The structure and stability of ecological networks

Benno Isaac Simmons

Clare College

April 2019

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Benno I. Simmons: The structure and stability of ecological networks

All species interact with other species to form complex networks of connections. Such networks are a powerful way to represent ecological communities because they describe (i) the roles of individual species and (ii) the structure of the community as a whole in a single framework amenable to mathematical and computational analysis. In this thesis I consider a number of outstanding problems in network ecology. In Chapter 1, I examine the consequences for network structure of removing non-mutualistic interactions from plant-frugivore visitation networks. I find that plant-frugivore visitation networks act as a good proxy for mutualistic seed dispersal networks in terms of whole-network topology, but not when considering species-level structures. Chapter 2 deals with whether generalisation drives abundance or vice versa in plant-hummingbird pollination networks. I find evidence that abundance drives generalisation and use a simple model to show that neutral processes can explain broad patterns of species-level generalisation. In Chapter 3, I quantify the importance and vulnerability of mutualistic interactions to understand the risk that interaction extinction poses to communities. I conclude that (i) the interactions most important for community stability are those which are most vulnerable to extinction, and (ii) important and vulnerable interactions tend to be important and vulnerable wherever they occur. In Chapter 4, I consider motifs as an alternative to indices for characterising the structure of bipartite networks. I find that motifs capture significantly more information about network topology than indices and advocate adding bipartite motifs to the suite of analytical tools used by network ecologists. Chapter 5 describes a software package in R, MATLAB and Python for conducting motif analyses of bipartite networks. It uses novel mathematical formulations to dramatically reduce the computational time required for motif calculations compared to competing software.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the 60,000 word limit for the Biology Degree Committee.

As the data chapters were prepared for publication, and invariably involved collaborations, I use the pronoun 'we' rather than 'I' throughout.

Supplementary materials for each chapter (Appendices) are structured as follows. If a chapter refers to an appendix or supplementary figure or table, proceed to the relevant chapter named 'Supplementary materials related to Chapter X' and find the reference. For example, Chapter 1 references Figure S4. Therefore, in this case, locate Figure S4 in 'Supplementary materials related to Chapter 1'.

The roles of people other than myself in this work are as follows:

Chapter 1
I conceived the idea with Juan Pedro González-Varo. Jörg Albrecht, Nina Farwig, Daniel García and Pedro Jordano provided empirical network data. Juan Pedro González-Varo (i) compiled data on the outcome of plant-frugivore interactions, (ii) wrote a first draft of the introduction, which I subsequently edited and adapted, (iii) produced Figure S1,

and (iv) provided advice and guidance throughout. William Sutherland, Lynn Dicks, Jörg Albrecht, Nina Farwig, Daniel García, Pedro Jordano and Juan P. González-Varo commented on manuscript drafts.

Chapter 2

Jeferson Vizentin-Bugoni, Pietro Maruyama, Peter Cotton, Oscar Marin-Gomez, Carlos Lara, Liliana Lasprilla, Maria Maglianesi, Raul Ortiz-Pulido, Marcia Rocca, Licleia Rodrigues, Boris Tinoco, Marcelo Vasconcelos, Marlies Sazima, Ana Martin Gonzalez, Jesper Sonne, Carsten Rahbek and Bo Dalsgaard provided and/or helped compile network data with associated abundance estimates and commented on manuscript drafts. Lynn V. Dicks and William J. Sutherland also commented on manuscript drafts.

Chapter 3

Hannah Wauchope assisted with using the Cambridge High Performance Computing facilities. Tatsuya Amano advised on statistical analyses to assess taxonomic consistency. Vasilis Dakos advised on analyses and writing throughout the project. Hannah Wauchope, Tatsuya Amano, Lynn Dicks, William Sutherland and Vasilis Dakos commented on manuscript drafts and/or advice during the project.

Chapter 4

I designed analyses and interpreted results in collaboration with Alyssa Cirtwill and Daniel Stouffer. Nick Baker performed the original analysis demonstrating meso-scale dissimilarity in networks with similar network-level properties, supervised by Daniel Stouffer. Alyssa Cirtwill adapted this analysis for the paper. I subsequently rewrote the code for the analysis from scratch, as well as extending and modifying the analysis. Hannah Wauchope assisted with using the Cambridge High Performance Computing facilities. Alyssa Cirtwill, Nick Baker, Hannah Wauchope, Lynn Dicks, Daniel Stouffer and William Sutherland commented on manuscript drafts.

Chapter 5

Building on my initial code, Riccardo Di Clemente developed mathematical formulae for counting the number of times two- to five-node bipartite motifs occur in a network.

Michelle Sweering extended these formulae to six-node motifs, and additionally developed formulae to count the number of times each node in a network occurs in each unique position within bipartite motifs. Michelle was supervised by me as part of the Cambridge Faculty of Mathematics Post-Masters Placements (now Cambridge Mathematics Placements) scheme. Maybritt Schillinger was another student I supervised through this scheme, and she wrote descriptions of the formulae which are part of the appendix of this chapter. Maybritt additionally made useful code suggestions for the final package. I developed the R package, wrote the manuscript draft and conducted all analyses. Riccardo Di Clemente developed the MATLAB and Python version of the package, which I documented. Michelle Sweering, Maybritt Schillinger, Lynn Dicks, William Sutherland and Riccardo Di Clemente commented on manuscript drafts.

Publications

The following publications form the basis of this thesis:

Chapter 1: Simmons, B. I., Sutherland W. J., Dicks, L. V., Albrecht, J., Farwig, N., García, D., Jordano, P. and González-Varo, J. P. (2018). Moving from frugivory to seed dispersal: incorporating the functional outcomes of interactions in plant-frugivore networks. *Journal of Animal Ecology*, 87, 995–1007.

Chapter 2: Simmons, B. I., Vizentin-Bugoni, J., Maruyama, P. K., Cotton, P. A., Marín-Gómez, O. H., Lara, C., Rosero-Lasprilla, L., Maglianesi, M. A., Ortiz-Pulido, R., Rocca, M. A., Rodrigues, L. C., Tinoco, B., Vasconcelos, M. F., Sazima, M., Martín González, A. M., Sonne, J., Rahbek, C., Dicks, L. V., Dalsgaard, B., Sutherland, W. J. Abundance drives broad patterns of generalisation in hummingbird-plant pollination networks. *Oikos*. (accepted)

Chapter 3: Simmons, B. I., Wauchope, H. S., Amano, T., Dicks, L. V., Sutherland, W. J. and Dakos, V. Vulnerable mutualistic interactions are consistently important across networks. *PNAS* (submitted)

Chapter 4: Simmons, B. I., Cirtwill, A. R., Baker, N. J., Wauchope, H. S., Dicks, L. V., Stouffer, D. B. and Sutherland, W. J. (2019). Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, 128: 154-170

Chapter 5: Simmons, B. I., Sweering, M. J. M., Schillinger, M., Dicks, L. V., Sutherland, W. J. and Di Clemente, R. (2019). bmotif: a package for motif analyses of bipartite networks. *Methods in Ecology and Evolution*. (in press)

In addition, I contributed to the following papers during my PhD:

Poisot, T., LaBrie, R., Larson, E., Rahlin, A. and **Simmons, B. I.** Data-based, synthesis-driven: setting the agenda for computational ecology. *Ideas in Ecology and Evolution* (submitted)

Rose, D. C., Amano, T., González-Varo, J. P., Mukherjee, N., Robertson, R. J., **Simmons, B. I.**, Wauchope, H. S. and Sutherland, W. J. Insurmountable barriers, bad solutions, or lack of implementation? A new agenda for conservation science-policy research. *Biological Conservation* (in review)

Simmons, B. I., Martin, P. A., Worthington, T. A. Improve preprint readability. *Nature.* (in review)

Simmons, B. I., Balmford, A., Bladon, A. J., Christie, A. P., De Palma, A., Dicks, L. V., Gallego-Zamorano, J., Johnston, A., Martin, P. A., Purvis, A., Rocha, R., Wauchope, H. S., Wordley, C. F. R., Worthington, T. A. and Finch, T. (2019). Worldwide insect declines: an important message, but interpret with caution. *Ecology and Evolution* (in press)

Tew, E. R., **Simmons B. I.**, Sutherland, W. J. (2019). Quantifying cultural ecosystem services: disentangling the effects of management from landscape features. *People and Nature* (in press)

Simmons, B. I., Hoeppke, C. and Sutherland, W. J. (2019). Beware greedy algorithms. *Journal of Animal Ecology* (in press).

Balmford, A., Amano, T., Bartlett, H., Chadwick, D., Collins, A., Edwards, D., Field, R., Garnsworthy, P., Green, R., Smith, P., Waters, H., Whitmore, A., Broom, D., Chara, J., Finch, T., Garnett, E., Gathorne-Hardy, A., Hernandex-Medrano, J., Herrero, M., Hua, F., Latawiec, A., Misselbrook, T., Phalan, B., **Simmons, B. I.**, Vause, J., Ermgassen, E., Eisner, R. (2018). The environmental costs and benefits of high-yield farming. *Nature Sustainability*, 1, 477-485.

Jucker, T., Wintle, B., Shackelford, G., Bocquillon, P., Geffert, J. L., Kasoar, T., Jovacs, E., Mumby, H. S., Orland, C., Schleicher, J., Tew, E., Zabala, A., Amano, T., Bell, A., Bongalov, B., Chambers, J. M., Corrigan, C., Durán, A. P., Duvic-Paoli, L-A., Emilson, C., da Silva, J. F., Garnett, E. E., Green, E. J., Guth, M. K., Hacket-Pain, A., Hinsley, A., Igea, J., Kunz, M., Luke, S. H., Lynam, W., Martin, P. A., Nunes, M. H., Ockendon, N., Pavitt, A., Payne, C. L. R., Plutshack, V., Rademacher, T., Robertson, R. J., Rose, D. C., Serban, A., **Simmons, B. I.**, Szkokan-Emilson, E. J., Tayleur, C., Wordley, C. and Mukherjee, N. (2018). Ten-year assessment of the 100 priority questions for global biodiversity conservation. *Conservation Biology* (in press)

Rose, D. C., Sutherland, W. J., Amano, T. , González-Varo, J. P., Robertson, R. J., **Simmons, B. I.**, Wauchope, H. S., Kovacs, E. , Durán, A. P., Vadrot, A. B., Wu, W., Dias, M. P., Di Fonzo, M. M., Ivory, S., Norris, L., Nunes, M. H., Nyumba, T. O., Steiner, N., Vickery, J. and Mukherjee, N. (2018). The major barriers to evidence-informed conservation policy and possible solutions. *Conservation Letters*. (in press)

Dalsgaard, B., Kennedy, J., **Simmons, B. I.**, Baquero, A., González, A. M. M., Timmermann, A., McGuire, J. A., Ollerton, J., Sutherland, W. J. and Rahbek, C. (2018). Trait evolution, nectar-feeding specialization and vulnerability to plant extinctions among hummingbirds on Antillean islands: the role of mountains and plant richness. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20172754.

Ratto, F., **Simmons, B. I.**, Spake, R., Gutierrez, V. Z., McDonald, M., Merriman, J. C., Poppy, G. M., Peh, K. S.-H. and Dicks, L. V. (2018). Global importance of vertebrate pollinators for plant reproductive success: a meta-analysis. *Frontiers in Ecology and the Environment*, 16(2), 82-90.

Rose, D. C., Mukherjee, N., **Simmons, B. I.**, Tew, E. R., Robertson, R. J., Vadrot, A. B. M., Doubleday, R. and Sutherland, W. J. (2017). Policy windows for the environment: Tips for improving the uptake of scientific knowledge. *Environmental Science and Policy*. (in press)

Acknowledgements

Summary

All species interact with other species to form complex networks of connections. Such networks are a powerful way to represent ecological communities because they describe (i) the roles of individual species and (ii) the structure of the community as a whole in a single framework amenable to mathematical and computational analysis. In this thesis I consider a number of outstanding problems in network ecology. In Chapter 1, I examine the consequences for network structure of removing non-mutualistic interactions from plant-frugivore visitation networks. I find that plant-frugivore visitation networks act as a good proxy for mutualistic seed dispersal networks in terms of whole-network topology, but not when considering species-level structures. Chapter 2 deals with whether generalisation drives abundance or vice versa in plant-hummingbird pollination networks. I find evidence that abundance drives generalisation and use a simple model to show that neutral processes can explain broad patterns of species-level generalisation. In Chapter 3, I quantify the importance and vulnerability of mutualistic interactions to understand the risk that interaction extinction poses to communities. I conclude that (i) the interactions most important for community stability are those which are most vulnerable to extinction, and (ii) important and vulnerable interactions tend to be important and vulnerable wherever they occur. In Chapter 4, I consider motifs as an alternative to indices for characterising the structure of bipartite networks. I find that motifs capture significantly more information about network topology than indices and advocate adding bipartite motifs to the suite of analytical tools used by network ecologists. Chapter 5 describes a software package in R, MATLAB and Python for conducting motif analyses of bipartite networks. It uses novel mathematical formulations to dramatically reduce the computational time required for motif calculations compared to competing software.
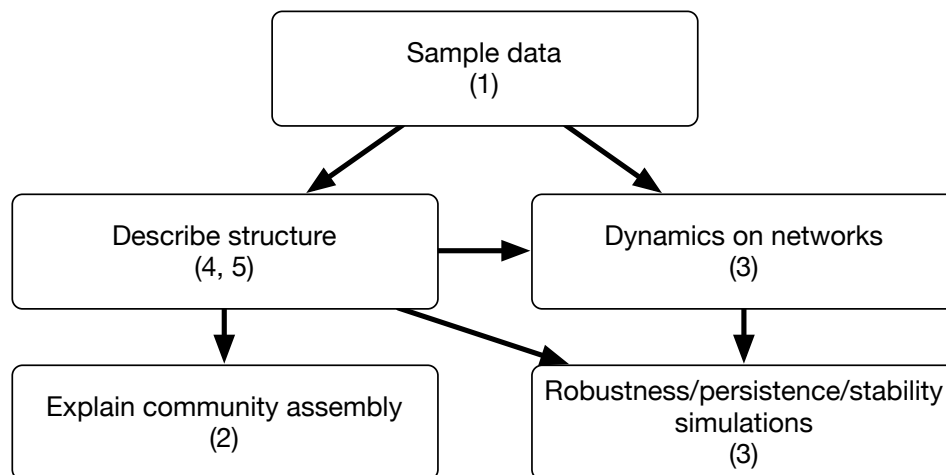
Contents

Introduction

Why study ecological networks?

All species interact with other species. These interactions form complex networks of connections that provide the wireframe for a wide variety of ecological functions, such as pollination, seed dispersal, parasitism, herbivory and predation. The functions mediated by ecological networks play an important role in regulating community dynamics and maintaining biodiversity, and thus studying ecological networks can yield important insights into these fundamental processes. More specifically, studying the *structure* of ecological networks can shed light on the processes governing species interactions, while studying the *stability* of networks can help understand the dynamic implications of this structure.

Mutualistic networks

Mutualistic networks are networks depicting interactions between two species, often a plant and an animal, where both species benefit from the exchange. For example, in pollination, the animal gains energy from consuming nectar, while the plant has its pollen dispersed for reproduction. Two mutualistic interactions – pollination and seed dispersal – are the focus of this thesis (though not exclusively). These interactions are fundamental to most terrestrial ecosystems and have been particularly well studied, with a large amount of network data freely available.

Mapping network ecology

```
                        ┌─────────────────┐
                        │   Sample data   │
                        │       (1)       │
                        └─────────────────┘
                         ╱               ╲
                        ╱                 ╲
        ┌──────────────────────┐     ┌──────────────────────┐
        │  Describe structure  │────▶│  Dynamics on networks│
        │       (4, 5)         │     │         (3)          │
        └──────────────────────┘     └──────────────────────┘
              │          ╲                       │
              ▼           ╲                      ▼
  ┌──────────────────────┐ ╲  ┌───────────────────────────────┐
  │ Explain community    │  ╲ │ Robustness/persistence/stability│
  │ assembly             │   ╲│         simulations           │
  │       (2)            │    │            (3)                │
  └──────────────────────┘    └───────────────────────────────┘
```

Above is a simplified schematic of network ecology, with chapter numbers given in brackets. Network ecology starts by observing nature to generate sampled data. The structure of these sampled networks can then be described, either a single network, or multiple networks over space and time. Process-based models may then be used to try and explain how observed structure arises. Rather than characterising structure directly, dynamic processes operating *on* networks may be studied, possibly empirically, but most often using simulations. These dynamic studies examine how populations change over time, using the network as a skeleton to determine which species interact. Dynamics models are often used to study the stability or persistence of networks, though these questions can also be tackled from a topological perspective by, for example, deleting nodes or links to assess network robustness. Of course, each of these domains is interrelated, with studies relating sampling methods to structure, or structure to stability, for example.

Thesis structure

Studies of ecological networks are only as good as the data on which they depend. For several types of interaction, most notably pollination and seed dispersal, networks are most often based on visits; that is, a plant and an animal are recorded as interacting when the animal visits the plant. Visitation is not necessarily equivalent to true

pollination or seed dispersal, however, as visits can involve a wide range of interactions that span the mutualistic-antagonistic spectrum. For example, animals can act as 'nectar robbers', piercing flowers to drink nectar without providing any reproductive benefit to the plant, while frugivores can consume seeds, destroying them and eliminating any chance of successful plant establishment. Visitation and pollination have been shown to not necessarily be equivalent (King *et al.* 2013; Ballantyne *et al.* 2015), but the same issue has not been evaluated for seed dispersal. We address this gap in Chapter 1.

Once network data has been obtained, it can be used to answer questions about the processes underpinning community structure. For mutualistic communities, one of the most frequently-observed patterns is a positive correlation between a species' abundance and its generalisation. This pattern has previously been described as a 'chicken-and-egg dilemma' as there are valid *a priori* explanations for both directions of the relationship; that is, that abundance drives generalisation or that generalisation drives abundance. In Chapter 2 we resolve this dilemma, for the first time using independent data on animal abundance.

While Chapter 2 considers the *causes* of structure, Chapter 3 considers its *consequences*. There has been an abundance of important and insightful studies characterising the relationship between mutualistic network structure and dynamic processes operating on such networks (Thébault & Fontaine 2010; Saavedra *et al.* 2011; Rohr *et al.* 2014). However, these studies have tended to focus on the implications of whole-network structure or the role of individual nodes, while links between nodes have been largely neglected. We redress this balance in Chapter 3, by considering the contribution of individual links to the structural stability of networks, and how this contribution relates to link vulnerability.

All the above chapters operate within the dominant methodological paradigm of mutualistic network research: characterising network structure with one-dimensional indices, such as nestedness, modularity or degree. This paradigm has prevailed for over 30 years, and is the basis of many seminal insights into community structure (Bascompte *et al.* 2003; Olesen *et al.* 2007). However, the advantages of indices lie in

their intuitiveness and interpretability, rather than their ability to capture network topology in great detail. In Chapter 4, we consider the amount of structural detail captured by indices relative to an alternative way of describing networks: motifs. Chapter 5 builds on this work, describing a software package in R, MATLAB and Python that enables very fast motif analyses. The discussion then draws some overall conclusions and looks at future directions.

References

Ballantyne, G., Baldock, K.C.R. & Willmer, P.G. (2015). Constructing more informative plantpollinator networks: Visitation and pollen deposition networks in a heathland plant community. *Proc. R. Soc. B Biol. Sci.*, 282, 20151130.

Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci.*, 100, 9383–9387.

King, C., Ballantyne, G. & Willmer, P.G. (2013). Why flower visitation is a poor proxy for pollination: Measuring single-visit pollen deposition, with implications for pollination networks and conservation. *Methods Ecol. Evol.*, 4, 811–818.

Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. (2007). The modularity of pollination networks. *Proc. Natl. Acad. Sci.*, 104, 19891–19896.

Rohr, R.P., Saavedra, S. & Bascompte, J. (2014). On the structural stability of mutualistic systems. *Science (80-.).*, 345, 1253497–1253497.

Saavedra, S., Stouffer, D.B., Uzzi, B. & Bascompte, J. (2011). Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478, 233–235.

Thébault, E. & Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science (80-.).*, 329, 853–856.

1 | Moving from frugivory to seed dispersal: incorporating the functional outcomes of interactions in plant-frugivore networks

This chapter was published as Simmons, B. I., Sutherland W. J., Dicks, L. V., Albrecht, J., Farwig, N., García, D., Jordano, P. and González-Varo, J. P. (2018). Moving from frugivory to seed dispersal: incorporating the functional outcomes of interactions in plant-frugivore networks. Journal of Animal Ecology, 87, 995–1007.

Abstract

1. There is growing interest in understanding the functional outcomes of species interactions in ecological networks. For many mutualistic networks, including pollination and seed dispersal networks, interactions are generally sampled by recording animal foraging visits to plants. However, these visits may not reflect actual pollination or seed dispersal events, despite these typically being the ecological processes of interest.

2. Frugivorous animals can act as seed dispersers, by swallowing entire fruits and dispersing their seeds, or as pulp peckers or seed predators, by pecking fruits to consume pieces of pulp or seeds. These processes have opposing consequences for plant reproductive success. Therefore, equating visitation with seed dispersal could lead to biased inferences about the ecology, evolution and conservation of seed dispersal mutualisms.

3. Here we use natural history information on the functional outcomes of pairwise bird-plant interactions to examine changes in the structure of seven European plant-frugivore visitation networks after non-mutualistic interactions (pulp-pecking and seed predation) have been removed. Following existing knowledge of the contrasting structures of mutualistic and antagonistic networks, we hypothesised a number of

changes following interaction removal, such as increased nestedness and lower specialisation.

4. Non-mutualistic interactions with pulp peckers and seed predators occurred in all seven networks, accounting for 21–48% of all interactions and 6–24% of total interaction frequency. When non-mutualistic interactions were removed, there were significant increases in network-level metrics such as connectance and nestedness, while robustness decreased. These changes were generally small, homogenous and driven by decreases in network size. Conversely, changes in species-level metrics were more variable and sometimes large, with significant decreases in plant degree, interaction frequency, specialisation and resilience to animal extinctions, and significant increases in frugivore species strength.

5. Visitation data can overestimate the actual frequency of seed dispersal services in plant-frugivore networks. We show here that incorporating natural history information on the functions of species interactions can bring us closer to understanding the processes and functions operating in ecological communities. Our categorical approach lays the foundation for future work quantifying functional interaction outcomes along a mutualism–antagonism continuum, as documented in other frugivore faunas.

Introduction

Interspecific interactions play a crucial role in the ecological and evolutionary dynamics of populations and communities (Roemer, Donlan & Courchamp, 2002; Thompson, 2009), determining energy fluxes and mediating key ecological functions, such as mycorrhizal-mediated mineral nutrition and animal-mediated pollination and seed dispersal (Bascompte & Jordano, 2013). During the last decade, networks have increasingly been used to study the complex web of interactions that structure ecological communities (Heleno et al., 2014). The network approach allows ecologists to simultaneously 'see the forest and the trees' (Heleno et al., 2014); that is, to analyse emergent properties at the community level while also assessing the functional role of individual species within communities. For example, the analysis of network-level

metrics has shown that mutualistic networks are more nested than antagonistic networks (Thébault & Fontaine, 2010), that specialization of pollination and frugivory networks decreases with latitude (Schleuning et al., 2012), and that non-native frugivores have more connected and generalized interactions with local fleshy-fruited plant communities than native frugivores (García, Martínez, Stouffer & Tylianakis, 2014). Species-level metrics have revealed, for example, that the role of invasive species within plant-pollinator networks can be predicted by their role in networks from their native range (Emer, Memmott, Vaughan, Montoya & Tylianakis, 2016), and that dependence of frugivore species on fruits is positively related to their strength in seed dispersal networks (Fricke, Tewksbury, Wandrag & Rogers, 2017).

There is, however, growing interest in understanding the functional role of species interactions in ecological networks (Ballantyne, Baldock & Willmer, 2015; Heleno et al., 2014; King, Ballantyne & Willmer, 2013). However, many networks are sampled by direct observation (Jordano, 2016). For example, pollination and seed dispersal networks are generally sampled by observing animals visiting plants to feed on their flowers or fruits (Chacoff et al., 2012; Plein et al., 2013). In both these mutualisms, visits describe food intake in animals, but not necessarily pollination or seed dispersal in plants. This issue has recently been examined for plant-pollinator interactions, showing that visitation does not necessarily mean effective pollination (Ballantyne et al., 2015; King et al., 2013). Consequently, network structure can change when incorporating detailed information on the functional outcomes of species interactions (Ballantyne et al., 2015; Carlo & Yang, 2011).

To our knowledge, no study has evaluated this issue in plant-frugivore networks (but see Genrich, Mello, Silveira, Bronstein & Paglia, 2017; Montesinos-Navarro, Hiraldo, Tella & Blanco, 2017), despite research suggesting that it could be important (Albrecht, Neuschulz & Farwig, 2012; Farwig, Schabo & Albrecht, 2017; González-Varo, 2010; Jordano, 1994; Jordano & Schupp, 2000; Snow & Snow, 1988). For plants, fleshy fruits represent the reward they offer for effective seed dispersal by animals (endozoochory), while for animals, fruits and seeds represent sources of food (Herrera, 2002; Janzen, 1983; Jordano, 2013). Frugivorous animals, notably birds and mammals, can process

fleshy fruits by either (i) swallowing entire fruits and defecating or regurgitating viable seeds (legitimate seed dispersers), or (ii) pecking or biting fruits for their pulp (pulp peckers) or seeds (seed predators) (see Snow & Snow (1988)). Legitimate seed dispersers are true mutualists as they disperse plant progenies away from the maternal environment and allow the colonization of new sites (Traveset, Heleno & Nogales, 2014). Conversely, seed predators are antagonists that destroy plant progeny (up to approximately 80% in some plant populations; (González-Varo, 2010)). Pulp peckers are between these two extremes (Fig. 1a) because they neither disperse nor destroy seeds; they usually peck fruit and the seed eventually drops to the ground (Jordano & Schupp, 2000). Some frugivore species may exhibit combinations of these behaviours when feeding on specific fruit species, falling into a continuum of interaction outcomes (Perea, Delibes, Polko, Suárez-Esteban & Fedriani, 2013). Clearly, frugivore visitation and seed dispersal are not equivalent, and plant reproductive success can be strongly influenced by the relative frequency of each type of interaction with frugivores (Schupp, Jordano & Gomez, 2010). We may envisage a gradient of outcomes, depending on the particular pairwise interaction; the above categories representing a categorical summary of variable, context-dependent outcomes.
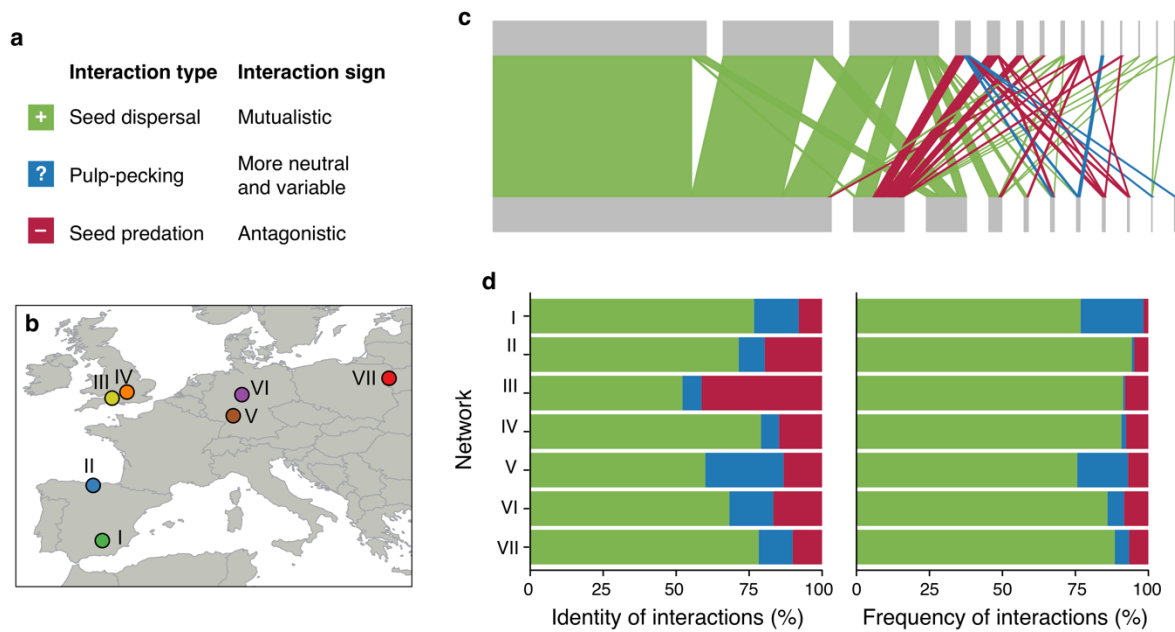
Figure 1. (a) Types of interactions between avian frugivores and fleshy fruits, and sign of the interaction from the plant's perspective. (b) Location and codes (roman numbers and colours) of the bird-fruit visitation networks included in this study (I: Jordano *unpublished*; II: García 2016; III: Sorensen 1981; IV: Snow & Snow 1988; V: Plein *et al.* 2013; VI: Stiebel & Bairlein 2008; VII: Farwig *et al.* 2017). (c) Representation of one of the studied networks (III); note that some frugivore species can have different interaction types depending on the plant species they feed on. (d) Frequency (%) of the different interaction types in the studied networks in terms of identity and quantity.

Importantly, most plant-frugivore networks analysed in recent studies, and those available in open-access network repositories, such as the Web of Life (www.web-of-life.es), are visitation networks (e.g. 16 out of 18 in Schleuning et al., 2014), which include both pulp-pecking and seed-predation interactions (see Fig. 1). This may not be a problem for questions related to the trophic specialization of frugivores (Dalsgaard et al., 2017). However, many studies using these networks aim to understand seed dispersal at the community level (Pigot et al., 2016; Schleuning et al., 2012; Schleuning et al., 2014) and its resilience to global change pressures (Fortuna & Bascompte, 2006; Schleuning et al., 2016), as well as identifying frugivore species that contribute the core of seed dispersal services (Fricke et al., 2017). Therefore, assessing structural differences between plant-frugivore visitation networks and true seed dispersal networks is important because strong biases might lead to incorrect inferences about the ecology, evolution and conservation of this mutualism.

Here, we classify all pairwise 'bird-fruit' interactions in seven European 'bird-fruit' visitation networks, as seed dispersal, pulp pecking or seed predation. We then evaluate how network structure and species structural roles (see Table 1) changed after removing the non-mutualistic interactions (seed predation and pulp pecking). We focus on European networks because they share a biogeographical region (Western Palearctic; Fig. 1b) and there is detailed natural history information available on the functional outcome of each pairwise bird-fruit interaction (e.g. Snow & Snow, 1988). Such information is crucial because the functional role of some bird species can change depending on the fruit species they feed on. It is important to note that our approach is primarily focussed on the fruit removal stage (the 'departure stage') of plant-frugivore interactions (Herrera, 2002), an easily-obtainable proxy for actual seed dispersal success. However, true dispersal not only requires viable seeds to be removed from a plant, but also for seeds to be dispersed to suitable locations. Therefore, a complete assessment of seed dispersal effectiveness requires consideration of post-removal processes, from seed deposition to seedling establishment (Schupp, Jordano & Gómez, 2017; Spiegel & Nathan, 2010; Wenny & Levey, 1998).

Materials and Methods

Study networks

We assembled a database of 1051 plant-frugivore interactions from seven European quantitative visitation networks (Fig. 1b). Some interactions occurred in more than one network, resulting in 681 unique interactions between 62 bird species spanning 19 families, and 69 plant species from 23 families. All interactions were between plants and birds, except four plant-mammal interactions in network VII, which were excluded from subsequent analyses.

In five networks (I-V), interaction weights were visitation frequency. In the other two networks (VI and VII) weights were visitation rates, which were converted to visitation frequency by multiplying the rate for a plant species by time spent sampling that species. Where this did not result in an integer, values were rounded to the nearest

whole number. We used visits as interaction weights because (i) it allowed use of quantitative null models, which require integer data, and (ii) weights had to be standardised across all seven networks, and visitation rates were not available for networks I-V.

Interaction classification

We classified each bird-plant interaction as 'seed dispersal', 'pulp-pecking' or 'seed predation'. Generally, a given bird species fits into one of these categories (Herrera, 1984). However, we did the classification at the interaction level because a bird species can have different interaction types depending on the plant species it feeds on (Fig. 1c) (Snow & Snow, 1988). For example, the Woodpigeon (*Columba palumbus*) can disperse large seeds with a hard coat (González-Varo, Carvalho, Arroyo & Jordano, 2017), but its gut typically destroys smaller and weaker seeds (Snow & Snow, 1988). Some bird species can even have different interaction types with the same plant species (Jordano & Schupp, 2000; Snow & Snow, 1988), but detailed data at the fruit level from network VII allowed us to validate our three-category classification according to the predominant interaction type (Fig. S1).

Classification data were directly available for 498 unique interactions (73.1%). Data came from four of the seven studied networks, namely *Birds and Berries* (Snow & Snow, 1988) for network IV and unpublished information from networks I (Jordano *unpublished*), II (García, 2016) and VII (Farwig et al., 2017). For the remaining 183 unique interactions (26.9%), we inferred the interaction type from the above sources and other references (Simmons et al., 2018). Inference was based on interactions with congeneric species and/or interactions with plant species with similar fruits and seeds (such as drupe or berry, fruit size). For example, we inferred that the Greenfinch (*Carduelis chloris*) consumed *Sorbus aria* seeds because one data source (Snow & Snow, 1988) classified greenfinches as predators of similar *Sorbus aucuparia* seeds.

Network-level analysis

We first assessed how the removal of non-mutualistic interactions changed network structure at the whole-network level. We evaluated changes in six network-level metrics

commonly used in ecological studies (network size, weighted connectance, weighted nestedness, H_2', modularity and robustness) each of which we hypothesised to change in a certain direction following interaction removal (see Table 1 for metric definitions and their associated hypotheses). For each metric, we calculated its value (i) for the original visitation network with all interactions, and (ii) after the removal of the non-mutualistic interactions (predatory and pulp pecking interactions). Many network metrics are sensitive to changes in network size. To control for this, we additionally used a null model approach, where metric values were -transformed. In -transformation, the mean value of a metric across 1000 null networks is subtracted from the empirical network metric value, to describe the extent to which the metric deviates from a random expectation (Dalsgaard et al., 2017). We used two null models: the Patefield model, which constrains network size and marginal totals, and 'quasiswap_count', which constrains network size, marginal totals and connectance (the proportion of species pairs that interact in the network) (Oksanen et al., 2016). The Patefield algorithm can generate unrealistic degree distributions and inflated Type II error rates (Bascompte and Jordano, 2013). However, the issue of null model building for networks is still unresolved and currently there is no better alternative than running different null models, some more conservative, others less conservative. That is the approach we have taken here: we use both the Patefield algorithm and the less conservative 'quasiswap_count' algorithm.

We used one-tailed Wilcoxon paired rank tests to determine whether network metrics consistently decreased or increased following the removal of non-mutualistic interactions. We used one-tailed tests because we adopted a hypothesis-driven approach to test directional changes in network metrics. For example, we did not test whether nestedness changed in any direction after interaction removal; instead we explicitly tested whether nestedness increased. This is because we hypothesized an increase in nestedness following the removal of non-mutualistic interactions as mutualistic systems tend to be nested (Fontaine et al., 2011; Thébault & Fontaine, 2010). We adopted this approach for all metrics, with the hypothesized direction of change (and consequently the direction of the one-tailed tests) given in Table 1. Additionally, we carried out one-tailed Spearman's rank correlation tests to test whether the ranking

of networks for each metric differed following interaction removal. A positive Spearman's correlation between metric values before and after removal is expected if there are no changes in ranks (assemblages respond to the removal of non-mutalistic interactions in a consistent way), whereas such correlation is not expected if there are significant changes in ranks. Therefore, the direction of the tests was informed by the null hypothesis of no change in the ranks (an expected positive relationship). We consider a non-significant Spearman's ρ to indicate a change in the ranks across networks. All these analyses were performed for the absolute metric values and the two sets of null-corrected values.

Table 1. Network- and species-level metrics considered in this study.

Metrics (level)	Definition	Hypothesised change after removal of non-mutualistic interactions
Network level		
Size	The total number of species in the network	Decrease: due to the removal of exclusively non-mutualistic frugivore species and plant species that only interacted with non-mutualistic frugivores.
Weighted connectance	Linkage density divided by the total number of species in the network (Tylianakis, Tscharntke & Lewis, 2007)	Increase: due to (i) decrease in network size, and (ii) because antagonists are expected to have a narrower niche than mutualists, and therefore lower degree, suggesting that their removal should result in connectance increase (Blüthgen et al., 2007).
Weighted nestedness	Weighted NODF: a quantitative index for nestedness. Higher values indicate greater nestedness (Almeida-Neto & Ulrich, 2011).	Increase: Mutualistic systems tend to be nested, while antagonistic systems tend to be modular (Fontaine et al., 2011; Thébault & Fontaine, 2010). Therefore, when removing antagonistic interactions, we expect an increase in nestedness.
H_2'	A measure of network specialisation. It ranges between 0 (no specialisation) and 1 (complete specialisation).	Decrease: Predators tend to be more specialised than mutualists, therefore specialisation decreases when they are removed (Fontaine et al., 2011; Morris et al., 2014).
Weighted modularity	The LPAwb+ algorithm, a measure of community partitioning in quantitative networks (Beckett, 2016).	Decrease: Antagonistic systems tend to be more modular than mutualistic systems (Fontaine et al., 2011; Thébault & Fontaine, 2010). Therefore, when removing antagonistic interactions, we expect a decrease in modularity.
Robustness	Area under the curve of bird species removed versus plant species remaining.	Decrease: With fewer animal partners, on average plants will have less redundancy and undergo dispersal failure sooner. Therefore, robustness will be lower.
Species level (plants)		
Degree	The number of species a given plant species interacts with	Decrease: Any plant species with non-mutualistic partners will undergo a decrease in degree. Plant species which exclusively interact with mutualistic partners will undergo no change in degree. Therefore, on average, a decrease is expected.
Interaction frequency	The total interaction frequency of a given species	Decrease: Any plant species with non-mutualistic partners will undergo a decrease in interaction frequency due to a decrease in degree. Plant species which exclusively interact with mutualistic partners will undergo no change in interaction frequency. Therefore, on average, a decrease is expected.
d'	Specialisation of a species, measured as deviation from a random selection of its partners (Blüthgen et al., 2006).	Decrease: Predators tend to be more specialised than mutualists, therefore specialisation decreases when they are removed (Blüthgen et al., 2007).
Resilience (R_{75})	The number of animal partners that are lost before a given plant species undergoes dispersal failure.	Decrease: Decreases in degree and interaction frequency mean that fewer partners will need to be removed until a plant species undergoes dispersal failure, resulting in a resilience decrease.
Species level (frugivores)		
Species strength	Sum of dependencies of plant species (Bascompte et al., 2006). It quantifies a frugivore species' relevance across all the fleshy-fruited plant community.	Increase: Plants will depend more on seed dispersers because dependencies in the original networks are distributed among mutualists and non-mutualists; after the removal of non-mutualistic interactions, dependencies will be spread among fewer partners and will therefore, on average, be higher.

To understand the processes driving changes in network metrics, we again used one-tailed Spearman's rank correlation to test whether the magnitude of the change in network metrics following interaction removal correlated with the proportion of non-mutualistic links removed from the networks. The direction of the one-tailed test was determined by the hypotheses in Table 1.

As we conduct multiple tests, there is an increased probability of incorrectly rejecting the null hypothesis of no change in network metrics (Type I errors). We used the equation given by Moran (2003), based on a Bernoulli process, to calculate the probability of a given number of significant tests from a given number of trials. The probability, p, is given by the equation $p = [N!/(N - K)!K!] \times \alpha^K(1 - \alpha)^{N - K}$, where N is the number of tests conducted and K is the number of tests below the significance level α.

Species-level analysis

We assessed how the removal of non-mutualistic interactions affected individual species, by examining changes in five species-level metrics: four involving plant species (degree, interaction frequency, d' and resilience) and one involving frugivore species (species strength) (see Table 1 for metric definitions and their associated hypotheses). We calculated metric values for all species in all networks (i) in the original visitation networks with all interactions, and (ii) after the removal of all non-mutualistic interactions. If interaction removal caused a species to lose all its links, it has a degree of zero and an interaction frequency of zero. We retained metric values of degree and interaction frequency for species that lost all links as excluding these would lead to an underestimation of mean changes in both metrics. However, the other metrics used in our analyses have a value of NA for a species with no links. We therefore excluded these NA metric values. We used one-tailed Wilcoxon signed rank tests to determine whether metrics changed significantly following interaction removal. We performed tests for all species pooled together, and separately for each network. The direction of the tests was informed by the hypothesised direction of change in each metric, as stated in Table 1. We additionally tested whether the ranking of species for each metric differed following interaction removal using one-tailed Spearman's rank correlation tests. The direction of the tests was informed by the null hypothesis of no change in the ranks, therefore, an

expected positive relationship. We consider a non-significant Spearman's ρ to indicate a change in the ranks across networks.

Metric calculation

All network metrics and null models, except modularity, Robustness and Resilience$_{75}$, were calculated using the R package 'bipartite' version 2.06.1 (Dormann, Fründ, Blüthgen & Gruber, 2009; R Core Team, 2015). Modularity was calculated using the LPAwb+ code available on GitHub (https://github.com/sjbeckett/weighted-modularity-LPAwbPLUS) (Beckett 2016). For each modularity calculation, the LPAwb+ algorithm was run once. However, due to the stochastic nature of the algorithm, we also repeated our analyses, running the LPAwb+ algorithm 1000 times for each modularity calculation. All results were unchanged by this additional analysis. Robustness and Resilience$_{75}$ were calculated using a topological coextinction model, similar to that developed by Schleuning et al. (2016). In this model, we removed bird species in order of least to most interaction frequency (a proxy for abundance), as low abundance species are likely to be most vulnerable to anthropogenic pressures (Pimm, Lee Jones & Diamond, 1988). Plant species were considered to have undergone dispersal failure when they had lost 75% of their interaction frequency. Robustness was measured as the area under the curve of bird species removed versus plant species remaining, producing a value between 0 and 1 (Burgos et al., 2007; Pocock, Evans & Memmott, 2012). Resilience of a given plant species was measured as the proportion of bird species that had to be removed from the network for it to undergo dispersal failure (Resilience$_{75}$).

Removing only truly antagonistic (seed-predation) interactions

We also performed all the analyses described above when only removing predatory interactions from the original visitation networks (leaving pulp-pecking and seed dispersal interactions). This was because several of our hypotheses consider true antagonism (Table 1), whereas pulp peckers can be considered cheaters rather than antagonists because they do not destroy seeds and may sometimes disperse the seeds (Jordano & Schupp, 2000; Snow & Snow, 1988). This could affect changes in network metrics as the extent of modularity and nestedness in antagonistic networks is closely related to the degree of interaction intimacy (Pires & Guimarães, 2013). This is very

generalized for pulp peckers (insectivores) but specialized for seed predators (granivores), like finches, whose bill morphology determines the size of seeds they can break and eat (Newton, 1967).

