

# Revealing biases in the sampling of ecological interaction networks

Marcus A.M. de Aguiar<sup>1</sup>, Erica A. Newman<sup>2</sup>, Mathias M. Pires<sup>3</sup>, Justin D. Yeakel<sup>4,5</sup>, Carl Boettiger<sup>6</sup>, Laura A. Burkle<sup>7</sup>, Dominique Gravel<sup>8</sup>, Paulo R. Guimarães Jr<sup>9</sup>, James L. O'Donnell<sup>10</sup>, Timothée Poisot<sup>11,12</sup>, Marie-Josée Fortin<sup>13</sup> and David H. Hembry<sup>2,14</sup>

<sup>1</sup> Instituto de Física “Gleb Wataghin”, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

<sup>2</sup> Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

<sup>3</sup> Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

<sup>4</sup> School of Natural Sciences, University of California, Merced, CA, USA

<sup>5</sup> Santa Fe Institute, Santa Fe, NM, USA

<sup>6</sup> Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA

<sup>7</sup> Department of Ecology, Montana State University, Bozeman, MT, USA

<sup>8</sup> Département de Biologie, Université de Sherbrooke, Sherbrooke, QC, Canada

<sup>9</sup> Departamento de Ecología, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

<sup>10</sup> School of Marine and Environmental Affairs, University of Washington, Seattle, WA, USA

<sup>11</sup> Département de Sciences Biologiques, Université de Montréal, Montréal, QC, Canada

<sup>12</sup> Québec Centre for Biodiversity Sciences, Montréal, QC, Canada

<sup>13</sup> Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

<sup>14</sup> Department of Entomology, Cornell University, Ithaca, NY, USA

## ABSTRACT

The structure of ecological interactions is commonly understood through analyses of interaction networks. However, these analyses may be sensitive to sampling biases with respect to both the interactors (the nodes of the network) and interactions (the links between nodes), because the detectability of species and their interactions is highly heterogeneous. These ecological and statistical issues directly affect ecologists' abilities to accurately construct ecological networks. However, statistical biases introduced by sampling are difficult to quantify in the absence of full knowledge of the underlying ecological network's structure. To explore properties of large-scale ecological networks, we developed the software *EcoNetGen*, which constructs and samples networks with predetermined topologies. These networks may represent a wide variety of communities that vary in size and types of ecological interactions. We sampled these networks with different mathematical sampling designs that correspond to methods used in field observations. The observed networks generated by each sampling process were then analyzed with respect to the number of components, size of components and other network metrics. We show that the sampling effort needed to estimate underlying network properties depends strongly both on the sampling design and on the underlying network topology. In particular, networks with random or scale-free modules require more complete sampling to reveal their structure, compared to networks whose modules are nested or bipartite. Overall, modules with nested structure were the easiest to detect, regardless of the sampling design used. Sampling a network starting with any species that had a high

Submitted 23 April 2019

Accepted 29 July 2019

Published 2 September 2019

Corresponding authors

Marcus A.M. de Aguiar,  
aguiar@ifi.unicamp.br  
Erica A. Newman,  
newmane@email.arizona.edu

Academic editor

Leonardo Gollo

Additional Information and  
Declarations can be found on  
page 18

DOI 10.7717/peerj.7566

© Copyright  
2019 de Aguiar et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

degree (e.g., abundant generalist species) was consistently found to be the most accurate strategy to estimate network structure. Because high-degree species tend to be generalists, abundant in natural communities relative to specialists, and connected to each other, sampling by degree may therefore be common but unintentional in empirical sampling of networks. Conversely, sampling according to module (representing different interaction types or taxa) results in a rather complete view of certain modules, but fails to provide a complete picture of the underlying network. To reduce biases introduced by sampling methods, we recommend that these findings be incorporated into field design considerations for projects aiming to characterize large species interaction networks.

**Subjects** Biodiversity, Ecology, Ecosystem Science, Mathematical Biology, Data Science

**Keywords** Ecological networks, Modularity, Network metrics, Species interaction networks, Network topology, Nestedness, Food webs, Field sampling design

## INTRODUCTION

Network theory provides an efficient way to represent and characterize the structure of ecological systems by organizing the complex relationships between species as graphs, where nodes represent the species, and links represent their interactions (*Pascual & Dunne, 2006*). Modularity (or compartmentalization) is a tendency of networks to form clusters of highly connected nodes, with weaker connections among clusters, which can be revealed using analytical methods (*Marquitti et al., 2014*). These clusters are referred to as modules or compartments. Modularity is a common property of ecological interaction networks (*Newman, 2006; Olesen et al., 2007*), but is a broad concept that can apply to clustering among nodes in any type of network. In an ecological network, modularity may arise within communities of co-occurring species or describe clustering in space, may or may not correspond to taxonomic or functional groups of species, and may also arise in intraspecific interactions (e.g., family groups). In this paper, we primarily focus on modular networks that represent the clustering of species and the interactions between them.

Empirical networks often focus on a limited subset of interacting species (in practice corresponding to one or a few modules), as field observations used to construct species interaction networks can be effort-intensive. Furthermore, a field ecologist may attempt to exhaustively sample the species interacting in a delimited area, while ignoring interactions and species that occur outside that area. These approaches leave many species and interactions unsampled, both in the community and in adjacent communities, and may therefore leave a large number of species or entire modules that interact with sampled species undetected.

Empirical networks may have limitations other than unsampled links or nodes. For example, the total size of the empirical network may influence how much of the underlying topology is sampled, and therefore, what we conclude about the overall topology of an interaction network. Because empirical interaction networks are often constructed with a focus on a given type of interaction and by sampling interactions of a particular taxonomic

**Table 1 Variations in structure among ecological interaction networks.** Network topology can vary greatly from one part of the network to another, and influence the conclusions drawn about the underlying network. For example, interactions among certain groups of species form subnetworks characterized by high degrees of modularity and reciprocal specialization, as is the case with some ant-myrmecophyte networks (Guimarães *et al.*, 2007), clownfish-anemone networks (Fautin & Allen, 1997; Ollerton *et al.*, 2007), and other networks where interactions are symbiotic (Hembry *et al.*, 2018). Conversely, mutualisms such as those between plants and their pollinators or seed dispersers (Bascompte *et al.*, 2003), or the interactions between generalized predators or herbivores with the resources they consume (Pires & Guimarães, 2013) are highly nested, where the interactions of specialists are a nested subset within those of generalists. Yet, as we look at broader scales that include multiple habitats, taxonomic groups, and/or interaction types, a modular organization tends to emerge (Olesen *et al.*, 2007; Donatti *et al.*, 2011), with each module having unique structural properties (Lewinsohn *et al.*, 2006; Fontaine *et al.*, 2011). More complete ecological networks may emerge from the aggregation of multiple types of interactions, as well as the various and sometime unique structures such interactions form, and can be represented as large, modular networks. However, it is unknown whether and to what extent different sampling strategies might bias our understanding of the underlying network structure (Jordano, 2016; Fründ, McCann & Williams, 2016; Vizentin-Bugoni *et al.*, 2016). Below, simulated modules represent certain types of ecological interactions, each of which are known to have different associated structural characteristics. These structures may provide null models for empiricists, or guide initial guesses at true, underlying network structure. Individual module structure is considered, though real networks may be composed of many modules of varying types.

Module structure	Ecological interaction type commonly represented
Random	Null model or random interactions
Scale-free	Null model for preferential attachment
Nested	Consumer–resource interactions (e.g., predator–prey interactions)
Bipartite nested	Consumer–resource interactions with nonoverlapping sets (e.g., plant–pollinator interactions)
Bipartite random	Null model for plant–pollinator interactions
Tripartite nested	Plant–pollinator interactions with added nested trophic level, such as birds–plant–bats where bird–plants and bat–plants are nested
Tripartite random	Plant–pollinator interactions with added random trophic level, such as birds–plant–bats where bird–plants are nested and bat–plants are random

group within a locality (Hall & Raffaelli, 1993; Bascompte & Jordano, 2007), the largest empirical ecological interaction networks typically include no more than a few hundred species; often many fewer. Nevertheless, these empirical networks represent subnetworks within a more complete ecological network. The underlying network will also necessarily include many more species interacting in multiple qualitatively different ways (Fontaine *et al.*, 2011; Pilosof *et al.*, 2017). For example, a plant–pollinator network focusing on insects is one or a few modules of a larger network that includes pollinators from other taxonomic groups, as well as the consumers of these species and their parasites, and so on (Table 1).

Establishing the relationships between the topology of a sampled network and the true, underlying network is a fundamental challenge in community ecology as a whole: establishing the boundaries of the system of interest (Morin, 2009). Previous studies have examined the effects of sampling on the inferred structure of networks (Jordano, 1987, 2016; Stang *et al.*, 2009; Gibson *et al.*, 2011; Cirtwill *et al.*, 2019). Network structure, in particular the relative abundance of species, may be affected by species specialization (Blüthgen, Menzel & Blüthgen, 2006). Detectability of ecological interactions is also known to influence what network topology is inferred from empirical sampling (Bartomeus, 2013; Graham & Weinstein, 2018). Although there are many known issues with empirical sampling of networks, an idealized mathematical approach can lend insights into what underlying biases are introduced to inferred network structure through sampling.

