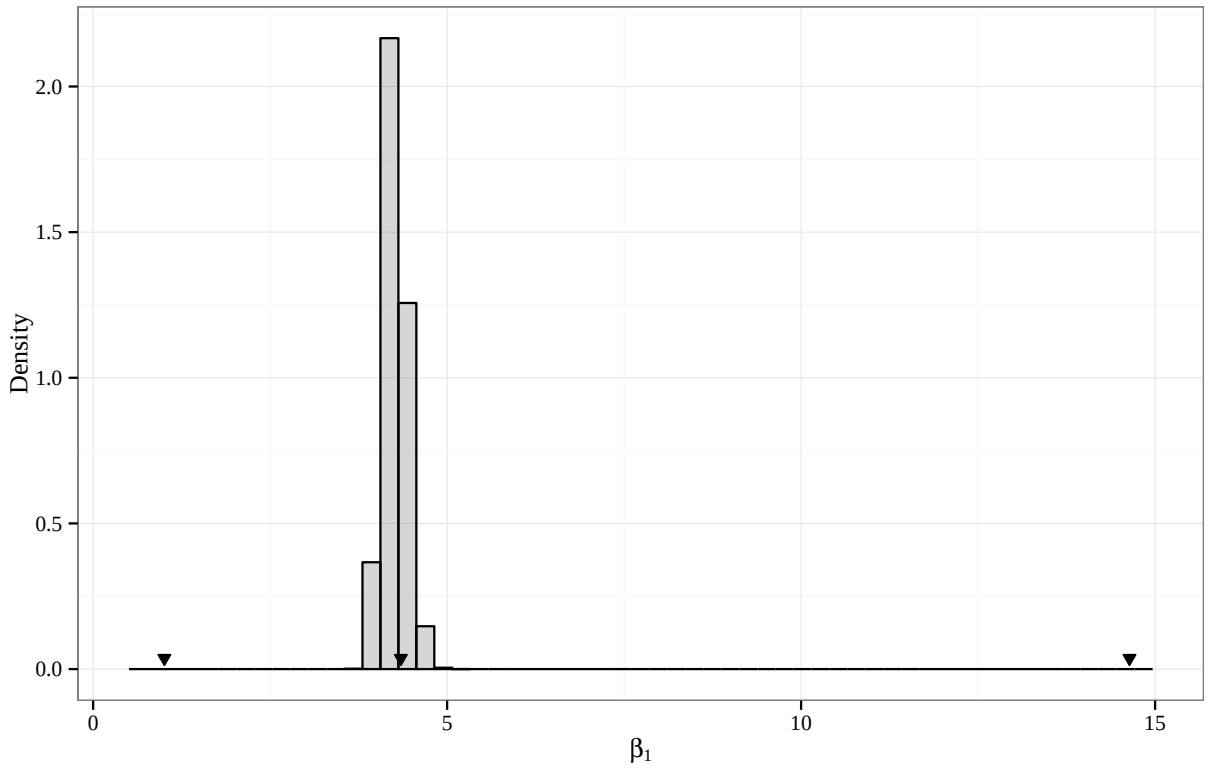


Figure 4.14: Distributions of the rightmost eigenvalue β_1 of $\mathbf{B}$ in a 50-species community. The histogram is generated by 100,000 random permutations of $\mathbf{B}$'s coefficients. The black arrows indicate the rightmost eigenvalues of the most stabilized, random, and least stabilized configurations, respectively (as in fig. 4.10). The first and last of which were found using a genetic optimization algorithm. The most stabilized matrix has been sorted according to the eigenvector associated with the left-most eigenvalue; the other two are sorted by the eigenvector associated with the right-most eigenvalue. Interspecific competition coefficients are uniformly sampled from $[-1, 0]$.

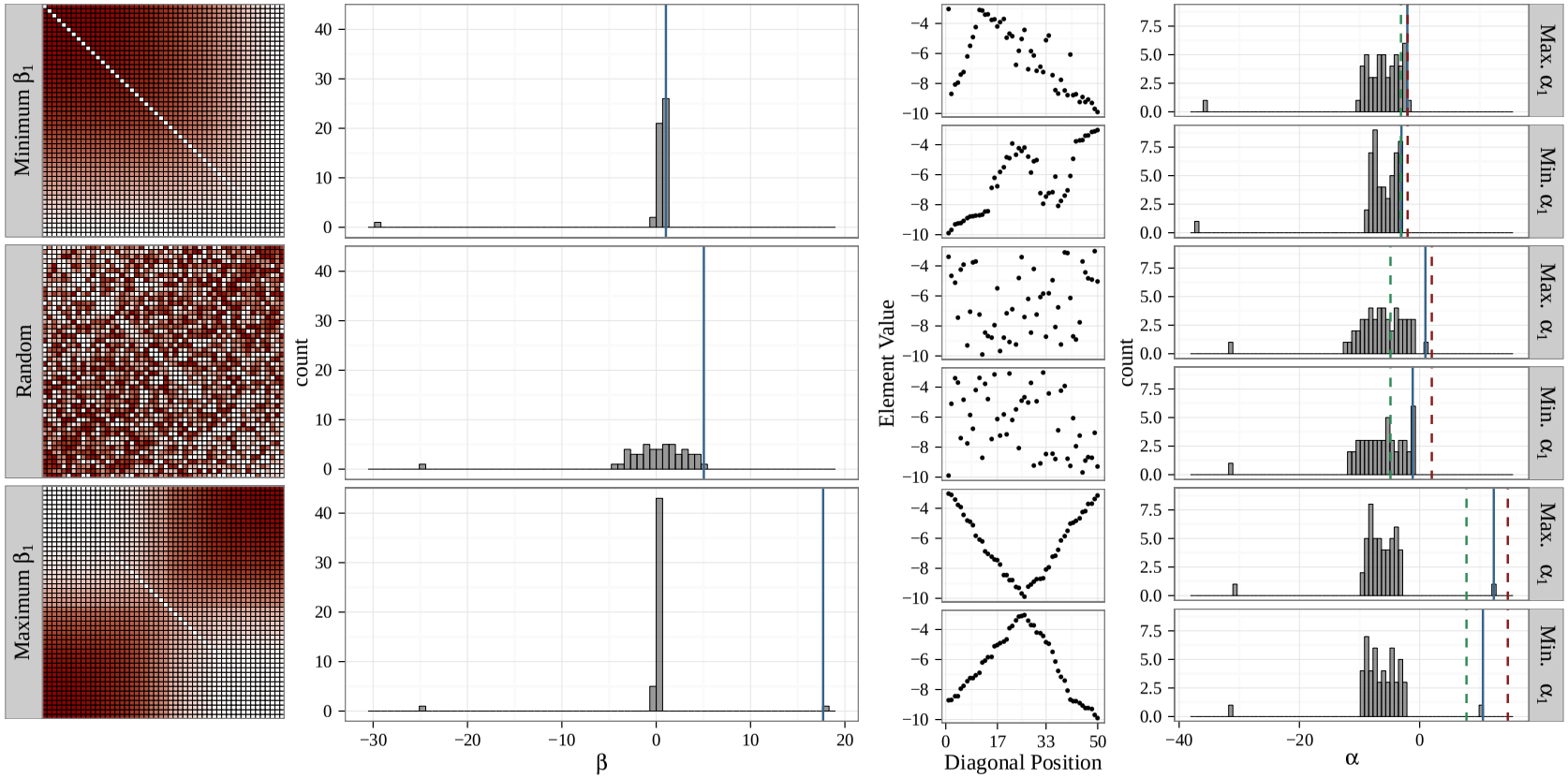


Figure 4.15: As fig. 4.1 in the main text, except the interspecific competition coefficients are sampled from a beta distribution with parameters (1/2, 1/2), and the intraspecific coefficients are uniformly sampled from $[-10, -3]$.

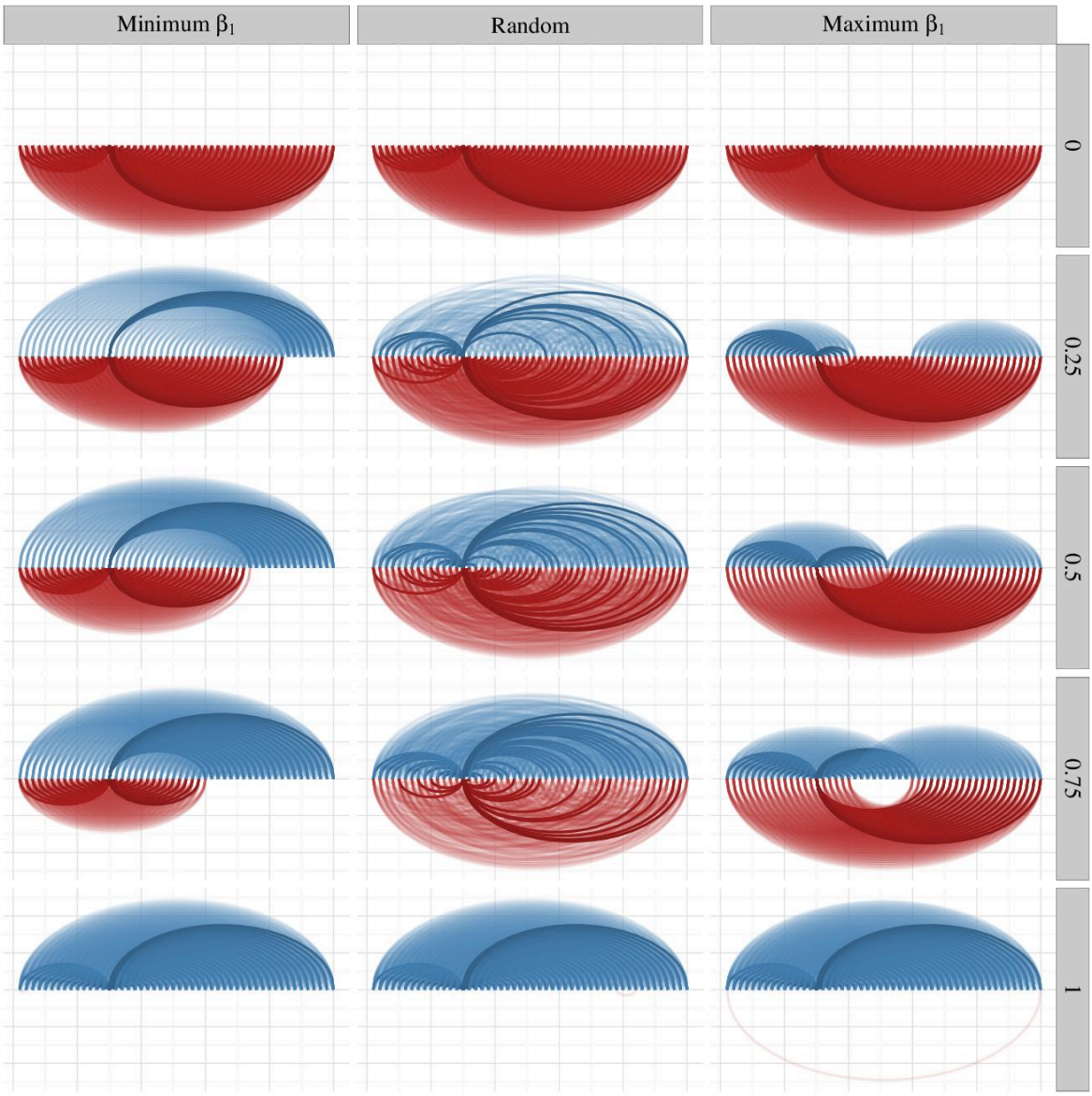


Figure 4.16: As fig. 4.2 in the main text, except with the three interspecific interaction matrices in fig. 4.15.