Results

Prevalence of non-mutualistic interactions

We found that both predatory and pulp-pecking interactions occurred in all seven communities, though their prevalence varied among networks (Fig. 1). Non-mutualistic interactions comprised between 21% and 48% of links and between 5.7% and 24% of interaction frequency (Fig. 1). Predatory interactions comprised between 8% and 41% of links and between 1.6% and 8.3% of interaction frequency. Pulp-pecking interactions comprised between 6.2% and 26% of links and between 0.6% and 22% of interaction frequency.

At the species level, we found that 63.2% of plant species were involved in non-mutualistic interactions (between 48.0% and 90.9% of species in each network) (Fig. 2a, c). For birds, we found that 45.6% of species were involved in non-mutualistic interactions (between 26.7% and 62.1% of species in each network) (Fig. 2b, d). The proportion of species' links and interaction frequency that were seed dispersal, pulp-pecking and seed predation is shown in Fig. 2; the distribution of non-mutualistic interactions is negatively skewed, but for many species constitute a meaningful proportion. This is particularly true for bird species where 34.7% of species have no seed dispersal interactions.

Almost 80% of the interaction frequency with seed dispersers involved just two bird families (Turdidae: 66.9%; Sylviidae: 12.3%). Two bird families also accounted for 77–78% of the interaction frequency with pulp peckers (Paridae: 49.2%; Fringillidae: 28.5%) and seed predators (Paridae: 27.1%; Fringillidae: 50.0%).
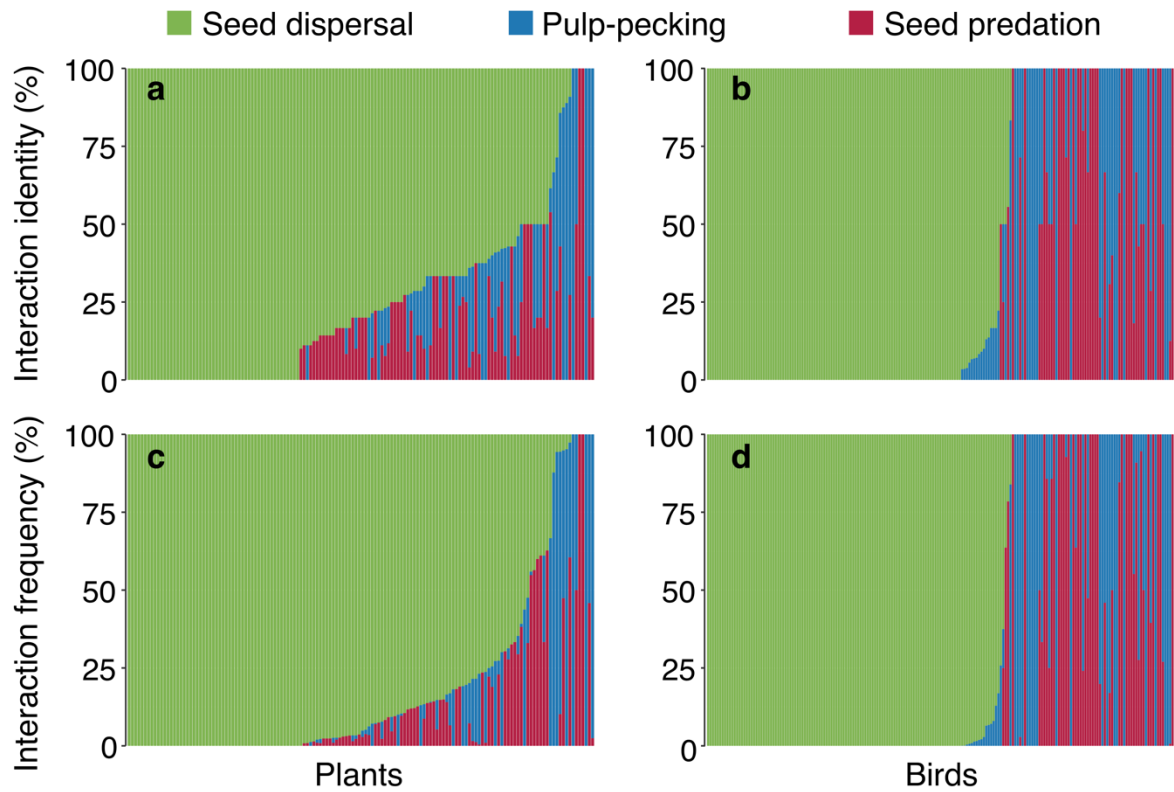
Figure 2. The composition of species' links (a, b) and interaction frequency (c, d) for each plant (a, c) and bird (b, d) species across all networks. Each bar shows the proportion of a species' links or interaction frequency which are seed dispersal (mutualistic), pulp-pecking or seed predation (non-mutualistic). Species are placed in order of decreasing proportion of links or interaction frequency which are seed dispersal.

Changes in network-level metrics

We found small, but consistent, changes in four network-level metrics after removing non-mutualistic interactions (Fig. 3; Table 2). Network size (Fig. 3a) and robustness (Fig. 3f) decreased significantly when interactions were removed, while weighted connectance (Fig. 3b) and weighted nestedness (Fig. 3c) increased significantly. No significant changes were found in H_2' (Fig. 3d) or modularity (Fig. 3e). The probability of finding four significant changes from six trials at a 0.05 significance level is 0.0000846 (Moran, 2003). Therefore, despite the inflated Type I error rate resulting from multiple tests, the number of significant results we found was substantially greater than expected from chance alone. In addition, we found a non-significant rank correlation between the original and the modified network for weighted nestedness, indicating that removal of non-mutualistic interactions led to changes in ranks across networks (Fig. 3c).

When null models were used to control for changes in network size, changes in weighted connectance and weighted nestedness were not significant (Fig. S4). This indicates that the significant changes in these metrics were driven by the decrease in network size. Conversely, decreases in robustness were still significant when corrected using both null models (Fig. S4). This indicates that changes in robustness were more than expected from the size decrease alone, and were driven by structural changes beyond those in connectance, as the 'quasiswap_count' null model algorithm constrains size, marginal totals and connectance.
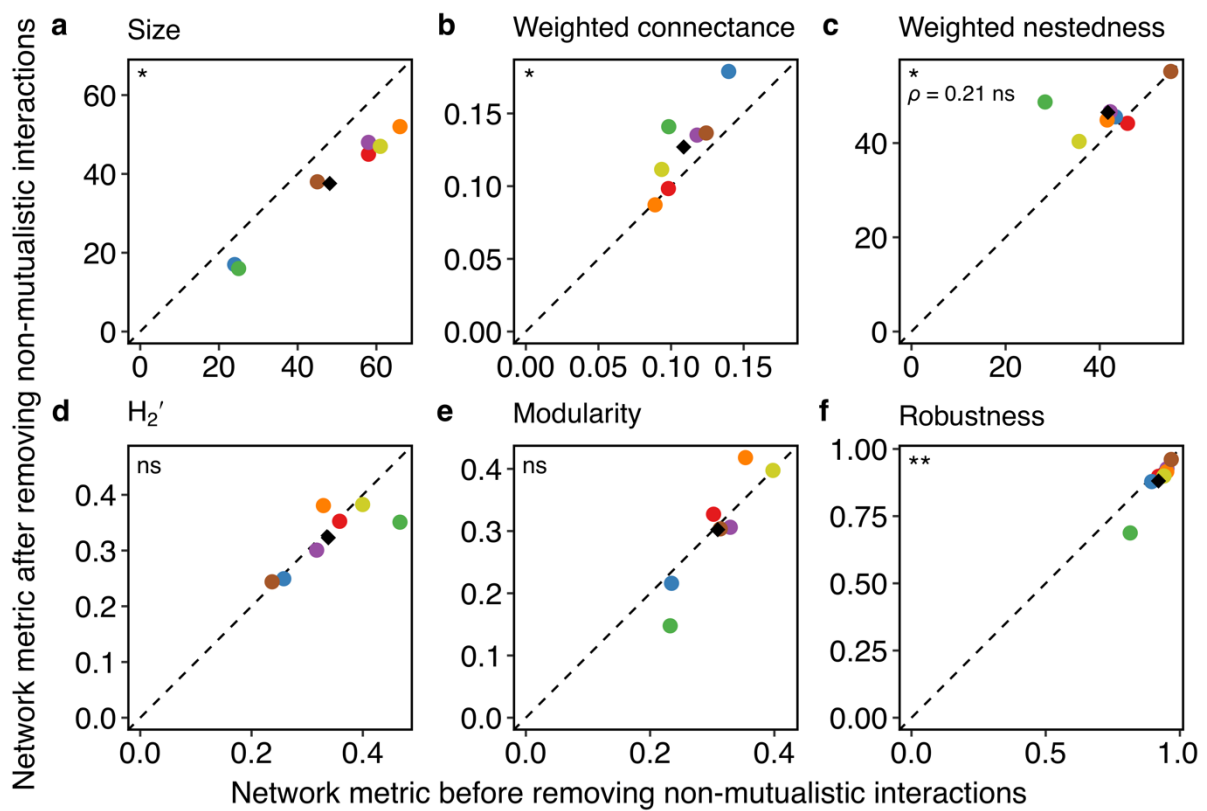

Figure 3. Changes (y axes) in the studied network-level metrics after the removal of non-mutualistic interactions (seed predation and pulp pecking). Colour codes denote network identity (see Fig. 1b). The black diamonds are mean values across networks. The dashed line is $y = x$, indicating the position of points if there was no change in metric values. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*ns*: non-significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$). Unless specified, all Spearman's ρ are significant ($\rho \geq 0.75$, $P < 0.05$); we consider a non-significant ρ to indicate a change in the ranks across networks.

In general, the magnitude of the changes was not significantly related to the proportion of links removed (Fig. S2). The exception was robustness, which significantly decreased with proportion of links removed (Spearman's ρ = –0.71, P = 0.044; Fig. S2).

Table 2. Mean change and variation in raw network-level metrics following the removal of non-mutualistic interactions. Significant changes are shown in bold (0.05 significance level).

Metric	Mean (absolute)	Mean (%)	Range across networks	Coefficient of variation (%)
Size	**–10.57**	**–23.5**	–14 to –7	29
Weighted connectance	**0.02**	**16.2**	0.00 to 0.04	95
Weighted nestedness	**4.77**	**15.0**	–1.72 to 20.33	152
H_2'	–0.01	–2.9	–0.12 to 0.05	336
Modularity	–0.01	–4.0	–0.08 to 0.06	698
Robustness	**–0.04**	**–4.5**	–0.13 to –0.01	103

Changes in species-level metrics

At the species level, we found that several metrics significantly changed following the removal of non-mutualistic interactions (Fig. 4), and that these results were generally consistent across networks (see Table 3, Table 4). On average, species lost 2.1 partners (26.4%). Remarkably, the maximum change in plant species degree was –11. Additionally, some plant species lost all their links: this phenomenon occurred in four networks, affecting between 3.3% and 27.0% of plant species. There were significant decreases in plant degree, interaction frequency, d' and Resilience$_{75}$ (Fig. 4a-d), while frugivore species strength significantly increased (Fig. 4f). Results for each network separately largely agree with the overall Wilcoxon results (Table 4), though a few metrics in some networks were unaffected by removal of non-mutualistic interactions (Table 4). Finally, in one network, one metric differed in its rank following interaction removal: the Spearman's rank test for d' in network III was not significant (ρ = 0.50, P = 0.108), indicating that species' relative values of d' changed following removal of non-mutualistic interactions.
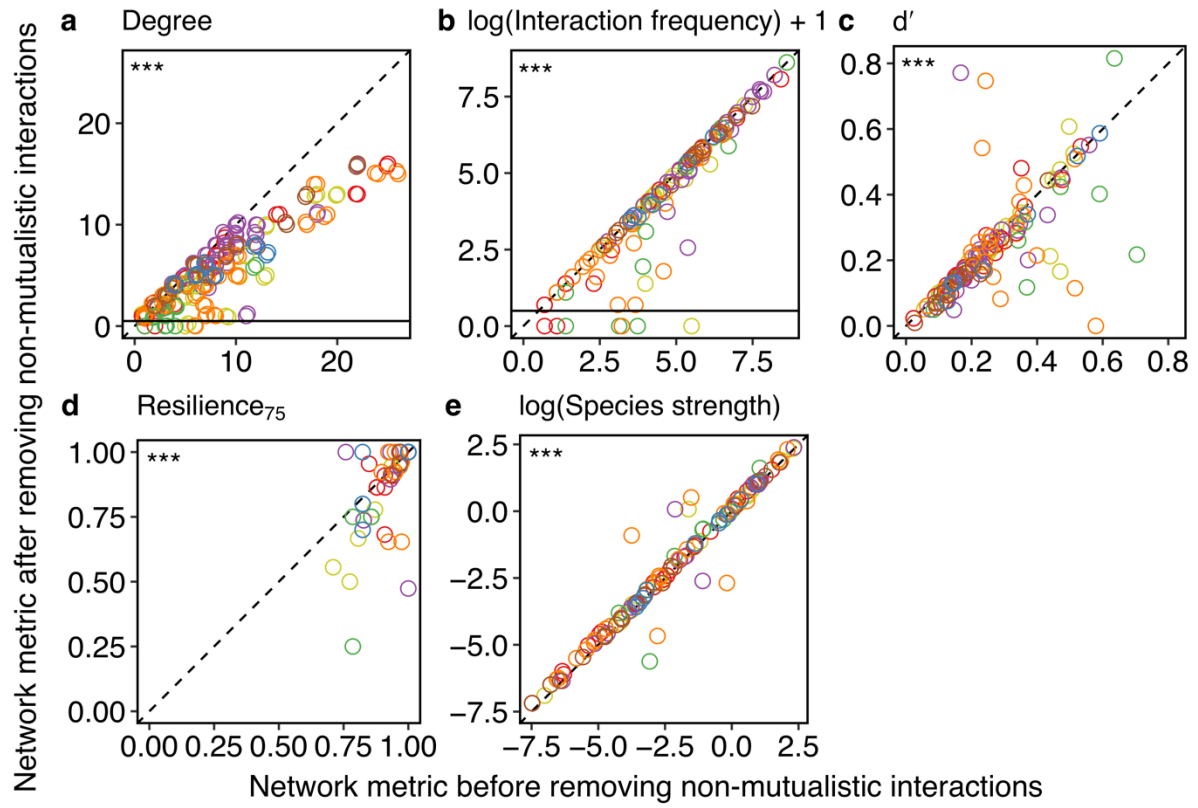
Figure 4. Changes (*y* axes) in species-level metrics for plants (a-d) and frugivores (e) after the removal of non-mutualistic interactions (seed predation and pulp pecking). Colour codes denote network identity (see Fig. 1b). The dashed line is *y = x*, indicating the position of points if there was no change in metric values. Points below the horizontal black lines in panels (a) and (b) highlight those species that lose all their partners (a: degree) and interactions (b: frequency) after pruning. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*ns*: non-significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

Table 3: Changes and variation in species-level metrics following the removal of non-mutualistic interactions. Significant changes are shown in bold (0.05 significance level). The mean change in each metric for each network was calculated and, then, an overall mean was obtained by calculating the mean of the mean changes in each network. The range of the mean change across networks is also reported, as well as the range of change across species in parentheses. The coefficient of variation was calculated across all species in all networks; in parentheses we show the range of coefficients of variation when calculated for each network separately.

Metric	Mean (absolute)	Mean (%)	Range across networks (species)	Coefficient of variation (%)
Degree (plants)	**–2.10**	**–26.4**	–3.26 to –1.40 (–11 to 0)	121 (87 to 177)
Interaction frequency (plants)	**–44.26**	**–19.8**	–69.34 to –7.86 (–1373 to 0)	294 (78 to 423)
d' (plants)	**–0.03**	**–11.4**	–0.12 to –0.01 (–0.58 to 0.60)	464 (80 to 1254)
Resilience$_{75}$ (plants)	**–0.03**	**–3.6**	–0.13 to 0.00 (–0.54 to 0.24)	346 (138 to 2103)
Species strength (frugivores)	**0.14**	**35.40**	0.04 to 0.33 (–0.76 to 2.19)	282 (143 to 305)

Table 4. Results of species-level Wilcoxon tests per network (I – VII) for each metric. '+' indicates that the metric increased following the removal of non-mutualistic interactions, while '–' indicates a decrease. *, ** and *** denote $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively (ns: non-significant differences).

Metric	Change	I	II	III	IV	V	VI	VII
Degree (plants)	–	**	*	**	***	***	***	**
Interaction frequency (plants)	–	**	*	**	***	***	***	**
d' (plants)	–	*	**	*	***	ns	***	***
Resilience$_{75}$ (plants)	–	**	ns	*	ns	*	**	ns
Species strength (frugivores)	+	***	**	*	**	**	***	***

Removal of seed predation interactions

When only seed predator interactions were removed, changes in network- and species-level metrics followed the same direction as when all non-mutualistic interactions were removed, although they were smaller in magnitude (see Appendix S1). Results of 'quasiswap_count' null models showed that changes in H_2' and weighted nestedness were driven by changes in network structure rather than just decreases in network size (Appendix S1). Moreover, changes in H_2' ($\rho = -0.86$, $P = 0.012$) and modularity ($\rho = -0.79$, $P = 0.024$) were significantly negatively correlated with the proportion of links removed from the original visitation networks (Appendix S1). Weighted connectance and weighted nestedness were positively related with the proportion of links removed from the original networks, yet these correlations were marginally significant ($\rho = 0.61$ and $P = 0.083$ for both metrics; Appendix S1).

Discussion

Here we disentangled seed-dispersal interactions (mutualism) from pulp-pecking (exploitation) and seed predation (antagonism) interactions in European plant-frugivore networks and evaluated changes in network properties when removing non-mutualistic interactions. We found that, at the network level, the magnitude of most changes was small and relatively uniform, suggesting that studies treating plant-frugivore visitation networks as seed dispersal networks are likely robust if they only use network-level metrics (though consideration of processes acting after fruit removal is strictly necessary to infer true dispersal). However, for species-level metrics, changes were generally larger and more variable, indicating the importance of considering natural history to gain insights into the functional roles of species in seed dispersal mutualisms. Ignoring such functional outcomes in visitation networks at the species level may, for instance, overestimate the potential for functional redundancy across frugivore species in the assemblage and the potential for interaction rewiring after loss of frugivore partners.

Changes in network-level metrics

European seed predators and pulp peckers feed on fleshy fruits less frequently than legitimate seed dispersers, which likely explains why non-mutualistic interactions were generally more important in qualitative than quantitative terms. Seed predators have bill morphologies poorly adapted for frugivory, which increases fruit-handling times and lowers energy intake, while pulp peckers' long gut passage time makes fruit an inefficient food source due to its low nutrient content per unit mass (Herrera, 1984). Instead, non-disperser species primarily feed on seeds from dry fruits or insects (Herrera, 1984). This tendency for predatory and pulp-pecking interactions to constitute a relatively small proportion of interaction frequency may explain why network-level metrics generally undergo only small changes after removing non-mutualistic interactions: we have used weighted versions of network-level metrics, where weaker interactions exert less influence on metric values than stronger interactions. Our results therefore suggest that macroecological studies comparing weighted network-level metrics between multiple plant-frugivore visitation networks (e.g. Schleuning et al., 2012) are likely to be robust to the presence of non-mutualistic interactions, especially when comparing H_2' and modularity. For example, Dalsgaard et al. (2017) examined latitudinal patterns in network specialisation (H_2'), finding values ranging between 0.18 and 0.48. Such values are an order of magnitude greater than the mean change in H_2' we observed when removing non-mutualistic interactions and so any biases are unlikely to affect the general conclusions of such studies. Studies are likely robust to changes in weighted and unweighted connectance and nestedness too, given the small magnitude of the mean absolute changes we found in these values (Fig. S3, Table S1).

Removal of non-mutualistic interactions can lead to decreases in network size in two ways: (i) frugivore species can be lost if they only form non-mutualistic interactions with plant species, and (ii) plant species can be lost if, during the sampling period, they only interact with frugivores that destroy their seeds or peck their pulp. The loss of purely non-mutualistic frugivores was the main driver of changes in network size, though plant loss did affect four of the seven networks. The loss of frugivores also helps to explain the decrease in robustness when non-mutualistic interactions were removed:

with fewer frugivore species, plants have fewer partners and less redundancy, meaning that the removal of a single bird species causes plants to lose a greater proportion of their interaction frequency than in the pre-removal network. Our results suggest that studies that do not consider interaction types may overestimate robustness and the redundancy of seed dispersal mutualisms and that this overestimate increases with the proportion of non-mutualistic interactions in a community (Fig. S2). Therefore, inferences about the sensitivity of seed dispersal processes to species loss need to carefully account for the natural history of pair-wise interactions.

When removing only predatory interactions, changes in weighted nestedness and H_2' were more than expected from the decrease in network size alone and were likely related to the antagonistic nature of the removed interactions. For example, the decrease in H_2' may be explained by antagonists forming more specialised interactions than mutualists (Fontaine et al., 2011; Morris, Gripenberg, Lewis & Roslin, 2014). This is expected for seed predators because bill size (depth) determines the size of seeds that predators can break and eat (Newton, 1967). Similarly, the increase in nestedness is supported by a number of previous studies that found nested architectures to be more common in mutualistic than antagonistic networks (Fontaine et al., 2011; Thébault & Fontaine, 2010), a pattern driven by multiple ecological and evolutionary processes (Bascompte, 2010; Vázquez, Chacoff & Cagnolo, 2009). For example, nestedness has been shown to stabilise mutualistic communities, but have a negative effect on the stability of antagonistic systems (Okuyama & Holland, 2008; Thébault & Fontaine, 2010). Conversely, high connectance is associated with stability in mutualistic systems, while antagonistic communities favour less connected architectures (Thébault & Fontaine, 2010). Therefore, with fewer antagonistic interactions and species, connectance and nestedness increase and specialisation decreases. This suggests that even though non-mutualistic interactions make up only a fraction of plant-frugivore networks, the structure of these networks seems to have an imprint of the antagonistic interactions. Finally, the change in relative weighted nestedness that followed interaction removal could be partially due to variations in the prevalence of non-mutualistic interactions (see trend in Fig. 3c), and suggests that comparisons between networks can be confounded by such changes.

Changes in species-level metrics

Changes in species-level metrics were most clear for plant degree and interaction frequency, with many species losing interaction partners (Fig. 4). In some cases the loss of degree was extreme and so, particularly for some species, incorporating information on the functional outcomes of interactions greatly changes inferences about their ecology and evolution. These results suggest that plant species have weaker dispersal interactions with fewer partners than previously recognised. Overall, these differences translated into a small but significant decrease in mean plant resilience to animal removal of −0.03. This value indicates that, after interaction removal, the average percentage of animal species that had to be removed from the network for plant species to undergo dispersal failure decreased by 3%. However, this mean value masks some heterogeneity in species responses (Table 3). Resilience can increase despite plant species having fewer partners on average if the removal of interactions changed the animal removal sequence or if non-mutualistic interactions constituted a large proportion of a species' interaction frequency in the original networks. Therefore, we conclude that, while most estimates of resilience are relatively unchanged by incorporating natural history information, some plant species underwent more major changes, revealing them as more susceptible to global change pressures, including climate change (Schleuning et al., 2016) or disperser extinction (Rumeu et al., 2017).

While decreases in plant d' may initially seem counterintuitive, d' is a measure of the extent to which a species deviates from randomly sampling all available partners and so does not necessarily correlate with measures of specificity, such as degree (Blüthgen, Menzel & Blüthgen, 2006). Instead, with d', species with one partner can be less specialised than species with two partners. For example, if a plant is only visited by one frugivore species, but this frugivore is highly dominant in the community, the plant would have a low d' value. Conversely, if a plant is visited by two very rare frugivores, it would have a high d' value. Antagonistic relationships, such as those between predators and prey or between hosts and parasites, tend to have higher levels of specialisation than mutualistic systems because hosts and prey deploy defences, which constrain the available partners of their enemies (Blüthgen, Menzel, Hovestadt, Fiala & Blüthgen, 2007; Jaenike, 1990).

Increases in frugivore species strength (the sum of dependencies of plant species on frugivores (Bascompte, Jordano & Olesen, 2006) occurred because, before non-mutualistic interactions were removed, plants distributed their dependencies among all avian frugivores. However, once non-mutualistic interactions were removed, plant dependencies shifted entirely to the seed dispersers, thereby increasing their strength values.

Generalizations and limitations

Our analyses represent an attempt to disentangle the variety of mutualistic and antagonistic processes present in plant-frugivore networks to focus on the seed dispersal of plant communities by legitimate seed dispersers. Most 'bird-fruit' interactions involving European frugivorous birds can be easily classified as 'seed dispersal', 'pulp pecking' and 'seed predation' thanks to (i) the availability of necessary data and (ii) the fact that most birds fall within one category (Herrera, 1984; Snow & Snow, 1988). Exceptionally, a few frugivore species may exhibit dual roles (such as European nuthatch *Sitta europaea*; (Jordano & Schupp, 2000; see also Fig. S1); some pulp peckers may pluck fruits and peck them in the branch of a nearby tree, dispersing the seed a few meters (e.g. Great tit *Parus major*; (Jordano & Schupp, 2000; Snow & Snow, 1988). Additionally, certain frugivores that predominantly act as pulp peckers (e.g. Great tit) and seed predators (e.g. Chaffinch *Fringilla coelebs*) have been reported to disperse seeds of fleshy fruits internally, through endozoochory (Cruz, Ramos, da Silva, Tenreiro & Heleno, 2013). However, evidence from the gut content of road-killed (Debussche & Isenmann, 1989) and mist-netted birds (Olesen et al., 2011), and more recently from DNA barcoding applied to dispersed seeds (González-Varo, Arroyo & Jordano, 2014), demonstrates that seed predators and pulp peckers are virtually absent from true seed dispersal networks. These results suggest that networks sampled using methods other than observations of visits, such as by identifying seeds and/or pulp remains recovered from faeces, are likely to be closer in structure to the 'true' seed dispersal networks revealed by removing non-mutualistic interactions than the raw visitation networks containing non-mutualistic interactions. Thus, analysing non-visitation networks, such as seed deposition networks, could be a useful way to circumvent some of the issues

36

raised by this study, bringing us closer to a description of plant-seed disperser community structure (Wang & Smith, 2002).

While our dataset covers a large spatial extent in Europe, further research with a larger database of networks covering other regions, would help assess whether our conclusions hold for other parts of the world. However, we are aware that such simplistic classification may not work in other frugivore groups that fall into a mutualism–antagonism continuum, as occur with tanagers in the neotropics (Moermond & Denslow, 1985), parrots (Montesinos-Navarro et al., 2017) and ungulate mammals (Perea et al., 2013). Whenever dealing with visitation data, the challenge in these groups is to quantify the frequency of different interaction outcomes with multiple plant species in order to incorporate weights of seed dispersal effectiveness into the links of the networks (Schupp et al., 2017).

Finally, while here we have incorporated information on fruit removal, it is important to remember that there remain other natural history details not included in this study. For example, birds of different sizes remove different quantities of seeds in a given visit, and therefore one visit of a small bird is not equivalent to one visit of a larger bird in terms of seed removal (Carlo & Yang, 2011). While visitation frequency is a main component of interaction outcome in generalized plant-frugivore networks, per-visit effects may overcome differences in visitation and alter frugivore effectiveness in significant ways.

Conclusions

Ecological networks constitute a powerful tool to analyse complex interactions between multiple species. We show here that adding more natural history details on the nature of species interactions can bring us closer to understanding the ecological processes and functions they mediate; here, seed dispersal mediated by frugivorous animals. After removing non-mutualistic interactions, changes in network-level metrics were generally small (particularly for H_2', modularity and robustness) and consistent. Importantly, consistent changes at the network level still allow for valid comparisons

among networks. However, at the species level, changes tended to be larger and more variable. This makes it harder to anticipate how individual species might respond if non-mutualistic interactions were removed: while some species may be unaffected, others are highly affected. Importantly, our results show that plants have less frequent interactions with fewer frugivores than previously recognised, and with more limited ecological redundancy. Consequently, we advise caution when using species-level metrics on plant-frugivore visitation networks whenever seed dispersal is the studied ecological process.

References

Albrecht, J., Neuschulz, E.L. & Farwig, N. (2012) Impact of habitat structure and fruit abundance on avian seed dispersal and fruit predation. *Basic and Applied Ecology,* **13,** 347-354.

Almeida-Neto, M. & Ulrich, W. (2011) A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling & Software,* **26,** 173-178.

Ballantyne, G., Baldock, K.C.R. & Willmer, P.G. (2015) Constructing more informative plant–pollinator networks: visitation and pollen deposition networks in a heathland plant community. *Proceedings of the Royal Society B: Biological Sciences,* **282,** 20151130.

Bascompte, J. (2010) Structure and dynamics of ecological networks. *Science,* **329,** 765-766.

Bascompte, J. & Jordano, P. (2013) *Mutualistic networks*. Princeton University press, Princeton.

Bascompte, J., Jordano, P. & Olesen, J.M. (2006) Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science,* **312,** 431-433.

Beckett, S.J. (2016) Improved community detection in weighted bipartite networks. *Royal Society Open Science,* **3,** 140536.

Blüthgen, N., Menzel, F. & Blüthgen, N. (2006) Measuring specialization in species interaction networks. *BMC Ecology,* **6,** 9.

Blüthgen, N., Menzel, F., Hovestadt, T., Fiala, B. & Blüthgen, N. (2007) Specialization, constraints, and conflicting interests in mutualistic networks. *Current Biology,* **17,** 341-346.

Burgos, E., Ceva, H., Perazzo, R.P.J., Devoto, M., Medan, D., Zimmermann, M. & María Delbue, A. (2007) Why nestedness in mutualistic networks? *Journal of Theoretical Biology,* **249,** 307-313.

Carlo, T. & Yang, S. (2011) Network models of frugivory and seed dispersal: Challenges and opportunities. *Acta Oecologica,* **37,** 619-624.

Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012) Evaluating sampling completeness in a desert plant–pollinator network. *Journal of Animal Ecology,* **81,** 190-200.

Cruz, J.C., Ramos, J.A., da Silva, L.P., Tenreiro, P.Q. & Heleno, R.H. (2013) Seed dispersal networks in an urban novel ecosystem. *European Journal of Forest Research,* **132,** 887-897.

Dalsgaard, B., Schleuning, M., Maruyama, P.K., Dehling, D.M., Sonne, J., Vizentin-Bugoni, J., . . . Rahbek, C. (2017) Opposed latitudinal patterns of network-derived and dietary specialization in avian plant–frugivore interaction systems. *Ecography*, **40**, 1395-1401.

Debussche, M. & Isenmann, P. (1989) Fleshy fruit characters and the choices of bird and mammal seed dispersers in a Mediterranean region. *Oikos,* **56,** 327-338.

Dormann, C.F., Fründ, J., Blüthgen, N. & Gruber, B. (2009) Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal,* **2,** 7-24.

Emer, C., Memmott, J., Vaughan, I.P., Montoya, D. & Tylianakis, J.M. (2016) Species roles in plant–pollinator communities are conserved across native and alien ranges. *Diversity and Distributions,* **22,** 841-852.

Farwig, N., Schabo, D.G. & Albrecht, J. (2017) Trait-associated loss of frugivores in fragmented forest does not affect seed removal rates. *Journal of Ecology,* **105,** 20-28.

Fontaine, C., Guimarães, P.R., Kéfi, S., Loeuille, N., Memmott, J., van der Putten, W.H., . . . Thébault, E. (2011) The ecological and evolutionary implications of merging different types of networks. *Ecology Letters,* **14,** 1170-1181.

Fortuna, M.A. & Bascompte, J. (2006) Habitat loss and the structure of plant–animal mutualistic networks. *Ecology Letters,* **9,** 281-286.

Fricke, E.C., Tewksbury, J.J., Wandrag, E.M. & Rogers, H.S. (2017) Mutualistic strategies minimize coextinction in plant–disperser networks. *Proceedings of the Royal Society B: Biological Sciences,* **284**, 20162302.

García, D. (2016) Birds in ecological networks: insights from bird-plant mutualistic interactions. *Ardeola,* **63,** 151-180.

García, D., Martínez, D., Stouffer, D.B. & Tylianakis, J.M. (2014) Exotic birds increase generalization and compensate for native bird decline in plant–frugivore assemblages. *Journal of Animal Ecology,* **83,** 1441-1450.

Genrich, C.M., Mello, M.A.R., Silveira, F.A.O., Bronstein, J.L. & Paglia, A.P. (2017) Duality of interaction outcomes in a plant–frugivore multilayer network. *Oikos,* **126,** 361-368.

González-Varo, J.P. (2010) Fragmentation, habitat composition and the dispersal/predation balance in interactions between the Mediterranean myrtle and avian frugivores. *Ecography,* **33,** 185-197.

González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in Ecology and Evolution,* **5,** 806-814.

González-Varo, J.P., Carvalho, C.S., Arroyo, J.M. & Jordano, P. (2017) Unravelling seed dispersal through fragmented landscapes: Frugivore species operate unevenly as mobile links. *Molecular Ecology,* **26,** 4309-4321.

Heleno, R., García, C., Jordano, P., Traveset, A., Gómez, J.M., Blüthgen, N., . . . Olesen, J.M. (2014) Ecological networks: delving into the architecture of biodiversity. *Biology Letters,* **10**, 20131000.

Herrera, C.M. (1984) Adaptation to frugivory of Mediterranean avian seed dispersers. *Ecology,* **65,** 609-617.

Herrera, C.M. (2002) Seed dispersal by vertebrates. *Plant-animal interactions. An evolutionary approach* (eds C.M. Herrera & O. Pellmyr), pp. 185-208. Blackwell Science, Oxford.

Jaenike, J. (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics,* **21,** 243-273.

Janzen, D.H. (1983) Seed and pollen dispersal by animals: convergence in the ecology of contamination and sloppy harvest. *Biological Journal of the Linnean Society,* **20,** 103-113.

Jordano, P. (1994) Spatial and temporal variation in the avian-frugivore assemblage of *Prunus mahaleb*: patterns and consequences. *Oikos,* **71,** 479-491.

Jordano, P. (2013) Fruits and frugivory. *Seeds: the ecology of regeneration of plant communities* (ed. R.S. Gallagher), pp. 18-61. CABI, Wallingford, UK.

Jordano, P. (2016) Sampling networks of ecological interactions. *Functional Ecology,* **30,** 1883-1893.

Jordano, P. & Schupp, E.W. (2000) Seed disperser effectiveness: the quantity component and patterns of seed rain for *Prunus mahaleb. Ecological Monographs,* **70,** 591-615.

King, C., Ballantyne, G. & Willmer, P.G. (2013) Why flower visitation is a poor proxy for pollination: measuring single-visit pollen deposition, with implications for pollination networks and conservation. *Methods in Ecology and Evolution,* **4,** 811-818.

Moermond, T.C. & Denslow, J.S. (1985) Neotropical avian frugivores: patterns of behavior, morphology, and nutrition, with consequences for fruit selection. *Ornithological Monographs,* 865-897.

Montesinos-Navarro, A., Hiraldo, F., Tella, J.L. & Blanco, G. (2017) Network structure embracing mutualism–antagonism continuums increases community robustness. *Nature Ecology & Evolution,* **1,** 1661-1669.

Moran, M.D. (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, **100**, 403-405.

Morris, R.J., Gripenberg, S., Lewis, O.T. & Roslin, T. (2014) Antagonistic interaction networks are structured independently of latitude and host guild. *Ecology Letters,* **17,** 340-349.

Newton, I. (1967) The adaptive radiation and feeding ecology of some British finches. *Ibis,* **109,** 33-96.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R., . . . Wagner, H. (2016) Package 'vegan'. Community ecology package, version 2.4-0.

Okuyama, T. & Holland, J.N. (2008) Network structural properties mediate the stability of mutualistic communities. *Ecology Letters,* **11,** 208-216.

Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2011) Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society of London, Series B: Biological Sciences,* **278,** 725-732.

Perea, R., Delibes, M., Polko, M., Suárez-Esteban, A. & Fedriani, J.M. (2013) Context-dependent fruit–frugivore interactions: partner identities and spatio-temporal variations. *Oikos,* **122,** 943-951.

Pigot, A.L., Bregman, T., Sheard, C., Daly, B., Etienne, R.S. & Tobias, J.A. (2016) Quantifying species contributions to ecosystem processes: a global assessment of functional trait and phylogenetic metrics across avian seed-dispersal networks. *Proceedings of the Royal Society B: Biological Sciences,* **283**.

Pimm, S. L., Lee Jones, H. & Diamond, J. On the risk of extinction. *The American Naturalist,* **132**, 757-785.

Pires, M.M. & Guimarães, P.R. (2013) Interaction intimacy organizes networks of antagonistic interactions in different ways. *Journal of The Royal Society Interface,* **10**.

Plein, M., Längsfeld, L., Neuschulz, E.L., Schultheiß, C., Ingmann, L., Töpfer, T., . . . Schleuning, M. (2013) Constant properties of plant–frugivore networks despite fluctuations in fruit and bird communities in space and time. *Ecology,* **94,** 1296-1306.

Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The robustness and restoration of a network of ecological networks. *Science,* **335,** 973-977.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roemer, G.W., Donlan, C.J. & Courchamp, F. (2002) Golden eagles, feral pigs, and insular carnivores: How exotic species turn native predators into prey. *Proceedings of the National Academy of Sciences, USA,* **99,** 791-796.

Rumeu, B., Devoto, M., Traveset, A., Olesen, J.M., Vargas, P., Nogales, M. & Heleno, R. (2017) Predicting the consequences of disperser extinction: richness matters the most when abundance is low. *Functional Ecology,* **31,** 1910-1920.

Schleuning, M., Fründ, J., Klein, A.-M., Abrahamczyk, S., Alarcón, R., Albrecht, M., . . . Blüthgen, N. (2012) Specialization of Mutualistic Interaction Networks Decreases toward Tropical Latitudes. *Current Biology, 22,* 1925-1931.

Schleuning, M., Fründ, J., Schweiger, O., Welk, E., Albrecht, J., Albrecht, M., . . . Hof, C. (2016) Ecological networks are more sensitive to plant than to animal extinction under climate change. *Nature Communications, 7,* 13965.

Schleuning, M., Ingmann, L., Strauß, R., Fritz, S.A., Dalsgaard, B., Matthias Dehling, D., . . . Dormann, C.F. (2014) Ecological, historical and evolutionary determinants of modularity in weighted seed-dispersal networks. *Ecology Letters, 17,* 454-463.

Schupp, E.W., Jordano, P. & Gomez, J.M. (2010) Seed dispersal effectiveness revisited: a conceptual review. *New Phytologist, 188,* 333-353.

Schupp, E.W., Jordano, P. & Gómez, J.M. (2017) A general framework for effectiveness concepts in mutualisms. *Ecology Letters, 20,* 577-590.

Simmons, B. I., Albrecht, J., Farwig, N., García, D., Jordano, P., Dicks, L. V., Sutherland W. J. and González-Varo, J. P. (2018). Data from: Moving from frugivory to seed dispersal: incorporating the functional outcomes of interactions in plant-frugivore networks. *Dryad Digital Repository*, doi:10.5061/dryad.r3d70m9.

Snow, B. & Snow, D. (1988) *Birds and berries*. T and A D Poyser, Calton, UK.

Spiegel, O. & Nathan, R. (2010) Incorporating density dependence into the directed-dispersal hypothesis. *Ecology, 91,* 1538-1548.

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

Thompson, J.N. (2009) The coevolving web of life. *The American Naturalist, 173,* 125-140.

Traveset, A., Heleno, R. & Nogales, M. (2014) The ecology of seed dispersal. *Seeds: The Ecology of Regeneration in Plant Communities* (ed. R.S. Gallaguer), pp. 62-93. CAB Int., Oxfordshire, UK.

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

Vázquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple determinants of the structure of plant-animal mutualistic networks. *Ecology, 90,* 2039-2046.

Wang, B.C. & Smith, T.B. (2002) Closing the seed dispersal loop. *Trends in Ecology & Evolution, 17,* 379-386.

Wenny, D.G. & Levey, D.J. (1998) Directed seed dispersal by bellbirds in a tropical cloud forest. *Proceedings of the National Academy of Sciences,* **95,** 6204-6207.

2 | Abundance drives broad patterns of generalisation in plant-hummingbird pollination networks

Abstract

Abundant pollinators are often more generalised than rare pollinators. This could be because abundant species have more chance encounters with potential interaction partners. On the other hand, generalised species could have a competitive advantage over specialists, leading to higher abundance. Determining the direction of the abundance-generalisation relationship is therefore a 'chicken-and-egg' dilemma. Here we determine the direction of the relationship between abundance and generalisation in plant-hummingbird pollination networks across the Americas. We find evidence that hummingbird pollinators are generalised because they are abundant, and little evidence that hummingbirds are abundant because they are generalised. Additionally, most patterns of species-level abundance and generalisation were well explained by a null model that assumed interaction neutrality (interaction probabilities defined by species relative abundances). These results suggest that neutral processes play a key role in driving broad patterns of generalisation in animal pollinators across large spatial scales.

Introduction

Pollination and other mutualistic associations are crucial for the functioning and maintenance of ecological communities (Heithaus 1974, Rech et al. 2016, Ollerton 2017, Ratto et al. 2018). A common phenomenon in mutualistic communities is that more abundant species have more generalised interaction niches (Dupont et al. 2003, Vázquez and Aizen 2003, Olesen et al. 2008). However, the direction of the relationship between abundance and generalisation has been described as a 'chicken-and-egg' dilemma because there are valid *a priori* explanations for both directions (Fort et al. 2016, Dormann et al. 2017). On the one hand, high abundance could lead to high generalisation. For example, abundant species are more likely to encounter a greater number of potential interaction partners than rare species (Vázquez et al. 2007, 2009, Poisot et al. 2015). Additionally, in a given area, higher species abundance leads to greater conspecific competition for available resources, resulting in increased generalization as predicted by optimal foraging theory (Fontaine et al. 2008, Tinoco et al. 2017). On the other hand, generalisation can have a selective advantage over specialisation, leading to higher abundance (Batstone et al. 2018). For example, the wider diet breadth of generalist individuals could allow them to receive a more stable benefit over time in communities with high levels of variability or species turnover; generalisation increases the likelihood that a given mutualist will interact with the most beneficial partner; and generalists benefit from having diverse partners that occupy different niches but provide the same rewards via different mechanisms (complementarity) (Waser et al. 1996, Albrecht et al. 2012, CaraDonna et al. 2017, Batstone et al. 2018). Generalisation can also provide a better nutrient balance (Tasei and Aupinel 2008, Behmer 2009, Vaudo et al. 2015), improve species' pathogen resistance (Alaux et al. 2010, Di Pasquale et al. 2013), entail a large resource base, and afford functional redundancy that buffers against partner extinction (Biesmeijer et al. 2006).

Here we evaluate the direction of the abundance-generalisation relationship in plant-hummingbird pollination networks and use a null model to assess the extent to which

observed patterns of species-level generalisation can be explained by neutral effects. Plant-hummingbird interactions are a particularly interesting model system to answer these questions as they involve species spanning the entire specialisation-generalisation spectrum (Bleiweiss 1998, Martín González et al. 2015, Dalsgaard et al. 2018, Maruyama et al. 2018). Additionally, pollination by vertebrates is important, especially in the tropics (Bawa 1990, Vizentin-Bugoni et al. 2018), and is on average responsible for 63% of fruit or seed production in vertebrate-pollinated plants (Ratto et al. 2018). Therefore, understanding the abundance-generalisation relationship in vertebrate pollinators such as hummingbirds has important implications for understanding the processes maintaining tropical plant and vertebrate communities.