In many areas of complex systems, sampling is critically important, because the properties of massive graphs with millions or even billions of nodes simply cannot be computed (Leskovec & Faloutsos, 2006; Gerhard *et al.*, 2011; Hu & Lau, 2013;

([Levina & Priesemann, 2017](#)). Several methods have been developed recently to ensure that the properties of the sampled networks are similar to those of the original one. It has been found, in particular, that scale-free networks, when randomly sampled, do not produce a scale-free sampled network ([Stumpf & Wiuf, 2005](#)). It has also been found that sampling by random walks among nodes performs well in most cases, and that sampling by edges does not perform well in reproducing the structure of the original network ([Leskovec & Faloutsos, 2006](#)). Sampling of highly modular networks, however, has not yet been considered.

Indeed, the study of empirical ecological networks relies on the reasonable assumption that the ecological and evolutionary dynamics of each module (which may be considered a “compartment,” or subnetwork) can usually be investigated independently ([Lewinsohn et al., 2006](#)). Yet there are situations in which neglecting the effects of other interactions and species outside the delineated boundaries may lead to incomplete or incorrect conclusions ([Ings et al., 2009](#); [Fontaine et al., 2011](#); [Mello et al., 2011a, 2011b](#); [Rivera-Hutinel et al., 2012](#); [Fründ, McCann & Williams, 2016](#)). The increased interest in larger ecological networks encompassing several groups and types of networks, such as multilayered networks ([Pilosof et al., 2017](#)), and development of tools for their study makes this problem especially timely for ecology.

The question is then: how much of the underlying, complete ecological network can be observed by sampling a subset of its species (nodes) and their associated interactions (links)? The effectiveness of field sampling in capturing the underlying complete network may depend on (1) the underlying topology of the complete network, (2) the sampling technique itself ([Hegland et al., 2010](#); [Gibson et al., 2011](#)), and (3) the potential interplay between the network topology and the sampling strategy.

We investigated the above questions with the use of a new software, *EcoNetGen*, developed initially for this project in Python using Fortran, and now available in the R programming language ([De Aguiar et al., 2019](#)) and on CRAN. *EcoNetGen* contains the script *NetGen* (*netgen* in R), which generates interaction networks with predetermined properties including network size, structure of the modules within the network (or structure of the overall network, if it is considered to be a single module), as well as the frequency of modules with particular structures. *EcoNetGen* also contains the script *NetSampler* (*netsampler* in R), which samples sets of nodes from the full network according to a chosen technique, and then compares the observed network against the full, complete network. With the use of simulated networks, we have full knowledge of the size and structure of the underlying, complete network, and we can directly compare it to the size and structure of the sampled object.

With *EcoNetGen*, we examined how the interplay between network structure and sampling design alters our inference on the among-module connectedness in networks, and were able to draw conclusions about the sampling design that captures the most accurate picture of the complete network for a given topology. This allowed us to evaluate how different sampling strategies often used in the field might alter observational accuracy, identify whether specific sampling designs produce more reliable estimates of the underlying network structure, and determine to what extent such designs

can be confounded or enhanced by alternative arrangements of the underlying species interactions.

To better understand the factors that may affect the match between the topology of the sampled and underlying networks, we focused on simulated networks in which we can control the initial structure and properties such as size and connectance. We also focused on simplified modules that depict well-defined structures instead of trying to encompass all the variability that can be found in nature. The investigations presented here are theoretical investigations of a range of possible network topologies, rather than a comparison to field-sampled data, as any empirical network available for analyses is already a sampled version and contains the effects of sampling biases. By building upon simplified structures we have greater control over which variables may be affecting network properties. We expect that the range of network topologies illustrated here can adequately describe many empirically observed patterns and give insight into the sampling of real networks.

Our findings are fourfold. First, both the underlying pattern of species interactions and the strategy used to sample them had a large impact on the observed network structure. Sampling species according to the number of their interactions (starting with any node of high degree and constructing the sampled network outward from that node) consistently resulted in more accurate estimates of the underlying network structure. Second, through the use of simulated networks, we found that nested sets of interactions are easier to detect regardless of sampling strategy (consistent with [Nielsen & Bascompte, 2007](#)). Third, we found that the size of the observed network (measured as the fraction of the sampled network contained in its largest connected component, or “relative size of the largest connected component”) did not depend significantly on the sampling method, but does depend strongly on the underlying network topology. Fourth and finally, sampling according to module membership can produce good estimates of the structure within individual modules, but increases the risk of missing entire modules of species interactions altogether. Because all sampling schemes have some limitations, we recommend that empiricists consider an iterative sampling approach, where the sampling strategy can be adjusted as network properties are revealed.

## MATERIALS AND METHODS

### Generating networks

Networks with a specified number of modules (one or more) and a variety of module topologies can be constructed with the software script *NetGen* (*netgen* in R) in *EcoNetGen* (details and examples in [Appendix S1](#); scripts can be found in [Appendices S5](#) and [S6](#)). Networks with a single module are equivalent to a nonmodular network with a single topology, and networks with multiple modules can be constructed with each module having a specified topology. We assume here that the complete networks we generate can represent an ecological system encompassing multiple interacting taxonomic groups, and different habitats and guilds here represented as network modules ([Table 1](#)). Modularity may also be observed in interaction networks encompassing different types of interactions, but the results from *EcoNetGen* (which does not contain weighted links or interaction

types) must be interpreted with care for these systems. In the modeled ecological interaction network, inter-module interactions are those that indirectly connect a module or guild with another (e.g., an interaction between bees and a plant that is mostly pollinated by hummingbirds), or interactions of a species that inhabits two habitat types over its lifetime (e.g., amphibians, who live part of their life cycle in an aquatic subnetwork and the other part in a terrestrial one), or are species interactions involving species that connect different interaction modules (e.g., butterflies, which can shift over ontogeny from herbivores on certain plant species to pollinators of different plant species).

*EcoNetGen* allows the construction of modular networks where the total network size (or total number of nodes,  $N$ ), the average module size ( $M_{av}$ ), the average degree of the nodes ( $k$ ), and topology of modules can be controlled. The simulated network can have either one or multiple modules, the sizes of which ( $M_i$ ) are drawn from a negative exponential distribution with average value  $M_{av}$ . Modules can have different topologies: random, scale-free, nested, bipartite nested, bipartite random, tripartite nested, and tripartite random (Appendix S1: Fig. A1). Networks may be uniform, such that all modules have similar structures (e.g., all scale-free) or may contain modules of various topologies (e.g., a combination of random and nested). When the generated network is specified to contain modules of multiple types (i.e., “mixed modules”), each module type is randomly chosen with given probabilities. A generated network can be specified to contain modules of certain types if the probability of the other module types is set to zero. Once the modules have been created, the links between nodes within each module can be changed, or rewired, with specified probabilities ( $p_{local}$ ) to randomize the initial structures. The nodes of the full network can then further be rewired (with probability  $p_{rew}$ ) to create connections among the modules.

For this study, we generated sets of modular networks with five different module topologies: random, scale-free, nested, bipartite nested, and mixed. Although all network generation parameters can be specified by the user, we chose to fix the total network size at  $N = 500$ , with average module size of 25, and average node degree of  $k = 10$  in order to reduce the number of parameters in our analyses. Though the average degree is fixed, the degree distribution can vary significantly depending on module type. Once the modules were constructed according to a given algorithm, nodes were randomly rewired to other nodes within the module with probability  $p_{local} = 0.1$ . This value was chosen to preserve the identity of the modules, but remove their “exact” algorithmic form. Nodes were further rewired to any node of the network with probability  $p_{rew} = 0.1$ , to create connections among modules.

### Sampling simulated networks

The motivation for examining different sampling designs applied to a full network is to explore how the most common practices used by a researcher with limited time or resources will affect the conclusions they draw about the underlying network structure. The sampling procedure, carried out in the script *NetSampler* (*netsampler* in R), consists of picking  $m$  nodes that anchor the construction of the observed network, and then adding a number of first neighbor nodes ( $nfn$ , for “number of first neighbors”) to each of these

“anchor” (or “anchoring”) nodes ([Appendix S2](#)). Such a sampling design emulates a researcher studying a particular set of  $m$  species, and subsequently identifying those species that interact with the original set, as is often done when sampling animal–plant interactions (see [Jordano, 2016](#)). The anchoring nodes and their neighbors can be chosen in different ways, as described below. We emphasize that only the observed interactions between nodes are included in the observed (or sampled) network. Therefore, two anchoring nodes that are connected in the original network will be connected in the sampled network only if one of the nodes is selected as a first neighbor of the other in the sampling process. In other words, an existing link between anchoring nodes is not automatically passed to the observed network. In sampling the full network, if an anchoring node is selected more than once, it is only considered once in the network.

### Sampling anchoring nodes

Once a complete network is constructed, the anchoring nodes that will be the first points of sampling can be chosen according to different criteria. Anchoring nodes can be chosen at random, according to the node’s degree (the number of interactions), according to abundances that are attributed to the nodes, or by attributing weights to each module such that species in one module (representing a particular interaction type or a taxonomic group) can be more or less likely to be included in the sampled network over species in other modules. Once the anchoring nodes have been chosen, the sampled network is constructed outward from them in a procedure that adds some of the nearest first neighbors of those nodes (described in *Sampling interactions: choosing the neighbors of the anchoring nodes*).