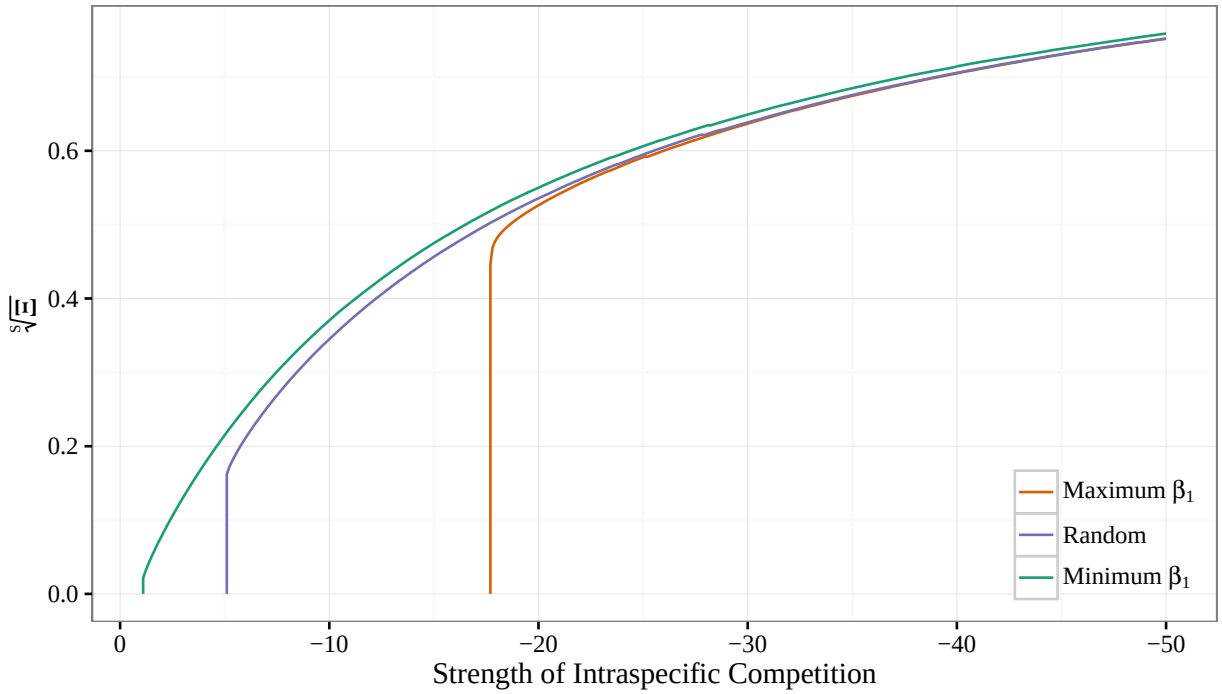


Figure 4.17: As fig. 4.3 in the main text, except the interspecific competition coefficients are sampled from a beta distribution with parameters (1/2, 1/2).

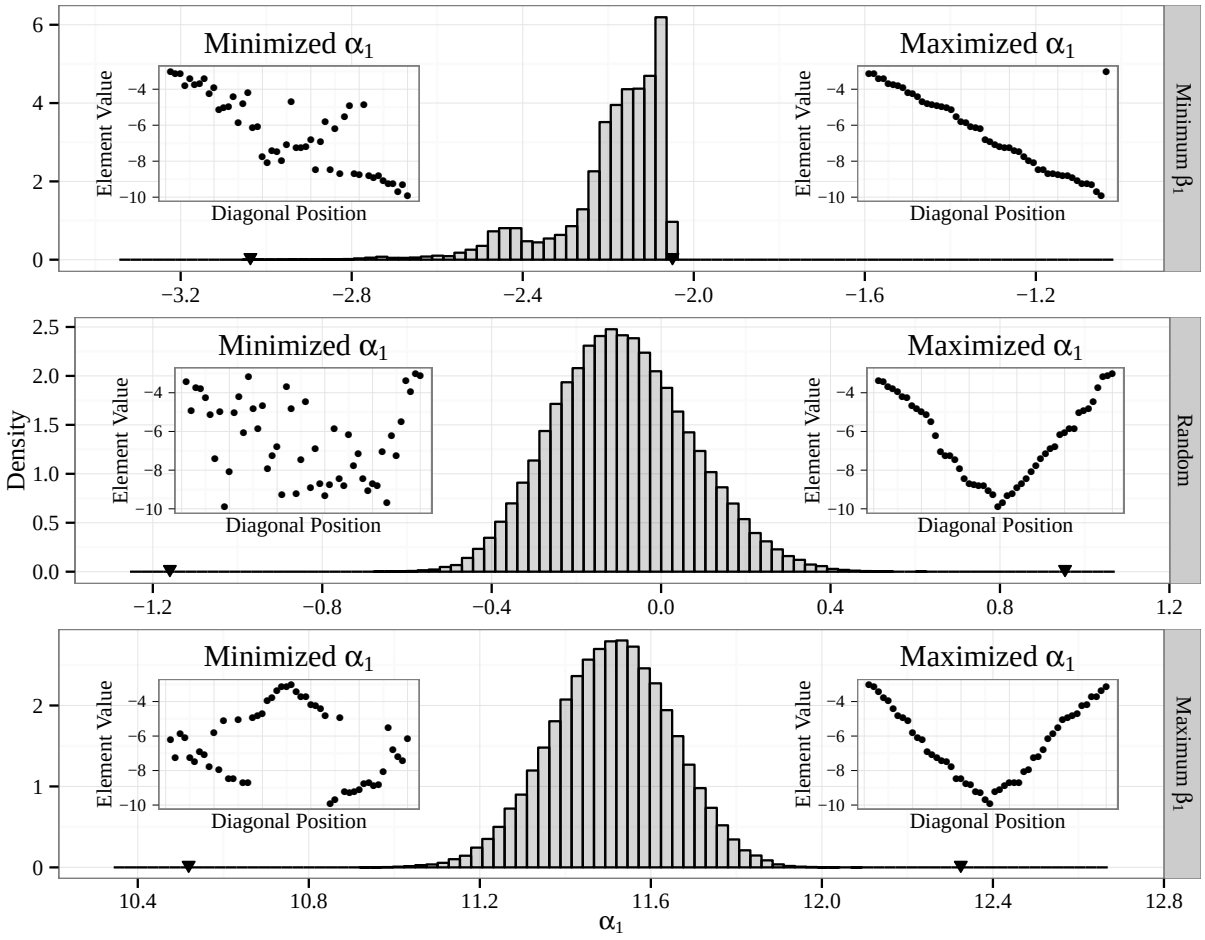


Figure 4.18: As fig. 4.13, except with interspecific competition coefficients are uniformly sampled from $[-1, 0]$, and the intraspecific coefficients are also uniformly sampled from $[-1, 0]$.

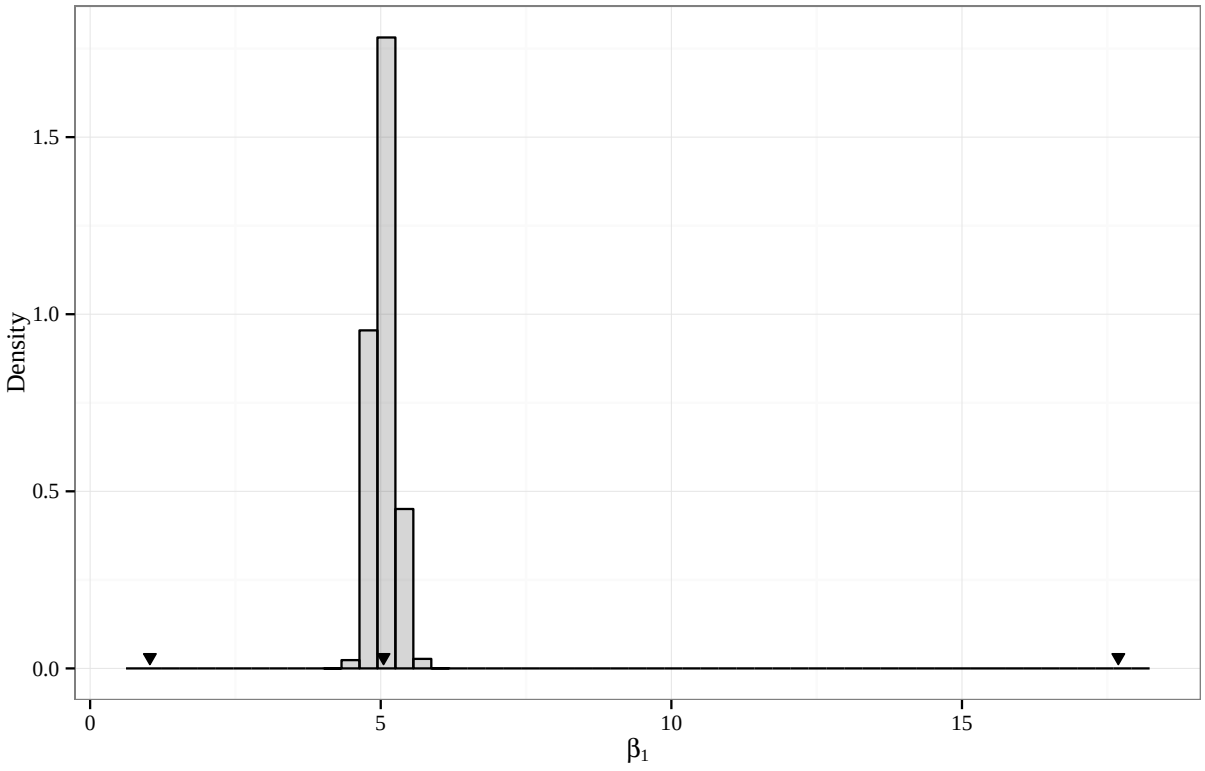


Figure 4.19: As fig. 4.14, except with interspecific competition coefficients are uniformly sampled from $[-1, 0]$.