Material and Methods

Dataset

We assembled a database of plant-hummingbird pollination networks with complementary information on hummingbird and plant abundance. In total, we gathered 19 quantitative networks, where link weights represent the number of observed hummingbird visits to plants. The database contained 103 hummingbird species and 403 plant species. For each of the 19 networks, hummingbird abundances were quantified as the mean number of individuals per species either recorded along transect counts within the sampling plots or caught using mist nets (Appendix 1). For four networks where not all species were recorded within the sampling plots during transect counts or mist netting, we used frequency of occurrence (the proportion of days of fieldwork in which a given species was recorded) as a proxy for relative abundances, as both measures are strongly correlated and frequency of occurrence is still independent from the network data (Vizentin-Bugoni et al. 2014). To test whether these four networks affected our results, we repeated all analyses excluding these data (Appendix 2). Plant abundances were quantified along transect counts or inside plots within the study areas and summarized as the number of flowers per species recorded over the sampling period. Species abundances and interactions were quantified several

times (typically, monthly) over at least a complete annual cycle in each community. Further details of each network are given in Appendix 1. The inclusion of independent abundance estimates is an important advance because all 35 pollination and seed dispersal networks analysed by Fort et al (2016) used estimates of animal abundance based on the interaction network data, and the authors had direct measures of plant abundance for only 29% of networks. Using species' interaction frequency as a proxy for animal abundance can lead to biased conclusions (Vizentin-Bugoni et al. 2014); by Fort et al's own admission, "These animal abundance data are arguably limited, as they are not independent from the interactions; but these are the best data available to evaluate our question." Conversely, ours is the first study where we have estimates of plant and animal abundance independent from the interaction observations for the majority of networks.

Measures of generalisation

We calculated the level of generalisation of all hummingbird species in all networks. We focus on hummingbird species, rather than plants, as plants may have non-hummingbird partners not included in our data that could result in misleading estimates of generalisation (Dalsgaard et al. 2008). To assess the sensitivity of our results to the choice of generalisation metric, we measured generalisation in three ways. First, species degree, which is simply the number of plant species a given hummingbird species interacts with. Second, normalised degree, which is equal to a species' degree divided by the total number of possible partners. Third, a generalisation index g, based on a widely used species-level measure of specialization (d') that quantifies the extent to which a species deviates from a random sampling of its available interaction partners (Blüthgen et al. 2006). We calculated d' using independent plant abundance data. To ensure that higher values of d' corresponded to higher levels of generalisation, we calculated the standardised generalisation index g, defined as $1-d'/d'_{max}$ where d'_{max} is the maximum possible value of d' (Fort et al. 2016). d' and d'_{max} were calculated using the 'dfun' function in the 'bipartite' R package (Dormann et al. 2009).

General approach

First, we tested whether there was a relationship between hummingbirds' abundance and their level of generalisation for each generalisation metric. The generalisation metric was the response variable, with log(abundance) and network identity as explanatory variables. A linear mixed effects model with a Gaussian distribution was used for the model with g as the response variable and network identity as a random effect. The model was fitted using the 'lme4' R package (Bates et al. 2015) and the significance of the fixed effect was calculated using Wald χ^2 tests available in the 'Anova' function of the 'car' R package (Fox and Weisberg 2002). We calculated both the marginal pseudo-$R^2_{(G)LMM(m)}$, which represents the variance explained by fixed effects, and the conditional pseudo-$R^2_{(G)LMM(c)}$, which represents the variance explained by both fixed and random effects (Nakagawa and Schielzeth 2013, Emer et al. 2016, Kaiser-Bunbury et al. 2017, Bartoń 2018). A zero-truncated negative binomial distribution was used for the model with degree as the response variable and a beta distribution was used for the model with normalised degree as the response variable. We used the zero-truncated negative binomial regression to account for overdispersion and zero-truncation in the degree data (no species had a degree of zero). A beta regression was used to model the normalised degree data because it accounts for overdispersion and is used for analysing continuous data greater than 0 and less than 1 (necessary for our analyses because no species had a normalised degree of zero). One data point in our dataset had a value of 1 and so we applied the standard correction following Smithson and Verkuilen (2006). These distributions are not available for mixed effects models, therefore the zero-truncated negative binomial model was fitted using the 'VGAM' R package (Yee and Wild 1996, Yee 2015) and the beta regression was fitted using the 'betareg' R package (Cribari-Neto and Zeileis 2010).

Having established that there is a relationship between abundance and generalisation, we used the approach of Fort *et al.* (2016) to determine whether abundance drives generalisation or generalisation drives abundance. This approach uses formal logic, specifically material implication, to derive expectations for broad species-level patterns of abundance and generalisation in ecological communities. To explain the approach, it

is useful to consider a simple example. Consider the proposition, *P*, "if it is a dodo, it is extinct". *P* is made up of two statements: (i) "it is a dodo" and (ii) "it is extinct". Given that each of these statements can either be true or false, we can derive four possible outcomes, as shown in Table 1. Outcome A is a dodo that is extinct. Outcome B is a non-dodo that is not extinct, such as the hummingbird species *Amazilia versicolor*. Outcome C is a non-dodo that is extinct, such as the dinosaur species *Tyrannosaurus rex*. Finally, outcome D is a dodo that is not extinct. We can only refute the proposition "if it is a dodo, it is extinct" when we observe outcome D to be true; that is, if we observe a living dodo. Conversely, observing an extinct dodo, an extant *Amazilia versicolor* individual, or an extinct *T. rex* specimen are all consistent with *P*.

There are four possible outcomes when applying this to the abundance-generalisation chicken-and-egg dilemma: abundant generalists, rare generalists, abundant specialists and rare specialists (Table 1). We can therefore derive two hypotheses:

1. If abundance implies generalisation, there should be no species which are abundant and specialist (outcome D: living dodos); we would only expect to observe abundant generalists (outcome A: extinct dodos), rare specialists (outcome B: a living *Amazilia versicolor*) and rare generalists (outcome C: extinct *T. rex*).

2. If generalisation implies abundance, there should be no generalist species that are rare; we would only expect to observe rare specialists, abundant specialists and abundant generalists.

Table 1: Truth table listing all possible outcomes for the propositions "if it is a dodo, it is extinct" and "if it is abundant, it is generalist". 'T' is 'True' and 'F' is 'False'.

Outcome	Dodo/Abundant	Extinct/Generalist
A	T	T
B	F	F
C	F	T
D	T	F

Therefore, by calculating the proportion of hummingbird species in each of the four abundance-generalisation categories (rare specialists, abundant specialists, rare generalists and abundant generalists; see below), it is possible to test these two hypotheses and determine whether the relationship between hummingbird abundance and generalisation is unidirectional (Fort et al. 2016). If hypothesis 1 is correct, the proportion of abundant specialists should be << the proportion of rare specialists, rare generalists, and abundant generalists; if hypothesis 2 is correct, the proportion of rare generalists should be << rare specialists, abundant specialists, and abundant generalists. We used contrasts within an ANOVA framework to test these hypotheses. To test hypothesis 1, we set abundant specialists as the reference contrast and tested whether it was significantly less than the other three categories. To test hypothesis 2, we set rare generalists as the reference contrast and tested whether it was significantly less than the other three categories.

Abundance and generalisation classification

To calculate the proportion of hummingbird species in each abundance-generalisation category, we developed a novel methodology to classify each species in a community as either rare or abundant and as either specialist or generalist. For each network, we first rescaled the abundance and generalisation values of all hummingbird species to range between 0 and 1 according to $(x - x_{min})/(x_{max} - x_{min})$, where x_{min} and x_{max} are the minimum and maximum values of abundance or generalisation (Aizen et al. 2012). We then conducted two Bernoulli trials for each species: (i) to classify a species as 'Abundant' or 'Rare' and (ii) to classify a species as 'Generalist or 'Specialist'. The probability of being classified as 'Abundant' in trial (i) was equal to the species' rescaled abundance; the probability of being classified as 'Generalist' in trial (ii) was equal to the species' rescaled generalisation. Therefore, a species with a rescaled abundance of 0.2 would have a 20% probability of being classified as abundant in a given iteration. Similarly, a species with a rescaled abundance of 0.8 would have an 80% probability of being classified as abundant. This was repeated 1000 times. The mean proportion of species in each of the four abundance-generalisation categories for each network was then calculated. This was repeated for each of the three generalisation metrics.

Our method offers a number of improvements over that used by Fort et al (2016), who used two methods to classify species. First, they classified species in a network as abundant or rare based on whether their abundance was greater than or less than the mean network abundance, respectively. Similarly, species were classified as generalised if their generalisation was greater than the mean network generalisation, and specialist otherwise. Delineating categories using a strict threshold such as this is problematic because it ignores the continuous nature of abundance and generalisation data: all values below the mean are treated as equivalent, as are all values above the mean. Consider a set of species with the following rescaled abundance values: 0.01, 0.02, 0.03, 0.04, 0.499, 0.501, 0.96, 0.97, 0.98, 0.99. Here the mean is 0.5. Therefore, using Fort et al's method, species with abundances of 0.01, 0.02, 0.03, 0.04 and 0.499 will always be classified as rare, while species with abundances of 0.501, 0.96, 0.97, 0.98 and 0.99 will always be classified as abundant. This is problematic because a species with 0.499 abundance is classified as rare, while one with 0.501 abundance is classified as abundant, despite there being a very small difference in the abundances of these two species. Conversely, species with very low or high abundances are treated as equal to those with medium abundances. For example, species with abundances between 0.01 and 0.04 are treated as equally rare to a species with an abundance of 0.499. Our method avoids these issues by using the full continuous range of the data to determine probabilities in the classification. For example, the species with an abundance of 0.499 and the species with an abundance of 0.501 both have similar probabilities of being classified as abundant. Similarly, the species with an abundance of 0.499 is 0.498 more likely to be classified as abundant than the species with an abundance of 0.01, thus more accurately accounting for abundance differences between these two species. Furthermore, given the highly-skewed nature of abundance and generalisation distributions, the mean threshold used by Fort et al could be misleading. Our method builds on this work to make no assumptions about the skewness of the data.

To remedy the problems with using the mean as a threshold, Fort et al also used a fuzzy logic classification, where species were classified as abundant or generalist if the value of abundance or generalisation was above the mean abundance or generalisation plus

one standard deviation. Species were classified as rare or specialist if the value of abundance or generalisation was below the mean abundance or generalisation minus one standard deviation. Species with measures between these values were given a linear class membership function, interpolated between 0 and 1. While this method overcomes some of the issues associated with categorisation based on a strict mean threshold, it still ignores continuous variation in abundance and generalisation values that are greater or less than one standard deviation from the mean. Conversely, our method considers the full range of the data, because the rescaled values simply determine probabilities of success in the Bernoulli trial. Additionally, the standard deviation could be a misleading measure given the highly-skewed distributions of abundance and generalisation. Our method makes no assumptions about skewness and works equally well for all distributions regardless of skewness. Finally, Fort et al's method assumes that a linear class membership function between the mean minus one standard deviation and the mean plus one standard deviation is appropriate, while our method requires no such assumptions.

Null model analysis

To assess the extent to which our results could be explained purely by neutral effects, we used a null model to generate 1000 randomised versions of each empirical network. The null model assumed interaction neutrality by assigning interactions according to a probability matrix, **A**, where element a_{ij} was the relative abundance of hummingbird species i multiplied by the relative abundance of plant species j (Vázquez et al. 2007, Maruyama et al. 2014, Vizentin-Bugoni et al. 2014, 2016). Therefore, the model assumes that two species with high abundance have a greater likelihood of interacting than two species with low abundance. The model constrained the number of links and ensured that each species had at least one interaction (Vázquez et al. 2007). We used independent plant and hummingbird abundance data to create the null networks, rather than relying on species marginal totals as a proxy for abundance. For each of the 1000 null versions of each of the 19 empirical networks, we repeated the permutational analysis described above ('Abundance and generalisation classification') to calculate the mean proportion of species in each of the four abundance-generalisation categories

predicted by the neutral model. We then compared these proportions based on neutrality to the empirical proportions: if the empirical proportions were within the 95% confidence intervals of the null model proportions then there were no significant differences between the null model and the observed values.

Results

We confirmed the positive relationship between abundance and generalisation in our dataset, finding a significant correlation between abundance and generalisation for degree ($P = < 0.001$; pseudo-$R^2 = 0.69$), normalised degree ($P = < 0.001$; pseudo-$R^2 = 0.63$) and the generalisation index g (Wald test: $\chi^2 = 10.7$; df = 1; $P = 0.001$; $R^2_{LMM(m)} = 0.06$; $R^2_{LMM(c)} = 0.44$).

Only a small proportion of species were abundant and specialist for all three generalisation metrics, while the proportion of species that were rare and generalist was consistently larger, particularly for the g generalisation metric (Figure 1). These differences were significant. We found that abundant specialists were significantly less common than rare specialists, rare generalists and abundant generalists for all generalisation metrics (Table 2). Conversely, for the degree and normalised degree metrics, we found that rare generalists were significantly less common than rare specialists, significantly more common than abundant specialists, and not significantly different to abundant generalists (Table 2). For the generalisation index (g), we found that rare generalists were not significantly different to rare specialists, and were significantly more common than abundant specialists and abundant generalists (Table 2). Overall, these findings support hypothesis 1, that abundance drives generalisation, and do not support hypothesis 2, that generalisation drives abundance.
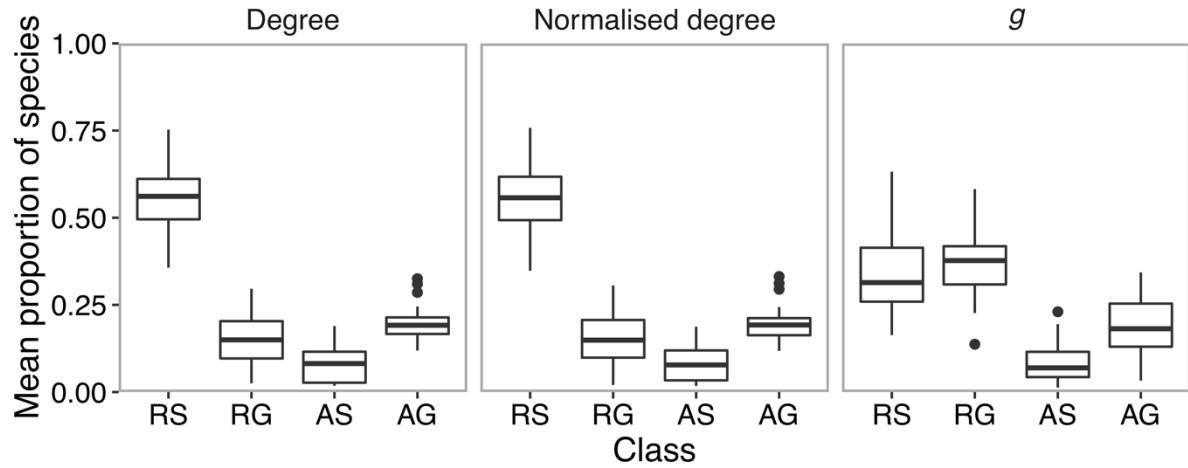
Figure 1: The mean proportion of hummingbird species classified as rare specialists ('RS'), rare generalists ('RG'), abundant specialists ('AS') and abundant generalists ('AG') across all networks, for three generalisation metrics: degree, normalised degree and g. The bold centre line in each box is the median; the lower and upper hinges are the first and third quartiles, respectively. The lower whisker indicates the smallest value no less than 1.5 times the inter-quartile range; the upper whisker indicates the largest value no greater than 1.5 times the inter-quartile range. Data outside the whiskers are outlying points plotted as solid black circles.

Table 2: Testing hypotheses 1 and 2 in an ANOVA framework, using abundant specialists and rare generalists as the reference contrast respectively. RS = rare specialist; RG = rare generalist; AS = abundant specialist; AG = abundant generalist. Significance codes: 0 '***' 0.001 '**', not significant 'ns'

Metric	Class	Estimate	t value	P	Significance
Hypothesis 1: Abundant specialist << rare specialist, rare generalist, abundant generalist					
Reference contrast = abundant specialist					
Degree	(Intercept)	0.08	4.88	0.00	***
	RS	0.48	19.70	0.00	***
	RG	0.07	2.87	0.01	**
	AG	0.11	4.70	0.00	***
Normalised degree	(Intercept)	0.08	4.77	0.00	***
	RS	0.48	19.00	0.00	***
	RG	0.07	2.81	0.01	**
	AG	0.11	4.57	0.00	***
g	(Intercept)	0.09	3.92	0.00	***
	RS	0.26	8.11	0.00	***
	RG	0.29	9.08	0.00	***
	AG	0.11	3.50	0.00	***
Hypothesis 2: Rare generalist << rare specialist, abundant generalist, abundant specialist					
Reference contrast = rare generalist					
Degree	(Intercept)	0.15	8.93	0.00	***
	RS	0.41	16.83	0.00	***
	AS	-0.07	-2.87	0.01	**
	AG	0.04	1.83	0.07	ns
Normalised degree	(Intercept)	0.16	8.75	0.00	***
	RS	0.41	16.19	0.00	***
	AS	-0.07	-2.81	0.01	**
	AG	0.04	1.76	0.08	ns
g	(Intercept)	0.37	16.77	0.00	***
	RS	-0.03	-0.97	0.33	ns
	AS	-0.29	-9.08	0.00	***
	AG	-0.18	-5.58	0.00	***

The proportion of species in each of the four abundance-generalisation categories predicted by the neutrality null model closely matched the empirical proportions, particularly for degree and normalised degree where there were no significant differences between observed and predicted proportions for the majority of networks (68–84% of networks; Figure 2). For *g*, the model correctly predicted the proportion of rare specialists and generalists for 79% of networks, but performed less well in predicting the proportion of abundant specialists and generalists, with predictions matching observed values for only 47% of networks (Figure 2).

All results were qualitatively the same and conclusions identical after the exclusion of the four networks where we used frequency of occurrence (the proportion of days of fieldwork in which a given species was recorded) as a proxy for relative abundances (Appendix 2).
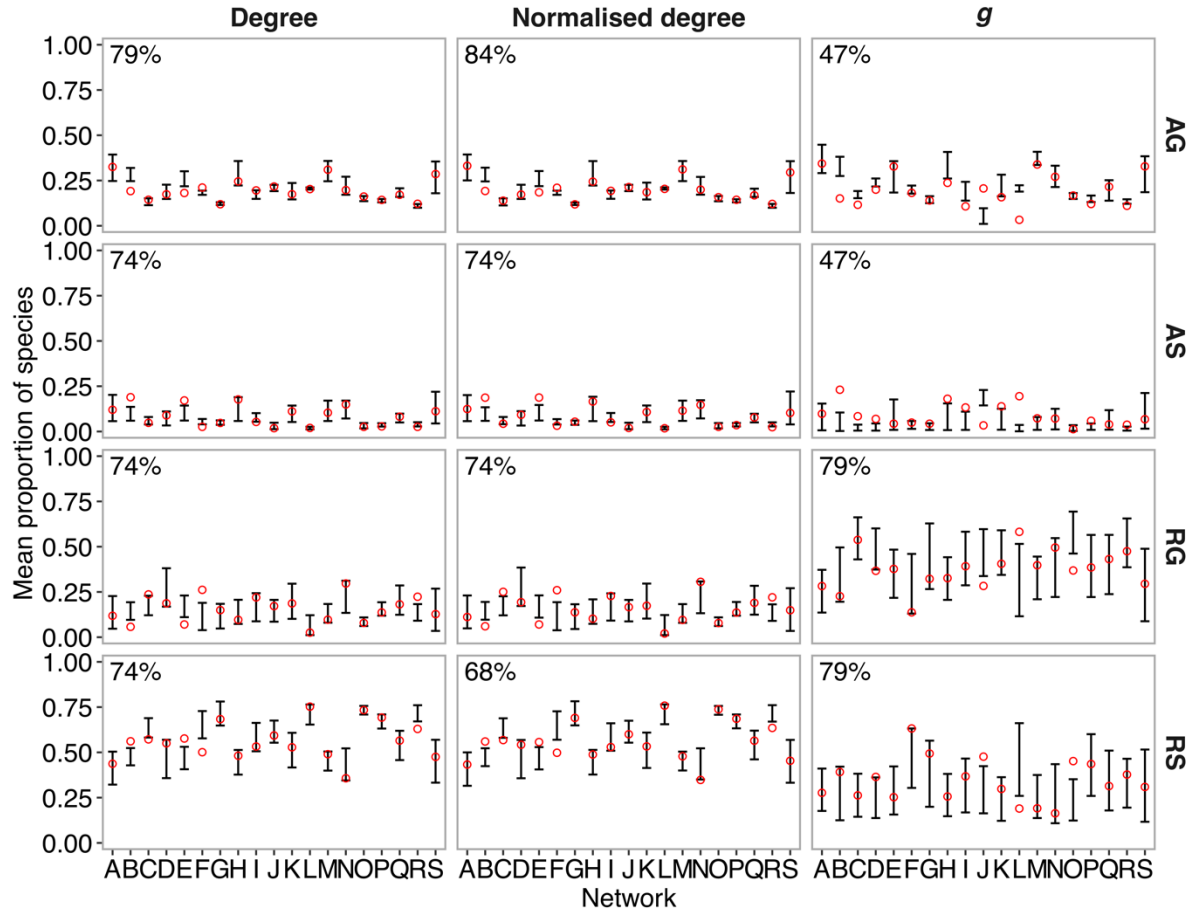
Figure 2: Comparisons between empirical networks (A-S) and null model networks in the proportions of species in each of the abundance-generalisation categories 'RS' (rare specialists), 'RG' (rare generalists), 'AS' (abundant specialists) and 'AG' (abundant generalists). Error bars represent the 95% confidence intervals of the mean proportion of hummingbird species in each abundance-generalisation category as predicted by 1000 null networks. Red circles show the empirically observed mean proportion of hummingbird species in each category. If the red circle is within the error bars, there were no significant differences between the observed proportions and the neutrality null model proportions. Percentages in the top left of each panel give the proportion of networks where empirical proportions were not significantly different from the null model proportions. Results are shown for each network (A-S) and for each generalisation metric (Degree, Normalised degree, *g*).

Discussion

The abundance-generalisation 'chicken and egg' dilemma concerns whether the widely observed positive relationship between abundance and generalisation is a consequence of abundance driving generalisation or generalisation driving abundance. Our analysis of plant-hummingbird communities sampled widely across the Americas provides

evidence of a unidirectional relationship, with hummingbird abundance driving hummingbird generalisation. Importantly, a null model assuming neutrality of interactions closely matched most empirical observations. This suggests that neutral effects have an important role in structuring broad patterns of species-level generalisation, even in a system such as plant-hummingbird pollination networks where phenotypical matching has a strong influence on the occurrence of pairwise interactions among species. Our results can be discussed in the context of *sufficient* and *necessary* conditions from formal logic. If we say that *P* is a *necessary* condition for *Q*, then in the absence of *P* there is also an absence of *Q*. However, if *P* is a *sufficient* condition for *Q*, then if we have *P*, *Q* must follow. For example, obtaining full marks on every question in an exam is a *sufficient*, but not *necessary*, condition for getting the top grade. Our results suggest abundance is a *sufficient* condition for generalisation as, if a species is abundant, it tends to also be a generalist. However, it is not a *necessary* condition as species can be generalist without being abundant. Conversely, our results suggest generalisation is a *necessary* condition for abundance as, if a species is a specialist, it tends to be rare. However, it is not a *sufficient* condition for abundance as, if a species is a generalist, this does not mean it is abundant. Therefore, our results agree with those of Fort *et al.* (2016) using pollination and seed dispersal networks, suggesting that abundance driving generalisation may be a general phenomenon that can be observed in mutualistic systems.

In all ecological studies it is worth asking whether sampling effort may impact the results. This is also the case for studies of species interaction networks, as sampling effects can influence the observed network structure (Fründ et al. 2016, Jordano 2016, Vizentin-Bugoni et al. 2016, Dalsgaard et al. 2017). Sampling is likely to result in missed detections of interactions for rare species, resulting in an underestimation of how generalised rare species are (Blüthgen 2010, Dorado et al. 2011). For this reason, Dormann et al. (2017) described sampling rare species with high generalisation as "impossible". This means that our results are unlikely to be a function of sampling effects, as the proportion of rare generalist species we observe is likely less than the true proportion: under theoretical perfect sampling, we would likely observe a larger proportion of species which are rare generalists, reinforcing our results (Dorado et al.

2011). Furthermore, sampling effects are likely to overestimate the proportion of species that are rare specialists as, even when rare species are observed, they are unlikely to be observed on all the plants they visit. This suggests that sampling effects will cause the generalisation level of rare species to be underestimated, and that consequently some species classified as rare specialists may actually be rare generalists (Blüthgen 2010, Dorado et al. 2011). Sampling effects are therefore not likely to impact our conclusions, because with perfect sampling we would expect the proportion of rare generalists to increase and the proportion of rare specialists to decrease, further increasing support for hypothesis 1 (many rare generalists, few abundant specialists) and refuting hypothesis 2 (few rare generalists, many abundant specialists). Additionally, we would not expect sampling artefacts to explain the low proportion of species which were abundant specialists because sampling effects tend to come from missing links for rare species rather than abundant species (Blüthgen 2010, Dorado et al. 2011, Fort et al. 2016). We also note that we do not consider the phylogenetic dependence of the hummingbird species within communities, which could cause an increase in Type I errors. While currently there are not ways to incorporate phylogenetic effects into our novel methodological framework, this is an important area for future research.

A frequent interpretation of the abundance-generalisation relationship is that abundant species are more generalised due to neutral effects; that is, they are more likely to encounter a greater number of interaction partners than less abundant species by chance alone (Vázquez et al. 2007). Our null model analysis supports this interpretation, particularly for degree and normalised degree: we found that the numbers of rare specialists, abundant specialists, rare generalists and abundant generalists were well predicted for the majority of networks by a null model that assumed interactions were formed entirely from neutral processes. This finding complements other recent studies of plant-hummingbird pollination networks showing the importance of morphological trait matching in predicting pairwise interactions at the network level (Maruyama et al. 2014, Vizentin-Bugoni et al. 2014, 2016, Weinstein and Graham 2017), while here we show that abundance predicts broad patterns of generalisation at the species level. Among Antillean hummingbirds, it was recently shown that local environmental conditions and floral richness, not hummingbirds'

morphological traits, determined species level nectar-feeding specialization (Dalsgaard et al. 2018). Combined with our findings, this might suggest a hierarchy of mechanisms structuring plant-hummingbird interactions, and more broadly whole pollination networks (Junker et al. 2013, Bartomeus et al. 2016, Vizentin-Bugoni et al. 2018): neutrality and local conditions govern broad patterns of generalisation, such as the number of plant partners, while morphological matching operates at a lower level to determine the identity of these plant partners. For the generalisation index g, the null model performed less well, predicting the proportion of abundant specialists and abundant generalists correctly in only 47% of networks. For the remaining 53% of networks, the model generally over predicted the number of abundant generalists and under predicted the number of abundant specialists. This may be due the nature of the g index itself: by accounting for the abundance of plants, g does not necessarily correlate with species degree (number of plant partners). For example, a hummingbird which visits one abundant plant could receive a higher value of g than a hummingbird that visits three rare plants. This means the null model may overestimate the number of abundant generalists and underestimate the number of abundant specialists as, in the model, an abundant hummingbird will have a higher probability of interacting with all plants, while in the empirical network it may be able to gain sufficient resources by only interacting with the most abundant plants.

Taken together, our study confirms that abundance is a sufficient, but not necessary, condition for generalisation in plant-hummingbird pollination networks; it is the first study to test this hypothesis in animals using independent data on species abundance encompassing a wide array of communities. Remarkably, our result corroborates the findings of Fort et al. (2016), giving further support that this may be a general phenomenon in mutualistic systems. Further research should investigate whether the relationships found here hold for other types of ecological systems, especially given evidence of the importance of neutral effects in structuring antagonistic host-parasite communities (Vázquez et al. 2005). We also find evidence that neutral effects are good predictors of coarse species-level patterns of generalisation, even in a system in which interactions are widely recognized to be constrained by species traits. This might suggest a hierarchy of mechanisms structuring plant-hummingbird interactions, with

neutral effects operating at a 'high level' to determine coarse patterns of generalisation, such as the number of partners, while niche-based processes act at a lower level to determine the identity of these partners.

References

Aizen, M. A. et al. 2012. Specialization and rarity predict nonrandom loss of interactions from mutualist networks. - Science (80-.). 335: 1486–1489.

Alaux, C. et al. 2010. Diet effects on honeybee immunocompetence. - Biol. Lett. 6: 562–565.

Albrecht, M. et al. 2012. Diverse pollinator communities enhance plant reproductive success. - Proc. R. Soc. B Biol. Sci. 279: 4845–4852.

Bartomeus, I. et al. 2016. A common framework for identifying linkage rules across different types of interactions. - Funct. Ecol. 30: 1894–1903.

Bartoń, K. 2018. MuMIn: Multi-Model Inference.: version 1.40.4.

Bates, D. et al. 2015. Fitting Linear Mixed-Effects Models using lme4. - J. Stat. Softw. 67: 1–48.

Batstone, R. T. et al. 2018. Using niche breadth theory to explain generalization in mutualisms. - Ecology 99: 1039–1050.

Bawa, K. S. 1990. Plant-Pollinator Interactions in Tropical Rain Forests. - Annu. Rev. Ecol. Syst. 21: 399–422.

Behmer, S. T. 2009. Insect Herbivore Nutrient Regulation. - Annu. Rev. Entomol. 54: 165–187.

Biesmeijer, J. C. et al. 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. - Science (80-.). 313: 351–354.

Bleiweiss, R. 1998. Origin of hummingbird faunas. - Biol. J. Linn. Soc. 65: 77–97.

Blüthgen, N. 2010. Why network analysis is often disconnected from community ecology: A critique and an ecologist's guide. - Basic Appl. Ecol. 11: 185–195.

Blüthgen, N. et al. 2006. Measuring specialization in species interaction networks. - BMC Ecol. 6: 9.

CaraDonna, P. J. et al. 2017. Interaction rewiring and the rapid turnover of plant–

pollinator networks. - Ecol. Lett. 20: 385–394.

Cribari-Neto, F. and Zeileis, A. 2010. Beta Regression in R. - J. Stat. Softw. 34: 1–24.

Dalsgaard, B. et al. 2008. Pollination networks and functional specialization: A test using Lesser Antillean plant-hummingbird assemblages. - Oikos 117: 789–793.

Dalsgaard, B. et al. 2017. Opposed latitudinal patterns of network-derived and dietary specialization in avian plant–frugivore interaction systems. - Ecography (Cop.). 40: 1395–1401.

Dalsgaard, B. et al. 2018. Trait evolution, resource specialization and vulnerability to plant extinctions among antillean hummingbirds. - Proc. R. Soc. B Biol. Sci. 285: 20172754.

Di Pasquale, G. et al. 2013. Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? - PLoS One 8: e72016.

Dorado, J. et al. 2011. Rareness and specialization in plant-pollinator networks. - Ecology 92: 19–25.

Dormann, C. F. et al. 2009. Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. - Open Ecol. J. 2: 7–24.

Dormann, C. F. et al. 2017. Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. - Annu. Rev. Ecol. Evol. Syst. 48: 559–584.

Dupont, Y. L. et al. 2003. Structure of a plant--flower-visitor network in the high-altitude sub-alpine desert of Tenerife, Canary Islands. - Ecography (Cop.). 26: 301–310.

Emer, C. et al. 2016. Species roles in plant–pollinator communities are conserved across native and alien ranges. - Divers. Distrib. 22: 841–852.

Fontaine, C. et al. 2008. Generalist foraging of pollinators: Diet expansion at high density. - J. Ecol. 96: 1002–1010.

Fort, H. et al. 2016. Abundance and generalisation in mutualistic networks: Solving the chicken-and-egg dilemma. - Ecol. Lett. 19: 4–11.

Fox, J. and Weisberg, S. 2002. An R Companion to Applied Regression. - Sage.

Fründ, J. et al. 2016. Sampling bias is a challenge for quantifying specialization and network structure: Lessons from a quantitative niche model. - Oikos 125: 502–513.

Heithaus, E. R. 1974. The Role of Plant-Pollinator Interactions in Determining Community Structure. - Ann. Missouri Bot. Gard. 61: 675–691.

Jordano, P. 2016. Sampling networks of ecological interactions. - Funct. Ecol. 30: 1883–1893.

Junker, R. R. et al. 2013. Specialization on traits as basis for the niche-breadth of flower visitors and as structuring mechanism of ecological networks. - Funct. Ecol. 27: 329–341.

Kaiser-Bunbury, C. N. et al. 2017. Ecosystem restoration strengthens pollination network resilience and function. - Nature 542: 223–227.

Martín González, A. M. et al. 2015. The macroecology of phylogenetically structured hummingbird-plant networks. - Glob. Ecol. Biogeogr. 24: 1212–1224.

Maruyama, P. K. et al. 2014. Morphological and spatio-temporal mismatches shape a neotropical savanna plant-hummingbird network. - Biotropica 46: 740–747.

Maruyama, P. K. et al. 2018. Functional diversity mediates macroecological variation in plant–hummingbird interaction networks. - Glob. Ecol. Biogeogr. 27: 1186–1199.

Nakagawa, S. and Schielzeth, H. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. - Methods Ecol. Evol. 4: 133–142.

Olesen, J. M. et al. 2008. Temporal dynamics in a pollination network. - Ecology 89: 1573–1582.

Ollerton, J. 2017. Pollinator Diversity: Distribution, Ecological Function, and Conservation. - Annu. Rev. Ecol. Evol. Syst. 48: 353–376.

Poisot, T. et al. 2015. Beyond species: Why ecological interaction networks vary through space and time. - Oikos 124: 243–251.

Ratto, F. et al. 2018. Global importance of vertebrate pollinators for plant reproductive success: a meta-analysis. - Front. Ecol. Environ. 16: 82–90.

Rech, A. R. et al. 2016. The macroecology of animal versus wind pollination: ecological factors are more important than historical climate stability. - Plant Ecol. Divers. 9: 253–262.

Smithson, M. and Verkuilen, J. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. - Psychol. Methods 11: 54–71.

Tasei, J.-N. and Aupinel, P. 2008. Nutritive value of 15 single pollens and pollen mixes tested on larvae produced by bumblebee workers (*Bombus terrestris*, Hymenoptera: Apidae). - Apidologie 39: 397–409.

Tinoco, B. A. et al. 2017. Effects of hummingbird morphology on specialization in pollination networks vary with resource availability. - Oikos 126: 52–60.

Vaudo, A. D. et al. 2015. Bee nutrition and floral resource restoration. - Curr. Opin. Insect Sci. 10: 133–141.

Vázquez, D. P. and Aizen, M. A. 2003. Null model analyses of specialization in plant-pollinator interactions. - Ecology 84: 2493–2501.

Vázquez, D. P. et al. 2005. Species abundance and the distribution of specialization in host-parasite interaction networks. - J. Anim. Ecol. 74: 946–955.

Vázquez, D. P. et al. 2007. Species abundance and asymmetric interaction strength in ecological networks. - Oikos 116: 1120–1127.

Vázquez, D. P. et al. 2009. Uniting pattern and process in plant-animal mutualistic networks: A review. - Ann. Bot. 103: 1445–1457.

Vizentin-Bugoni, J. et al. 2014. Processes entangling interactions in communities: forbidden links are more important than abundance in a hummingbird-plant network. - Proc. R. Soc. B Biol. Sci. 281: 20132397.

Vizentin-Bugoni, J. et al. 2016. Influences of sampling effort on detected patterns and structuring processes of a Neotropical plant-hummingbird network. - J. Anim. Ecol. 85: 262–272.

Vizentin-Bugoni, J. et al. 2018. Plant-Pollinator Networks in the Tropics: A Review. - In: Dáttilo, W. and Rico-Gray, V. (eds), Ecological Networks in the Tropics: An Integrative Overview of Species Interactions from Some of the Most Species-Rich Habitats on Earth. Springer International Publishing, pp. 73–91.

Waser, N. M. et al. 1996. Generalization in pollination systems, and why it matters. - Ecology 77: 1043–1060.

Weinstein, B. G. and Graham, C. H. 2017. Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions. - Ecol. Lett. 20: 326–335.

Yee, T. W. 2015. Vector Generalized Linear and Additive Models: With an Implementation in R. - Springer.

Yee, T. W. and Wild, C. J. 1996. Vector Generalized Additive Models. - J. R. Stat. Soc. Ser. B 58: 481–493.

3 | Vulnerable species interactions are important for the stability of mutualistic networks

Abstract

Species are central to ecology and conservation. However, it is the interactions between species that generate the functions on which ecosystems and humans depend. Despite the importance of interactions, we lack an understanding of the risk that their loss poses to ecological communities. Here, we quantify risk as a function of the vulnerability (likelihood of loss) and importance (contribution to network stability in terms of species coexistence) of 4330 mutualistic interactions from 41 empirical pollination and seed dispersal networks across six continents. Remarkably, we find that more vulnerable interactions are also more important: the interactions that contribute most to network stability are those that are most likely to be lost. Furthermore, most interactions tend to have more similar vulnerability and importance across networks than expected by chance, suggesting that vulnerability and importance may be intrinsic properties of interactions, rather than only a function of ecological context. These results provide a starting point for prioritising interactions for conservation in species interaction networks and, in areas lacking network data, could allow interaction properties to be inferred from taxonomy alone.

Introduction

Species are the predominant biological unit of interest across ecology and conservation. However, it is interactions between species, rather than species themselves, that

mediate the ecological functions that drive community dynamics and support biodiversity (Poisot, Stouffer, & Gravel, 2015). For example, pollination interactions shape co-evolution in diverse plant-animal communities (Guimarães, Pires, Jordano, Bascompte, & Thompson, 2017), while seed dispersal maintains spatial patterns of diversity (Wandrag, Dunham, Duncan, & Rogers, 2017). Given the importance of interactions for ecosystem functioning, their loss could have reverberating effects on entire communities and, ultimately, the ecosystem services they deliver (Díaz et al., 2013; Valiente-Banuet et al., 2015).

Interactions are thus a vital component of biodiversity, but they remain largely neglected (Janzen, 1977). Studies tend to focus on the impact of anthropogenic stressors on single interactions at single sites (Tylianakis, Didham, Bascompte, & Wardle, 2008), while the few studies that have considered interaction loss at the community level are either at local scales (Aizen, Sabatino, & Tylianakis, 2012), based on hypothetical network structures (Harvey, Gounand, Ward, & Altermatt, 2017) or only consider aggregate properties of interactions, rather than considering them individually (Santamaría, Galeano, Pastor, & Méndez, 2016). There is thus an urgent need to incorporate interactions into studies assessing community responses to environmental change. Specifically, we lack a quantitative understanding of the risk that interaction loss poses to communities, which, in turn, limits our ability to make conservation decisions.

Here, we address this gap by quantifying the risk of interaction loss to 41 pollination and seed dispersal communities that, combined, comprise a global dataset of 4330 species-species links (see Materials and Methods). Such mutualisms are fundamental to the functioning of most communities. The loss of pollination can lead to pollen limitation, potentially compromising reproduction for the vast majority of plant species that rely, to some extent, on animal pollinators (Ollerton, Winfree, & Tarrant, 2011; Ratto et al., 2018). Similarly, the disruption of seed dispersal can have deleterious, cascading consequences for those woody plant species that depend on frugivores, which can exceed 90% in biodiverse ecosystems such as tropical rainforests (Jordano, 2016).

Conventionally, risk is a function of both the likelihood of an event occurring and the severity of the impacts if it did occur (Rausand, 2013). For ecological networks, we therefore start by reasoning that the risk of losing a particular link is a function of (i) the likelihood of that link being lost (link vulnerability, V), and (ii) the severity of the consequences to the community if the link is lost (link importance, I). Using novel quantitative methods, we calculate the vulnerability and importance of all links in our dataset (Figure 1) and show that the most vulnerable links in a community are also those that contribute most to its structural stability (Materials and Methods). We next examine whether vulnerability and importance are intrinsic attributes of interactions, rather than functions of ecological context, by testing whether an interaction's vulnerability and importance is more similar across occurrences than expected by chance. Our aim is to explore the risk of interaction loss in mutualistic communities and to inform their conservation. Hereafter we distinguish between the terms *interaction* and *link*: *interaction* refers to all occurrences of a given taxon-taxon interaction identity, while *link* refers to a single occurrence of an *interaction* in a particular network. Thus, for example, the *Bombus pratorum – Leucanthemum vulgare interaction* is present in our data, while *Bombus pratorum – Leucanthemum vulgare links* occur in two networks.
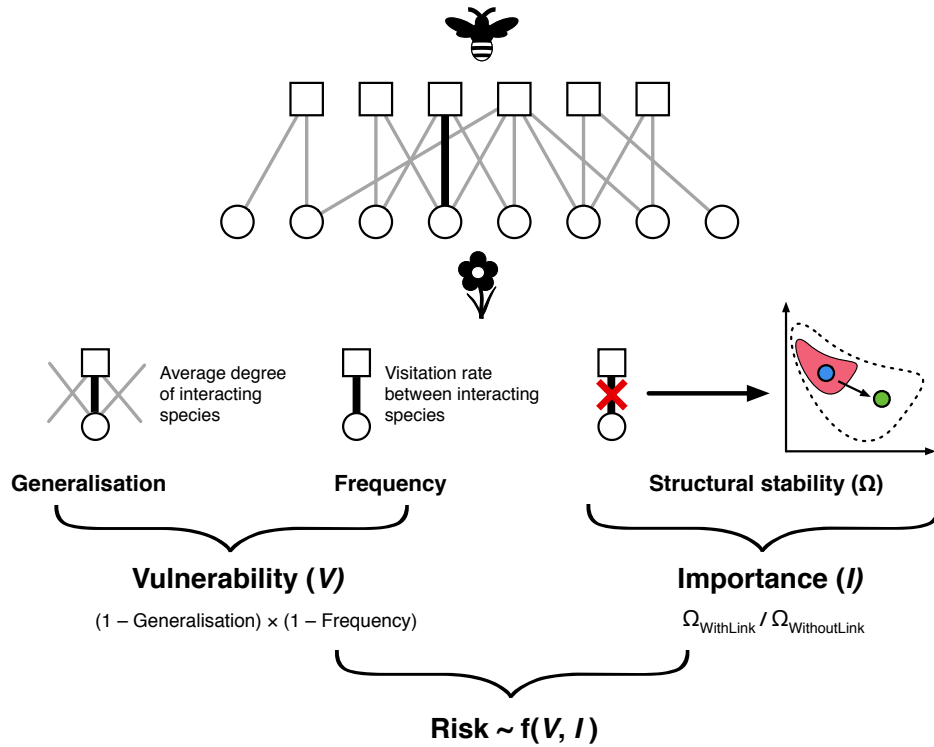
Figure 1: The quantities used in the analysis. Top: an illustrative network depicting interactions between plants and pollinators. A focal link is highlighted in black. The generalisation of a link is the average degree of the interacting species. The frequency of the link is the visitation rate between the interacting species. Combined, these measures determine the vulnerability of a link, such that vulnerable links are low frequency interactions between specialists. The structural stability of the network was measured with the focal link and without the focal link. The ratio of these two values is the importance of the link. This is represented by the graph on the right of the figure. The structural stability of a network is defined as the parameter space of intrinsic growth rates in which all species in a community can have positive abundances. This is represented for the whole community by the dotted outline and for the whole community without the focal link by the pink shape. In this case, removing the link has reduced the structural stability of the community (reduced the size of the shape). This means that a perturbation that moves the community from the initial state (blue circle) to a final state (green circle) will result in extinctions without the focal link because the community moves outside the pink feasibility domain. However, no extinctions would occur under the same perturbation in the original community before the focal link was removed, because the final state of the community (green circle) is within the original feasibility domain (dotted outline). Together, vulnerability and importance describe the risk to a community of losing a particular link.