Sampling of anchoring nodes is implemented through *NetSampler* (*netsampler* in R). An example of sampling with *NetSampler* (*netsampler* in R) is given in [Appendix S2](#), and the mathematical forms of these sampling distributions are available in [Appendix S3](#).

1. Random:  $m$  attempts to select nodes at random from the network are performed. The actual number of distinct anchoring nodes might turn out to be smaller than  $m$ , because the same node can be selected more than once. This sampling design represents a benchmark with which other sampling methods can be compared.
2. Degree of the node,  $k$ : the probability that a node is selected is proportional to its degree. The higher the degree of the node, the higher the chances are that it will be included in the observed network. Again,  $m$  attempts are made, but fewer than  $m$  nodes might actually be included. The reasoning for such a sampling process is that a field biologist could choose to study a generalist species, whose interactions might be more likely to be observed because degree is sometimes correlated with abundance ([Vázquez et al., 2009](#)). Although in this sampling design we make no particular assumption about abundances, this relationship could be thought of as the underlying reason in practice why interactions of species with higher degree may be more easily detected.
3. Abundances: an abundance value is attributed to each species (node) following three possible distributions: exponential, Fisher log-series, and lognormal (specified in *NetGen* (*netgen* in R), see equations C1, C6, and C8 in [Appendix S3](#)). In field sampling, any

sampling by transect, by plot, or by timed observations will be equivalent to sampling according to a species-abundance distribution. Species-abundance distributions often have the form of a log-normal or Fisher log-series distribution ([McGill et al., 2007](#); [White, Thibault & Xiao, 2012](#); [Harte & Newman, 2014](#); [Baldridge et al., 2016](#); [Newman et al., 2018](#)). Once the abundances have been attributed,  $m$  attempts to select nodes are made, and the probability that a node is selected is proportional to its abundance.

We note that abundances are assigned in the generation of the networks (as a specific abundance distribution), if and only if the networks are to be sampled by their abundances. In this sense, we consider the abundance distributions to be part of the sampling routine. Abundances are attributed to each module independently and are not correlated to the degree of nodes or any other network properties (if this is desired, a user can upload their own matrix with other specified properties to use with *NetSampler* (*netsampler* in R)). This simulates a sampling process where the likelihood of sampling depends on abundances, and therefore favors the most abundant species of each module to be selected as anchoring nodes. This scheme differs from random sampling, where nodes have the same chance of being selected irrespective of their module membership. Here each module is likely to have an anchoring node represented by its most abundant species. The process thus promotes uniform sampling across modules, and random sampling within modules.

Restated in a slightly different way, the probability of selecting an anchor node is not deterministic, that is, the most abundant nodes are not necessarily selected first. The sampling of a node is proportional to its abundance, so that the most abundant nodes are more likely to be selected in any given abundance distribution.

4. Module: sampling probabilities are assigned to the network modules, and within each module the probabilities associated with the nodes are uniform. In this way, species in some modules have a higher probability of being sampled than those in other modules, while the sampling probability is uniform within a given module. This incorporates the notion that some groups of species are easier to observe than others or that some researchers focus on particular types of interactions or taxonomic groups (see equation C9 in [Appendix S3](#)).

### Sampling interactions: choosing the neighbors of the anchoring nodes

Once the anchoring nodes have been selected, a subset of their interactions is sampled from the complete network to construct the observed network. Interactions are sampled in two ways: by specifying a maximum  $nfn$  ( $nfn$  is an integer and  $>1$ ), or by specifying a fraction of the total number of neighbors per node ( $nfn < 1$ ). Similar to the parameter  $m$  (the number of anchoring nodes),  $nfn$  specifies the number of attempts to include neighbors: if a neighbor is selected twice, one attempt is lost. This is analogous to performing field observations for a limited time and observing several interactions between the same pair of species ([Jordano, 2016](#)). If  $nfn = 4$ , for example, a node with two links will very likely have its two neighbors included, whereas a node with eight links will have at

most four of its neighbors included (with a range of one to four neighbors actually included). If  $nfn = 0.5$ , on the other hand, the number of attempts per node is equal to half the number of its neighbors. Because  $nfn$  is defined in this way, we define adding all neighbors as  $nfn = 1$ , and distinguish it in *NetGen* (*netgen* in R) code from adding a single neighbor (which is specified by  $nfn = 1.1$ ). Once the method for sampling interactions has been chosen, the actual neighboring nodes can then either be selected: (1) with uniform probability, or (2) with varying probabilities following an exponential distribution. In the latter case, these probabilities can be thought of as weights that represent interaction frequency, abundance, or a convolution of the two (Vázquez, Morris & Jordano, 2005).

For each network, we sampled  $m$  anchoring nodes and randomly added  $nfn$  first neighbors for  $m = 10, 20, \dots, 100$  and  $nfn = 5$  or 10. This process simulates the sampling design used to build interaction networks from field data, where only a subset of species is repeatedly surveyed for their interactions. To demonstrate these methods, we performed 1,000 replicates of sampling according to each scheme described above on each generated network.

### Network metrics

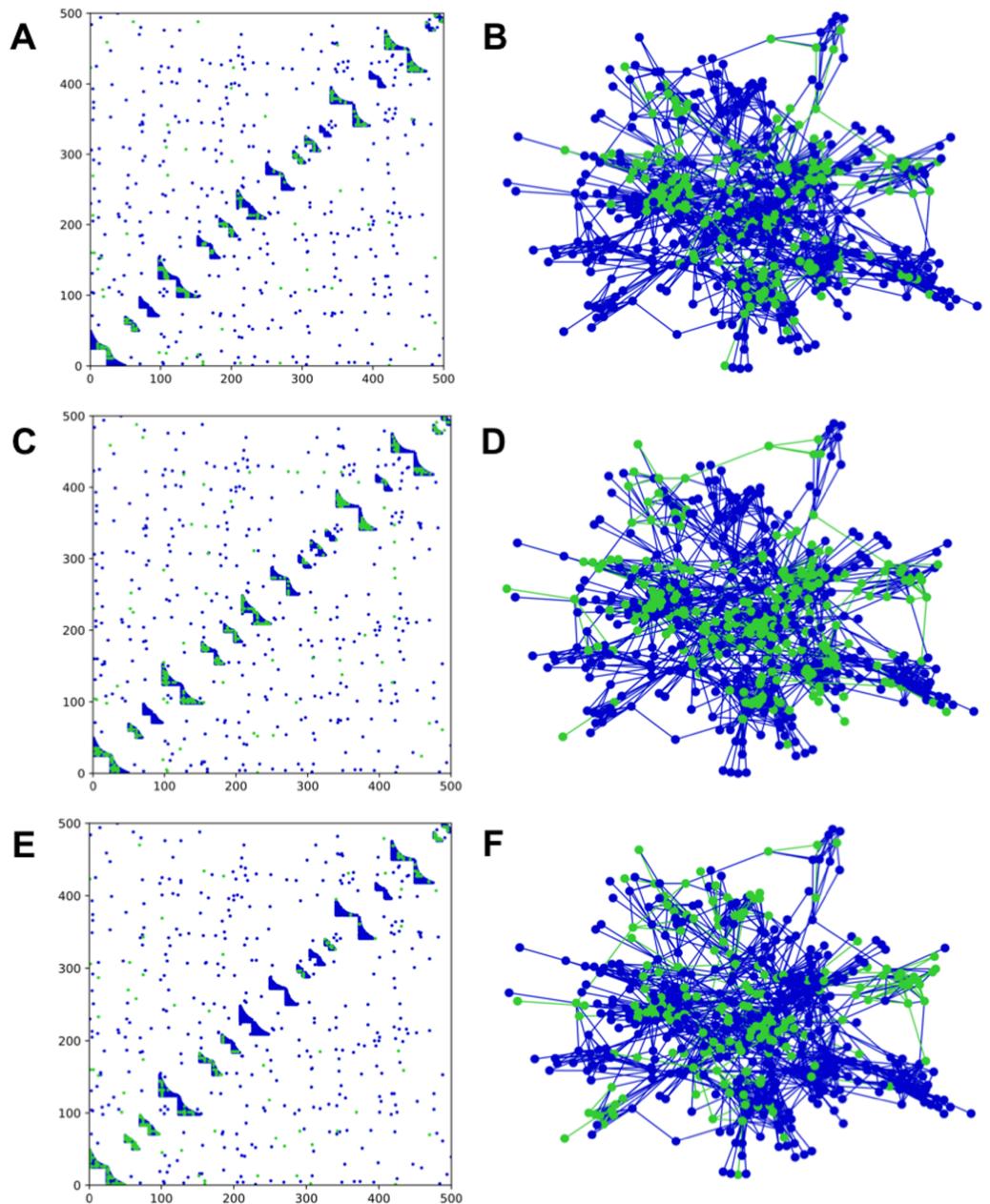
The number of components of the sampled network, together with the size distribution of these components, measure how well the between-module connectedness has been captured by the sampling procedure. Ideally a single component should emerge, matching the complete network. Therefore, for each sampling design, we calculated: (a) the size of the sampled network, i.e., the total number of sampled nodes; (b) the number of components of the sampled network and; (c) the size of the largest component divided by the size of the sampled network, i.e., the relative size of the largest component (RSLC). This last quantity measures the fraction of sampled network contained in its largest connected component.