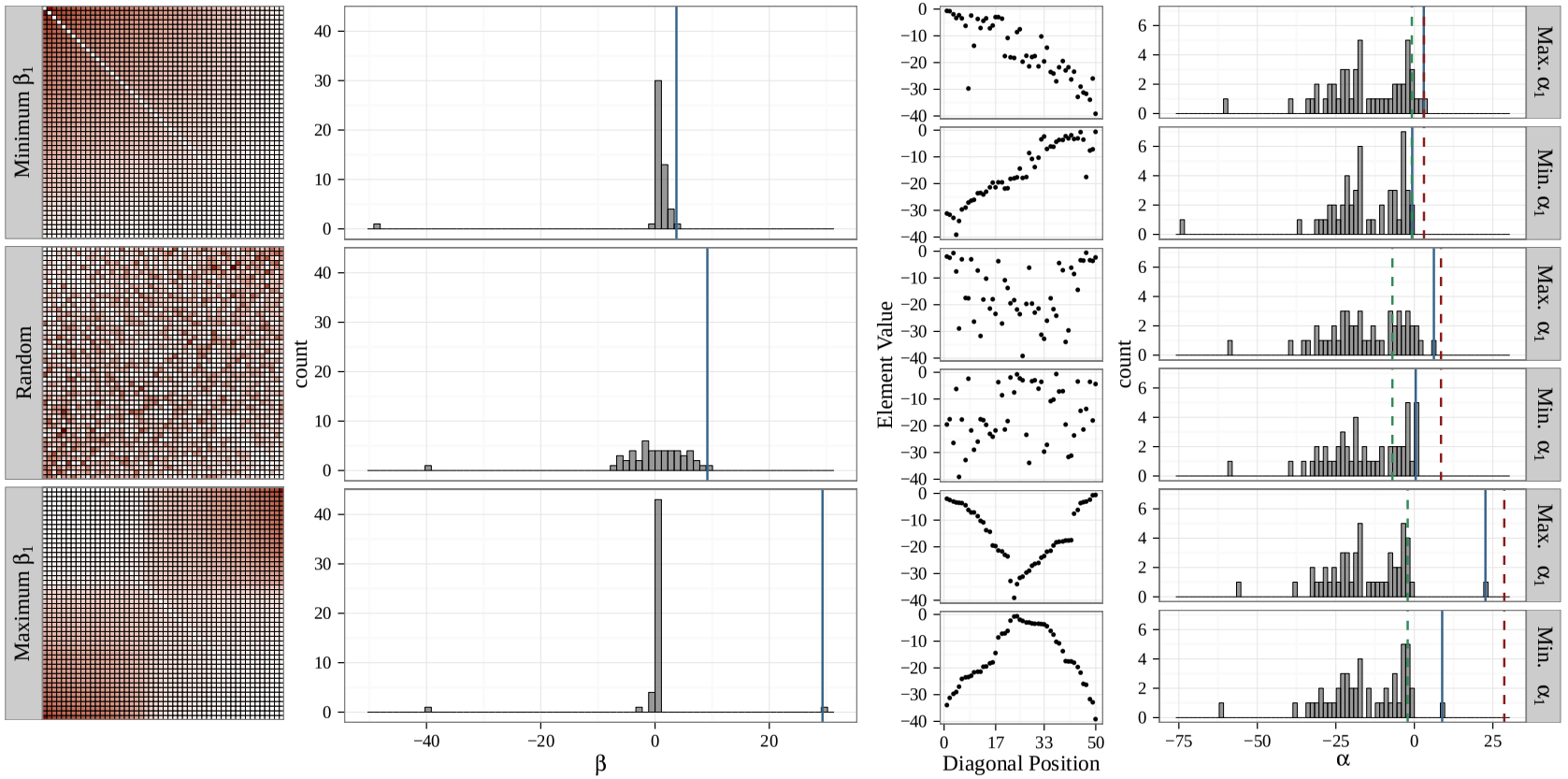


Figure 4.20: As fig. 4.1 in the main text, except the interspecific competition coefficients are sampled from a half-normal distribution with $\sigma = 1$, and the intraspecific coefficients are sampled from a half-normal distribution with $\sigma = 20$.

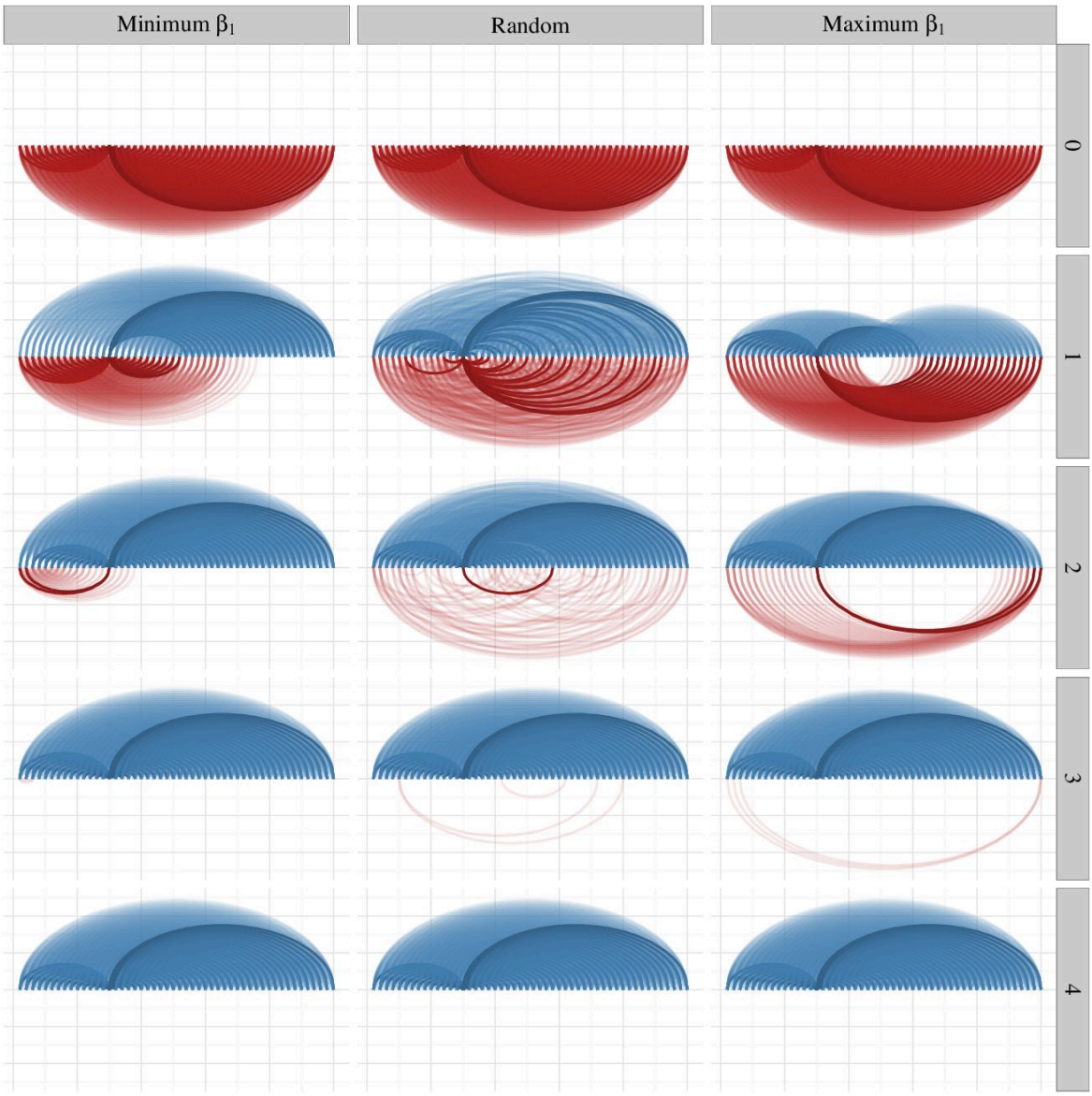


Figure 4.21: As fig. 4.2 in the main text, except with the three interspecific interaction matrices in fig. 4.20.

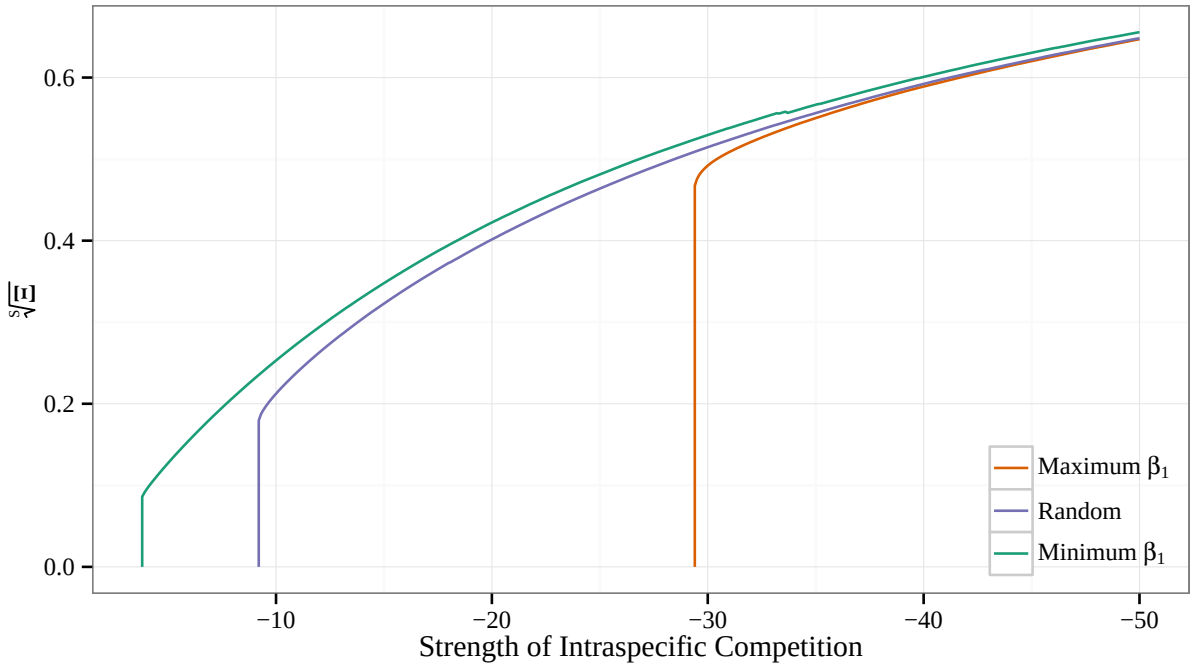


Figure 4.22: As fig. 4.3 in the main text, except the interspecific competition coefficients are sampled from a half-normal distribution with $\sigma = 1$.

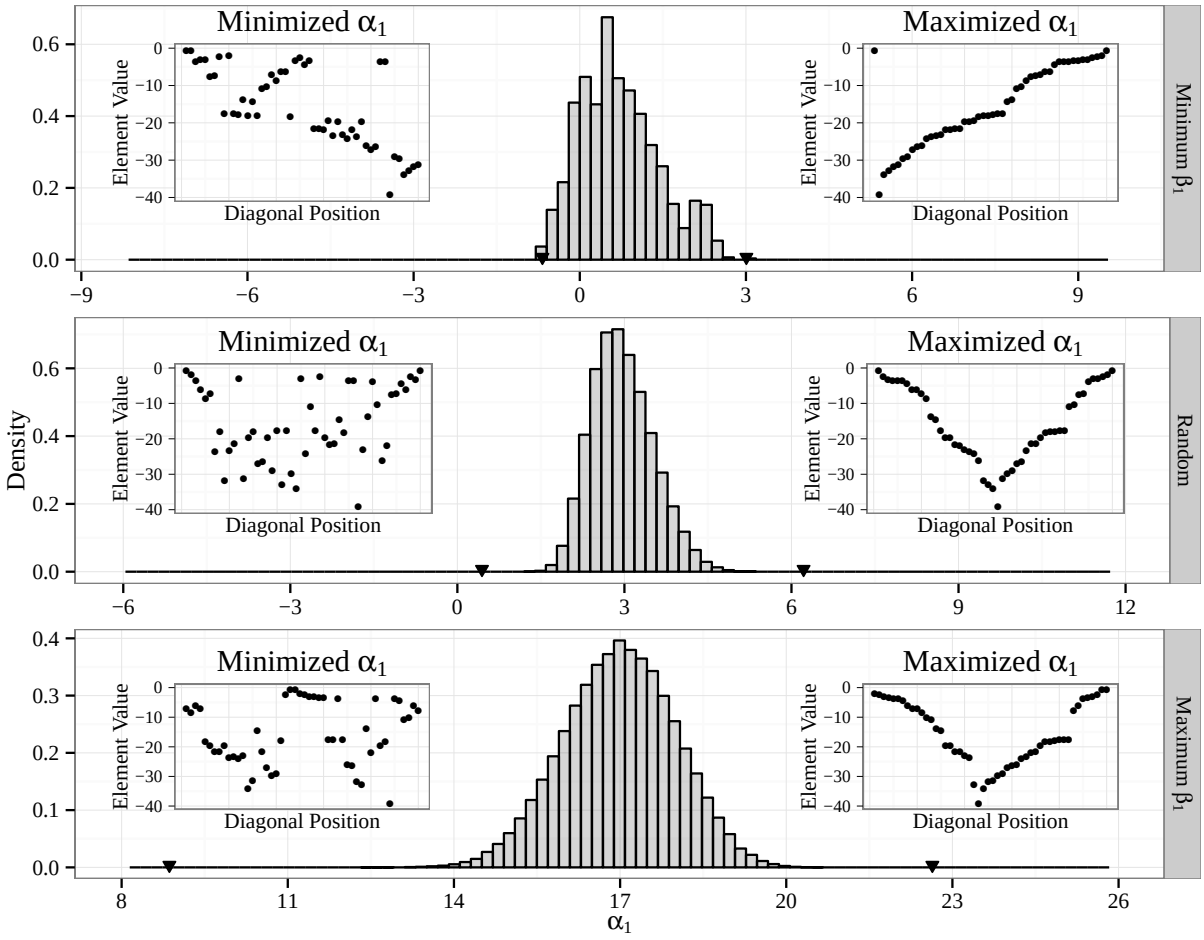


Figure 4.23: As fig. 4.13, except with interspecific competition coefficients are uniformly sampled from $[-1, 0]$, and the intraspecific coefficients are also uniformly sampled from $[-1, 0]$.