Results

Relationship between link vulnerability and importance

We calculated the vulnerability and importance of 4330 links from a global dataset of 29 plant-pollinator and 12 plant-seed disperser networks (Figure 1). Vulnerability was measured as a function of link frequency (how often the two species involved in the link interact) and link generalisation (the mean number of links [the mean degree] of the two species involved in the link) (Aizen et al., 2012). This means that weak (less frequent) links between specialists were more vulnerable than strong (more frequent) links between generalists (Aizen et al., 2012). Importance was defined as the contribution of a given link to the feasibility of a network, where feasibility is a measure of a network's ability to withstand environmental variation without leading to species extinctions (Rohr, Saavedra, & Bascompte, 2014; Song, Rohr, & Saavedra, 2018). Important links were those that, when removed, lowered a network's feasibility; that is, when removed, they reduced the amount of environmental variability a network can tolerate before species extinctions took place (Figure 1; see Materials and Methods). We found that there was a significantly positive correlation between the vulnerability and importance of links across the 41 networks (Wald test: χ^2 = 53.71, df = 1, P = < 0.001) (Figure 2). This correlation indicates that the links that contribute most to the structural stability of communities are those links that are most likely to be lost in the face of environmental changes.
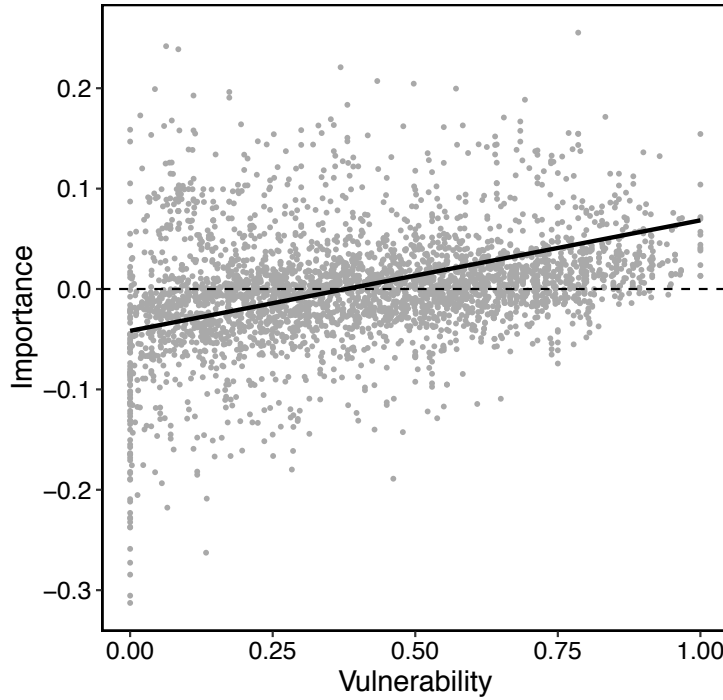
Figure 2: The relationship between vulnerability (the likelihood of a link being lost) and importance (the contribution of a link to a network's structural stability) for all 3391 species-species links across 41 mutualistic networks. Best fit line is from a mixed effects model with importance as the response variable, vulnerability as a fixed effect, and network identity as a random effect.

Taxonomic consistency of link vulnerability and importance

A consistent, positive relationship between vulnerability and importance could arise if links tend to have the same vulnerability and importance independent from the network in which they occur. This would imply some form of evolutionary conservatism in interaction properties. We tested this hypothesis by assessing the extent to which vulnerability and importance exhibited taxonomic consistency: the tendency for an interaction's vulnerability and importance to be more similar across all the networks in which the interaction occurs than expected by chance. If vulnerability and importance exhibit taxonomic consistency, then all occurrences of a given interaction should have similar levels of these properties. For each interaction, we compared the variance in vulnerability and importance to a null expectation where links were sampled randomly (see Materials and Methods). We carried out analyses at genus, family and order levels, but not at the species level, because very few interactions at the species level occurred more than once in the data. We found a strong tendency towards consistency for both vulnerability and importance at all taxonomic levels (between 76% and 83% of

interactions had more similar values of vulnerability and importance than expected by chance; Figure 3). Considering vulnerability, there was significant taxonomic consistency for 18% of genus, 17% of family and 30% of order interactions (see Materials and Methods). For importance, interactions had significant taxonomic consistency for 14% of genus, 20% of family and 33% of order level interactions. Conservatism was observed across large geographic scales, with many significantly-consistent interactions comprised of links occurring in different regions or continents. These results suggest that vulnerability and importance may be, to some extent, intrinsic properties of interactions and not only a function of ecological context.
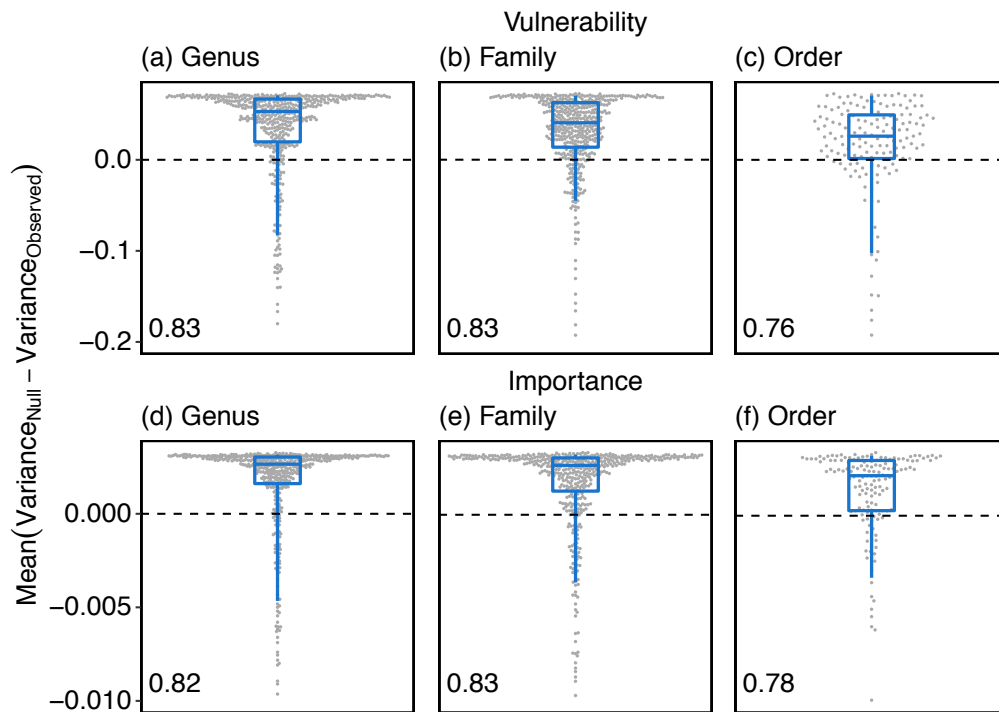

Figure 3: The degree of taxonomic consistency for each interaction at genus ($n = 469$), family ($n = 466$) and order ($n = 151$) levels, for both vulnerability (likelihood of a link being lost) and importance (contribution of a link to a network's structural stability). Taxonomic consistency is the tendency for properties of an interaction to be more similar across occurrences than expected by chance. Points represent individual interactions. Boxplots represent 5%, 25%, 50%, 75% and 95% quantiles of the same data, moving from the bottom whisker to the top whisker. Number in bottom left of each panel is the proportion of interactions which exhibited positive consistency (Variance$_{Observed}$ < Variance$_{Null}$). For visualisation, a small number of points with low values were removed. The percentage of points with values lower than the y-axis minimum are as follows for each panel: (a) 1.5%, (b) 1.1%, (d) 3.2%, (e) 1.5%, (f) 1.3%.

Mapping the risk of link loss

The two components of risk – vulnerability and importance – reflect, respectively, how likely a link is to be lost and how serious the consequences of that loss are for the community. Using an illustrative plant-seed disperser network, we coloured links based on their vulnerability and importance, and highlighted those that exhibited taxonomic consistency (Figure 4). As expected, given the positive correlation between vulnerability and importance, we find that a substantial proportion of links are either highly vulnerable and contribute strongly to structural stability (22.5%; dark red in Figure 4) or have low vulnerability and contribute negatively to stability (19.4%; light purple in Figure 4). Conversely, few links are of low vulnerability and positive importance (5.4%; light yellow in Figure 4) or high vulnerability and negative importance (2.3%; dark purple in Figure 4). Identifying those links that are vulnerable and which benefit community stability as a whole provides a potential starting point when deciding which links should be priorities for conservation (Figure 4). If two links have similar vulnerability and importance, but one exhibits stronger taxonomic consistency than the other, then the more consistent link may be of higher concern as it could be important and susceptible to extinction across communities in different geographical regions.
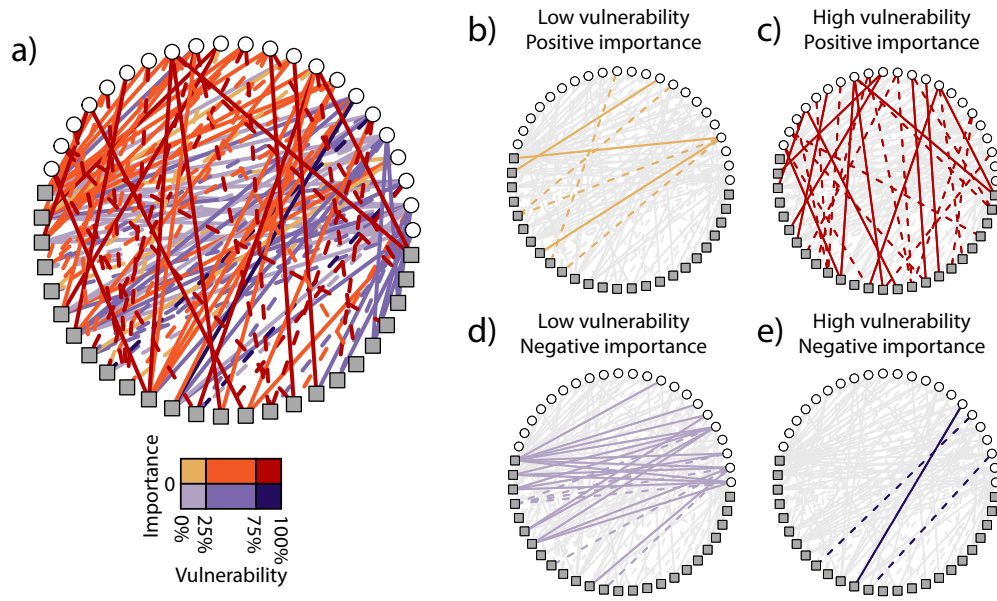
Figure 4: Mapping vulnerability, importance and conservatism onto an example plant-seed disperser network (squares are plant species; circles are seed disperser species). Solid lines indicate significant taxonomic consistency at the genus, family or order level. Links are categorised based on (i) whether they positively or negatively contribute to the feasibility of the community; that is whether their importance (contribution to network feasibility) is positive or negative, and (ii) whether their vulnerability is in the bottom 25% of vulnerability values, the middle 50% or the top 25%. (a) Shows all links together; (b) highlights low vulnerability, positive importance links; c) highlights high vulnerability, positive importance links; (d) highlights low vulnerability, negative importance links; (e) highlights high vulnerability, negative importance links. We suggest that consistently-important and vulnerable links (solid, dark red lines) should be high priorities in conservation.

Discussion

Our analysis represents the first attempt to quantify the risk that link loss poses to ecological communities. We find that, across 41 ecological networks, the links that contribute most to a network's ability to tolerate environmental perturbations are the same links that are most likely to be lost in the face of such perturbations (Figure 2). We additionally find that there is a strong tendency for interactions to have more similar vulnerability and importance across occurrences than expected by chance, with a substantial proportion of interactions exhibiting this signal significantly (Figure 3). By combining these results, we are able to map the risk of interaction loss onto empirical communities, which could be used to guide conservation efforts (Figure 4).

The positive relationship between vulnerability and importance means that the more vulnerable a link is, the more likely it is to have a negative impact on network feasibility if it is lost. Vulnerability is therefore an important indicator of the extent to which a link supports or hinders a community's ability to tolerate environment variation and, thus, species' long-term persistence. From a conservation perspective, this result is concerning as it suggests that losing vulnerable interactions reduces the ability of mutualistic networks to absorb future stressors. However, it also suggests that our proposed link vulnerability measure enables estimates of how much a link benefits a community (which may be difficult to measure otherwise) using only simple topological information. Aizen et al (Aizen et al., 2012) found that vulnerable links were found in places with greater habitat loss. Combined with our results, this suggests environmental stressors like habitat loss may be detrimental for whole-community stability, not just those links which are vulnerable.

That those links that are stronger contributors to structural stability are more vulnerable to extinction mirrors findings that *nodes* that contribute most to a network's nestedness (and thus persistence) are those that are most likely to go extinct (Serguei Saavedra, Stouffer, Uzzi, & Bascompte, 2011). The positive relationship between link vulnerability and importance suggests that links have a tendency to fall into one of two categories: low vulnerability and low importance or high vulnerability and high importance. Thus, some links have a high probability of survival to the detriment of the network as a whole, while others contribute to the collective good at the expense of their own viability (Leigh, 1977; Serguei Saavedra et al., 2011). While the causes of these patterns are unclear, perhaps there is a tendency for some species to maximise their fitness by being involved in a mixture of 'selfless' links that ensure the community as a whole remains intact, and 'greedy' links that provide stable benefit to the species over time. Determining how such 'selfless links' arise is an important area for future research as characterising the conditions compatible with their genesis could aid the design or maintenance of resilient ecosystems, and cooperative systems more broadly.

We found that many interactions tend to have similar vulnerability and importance values across occurrences, implying a form of evolutionary conservatism in properties

of species interactions. The consistency of vulnerability could be driven by conserved patterns of generalisation or abundance of the interacting partners. For example, pollinator species have been shown to have similar levels of generalisation across their range (Emer, Memmott, Vaughan, Montoya, & Tylianakis, 2016). Similarly, (Rodríguez-Flores, Ornelas, Wethington, & Arizmendi, 2019) found that particular pollinator clades tended to be generalist, while (Martín González et al., 2015) found significant phylogenetic signal in pollinator interactions. Meanwhile, the taxonomic consistency of importance has substantial implications. While here we consider the importance of individual links, such importance values are governed by the whole network structure, not just the roles of the two partner species involved in the link. To illustrate this, consider two pollinators, i and j, and two plants, m and n that all interact with each other; that is i and j both interact with m and n. If j leaves the network, the importance of all remaining links will change (Baker, Kaartinen, Roslin, & Stouffer, 2015). Thus, that link importance is consistent across occurrences suggests that links, and their partner species, are embedded in networks in similar ways, as has been found for species in antagonistic networks (Baker et al., 2015; Stouffer, Sales-Pardo, Sirer, & Bascompte, 2012).

Our results have significant conservation implications. Differences in links' vulnerability, importance and taxonomic consistency could be used to guide proactive conservation efforts: links with high vulnerability and importance could provide a useful starting point to inform prioritisation before any links are lost (Figure 4). Similarly, highly-important links that are not currently vulnerable could be a focus of monitoring efforts in case they become vulnerable in the future. Importantly, by explicitly focusing on links themselves, our methods may identify high-priority links that are not expected to be so based only on assessments of species extinction risk. Conversely, our results may be able to inform species conservation if high-priority links tend to involve species that are also of high priority. Determining the relationship between the conservation priority of species and the links they form is an important area for future research. Our finding of widespread taxonomic consistency potentially allows properties of interactions to be inferred in regions without network data, even if such properties are only known for congeneric, confamiliar or conorder interactions.

This is important because species interaction networks are often cost- and time-intensive to collect, and coverage is highly biased geographically (Cameron et al., 2019; Kaiser-Bunbury & Blüthgen, 2015).

Conserving links is perhaps even more challenging than conserving species. While species conservation requires one species to remain extant, link conservation requires that two species remain at sufficiently high abundance to still significantly interact: two species must be prevented from going *ecologically* extinct (Galetti et al., 2013; Redford, 1992; Säterberg, Sellman, & Ebenman, 2013). Moreover, because interaction extinction often precedes species loss, conservation actions must occur sooner. While link conservation has attracted little attention so far, the importance of interactions like pollination is now widely recognised. Thus, we hope our results can help guide future research in this nascent and important field, because ultimately it is links that support the ecological functions and services that communities provide.

Materials and Methods

Data

We assembled a dataset of 4330 plant-animal links from 41 quantitative mutualistic networks spanning a broad geographical range, with data in tropical and non-tropical areas from both islands and mainlands (www.web-of-life.es, [29, 30]). The database spanned two types of mutualism, comprising 3182 pollination links from 29 pollination networks and 1148 seed dispersal links from 12 seed dispersal networks. The data contained 551 plant species and 1151 animal species.

Interaction properties

Link Vulnerability

We developed a measure of network link vulnerability following Aizen *et al.* (Aizen et al., 2012). They identified two factors that determine the vulnerability of a mutualistic link between a plant (i) and animal (j) species: link frequency (hereafter 'frequency') and link degree (hereafter 'generalisation') (Aizen et al., 2012). Frequency is how often a link occurs between i and j (such as the number of times a pollinator species visits a

particular flower species), while generalisation is defined as the mean degree of the two species involved in a link, that is, the average number of species with which species i and j interact (Aizen et al., 2012). This notion of vulnerability aims to capture the sensitivity of a network to the loss of a given link (Figure 1).

We calculated the frequency and generalisation of all links in our dataset. Following Aizen *et al.* (Aizen et al., 2012), we first $\log_{10}$ transformed all frequencies. Second, to make results between networks comparable, we standardised frequency and generalisation to between 0 and 1 at the network level. Finally, we calculated the vulnerability of a link between species i and j, as $V_{ij} = (1 - f_{ij})(1 - D_{ij})$, where f_{ij} is the standardised link frequency and D_{ij} is the standardised link generalisation. In this formulation, the index can take values between 0 (least vulnerable) and 1 (most vulnerable), which means that it categorizes weak links between specialist partners as more vulnerable than strong links between generalists.

Link Importance

Feasibility is defined as the range of conditions under which all species in a community can stably coexist (Song et al., 2018). Feasibility can therefore be thought of as the 'safe operating space' of ecological communities: it is an indicator of how much environmental stress a community can tolerate before extinction of any of its constituent species. Formally, feasibility is defined as the volume of the parameter space of intrinsic growth rates in which all species in a community can have positive abundances (Grilli et al., 2017; Serguei Saavedra, Rohr, Olesen, & Bascompte, 2016). Feasibility is essential for understanding how communities might respond to future environmental changes. For example, in a very feasible community, there is a large range of conditions under which all species stably coexist. Therefore, in the presence of an environmental perturbation, such as climate change or habitat loss, it is less likely that any of the species in the community decline to extinction. Conversely, in a community with low feasibility, there is a small range of conditions under which all species stably coexist. Therefore, perturbations are more likely to result in species extinctions.

We measured the importance of a link between species i and j (I_{ij}) as its contribution to the feasibility of a network, defined as the ratio between the feasibility of the network with (O) and without (R) the focal link: $I_{ij} = \Omega_O / \Omega_R$, where Ω is the feasibility (Figure 1). Importance values were expressed as $((100 * I_{ij}) - 100)$, such that $I_{ij} = 0$ if the feasibility of the network was identical with and without the focal link. We calculated feasibility following (Song et al., 2018). Full details of the mutualistic model and equations used can be found in (S Saavedra, Rohr, Olesen, & Bascompte, 2016; Serguei Saavedra et al., 2016; Song et al., 2018), but we outline these briefly below.

A generalized Lotka-Volterra model of the following form was used:

$$\begin{cases} \dfrac{dP_i}{dt} = P_i(r_i^{(P)} - \sum_j \alpha_{ij}^{(P)} P_j + \sum_j \gamma_{ij}^{(P)} A_j) \\ \dfrac{dA_i}{dt} = A_i(r_i^{(A)} - \sum_j \alpha_{ij}^{(A)} A_j + \sum_j \gamma_{ij}^{(A)} P_j) \end{cases}$$

where P_i and A_i give the abundance of plant and animal species i, respectively; r_i denotes the intrinsic growth rates; α_{ij} represents intraguild competition; and γ_{ij} is the mutualistic benefit. The mutualistic benefit follows the equation $\gamma_{ij} = \gamma_0 \gamma_{ij}/d_i^{\delta}$, where $\gamma_{ij} = 1$ if there is a link between species i and j and zero if there is no link; d_i is the degree of species i; δ is the mutualistic trade-off (Serguei Saavedra, Rohr, Dakos, & Bascompte, 2013); and γ_0 is the overall level of mutualistic strength. A mean field approximation was used for the intraguild competition parameters, setting $\alpha_{ii}^{(P)}$ and $\alpha_{ii}^{(A)}$ equal to 1 and $a_{ij}^{(P)}$ and $a_{ij}^{(A)}$ equal to $\rho(i \neq j)$. We estimated the mutualistic trade-off, δ, empirically across all networks in our dataset. δ is given by the slope of two linear regressions (Rohr et al., 2014)

$$\log(f_{ij}/d_i^P d_j^A) = a^P - \delta\log(d_i^P) \text{ and } \log(f_{ij}/d_i^A d_j^P) = a^A - \delta\log(d_i^A),$$

where f_{ij} is the link frequency between animal species j and plant species i, a^P is the intercept for plants and a^A is the intercept for animals. These regressions were performed together on the whole dataset. We obtained a value for δ of 0.339, which was consequently used in all simulations. To focus on mutualistic effects, we ran analyses with zero interspecific competition ($\rho = 0$), following (Serguei Saavedra et al., 2016).

Results were qualitatively identical using weak competition ($\rho = 0.01$) (Supplementary Material Figure S1). The average mutualistic strength was set as half the average mutualistic strength at the stability threshold. Contribution to feasibility could not be measured for 931 links which, when removed, resulted in at least one species having no connections; these were excluded from the dataset.

To examine the relationship between interaction vulnerability and importance, we used a linear mixed-effects model, with importance as the response variable, vulnerability as a fixed effect, and network identity as a random effect. Linear mixed-effects models were run and analysed using the 'lme4', 'car' and 'MuMIn' R packages (Bartoń, 2018; Bates, Machler, Bolker, & Walker, 2015; Fox & Weisberg, 2002; R Core Team, 2015).

Taxonomic consistency of vulnerability and importance

We next assessed the extent to which vulnerability and importance exhibited taxonomic consistency: the tendency for an interaction's vulnerability and importance to be more similar across all the networks in which it occurs than expected by chance. Significant taxonomic consistency would imply a form of evolutionary conservatism: that vulnerability and importance are intrinsic properties of interactions, rather than a function of the ecological context in which they occur (Stouffer et al., 2012). We made comparisons at three levels of taxonomic aggregation – genus, family and order – but not at the species level, as only 180 (5.8%) of interactions at the species level occurred more than once in the data (Stouffer et al., 2012). For each level of taxonomic aggregation, we excluded all interactions that only occurred once.

If vulnerability and importance exhibit taxonomic consistency, then all occurrences of a given interaction should have similar levels of vulnerability and importance. More specifically, variance in vulnerability and importance across all links of a given interaction should be low. Therefore, for each interaction, we first calculated the variance in vulnerability and importance across all the networks in which the interaction occurred. We then created a corresponding 'null interaction', comprising the same number of links as the empirical focal interaction, but consisting of links sampled randomly without replacement from across the dataset. Links that were part

of the focal interaction were excluded from this sample. To ensure vulnerability and importance values were comparable between networks, and to control for any network-level effects, we used only relative values of vulnerability and importance. Vulnerability values were rescaled between 0 and 1 at the network level, while importance values were already relative (see definition above). Relative values are more relevant for our study because we were interested in whether interactions tend to have the same relative roles in all communities in which they occur, rather than if they have the same absolute values of a particular property. For example, we wanted to know whether a given link was always the most vulnerable link in a community, rather than if it always has an absolute vulnerability value of, say, 0.7. For each taxon-taxon interaction, at each taxonomic level, we sampled 10,000 null interactions and recorded the mean paired difference between the observed and null variance in vulnerability and importance. If an interaction exhibits taxonomic consistency, the mean paired difference ($Variance_{Null}$ – $Variance_{Observed}$) will be positive, because the observed variance would be lower than that expected by chance. P was the probability that a null interaction had lower variance in vulnerability or importance than the observed interaction. Interactions had significant consistency when $P < 0.05$. Taxonomic consistency results were qualitatively identical using weak competition ($\rho = 0.01$) (Supplementary Material Figure S2).

References

Aizen, M. A., Sabatino, M., & Tylianakis, J. M. (2012). Specialization and rarity predict nonrandom loss of interactions from mutualist networks. *Science, 335*(6075), 1486–1489. doi:10.1126/science.1215320

Baker, N. J., Kaartinen, R., Roslin, T., & Stouffer, D. B. (2015). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography, 38*(2), 130–139. doi:10.1111/ecog.00913

Bartoń, K. (2018). MuMIn: Multi-Model Inference. Retrieved from https://cran.r-project.org/package=MuMIn

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

doi:10.18637/jss.v067.io1

Cameron, E. K., Sundqvist, M. K., Keith, S. A., CaraDonna, P. J., Mousing, E. A., Nilsson, K. A., ... T, C. A. (2019). Uneven global distribution of food web studies under climate change. *Ecosphere*, *10*(3), e02645.

Díaz, S., Purvis, A., Cornelissen, J. H. C., Mace, G. M., Donoghue, M. J., Ewers, R. M., ... Pearse, W. D. (2013). Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecology and Evolution*, *3*(9), 2958–2975. doi:10.1002/ece3.601

Emer, C., Memmott, J., Vaughan, I. P., Montoya, D., & Tylianakis, J. M. (2016). Species roles in plant–pollinator communities are conserved across native and alien ranges. *Diversity and Distributions*, *22*(8), 841–852. doi:10.1111/ddi.12458

Fox, J., & Weisberg, S. (2002). *An R Companion to Applied Regression* (Second edi). Thousand Oaks, California: Sage. doi:10.1177/0049124105277200

Galetti, M., Guevara, R., Côrtes, M. C., Fadini, R., Von Matter, S., Leite, A. B., ... Jordano, P. (2013). Functional extinction of birds drives rapid evolutionary changes in seed size. *Science*. doi:10.1126/science.1233774

Grilli, J., Adorisio, M., Suweis, S., Barabás, G., Banavar, J. R., Allesina, S., & Maritan, A. (2017). Feasibility and coexistence of large ecological communities. *Nature Communications*, *8*. doi:10.1038/ncomms14389

Guimarães, P. R., Pires, M. M., Jordano, P., Bascompte, J., & Thompson, J. N. (2017). Indirect effects drive coevolution in mutualistic networks. *Nature*, *550*(7677), 511–514. doi:10.1038/nature24273

Harvey, E., Gounand, I., Ward, C. L., & Altermatt, F. (2017). Bridging ecology and conservation: from ecological networks to ecosystem function. *Journal of Applied Ecology*. doi:10.1111/1365-2664.12769

Janzen, D. H. (1977). Promising directions of study in tropical animal-plant interactions. *Annals of the Missouri Botanical Garden*, *64*(4), 706–736.

Jordano, P. (2016). Chasing Ecological Interactions. *PLoS Biology*, *14*(9). doi:10.1371/journal.pbio.1002559

Kaiser-Bunbury, C. N., & Blüthgen, N. (2015). Integrating network ecology with applied

conservation: A synthesis and guide to implementation. *AoB PLANTS*. doi:10.1093/aobpla/plv076

Leigh, E. G. (1977). How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences, 74*(10), 4542–4546. doi:10.1073/pnas.74.10.4542

Martín González, A. M., Dalsgaard, B., Nogués-Bravo, D., Graham, C. H., Schleuning, M., Maruyama, P. K., … Martinez, N. D. (2015). The macroecology of phylogenetically structured hummingbird-plant networks. *Global Ecology and Biogeography, 24*(11), 1212–1224. doi:10.1111/geb.12355

Ollerton, J., Winfree, R., & Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos, 120*(3), 321–326. doi:10.1111/j.1600-0706.2010.18644.x

Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos, 124*(3), 243–251. doi:10.1111/oik.01719

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ratto, F., Simmons, B. I., Spake, R., Zamora-Gutierrez, V., MacDonald, M. A., Merriman, J. C., … Dicks, L. V. (2018). Global importance of vertebrate pollinators for plant reproductive success: a meta-analysis. *Frontiers in Ecology and the Environment*. doi:10.1002/fee.1763

Rausand, M. (2013). *Risk assessment: theory, methods, and applications* (Vol. 115). John Wiley & Sons.

Redford, K. H. (1992). The Empty Forest. *BioScience, 42*(6), 412–422. doi:10.2307/1311860

Rodríguez-Flores, C. I., Ornelas, J. F., Wethington, S., & Arizmendi, M. del C. (2019). Are hummingbirds generalists or specialists? Using network analysis to explore the mechanisms influencing their interaction with nectar resources. *PLOS ONE, 14*(2), 1–32. doi:10.1371/journal.pone.0211855

Rohr, R. P., Saavedra, S., & Bascompte, J. (2014). On the structural stability of mutualistic systems. *Science, 345*(6195), 1253497–1253497.

doi:10.1126/science.1253497

Saavedra, S., Rohr, R. P., Dakos, V., & Bascompte, J. (2013). Estimating the tolerance of species to the effects of global environmental change. *Nature Communications*, *4*, 2350. doi:10.1038/ncomms3350

Saavedra, S., Rohr, R. P., Olesen, J. M., & Bascompte, J. (2016). Data from: Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and Evolution*. Dryad Digital Repository. doi:doi:10.5061/dryad.3pk73

Saavedra, S., Rohr, R. P., Olesen, J. M., & Bascompte, J. (2016). Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and Evolution*, *6*(4), 997–1007. doi:10.1002/ece3.1930

Saavedra, S., Stouffer, D. B., Uzzi, B., & Bascompte, J. (2011). Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, *478*(7368), 233–235. doi:10.1038/nature10433

Santamaría, S., Galeano, J., Pastor, J. M., & Méndez, M. (2016). Removing interactions, rather than species, casts doubt on the high robustness of pollination networks. *Oikos*, *125*(4), 526–534. doi:10.1111/oik.02921

Säterberg, T., Sellman, S., & Ebenman, B. (2013). High frequency of functional extinctions in ecological networks. *Nature*, *499*, 468–470. doi:10.1038/nature12277

Song, C., Rohr, R. P., & Saavedra, S. (2018). A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *Journal of Theoretical Biology*, *450*, 30–36. doi:10.1016/j.jtbi.2018.04.030

Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, *335*(6075), 1489–1492. doi:10.1126/science.1216556

Trøjelsgaard, K., Jordano, P., Carstensen, D. W., & Olesen, J. M. (2015). Data from: Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. *Proceedings of the Royal Society B*. Dryad Digital Repository. doi:doi:10.5061/dryad.76173

Trøjelsgaard, K., Jordano, P., Carstensen, D. W., & Olesen, J. M. (2015). Geographical variation in mutualistic networks: Similarity, turnover and partner fidelity. *Proceedings of the Royal Society B: Biological Sciences, 282*(1802), 1–9. doi:10.1098/rspb.2014.2925

Tylianakis, J. M., Didham, R. K., Bascompte, J., & Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters, 11*(12), 1351–1363.

Valiente-Banuet, A., Aizen, M. A., Alcántara, J. M., Arroyo, J., Cocucci, A., Galetti, M., ... others. (2015). Beyond species loss: the extinction of ecological interactions in a changing world. *Functional Ecology, 29*(3), 299–307.

Wandrag, E. M., Dunham, A. E., Duncan, R. P., & Rogers, H. S. (2017). Seed dispersal increases local species richness and reduces spatial turnover of tropical tree seedlings. *Proceedings of the National Academy of Sciences, 114*(40), 10689–10694. doi:10.1073/pnas.1709584114

4 | Motifs in bipartite ecological networks: uncovering indirect interactions

This chapter was published as Simmons, B. I., Cirtwill, A. R., Baker, N. J., Wauchope, H. S., Dicks, L. V., Stouffer, D. B. and Sutherland, W. J. (2019), Motifs in bipartite ecological networks: uncovering indirect interactions. Oikos, 128: 154-170.

Abstract

Indirect interactions play an essential role in governing population, community and coevolutionary dynamics across a diverse range of ecological communities. Such communities are widely represented as bipartite networks: graphs depicting interactions between two groups of species, such as plants and pollinators or hosts and parasites. For over thirty years, studies have used indices, such as connectance and species degree, to characterise the structure of these networks and the roles of their constituent species. However, compressing a complex network into a single metric necessarily discards large amounts of information about indirect interactions. Given the large literature demonstrating the importance and ubiquity of indirect effects, many studies of network structure are likely missing a substantial piece of the ecological puzzle. Here we use the emerging concept of bipartite motifs to outline a new framework for bipartite networks that incorporates indirect interactions. While this framework is a significant departure from the current way of thinking about bipartite ecological networks, we show that this shift is supported by analyses of simulated and empirical data. We use simulations to show how consideration of indirect interactions can highlight differences missed by the current index paradigm that may be ecologically important. We extend this finding to empirical plant-pollinator communities, showing how two bee species, with similar direct interactions, differ in how specialised their competitors are. These examples underscore the need to not rely solely on network- and species-level indices for characterising the structure of bipartite ecological networks.

Introduction

Ecological communities are widely represented as bipartite networks that depict interactions between two groups of species, such as plants and pollinators. These networks are used to answer a diverse range of questions about community structure, such as whether antagonistic and mutualistic communities have different architectures (Fontaine et al. 2011, Morris et al. 2014); how plant-frugivore communities at forest edges differ from those in forest interiors (Menke et al. 2012); whether fluctuations in species and interactions over time alter network structure (Petanidou et al. 2008); and whether individual pollinators vary in their use of floral patches (Dupont et al. 2014).

For over thirty years, the framework for characterising the structure of bipartite networks has remained unchanged: indices, such as nestedness and species degree, are used to describe either whole-network topology or the roles of individual species with a single summary statistic. However, while these network- and species-level indices have greatly improved our understanding of community structure, they also suffer from a substantial, but largely ignored, ecological limitation: reducing a complex network to a handful of one-dimensional metrics necessarily involves a loss of information. This is because network and species-level indices are insensitive to changes in pairwise species interactions: different network configurations can have identical index values (Olito and Fox 2015). Often this means discarding important detail about indirect interactions. For example, let there be two communities: in the first community, plant i is pollinated by one species, j; in the second community, i is still only pollinated by j, but j also pollinates plants k, l and m. We cannot distinguish the two situations by examining, for example, the degree of i because degree discards all information on indirect interactions: we know that i has a direct interaction with j, but we do not know whether j is an obligate specialist on i or a generalist visiting several other plants.

The loss of ecological detail resulting from the use of network and species-level indices is concerning as it puts many studies describing network structure directly at odds with a large literature that has repeatedly documented important and widespread indirect effects in nature (Wootton 2002). For example, in mutualistic networks, dynamical

models (which use the whole network as the skeleton of dynamics and therefore incorporate indirect interactions) have shown that indirect effects are a major process governing coevolution (Guimarães et al. 2017), while in host-parasitoid communities, apparent competition and even apparent mutualism can occur when herbivorous insects influence each other through shared natural enemies (Morris et al. 2004, Frank van Veen et al. 2006, Tack et al. 2011). Similarly, indirect effects between co-flowering plant species in pollinator communities can range from facilitation, where the presence of one plant increases the frequency of pollinator visits to another, to competition, where one plant attracts pollinators away from another (Mitchell et al. 2009, Morales and Traveset 2009, Carvalheiro et al. 2014). Indirect interactions are therefore a fundamental component of ecosystems, driving ecological and evolutionary processes to an equal, or greater, extent than direct interactions (Vandermeer et al. 1985, Strauss 1991, Bailey and Whitham 2007, Martínez et al. 2014, Guimarães et al. 2017). Widespread, uncritical use of network and species-level indices as the sole method for characterising network structure risks missing all or part of this component.

Here we advocate a new way of thinking about bipartite networks that complements existing index-based approaches by incorporating more explicit detail on realised direct interactions and thus potential indirect interactions. We argue for conceptualising networks as a collection of constituent parts or 'building blocks' using the emerging concept of bipartite motifs (subgraphs representing patterns of interactions between a small number of species). We outline the theory, applications and future directions of this framework. We show that a motif conceptualisation of networks is well supported by simulated and empirical data, using three analyses to demonstrate the importance of the local-scale topological detail captured by motifs. First, we use three six-species networks to show that indirect interactions are necessary to accurately describe a species' role in even a small community. Through simulation, we then generalise this finding to a large ensemble of networks with diverse sizes and structures to establish and quantify how communities and species with similar overall properties can exhibit remarkable dissimilarity in their indirect interaction structures. Finally, we demonstrate these results in an empirical context, highlighting how indirect interactions can result in ecologically important differences between two pollinator

species with similar direct interactions. We also assess the robustness of the framework to sampling effort and propose several hypotheses about how our understanding of ecological communities might change if indirect interactions were incorporated. While our focus here is on bipartite mutualistic networks, such as those representing plant-pollinator interactions, we note that the ideas presented here are also applicable to non-mutualistic bipartite networks, such as plant-pest, host-parasitoid or plant-herbivore networks, and so we also provide some applications of our framework to these non-mutualistic systems. Importantly, we do not advocate removing network and species-level indices from the network ecologists 'toolbox'. Instead we hope to raise awareness of network and species-level indices' limitations and promote bipartite motifs as a complementary framework for characterising network structure. We anticipate a future where index- and motif-based approaches coexist in both empirical and theoretical studies of ecological networks.

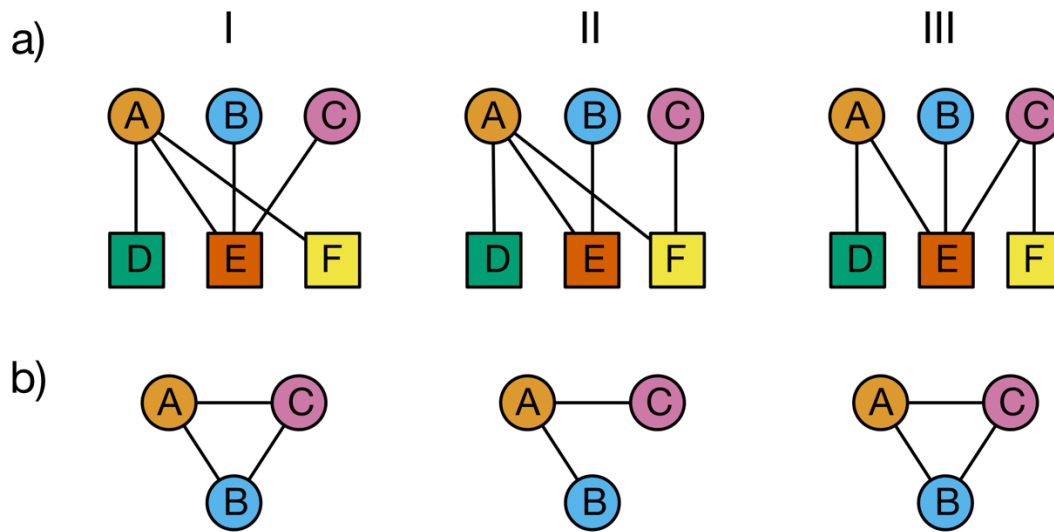## Indirect interactions and the index paradigm

Figure 1: Three example networks (a) and the corresponding one-mode projections for species in the upper level (b).

We define indirect interactions as the impact of one species on another, mediated by one or more intermediary species (Wootton 1994, 2002). There are two main types of indirect effects (Wootton 1994, 2002) (Supplementary Fig. 4). First, changes in the

abundance of a donor species can influence the abundance of a recipient species by affecting the abundance of an intermediary species that interacts with both (Wootton 1994). This is known as an interaction chain (Supplementary Fig. 4). Chains can comprise multiple steps, with more than one intermediary species: for example, in network I in Fig 1a., a change in the abundance of species *D* could indirectly affect the abundance of species *C*, through changes in the abundances of species *A* and *E*. Classic examples of interaction chains include apparent competition, exploitative competition, omnivory and tri-trophic chains (also known as trophic cascades). The second type of indirect interaction is where the interaction between two species is affected by a third species (Supplementary Fig. 4). This is known as an interaction modification. For example, a predator may refrain from feeding in a patch due to the presence of a defended plant species, therefore reducing consumption of another undefended plant species (Hay 1986, Pfister and Hay 1988, Wootton 1994). Here we focus on interaction chains, as these are represented by the topology of the network and are therefore captured explicitly by motifs. For example, among studies of unipartite (one type of node) food webs, where motif analyses are more common, motifs have been used to capture classic interaction chains like apparent competition, exploitative competition, omnivory and tri-trophic chains (Fig. 2a) (Camacho et al. 2007, Kondoh 2008, Bascompte and Stouffer 2009, Stouffer and Bascompte 2010).
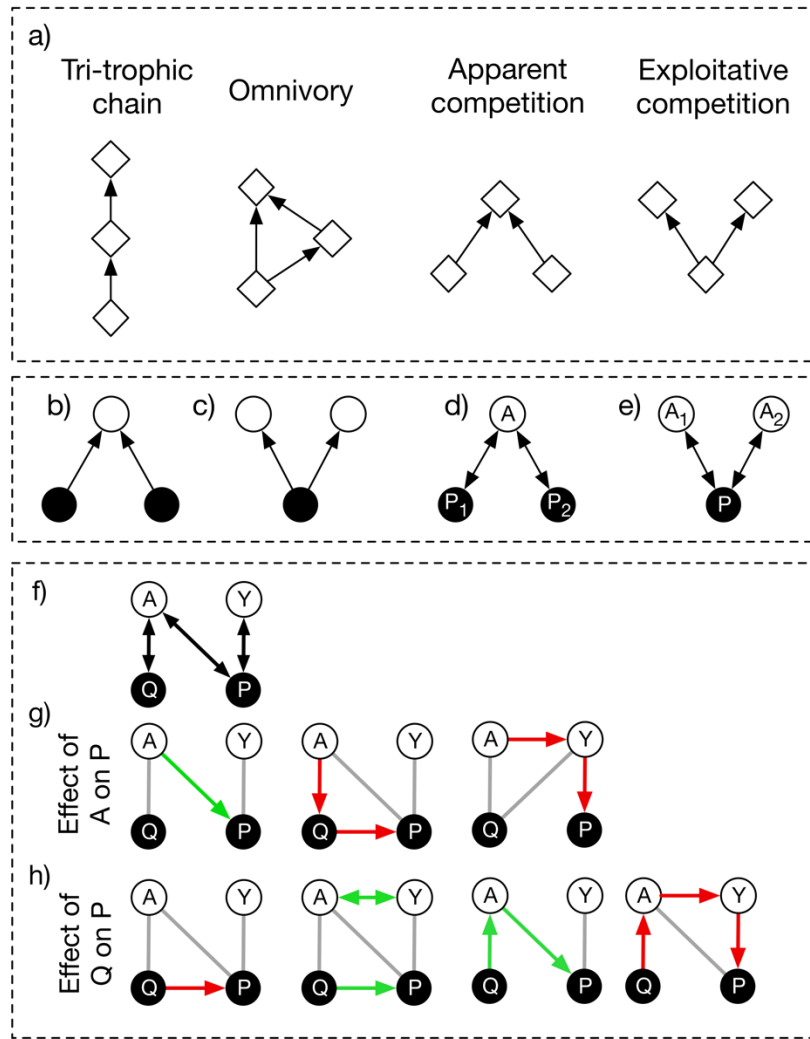
Figure 2: A selection of unipartite and bipartite motifs and their indirect effects. (a) Four classic unipartite motifs representing well-studied interaction chains. Motif 2 (b) and motif 3 (c) in an antagonistic network, with directed interactions from the lower to the higher trophic level. Motif 2 (d) and motif 3 (e) in a mutualistic network, with bidirectional interactions representing the mutual benefit that species receive from each other. P_1, P_2 and P represent plant species, while A, A_1 and A_2 represent animal species. (f) Motif 5 with mutualistic interactions as discussed in Vázquez et al (2015). (g) and (h) represent the direct and indirect effects of A on P (g) and Q on P (h). Paths resulting in a positive effect are shown in green, while paths resulting in a negative effect are shown in red. Adapted from Vázquez et al (2015).