Because we were interested in the overall topology of the network, we focused on metrics describing the size and number of components instead of assessing the internal structure of each component. Because the sizes of most components were typically small, measures such as average degree, clustering, average path length, or degree distribution would provide much information about the observed structures. However, since degree distribution is such a basic descriptor of networks, we explored how the degree of sampled nodes (in the observed network) correlate with their value in the complete network for different network topologies and sampling strategies.

## RESULTS

We investigated how the incomplete sampling of large networks formed by several modular structures affected conclusions drawn about the underlying network structure, depending on the structure of the network, sampling intensity, sampling procedure, and the interaction between sampling procedure and network structure.

We interacted three sampling methods (random, degree, and module) with nested bipartite modules, shown in Fig. 1. Bipartite networks were generated with *NetGen* (*netgen* in R) and the resulting sampled networks created by *NetSampler* (*netsampler* in R). From these simulations, we found that sampling by module can leave entire modules hidden

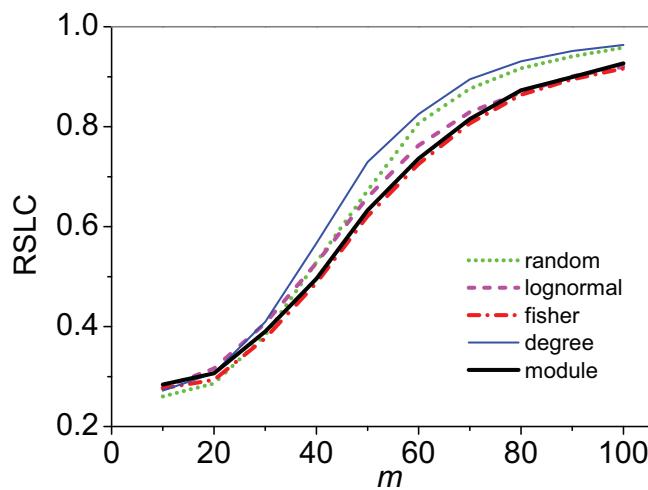


**Figure 1** Adjacency matrices and network structure for a network with bipartite nested modules. Sampling occurred on  $m = 50$  anchoring nodes, adding up to 10 nearest first neighbors. The complete network has 16 modules with average degree of 7.5 and average module size of 31.25. Anchoring nodes were chosen randomly in (A) and (B); according to degree in (C) and (D); or by module preference in (E) and (F). Nodes and links in green represent the sampled species and interactions in each case. The number of connected components in each case is 12, 6, and 13, respectively.

Full-size DOI: 10.7717/peerj.7566/fig-1

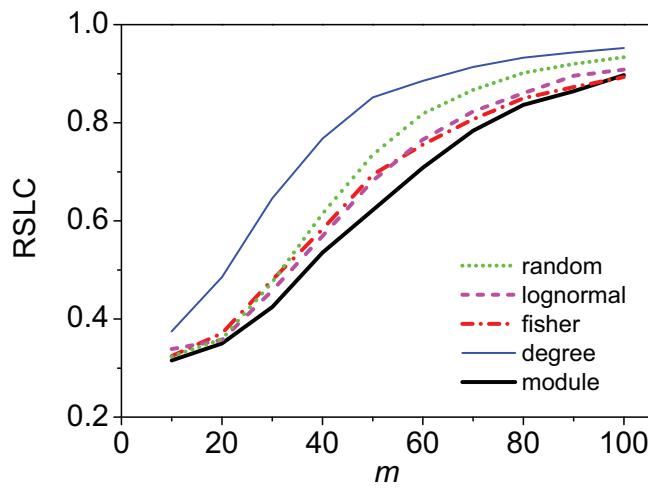
from the observer. Analogous results for a network with mixed modules are shown in Appendix S4.

Clustering of sampled nodes is apparent and visually different between sampling methods, which can be analyzed according to network metrics described in the ‘Methods’



**Figure 2** Relative size of the largest component (RSLC) analysis for a bipartite, nested network. RSLC is shown for a network with bipartite nested modules as a function of  $m$  (the number of anchoring nodes, which is the number of attempts to select nodes at random from the network) for  $nfn = 10$  (where  $nfn$  is maximum number of first neighbors added to the anchoring nodes), with line color indicating sampling design.

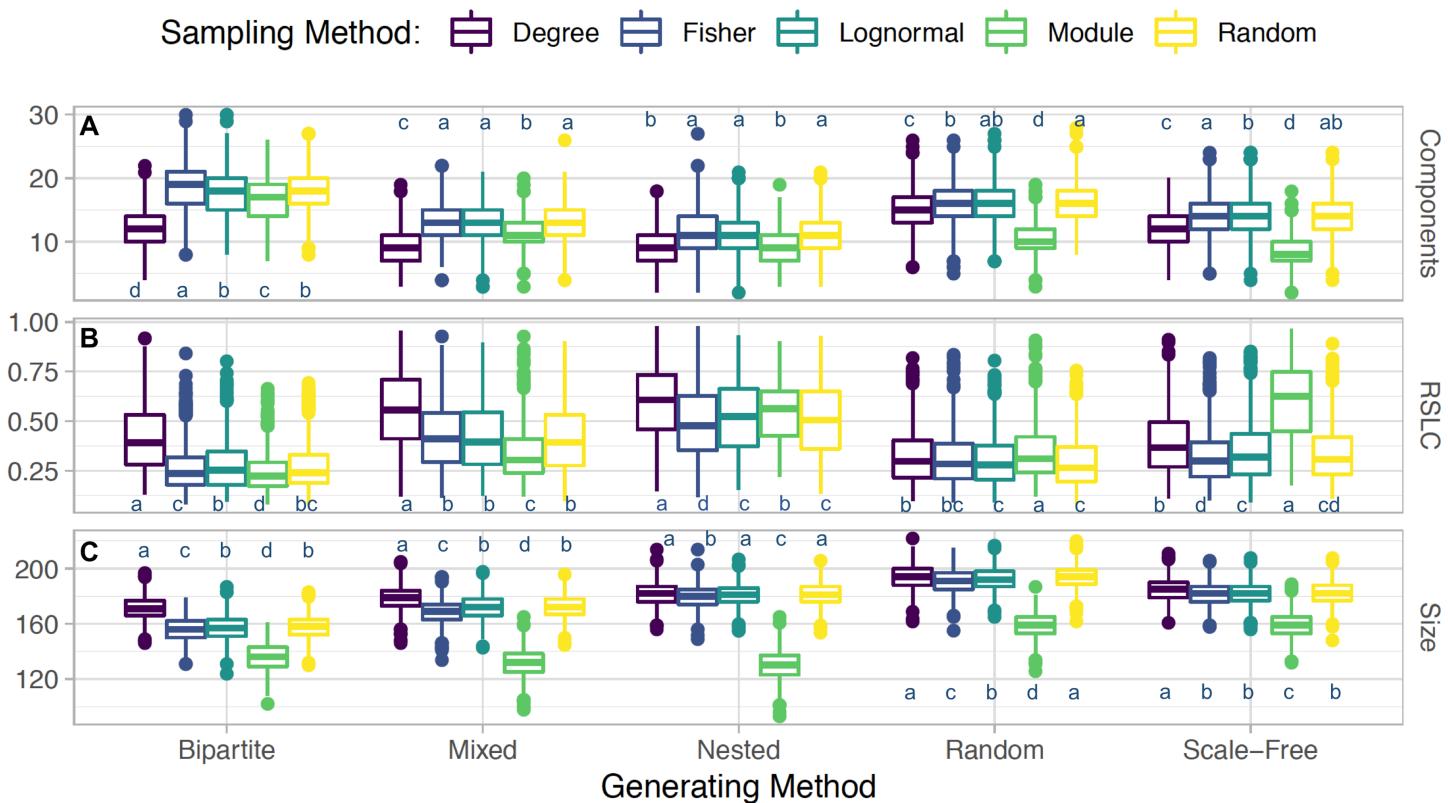
Full-size DOI: [10.7717/peerj.7566/fig-2](https://doi.org/10.7717/peerj.7566/fig-2)



**Figure 3** Relative size of the largest component (RSLC) analysis for a mixed modular network. The relative size of the largest component (RSLC) is shown for a network with mixed modules as a function of  $m$  (the number of anchoring nodes, which is the number of attempts to select nodes at random from the network) for  $nfn = 10$  (where  $nfn$  is maximum number of first neighbors added to the anchoring nodes), with line color indicating sampling design.