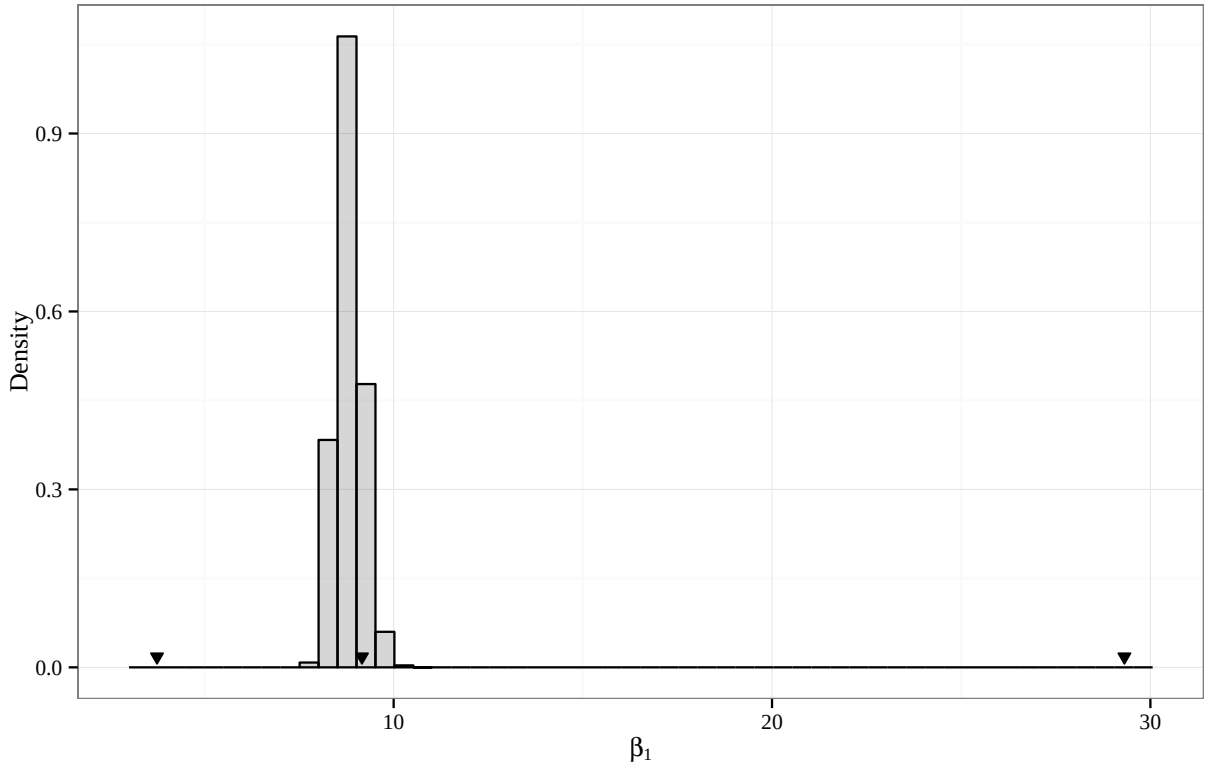


Figure 4.24: As fig. 4.14, except the interspecific competition coefficients are sampled from a half-normal distribution with $\sigma = 1$, and the intraspecific coefficients are sampled from a half-normal distribution with $\sigma = 20$.

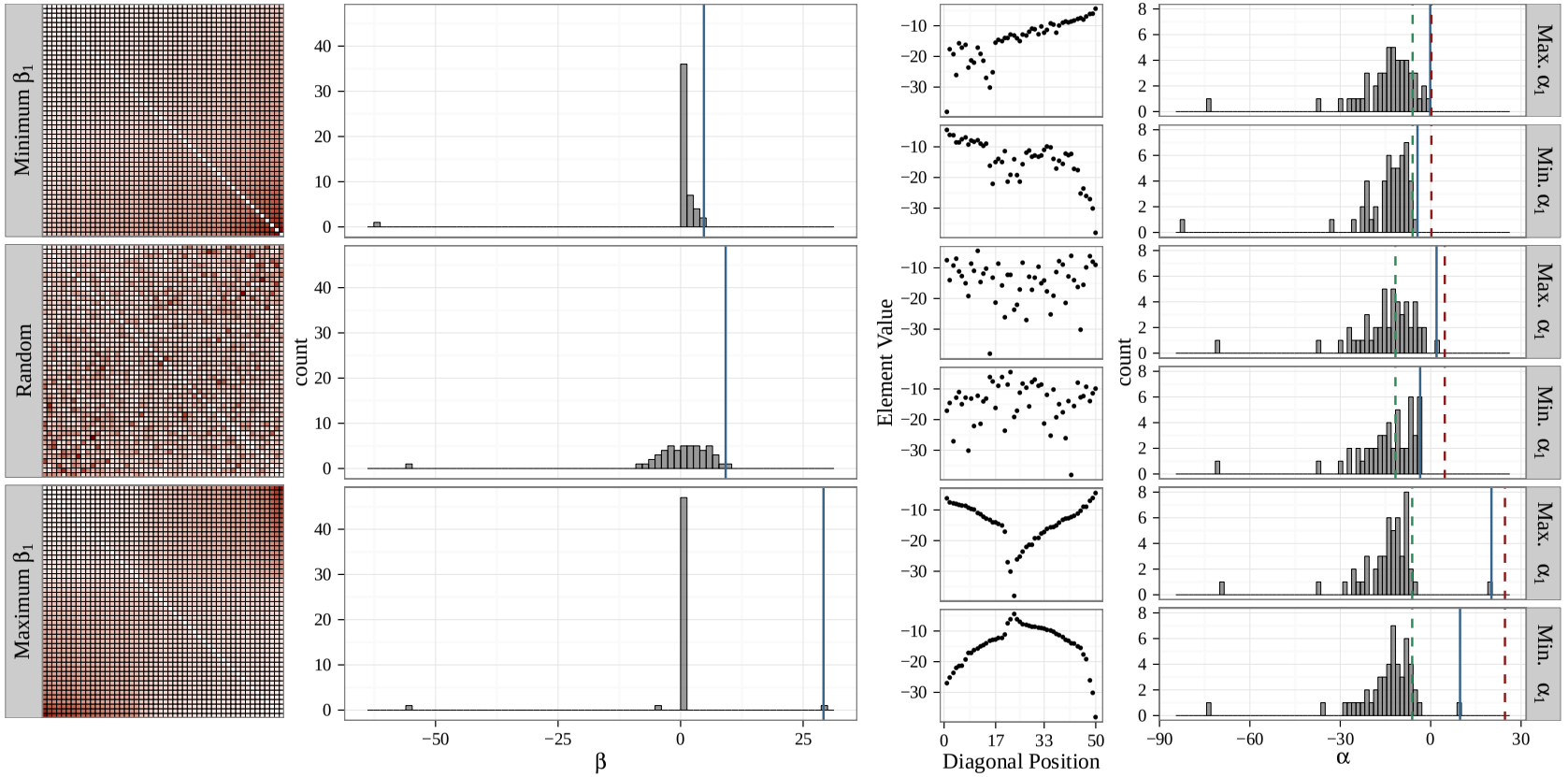


Figure 4.25: As fig. 4.1 in the main text, except the interspecific competition coefficients are sampled from a lognormal distribution with parameters $\mu = 0$, $\sigma = 0.5$, and the intraspecific coefficients are sampled from a lognormal distribution with parameters $\mu = 2.5$, $\sigma = 0.5$.

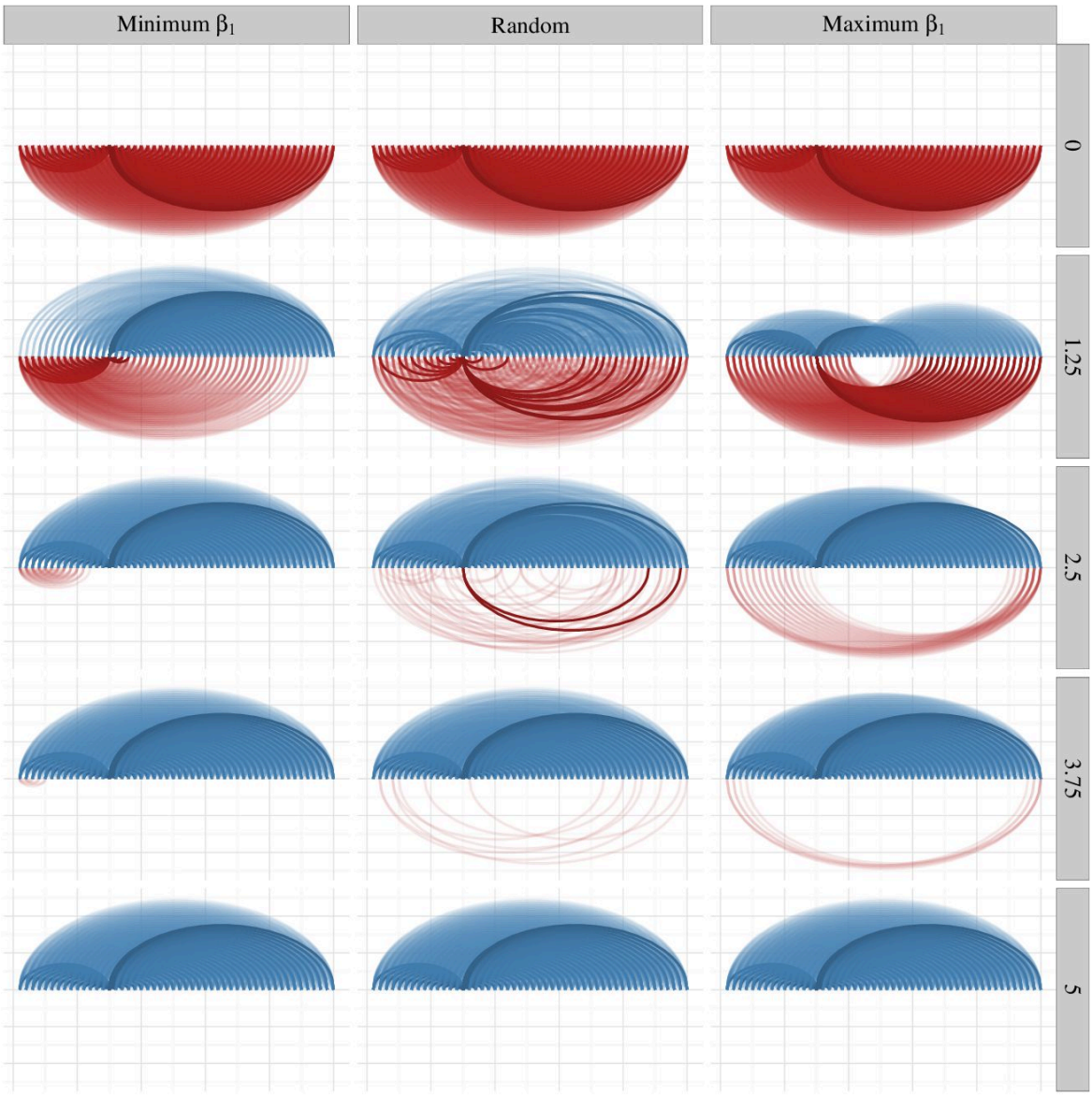


Figure 4.26: As fig. 4.2 in the main text, except with the three interspecific interaction matrices in fig. 4.25.