To illustrate the importance of interaction chains, consider the example of species B in the three communities shown in Fig. 1a. Based only on its direct interactions, the role of B is identical in all three communities: B interacts with E. However, by considering the interactions of B's partner E, B's role in networks I and III can be distinguished from its role in network II: in networks I and III, B competes with A and C for the shared

resource *E*, while in network II, *B* competes with only *A*. In other words, in networks I and III, *C* can indirectly influence *B* through a short interaction chain with only one intermediary species (*E*), while in network II, the interaction chain between *C* and *B* is longer, involving three intermediary species (*F*, *A* and *E*). Furthermore, by considering the interactions of *A* and *C* (*B*'s partners' partners), the roles of *B* in networks I and III can also be distinguished: in network I, *B*'s competitor *C* is a specialist on resource *E*, while in network III, *C* also visits *F*. Similarly, while *A* is a super-generalist in network I, visiting every resource in the community, in network III it has a narrower diet breadth, visiting only *D* and *E*.

This simple example shows how indirect interactions are necessary to give a complete picture of a species' role, even in a small community: between the three networks, *B* differs both in the number of competitors it has and in how specialised these competitors are on the shared resource *E*. Such differences are likely to have important ecological consequences. To capture this detail, it was not sufficient to consider only *B*'s direct interactions, or even the interactions of *B*'s partner; rather we had to go 'deeper' and consider the interactions of *B*'s partners' partners to differentiate its role in all three networks.

Many indices that capture interaction patterns at the level of individual species – such as degree, dependence (the strength of an interaction between species *i* and *j* as a proportion of *i*'s total interaction strength) or species strength (sum of dependencies on a species) – are largely based on direct interactions and so do not give a complete picture of *B*'s role in the three communities (Bascompte et al. 2006). Other species-level indices, such as *z*- and *c*- scores, do consider indirect interactions, but only with respect to modules (groups of species connected more to each other than to other species in the network). It is important to note that modules and motifs are not equivalent: all networks contain motifs, but not all networks contain modules. This is because motifs have no requirement that their nodes must be more connected to each other than to other nodes in the network, and are simply a decomposition of a network into its constituent subgraphs. *z*-scores quantify a species' connectivity within modules, while *c*-scores (also known as the participation coefficient) quantify a species connectivity among modules. In all three networks, *B* has identical *c*-scores, while *z*-scores for *B* are

identical in networks I and III. Various centrality indices also incorporate indirect interactions, but rely on the one-mode projection of the bipartite network, where species in one set are linked when they share one or more partners in the other set (Jordán et al. 2007) (Fig. 1b). This compression necessarily leads to a substantial loss of information (Zhou et al. 2007, Saracco et al. 2017). While more sophisticated one-mode projections are available, and these would likely lose less information than the simple projection detailed above, projecting a bipartite network into a unipartite network will always lose some detail. One such detail is that interactions with specialist species such as D will not be considered. This is because specialists only interact with one species and therefore cannot be shared between multiple partners. Given that many one-mode projections are fundamentally based on linking species when they share a partner, singleton species such as D are not accounted for. Consequently, betweenness centrality (the number of shortest paths between two species passing through a focal species) and closeness centrality (the mean shortest path between the focal species and all other species) values for B are identical in all three networks. Finally, even multivariate combinations of common indices describing whole-network structure cannot distinguish between these three situations because all three communities have identical connectance, nestedness and modularity.

This is an example of the Goldilocks principle: by accounting for all interactions simultaneously, indices characterising whole-network patterns can be too coarse to detect fine differences. Conversely, by considering too little of the indirect interaction structure, indices describing individual species roles can miss differences beyond their local scope. In both cases, indirect interactions occurring at a level between these whole-network- and species-scales – that is, at the meso-scale – may be missed. This is not to say that network and species-level indices cannot capture *any* information about indirect interactions. For example, in bipartite networks, species can be involved in indirect interactions either because they are generalist or because their partner is generalist. Therefore, by measuring the extent to which two sets of species are asymmetrically specialised, nestedness is able to capture aspects of indirect interactions (Bastolla et al. 2009). Similarly, compartmentalisation describes how indirect competition in food webs, or even energy pathways (Zhao et al. 2018), may be

constrained to some specific groups (Stouffer and Bascompte 2011). The related species-level metrics (c- and z-scores) highlight how species may contribute to these. Therefore, network and species-level indices can provide useful information on indirect interactions. However, because network and species-level indices are one-dimensional, this information can only capture specific aspects of network structure or a species' role rather than a complete picture of how each species is embedded in the community. Given the importance and ubiquity of indirect interactions, there is room for an alternative framework for describing network structure that uncovers the indirect interactions present in the meso-scale topology of networks.

A framework for indirect interactions

We start by recognising the fact that any given network made up of S species can be broken down into a series of smaller subnetworks containing n species (where $n < S$ and all species have at least one interaction). For example, network I in Fig. 1a includes five subnetworks containing two species (A–D, A–E, A–F, B–E, C–E) and six subnetworks containing three species (D–A–E, D–A–F, E–A–F, A–E–B, A–E–C, B–E–C). As there are a finite number of ways to arrange interactions between n species, there are also a finite number of possible subnetworks of size n that a network can contain. In other words, all bipartite networks, regardless of their complexity, are assembled from a limited number of parts or building blocks known as 'bipartite motifs' (Baker et al. 2015). For example, Fig. 3 shows all 44 possible motifs containing between two and six species. We argue that an understanding of these basic structural elements captures the details of indirect interactions beyond the global and local features captured by network and species-level indices (Milo et al. 2002).
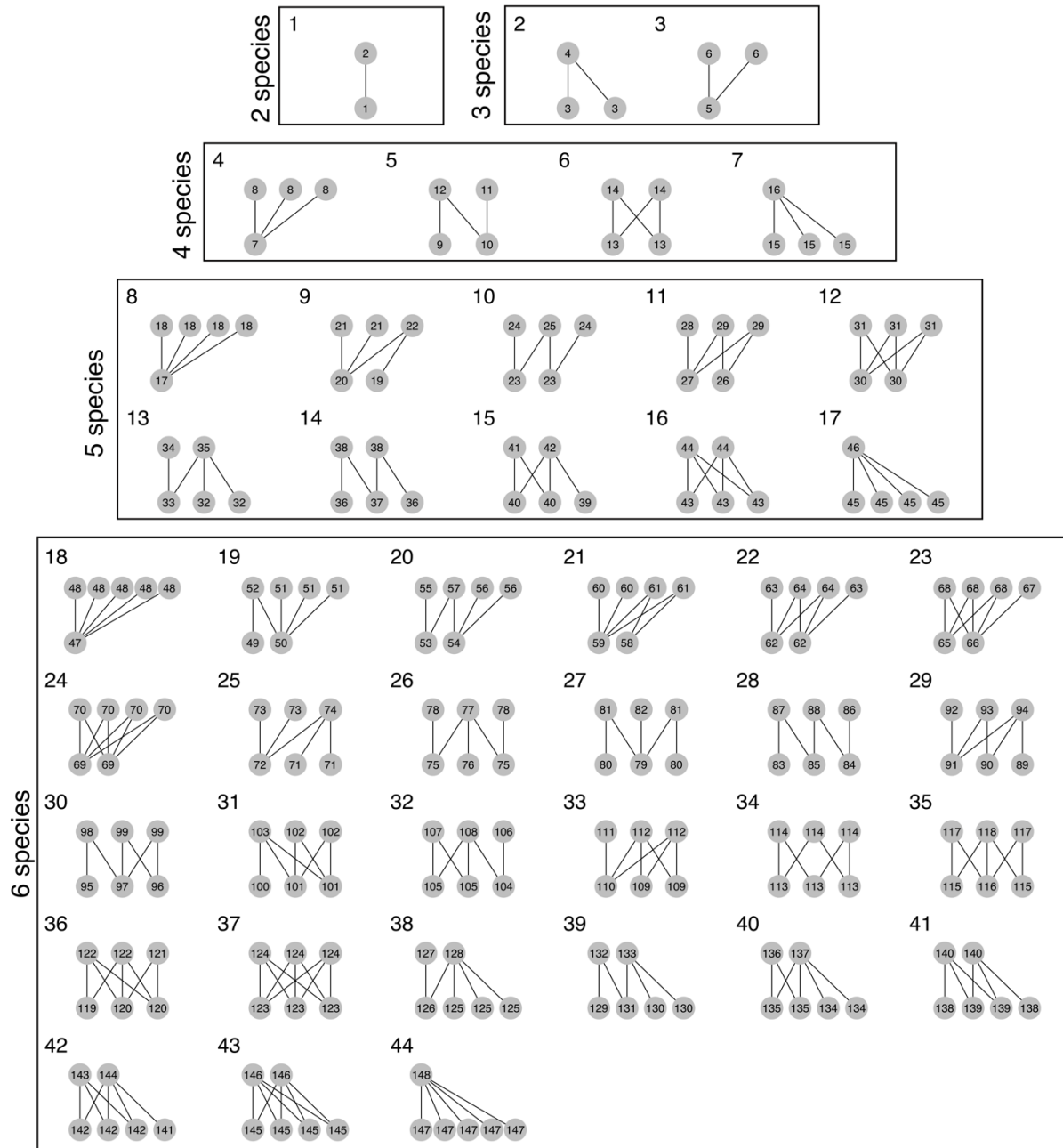
Figure 3: All possible two- to six-species bipartite motifs. Large numbers represent individual motifs. Small numbers within nodes represent unique position within motifs. In total there are 148 positions across 44 motifs.

While motifs are not yet widely adopted in studies of bipartite networks, they have seen much greater uptake for describing indirect interactions in studies of unipartite food webs. Not all unipartite motifs have a simple ecological interpretation, but some represent classic, and well-studied, examples of indirect effects. Fig. 2a shows four such motifs. The tri-trophic chain, or trophic cascade, motif represents a situation where the

species at the lowest trophic level benefits from the decrease in predation that results from the predator at the top level reducing the abundance of the species in the middle level. In the omnivory or intraguild predation motif, a predator and its prey compete for the same resource. In the apparent competition motif, two prey species share a predator: if one prey species increases in abundance this can lead to an increase in abundance of the common predator which, in turn, reduces the abundance of the other prey species. Finally, in the exploitative competition motif, two predators compete for the same prey: if one predator reduces the abundance of the shared resource, this indirectly reduces the abundance of the other predator. These simple interaction chains have been used to show that motifs which occur most often in empirical food webs are those which contribute the most to community persistence (Stouffer and Bascompte 2010); that food webs can be regarded as a collection of interconnected motifs in a non-random configuration that enhances biodiversity maintenance (Kondoh 2008); and that simple models of food web structure can reproduce the local topology of networks (Camacho et al. 2007).

As with unipartite motifs, some bipartite motifs have simple ecological interpretations, while others represent more complex situations. For example, some of the classic unipartite motifs have direct bipartite counterparts: when bipartite networks represent antagonistic interactions, such as herbivory or parasitism, motifs 2 and 3 (Fig. 3) represent apparent competition and exploitative competition, respectively (Fig. 2b, c). When bipartite networks represent mutualistic interactions, these same motifs can represent indirect competitive (Waser 1983, Campbell 1985, Mitchell et al. 2009, Runquist and Stanton 2013, Ye et al. 2014) or facilitative interactions (Moeller 2004, Ghazoul 2006, Liao et al. 2011, Sieber et al. 2011) (Fig. 2d, e). For example, if species P_1 and P_2 in Fig. 2d represent plants, and species A represents a pollinator, plants may be involved in exploitative competition for finite pollinator resources, or interference competition through interspecific pollen deposition (Chittka and Schürkens 2001, Moeller 2004, Mitchell et al. 2009, Flanagan et al. 2010, Hochkirch et al. 2012, Ye et al. 2014). Conversely, facilitative effects can occur where an increase in the abundance of P_1 could indirectly benefit P_2 through providing a beneficial effect to A (Moeller 2004, Sotomayor and Lortie 2015). For example, the presence of one plant species could

increase pollinator visits to a coflowering species (Moeller 2004, Ghazoul 2006, Carvalheiro et al. 2014, Ye et al. 2014). A similar situation is found in Fig. 2e, where A_1 and A_2 represent pollinators and P represents a plant species: exploitative and interference competition between pollinators can occur, as can indirect facilitation, where an increase in the abundance of A_1 can provide a beneficial effect to A_2 through P (Rathcke 1983, Temeles et al. 2016). Similar ideas can be applied to other motifs where many specialists interact with a single generalist, such as motifs 4, 7, 8, 17, 18 and 44 (Fig. 3). These motifs extend the implications of the competition motifs discussed above (motifs 2 and 3) to having many species in the specialist group. These 'fan' motifs capture situations where all the specialists affect each other indirectly via their effect on the generalist, and are also notable for representing situations where all potential indirect interactions are on the same side of the network, in contrast to other motifs which are more 'balanced'.

Within motifs, species can occupy different positions (Kashtan et al. 2004). For example, in motif five there are four unique positions, as each species interacts with a unique set of partners (Fig. 3). Considering all bipartite motifs up to six species, there are 148 unique positions (Fig. 3). Note that, due to symmetry, there may be fewer than n unique positions in a motif with n species. For example, in motif six there are only two unique positions, as both species in the top level interact with both species in the bottom level (Fig. 3). Therefore, a bipartite motif with n species can include between 2 and n unique positions. These positions have distinct ecological meanings, with different positions corresponding to species with different direct and indirect interactions (Stouffer et al. 2012, Baker et al. 2015, Cirtwill and Stouffer 2015). Vázquez *et al.* (2015) highlight this point by detailing some of the different indirect effects that are captured by motif 5 (Fig. 2f, Fig. 3). For example, while animal A has a positive direct effect on plant P, A has two negative indirect effects on P, first by providing a beneficial effect to P's competitor Q, and second by suppressing animal Y's growth rate, which reduces the direct beneficial effect of Y on P (Fig. 2g). Considering the effect of Q on P highlights a similar mix of positive and negative effects (Fig. 2h). Vázquez et al. highlight four interaction pathways. First, Q has a direct short-term negative effect on P due to direct competition. Second, a net positive effect results from the negative of the product between the

mutually negative effects of *A* on *Y* and the negative effect of *Q* on *P*. Third, *Q* has a positive indirect effect on *P* through *A*. Fourth, *Q* has a negative effect on *P* by providing beneficial effects to *A* which suppresses *A*'s competitor *Y* which, in turn, has a negative effect on *P*. This mixture of positive and negative direct and indirect effects show how interactions between species can mitigate or cancel each other out. While here our focus is on the topology of these interactions, dynamic models or experimentation can be used to determine the overall effects of one species on another (Vázquez et al. 2015).

We can also derive some broad expectations for indirect interactions in motifs with particular structures. For example, in motifs where all species in one group interact with all species in the other group (such as motifs 6,16,24,37,43), we might expect indirect interactions to be stronger than those in 'fan' motifs because they can be transmitted via multiple routes at the same time. However we might also expect dynamics in these 'complete' motifs to be less predictable. For example, a decrease in the abundance of a pollinator species in a 'complete' motif would decrease the population of all plants in the motif, but also decrease the amount of competition for other pollinators using those plants. For 'asymmetric complete' motifs such as 11, 15, 31 and 42, where a specialist species is attached to a group of species which all interact, we might expect that generalists affect the specialists more strongly than *vice versa*, since the generalists have more choice of interactions and can buffer changes in one partner's abundance. It is also important to note that, because different trophic groups often live on different timescales, the difference between bottom-majority and top-majority motifs is important. For example, following a decrease in a pollinator species' abundance, the reduction in competition between pollinators would occur more quickly than a decrease in plant populations.

Above we have provided some examples of how bipartite motifs capture indirect interactions. While a detailed study of all 44 bipartite motifs is beyond the scope of this work, detailed dynamic modelling and interpretation of individual motifs is an important area for future research. Given the large number of papers that have been dedicated to studying individual motifs in unipartite food webs, we anticipate that this is a fruitful area for further studies. Ultimately, motifs are a way to capture the topology of interaction chains explicitly. Considering up to six-node motifs as we do here, this

means that motifs can capture interaction chains with up to four intermediary species between the donor and recipient species. For example, Supplementary Fig. 5 shows how even a small five-species motif captures 28 interaction chains.

We have discussed how motifs relate to indirect interactions, but it is also important to detail how motifs can actually be used to characterise the structure of ecological networks and their constituent species. To characterise network structure in the motif framework, networks are first decomposed into their constituent motifs, giving an inventory of the parts which make up the network. These simple lists show the frequency c_i with which each motif i occurs in a network. This provides an m-dimensional 'signature' of a network's structure, given by the vector $\overrightarrow{c_i} = \{c_1, c_2 \dots, c_m\}$, where m is the number of motifs counted. For example, Fig. 4 shows the constituent motifs of each example network from Fig. 1 and Fig. 5a shows the structural signature of each network. When viewed in this way, it becomes clear that each of these communities is made up of different parts, despite having similar or identical values of several common network-level indices.
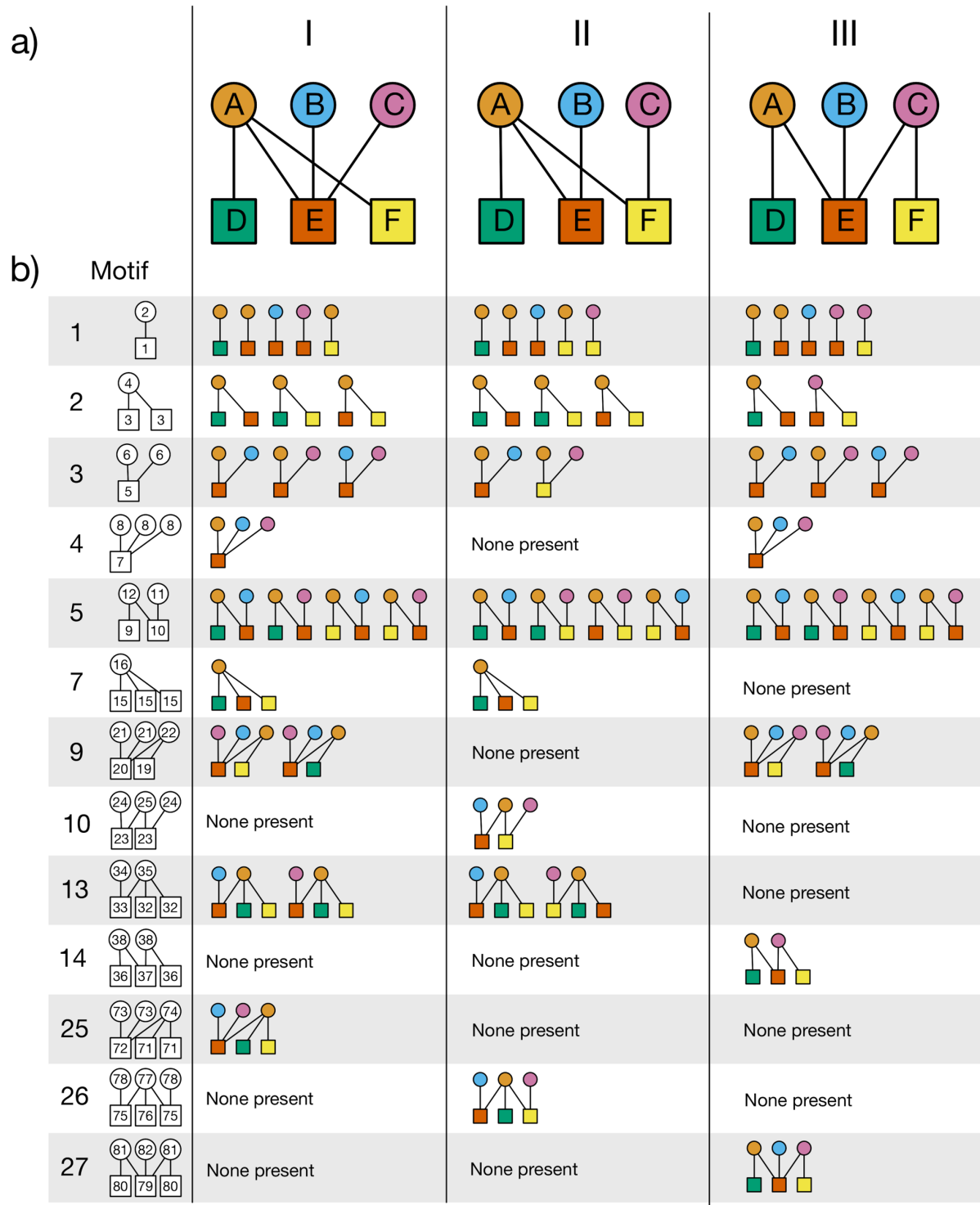
Figure 4: Decomposing three example networks into their constituent two- to six-node motifs. (a) Three example networks also used in Fig. 1a. (b) Table showing each network's constituent motifs. The first column shows the motif being counted: the large number refers to the ID of the motif as given in Fig. 2; the small number within each node refers to the unique positions species can occupy within each motif as given in Fig. 2. The second, third and fourth columns show the occurrences of each motif in networks I, II and III respectively. Node colours refer to the species involved in each motif. For visualisation purposes, we exclude motifs which do not occur in any network, such as motif 8.

As noted above, species can occupy different positions within motifs. As these positions have distinct ecological meanings, a species' role in a network can be defined by the frequency with which it occurs in each position (Stouffer et al. 2012, Baker et al. 2015, Cirtwill and Stouffer 2015). For example, Fig. 4b shows how, in network I, species B occurs once in position two, twice in position six, once in position eight, and so on. Generally, therefore, species roles are described by a vector $\overrightarrow{r_{xy}} = \{r_{x1}, r_{x2} \dots, r_{xp}\}$, where r_{xy} is the frequency with which species x occurs in position y and p is the number of positions counted. This vector can be thought of as a p-dimensional signature of a species' role, or its multidimensional 'interaction niche'. Fig. 5b shows the role signature of species B in the three networks from Fig. 1a. The roles are different in each network, demonstrating how this framework, by capturing indirect interactions at the meso-scale, distinguishes species roles that many network and species-level indices cannot. Additionally, this did not require an *a priori* selection of the particular aspect of network structure to examine for differences; rather the approach is general, simply providing a detailed view of how B is embedded in the community.
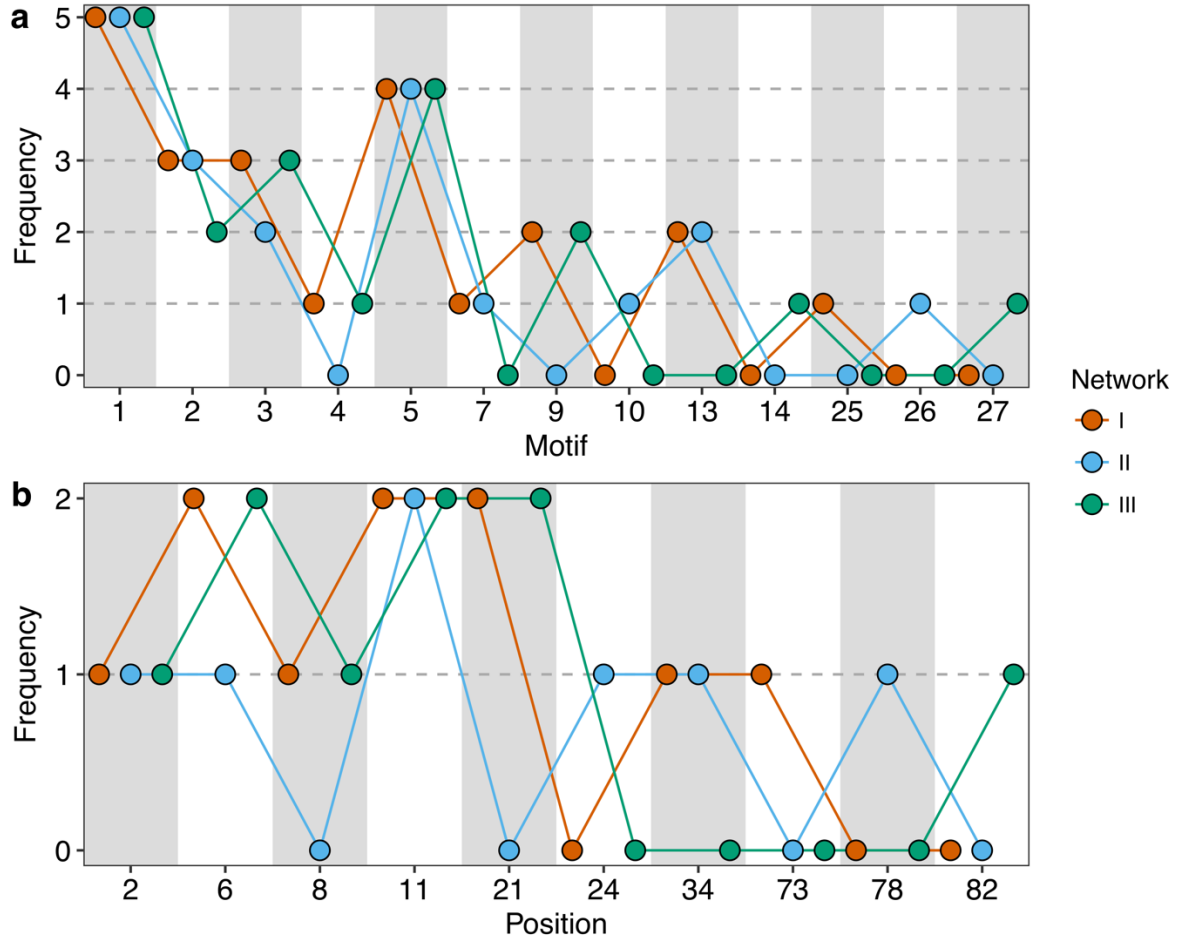
Figure 5: Motif and position frequencies of networks I, II and III from Fig. 1a and Fig. 3a. (a) The frequency with which each motif occurs in each network. For visualisation purposes, we exclude motifs which do not occur in any network. (b) The frequency with which species B occurs in each unique position within motifs in each network. For visualisation purposes, we exclude positions within which species B does not occur in any network. Motif and positions IDs correspond to those given in Fig. 2 and Fig. 3.

Here we have considered bipartite motifs containing between 2 and 6 species. Studies of unipartite networks usually consider only three-species motifs (Kondoh 2008, Stouffer and Bascompte 2010, Stouffer et al. 2012, Borrelli 2015). Since interactions in these unipartite networks are directed, there are 13 possible three-species motifs which alone can provide a good description of meso-scale network structure (Milo et al. 2002). Conversely, in the undirected bipartite networks we consider here, only two three-species motifs exist (Fig. 3), and therefore only considering motifs of this size would not capture much information about how interactions are distributed in networks. Similarly, while there are 199 possible four-node motifs in directed unipartite networks,

there are only four in undirected bipartite networks (Fig. 3). Therefore, for undirected bipartite networks, it is important to consider larger motifs in order to adequately capture meso-scale structure (Baker et al. 2015, Cirtwill et al. 2018).

Motifs can be of any size and a completely lossless description of network structure would require counting motifs containing up to n nodes, where n is the number of nodes in the network. However, this would be analytically intractable for all but the smallest networks, as it would involve a vast number of different motifs and be very difficult to enumerate. Therefore it is necessary to decide the maximum size of motifs to be included in a given analysis, given the computational and methodological constraints involved in counting larger motifs. These computational and methodological considerations are particularly important for modern network research, where analyses involving large numbers of calculations are common, such as using ensembles of null networks to control for network size and connectance. There is therefore a trade-off between capturing more structural detail with larger motifs, and the methodological and computational challenges associated with the discovery and counting of larger motifs and their constituent motif positions. We are not aware of any studies using bipartite motifs of more than six nodes. Additionally, to our knowledge, motif positions have never been defined for motifs greater than six nodes. Therefore, while more information may be embedded in 7-, 8-, or 9-species motifs, and so on, 6-species motifs strike a useful balance between capturing detail and being analytically tractable. Determining whether incorporating larger bipartite motifs adds significantly more information is an important area for future research, especially as computational capacity continually improves. Whatever the maximum size used, we note that a significant pattern at n species necessarily implies a pattern at sizes greater than n because they are composites of the level below.

Comparing network- and species-level indices to motifs

In Figs. 1, 4 and 5, we used simple six-species networks to demonstrate how network and species-level indices can mask potentially important meso-scale variation in indirect interactions. Here we generalise this effect to a large ensemble of networks of

varying sizes and structures using simulations. We first generated 20,000 bipartite networks containing 6 to 50 species in each set (giving 12 to 100 species) and with connectances ranging between the minimum required for each species to interact with at least one partner and 0.5. Networks were generated using the bipartite cooperation model (Saavedra et al. 2009). We chose this model because cooperative interactions are one of the main types of interaction represented as bipartite networks in ecology, with pollination and seed dispersal being two of the most popular examples (Bascompte and Jordano 2007). For our analysis, we needed to generate many networks with a range of structures, but we also wanted these networks to be realistic. The bipartite cooperation model was therefore suitable because it reproduces many important structural features of empirical ecological cooperative networks (Saavedra et al. 2009).

Network level

We characterised the structure of each of the 20,000 networks at the macro-scale, using three whole-network indices (connectance, nestedness, modularity). Nestedness was measured as NODF (Almeida-Neto et al. 2008) and modularity was calculated using using the 'computeModules' function from the 'bipartite' R package (Dormann et al. 2009, R Core Team 2015, Beckett 2016). We ranked networks according to each macro-scale index (connectance, nestedness, and modularity) in turn and divided networks into subsets of 50 according to this ranking. For example, when ranking networks by connectance there would be 400 subsets, each containing 50 networks with similar values of connectance.

For each network in each subset, we characterised its structure at both the macro- and meso-scale. At the macro-scale, networks were described with a vector containing the number of species in the first set, the number of species in the second set, and two of connectance, nestedness, and modularity (having excluded the metric used for ranking networks). For example, if connectance was used as the ranking property, the macro-scale structure vector would include the number of species in the first set, the number of species in the second set, nestedness and modularity. At the meso-scale, network structure was described using the frequencies of motifs containing between three and

six species, $\underset{c_i}{\rightarrow}$, as described above. Motif frequencies were calculated using the 'bmotif' R package (Simmons et al. 2018). We excluded the two-species motif representing a direct interaction between two species because the focus here is on indirect interactions.

For each subset, we calculated pairwise distances between all network structural vectors at both the macro- and meso-scale. At the macro-scale, we calculated distances between networks' macro-scale vectors, while at the meso-scale we calculated distances between networks' meso-scale vectors. Distances were calculated using the 'correlation' distance measure in the 'rdist' function from the 'rdist' R package (Blader 2018). This converts correlations ranging between –1 and 1 to distances ranging between 0 and 1, following $\sqrt{((1-r)/2)}$, where r is the correlation between two vectors. To control for the possibility that some subsets might have more variable structure than others, we then normalised distances by dividing by the maximum distance between any two networks, giving values between 0 (for identical networks) and 1 (for completely different networks).

For each subset, we then calculated the distance between each network's macro- or meso-scale structural vector and the subset centroid representing the 'typical' structure for each subset. These distances were calculated using the 'betadisper' function from the 'vegan' R package (Oksanen et al. 2016). Again, this was done separately for each scale: for the macro-scale analysis, the 'betadisper' function was fed the distance-based object based on the macro-scale vectors, while for the meso-scale analysis, the function was fed the distance-based object based on the meso-scale vectors.

We repeated this procedure using each of connectance, nestedness, and modularity as the ranking variable, to give three views of the variability of network structure at macro- and meso-scales. We then used paired Wilcoxon signed-rank tests to compare the differences in variation between macro- and meso-scale subsets.

The median paired differences in variation between macro- and meso-scale subsets were 0.13 (Wilcoxon: p < 0.0001) when ranked by connectance, 0.12 (Wilcoxon: p <

0.0001) when ranked by nestedness and 0.13 (Wilcoxon: p < 0.0001) when ranked by modularity. These results show that, for a given level of connectance, nestedness or modularity, networks that appear similar at the macro-scale can be composed of different indirect interaction structures: meso-scale structural signatures based on motifs generally showed significantly more dissimilarity than macro-scale measures of structure (Fig. 6). Specifically, for connectance, nestedness and modularity as the ranking variable respectively, the motif framework captured 69%, 62% and 57% more variation in indirect interactions on average than traditional whole-network indices.
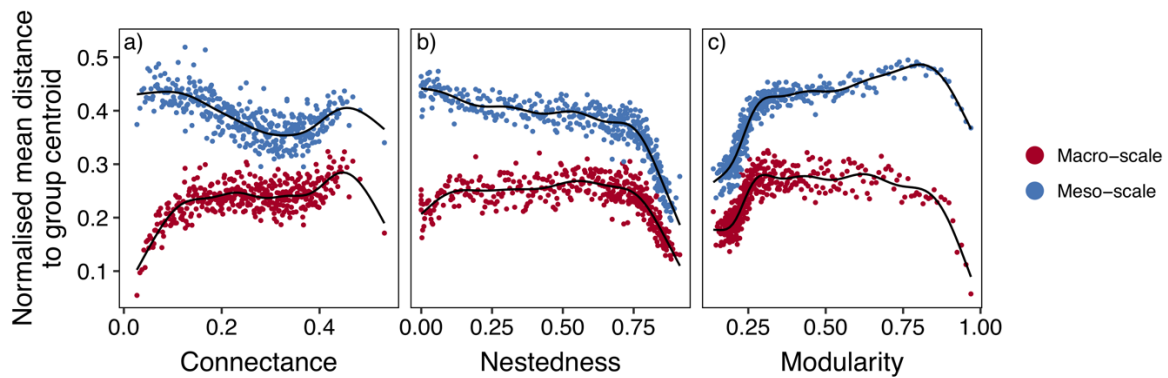

Figure 6: Network variation (normalised mean distance to group centroid) against mean connectance (a), nestedness (b) and modularity (c) for all networks. Points represent subsets of 50 networks.

Nestedness, modularity and motif frequencies can vary with network size and connectance. To control for this effect, we repeated the above analysis using a null model approach, where nestedness, modularity and motif frequencies were expressed relative to a suite of null networks that preserve some basic network properties (Appendix 1). Our results and conclusions from this analysis were qualitatively identical. Overall, therefore, the increased variation in indirect interactions highlights the problem of describing network structure by its macro-scale properties alone.

Species level

The above analysis compares meso- and macro-scale approaches. However, it is also important to understand how meso-scale descriptors relate to species-level (micro-scale) indices. We therefore conducted a similar analysis to that described above, but compared motifs to species-level indices instead of network-level indices.

Using five complementary species-level indices (degree, closeness centrality, betweenness centrality, c-score and z-score) (Emer et al. 2016), we characterised the roles of 55,094 species across 1000 networks generated using the bipartite cooperation model. Degree, closeness centrality and betweenness centrality were calculated using the 'bipartite' R package (Dormann et al. 2009), while c- and z-scores were calculated using the 'rnetcarto' R package (Doulcier and Stouffer 2015). We ranked species according to each micro-scale index (degree, closeness centrality, betweenness centrality, c-score and z-score) in turn and divided species into subsets of 50 according to this ranking. For example, when ranking species by closeness centrality each subset would contain 50 species with similar values of closeness centrality.

For each species in each subset, we characterised its role at both the micro- and meso-scale. At the micro-scale, species roles were described with a vector containing four of degree, closeness centrality, betweenness centrality, c-score and z-score (having excluded the metric used for ranking species). For example, if degree was used as the ranking property, the micro-scale structure vector would include closeness centrality, betweenness centrality, c-score and z-score. At the meso-scale, species roles were characterised with the vectors describing the frequency with which species occur in all unique positions across motifs containing between three and six species. Species motif roles were calculated using the 'bmotif' R package (Simmons et al. 2018).

For each subset, we calculated pairwise distances between all species role vectors at both the micro- and meso-scales, using the 'correlation' distance measure in the 'rdist' function from the 'rdist' R package (Blader 2018). Distances were normalised by dividing by the maximum distance between any two species, giving values between 0 (for identical species) and 1 (for completely different species).

Finally, for each subset at each scale, we calculated the distance between each species' micro- or meso-scale role vector and the subset centroid representing the 'typical' structure for each subset using the 'betadisper' function from the 'vegan' R package (Oksanen et al. 2016).

We repeated this procedure using each of degree, closeness centrality, betweenness centrality, *c*-score and *z*-score as the ranking variable, to give five views of the variability of species roles at micro- and meso-scales We then used paired Wilcoxon signed-rank tests to compare the differences in variation between macro- and meso-scale subsets. The median paired differences in variation between micro- and meso-scale subsets were 0.16 when ranked by degree (Wilcoxon: $p < 0.0001$), 0.32 when ranked by closeness centrality (Wilcoxon: $p < 0.0001$), 0.34 when ranked by betweenness centrality (Wilcoxon: $p < 0.0001$), 0.44 when ranked by *c*-score (Wilcoxon: $p < 0.0001$) and 0.40 when ranked by *z*-score (Wilcoxon: $p < 0.0001$). These results show that, for a given level of degree, closeness centrality, betweenness centrality, *c*-score or *z*-score, species that appear to have similar roles at the micro-scale can be have different indirect interaction structures: meso-scale structural signatures based on motifs generally showed significantly more dissimilarity than micro-scale measures of structure (Fig. 7). Specifically, for degree, closeness centrality, betweenness centrality, *c*-score or *z*-score as the ranking variable respectively, the motif framework captured 83%, 406%, 465%, 1076% and 610% more variation on average than traditional species-level indices. Overall, therefore, the increased variation in indirect interactions highlights the problem of describing species roles by their micro-scale properties alone.
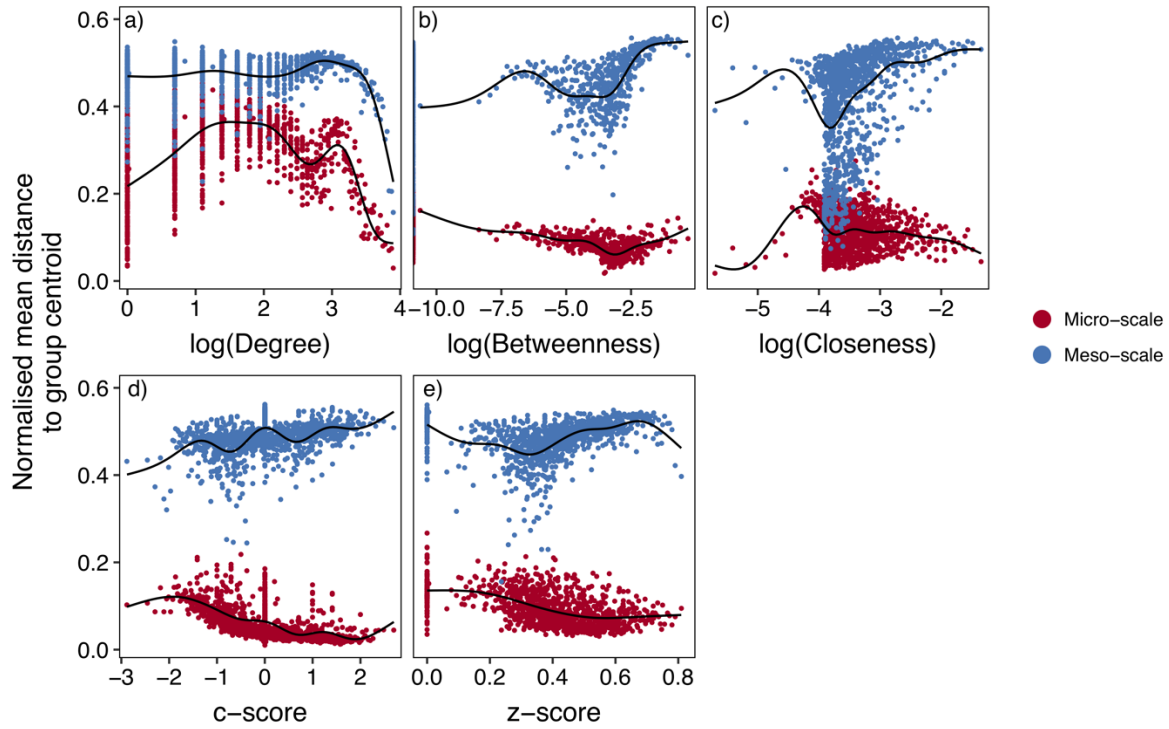
Figure 7: Species role variation (normalised mean distance to group centroid) against mean degree (a), betweenness (b), closeness (c), c-score (d) and z-score (e) for X species across 1000 networks. Points represent subsets of 50 species.