Full-size DOI: [10.7717/peerj.7566/fig-3](https://doi.org/10.7717/peerj.7566/fig-3)

section. RSLC varies with underlying network topology and sampling method (Figs. 2 and 3, with RSLC as a function of  $m$  for five sampling methods for bipartite nested modular networks and mixed module networks, respectively). Interactions between underlying network structure and sampling design were quantified by number of components, RSLC, and size of the sampled network (Fig. 4). The modular structure of the complete network led to sampled networks consisting of several disconnected components corresponding to nodes from a single module or from a small group of modules. In Fig. 4,



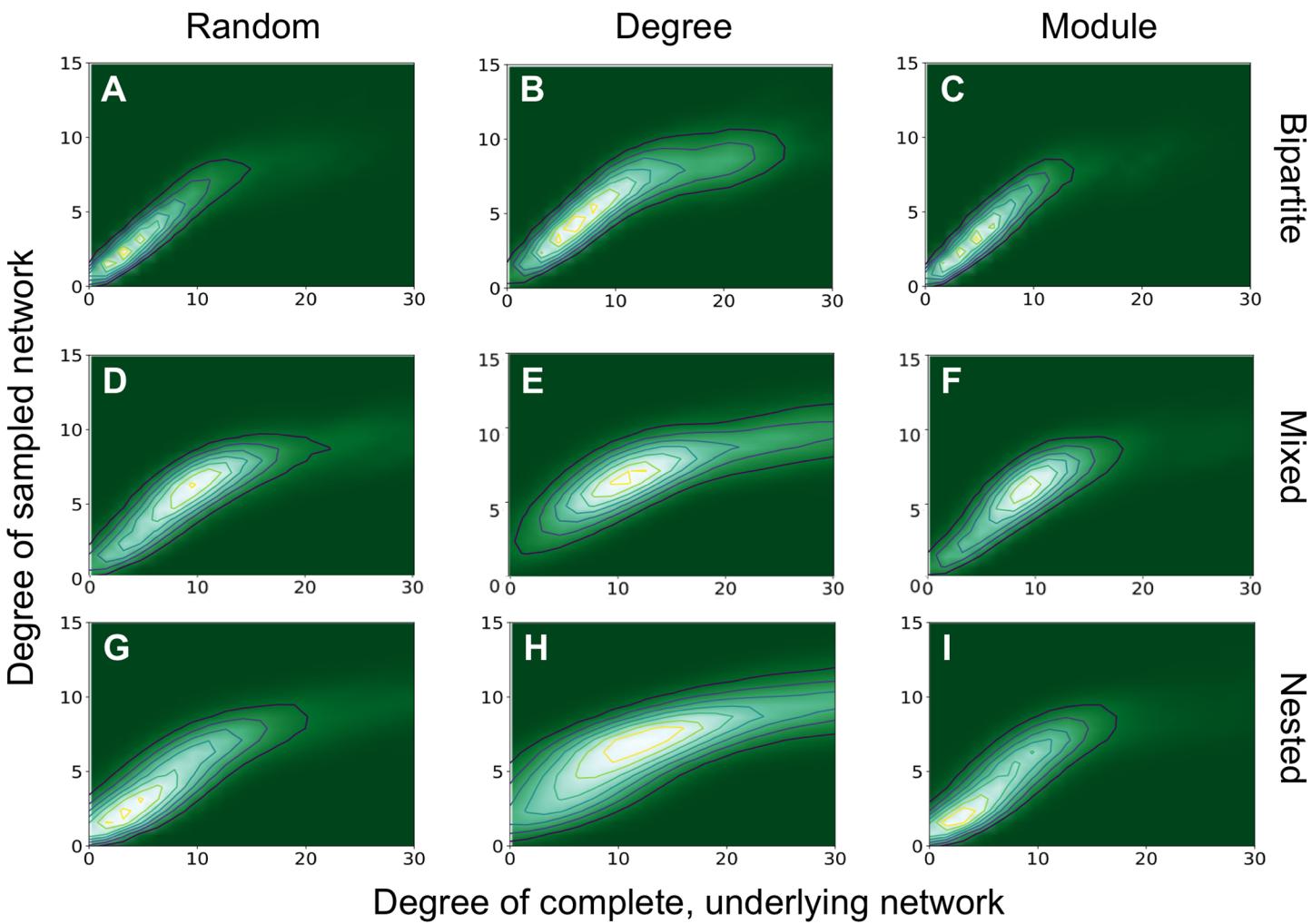
**Figure 4** Interacting sampling schema with network topology. Boxplots depicting average and standard error for (A) number of connected components, (B) relative size of largest components (RSLC), and (C) size of sampled network (measured by number of nodes) for generated networks with  $N = 500$ ,  $m = 50$ , and  $nfn = 5$ , sampled with multiple methods. For each method, the generated network was sampled with 1,000 replicates. With the Tukey HSD (honestly significant differences) test, we examine differences between sampling within a group (a generated network type and its outcome metric). Sampling methods within a group are statistically the same if they share a label.

Full-size DOI: 10.7717/peerj.7566/fig-4

for example ( $N = 500$ ,  $m = 50$ , and  $nfn = 5$ ), sampled networks typically had 12 disconnected components comprising approximately 150 sampled nodes.

We also found that the degree of sampled anchor nodes correlated with their *true* degree in the complete network (Fig. 5). We demonstrate this for the bipartite network with  $m = 50$ ,  $nfn = 10$  and three sampling methods. In all cases examined, there is a clear correlation between the true degree (corresponding to the complete network) and sampled degrees, represented by the straight ridge on the contour plots, at least for small degrees. Many of the patterns network ecologists are interested in are related to the degree of generalization or specialization of species, so even if the “true” degree cannot be determined by sampling, this demonstrates that all sampling schemes can capture reliable estimates of the degree of a species relative to others at low degrees of connectedness. Sampling by degree is one way to mimic a passive sampling where species with more interactions are more likely to be sampled. Only by sampling by degree by degree are we able to capture a saturation of the node degree, which is imposed here by the  $nfn$  added to the key nodes,  $nfn = 10$  (see point 4 below).

The variance of the distribution in Fig. 5 does not change much across different sampling methods for each type of network. More significant changes in the variance can



**Figure 5** Degree of the sampled network vs degree of the complete network. Degree of sampled and complete networks are investigated for bipartite (A–C), mixed (D–F), and nested (G–I) networks with  $m = 50$  and  $nfn = 10$ . Only the sampled anchoring nodes are shown. From left to right the sampling methods are: random, degree, and module. Contour plots were generated with 50 realizations of samplings on each network.

Full-size DOI: 10.7717/peerj.7566/fig-5

be observed for different network types, as they have different degree distributions. Bipartite nested networks have the smallest variance, which is related to their exponential degree distribution (see Eq. (A.4)), and to the fact that in bipartite networks, nodes in one type (one guild or class, for example) are only connected with nodes of the other type, and not directly to one another. The modules corresponding to each part are, therefore, smaller than in unipartite systems. When anchoring nodes are selected at random they will likely have small degree and all their neighbors will be included if  $nfn \sim k$ . Unipartite nested networks also show this effect, but not so strongly, since there is a single module and the degree distribution is less peaked. The degree distribution is a bit more uniform in unipartite networks than in bipartite networks, where each module is half the size of those in unipartite networks.

Sampling by degree, by definition, tends to find more anchoring nodes at the center of the modules and few in the periphery. Random sampling, on the other hand, picks

relatively more peripheral nodes, as can be seen by comparing Figs. 1B with 1D and 2B with 2D. These general features are also present in nested, scale-free, and random networks (not shown). The aggregate statistics of the sampled network properties we calculated are summarized in Fig. 4 for networks with five different module types and five sampling methods. Of the major metrics we investigated, we observed the following trends:

1. Number of connected components: Networks with nested modules have the smallest number of connected components and show the highest level of between-module connectedness. The other network types did not show significant variation in the number of connected components across the different sampling designs.
2. Relative size of the largest connected component: Networks with nested modules had largest sampled components containing up to 60% of the entire number of observed nodes, whereas the largest component of the other network types represents only 30% of the observed nodes, revealing a much lower degree of connectedness between modules than in nested networks. An exception is the mixed network sampled by degree, whose largest component contained 56% of the nodes of the observed network.
3. Size of sampled network: Networks with nested and bipartite nested modules always produced the smallest sampled networks (measured by RSLC), independent of the sampling procedure (Fig. 4). This is because in nested networks the probability that nearest neighbors of anchor nodes will overlap is large, resulting in multiple samplings of the same neighbor and creating the small observed sampled network sizes. Networks with scale-free modules, on the other hand, produced the largest observed networks, followed closely by those with random modules. On average, observed networks that were sampled from full networks with nested modules were 72% smaller than those sampled from networks with scale-free modules. This is because species with high degree in the scale free module will be in hubs that are also connected with species in other modules. Hence, once such species are sampled, we are likely to sample several other nodes from other modules. Mixed networks fell in between these two cases, as expected. However, irrespective of network type, if  $m \ll N$ , the expected network size is simply  $m^*nfn$  if  $nfn \ll k$  or  $m^*k$  if  $k \ll nfn$ , as all sampled neighbors would be included in this case.
4. Degree of sampled nodes: All sampling methods show linear correlation between the *true* and sampled degrees for anchor nodes if  $nfn \sim k$ , since most of the node's neighbors will be added to the sampled network. In this regime, only sampling by degree goes beyond the linear correlation and captures the saturation at  $nfn$ , as anchor nodes will typically have degrees larger than the average. If  $nfn \ll k$ , saturation at  $nfn$  will occur for all sampling methods. Saturation is not a benefit or drawback in itself, but an indication of the relationship between real and sampled networks. All sampling methods are limited by  $nfn$ , the  $nfn$  that can be potentially added to the sampled network. Sampling by degree implies that the selected anchor nodes have many connections. Saturation is an indication that for larger values of  $nfn$ , a greater sampling effort is required to recover the underlying structure. The extra effort is rewarded by a better sampled results (a larger sampled network with fewer components). Random sampling or sampling by module

(as illustrated in Fig. 5), on the other hand, selects portions of the original network with fewer connections, and therefore requires smaller values of  $nfn$ , but it also returns a poorer representation of the original network.