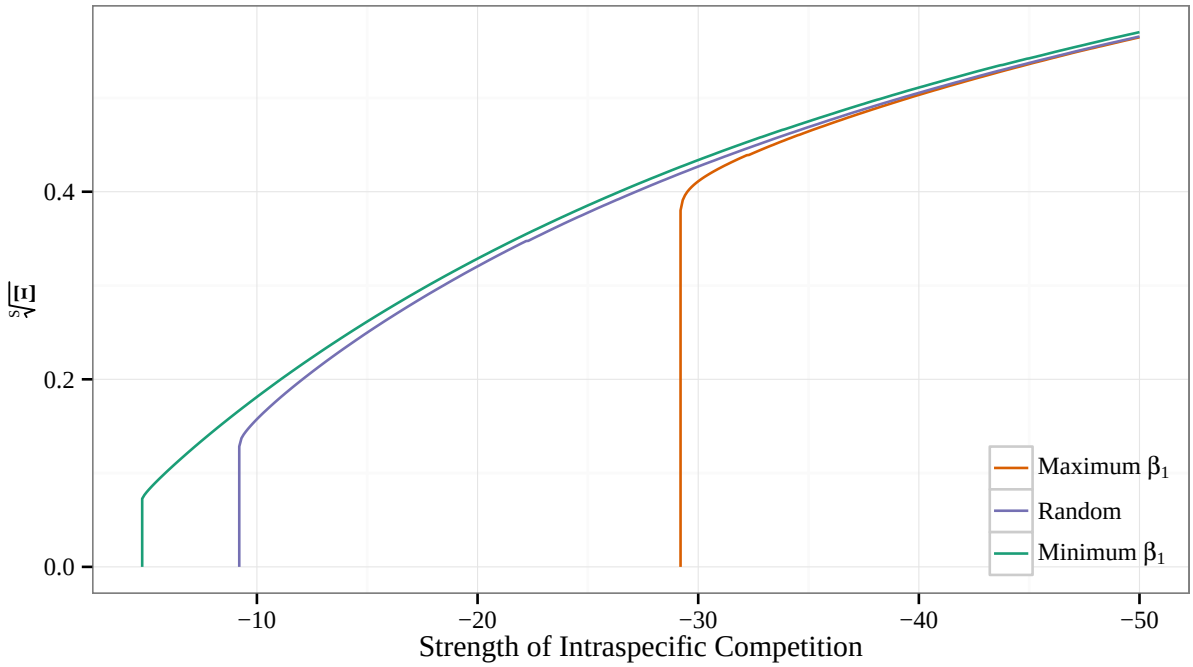


Figure 4.27: As fig. 4.3 in the main text, except the interspecific competition coefficients are sampled from a lognormal distribution with parameters $\mu = 0$, $\sigma = 0.5$.

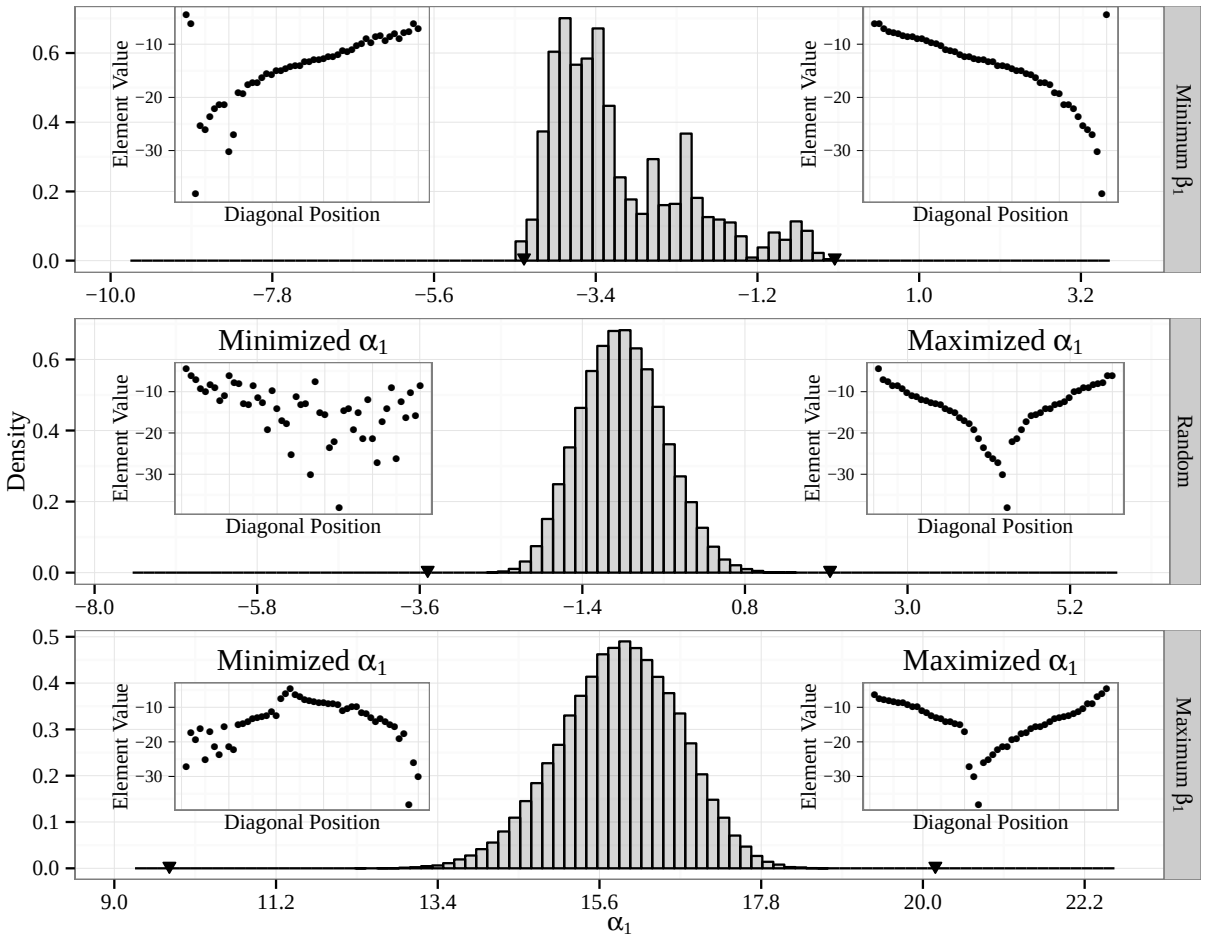


Figure 4.28: As fig. 4.13, except the interspecific competition coefficients are sampled from a lognormal distribution with parameters $\mu = 0$, $\sigma = 0.5$, and the intraspecific coefficients are sampled from a lognormal distribution with parameters $\mu = 2.5$, $\sigma = 0.5$.

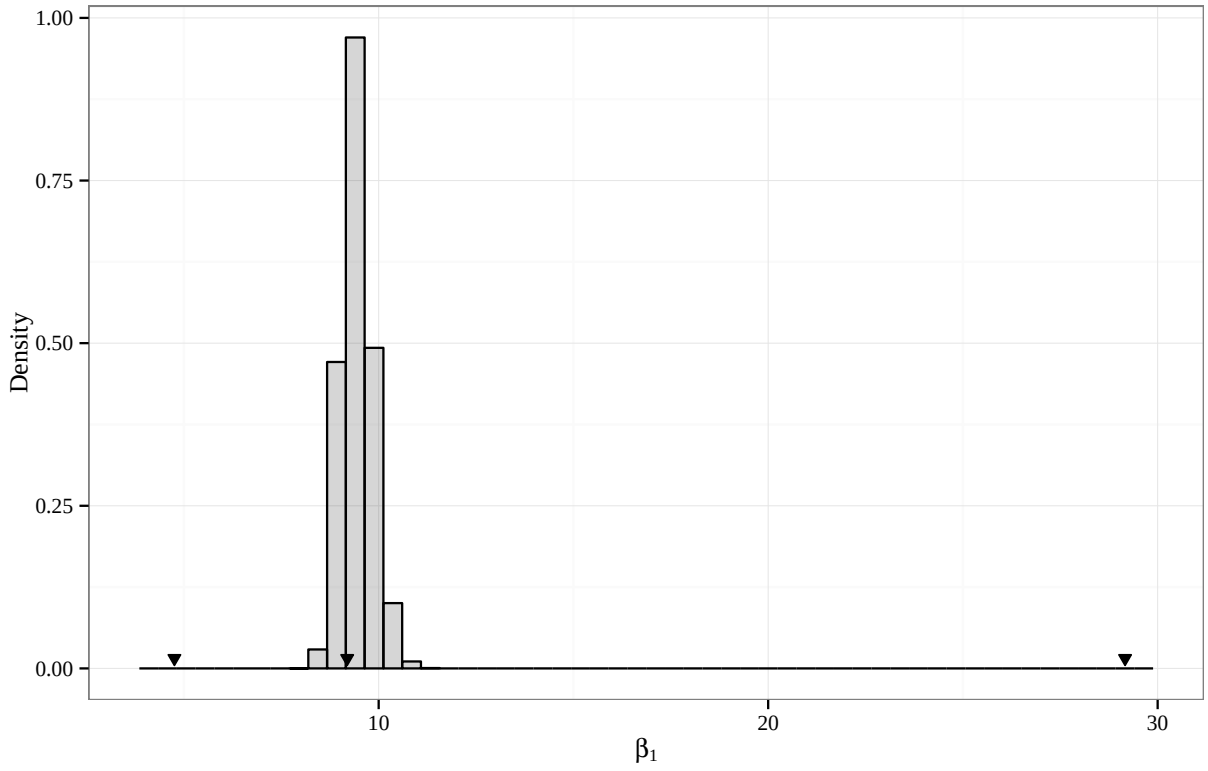


Figure 4.29: As fig. 4.14, except the interspecific competition coefficients are sampled from a lognormal distribution with parameters $\mu = 0$, $\sigma = 0.5$.