Robustness to sampling effort

A major challenge among studies of ecological networks is that completely sampling a web of interactions is difficult: both species and their interactions can be missed, especially if they are rare or hard to detect (Jordano 2016). Many network and species-level indices are sensitive to sampling effects (Dorado et al. 2011, Rivera-Hutinel et al. 2012, Fründ et al. 2016). To assess the sensitivity of the motif framework to sampling biases, we simulated different levels of sampling effort on 40 empirical, quantitative pollination and seed dispersal networks obtained from the Web of Life repository (www.web-of-life.es; Supplementary Table 1). The inclusion criteria we used for selecting these networks are as follows. We started with all 119 quantitative pollination and seed dispersal networks currently available from the Web of Life repository. We first removed networks which were part of a timeseries of networks from the same location (rather than independent datasets) to avoid pseudoreplication. This reduced the dataset to 47 networks. We then only included networks with more than five species in each level of the bipartite network to ensure that five-species motifs could be

calculated for all datasets. This resulted in a dataset of 42 networks. Finally, we removed two very large networks for which it was computationally infeasible to carry out our analysis, leaving a final dataset of 40 networks. In field studies, plant-animal interaction networks are usually sampled by observing plants and recording the animals that visit them (Jordano 2016). To replicate this process *in silico*, we sampled networks in two stages (de Aguiar et al. 2017). First, we sampled a proportion, p, of plant species to simulate the likely scenario that not all plant species are observed when surveying a site (Jordano 2016). Species with more partners had a higher chance of being sampled, as generalist species tend to be more abundant in mutualistic communities (Fort *et al.* 2016; though see Supplementary Fig. 6 for results where species had a random probability of being selected). Second, for each selected plant species, we sampled a proportion, q, of their interactions to simulate the fact that not all interactions are observed (Dormann et al. 2009, Poisot et al. 2012); stronger interactions, corresponding to more frequent visits between plants and animals, had a higher probability of being sampled (de Aguiar et al. 2017). We repeated this process for different values of p and q between 0.5 and 1, performing 1000 randomisations at each p-q combination. This choice of threshold was partly dictated by the constraints of the dataset: when more than 50% of species and 50% of links were removed, most networks became disconnected and trivial, meaning that motifs of larger size classes often did not occur. However, a 50% sampling threshold is also a realistic one for many networks: in several studies that have measured the sampling completeness of networks, over 50% of interactions and species are often recorded. In the first study to propose estimating ecological network sampling completeness, Chacoff et al (2012) detected 80% of pollinator species, 55% of interactions, and an average of 61% of plant interaction partners in a desert plant–pollinator network. Subsequent studies have found similar results. For example, Devoto et al (2012) detected an average of 57% of interactions in their plant-pollinator networks, while Traveset et al (2015) detected an average of 68% of bird interactions in bird-flower visitation networks. Therefore while we are sure some sampled networks contain less than 50% of the 'true' number of interactions or species, this does not seem to be such a widespread phenomenon as to make our choice of a 50% threshold unrealistic or unhelpful. Instead, many networks are estimated to contain more than 50% of interactions and species.

We decomposed each sampled network into its constituent motifs and recorded each network's motif structural signature and the motif role signatures of each species. We then measured R^2 between the network structural signature or species role signature of the sampled network and those of the corresponding 'true' network containing all species and interactions. Further details of the simulations are given in Appendix 3. We found that both network structural signatures and species role signatures were remarkably robust to sampling effects. Even when only 50% of plant species and 50% of their interactions were sampled, the mean R^2 between the sampled and 'true' network signatures was 0.87 (Fig. 8a). At this same level of sampling, the mean R^2 between sampled and 'true' species role signatures was 0.93 (Fig. 8b). That motifs appear robust to sampling effects is encouraging for future studies adopting this framework.
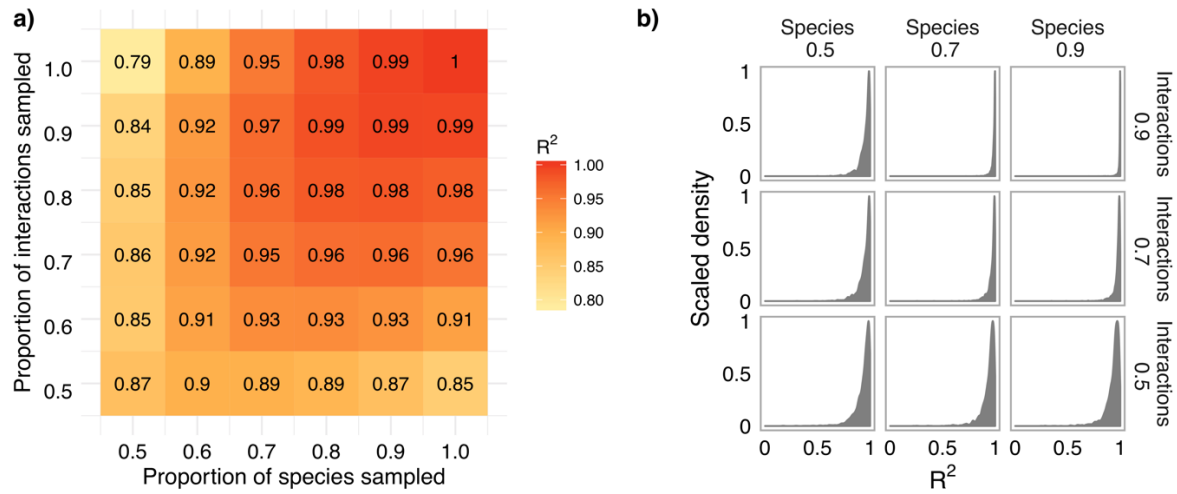

Figure 8: Results of simulations assessing the sensitivity of the motif framework to variation in sampling effort. (a) The mean R^2 between the network structural signatures of the sampled networks and the structural signatures of their corresponding 'true' networks for different levels of species and interaction removal. (b) Distribution of mean R^2 between species role signatures in sampled networks and species role signatures in their corresponding 'true' networks for different levels of species and interaction removal.

Indirect interactions in empirical plant-pollinator networks

Species roles

Here we present a case study comparing the roles of two pollinator species over time. Data was from four mountaintop plant-pollinator communities in the Seychelles, sampled over the flowering season in eight consecutive months between September 2012 and April 2013 (Kaiser-Bunbury *et al.* 2017; Supplementary Table 2). Restoration by removal of exotic plants from these communities resulted in pollinator species becoming more generalised. This pattern was driven largely by two abundant, highly generalist pollinator species, one native (*Lasioglossum mahense*) and one non-native (*Apis mellifera*) (Kaiser-Bunbury et al. 2017). These two abundant, super-generalist species could have similar strategies for partner selection and therefore play similar roles in the community. This is the result found in the original study: both species had similar levels of specialisation (quantified using the specialisation index *d'*, which measures the extent to which species deviate from a random sampling of available partners (Blüthgen et al. 2006)): 0.17 ± 0.10 and 0.22 ± 0.18 for *Lasioglossum mahense* and *Apis mellifera* respectively. Alternatively, two abundant, super-generalists could minimise competition by exploiting different areas of 'interaction niche space' and therefore have different roles. To test these alternatives, we calculated the motif role signatures of both species at each site in each monthly network, giving a detailed view of how each species is embedded in the community over time. We used permutational multivariate analysis of variance (PERMANOVA), stratified by site, to assess if there are significant differences between the roles the two species play in the four communities. PERMANOVA is similar to ANOVA but compares multivariate differences within and between groups without assuming normality or Euclidean distances (Anderson 2001). We used Bray-Curtis distance as the dissimilarity measure, as it is suitable for a variety of ecological data, including motifs (Faith et al. 1987, Anderson and Robinson 2003, Baker et al. 2015). PERMANOVAs were run with 10000 permutations.

The PERMANOVA analysis showed that *Lasioglossum mahense* and *Apis mellifera* had significantly different roles over time ($F_{1,62}$, p = 0.0496), exploiting different areas of

interaction niche space. This means that, while Kaiser-Bunbury *et al.* (2017) used the species-level metric d' to show that both species were super-generalists, a motif approach reveals that they are generalist in different ways. This result is visualised in Fig. 9. More positive values of the first NMDS axis are associated with motif positions where more specialist pollinators compete with generalist pollinators for a shared plant resource, while negative values are associated with positions where generalist pollinators visit specialist plants with little competition. More positive values of the second NMDS axis are associated with positions where pollinators visit plants which are also visited by generalist species; negative values are associated with positions where pollinators visit plants which are also visited by specialist species. *Lassioglossum mahense* generally occupies higher values of both NMDS axes than *Apis mellifera*. Therefore, while both species are generalists, *Lassioglossum mahense* is in greater competition with generalist pollinators than *Apis mellifera* which visits more specialist plants and competes with more specialist pollinators. These differences in indirect interactions are essential for understanding the ecology of these two species and are missed using the d' index alone. All PERMANOVA tests and NMDS analyses were conducted in the R package vegan (Oksanen et al. 2016).
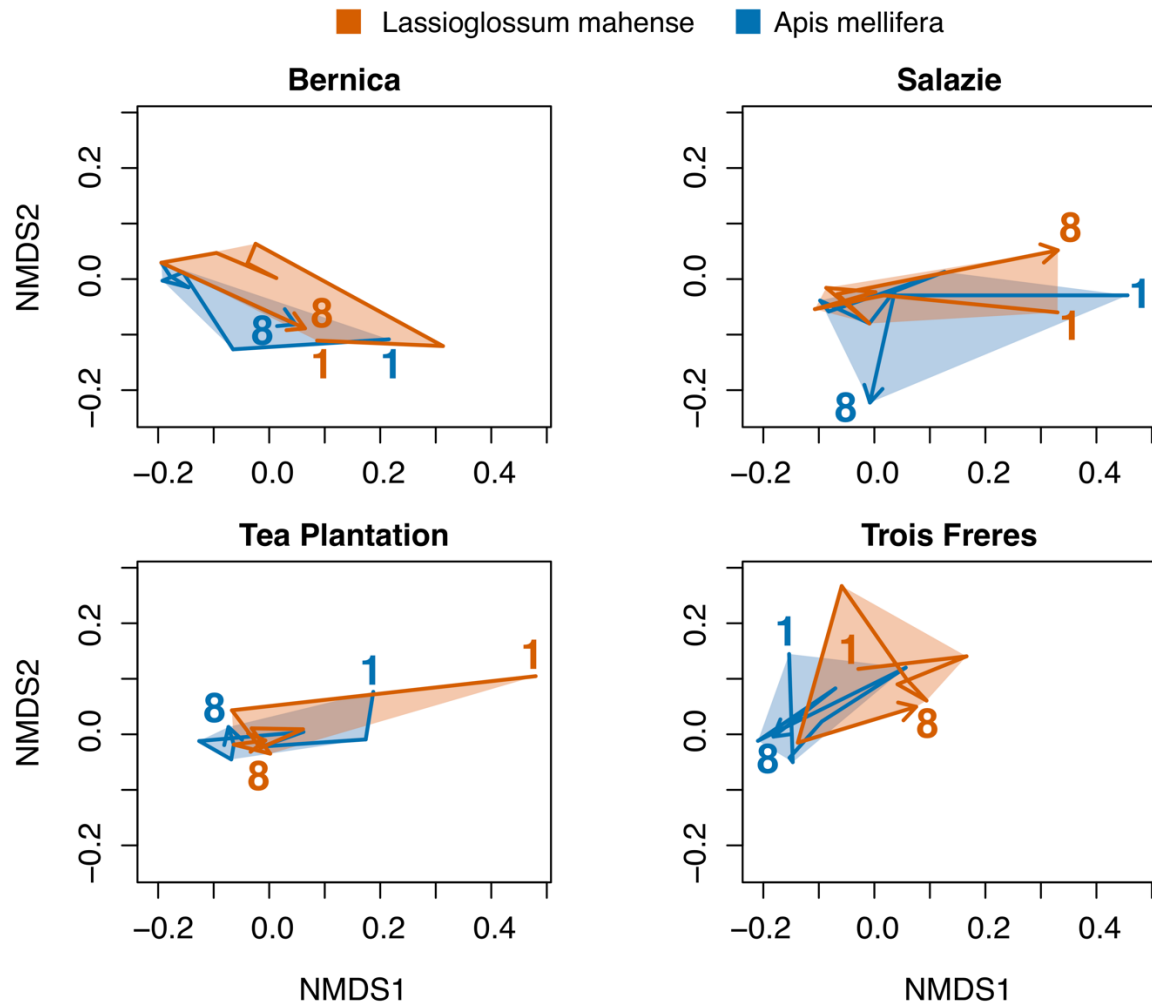
Figure 9: The movement of Lasioglossum mahense and Apis mellifera through interaction niche space over eight months in four sites (Bernica, Salazie, Tea Plantation and Trois Freres). Each vertex represents the role of a species in a monthly network. Numbers '1' and '8' indicate the first and last sampling month, respectively. Shaded polygons are convex hulls containing the vertices of each species.

Motif over- and under-representation

To determine whether particular motifs occur more or less than would be expected by chance, it is possible to compare empirical motif frequencies to those produced by a null model (Milo et al. 2002). We compared the motif distributions of 122 empirical pollination networks from the Web of Life repository (web-of-life.es) to 100 randomisations of a null model where the probability of a link occurring between plant *i* and animal *j* is equal to the mean of the normalised degree of species *i* and *j* (Bascompte et al 2003). In other words, the probability of a cell in the interaction matrix being occupied is equal to the mean of the occupancy of that cell's row and column (Bascompte et al 2003). Ecologically, this means that the likelihood of two species

interacting is proportional to their level of generalisation (degree). This null model is used to determine the significance of structural properties in pollination networks as it allows structure to be determined beyond that which results from the degree distribution alone (Bascompte et al 2003). The level of over- or under-representation of motif i was expressed as a z-score (Milo et al 2002; Stouffer et al 2007):

$$z_i = \frac{N_{Emp} - \overline{N_{Rand}}}{\sigma_{N_{Rand}}}$$

where $\overline{N_{Rand}}$ and $\sigma_{N_{Rand}}$ are the mean and standard deviation of the randomised motif counts, respectively.

Results are shown in Figure 10. Motifs 2, 7, 13, 14, 15, 16, 17 occur significantly more than random in the majority of networks, while motifs 3 and 10 occur significantly less than random in the majority of networks (though only in just above half of networks for motif 10). The over-represented motifs all involve one or two plants interacting with between two and four pollinators. This asymmetry in the number of species in each level suggests that there are a large number of indirect interactions between animal pollinators mediated through a smaller number of plants. Conversely, under-represented motifs 3 and 10 involve two or three plants interacting with one or two pollinators, respectively. This asymmetry suggests that indirect interactions between plants, mediated by one or two pollinators, are less common than would be expected from degree distribution alone. The structure of under-represented motif 10 is particularly interesting. Of all the motifs containing three plants and two pollinators, motif 10 has the lowest possible connectance. Other motifs with the same number of species in each level, but with more dense patterns of connections, are not under-represented. This implies that, when plants and pollinators form asymmetric local structures with more plants than pollinators, these are tightly connected with a high degree of cohesion. These more tightly connected motifs will also have indirect effects which are harder to predict because there are many possible pathways through which the mutualistic benefit could flow. This is in contrast to the under-represented motif 10, which has a relatively simple and predictable structure. This finding matches results by Carvalheiro et al (2014), who found that indirect interactions between plants via shared pollinators are more frequent among abundant plants that tend to be more generalised and are thus more likely to be

involved in more densely connected motifs. Overall, therefore, we conclude that pollination networks have high levels of indirect interactions between pollinators mediated by a small number of plants, and fewer indirect interactions between plants mediated by a small number of pollinators. However, when indirect interactions between plants mediated by pollinators do occur, these are in tightly connected clusters that may have complex, and possibly hard to predict, dynamics.
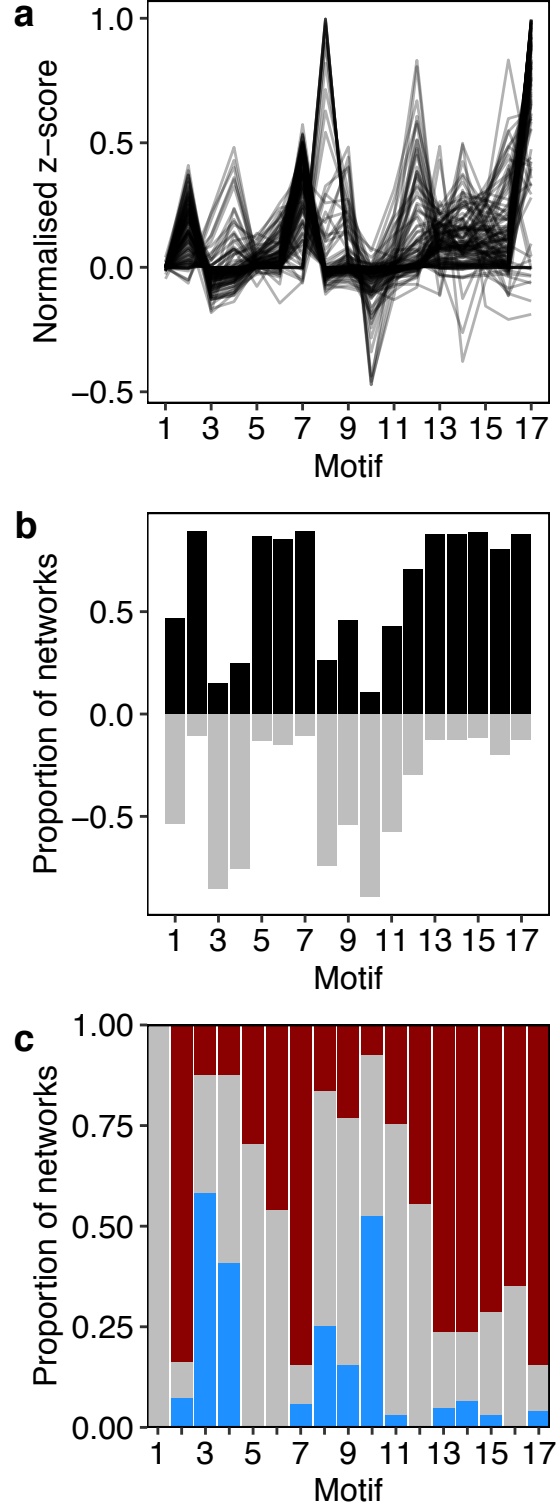
Figure 10: Patterns of motif over- and under-representation in pollination networks relative to a null model (Bascompte et al 2003). (a) Each line shows the pattern of motif over- and under-representation for a single network. z-scores were normalised following $P_i = z_i / \sqrt{\sum_j z_j^2}$ where P_i is the normalised profile of network i, z is a z-score and j is an index over motifs (Stouffer et al 2007). (b) For each motif, the figure shows the proportion of networks which had more (black) or less (grey) of that motif than

the null model. (c) For each motif, the figure shows the proportion of networks which had significantly (± 1.96 standard deviations) more (red) or less (blue) of the motif than the null model. Grey indicates motif frequencies were not significantly different to the null model.

Potential applications

Characterising the structure of species interaction networks is a key component of many areas of ecological research, such as robustness to extinctions (Kaiser-Bunbury et al. 2010), ecosystem functioning (Coux et al. 2016) and macroecology (Araújo and Luoto 2007, Staniczenko et al. 2017). It is essential to incorporate indirect interactions into all these analyses, suggesting that the framework presented here has wide applicability to a diverse range of topics, systems and interaction types. In particular, we suggest the motif framework may be beneficial for studies where the scale of interest is at the species level, such as examining how invasive species integrate into communities (Vilà et al. 2009, Stouffer et al. 2014); or when within-network phenomena are the focus, such as studies of rewiring and network variability over time (Olesen et al. 2008, Kaiser-Bunbury et al. 2010). As indirect interactions are likely to be of increased importance when investigating these types of questions, we caution against using only conventional network and species-level indices which can mask detail about these interactions. Adopting motif descriptions of network structure also opens up new ways to answer a diverse range of questions such as those concerning competitive exclusion, species packing and functional redundancy (Blonder et al. 2014). For example, does interaction distinctiveness correlate with functional distinctiveness? Do species have overlapping or disjoint roles? Do indirect interaction structures vary over space and time?

We have shown how easy it is for similar-looking networks to be composed of very dissimilar parts. We therefore expect that much valuable information on indirect interactions has been ignored, intentionally or unintentionally. This realisation yields a series of hypotheses about how our understanding of bipartite ecological networks may change if indirect interactions were incorporated. For example, uncovering indirect interactions could revise our understanding of how invariant network structure is across space and time. Several studies have shown that network structure is relatively stable in the presence of temporal and spatial turnover in species and interaction identity

(Petanidou et al. 2008, Dáttilo et al. 2013). However, these studies have considered only global descriptors of structure that likely mask meso-scale structural variation in indirect interactions. We anticipate that, if indirect interactions were considered, network structure may not be as invariant to compositional turnover as previously identified.

We also hypothesise that incorporating indirect interactions may improve predictions of network structure and our understanding of the mechanisms underpinning network assembly. Understanding the processes that govern the formation of species interactions is essential for predicting the structure of novel communities under global changes (Eklöf et al. 2013). However, current attempts often involve assessing how well different mechanisms, such as neutral effects, morphological matching and phenological overlap, predict different network and species-level indices (Vázquez et al. 2009, Verdú and Valiente-Banuet 2011, Sayago et al. 2013, Vizentin-Bugoni et al. 2014). For example, Vázquez et al. (2009) show that data on abundance and phenology can accurately predict network-level indices such as connectance and nestedness. This approach is problematic because many network and species-level indices are insensitive to changes in network topology. Models can therefore accurately predict index values while incorrectly predicting pairwise interactions (Fox 2006, Olito and Fox 2015). Such models may be of limited utility in helping to understand the processes underlying network structure. To improve models, structural signatures based on motifs could be used instead as a benchmark of predictive performance. As the motif framework is much more sensitive to changes in network topology than network and species-level indices are, it would be harder for models to accurately predict a structural signature while incorrectly predicting pairwise interactions. We therefore expect that adopting the motif framework could change both our understanding of the processes governing interactions and improve our ability to predict novel communities.

Finally, incorporating indirect interactions could improve understanding of the functional consequences of community structure (Thompson et al. 2012, Poisot et al. 2013). For example, pollinators with more distinct traits (traits furthest from the community average) tend to have fewer interaction partners (Coux et al. 2016). One

hypothesis for this pattern is a trade-off between reducing competition with other pollinators by having original traits and needing to retain interaction partners (Vamosi et al. 2014, Coux et al. 2016). The motif framework could explicitly test this hypothesis by assessing whether functionally original species appear primarily in motif positions where there is low competition between pollinators.

Limitations and challenges

Currently, bipartite motifs have only been used for qualitative networks, where interactions are present or absent. This contrasts with quantitative networks where interactions are weighted in proportion to their relative strength. Using only qualitative information, rare species or interactions can exert a disproportionate influence on network metrics (Banašek-Richter et al. 2004). The loss of detail on indirect interactions resulting from the use of conventional network and species-level indices is, however, likely to be equal to or greater than the loss of information resulting from using qualitative instead of quantitative networks. As shown in the example above, d', an index which uses quantitative information on interaction weights, could not distinguish the roles of *Lasioglossum mahense* and *Apis mellifera*, while qualitative motifs could. We also note that qualitative versions of many conventional metrics (such as connectance and degree) are frequently used to characterise quantitative networks instead of their weighted counterparts. While methods to enumerate weighted motifs are being developed (Bramon Mora et al. 2018), there are a number of tractable methods to incorporate quantitative information in motif analyses. For example, interactions within motifs can be classified as 'strong' or 'weak' depending on whether a given interaction's strength is greater or lesser, respectively, than the median strength (Rodríguez-Rodríguez et al. 2017). Alternatively, a suite of qualitative networks can be assembled by sampling a quantitative network in proportion to the strength of each interaction (Baker et al. 2015). This creates an ensemble of qualitative resampled networks where stronger interactions appear more frequently than weaker ones. Analyses can then be repeated using each of the resampled networks as input. This creates a distribution of p-values or effect sizes associated with a particular analysis,

which can then be compared to the results obtained using a binary version of the original quantitative network.

Finally, it is important to note that, like network and species-level indices, the motif framework also results in a loss of information when characterising network structure: transforming a network into a structural signature or ensemble of species' role signatures is unique, while the reverse is not. Some loss of information is inevitable so long as we must summarize networks in order to analyse them. However, motifs are substantially less interaction inelastic than network and species-level indices, and therein lies their advantage.

Concluding remarks

Indirect interactions are a widespread and important component of ecological communities, essential for understanding species roles and the structure of biotic interactions. However, to date the dominant paradigm has been to describe community structure using a wide variety of network and species-level indices that can mask indirect interaction detail. Here we have presented a framework that conceptualises networks as a series of component building blocks or 'motifs'. By thinking of networks in this way, we have shown that potential indirect interactions can be explicitly identified and quantified. We do not advocate widespread abandonment of network and species-level indices, but instead aim to raise awareness of their limitations. We hope that motifs will exist alongside network and species-level indices to form the basis of a new paradigm among studies of bipartite ecological networks. Given the increasingly large amount of ecological network data available, and the rapid growth in computational capacity to analyse these data, there is now a timely opportunity to make motifs a standard part of the analytical toolkit for studying bipartite systems. Such an approach could enable novel perspectives and insights into the ecology and evolution of many important communities.

References

Almeida-Neto, M. et al. 2008. A consistent metric for nestedness analysis in ecological

systems: Reconciling concept and measurement. - Oikos 117: 1227–1239.

Anderson, M. J. 2001. A new method for non parametric multivariate analysis of variance. - Austral Ecol. 26: 32–46.

Anderson, M. J. and Robinson, J. 2003. Generalized discriminant analysis based on distances. - Aust. New Zeal. J. Stat. 45: 301–318.

Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. - Glob. Ecol. Biogeogr. 16: 743–753.

Bailey, J. K. and Whitham, T. G. 2007. Biodiversity is related to indirect interactions among species of large effect. - In: Ohgushi, T. et al. (eds), Ecological communities: plant mediation in indirect interaction webs. Cambridge University Press, pp. 306–328.

Baker, N. J. et al. 2015. Species' roles in food webs show fidelity across a highly variable oak forest. - Ecography (Cop.). 38: 130–139.

Banašek-Richter, C. et al. 2004. Sampling effects and the robustness of quantitative and qualitative food-web descriptors. - J. Theor. Biol. 226: 23–32.

Bascompte, J., Jordano, P., Melián, C.J. and Olesen, J.M., 2003. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences, 100(16), pp.9383-9387.

Bascompte, J. and Jordano, P. 2007. Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. - Annu. Rev. Ecol. Evol. Syst. 38: 567–593.

Bascompte, J. and Stouffer, D. B. 2009. The assembly and disassembly of ecological networks. - Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364: 1781–7.

Bascompte, J. et al. 2006. Asymmetric coevolutionary networks facilitate biodiversity maintenance. - Science (80-.). 312: 431–433.

Bastolla, U. et al. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. - Nature 458: 1018–1020.

Beckett, S. J. 2016. Improved community detection in weighted bipartite networks. - R. Soc. Open Sci. in press.

Blader, N. 2018. rdist: Calculate Pairwise Distances. in press.

Blonder, B. et al. 2014. The n-dimensional hypervolume. - Glob. Ecol. Biogeogr. 23: 595–609.

Blüthgen, N. et al. 2006. Measuring specialization in species interaction networks. -

BMC Ecol. 6: 9.

Borrelli, J. J. 2015. Selection against instability: Stable subgraphs are most frequent in empirical food webs. - Oikos 124: 1583–1588.

Bramon Mora, B. et al. 2018. pymfinder: a tool for the motif analysis of binary and quantitative complex networks. - bioRxiv in press.

Camacho, J. et al. 2007. Quantitative analysis of the local structure of food webs. - J. Theor. Biol. 246: 260–268.

Campbell, D. R. 1985. Pollinator Sharing and Seed Set of Stellaria pubera: Competition for Pollination. - Ecology 66: 544–553.

Carvalheiro, L. G. et al. 2014. The potential for indirect effects between co-flowering plants via shared pollinators depends on resource abundance, accessibility and relatedness. - Ecol. Lett. 17: 1389–1399.

Chacoff, N. P. et al. 2012. Evaluating sampling completeness in a desert plant-pollinator network. - J. Anim. Ecol. 81: 190–200.

Chittka, L. and Schürkens, S. 2001. Successful invasion of a floral market: An exotic Asian plant has moved in on Europe's river-banks by bribing pollinators. - Nature 411: 653.

Cirtwill, A. R. and Stouffer, D. B. 2015. Concomitant predation on parasites is highly variable but constrains the ways in which parasites contribute to food web structure. - J. Anim. Ecol. 84: 734–744.

Cirtwill, A. R. et al. 2018. Between-year changes in community composition shape species' roles in an Arctic plant--pollinator network. - Oikos in press.

Coux, C. et al. 2016. Linking species functional roles to their network roles. - Ecol. Lett. 19: 762–770.

Dáttilo, W. et al. 2013. Spatial structure of ant-plant mutualistic networks. - Oikos 122: 1643–1648.

de Aguiar, M. A. M. et al. 2017. Revealing biases in the sampling of ecological interaction networks. - arXiv: 1708.01242.

Devoto, M. et al. 2012. Understanding and planning ecological restoration of plant-pollinator networks. - Ecol. Lett. 15: 319–328.

Dorado, J. et al. 2011. Rareness and specialization in plant-pollinator networks. - Ecology 92: 19–25.

Dormann, C. F. et al. 2009. Indices, Graphs and Null Models: Analyzing Bipartite
    Ecological Networks. - Open Ecol. J. 2: 7–24.

Doulcier, G. and Stouffer, D. 2015. Rnetcarto: Fast Network Modularity and Roles
    Computation by Simulated Annealing. R package version 0.2.4. in press.

Dupont, Y. L. et al. 2014. Spatial structure of an individual-based plant-pollinator
    network. - Oikos 123: 1301–1310.

Eklöf, A. et al. 2013. The dimensionality of ecological networks. - Ecol. Lett. 16: 577–
    583.

Emer, C. et al. 2016. Species roles in plant–pollinator communities are conserved
    across native and alien ranges. - Divers. Distrib. 22: 841–852.

Faith, D. P. et al. 1987. Compositional dissimilarity as a robust measure of ecological
    distance. - Vegetatio 69: 57–68.

Flanagan, R. J. et al. 2010. Increased relative abundance of an invasive competitor for
    pollination, Lythrum salicaria, reduces seed number in Mimulus ringens. -
    Oecologia 164: 445–454.

Fontaine, C. et al. 2011. The ecological and evolutionary implications of merging
    different types of networks. - Ecol. Lett. 14: 1170–1181.

Fort, H. et al. 2016. Abundance and generalisation in mutualistic networks: Solving the
    chicken-and-egg dilemma. - Ecol. Lett. 19: 4–11.

Fox, J. W. 2006. Current food web models cannot explain the overall topological
    structure of observed food webs. - Oikos 115: 97–109.

Frank van Veen, F. J. et al. 2006. Apparent competition, quantitative food webs, and
    the structure of phytophagous insect communities. - Annu. Rev. Entomol. 51: 187–
    208.

Fründ, J. et al. 2016. Sampling bias is a challenge for quantifying specialization and
    network structure: Lessons from a quantitative niche model. - Oikos 125: 502–513.

Ghazoul, J. 2006. Floral diversity and the facilitation of pollination. - J. Ecol. 94: 295–
    304.

Guimarães, P. R. et al. 2017. Indirect effects drive coevolution in mutualistic networks.
    - Nature 550: 511–514.

Hay, M. E. 1986. Associational Plant Defenses and the Maintenance of Species
    Diversity: Turning Competitors Into Accomplices. - Am. Nat. 128: 617–641.

Hochkirch, A. et al. 2012. Conspecific flowers of sinapis arvensis are stronger competitors for pollinators than those of the invasive weed bunias orientalis. - Naturwissenschaften 99: 217–224.

Jordán, F. et al. 2007. Quantifying positional importance in food webs: A comparison of centrality indices. - Ecol. Modell. 205: 270–275.

Jordano, P. 2016. Sampling networks of ecological interactions. - Funct. Ecol. 30: 1883–1893.

Kaiser-Bunbury, C. N. et al. 2010. The robustness of pollination networks to the loss of species and interactions: A quantitative approach incorporating pollinator behaviour. - Ecol. Lett. 13: 442–452.

Kaiser-Bunbury, C. N. et al. 2017. Ecosystem restoration strengthens pollination network resilience and function. - Nature 542: 223–227.

Kashtan, N. et al. 2004. Topological generalizations of network motifs. - Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 70: 12.

Kondoh, M. 2008. Building trophic modules into a persistent food web. - Proc. Natl. Acad. Sci. 105: 16631–16635.

Liao, K. et al. 2011. The presence of co-flowering species facilitates reproductive success of Pedicularis monbeigiana (Orobanchaceae) through variation in bumble-bee foraging behaviour. - Ann. Bot. 108: 877–884.

Martínez, D. et al. 2014. Consistency and reciprocity of indirect interactions between tree species mediated by frugivorous birds. - Oikos 123: 414–422.

Menke, S. et al. 2012. Plant-frugivore networks are less specialized and more robust at forest-farmland edges than in the interior of a tropical forest. - Oikos 121: 1553–1566.

Milo, R. et al. 2002. Network motifs: simple building blocks of complex networks. - Science (80-.). 298: 824–827.

Mitchell, R. J. et al. 2009. New frontiers in competition for pollination. - Ann. Bot. 103: 1403–1413.

Moeller, D. A. 2004. Facilitative interactions among plants via shared pollinators. - Ecology 85: 3289–3301.

Morales, C. L. and Traveset, A. 2009. A meta-analysis of impacts of alien vs. native plants on pollinator visitation and reproductive success of co-flowering native

plants. - Ecol. Lett. 12: 716–728.

Morris, R. J. et al. 2004. Experimental evidence for apparent competition in a tropical forest food web. - Nature 428: 310–313.

Morris, R. J. et al. 2014. Antagonistic interaction networks are structured independently of latitude and host guild. - Ecol. Lett. 17: 340–349.

Oksanen, J. et al. 2016. vegan: Community Ecology Package. R package version 2.4-0. in press.

Olesen, J. M. et al. 2008. Temporal dynamics in a pollination network. - Ecology 89: 1573–1582.

Olito, C. and Fox, J. W. 2015. Species traits and abundances predict metrics of plant-pollinator network structure, but not pairwise interactions. - Oikos 124: 428–436.

Petanidou, T. et al. 2008. Long-term observation of a pollination network: Fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization. - Ecol. Lett. 11: 564–575.

Pfister, C. A. and Hay, M. E. 1988. Associational plant refuges: convergent patterns in marine and terrestrial communities result from differing mechanisms. - Oecologia 77: 118–129.

Poisot, T. et al. 2012. A comparative study of ecological specialization estimators. - Methods Ecol. Evol. 3: 537–544.

Poisot, T. et al. 2013. Trophic complementarity drives the biodiversity-ecosystem functioning relationship in food webs. - Ecol. Lett. 16: 853–861.

R Core Team 2015. R: A language and environment for statistical computing. in press.

Rathcke, B. 1983. Competition and facilitation among plants for pollination. - In: Pollination biology. pp. 305–329.

Rivera-Hutinel, A. et al. 2012. Effects of sampling completeness on the structure of plant-pollinator networks. - Ecology 93: 1593–1603.

Rodríguez-Rodríguez, M. C. et al. 2017. Functional consequences of plant-animal interactions along the mutualism-antagonism gradient. - Ecology 98: 1266–1276.

Runquist, R. B. and Stanton, M. L. 2013. Asymmetric and frequency-dependent pollinator-mediated interactions may influence competitive displacement in two vernal pool plants. - Ecol. Lett. 16: 183–190.

Saavedra, S. et al. 2009. A simple model of bipartite cooperation for ecological and

organizational networks. - Nature 457: 463–466.

Saracco, F. et al. 2017. Inferring monopartite projections of bipartite networks: An entropy-based approach. - New J. Phys. 19: 053022.

Sayago, R. et al. 2013. Evaluating factors that predict the structure of a commensalistic epiphyte-phorophyte network. - Proc. R. Soc. B Biol. Sci. 280: 20122821–20122821.

Sieber, Y. et al. 2011. Do alpine plants facilitate each other's pollination? Experiments at a small spatial scale. - Acta Oecologica 37: 369–374.

Simmons, B. I. et al. 2018. bmotif: a package for counting motifs in bipartite networks. - bioRxiv in press.

Sotomayor, D. A. and Lortie, C. J. 2015. Indirect interactions in terrestrial plant communities: Emerging patterns and research gaps. - Ecosphere in press.

Staniczenko, P. P. A. et al. 2017. Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. - Ecol. Lett. 20: 693–707.

Stouffer, D.B., Camacho, J., Jiang, W. and Nunes Amaral, L.A., 2007. Evidence for the existence of a robust pattern of prey selection in food webs. Proceedings of the Royal Society B: Biological Sciences, 274(1621), pp.1931-1940.

Stouffer, D. B. and Bascompte, J. 2010. Understanding food-web persistence from local to global scales. - Ecol. Lett. 13: 154–161.

Stouffer, D. B. and Bascompte, J. 2011. Compartmentalization increases food-web persistence. - Proc. Natl. Acad. Sci. 108: 3648–3652.

Stouffer, D. B. et al. 2012. Evolutionary Conservation of Species' Roles in Food Webs. - Science (80-.). 335: 1489–1492.

Stouffer, D. B. et al. 2014. How exotic plants integrate into pollination networks. - J. Ecol. 102: 1442–1450.

Strauss, S. Y. 1991. Indirect Effects in Community Ecology - Their Definition, Study and Importance. - Trends Ecol. Evol. 6: 206–210.

Tack, A. J. M. et al. 2011. Can we predict indirect interactions from quantitative food webs? - An experimental approach. - J. Anim. Ecol. 80: 108–118.

Temeles, E. J. et al. 2016. Pollinator competition as a driver of floral divergence: An experimental test. - PLoS One in press.

Thompson, R. M. et al. 2012. Food webs: Reconciling the structure and function of

biodiversity. - Trends Ecol. Evol. 27: 689–697.

Traveset, A. et al. 2015. Bird-flower visitation networks in the Galápagos unveil a widespread interaction release. - Nat. Commun. in press.

Vamosi, J. C. et al. 2014. Pollinators visit related plant species across 29 plant-pollinator networks. - Ecol. Evol. 4: 2303–2315.

Vandermeer, J. et al. 1985. Indirect facilitation and mutualism. - In: Boucher, D. H. (ed), The Biology of Mutualism: Ecology and Evolution. Oxford University Press, pp. 326–343.

Vázquez, D. P. et al. 2009. Evaluating multiple determinants of the structure of plant-animal mutualistic networks. - Ecology 90: 2039–2046.

Vázquez, D. P. et al. 2015. A conceptual framework for studying the strength of plant-animal mutualistic interactions. - Ecol. Lett. 18: 385–400.

Verdú, M. and Valiente-Banuet, A. 2011. The relative contribution of abundance and phylogeny to the structure of plant facilitation networks. - Oikos 120: 1351–1356.

Vilà, M. et al. 2009. Invasive plant integration into native plant-pollinator networks across Europe. - Proc. R. Soc. B Biol. Sci. 276: 3887–3893.

Vizentin-Bugoni, J. et al. 2014. Processes entangling interactions in communities: forbidden links are more important than abundance in a hummingbird-plant network. - Proc. R. Soc. B Biol. Sci. 281: 20132397.

Waser, N. M. 1983. Competition for pollination and floral character differences among sympatric plant species: a review of the evidence. - In: Handbook of experimental pollination biology. pp. 277–293.

Wootton, J. T. 1994. The Nature and Consequences of Indirect Effects in Ecological Communities. - Annu. Rev. Ecol. Syst. 25: 443–466.

Wootton, J. T. 2002. Indirect effects in complex ecosystems: recent progress and future challenges. - J. Sea Res. 48: 157–172.

Ye, Z. M. et al. 2014. Competition and facilitation among plants for pollination: Can pollinator abundance shift the plant-plant interactions? - Plant Ecol. 215: 3–13.

Zhao, L. et al. 2018. Identifying compartments in ecological networks based on energy channels. - Ecol. Evol. 8: 309–318.

Zhou, T. et al. 2007. Bipartite network projection and personal recommendation. - Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 76: 046115.

5 | bmotif: a package for motif analyses of bipartite networks

This chapter was published as Simmons, B. I., Sweering, M. J. M., Schillinger, M., Dicks, L. V., Sutherland, W. J. and Di Clemente, R. (2019). bmotif: a package for motif analyses of bipartite networks. Methods in Ecology and Evolution. (in press).

Abstract

1. Bipartite networks are widely-used to represent a diverse range of species interactions, such as pollination, herbivory, parasitism and seed dispersal. The structure of these networks is usually characterised by calculating one or more indices that capture different aspects of network architecture. While these indices capture useful properties of networks, they are relatively insensitive to changes in network structure. Consequently, variation in ecologically-important interactions can be missed. Network motifs are a way to characterise network structure that is substantially more sensitive to changes in pairwise interactions, and is gaining in popularity. However, there is no software available in R, the most popular programming language among ecologists, for conducting motif analyses in bipartite networks. Similarly, no mathematical formalisation of bipartite motifs has been developed.

2. Here we introduce bmotif: a package for counting motifs, and species positions within motifs, in bipartite networks. Our code is primarily an R package, but we also provide MATLAB and Python code of the core functionality. The software is based on a mathematical framework where, for the first time, we derive formal expressions for motif frequencies and the frequencies with which species occur in different positions within motifs. This framework means that analyses with bmotif are fast, making motif methods compatible with the permutational approaches often used in network studies, such as null model analyses.

3. We describe the package and demonstrate how it can be used to conduct ecological analyses, using two examples of plant-pollinator networks. We first use motifs to examine the assembly and disassembly of an Arctic plant-pollinator community, and then use them to compare the roles of native and introduced plant species in an unrestored site in Mauritius.

4. bmotif will enable motif analyses of a wide range of bipartite ecological networks, allowing future research to characterise these complex networks without discarding important meso-scale structural detail.

Introduction

Bipartite networks have long been used to analyse complex systems (Diestel, 2000; Guillaume & Latapy, 2004; Newman, 2010). In ecology, they are widely used to study the structure of interactions between two groups of species, including plants and pollinators, hosts and parasitoids, and plants and seed dispersers. Studies of bipartite networks have yielded many new insights. For example, they have been used to uncover widespread nestedness in mutualistic communities (Bascompte, Jordano, Melián, & Olesen, 2003), and to show that community structure is stable despite turnover in species and interactions (Dáttilo, Guimarães, & Izzo, 2013). Such studies typically describe networks with one or more indices, such as connectance (the proportion of possible interactions which are realised), nestedness (the extent to which specialist species interact with subsets of the species generalist species interact with), degree (number of partners a species has) and d' (the extent to which a species' interactions deviate from a random sampling of its partners).

More recently, ecologists have been using bipartite motifs to characterise network structure. Bipartite motifs are subnetworks representing interactions between a given number of species (Fig. 1). These subnetworks can be considered the basic 'building blocks' of networks (Milo et al., 2002). Bipartite motifs are used in two main ways. First, to calculate how frequently different motifs occur in a network; Rodríguez-Rodríguez et al. (2017) used this approach to analyse the reproductive consequences of both

mutualistic and antagonistic interactions with animals. Second, to quantify species roles in a community by counting the frequency with which species occur in different positions within motifs; for example, Baker et al. (2015) used this method to demonstrate that species' roles in host-parasitoid networks are an intrinsic property of species. Moreover, studies of bipartite motifs in non-biological networks have been valuable to understand similarities in trade patterns (Saracco et al, 2015), gauge the effect of 2007 financial crisis on the world trade web (Saracco et al, 2016) and assess the similarity of stock market portfolios (Gualdi et al, 2016).

The advantage of motifs is that they are significantly more sensitive to changes in network structure than the indices traditionally used to describe bipartite ecological networks. In other words, a wide diversity of network configurations can have similar values of indices such as nestedness, but far fewer network configurations have similar motif compositions. A recent analysis found that, on average, motifs capture 63% more information about network structure than even multivariate combinations of popular network-level indices, and an average of 528% more information than multivariate combinations of species-level indices; this latter value rises to 1076% more information in the most extreme case (Simmons, Cirtwill, et al., 2018). Thus, while indices are useful, they also have important limitations. As a simple example, the degree of a plant might show it is visited by two pollinators, while motifs could reveal that one of these pollinators is a generalist visiting three other generalist plants, while the other is a specialist visiting only the focal plant. Such distinctions can have important consequences for understanding the ecology and evolution of communities and so are essential to incorporate in network analyses. However, while the motif framework is gaining in popularity, no software currently exists to conduct motif analyses of bipartite networks in R, the most popular programming language among ecologists.