## DISCUSSION

Although complete ecological networks found in nature may be incredibly large, efforts to understand the structure or dynamics of these empirical systems have focused on smaller, tractable subcomponents of the actual networks due to limitations of time, energy, and budget (Burkle & Alarcón, 2011). Moreover, most field-based attempts to quantify ecological networks limit the types of interactions being measured to particular species of interest. In our formalization, individual studies would generally tend to examine one or a few modules that exist within a larger universe of interactions, defined here as the “complete” or underlying network. For example, a plant–pollinator module is depicted as a tightly interconnected bipartite network (Bascompte & Jordano, 2007), where the trophic interactions of its constituent species are part of a trophic network that may be ignored.

The complete network, encompassing different types of interactions and taxonomic groups, as well as the structural heterogeneity depicted among its subnetworks (here represented by the distinct modules), is rarely addressed. Nonetheless, the ecological and evolutionary dynamics of these modules in the community are hardly independent of each other. Modules may emerge naturally due to the sparseness of interactions across space, time, or even as the result of coevolutionary forces (Olesen et al., 2007; Beckett & Williams, 2013; Andreazzi, Thompson & Guimarães, 2017). Still, the effects of interactions in one module can propagate across the system (McCann, Rasmussen & Umbanhowar, 2005; Rooney et al., 2006). The groups of species and types of interactions one targets when conducting fieldwork will define the type and size of the network studied. In practice, adding additional species to a sampled network may uncover additional, previously unrecognized modules (sometimes, but not always, corresponding to guilds or functional groups of species; Donatti et al., 2011). Likewise, interacting species in adjacent communities and unsampled layers (of a multilayer network) may contain or correspond to their own modules.

Analytical techniques that simultaneously address multiple types of interactions and ecological outcomes (i.e., multilayered networks; Pilosof et al., 2017; Genrich et al., 2016) will require a better understanding of the bias imparted by sampling strategies, in order to deal with the insurmountable diversity of organisms and interactions in real communities. Thus, if we desire to understand the relationships between structure and function, we should ultimately aim to obtain the most accurate depiction of a network’s structure that encompasses all elements potentially affecting its function.

Here, we attempted to quantify how much of the idealized network is observable, and what systematic biases may exist as a function of the interaction between the designs used to sample species interactions, and the topological structure of the complete network. To this end, we provide the software *EcoNetGen*, which can be used to explore many features of biases introduced by sampling methods. We investigated only a few of these features, related to the between-module connectedness of the observed network.

We highlight that the scripts contained in *EcoNetGen*, *NetGen* (*netgen* in R), and *NetSampler* (*netsampler* in R) may be used to form null models and simulated networks in other studies, for further explorations of network size, structure, and sampling.

## CONCLUSIONS

Our analyses using *EcoNetGen* point to four main results. First, sampling design has a large impact on the properties of the observed network. Sampling according to species degree seems to be the only method that consistently generates nearly complete networks, as it produces the largest and more connected observed networks, with the smallest number of components. Sampling by degree in the field context would involve identifying a highly connected organism, such as a generalist pollinator or seed disperser, measuring interaction links from that organism, and subsequently adding “nearest neighbor” species, the interactions they have, and so on. It is not necessary, in this sampling schema, to start with the most highly-connected species, as long as it is highly connected. For most ecological systems, natural history studies can provide intuition about which species are most likely to be the highly-connected in the network, and which species are highly specialized. Moreover, it is relatively easy to identify the highly-connected species through incomplete sampling schemes (Pires, Marquitti & Guimarães, 2017). Thus, a combination of natural history information and sampling schemes focused on highly-connected species may provide the most accurate description of ecological networks.

Further, networks comprised of bipartite nested and nested modules generally result in poorly sampled networks that are small and have several disconnected components. This suggests that networks with these types of structures demand greater sampling effort than networks with random or scale-free modules, for instance. Although real networks are in-between these different structural patterns explored here, this highlights that special care should be taken with sampling design, since the interplay between sampling and structure will affect how representative the sampled network will be. The positive message is that sampling from networks with bipartite nested modules results in observed networks with a large number of small components, meaning that each module is well sampled. The challenge to build more representative networks is to devise ways to sample the connections among modules, which are rarely observed by the different sampling schemes for these types of networks.

In comparing sampling designs, it is clear that sampling by module produces by far the smallest observed networks for all topologies (Fig. 4). Sampling by degree, in contrast, produces the largest sampled networks (as measured by the RSLC) and is therefore most representative of the underlying complete network. For networks with random and scale-free modules, sampling by degree produced similar results compared to random sampling, but for nested and bipartite nested networks, sampling by degree always produced significantly larger observed networks than random sampling. Interestingly, sampling by abundance (as by transects or by sweep netting in the field) does not seem to be appropriate for nested or bipartite nested networks, because the results are only slightly better than sampling by module. For random and scale-free networks, sampling by

abundance produces observed networks that are only slightly smaller than those produced by sampling by degree.

Second, nested modules are better represented in sampled networks than other module structures. Since the renewal of the interest in ecological networks in recent decades, nestedness has played a central role in the literature and has been reported in a wide variety of systems described as bipartite networks (*Bascompte et al., 2003; Guimarães et al., 2006; Joppa et al., 2010*). However, the relevance of nestedness has been contested (*Staniczenko, Kopp & Allesina, 2013; James, Pitchford & Plank, 2012*), and mechanisms such as abundance heterogeneity and sampling have been invoked as underlying causes of the pervasiveness of the nested pattern (*Vázquez et al., 2009*). The fact that nested modules are over-represented compared to other module types in sampled networks suggests that a major underlying reason for the ubiquity of nestedness in empirical networks is that sampling strategies employed in empirical studies are successful in thoroughly sampling nested subnetworks (*Nielsen & Bascompte, 2007*), but may not perform as well when sampling non-nested subnetworks. Similarly, networks with unipartite nested modules (*Cantor et al., 2017*) stand out as providing observed networks with the most closely connected of all topologies. Sampling by degree, i.e., with focus on those species likely to establish more interactions, is therefore the recommended procedure for sampling networks with mixed modules, but it may overestimate the relative frequency of nested modules because non-nested modules are harder to thoroughly sample. Testing additional sampling designs capable of identifying other structures will aid in understanding the relative frequency of the different structural patterns in real networks.

Third, given a set number of anchoring nodes,  $m$ , and nearest neighbors to be probed,  $nfn$ , the size of the observed network does not depend significantly on the sampling method, but does depend strongly on the underlying network topology. As shown in Fig. 1, the size of the network at  $m = 50$  is smaller for nested and bipartite nested networks, independent of the sampling criterion, whereas networks with scale-free modules produce the largest sampled networks. This happens because the degree distribution in nested networks is highly heterogeneous and more anchoring nodes will likely have fewer than  $nfn$  neighbors. Network size (measured by RSLC) does not change much for any sampling method, with the exception of sampling by module (which always produces small observed networks). This suggests that sampling the entire network will be difficult no matter which sampling strategy is chosen, and perhaps the best strategy is one of iterative sampling, where the structure of a partially sampled network is analyzed and sampling is resumed using the sampling design that best suits the structure that has been uncovered.

Fourth and finally, sampling according to module generally results in small observed networks with a small number of observed components. This means that this method may detect most of the inner structure of some modules, but will also miss most of the connections between modules, and may miss modules entirely. Sampling by module does thoroughly sample a part of the network, allowing for identification of interactions in multiple modules. This type of sampling is arguably the most pervasive in the network literature where a certain type of interaction, guild, or taxonomic group is exhaustively

sampled. Our simulations show that sampling by module may give a thorough depiction of the module but may also point to other modules, which can then be sampled according to the most adequate sampling design.

Because properties of the network that determine the most efficient sampling strategy (such as species degree and module topology) are not fully known prior to undertaking a study, we recommend an iterative sampling approach, where the strategy can be adjusted as network properties are revealed. In recent years, interest in ecological network theory has grown exponentially ([Costa et al., 2007](#); [Borrett, Moody & Edelmann, 2014](#); [Delmas et al., 2019](#)) while our understanding of empirical systems has lagged behind, in part due to the difficult and time-intensive nature of field data collection and sample processing. Only by integrating a formal understanding of how empirical efforts reflect or bias estimation of the underlying network of species interactions can we confront theoretical models with our observations of natural systems.

## ACKNOWLEDGEMENTS

We thank the National Institute for Mathematical and Biological Synthesis, as well as three anonymous reviewers of this manuscript.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was conducted as a part of the Ecological Network Dynamics Working Group at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville, and the International Centre for Theoretical Physics ICTP-SAIFR #2016/01343-7 FAPESP. Marcus A.M. de Aguiar was supported by FAPESP (grants #2016/06054-3 and #2016/01343-7) and CNPq (grant #302049/2015-0). Erica Newman was supported by the University of Arizona Bridging Biodiversity and Conservation Science program. Publication fees were provided by the Berkeley Research Impact Initiative (BRII) sponsored by the UC Berkeley Library. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Ecological Network Dynamics Working Group at the National Institute for Mathematical and Biological Synthesis.

National Science Foundation: #DBI-1300426.

University of Tennessee, Knoxville, and the International Centre for Theoretical Physics ICTP-SAIFR #2016/01343-7 FAPESP.