CHAPTER 5

CONCLUSION

The study of ecological networks is challenging both due a diversity of approaches to the construction of ecological networks (in which nodes can vary in taxonomic resolution, have spatial and temporal variation, and have intricate age or class substructure, and edges can depict various types of interactions, movements, or transformations) and due to the large variation that seems to be inherent to ecological systems. Yet, I hope that the preceding chapters provide some evidence that there are nevertheless detectable patterns in ecological network structure that yield insight into real, biological processes at both the local and whole-network levels. At first glance, these chapters might seem to have little in common, but each is an example of taking patterns in network structure and drawing out ecological relevance (or *vice versa*). This intimate relationship between the actual ecology underlying what we choose to depict with matrices and networks and the mathematics we can apply to data in those forms is one of the most fascinating aspects of science in general, but for me is particularly impressive when considering ecological systems.

5.1 Future Directions

The growth of research into network analyses continues to accelerate both within Ecology and in science more generally. This research includes new metrics to summarize network structure, *etc.*, but it also includes new ways of viewing complex systems through a network lens.

Two areas that have received particular attention in recent years are multi-layer networks and dynamic networks. The former arose in Ecology in part from the observation that while Ecologists have for a long time been storing network data in webs depicting a particular type of interaction (*e.g.* pollination or parasitism), the same species are often involved in a number

of different interaction types with one another (and sometimes multiple types of interactions between the same pair of species). This has led to different approaches of incorporating these multiple types of interactions into a single network framework.

One approach has been to just include different interaction strengths corresponding to different interactions (*i.e.* positive values for mutualisms and negative values for antagonistic interactions). This, however has a shortcoming of not being able to represent a species that has both types of interactions with the same partner (*e.g.* consider a butterfly which consumes plants in its caterpillar phase that it later pollinates after metamorphosis).

The way around this is to allow multiple interactions between two nodes of the network, and one way to organize these different interactions is in splitting the network into "layers" where each layer could correspond to a type of interaction, or a time of year, or a location in space, *etc.* These networks make sense intuitively and are graphically appropriate representations, but how to apply traditional network metrics to these constructions is not obvious.

Dynamic networks face a similar hurdle. In these networks, the strength or even existence of nodes and links changes through time. While undoubtedly, this change makes the networks a more accurate portrayal of reality in which connections are inevitably spatially and temporally constrained, there has been little research into how this change affects dynamical properties nor into when and whether this added complexity is worth it from a model selection perspective.

5.2 Final Remarks

As ecology moves forward, the volume and quality of data are increasing at a rapid pace. It will be important to develop tools and methods for both the storage/management of large data sets as well as their analysis. In this vein, computation and mathematics will be key assets for ecologists looking to make an impact. Even more important is an appreciation for

the progress made in network analyses across a range of other disciplines. While research is often focused on the *de novo* development of ideas that have never been considered before, in my experience, a much more common source of innovation is the integration of two or more ideas, well established in their own rights, but which no one has thought to conjoin before. As networks become more studied not just in Ecology, but in Science generally, there will increasingly be opportunity for applying foreign analyses to ecological data. This has great potential for catalyzing ecological research, but also requires extra consideration into what the ecological interpretations of such approaches and their results, as well as a vigilance about the appropriateness of any such analysis in addressing a given question.

REFERENCES

[1] P. A. Abrams and W. G. Wilson. Coexistence of competitors in metacommunities due to spatial variation in resource growth rates; does R^* predict the outcome of competition? *American Naturalist*, 7:929–940, 2004.

[2] T. P. Adams, D. W. Purves, and S. W. Pacala. Understanding height-structured competition in forests: is there an R^* for light? *Proceedings of the Royal Society B*, 274:3039–3047, 2007.

[3] P. B. Adler, S. P. Ellner, and J. M. Levine. Coexistence of perennial plants: an embarrassment of niches. *Ecology Letters*, 13:1019–1029, 2010.

[4] P. B. Adler, J. Hillerislambers, and J. M. Levine. A niche for neutrality. *Ecology Letters*, 10:95–104, 2007.

[5] S. Allesina and S. Tang. The stability-complexity relationship at age 40: a random matrix perspective. *Population Ecology*, 57:63–75, 2015.

[6] Stefano Allesina and Mercedes Pascual. Network structure, predator-prey modules, and stability in large food webs. *Theoretical Ecology*, 1(1):55–64, 2008.

[7] Stefano Allesina and Mercedes Pascual. Food web models: a plea for groups. *Ecology Letters*, 12(7):652–62, jul 2009.

[8] Mário Almeida-Neto, Paulo Guimaraes, Paulo R Guimarães, Rafael D Loyola, and Werner Ulrich. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239, 2008.

[9] David M. Althoff. Does parasitoid attack strategy influence host specificity? A test with New World braconids. *Ecological Entomology*, 28(4):500–502, 2003.

[10] P. Amarasekare, M. F. Hoopes, N. Mouquet, and M. Holyoak. Mechanisms of coexistence in competitive metacommunities. *American Naturalist*, 164:310–326, 2004.

[11] Per-arne Amundsen, Kevin D Lafferty, Rune Knudsen, Raul Primicerio, Anders Klemetsen, and Armand M Kuris. Food web topology and parasites in the pelagic zone of a subarctic lake. *Journal of Animal Ecology*, 78:563–572, 2009.

[12] Wirt Atmar and Bruce D Patterson. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, 96(3):373–382, 1993.

[13] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, 2009.

[14] G. Barabás, S. Pigolotti, M. Gyllenberg, U. Dieckmann, and G. Meszéna. Continuous coexistence or discrete species? A new review of an old question. *Evolutionary Ecology Research*, 14:523–554, 2012.

[15] Sébastien Barot. Mechanisms promoting plant coexistence: can all the proposed processes be reconciled? *Oikos*, 106:185–192, 2004.

[16] Jordi Bascompte and Pedro Jordano. Plant-animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, 38:567–593, 2007.

[17] Jordi Bascompte, Pedro Jordano, and Jens M Olesen. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312(5772):431–433, 2006.

[18] Jordi Bascompte and Carlos J Melián. Simple trophic modules for complex food webs. *Ecology*, 86(11):2868–2873, 2005.

[19] Eric L Berlow, Jennifer A Dunne, Neo D Martinez, Philip B Stark, Richard J Williams, Ulrich Brose, and Simon A Levin. Simple Prediction of Interaction Strengths in Complex Food Webs. *Proceedings of the National Academy of Sciences*, 106(1):187–191, 2009.

[20] Nico Blüthgen. Why network analysis is often disconnected from community ecology: a critique and an ecologist's guide. *Basic and Applied Ecology*, 11(3):185–195, 2010.

[21] B. M. Bolker, S. W. Pacala, and C. Neuhauser. Spatial dynamics in model plant communities: What do we really know? *American Naturalist*, 162:135–148, 2003.

[22] Elizabeth T Borer, Kurt Anderson, Carol A Blanchette, Bernardo Broitman, Scott D Cooper, Benjamin S Halpern, Eric W Seabloom, and Jonathan B Shurin. Topological approaches to food web analyses: a few modifications may improve our insights. *Oikos*, 99(2):397–401, 2002.

[23] Stephen P Borgatti. Centrality and network flow. *Social networks*, 27(1):55–71, 2005.

[24] Jonathan J Borrelli. Selection against instability: stable subgraphs are most frequent in empirical food webs. *Oikos*, 124(12):1583–1588, 2015.

[25] Jonathan J Borrelli, Stefano Allesina, Priyanga Amarasekare, Roger Arditi, Ivan Chase, John Damuth, Robert D Holt, Dmitrii O Logofet, Mark Novak, Rudolf P Rohr, et al. Selection on stability across ecological scales. *Trends in ecology & evolution*, 30(7):417–425, 2015.

[26] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. On modularity clustering. *IEEE transactions on knowledge and data engineering*, 20(2):172–188, 2008.

[27] Ulrich Brose, Richard J. Williams, and Neo D. Martinez. Allometric scaling enhances stability in complex food webs. *Ecology Letters*, 9(11):1228–1236, 2006.

[28] Enrique Burgos, Horacio Ceva, Roberto PJ Perazzo, Mariano Devoto, Diego Medan, Martín Zimmermann, and Ana María Delbue. Why nestedness in mutualistic networks? *Journal of theoretical biology*, 249(2):307–313, 2007.

[29] Lorenzo Camerano. Dell'equilibrio dei viventi mercé la reciproca distruzione. *Accademia delle Scienze di Torino*, 15:393–414, 1880.

[30] T. J. Case. Invasion resistance arises in strongly interacting species-rich model competition communities. *Proceedings of the National Academy of Sciences USA*, 87:9610–9614, 1990.

[31] T. J. Case. *An Illustrated Guide to Theoretical Ecology.* Oxford University Press, New York, 2000.

[32] H. Caswell. *Matrix population models: Construction, analysis, and interpretation. 2nd edition.* Sinauer Associates, 2001.

[33] J. M. Chase, P. A. Abrams, J. P. Grover, S. Diehl, P. Chesson, R. D. Holt, S. A. Richards, R. M. Nisbet, and T. J. Case. The interaction between predation and com-

petition: a review and synthesis. *Ecology Letters*, 5:302–315, 2002.

[34] P. Chesson. Multispecies competition in variable environments. *Theoretical Population Biology*, 45:227–276, 1994.

[35] P. Chesson. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31:343–366, 2000.

[36] P. Chesson. Scale transition theory with special reference to species coexistence in a variable environment. *Journal of Biological Dynamics*, 3:149–163, 2009.

[37] P. Chesson. Ecological niches and diversity maintenance. In *Research in Biodiversity-Models and Applications*, pages 43–60. InTech, Rijeka, Croatia, 2011.

[38] P. Chesson and S. Ellner. Invasibility and stochastic boundedness in two-dimensional competition models. *Journal of Mathematical Biology*, 27:117–138, 1989.

[39] P. Chesson, R. L. E. Gebauer, S. Schwinning, N. Huntly, K. Wiegand, M. S. K. Ernest, A. Sher, A. Novoplanski, and J. F. Weltzin. Resource pulses, species interactions, and diversity maintenance in arid and semi-arid environments. *Oecologia*, 141:236–253, 2004.

[40] Alyssa R Cirtwill and Daniel B Stouffer. Concomitant predation on parasites is highly variable but constrains the ways in which parasites contribute to food web structure. *Journal of Animal Ecology*, 84(3):734–744, 2015.

[41] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[42] Joel Cohen and C.M. Newman. A stochastic theory of community food webs I . Models and aggregated data. *Proceedings of the Royal Society of London B*, 224(1237):421–448, 1985.

[43] C. Darwin. *On The Origin of Species by means of natural selection. First Edition.* John Murray, Albemarle Street, London, 1859.

[44] S. Diederich and M. Opper. Replicatiors with random interactions: a solvable model. *Physical Review A*, 39:4333–4336, 1989.

[45] JA Dunne, KD Lafferty, AP Dobson, RF Hechinger, AM Kuris, ND Martinez, JP McLaughlin, KN Mouritsen, R Poulin, K Reise, DB Stouffer, DW Thieltges, RJ Williams, and CD Zander. Data from: Parasites affect food web structure primarily through increased diversity and complexity, 2013.

[46] Jennifer A Dunne, Kevin D Lafferty, Andrew P Dobson, Ryan F Hechinger, Armand M Kuris, Neo D Martinez, John P McLaughlin, Kim N Mouritsen, Robert Poulin, Karsten Reise, Daniel B Stouffer, David W Thieltges, Richard J Williams, and Claus Dieter Zander. Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biology*, 11(6):e1001579, jan 2013.