To fill this gap, we introduce bmotif: an R package, based on a formal mathematical framework, for counting motifs, and species positions within motifs, in bipartite networks. While bmotif is primarily an R package, we additionally provide MATLAB and Python code that replicates the core package functionality. Here, we introduce the motifs and motif positions counted by bmotif and describe the package's main functions

and performance. We then provide two examples showing how bmotif can be used to answer questions about ecological communities. While here we focus on mutualistic bipartite networks, our methods are general so can also be applied to other types of interaction, such as parasitism and herbivory, and even non-biological systems, such as trade networks, finance networks and recommendation systems.

Description

Defining bipartite motifs

In a bipartite network containing N species, a motif is a subnetwork comprising n species and their interactions (where $n < N$ and all species have at least one interaction). Fig. 1 shows the motifs included in bmotif: all 44 possible motifs containing up to six nodes. Large numbers represent the identity of each motif. Within motifs, species can appear in different positions. Nodes in a motif share the same position if there exists a permutation of these nodes, together with their links, that preserves the motif structure (see Appendix S1 for formal definition) (Kashtan, Itzkovitz, Milo, & Alon, 2004). For example, in motif 9, the left and centre nodes in the top level can be swapped without changing the motif structure, but the centre and right nodes cannot (Fig. 1). The 148 unique positions a species can occupy across all motifs up to six nodes are shown in Fig. 1 as small numbers associated with each node. These positions are important because each represents a different ecological situation with a unique set of direct and indirect interactions. For example, in motif 6 both species in the top level are in the same position (position 14), indicating that they have identical topological roles: both have a single interaction with the shared resource in position 5. Conversely, in motif 5, both top-level species are in different positions (12 and 11), which can have important functional consequences. For example, while the species in position 11 is a specialist on the resource in position 10, the species in position 12 has a wider diet breadth, interacting with species in both position 9 and 10 and thus having greater redundancy in its partners. Motifs and positions are ordered as in Baker et al. (2015) Appendix 1.
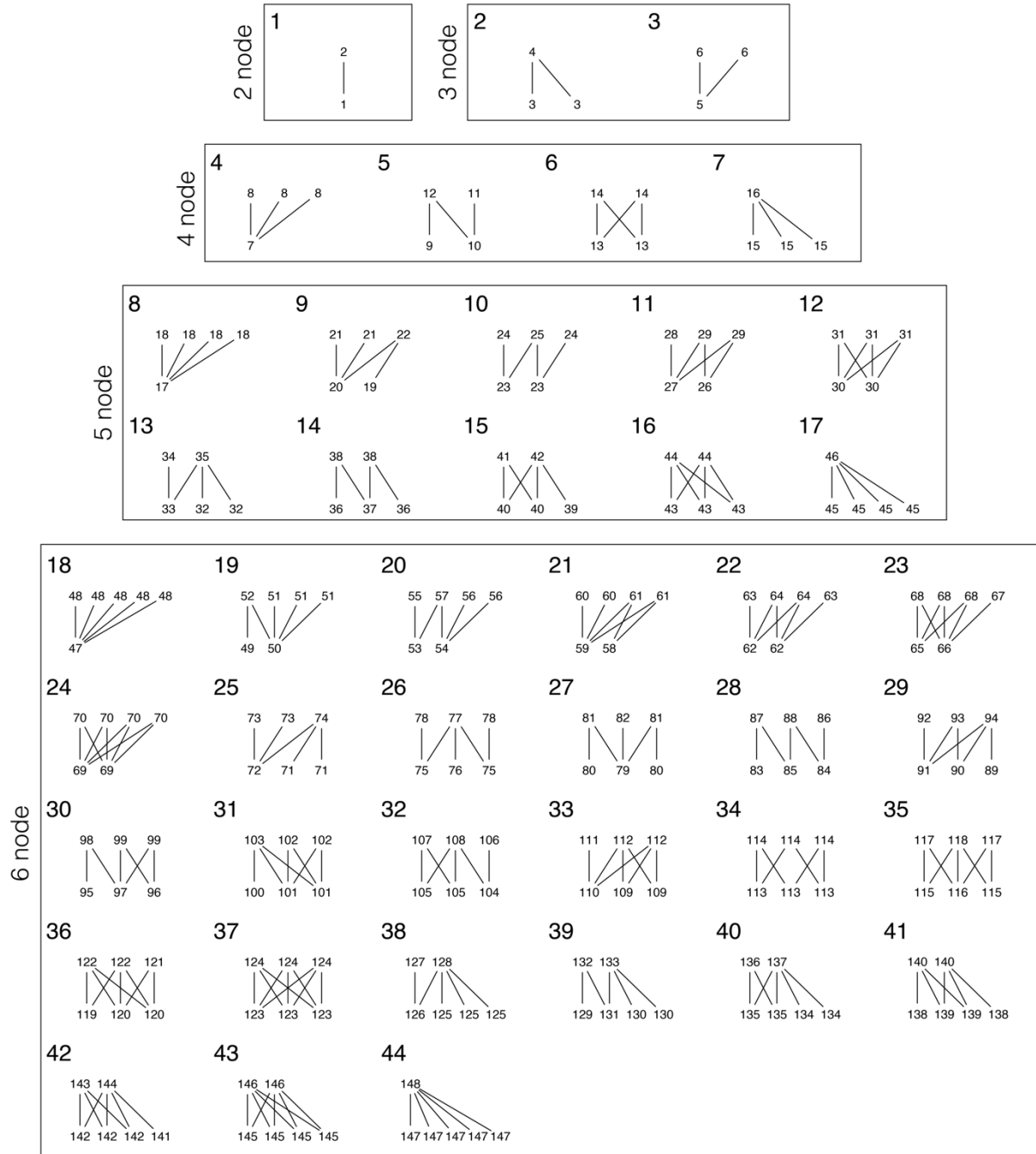
Figure 1: All bipartite motifs containing up to 6 nodes (species). Large numbers identify each motif. Small numbers represent the unique positions species can occupy within motifs, following Appendix 1 in Baker et al. (2015). Lines between small numbers indicate undirected species interactions. There are 44 motifs containing 148 unique positions.

Networks in bmotif are represented as biadjacency matrices (**M**), with one row for each species in the first set (such as pollinators) and one column for each species in the second set (such as plants). When species i and j interact, $m_{ij} = 1$; if they do not interact $m_{ij} = 0$. This widely-used representation was chosen for compatibility with other

packages and open-access network repositories, such as the Web of Life (www.web-of-life.es). Species in rows correspond to nodes in the top level of the motifs in Fig. 1; species in columns correspond to nodes in the bottom level. Appendix S2 shows how each motif is represented in a biadjacency matrix.

Main functions

bmotif has two functions: (i) *mcount*, for calculating how frequently different motifs occur in a network, and (ii) *node_positions*, for calculating the frequency with which species (nodes) occur in different positions within motifs to quantify a species' structural role. To enumerate motif frequencies and species position counts, bmotif uses mathematical operations directly on the biadjacency matrix: for the first time, we derive 44 expressions for each of the 44 motifs and 148 expressions for each of the 148 positions within motifs (Appendix S3). The advantage of this approach is that analyses with bmotif are fast: using a dataset of 175 empirical pollination and seed dispersal networks, *mcount* completed in 0.01 seconds and *node_positions* completed in 0.32 seconds for a network with 78 species (close to the mean network size of 77.1 species) and for motifs up to 6 nodes. Appendix S4 gives full details and analyses of bmotif's computational performance while Appendix S5 provides a detailed description of the outputs returned by the two functions.

Example analyses

Comparing community structures

Here we use bmotif to examine the assembly and disassembly of an Arctic plant-pollinator community. Networks were sampled daily, when weather conditions allowed, at the Zackenberg Research Station in northeastern Greenland, across two full seasons in 1996 (24 days) and 1997 (26 days) (Olesen, Bascompte, Elberling, & Jordano, 2008). While these networks use the frequency of animal visits to plants as a surrogate for true pollination, this has been shown to be a reasonable proxy in mutualistic networks (Vázquez, Morris, & Jordano, 2005; Simmons, Sutherland, et al., 2018). Data were obtained from Saavedra et al. (2016). We used *mcount* to calculate motif frequencies in each daily network in both years, normalised using 'normalise_nodesets', which

expresses the frequency of each motif as the number of sets of species that form the motif as a proportion of the number of sets of species that could form that motif (Poisot & Stouffer, 2016). Days 1 and 24 in 1996, and days 1 and 26 in 1997, were excluded from the analysis as they were too small for some motifs to occur. Table 1 shows the data frame returned by *mcount* for an example daily network (day 12 in 1996), and Fig. 2b visualises the distribution of motifs in this network. Using nonmetric multidimensional scaling (NMDS), we visualised how the community structure changed from assembly after the last snow melt to disassembly at the first snow fall, in two consecutive years (Fig. 2a). NMDS is an ordination technique that attempts to represent the pairwise dissimilarities between multidimensional data in a lower-dimensional space as accurately as possible (Kruskal, 1964). NMDS can be used with any dissimilarity measure and is regarded as one of the most robust ordination techniques in ecology (Minchin, 1987). NMDS analyses were conducted with the metaMDS function in the R package vegan using Bray-Curtis dissimilarities (Oksanen et al., 2016). We used Bray-Curtis dissimilarity as it is a robust dissimilarity measure for a wide range of community traits, including motifs (Baker et al., 2015; Simmons, Cirtwill, et al., 2018). More positive values of the first NMDS axis are associated with motifs where generalist pollinators compete for generalist plants, while negative values are associated with motifs where more specialist pollinators have greater complementarity in the specialist plants they visit. More positive values of the second NMDS axis are associated with loosely connected motifs containing specialist plants interacting with both specialist and generalist pollinators, while negative values are associated with highly connected motifs containing pollinators competing for generalist plants. While the community was relatively stable over time in the 1996 season, there were larger structural changes in 1997, with a largely monotonic shift from high competition between generalist pollinators at the start of the season, to lower competition between more specialist pollinators at the end of the season, with a more complementary division of plant resources (Fig. 2). Thus while network structure may appear stable when analysed with traditional indices such as connectance (Olesen et al., 2008), motifs reveal the presence of complex, ecologically-important structural dynamics. Additionally, it is clear that, even in consecutive years, the community followed different structural trajectories, emphasising the danger of treating networks as static entities.

motif	nodes	frequency	normalise_sum	normalise_sizeclass	normalise_nodesets
1	2	140	0.00200194	1	0.34313725
2	3	621	0.00888005	0.57393715	0.13235294
3	3	461	0.00659212	0.42606285	0.14123775
4	4	1153	0.01648744	0.1370661	0.07064951
5	4	4486	0.06414803	0.53328578	0.11951194
6	4	831	0.01188297	0.09878745	0.02213875
7	4	1942	0.02776983	0.23086068	0.05644036
8	5	2393	0.03421896	0.03968623	0.04189426
9	5	10689	0.15284848	0.17726956	0.05695332
10	5	5243	0.07497283	0.08695147	0.02793585
11	5	5941	0.08495396	0.09852731	0.03165494
12	5	901	0.01288394	0.01494245	0.00480072
13	5	12815	0.18324944	0.21252778	0.04655531
14	5	8564	0.12246182	0.14202793	0.03111195
15	5	8002	0.11442544	0.13270755	0.02907027
16	5	1096	0.01567237	0.01817639	0.00398163
17	5	4654	0.06655036	0.07718332	0.02576367

Table 1: The data frame returned by *mcount* for an example daily network from Zackenberg Research Station in northeastern Greenland (day 12 in 1996). Details of the different columns are given in Appendix S5.
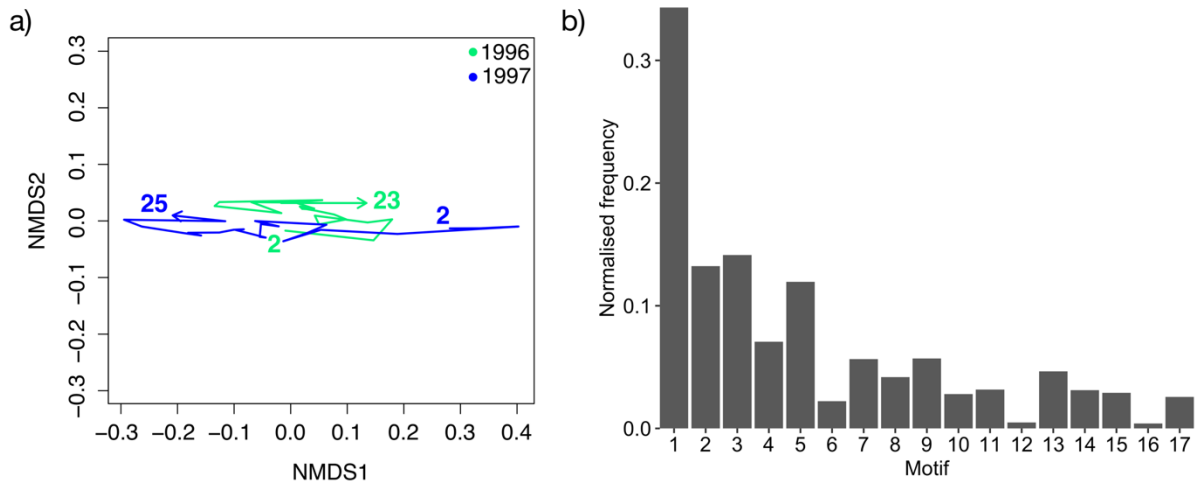

Figure 2: (a) Nonmetric multidimensional scaling plot (NMDS) showing change in Arctic plant-pollinator network structure over the 1996 and 1997 seasons, quantified using motifs. Numbers represent the days of sampling. (b) The normalised frequency of motifs in one time slice network (day 12 in 1996).

Comparing species' structural roles

We use *node_positions* to compare the roles of native and introduced plant species in a plant-pollinator community sampled in Mauritius in November 2003 (Kaiser-Bunbury, Memmott & Müller 2009; 48 species, 75 interactions, connectance = 0.134). Network data were obtained from the Web of Life dataset (www.web-of-life.es) and information on plant origin was obtained from Kaiser-Bunbury et al., 2009 Appendix II. We calculated the sum-normalised roles of all plant species (16 native and 4 introduced; see Table 2 for the data frame returned by *node_positions* and Fig. 3b for the motif composition of the network) and plotted them on two NMDS axes (Fig. 3a). This figure shows three striking features. First, there is almost no overlap between native and introduced species' interaction niches. Similar to research showing that non-native species can occupy different functional niches to native species (Ordonez, Wright, & Olff, 2010), these results suggest they may also occupy unexploited interaction niches. This aligns with previous studies showing differences in species-level network indices between native and invasive plant species, such as higher generalisation (Albrecht, Padrón, Bartomeus, & Traveset, 2014) and species strength (Maruyama et al., 2016). Further research could use motifs to investigate whether introduced species 'pushed' native species out of previously occupied interaction niche space, or whether introduced species colonised previously-unused space. If the latter is true, the size of a community's unused 'role space' could potentially inform predictions of its vulnerability to invasion. Second, the interaction niche of introduced species is much smaller than that of native species: the four introduced species all occupy similar areas of motif space, possibly suggesting a single 'invader role'. This could have important implications for predicting the effects of invasive species on community structure, an important challenge especially in the face of global changes. While previous studies have identified species and community traits that predict the identity of invasive species, or communities vulnerable to invasion, it has recently been argued that species topological roles are a more practical predictor of how species could affect communities because they are comparatively easier to sample. (Emer, Memmott, Vaughan, Montoya, & Tylianakis, 2016). Thus, our finding could lay the foundation for future work predicting which species will become invasive based on their motif roles alone, especially given evidence that species roles are conserved across native and alien ranges (Emer et al.,

2016). Third, introduced species occupy lower values on the second NMDS axis, corresponding to motif positions where they are visited by generalist pollinator species, possibly due to the absence of co-evolutionary associations with specialists.

	np1	np2	np3	np4	np5	np6	...	np46
Sideroxylon puberulum	0.000000	0.003380	0.000000	0.010140	0.000000	0.011589		0.016900
Grangeria borbonica	0.000000	0.002259	0.000000	0.007905	0.000000	0.008752		0.019763
Badula platiphylla	0.000000	0.002629	0.000000	0.005258	0.000000	0.009989		0.002629
Helichrysum proteoides	0.000000	0.001903	0.000000	0.011415	0.000000	0.005854		0.104639
Myonima violacea	0.000000	0.002358	0.000000	0.001179	0.000000	0.014151		0.000000
Harungana madagascariensis	0.000000	0.002494	0.000000	0.002494	0.000000	0.012469		0.000000
Stillingia lineata	0.000000	0.001832	0.000000	0.000916	0.000000	0.010989		0.000000
Ochna mauritiana	0.000000	0.001793	0.000000	0.002689	0.000000	0.012550		0.000448
Olea lancea	0.000000	0.001768	0.000000	0.000884	0.000000	0.011494		0.000000
Psiadia terebinthina	0.000000	0.002208	0.000000	0.007728	0.000000	0.008832	...	0.019321
Aphloia theiformis	0.000000	0.001570	0.000000	0.000000	0.000000	0.014129		0.000000
Psidium cattleianum	0.000000	0.002469	0.000000	0.002469	0.000000	0.009877		0.000000
Coffea macrocarpa	0.000000	0.002847	0.000000	0.004270	0.000000	0.012100		0.000712
Homalanthus populifolius	0.000000	0.001832	0.000000	0.000916	0.000000	0.010989		0.000000
Faujasiopsis flexuosa	0.000000	0.001605	0.000000	0.001605	0.000000	0.012841		0.000000
Gaertnera sp1	0.000000	0.002956	0.000000	0.004435	0.000000	0.013304		0.000739
Coffea mauritiana	0.000000	0.011236	0.000000	0.000000	0.000000	0.022472		0.000000
Gaertnera rotundifolia	0.000000	0.004975	0.000000	0.000000	0.000000	0.014925		0.000000
Warneckea trinervis	0.000000	0.001570	0.000000	0.000000	0.000000	0.014129		0.000000
Wikstroemia indica	0.000000	0.001020	0.000000	0.000000	0.000000	0.012245		0.000000

Table 2: The data frame returned by *node_positions* for the Mauritius plant-pollinator network. Details of this output are given in Appendix S5. For visualisation purposes only columns 1–6 and 46 are shown.
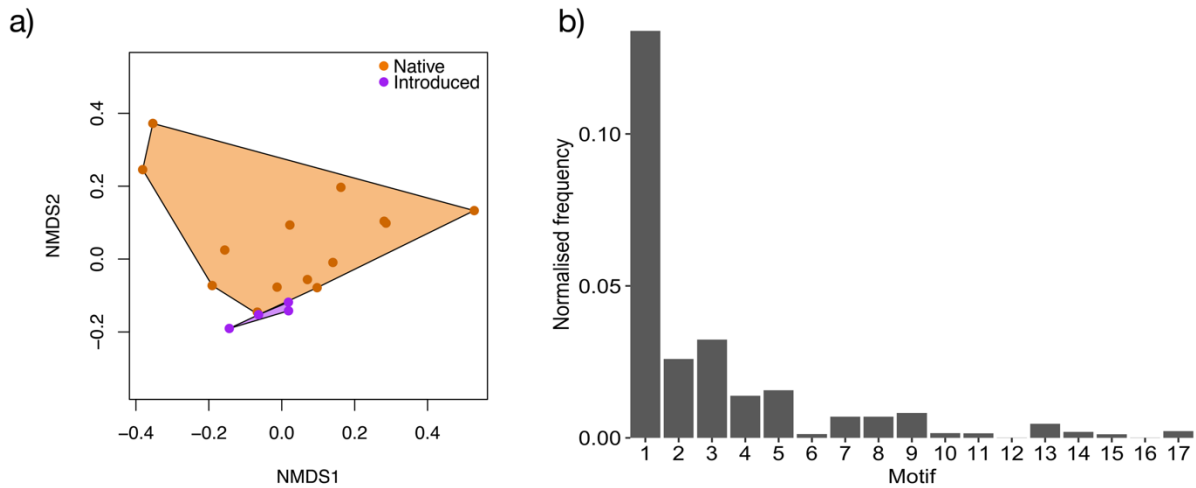
Figure 3: (a) The roles of native and introduced species in a plant-pollinator network. Each point represents the role of a species in the network. Shaded polygons are convex hulls either containing all introduced species or all alien species. (b) The normalised frequency of motifs in the network.

Implementation and availability

The bmotif package is available for the R programming language. The package can be installed in R using install.packages("bmotif"). This paper describes version 1.0.0 of the software. The source code of the package is available at https://github.com/SimmonsBI/bmotif. Any problems can be reported using the *Issues* system. The code is version controlled with continuous integration and has code coverage of approximately 98%. MATLAB and Python code replicating the core package functionality is available at https://github.com/SimmonsBI/bmotif-matlab and https://github.com/SimmonsBI/bmotif-python respectively. All code is released under the MIT license.

Conclusions

bmotif is an R package and set of mathematical formulae enabling motif analyses of bipartite networks. Specifically, bmotif provides functions for two key analyses: (i) enumerating the frequency of different motifs in a network, and (ii) calculating how often species occur in each position within motifs. These two techniques capture important information about network structure that may be missed by traditional methods. As an illustration, by analysing the roles of native and introduced plant species

in a plant-pollinator network, we found that introduced species adopted similar roles in the community that differed from those of native species. Motif approaches represent a new addition to the network ecologists 'toolbox' for use alongside other techniques to analyse bipartite networks. We hope bmotif encourages further uptake of the motif approach to shed light on the ecology and evolution of ecological communities.

Data accessibility

All networks are available from the Web of Life repository (www.web-of-life.es), with the exception of the Greenland plant-pollinator networks which are available from Data Dryad (Saavedra et al., 2016). Plant origin data for Mauritius networks was from Kaiser-Bunbury et al., 2009 Appendix II.

References

Albrecht, M., Padrón, B., Bartomeus, I., & Traveset, A. (2014). Consequences of plant invasions on compartmentalization and species' roles in plant-pollinator networks. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140773. doi:10.1098/rspb.2014.0773

Baker, N. J., Kaartinen, R., Roslin, T., & Stouffer, D. B. (2015). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography*, *38*(2), 130–139. doi:10.1111/ecog.00913

Bascompte, J., Jordano, P., Melián, C. J., & Olesen, J. M. (2003). The nested assembly of plant--animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387.

Dáttilo, W., Guimarães, P. R., & Izzo, T. J. (2013). Spatial structure of ant-plant mutualistic networks. *Oikos*, *122*(11), 1643–1648. doi:10.1111/j.1600-0706.2013.00562.x

Diestel, R. (2000). *Graph Theory (Graduate Texts in Mathematics). Springer-Verlag New York*. doi:10.1109/IEMBS.2010.5626521

Emer, C., Memmott, J., Vaughan, I. P., Montoya, D., & Tylianakis, J. M. (2016). Species roles in plant–pollinator communities are conserved across native and alien ranges. *Diversity and Distributions*, *22*(8), 841–852. doi:10.1111/ddi.12458

Gualdi, S., Cimini, G., Primicerio, K., Di Clemente, R., & Challet, D. (2016). Statistically

validated network of portfolio overlaps and systemic risk. *Scientific Reports, 6*.
doi:10.1038/srep39467

Guillaume, J. L., & Latapy, M. (2004). Bipartite structure of all complex networks.
Information Processing Letters, 90(5), 215–221. doi:10.1016/j.ipl.2004.03.007

Kaiser-Bunbury, C. N., Memmott, J., & Müller, C. B. (2009). Community structure of
pollination webs of Mauritian heathland habitats. *Perspectives in Plant Ecology,
Evolution and Systematics, 11*(4), 241–254. doi:10.1016/j.ppees.2009.04.001

Kashtan, N., Itzkovitz, S., Milo, R., & Alon, U. (2004). Topological generalizations of
network motifs. *Physical Review E - Statistical Physics, Plasmas, Fluids, and
Related Interdisciplinary Topics, 70*(3), 12. doi:10.1103/PhysRevE.70.031909

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a
nonmetric hypothesis. *Psychometrika, 29*(1), 1–27. doi:10.1016/B978-0-12-385157-
4.00313-4

Maruyama, P. K., Vizentin-Bugoni, J., Sonne, J., Martín González, A. M., Schleuning,
M., Araujo, A. C., … Dalsgaard, B. (2016). The integration of alien plants in
mutualistic plant-hummingbird networks across the Americas: The importance of
species traits and insularity. *Diversity and Distributions, 22*(6), 672–681.
doi:10.1111/ddi.12434

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002).
Network motifs: simple building blocks of complex networks. *Science, 298*(5594),
824–827.

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for
ecological ordination. *Vegetatio, 69*(1–3), 89–107. doi:10.1007/BF00038690

Newman, M. E. . (2010). *Networks. An introduction. Oxford University Press*.
doi:10.1111/j.1468-5922.2010.01872_2.x

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., …
Wagner, H. (2016). vegan: Community Ecology Package. R package version 2.4-0.

Olesen, J. M., Bascompte, J., Elberling, H., & Jordano, P. (2008). Temporal dynamics in
a pollination network. *Ecology, 89*(6), 1573–1582. doi:10.1890/07-0451.1

Ordonez, A., Wright, I. J., & Olff, H. (2010). Functional differences between native and
alien species: A global-scale comparison. *Functional Ecology, 24*(6), 1353–1361.
doi:10.1111/j.1365-2435.2010.01739.x

Poisot, T., & Stouffer, D. (2016). How ecological networks evolve. *BioRxiv*. Retrieved from http://biorxiv.org/content/early/2016/08/29/071993.abstract

Rodríguez-Rodríguez, M. C., Jordano, P., & Valido, A. (2017). Functional consequences of plant-animal interactions along the mutualism-antagonism gradient. *Ecology*, *98*(5), 1266–1276. doi:10.1002/ecy.1756

Saavedra, S., Rohr, R. P., Olesen, J. M., & Bascompte, J. (2016). Data from: Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and Evolution*. Dryad Digital Repository. doi:doi:10.5061/dryad.3pk73

Saracco, F., Di Clemente, R., Gabrielli, A., & Squartini, T. (2015). Randomizing bipartite networks: the case of the World Trade Web. *Scientific Reports*, 5, 10595. Retrieved from http://dx.doi.org/10.1038/srep10595

Saracco, F., Di Clemente, R., Gabrielli, A., & Squartini, T. (2016). Detecting early signs of the 2007-2008 crisis in the world trade. *Scientific Reports*, *6*. doi:10.1038/srep30286

Simmons, B. I., Cirtwill, A. R., Baker, N. J., Wauchope, H. S., Dicks, L. V., Stouffer, D. B., & Sutherland, W. J. (2018). Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, (in press). doi:10.1111/oik.05670

Simmons, B. I., Sutherland, W. J., Dicks, L. V, Albrecht, J., Farwig, N., García, D., … González-Varo, J. P. (2018). Moving from frugivory to seed dispersal: incorporating the functional outcomes of interactions in plant-frugivore networks. *Journal of Animal Ecology*.

Vázquez, D. P., Morris, W. F., & 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

Discussion

The projects in this thesis were mostly conducted independently, rather than as part of a sequential progression, and thus it is useful to first summarise the key findings of each chapter.

Chapter 1: A large proportion of interactions in plant-frugivore visitation networks are non-mutualistic. These networks are a good proxy for true seed dispersal networks when considering whole-network structure, but less so for species-level structure.

Chapter 2: The widespread positive relationship between abundance and generalisation in mutualistic networks appears to be unidirectional, with abundance driving generalisation rather than the other way round. Despite the importance of morphological matching in governing plant-hummingbird interactions, neutral effects also play a role.

Chapter 3: The mutualistic interactions which are most vulnerable to extinction are also those which contribute most to the stability of a network. Moreover, for many interactions, vulnerability and contribution to stability are determined by the identity of the taxa involved in the interaction, rather than ecological context. This means that the vulnerability and contribution to stability of an interaction is similar across different networks, implying a form of evolutionary conservatism.

Chapter 4: Indices are a useful tool to characterise network structure and species roles within networks, but we need to be aware of their limitations. Other methods of characterising network structure, like motifs, are less lossy and could be particularly useful for questions concerning interaction turnover or species roles.

Chapter 5: Counting motifs in bipartite networks, or counting the number of times species occur in unique positions within motifs, can be done quickly in R, MATLAB and Python. There is preliminary evidence that invasive species may occupy a single 'invader role' in mutualistic networks.

The future of network ecology

While the above chapters can be considered independently, they also all fall under a unifying theme: the structure and stability of ecological networks. Having explored a wide range of questions within this theme, some conclusions can be drawn about where network ecology might be heading.

The panacea of network ecology is perhaps to have networks for all species and all interaction types, everywhere and over time. While this goal is of course impossible, recent progress in methods and data are bringing us closer. For example, multilayer networks offer the methodological framework with which to move beyond single-site networks comprising one interaction type, to instead consider networks of multiple interaction types over space and time (Pilosof *et al.* 2017). Data are increasingly available to tackle these issues too, with studies of 'networks of networks' incorporating different interaction types and eDNA analyses revealing 'mobile link' species that connect networks over space and time (Pocock *et al.* 2012; González-Varo *et al.* 2017). Having networks over larger spatial and temporal scales would enable analyses of how stressors, such as climate change and habitat loss, affect whole, connected ecosystems rather than individual communities.

Networks with multiple interaction types over space and time

As ecology moves towards these larger and more varied networks, Chapter 1 highlights the need to keep in mind issues of data quality. Networks are a tool to answer interesting ecological questions, rather than an end unto themselves, and thus we must continue to remember how data are collected, even as the data themselves and the tools to analyse them grow in complexity.

The eternal call to move from pattern to process continues to resonate in network ecology. Chapter 2 highlights that there are still interesting processual relationships to uncover, even in well-studied systems, and we must be mindful of this as networks with multiple interaction types over space and time become more common. Admittedly, these new networks will also have interesting structures to uncover, and so new methods are necessary to characterise their topology. Chapters 5 and 6 draw attention to the large amount of structural detail missed by multivariate combinations of network indices in even simple networks. Thus complex multilayer networks require more than just adaptations of traditional indices, but also adaptations of other structural measures like motifs and perhaps new measures entirely. Extinction models, too, will have to be adapted, both dynamic and topological. These new methods and models must do justice to the rich data they are applied to. The ground is therefore fertile for methodological development in the coming years, and computational ecologists and theoreticians can play a key role as 'importers' of methods from mathematics and physics.

An eye to conservation

Network ecology has much to offer conservation (Kaiser-Bunbury & Blüthgen 2015). However, a largely unexplored avenue is the conservation of interactions, rather than species or whole communities. Chapter 3 highlights the risk interaction loss poses to communities, and it is important that this area continues to be researched. Particularly, research is needed into how the requirements of interaction conservation – two species to be maintained at sufficiently high abundance to interact – affect the practices and policies of conservation on the ground.

Interaction conservation is important because interaction diversity provides many benefits to ecological communities. For example, higher interaction richness is associated with an increase in the rate of ecosystem processes, which can improve seed set in mutualistic communities through higher functional complementarity (Tylianakis et al 2010). Interaction diversity can also promote response diversity and system resilience: with higher link diversity, species' average generalism will increase, allowing species to receive and deliver stable mutualistic benefit over time even in the presence of environmental perturbations (Sole and Montoya 2001; Rezende et al 2007). Finally,

interaction diversity can also mean improved response diversity (Tylianakis et al 2010), defined as the range of responses species have to a stressor. While there are few studies of response diversity in an interaction network context, response diversity has been shown to be important for pollination rates in agricultural contexts (Winfree and Kremen, 2008).

It is important to note that in most empirical species interaction networks, a large proportion of interactions will be missed due to sampling effects. This is likely to be rare interactions between low abundance species (de Aguiar et al 2017), but as Chapter 3 shows, such interactions may be important for community stability. An essential first step for interaction conservation, therefore, is to find these missing interactions. While greater sampling effort could go some way towards this goal, it is unlikely to be a viable solution due to the high financial and time costs associated with fieldwork. Instead, modelling approaches will be required to 'fill in the gaps' in ecological networks. For example, Weinstein and Graham use Bayesian hierarchical models to infer missing interactions in plant-hummingbird networks, while Terry and Lewis (2019) find that structural models produce good predictive performance. Improving and refining these methods will uncover the up to 50% of interactions that are missed during sampling (Dorado et al 2011; Rivera-Hutinel et al 2012).

Given that the interactions lost from a community are likely to be non-random (Aizen et al 2012), this loss could lead to non-random changes in higher-level network structure. In turn, this could have negative effects on emergent properties of communities, such as stability. This is particularly true if metrics such as modularity of nestedness are lowered, as these have been repeatedly demonstrated to have stabilising effects on ecological communities (Thebault and Fontaine, 2010). However, the effect of realistic interaction loss scenarios on macro-scale network topology is as yet not known. Perhaps, because weak interactions are more vulnerable (see Chapter 3), interaction extinction could negatively affect nestedness, because nestedness relies on asymmetric specialisation: if weak links disappear, this asymmetry may be eroded (Tylianakis et al 2010). Non-random interaction loss could also cause shifts in the distribution of interaction strength within a network. This can be important because

weak interactions have repeatedly been shown to have a stabilising effect on population dynamics within networks because they dampen strong oscillations (McCann et al 1998). The loss of weak links could also lower asymmetric dependence (a quantitative measure of the proportion of interaction strength a species allocates to another) in networks, which Bascompte et al (2006) showed can buffer communities against secondary extinctions.

This thesis has also discussed motifs. While the dynamic implications of motifs in bipartite networks are as yet unknown, this is an important area for future research. Above, I discussed how interaction loss could change macro-scale network properties important for conservation. However, as shown in Chapter 4, these macro-properties mask a large amount of detail in network structure. By evaluating the dynamic implications of local-scale network structure – that is, motifs – it will be possible to understand the stability and resilience of communities in more detail. However, bipartite motifs remain a nascent idea in the literature, so it is difficult to evaluate their conservation implications at this time. However, this is likely to be a promising area moving forward.

Overall, network ecology has much to offer conservation, through capturing species, their interactions and community structure as a whole. This thesis shines light on the conservation of interactions in particular, which need to be adequately described, and which could have negative impacts on communities if lost. Motifs, too, may have something to offer conservation, in terms of linking dynamics to a detailed description of structure.

Conclusion

Despite its decades of research, network ecology remains a field with much to discover. To date, the field (including this thesis) has been dominated by single-site and single-interaction networks, alone or in synthesis. These are undoubtedly useful, but a 'new' network ecology will use more expansive datasets to consider multiple interaction types over space and time. As these multilayer networks become more common, we must develop new structural measures and keep an eye on process. Ultimately, networks

remain a powerful representation of ecological communities that enable understanding of both community structure and individual species roles. Exciting methodological developments and datasets suggest that network ecology has a long and interesting future ahead.

References

Aizen, M.A., Sabatino, M. and Tylianakis, J.M., 2012. Specialization and rarity predict nonrandom loss of interactions from mutualist networks. Science, 335(6075), pp.1486-1489.

Bascompte, J., Jordano, P. and Olesen, J.M., 2006. Asymmetric coevolutionary networks facilitate biodiversity maintenance. Science, 312(5772), pp.431-433.

de Aguiar, M.A., Newman, E.A., Pires, M.M., Yeakel, J.D., Hembry, D.H., Burkle, L., Gravel, D., Guimaraes Jr, P.R., O'Donnell, J., Poisot, T. and Fortin, M.J., 2017. Revealing biases in the sampling of ecological interaction networks. arXiv preprint arXiv:1708.01242.

Dorado, J., Vázquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and specialization in plant–pollinator networks. *Ecology*, 92, 19–25.

González-Varo, J.P., Carvalho, C.S., Arroyo, J.M. & Jordano, P. (2017). Unravelling seed dispersal through fragmented landscapes: Frugivore species operate unevenly as mobile links. *Mol. Ecol.*, 26, 4309–4321.

Kaiser-Bunbury, C.N. & Blüthgen, N. (2015). Integrating network ecology with applied conservation: A synthesis and guide to implementation. *AoB Plants*.

McCann, K., Hastings, A. and Huxel, G.R., 1998. Weak trophic interactions and the balance of nature. Nature, 395(6704), p.794.

Pilosof, S., Porter, M.A., Pascual, M. & Kéfi, S. (2017). The Multilayer Nature of Ecological Networks. *Nat. Ecol. Evol.*, 1, 0101.

Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012). The robustness and restoration of a network of ecological networks. *Science (80-.).*, 335, 973–7.

Rezende, E.L., Lavabre, J.E., Guimarães, P.R., Jordano, P. and Bascompte, J., 2007. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448(7156), p.925.

Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects of sampling completeness on the structure of plant–pollinator networks. *Ecology*, 93, 1593–1603.

Sole, R.V. and Montoya, M., 2001. Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1480), pp.2039-2045.

Terry , J. C. and Lewis, O. 2019. Finding missing links in interaction networks, *bioRxiv* 695726

Thébault, E. and Fontaine, C., 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science, 329(5993), pp.853-856.

Tylianakis, J.M., Laliberté, E., Nielsen, A. and Bascompte, J., 2010. Conservation of species interaction networks. Biological conservation, 143(10), pp.2270-2279.

Vieira, M.C. and Almeida-Neto, M., 2015. A simple stochastic model for complex coextinctions in mutualistic networks: robustness decreases with connectance. Ecology Letters, 18(2), pp.144-152.

Weinstein, B.G. and Graham, C.H., 2017. Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions. Ecology letters, 20(3), pp.326-335.

Winfree, R. and Kremen, C., 2008. Are ecosystem services stabilized by differences among species? A test using crop pollination. Proceedings of the Royal Society B: Biological Sciences, 276(1655), pp.229-237.