FAPESP: #2016/06054-3 and #2016/01343-7.

CNPq: #302049/2015-0.

## Competing Interests

The authors declare that they have no competing interests.

## Author Contributions

- Marcus A.M. de Aguiar conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft, developed Fortran code, and Python scripts “*NetGen* (*netgen* in R)” and “*NetSampler* (*netsampler* in R)”; wrote [Supplemental Material](#); developed R package “EcoNetGen.”
- Erica A. Newman conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft, wrote [Supplemental Material](#); developed R package “EcoNetGen.”
- Mathias M. Pires conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft, developed R package “EcoNetGen.”
- Justin D. Yeakel authored or reviewed drafts of the paper, approved the final draft.
- Carl Boettiger performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft, developed R package “EcoNetGen.”
- Laura A. Burkle authored or reviewed drafts of the paper, approved the final draft.
- Dominique Gravel authored or reviewed drafts of the paper, approved the final draft, organized NIMBioS Working Group and secured funding.
- Paulo R. Guimarães Jr. conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft, organized NIMBioS Working Group and secured funding.
- James L. O'Donnell performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft, organized NIMBioS Working Group and secured funding.
- Timothée Poisot authored or reviewed drafts of the paper, approved the final draft.
- Marie-Josée Fortin authored or reviewed drafts of the paper, approved the final draft.
- David H. Hembry authored or reviewed drafts of the paper, approved the final draft, organized NIMBioS Working Group and secured funding.

## Data Availability

The following information was supplied regarding data availability:

We made our code available as an R package called “EcoNetGen,” which is now available on CRAN (<https://CRAN.R-project.org/package=EcoNetGen>).

Python scripts are available in [Appendix S6](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7566#supplemental-information>.

## REFERENCES

- Andreazzi CS, Thompson JN, Guimarães PR Jr.** 2017. Network structure and selection asymmetry drive coevolution in species-rich antagonistic interactions. *American Naturalist* **190**(1):99–115 DOI [10.1086/692110](https://doi.org/10.1086/692110).
- Baldridge E, Harris DJ, Xiao X, White EP.** 2016. An extensive comparison of species-abundance distribution models. *PeerJ* **4**:e2823 DOI [10.7717/peerj.2823](https://doi.org/10.7717/peerj.2823).
- Bartomeus I.** 2013. Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. *PLOS ONE* **8**(7):e69200 DOI [10.1371/journal.pone.0069200](https://doi.org/10.1371/journal.pone.0069200).
- Bascompte J, Jordano P.** 2007. Plant-animal mutualistic networks: the architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **38**(1):567–593 DOI [10.1146/annurev.ecolsys.38.091206.095818](https://doi.org/10.1146/annurev.ecolsys.38.091206.095818).
- Bascompte J, Jordano P, Melián CJ, Olesen JM.** 2003. The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences of the United States of America* **100**(16):9383–9387 DOI [10.1073/pnas.1633576100](https://doi.org/10.1073/pnas.1633576100).
- Beckett SJ, Williams HTP.** 2013. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. *Interface Focus* **3**(6):20130033 DOI [10.1098/rsfs.2013.0033](https://doi.org/10.1098/rsfs.2013.0033).
- Blüthgen N, Menzel F, Blüthgen N.** 2006. Measuring specialization in species interaction networks. *BMC Ecology* **6**(1):9 DOI [10.1186/1472-6785-6-9](https://doi.org/10.1186/1472-6785-6-9).
- Borrett SR, Moody J, Edelmann A.** 2014. The rise of network ecology: maps of the topic diversity and scientific collaboration. *Ecological Modelling* **293**:111–127 DOI [10.1016/j.ecolmodel.2014.02.019](https://doi.org/10.1016/j.ecolmodel.2014.02.019).
- Burkle LA, Alarcón R.** 2011. The future of plant-pollinator diversity: understanding interaction networks across time, space, and global change. *American Journal of Botany* **98**(3):528–538 DOI [10.3732/ajb.1000391](https://doi.org/10.3732/ajb.1000391).
- Cantor M, Pires MM, Marquitti FMD, Raimundo RLG, Sebastián-González E, Coltri PP, Perez SI, Barneche DR, Brandt DYC, Nunes K, Daura-Jorge FG, Floeter SR, Guimarães PR Jr.** 2017. Nestedness across biological scales. *PLOS ONE* **12**(2):e0171691 DOI [10.1371/journal.pone.0171691](https://doi.org/10.1371/journal.pone.0171691).
- Cirtwill AR, Eklöf A, Roslin T, Wootton K, Gravel D.** 2019. A quantitative framework for investigating the reliability of empirical network construction. *Methods in Ecology and Evolution* **10**(6):902–911 DOI [10.1111/2041-210X.13180](https://doi.org/10.1111/2041-210X.13180).
- Costa LDF, Rodrigues FA, Travieso G, Villas Boas PR.** 2007. Characterization of complex networks: a survey of measurements. *Advances in Physics* **56**(1):167–242 DOI [10.1080/00018730601170527](https://doi.org/10.1080/00018730601170527).
- De Aguiar MAM, Newman EA, Pires M, Boettiger C.** 2019. EcoNetGen: simulate and sample from ecological interaction networks (version v0.2.3). Available at <https://CRAN.R-project.org/package=EcoNetGen>.
- Delmas E, Besson M, Brice M-H, Burkle LA, Dalla Riva GV, Fortin M-J, Gravel D, Guimarães PR Jr, Hembry DH, Newman EA, Olesen JM, Pires MM, Yeakel JD, Poisot T.** 2019. Analysing ecological networks of species interactions. *Biological Reviews* **94**(1):16–36 DOI [10.1111/brv.12433](https://doi.org/10.1111/brv.12433).
- Donatti CI, Guimarães PR, Galetti M, Pizo MA, Marquitti FM, Dirzo R.** 2011. Analysis of a hyper-diverse seed dispersal network: modularity and underlying mechanisms. *Ecology Letters* **14**(8):773–781 DOI [10.1111/j.1461-0248.2011.01639.x](https://doi.org/10.1111/j.1461-0248.2011.01639.x).

- Fautin DG, Allen GR.** 1997. Life history of Anemonefishes. In: Fautin DG, Allen GR, eds. *Anemone Fishes and Their Host Sea Anemones*. Vol. 600. Perth: Western Australian Museum, 1–142.
- Fontaine C, Guimarães PR Jr, Kéfi S, Loeuille N, Memmott J, Van Der Putten WH, Van Veen FJF, Thébaud E.** 2011. The ecological and evolutionary implications of merging different types of networks. *Ecology Letters* **14**(11):1170–1181 DOI [10.1111/j.1461-0248.2011.01688.x](https://doi.org/10.1111/j.1461-0248.2011.01688.x).
- Fründ J, McCann KS, Williams NM.** 2016. Sampling bias is a challenge for quantifying specialization and network structure: lessons from a quantitative niche model. *Oikos* **125**(4):502–513 DOI [10.1111/oik.02256](https://doi.org/10.1111/oik.02256).
- Genrich CM, Mello MAR, Silveira FAO, Bronstein JL, Paglia AP.** 2016. Duality of interaction outcomes in a plant-frugivore multilayer network. *Oikos* **126**(3):361–368 DOI [10.1111/oik.03825](https://doi.org/10.1111/oik.03825).
- Gerhard F, Pipa G, Lima B, Neuenschwander S, Gerstner W.** 2011. Extraction of network topology from multi-electrode recordings: is there a small-world effect? *Frontiers in Computational Neuroscience* **5**:4.
- Gibson RH, Knott B, Eberlein T, Memmott J.** 2011. Sampling method influences the structure of plant-pollinator networks. *Oikos* **120**(6):822–831 DOI [10.1111/j.1600-0706.2010.18927.x](https://doi.org/10.1111/j.1600-0706.2010.18927.x).
- Graham CH, Weinstein BG.** 2018. Towards a predictive model of species interaction beta diversity. *Ecology Letters* **21**(9):1299–1310 DOI [10.1111/ele.13084](https://doi.org/10.1111/ele.13084).
- Guimarães PR Jr, Rico-Gray V, Dos Reis SF, Thompson JN.** 2006. Asymmetries in specialization in ant-plant mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences* **273**(1597):2041–2047 DOI [10.1098/rspb.2006.3548](https://doi.org/10.1098/rspb.2006.3548).
- Guimarães PR Jr, Sazima C, Dos Reis SF, Sazima I.** 2007. The nested structure of marine cleaning symbiosis: is it like flowers and bees? *Biology Letters* **3**(1):51–54 DOI [10.1098/rsbl.2006.0562](https://doi.org/10.1098/rsbl.2006.0562).
- Hall SJ, Raffaelli DG.** 1993. Food webs: theory and reality. *Advances in Ecological Research* **24**:187–239 DOI [10.1016/S0065-2504\(08\)60043-4](https://doi.org/10.1016/S0065-2504(08)60043-4).
- Harte J, Newman EA.** 2014. Maximum information entropy: a foundation for ecological theory. *Trends in Ecology & Evolution* **29**(7):384–389 DOI [10.1016/j.tree.2014.04.009](https://doi.org/10.1016/j.tree.2014.04.009).
- Hegland SJ, Dunne J, Nielsen A, Memmott J.** 2010. How to monitor ecological communities cost-efficiently: the example of plant-pollinator networks. *Biological Conservation* **143**(9):2092–2101 DOI [10.1016/j.biocon.2010.05.018](https://doi.org/10.1016/j.biocon.2010.05.018).
- Hembry DH, Raimundo RLG, Newman EA, Atkinson L, Guo C, Guimarães PR Jr, Gillespie RG.** 2018. Does biological intimacy shape ecological network structure? A test using a brood pollination mutualism on continental and oceanic islands. *Journal of Animal Ecology* **87**(4):1160–1171 DOI [10.1111/1365-2656.12841](https://doi.org/10.1111/1365-2656.12841).
- Hu P, Lau WC.** 2013. A survey and taxonomy of graph sampling. *arXiv Preprint arXiv:1308.5865*.
- Ings TC, Montoya JM, Bascompte J, Blüthgen N, Brown L, Dormann CF, Edwards F, Figueroa D, Jacob U, Jones JI, Lauridsen RB, Ledger ME, Lewis HM, Olesen JM, Van Veen FJF, Warren PH, Woodward G.** 2009. Review: ecological networks-beyond food webs. *Journal of Animal Ecology* **78**(1):253–269 DOI [10.1111/j.1365-2656.2008.01460.x](https://doi.org/10.1111/j.1365-2656.2008.01460.x).
- James A, Pitchford JW, Plank MJ.** 2012. Disentangling nestedness from models of ecological complexity. *Nature* **487**(7406):227–230 DOI [10.1038/nature11214](https://doi.org/10.1038/nature11214).
- Joppa LN, Montoya JM, Solé R, Sanderson J, Pimm SL.** 2010. On nestedness in ecological networks. *Evolutionary Ecology Research* **12**:35–46.
- Jordano P.** 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *American Naturalist* **129**(5):657–677 DOI [10.1086/284665](https://doi.org/10.1086/284665).