[47] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12917–22, oct 2002.

[48] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and

biodiversity loss in food webs : robustness increases with connectance. *Ecology Letters*, 5:558–567, 2002.

[49] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology letters*, 5(4):558–567, 2002.

[50] Jennifer A. Dunne, Richard J. Williams, and Neo D. Martinez. Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, 5(4):558–567, 2002.

[51] L A Dyer, M S Singer, J T Till, J O Stireman, G L Gentry, R J Marquis, R E Ricklefs, H F Greeney, D L Wagner, H C Morais, I R Diniz, T A Kursar, and P D Coley. Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, 448(August):696–700, 2007.

[52] L. Edelstein-Keshet. *Mathematical models in biology*. Random House, New York, USA, 1988.

[53] K. F. Edwards and S. J. Schreiber. Preemption of space can lead to intransitive coexistence of competitors. *Oikos*, 119:1201–1209, 2010.

[54] Anna Eklöf, Matthew R Helmus, M Moore, and Stefano Allesina. Relevance of evolutionary history for food web structure. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20112149, 2011.

[55] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[56] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, pages 128–140, 1741.

[57] Colin Fontaine, Paulo R Guimarães, Sonia Kéfi, Nicolas Loeuille, Jane Memmott, Wim H van Der Putten, Frank JF van Veen, and Elisa Thébault. The ecological and evolutionary implications of merging different types of networks. *Ecology letters*, 14(11):1170–1181, 2011.

[58] G. E. Forrester, B. Evans, M. A. Steele, and R. R. Vance. Assessing the magnitude of intra- and interspecific competition in two coral reef fishes. *Oecologia*, 148:632–640, 2006.

[59] W. Fulton. Eigenvalues, invariant factors, highest weights, and Schubert calculus. *Bulletin of the American Mathematical Society*, 37:209–249, 2000.

[60] T. Galla. Random replicators with asymmetric couplings. pages arXiv:cond–mat/0508174v3 [cond–mat.dis–nn], 2006.

[61] N. J. Gotelli. *A primer of ecology*. Sinauer Associates, Sunderland, Massachusetts, USA, 2008.

[62] J. Grilli, M. Adorisio, S. Suweis, G. Barabás, J. R. Banavar, S. Allesina, and A. Maritan. The geometry of coexistence in large ecosystems. page arXiv:1507.05337, 2015.

[63] Jacopo Grilli, Tim Rogers, and Stefano Allesina. Modularity and stability in ecological communities. *Nature communications*, 7:12031, 2016.

[64] Jan O. Haerter, Namiko Mitarai, and Kim Sneppen. Food Web Assembly Rules for Generalized Lotka-Volterra Equations. *PLoS Computational Biology*, 12(2):1–17, 2016.

[65] K. Hasegawa, K. Morita, K. Ohkuma, T. Ohnuki, and Y. Okamoto. Effects of hatchery chum salmon fry on density-dependent intra- and interspecific competition between wild chum and masu salmon fry. *Canadian Journal of Fisheries and Aquatic Sciences*, 71:1475–1482, 2014.

[66] R. F. Hechinger, K. D. Lafferty, J. P. McLaughlin, et al. Food webs including parasites, biomass, body sizes, and life stages for three california/baja california estuaries: Ecological archives e092-066. *Ecology*, 92:791–791, 2011.

[67] E. Hernández-García, C. López, S. Pigolotti, and K. H. Andersen. Species competition: coexistence, exclusion and clustering. *Philosophical Transactions of the Royal Society London, Series A*, 367:3183–3195, 2009.

[68] J. Hofbauer and K. Sigmund. *The Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge, 1988.

[69] R. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 2012.

[70] T Horton, A Kroh, N Bailly, N Boury-Esnault, S N Brandão, M J Costello, S Gofas, F Hernandez, J Mees, G Paulay, G Poore, G Rosenberg, S Stöhr, W Decock, S Dekeyzer, L Vandepitte, B Vanhoorne, K Verfaille, S Vranken, M J Adams, R Adlard, P Adriaens, S Agatha, K J Ahn, S Ahyong, N Akkari, B Alvarez, G Anderson, M Angel, C Arango, T Artois, S Atkinson, R Bank, A Barber, J P Barbosa, I Bartsch, D Bellan-Santini, A Berta, R Bieler, S Blanco, I Blasco-Costa, M Błażewicz, P Bock, R Böttger-Schnack, P Bouchet, C B Boyko, R Bray, N L Bruce, S Cairns, T N Campinas Bezerra, P Cárdenas, E Carstens, B K Chan, T Y Chan, L Cheng, M Churchill, C O Coleman, A G Collins, R Cordeiro, M Coste, K A Crandall, T Cribb, S Cutmore, F Dahdouh-Guebas, M Daly, M Daneliya, J C Dauvin, P Davie, S De Grave, V de Mazancourt, P Decker, D Defaye, J L D'Hondt, H Dijkstra, M Dohrmann, J Dolan, D Domning, R Downey, I Drapun, L Ector, U Eisendle-Flöckner, M Eitel, S.C.d. Encarnação, H Enghoff, J Epler, C Ewers-Saucedo, M Faber, S Feist, J Finn, C Fišer, G Fonseca, E Fordyce, W Foster, J H Frank, C Fransen, H Furuya, H Galea, O Garcia-Alvarez, R Garic, R Gasca, S Gaviria-Melo, S Gerken, H Gheerardyn, D Gibson, J Gil, A Gittenberger, C Glasby, A Glover, D González-Solís, D Gordon, M Grabowski, C Gravili, J M Guerra-García, R Guidetti, K Guilini, M D Guiry, K A Hadfield, E Hajdu, J Hallermann, B Hayward, E Hendrycks, A Herrera Bachiller, J.s. Ho, J Høeg, O Holovachov, J Hooper, R Houart, L Hughes, W Hummon, M Hyzny, L F M Iniesta, T Iseto, S Ivanenko, M Iwataki, G Jarms, D Jaume, K Jazdzewski, I Karanovic, B Karthick, Y H Kim, R King, P M Kirk, J P Kociolek, J Kolb, A Kotov, T Krapp-Schickel, A Kremenetskaia, R Kristensen, M Kulikovskiy, S Kullander, R La Perna, G Lambert, D Lazarus, F Le Coze, S LeCroy, D Leduc, E J Lefkowitz, R Lemaitre, Y Liu, A N Lörz, J Lowry, T Ludwig, N Lundholm, E Macpherson, L Madin, C Mah, T Mamos, R Manconi, G Mapstone, P E Marek, B Marshall, D J Marshall, P Martin, S McInnes, T Meidla, K Meland, K Merrin, R Mesibov, C Messing, D Miljutin,

C Mills, Ø Moestrup, V Mokievsky, T Molodtsova, F Monniot, R Mooi, A C Morandini, R da Rocha, F Moretzsohn, J Mortelmans, J Mortimer, L Musco, T A Neubauer, E Neubert, B Neuhaus, P Ng, A D Nguyen, C Nielsen, T Nishikawa, J Norenburg, T O'Hara, H Okahashi, D Opresko, M Osawa, Y Ota, D Patterson, H Paxton, V Perrier, W Perrin, I Petrescu, B Picton, J F Pilger, A Pisera, D Polhemus, M Potapova, P Pugh, J D Reimer, H Reip, M Reuscher, J W Reynolds, F Rimet, P Rios Lopez, M Rius, K Rützler, A Rzhavsky, J Saiz-Salinas, S Sala, S Santos, E Sar, A F Sartori, A Satoh, H Schatz, B Schierwater, A Schmidt-Rhaesa, S Schneider, C Schönberg, P Schuchert, A R Senna, C Serejo, S Shamsi, J Sharma, W A Shear, N Shenkar, A Shinn, M Short, J Sicinski, V Siegel, P Sierwald, E Simmons, F Sinniger, D Sivell, B Sket, H Smit, N Smit, N Smol, J F Souza-Filho, J Spelda, S N Stampar, W Sterrer, E Stienen, P Stoev, M Strand, E Suárez-Morales, M Summers, C Suttle, B J Swalla, S Taiti, M Tanaka, A H Tandberg, D Tang, M Tasker, J Taylor, J Taylor, A Tchesunov, H ten Hove, J J ter Poorten, J Thomas, E V Thuesen, M Thurston, B Thuy, J T Timi, T Timm, A Todaro, X Turon, S Tyler, P Uetz, S Utevsky, J Vacelet, W Vader, R Väinölä, B de Vijver, S E van der Meij, T van Haaren, R van Soest, R Van Syoc, V Venekey, R Vonk, C Vos, G Walker-Smith, T C Walter, L Watling, M Wayland, T Wesener, C Wetzel, C Whipps, K White, D Williams, G Williams, R Wilson, A Witkowski, J Witkowski, N Wyatt, C Wylezich, M Yasuhara, J Zanol, and W Zeidler. World Register of Marine Species (WoRMS), 2017.

[71] G. R. Houseman. Aggregated seed arrival alters plant diversity in grassland communities. *Journal of Plant Ecology*, 7:51–58, 2014.

[72] M Huxham, S Beaney, and D Raffaelli. Do parasites reduce the chances of triangulation in a real food web? *Oikos*, pages 284–300, 1996.

[73] Mark Huxham, Dave Raffaelli, and Alan Pike. Parasites and food web patterns. *Journal of Animal Ecology*, 64(2):168–176, 1995.

[74] Thomas C Ings and Joseph E Hawes. The history of ecological networks. In *Ecological Networks in the Tropics*, pages 15–28. Springer, 2018.

[75] J. Justus. Loop analysis and qualitative modeling: limitations and merits. *Biology and Philosophy*, 21:647–666, 2006.

[76] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, aug 2011.

[77] Sonia Kéfi, Vincent Miele, Evie A. Wieters, Sergio A. Navarrete, and Eric L. Berlow. How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience. *PLOS Biology*, 14(8):e1002527, 2016.

[78] T. Kohyama. Size-structured tree populations in gap-dynamic forest—the forest architecture hypothesis for the stable coexistence of species. *Journal of Ecology*, 81:131–143, 1993.

[79] T. Kohyama. The effect of patch demography on the community structure of forest

trees. *Ecological Research*, 21:346–355, 2006.

[80] Armand M. Kuris, Ryan F. Hechinger, Jenny C. Shaw, Kathleen L. Whitney, Leopoldina Aguirre-Macedo, Charlie A. Boch, Andrew P. Dobson, Eleca J. Dunham, Brian L. Fredensborg, Todd C. Huspeni, Julio Lorda, Luzviminda Mababa, Frank T. Mancini, Adrienne B. Mora, Maria Pickering, Nadia L. Talhouk, Mark E. Torchin, and Kevin D. Lafferty. Ecosystem energetic implications of parasite and free-living biomass in three estuaries. *Nature*, 454(7203):515–518, 2008.

[81] Kevin D Lafferty, Stefano Allesina, Matias Arim, Cherie J Briggs, Giulio De Leo, Andrew P Dobson, Jennifer A Dunne, Pieter T J Johnson, Armand M Kuris, David J Marcogliese, Neo D Martinez, Jane Memmott, Pablo a Marquet, John P McLaughlin, Erin a Mordecai, Mercedes Pascual, Robert Poulin, and David W Thieltges. Parasites in food webs : the ultimate missing links. *Ecology Letters*, 11(6):533–46, jun 2008.

[82] Kevin D Lafferty, Andrew P Dobson, and Armand M Kuris. Parasites dominate food web links. *Proceedings of the National Academy of Sciences of the United States of America*, 103(30):11211–6, jul 2006.

[83] R. Law and R. D. Morton. Permanence and the assembly of ecological communities. *Ecology*, 77:762–775, 1996.

[84] M. A. Leibold and M. A. McPeek. Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, 87:1399–1410, 2006.

[85] J. M. Levine and J. HilleRisLambers. The importance of niches for the maintenance of species diversity. *Nature*, 461:254–258, 2009.

[86] J. M. Levine and M. Rees. Effects of temporal variability on rare plant persistence in annual systems. *American Naturalist*, 164:350–363, 2004.

[87] R. Levins. *Evolution in changing environments*. Princeton University Press, Princeton, 1968.

[88] R. Levins. Qualitative analysis of partially specified systems. *Ann. NY Acad. Sci.*, 231:123–138, 1974.

[89] R. Levins. Evolution in communities near equilibrium. In M. Cody and J. M. Diamond, editors, *Ecology and Evolution of Communities*, pages 16–50. Harvard University Press, Cambridge, 1975.

[90] Thomas M Lewinsohn, Paulo Inácio Prado, Pedro Jordano, Jordi Bascompte, and Jens M Olesen. Structure in plant–animal interaction assemblages. *Oikos*, 113(1):174–184, 2006.

[91] Joseph J Luczkovich, Stephen P Borgatti, Jeffrey C Johnson, and Martin G Everett. Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of Theoretical Biology*, 220(3):303–321, 2003.

[92] R. H. MacArthur. Species packing and competitive equilibria for many species. *Theoretical Population Biology*, 1:1–11, 1970.

[93] R. H. MacArthur. *Geographical ecology*. Harper & Row, New York, 1972.

[94] R. H. MacArthur and R. Levins. Limiting similarity, convergence, and divergence of

coexisting species. *American Naturalist*, 101(921):377–385, 1967.

[95] J. Mallet. The struggle for existence: how the notion of carrying capacity, k, obscures the links between demography, darwinian evolution, and speciation. *Evolutionary Ecology Research*, 14:627–665, 2012.

[96] David J Marcogliese. Food webs and biodiversity: are parasites the missing link? *Journal of Parasitology*, 89(6):S106–S113, 2003.

[97] David J Marcogliese and David K Cone. Food webs: a plea for parasites. *Trends in Ecology & Evolution*, 12(8):320–324, 1997.

[98] R. M. May. *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, 1973.

[99] R. M May and W. J. Leonard. Nonlinear aspects of competition between three species. *SIAM Journal on Applied Mathematics*, 29:243–253, 1975.

[100] Robert M May. *Stability and complexity in model ecosystems*. Princeton University Press, 1973.

[101] Daniel S Maynard, Carlos A Serván, and Stefano Allesina. Network spandrels reflect ecological assembly. *Ecology letters*, 21(3):324–334, 2018.

[102] B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10:995–1015, 2007.

[103] Jane Memmott. The structure of a plant-pollinator food web. *Ecology letters*, 2(5):276–280, 1999.

[104] G. Meszéna, M. Gyllenberg, L. Pásztor, and J. A. J. Metz. Competitive exclusion and limiting similarity: a unified theory. *Theoretical Population Biology*, 69:68–87, 2006.

[105] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[106] G. G. Mittelbach. *Community Ecology*. Sinauer Associates, Sunderland, Massachusetts, USA, 2012.

[107] Akihiko Mougi and Michio Kondoh. Diversity of interaction types and ecological community stability. *Science*, 337(6092):349–351, 2012.

[108] K. N. Mouritsen, R. Poulin, J. P. McLaughlin, and D. W. Thieltges. Food web including metazoan parasites for an intertidal ecosystem in New Zealand: Ecological Archives E092-173. *Ecology*, 92(10):2006–2006, 2011.

[109] D. J. Murrell. When does local spatial structure hinder competitive coexistence and reverse competitive hierarchies? *Ecology*, 91:1605–1616, 2010.

[110] M. G. Neubert and H. Caswell. Alternatives to resilience for measuring the responses of ecological systems to perturbations. *Ecology*, 78:653–665, 1997.

[111] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of*

the national academy of sciences, 103(23):8577–8582, 2006.

[112] Mark EJ Newman. Spectral methods for community detection and graph partitioning. *Physical Review E*, 88(4):042822, 2013.

[113] John Novembre, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R Boyko, Adam Auton, Amit Indap, Karen S King, Sven Bergmann, Matthew R Nelson, et al. Genes mirror geography within europe. *Nature*, 456(7218):98, 2008.

[114] M. Opper and S. Diederich. Phase transition and $1/f$ noise in a game dynamical model. *Physical Review Letters*, 69:1616–1619, 1992.

[115] Sean O'Rourke and David Renfrew. Low rank perturbations of large elliptic random matrices. *Electronic Journal of Probability*, 19:1–65, 2014.

[116] Owen L Petchey, Andrew P Beckerman, Jens O Riede, and Philip H Warren. Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11):4191–4196, 2008.

[117] Evelyn C Pielou. Species-diversity and pattern-diversity in the study of ecological succession. *Journal of theoretical biology*, 10(2):370–383, 1966.

[118] SL Pimm. Food webs, 2002.

[119] Stuart L Pimm. The structure of food webs. *Theoretical population biology*, 16(2):144–158, 1979.

[120] Stephen R Proulx, Daniel EL Promislow, and Patrick C Phillips. Network thinking in ecology and evolution. *Trends in ecology & evolution*, 20(6):345–353, 2005.

[121] Haseeb Sajjad Randhawa and Robert Poulin. Determinants and consequences of interspeci W c body size variation in tetraphyllidean tapeworms. *Oecologia*, 161:759–769, 2009.

[122] R. P. Rohr, S. Saavedra, and J. Bascompte. On the structural stability of mutualistic systems. *Science*, 345, 2014.

[123] Sonia Romero-Romero, Axayacatl Molina-Ramírez, Juan Höfer, and José Luis Acuña. Body size-based trophic structure of a deep marine ecosystem. *Ecology*, 97(1):171–181, 2016.

[124] J. Roughgarden. *Theory of Population Genetics and Evolutionary Ecology: an Introduction*. Macmillan, New York, 1979.

[125] Serguei Saavedra, Rudolf P Rohr, Jens M Olesen, and Jordi Bascompte. Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and evolution*, 6(4):997–1007, 2016.

[126] Elizabeth L Sander, J Timothy Wootton, and Stefano Allesina. What can interaction webs tell us about species roles? *PLoS computational biology*, 11(7):e1004330, 2015.

[127] Elizabeth L Sander, J Timothy Wootton, and Stefano Allesina. What can interaction webs tell us about species roles? *PLoS Computational Biology*, 11(7):1–22, 2015.

[128] A. M. Siepielski, K. L. Hung, E. E. B. Bein, and M. A. McPeek. Experimental evidence for neutral community dynamics governing an insect assemblage. *Ecology*, 91:847–857, 2010.

[129] A. M. Siepielski and M. A. McPeek. On the evidence for species coexistence: a critique of the coexistence program. *Ecology*, 91:3153–3164, 2010.

[130] S. Smale. On the differential equations of species in competition. *Journal of Mathematical Biology*, 3:5–7, 1976.

[131] Tom A. B. Snijders and Nowicki Krzysztof. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14:75–100, 1997.

[132] Ricard V Solé and Sergi Valverde. Are network motifs the spandrels of cellular complexity? *Trends in ecology & evolution*, 21(8):419–422, 2006.

[133] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein. Spectrum of large random asymmetric matrices. *Physical Review Letters*, 60:1895–1898, 1998.

[134] P. P. A. Staniczenko, J. C. Kopp, and S. Allesina. The ghost of nestedness in ecological networks. *Nature Communications*, 4:1391, 2013.

[135] Phillip Staniczenko, Matthew J Smith, and Stefano Allesina. Selecting food web models using normalized maximum likelihood. *Methods in Ecology and Evolution*, 5(6):551–562, 2014.

[136] Phillip PA Staniczenko, Jason C Kopp, and Stefano Allesina. The ghost of nestedness in ecological networks. *Nature communications*, 4:1391, 2013.

[137] Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

[138] Giovanni Strona, Domenico Nappo, Francesco Boccacci, Simone Fattorini, and Jesus San-Miguel-Ayanz. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature communications*, 5:4114, 2014.

[139] Giovanni Strona and Joseph A Veech. A new measure of ecological network structure based on node overlap and segregation. *Methods in Ecology and Evolution*, 6(8):907–915, 2015.

[140] Y. M. Svirezhev and D. O. Logofet. *Stability of Biological Communities*. Mir Publishers, Moscow, Russia, 1983.

[141] S. Tang and S. Allesina. Reactivity and stability of large ecosystems. *Frontiers in Ecology and Evolution*, 2014.

[142] Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

[143] D. W. Thieltges, K. Reise, K. N. Mouritsen, J. P. McLaughlin, and R. Poulin. Food web including metazoan parasites for a tidal basin in Germany and Denmark: Ecological Archives E092-172. *Ecology*, 92(10):2005–2005, 2011.

[144] John N Thompson. *The coevolutionary process*. University of Chicago Press, 1994.

[145] John N Thompson. *The geographic mosaic of coevolution*. University of Chicago Press, 2005.

[146] Ross M Thompson, Kim N Mouritsen, and Robert Poulin. Importance of parasites

and their life cycle characteristics in determining the structure of a large marine food web. *Journal of Animal Ecology*, 74:77–85, 2005.

[147] Peter H Thrall, Michael E Hochberg, Jeremy J Burdon, and James D Bever. Coevolution of symbiotic mutualists and parasites in a community context. *Trends in Ecology & Evolution*, 22(3):120–126, 2007.

[148] K. Tokita. Species abundance patterns in complex evolutionary dynamics. *Physical Review Letters*, 93:178102, 2004.

[149] Werner Ulrich, Mário Almeida-Neto, and Nicholas J Gotelli. A consumer's guide to nestedness analysis. *Oikos*, 118(1):3–17, 2009.

[150] J. H. Vandermeer. Interspecific competition: A new approach to the classical theory. *Science*, 188:253–255, 1975.

[151] J. H. Vandermeer and D. E. Goldberg. *Population ecology: First principles. 2nd edition.* Princeton University Press, Princeton, New Jersey, 2013.

[152] J. A. Vano, J. C. Wildenberg, M. B. Anderson, J. K. Noel, and J. C. Sprott. Chaos in low-dimensional Lotka–Volterra models of competition. *Nonlinearity*, 19:2391–2404, 2006.

[153] Richard J Williams and Neo D Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–3, mar 2000.

[154] David H Wright and Jaxk H Reeves. On the meaning and measurement of nestedness of species assemblages. *Oecologia*, 92(3):416–428, 1992.

[155] Y. Yoshino, T. Galla, and K. Tokita. Rank abundance relations in evolutionary dynamics of random replicators. pages arXiv:0803.1082v2 [q–bio.PE], 2008.

[156] C. D. Zander, N. Josten, K. C. Detloff, R. Poulin, J. P. McLaughlin, and D. W. Thieltges. Food web including metazoan parasites for a brackish shallow water ecosystem in Germany and Denmark: Ecological Archives E092-174. *Ecology*, 92(10):2007–2007, 2011.

[157] M. L. Zeeman. Hopf bifurcations in competitive three-dimensional Lotka–Volterra systems. *Dynamics and Stability of Systems*, 8:189–216, 1993.