- Jordano P.** 2016. Sampling networks of ecological interactions. *Functional Ecology* 30(12):1883–1893 DOI [10.1111/1365-2435.12763](https://doi.org/10.1111/1365-2435.12763).
- Leskovec J, Faloutsos C.** 2006. Sampling from large graphs. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM, 631–636.
- Levina A, Priesemann V.** 2017. Subsampling scaling. *Nature Communications* 8(1):15140 DOI [10.1038/ncomms15140](https://doi.org/10.1038/ncomms15140).
- Lewinsohn TM, Inácio Prado P, Jordano P, Bascompte J, Olesen JM.** 2006. Structure in plant-animal interaction assemblages. *Oikos* 113(1):174–184 DOI [10.1111/j.0030-1299.2006.14583.x](https://doi.org/10.1111/j.0030-1299.2006.14583.x).
- Marquitti FMD, Guimarães PR Jr, Pires MM, Bittencourt LF.** 2014. MODULAR: software for the autonomous computation of modularity in large network sets. *Ecography* 37(3):221–224 DOI [10.1111/j.1600-0587.2013.00506.x](https://doi.org/10.1111/j.1600-0587.2013.00506.x).
- McCann KS, Rasmussen JB, Umbanhowar J.** 2005. The dynamics of spatially coupled food webs. *Ecology Letters* 8(5):513–523 DOI [10.1111/j.1461-0248.2005.00742.x](https://doi.org/10.1111/j.1461-0248.2005.00742.x).
- McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, Dornelas M, Enquist BJ, Green JL, He F, Hurlbert AH, Magurran AE, Marquet PA, Maurer BA, Ostling A, Soykan CU, Ugland KI, White EP.** 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10(10):995–1015 DOI [10.1111/j.1461-0248.2007.01094.x](https://doi.org/10.1111/j.1461-0248.2007.01094.x).
- Mello MAR, Marquitti FMD, Guimarães PR Jr, Kalko EKV, Jordano P, De Aguiar MAM.** 2011a. The missing part of seed dispersal networks: structure and robustness of bat-fruit interactions. *PLOS ONE* 6(2):e17395 DOI [10.1371/journal.pone.0017395](https://doi.org/10.1371/journal.pone.0017395).
- Mello MAR, Marquitti FMD, Guimarães PR Jr, Kalko EKV, Jordano P, De Aguiar MAM.** 2011b. The modularity of seed dispersal: differences in structure and robustness between bat- and bird- fruit networks. *Oecologia* 167(1):131–140 DOI [10.1007/s00442-011-1984-2](https://doi.org/10.1007/s00442-011-1984-2).
- Morin PJ.** 2009. *Community ecology*. Second Edition. Hoboken: John Wiley & Sons.
- Newman MEJ.** 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* 103(23):8577–8582 DOI [10.1073/pnas.0601602103](https://doi.org/10.1073/pnas.0601602103).
- Newman EA, Wilber MQ, Kopper KE, Moritz MA, Falk DA, McKenzie D, Harte J.** 2018. Disturbance macroecology: integrating disturbance ecology and macroecology in different-age post-fire stands of a closed-cone pine forest. *BioRxiv* 309419 DOI [10.1101/309419](https://doi.org/10.1101/309419).
- Nielsen A, Bascompte J.** 2007. Ecological networks, nestedness and sampling effort. *Journal of Ecology* 95(5):1134–1141 DOI [10.1111/j.1365-2745.2007.01271.x](https://doi.org/10.1111/j.1365-2745.2007.01271.x).
- Olesen JM, Bascompte J, Dupont YL, Jordano P.** 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the United States of America* 104(50):19891–19896 DOI [10.1073/pnas.0706375104](https://doi.org/10.1073/pnas.0706375104).
- Ollerton J, McCollin D, Fautin DG, Allen GR.** 2007. Finding NEMO: nestedness engendered by mutualistic organisation in anemonefish and their hosts. *Proceedings of the Royal Society B: Biological Sciences* 274(1609):591–598 DOI [10.1098/rspb.2006.3758](https://doi.org/10.1098/rspb.2006.3758).
- Pascual M, Dunne J.** 2006. Ecological networks: linking structure to dynamics in food webs. *Proceedings of Workshop at the Santa Fe Institute*. Oxford: SFI and Oxford University Press.
- Pilosof S, Porter MA, Pascual M, Kéfi S.** 2017. The multilayer nature of ecological networks. *Nature Ecology & Evolution* 1:0101.
- Pires MM, Guimarães PR Jr.** 2013. Interaction intimacy organizes networks of antagonistic interactions in different ways. *Journal of the Royal Society Interface* 10(78):20120649 DOI [10.1098/rsif.2012.0649](https://doi.org/10.1098/rsif.2012.0649).

- Pires MM, Marquitti FMD, Guimarães PR Jr.** 2017. The friendship paradox in species-rich ecological networks: implications for conservation and monitoring. *Biological Conservation* **209**:245–252 DOI [10.1016/j.biocon.2017.02.026](https://doi.org/10.1016/j.biocon.2017.02.026).
- Rivera-Hutinel A, Bustamante RO, Marín VH, Medel R.** 2012. Effects of sampling completeness on the structure of plant-pollinator networks. *Ecology* **93**(7):1593–1603 DOI [10.1890/11-1803.1](https://doi.org/10.1890/11-1803.1).
- Rooney N, McCann K, Gellner G, Moore JC.** 2006. Structural asymmetry and the stability of diverse food webs. *Nature* **442**(7100):265–269 DOI [10.1038/nature04887](https://doi.org/10.1038/nature04887).
- Stang M, Klinkhamer PGL, Waser NM, Stang I, Van Der Meijden E.** 2009. Size-specific interaction patterns and size matching in a plant-pollinator interaction web. *Annals of Botany* **103**(9):1459–1469 DOI [10.1093/aob/mcp027](https://doi.org/10.1093/aob/mcp027).
- Staniczenko PPA, Kopp JC, Allesina S.** 2013. The ghost of nestedness in ecological networks. *Nature Communications* **4**(1):1391 DOI [10.1038/ncomms2422](https://doi.org/10.1038/ncomms2422).
- Stumpf MPH, Wiuf C.** 2005. Sampling properties of random graphs: the degree distribution. *Physical Review E* **72**(3):036118 DOI [10.1103/PhysRevE.72.036118](https://doi.org/10.1103/PhysRevE.72.036118).
- Vázquez DP, Blüthgen N, Cagnolo L, Chacoff NP.** 2009. Uniting pattern and process in plant-animal mutualistic networks: a review. *Annals of Botany* **103**(9):1445–1457 DOI [10.1093/aob/mcp057](https://doi.org/10.1093/aob/mcp057).
- Vázquez DP, Morris WF, Jordano P.** 2005. Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecology Letters* **8**(10):1088–1094 DOI [10.1111/j.1461-0248.2005.00810.x](https://doi.org/10.1111/j.1461-0248.2005.00810.x).
- Vizentin-Bugoni J, Maruyama PK, Debastiani VJ, Duarte LDS, Dalsgaard B, Sazima M.** 2016. Influences of sampling effort on detected patterns and structuring processes of a Neotropical plant-hummingbird network. *Journal of Animal Ecology* **85**(1):262–272 DOI [10.1111/1365-2656.12459](https://doi.org/10.1111/1365-2656.12459).
- White EP, Thibault KM, Xiao X.** 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* **93**(8):1772–1778 DOI [10.1890/11-2177.1](https://doi.org/10.1890/11-2177.1).