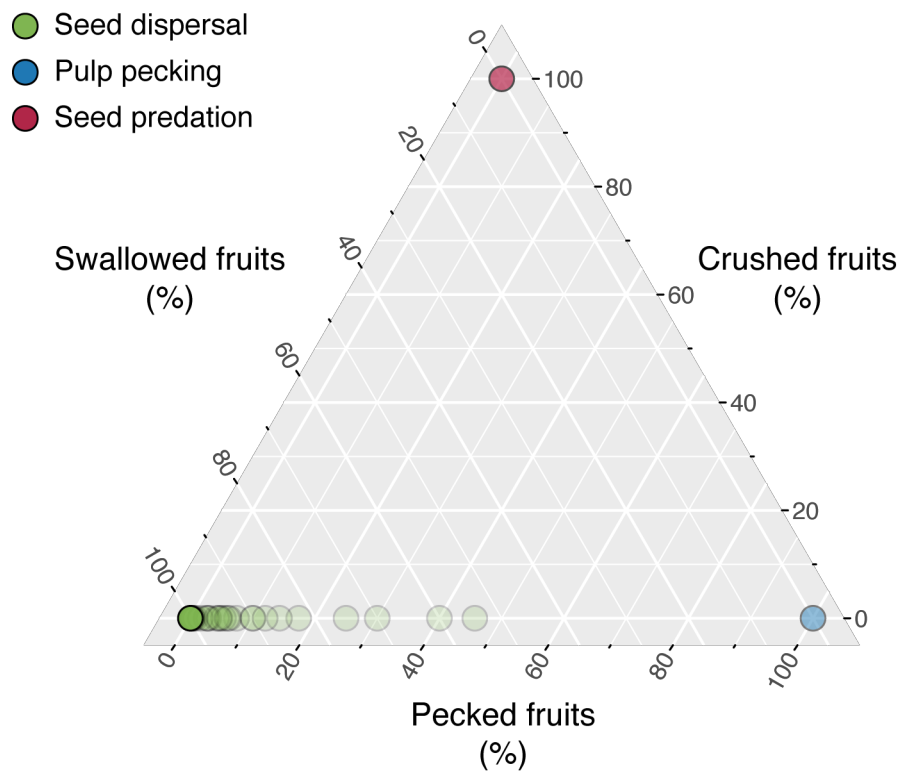# Supplementary materials related to Chapter 1

Figure S1. Ternary plot showing the percentage of fruits that were swallowed, pecked or crushed in 77 pairwise fruit-bird interactions from network VII (Farwig *et al*. 2017) that were classified as 'seed dispersal' ($n = 69$), 'pulp pecking' ($n = 3$) or 'seed predation' ($n = 5$). These 77 interactions – out of a total of 129 – were those for which the number of fruits observed to obtain the percentages was ≥ 6 (range = 6–1794 fruits, median = 46, total = 10,997). The figure illustrates that our categorical classification is, in general, very consistent with detailed information on the specific role of each pairwise interaction. In particular, when the interactions were classified as 'seed predation' there was no 'seed dispersal' or 'pulp pecking'. Moreover, interactions classified as 'seed dispersal' only included small fractions of 'pulp pecking' (median = 0%, mean = 3.7%), and only in few cases this fraction was higher than 20% (max. = 45.7%). Plant-frugivore interactions were observed using binoculars from camouflage tents. During the observations all frugivore species visiting the individual plants were recorded, as well as the number of frugivore individuals, the duration of frugivore visits and fruit-handling behaviour (interaction outcome). Swallowing, (seed) crushing and (pulp) pecking were distinguished. Thereby, single visitors could handle fruits in various ways, so that some fruits were swallowed, crushed or pecked during the same visit. If a group of conspecific frugivores visited a plant

and individual behaviour could not be observed simultaneously, the most visible individual was focused on. If the behaviour of different species could not be observed simultaneously, the rarer species was focused on.
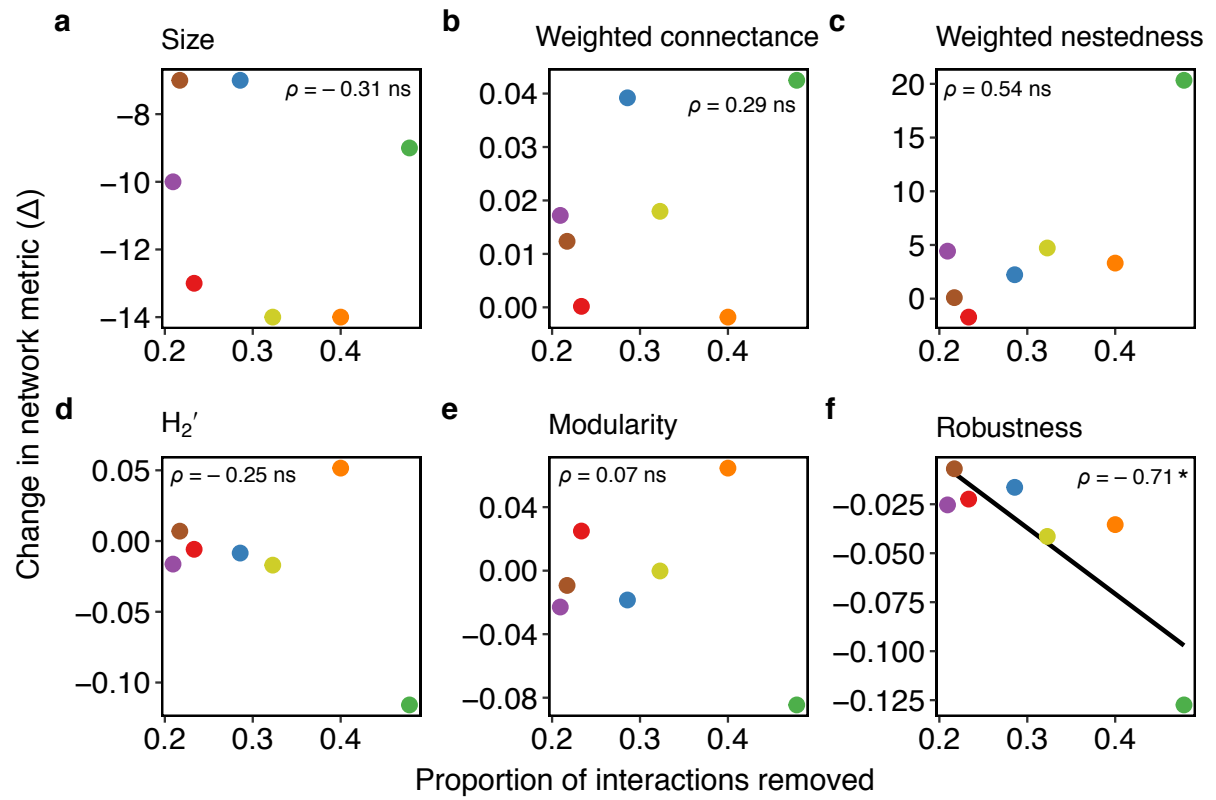
Figure S2. Relationship between the difference (Δ) in the network-level metrics after the removal of non-mutualistic interactions (seed predation and pulp pecking), and the proportion of interactions removed in the original networks. For illustrative purposes, we show a regression line in significant Spearman's rank correlations (*ns*: non-significant; * $P < 0.05$).
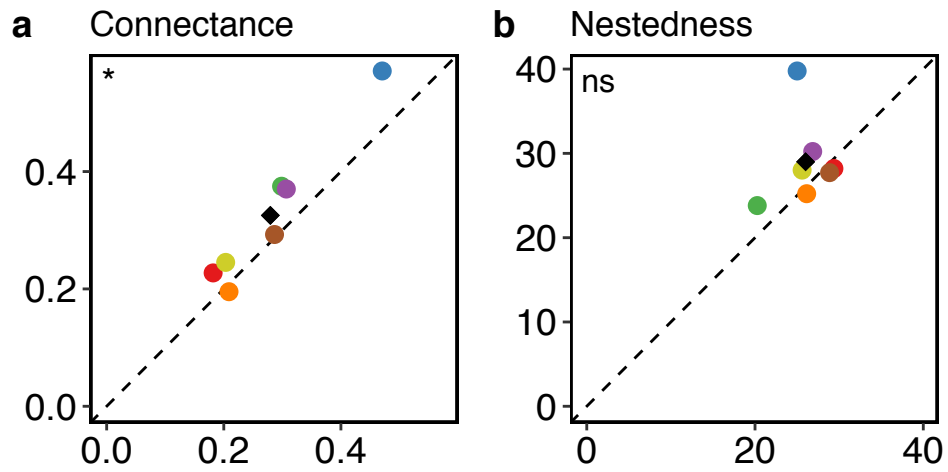
Figure S3. Changes in unweighted (not incorporating information on interaction frequency) connectance and nestedness after the removal of non-mutualistic interactions. Colour codes denote network identity (see Fig. 1b). The black diamonds are mean values across networks. The dashed line is $y = x$, indicating the position of points if there was no change in metric values. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*ns*: non-significant; * $P <$ 0.05).
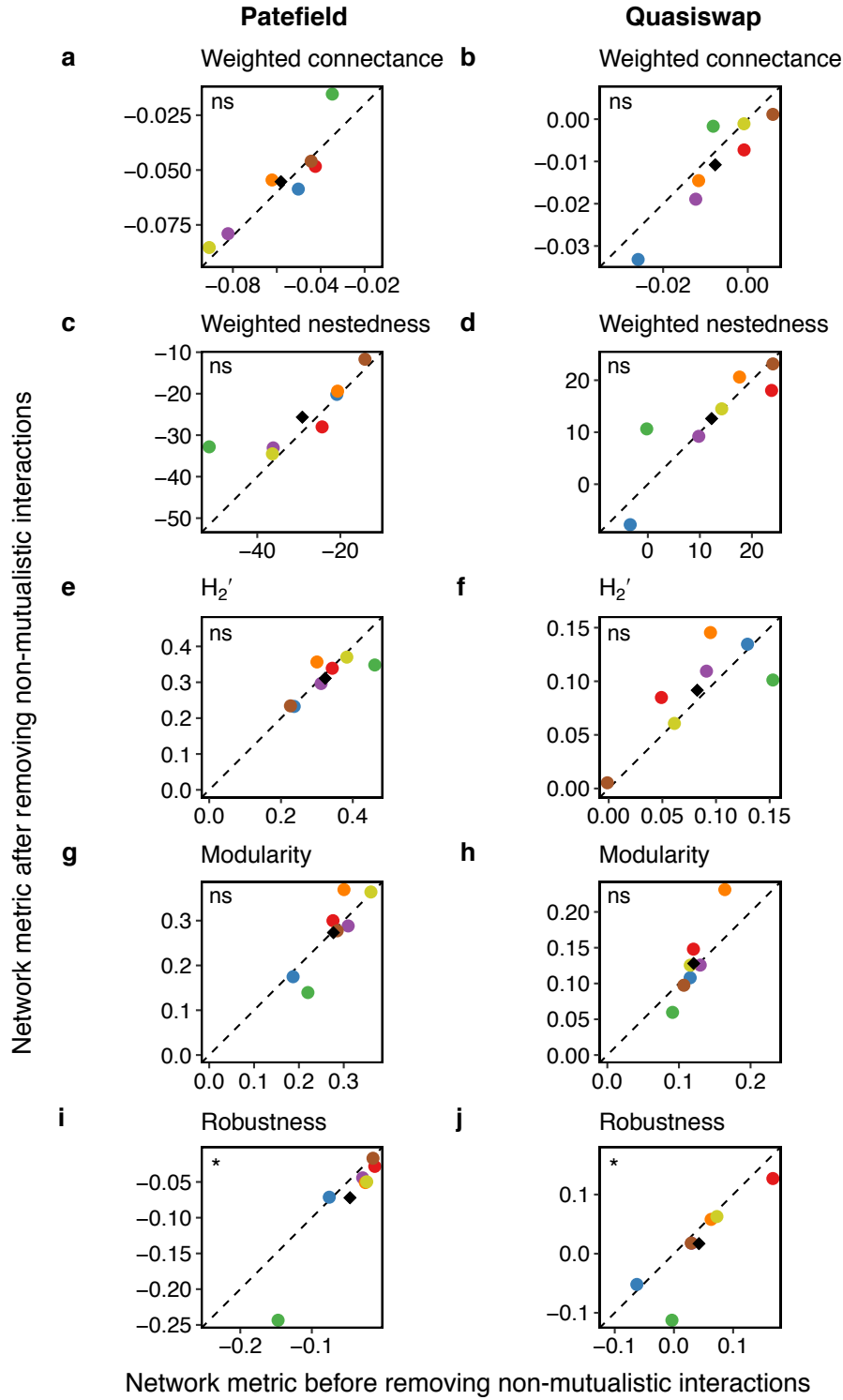
Figure S4 Changes in the null-corrected (Δ-transformed) network-level metrics after the removal of non-mutualistic interactions (seed predation and pulp pecking). Colour codes denote network identity (see Fig. 1b). The black diamonds are mean values across networks. The dashed line is *y* = *x*, indicating the position of points if there was no change in metric values. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*ns*: non-significant; * *P* < 0.05).

Table S1. Mean change and variation in weighted and unweighted connectance and nestedness following the removal of non-mutualistic interactions. Significant changes are shown in bold.

Metric	Type	Change in metric	Change in metric (%)	Range	Coefficient of variation (%)
Connectance	Unweighted	**0.05**	**15.6**	–0.01 to 0.1	86
	Weighted	**0.02**	**16.2**	0.00 to 0.04	95
Nestedness	Unweighted	2.98	12.4	–1.17 to 14.77	188
	Weighted	**4.77**	**15.0**	–1.72 to 20.33	152

Appendix S1: Removing seed predation interactions only

In this Appendix, we repeated our analyses from the main text, but removing only seed predation interactions (leaving pulp-pecking and seed dispersal interactions). At the network level, changes were similar to removing all non-mutualistic interactions, though the magnitude of changes was smaller (Fig A1.1; Table A1.1). Size and H_2' decreased significantly when predatory interactions were removed, while weighted connectance and weighted nestedness significantly increased (Table A1.1). Change in H_2' and modularity were significantly related to the proportion of seed predator interactions that were removed from a network (Fig. A1.2), which likely reflects the more modular and specialised architecture of antagonistic systems. Weighted connectance, weighted nestedness and H_2' all significantly decreased even when corrected using null models, indicating that changes in these metrics were not solely driven by network size (Fig A1.3). This again is likely due to the contrasting structure of mutualistic and antagonistic systems. At the species level, changes were in the same direction as when all non-mutualistic interactions were removed but of smaller magnitude (Fig. A1.4; Table A1.2). Plant degree, interaction frequency, d' and Resilience$_{75}$ all decreased significantly, while frugivore species strength significantly increased. Changes were generally consistent across networks, though some metrics did not change significantly in some networks, despite having significant changes overall across all networks (Table A1.3).
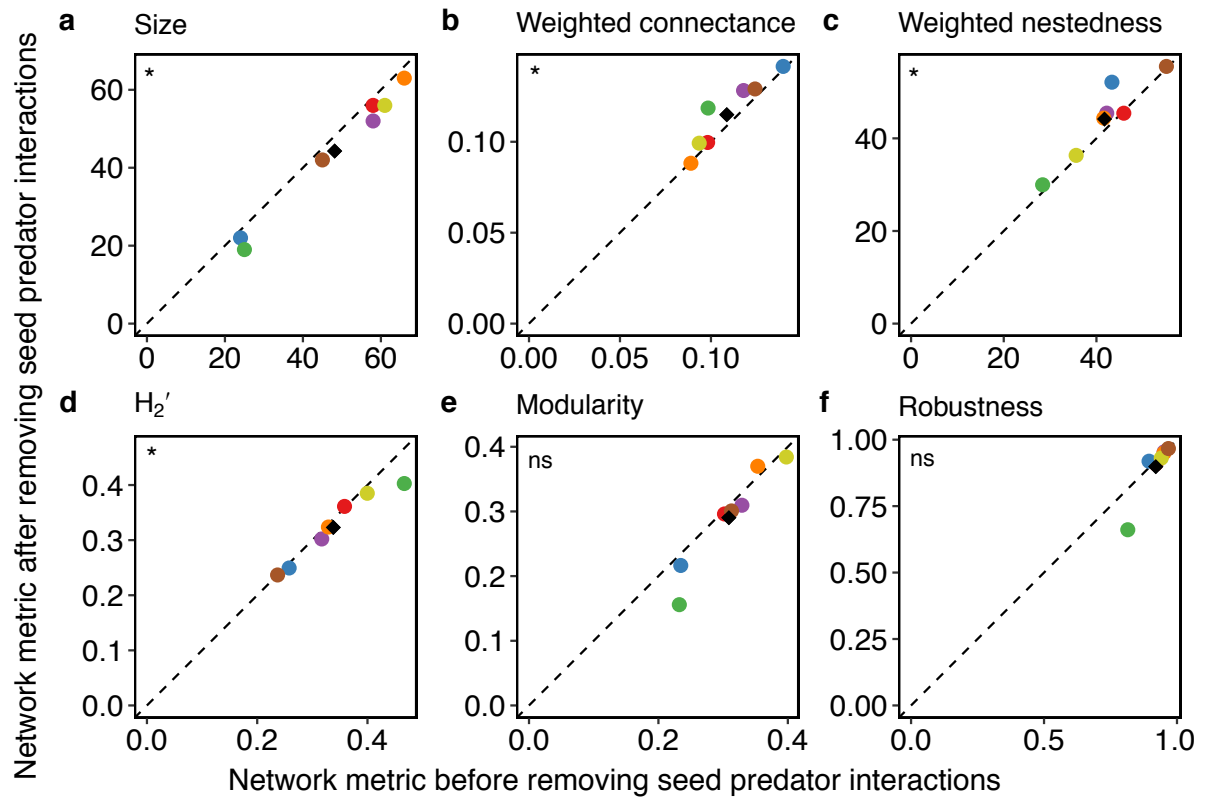
Figure A1.1. Changes in the studied network-level metrics after the removal of seed predator interactions. Colour codes denote network identity (see Fig. 1b). The black diamonds are mean values across networks. The dashed line is $y = x$, indicating the position of points if there was no change in metric values. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*ns*: non-significant; * $P < 0.05$). Unless specified, all Spearman's ρ are significant ($\rho \geq 0.89$, $P < 0.05$); we consider a non-significant ρ to indicate a change in the ranks across networks.
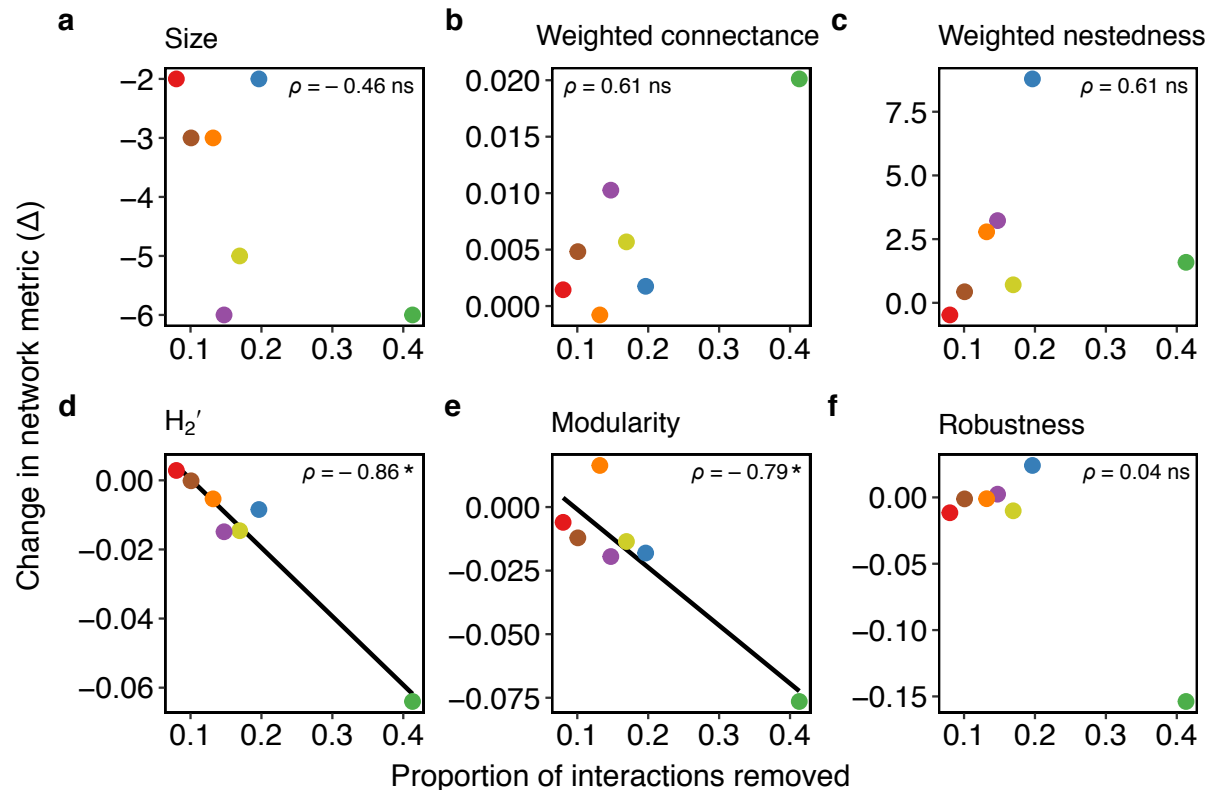
Figure A1.2. Relationship between the difference (Δ) in the network-level metrics after the removal of seed predator interactions, and the proportion of interactions removed in the original networks. For illustrative purposes, we show a regression line in significant Spearman's rank correlations (*ns*: non-significant; * $P < 0.05$).
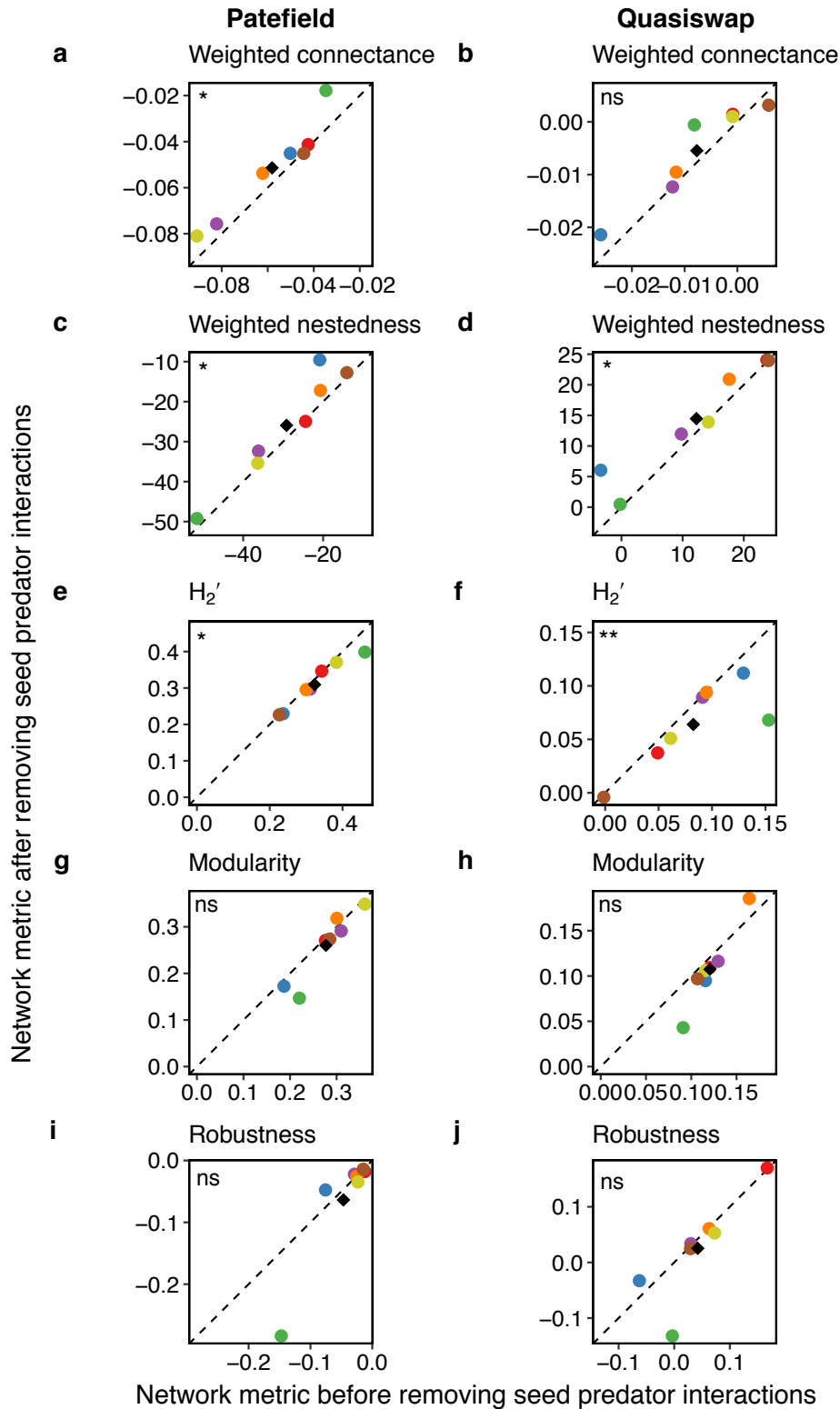
Figure A1.3. Changes in the null-corrected (Δ-transformed) network-level metrics after the removal of seed predator interactions. Colour codes denote network identity (see Fig. 1b). The black diamonds are mean values across networks. The dashed line is $y = x$, indicating the position of points if there was no change in metric values. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*ns*: non-significant; * $P < 0.05$; ** $P < 0.01$).
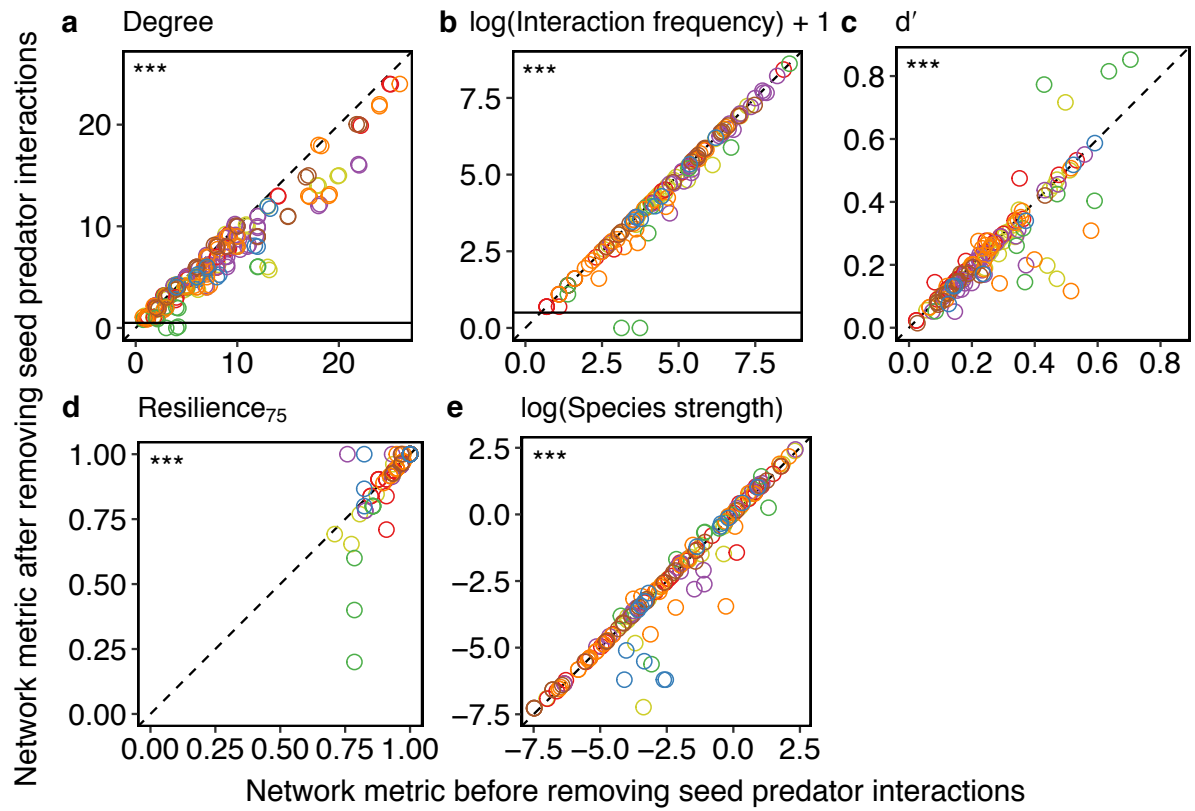
Figure A1.4. Changes in species-level metrics for plants (a-d) and frugivores (e) after the removal of seed predator interactions. Colour codes denote network identity (see Fig. 1b). The dashed line is $y = x$, indicating the position of points if there was no change in metric values. Points below the horizontal black lines in panels (a) and (b) highlight those species that lose all their partners (a: degree) and interactions (b: frequency) after pruning. The significance of Wilcoxon matched-pairs tests is shown in the top-left corner of the panels (*** $P < 0.001$).

Table A1.1. Mean change and variation in network-level metrics following the removal of non-mutualistic interactions. Significant changes are shown in bold.

Metric	Change in metric	Change in metric (%)	Range	Coefficient of variation (%)
Size	**–3.86**	**–9.4**	–6 to –2	46
Weighted connectance	**0.01**	**5.8**	0 to 0.02	115
Weighted nestedness	**2.44**	**6.0**	–0.48 to 8.79	127
H_2'	**–0.01**	**–3.7**	–0.06 to 0	152
Modularity	–0.02	–7.3	–0.08 to 0.02	153
Robustness	–0.02	–2.6	–0.15 to 0.02	276

Table A1.2. Changes and variation in species-level metrics following the removal of predatory interactions. The mean change in each metric for each network was calculated. An overall mean was obtained by calculating the mean of the mean changes in each network. The range of the mean change across networks is also reported as well as the range of change across species in parentheses. The coefficient of variation was calculated across all species in all networks; in parentheses we show the range of coefficients of variation when calculated for each network separately.

Metric	Mean (absolute)	Range	Coefficient of variation (%)
Degree (plants)	–1.16	–1.73 to –0.48 (–7 to 0)	138 (89 to 154)
Interaction frequency (plants)	–25.38	–57.38 to –4.60 (–476 to 0)	288 (86 to 308)
d' (plants)	–0.02	–0.04 to 0.01 (–0.40 to 0.34)	421 (96 to 3092)
Resilience$_{75}$ (plants)	–0.02	–0.15 to 0.03 (–0.59 to 0.24)	689 (131 to 14986)
Species strength (frugivores)	0.02	–0.07 to 0.09 (–2.46 to 1.31)	1056 (307 to 3778)

Table A1.3. Results of species-level Wilcoxon tests per network for each metric; '+' indicates that the metric increased following the removal of predatory interactions, while '–' indicates a decrease. *, ** and *** denote $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively (ns: non-significant differences).

Metric	Change	I	II	III	IV	V	VI	VII
Degree (plants)	–	**	*	**	***	***	***	*
Interaction frequency (plants)	–	**	*	**	***	***	***	*
d' (plants)	–	ns	**	ns	***	*	***	***
Resilience$_{75}$ (plants)	–	*	ns	*	ns	**	*	ns
Species strength (frugivores)	+	***	ns	ns	*	**	ns	***

Supplementary materials related to Chapter 2

ID	Latitude	Longitude	Bird counts	Flower counts	Interactions recorded	Site description and general location	Source
A	-19.16	-48.39	transects	transects	observation during transects	Cerrado, central Brazil	Maruyama P.K., Vizentin-Bugoni J., Oliveira G.M., Oliveira P.E., & Dalsgaard B. (2014) Morphological and spatio-temporal mismatches shape a Neotropical savanna plant-hummingbird network. Biotropica, 46, 740–747.
B	-23.28	-45.05	transects	transects	focal observation	Montane Atlantic forest, SE Brazil	Vizentin-Bugoni, J., Maruyama, P.K., Debastiani, V.J., Duarte, L.D.S., Dalsgaard, B. & Sazima, M. (2016). Influences of sampling effort on detected patterns and structuring processes of a Neotropical plant-hummingbird network. J. Anim. Ecol., 85, 262–272.
C	-3.82	-70.27		transects	focal observation	Amazonian rainforest, SE Colombia	Cotton P.A. (1998) The hummingbird community of a lowland Amazonian rainforest. *Ibis*, **140**, 512-521
D	19.23	-98.97	mist netting	transects	focal observation	Highland temperate forest, Mexico	Lara C. (2006) Temporal dynamics of flower use by hummingbirds in a highland temperate forest in Mexico. *Ecoscience*, **13**, 23-29
E	0.07	-72.45	mist netting	transects	focal observation	Tropical rainforest, Costa Rica	Rosero-Lasprilla L. (2003) *Interações planta/beija-flor em três comunidades vegetais da parte sul do Parque Nacional Natural Chiribiquete, Amazonas (Colombia)*. PhD Thesis. Universidade Estadual de Campinas, Brasil
F	-19.25	-43.52	mist netting	transects	focal observation	Rocky outcrops, central Brazil	Rodrigues L.C. & Rodrigues M. (2014) Flowers visited by hummingbirds in the open habitats of the southeastern Brazilian mountaintops: species composition and seasonality. *Brazilian Journal of Biology*, **74**, 659-676; Rodrigues, L.C. and Rodrigues, M., 2015. Floral resources and habitat affect the composition of hummingbirds at the local scale in tropical mountaintops. *Brazilian Journal of Biology*, 75(1), pp.39-48.
G	10.44	-84.01	mist netting	transects	focal observation	Tropical rainforest, Costa Rica	Maglianesi M.A., Blüthgen N., Böhning-Gaese K., & Schleuning, M. (2014) Morphological traits determine specialization and resource use in plant-hummingbird networks in the neotropics. *Ecology*, **95**, 3325-3334
H	10.27	-84.08	mist netting	transects	focal observation	Tropical rainforest, Costa Rica	Maglianesi M.A., Blüthgen N., Böhning-Gaese K., & Schleuning, M. (2014) Morphological traits determine specialization and resource use in plant-hummingbird networks in the neotropics. *Ecology*, **95**, 3325-3334
I	10.18	-84.11	mist netting	transects	focal observation	Tropical rainforest, Costa Rica	Maglianesi M.A., Blüthgen N., Böhning-Gaese K., & Schleuning, M. (2014) Morphological traits determine specialization and resource use in plant-hummingbird networks in the neotropics. *Ecology*, **95**, 3325-3334
J	20.13	-98.71	plots	plots	focal observation	Temperated highlands, central Mexico	Román Díaz-Valenzuela & Ortiz-Pulido, R. Unpublished data.

K	-23.32	-44.94	transects	transects	focal observation	Restinga, Atlantic forest, SE Brazil	Maruyama P.K, Vizentin-Bugoni J., Dalsgaard B., Sazima I. & Sazima. M. (2015) Nectar robbery by a hermit hummingbird: association to floral phenotype and its influence on flowers and network structure. *Oecologia*: **178**, 783-793.
L	-23.35	-44.83	transects	transects	focal observation	Secondary Atlantic forest, SE Brazil	Maruyama P.K, Vizentin-Bugoni J., Dalsgaard B., Sazima I. & Sazima. M. (2015) Nectar robbery by a hermit hummingbird: association to floral phenotype and its influence on flowers and network structure. *Oecologia*: **178**, 783-793.
M	-23.36	-44.85	transects	transects	focal observation	Coastal Atlantic forest, SE Brazil	Maruyama P.K, Vizentin-Bugoni J., Dalsgaard B., Sazima I. & Sazima. M. (2015) Nectar robbery by a hermit hummingbird: association to floral phenotype and its influence on flowers and network structure. Oecologia: 178, 783-793.
N	4.5	-75.6	mist netting	transects	focal observation	Secondary Andean forest, Colombia	Marín-Gómez, O.H. Unpublished data.
O	-24.18	-47.93	transects	transects	focal observation	Atlantic forest, SE Brazil	Rocca-de-Andrade M.A. (2006) *Recurso floral para aves em uma comunidade de Mata Atlântica de encosta: sazonalidade e distribuição vertical* . PhD Thesis. Universidade Estadual de Campinas, Brasil
P	-2.96	-79.1	mist netting	transects	focal observation	Cattle ranching (former Andean forest), Ecuador	Tinoco, B. A., Graham, C. H., Aguilar, J. M. and Schleuning, M. (2016), Effects of hummingbird morphology on specialization in pollination networks vary with resource availability. Oikos. doi: 10.1111/oik.02998
Q	-2.87	-79.12	mist netting	transects	focal observation	Highland Andean forest, Ecuador	Tinoco, B. A., Graham, C. H., Aguilar, J. M. and Schleuning, M. (2016), Effects of hummingbird morphology on specialization in pollination networks vary with resource availability. Oikos. doi: 10.1111/oik.02998
R	-2.84	-79.16	mist netting	transects	focal observation	Shrubland, Ecuador	Tinoco, B. A., Graham, C. H., Aguilar, J. M. and Schleuning, M. (2016), Effects of hummingbird morphology on specialization in pollination networks vary with resource availability. Oikos. doi: 10.1111/oik.02998
S	-19.95	-43.9	transects	transects	observation during transects	Rocky outcrops, central Brazil	Vasconcelos M.F. & Lombardi J.A. (1999) Padrão sazonal na ocorrência de seis espécies de beija-flores (Apodiformes: Trochilidae) em uma localidade de campo rupestre na Serra do Curral, Minas Gerais. *Ararajuba*, **7**, 71-79

Appendix 1: Networks used in this study.

Appendix 2

We repeated all analyses excluding four networks where we used frequency of occurrence (the proportion of days of fieldwork in which a given species was recorded) as a proxy for relative abundances, because species were not recorded within the sampling plots during transect counts or mist netting for these networks. We note that frequency of occurrence and relative abundance are strongly correlated and frequency of occurrence is still independent from the network data (Vizentin-Bugoni et al. 2014).

Results were qualitatively the same and conclusions identical after the exclusion of these networks. We confirmed the positive relationship between abundance and generalisation in our dataset, finding a significant correlation between abundance and generalisation for degree ($P = < 0.001$; pseudo-$R^2 = 0.65$), normalised degree ($P = < 0.001$; pseudo-$R^2 = 0.62$) and the generalisation index g (Wald test: $\chi^2 = 14.94$; df = 1; $P < 0.001$; $R^2_{LMM(m)} = 0.10$; $R^2_{LMM(c)} = 0.50$).

Only a small proportion of species were abundant and specialist for all three generalisation metrics, while the proportion of species that were rare and generalist was consistently larger, particularly for the g generalisation metric (Figure A1). These differences were significant. We found that abundant specialists were significantly less than rare specialists, rare generalists and abundant generalists for all generalisation metrics (Table A1). Conversely, we found that rare generalists were significantly less than rare specialists, significantly greater than abundant specialists, and not significantly different to abundant generalists, for the degree and normalised degree metrics (Table A1). For the g generalisation index we found that rare generalists were not significantly different to rare specialists, and were significantly greater than abundant specialists and abundant generalists (Table A1). Overall, these findings support hypothesis 1, that abundance drives generalisation, and do not support hypothesis 2, that generalisation drives abundance.
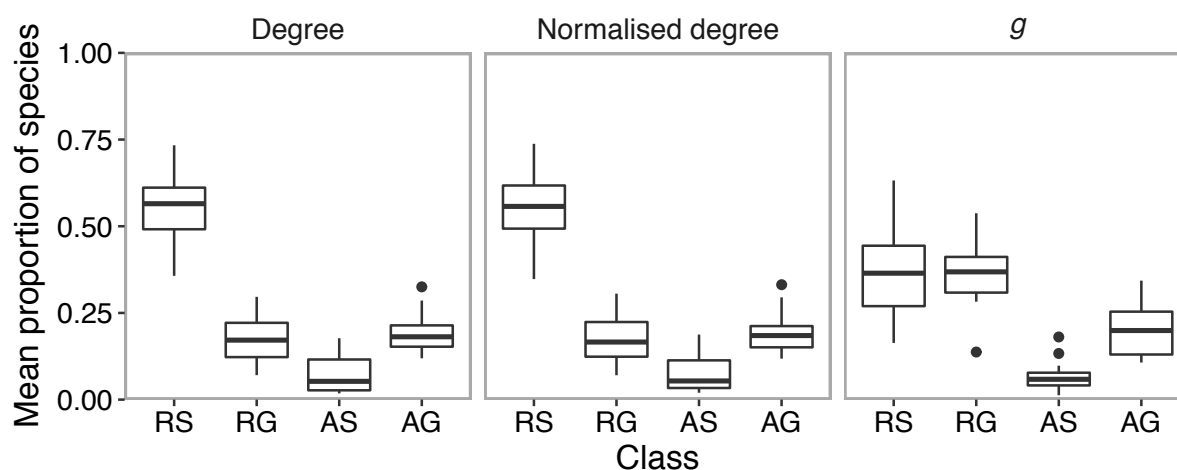

Figure A1: The mean proportion of hummingbird species classified as rare specialists ('RS'), rare generalists ('RG'), abundant specialists ('AS') and abundant generalists ('AG') across a subset of networks, for three generalisation metrics: degree, normalised degree and g. The subset excluded four networks where we used frequency of occurrence (the proportion of days of fieldwork in which a given species was recorded) as a proxy for relative abundances. The bold centre line in each box is the median; the lower and upper hinges are the first and third quartiles, respectively. The lower whisker indicates the smallest value no less than 1.5 times the inter-quartile range; the upper whisker indicates the largest value no greater than 1.5 times the inter-quartile range. Data outside the whiskers are outlying points plotted as solid black circles.

Table A1: Testing hypotheses 1 and 2 in an ANOVA framework, using abundant specialists and rare generalists as the reference contrast respectively. RS = rare specialist; RG = rare generalist; AS = abundant specialist; AG = abundant generalist. Significance codes: 0 '***' 0.001 '**', not significant 'ns'

Metric	Class	Estimate	t value	P	Significance
Hypothesis 1: Abundant specialist << rare specialist, rare generalist, abundant generalist					
Reference contrast = abundant specialist					
Degree	(Intercept)	0.08	4.15	0.00	***
	RS	0.48	18.04	0.00	***
	RG	0.09	3.46	0.00	**
	AG	0.11	4.30	0.00	***
Normalised degree	(Intercept)	0.08	4.06	0.00	***
	RS	0.48	17.36	0.00	***
	RG	0.09	3.44	0.00	**
	AG	0.11	4.15	0.00	***
g	(Intercept)	0.07	2.91	0.01	***
	RS	0.29	8.88	0.00	***
	RG	0.30	8.97	0.00	***
	AG	0.14	4.12	0.00	***
Hypothesis 2: Rare generalist << rare specialist, abundant generalist, abundant specialist					
Reference contrast = rare generalist					
Degree	(Intercept)	0.17	9.04	0.00	***
	RS	0.39	14.57	0.00	***
	AS	-0.09	-3.46	0.00	**
	AG	0.02	0.84	0.40	ns
Normalised degree	(Intercept)	0.17	8.92	0.00	***
	RS	0.38	13.92	0.00	***
	AS	-0.09	-3.44	0.00	**
	AG	0.02	0.71	0.48	ns
g	(Intercept)	0.36	15.59	0.00	***
	RS	0.00	-0.08	0.94	ns
	AS	-0.30	-8.97	0.00	***
	AG	-0.16	-4.84	0.00	***

The proportion of species in each of the four abundance-generalisation categories predicted by the neutrality null model closely matched the empirical proportions, particularly for degree and normalised degree where there were no significant differences between observed and predicted proportions for the majority of networks (67–87% of networks; Fig. A2). For *g*, the model correctly predicted the proportion of rare specialists and generalists for 80% of networks, but performed less well in predicting the proportion of abundant specialists and generalists, with predictions matching observed values for only 53% of networks (Fig. A2).
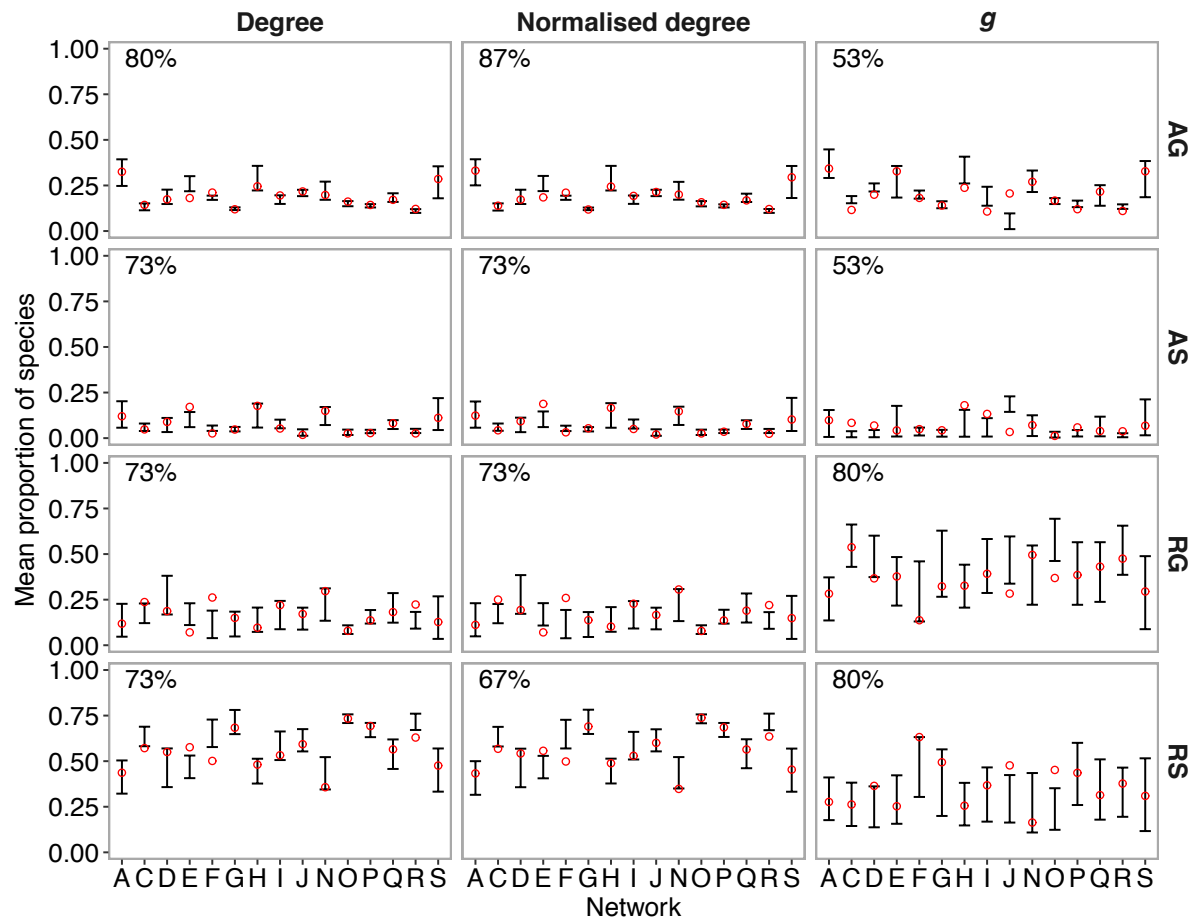
Figure A2: Comparisons between empirical networks and null model networks in the proportions of species in each of the abundance-generalisation categories 'RS' (rare specialists), 'RG' (rare generalists), 'AS' (abundant specialists) and 'AG' (abundant generalists). Four networks (B, K, L, M) where we used frequency of occurrence (the proportion of days of fieldwork in which a given species was recorded) as a proxy for relative abundances were excluded. Error bars represent the 95% confidence intervals of the mean proportion of hummingbird species in each abundance-generalisation category as predicted by 1000 null networks. Red circles show the empirically observed mean proportion of hummingbird species in each category. If the red circle is within the error bars, there were no significant differences between the observed proportions and the neutrality null model proportions. Percentages in the top left of each panel give the proportion of networks where empirical proportions were not significantly different from the null model proportions. Results are shown for each network (A-S, excluding B, K, L and M) and for each generalisation metric (Degree, Normalised degree, *g*).

References

Vizentin-Bugoni, J. et al. 2014. Processes entangling interactions in communities: forbidden links are more important than abundance in a hummingbird-plant network. - Proc. R. Soc. B Biol. Sci. 281: 20132397.

Supplementary materials related to Chapter 3

Figure S1: Relationship between link vulnerability and importance ($\rho = 0.01$)
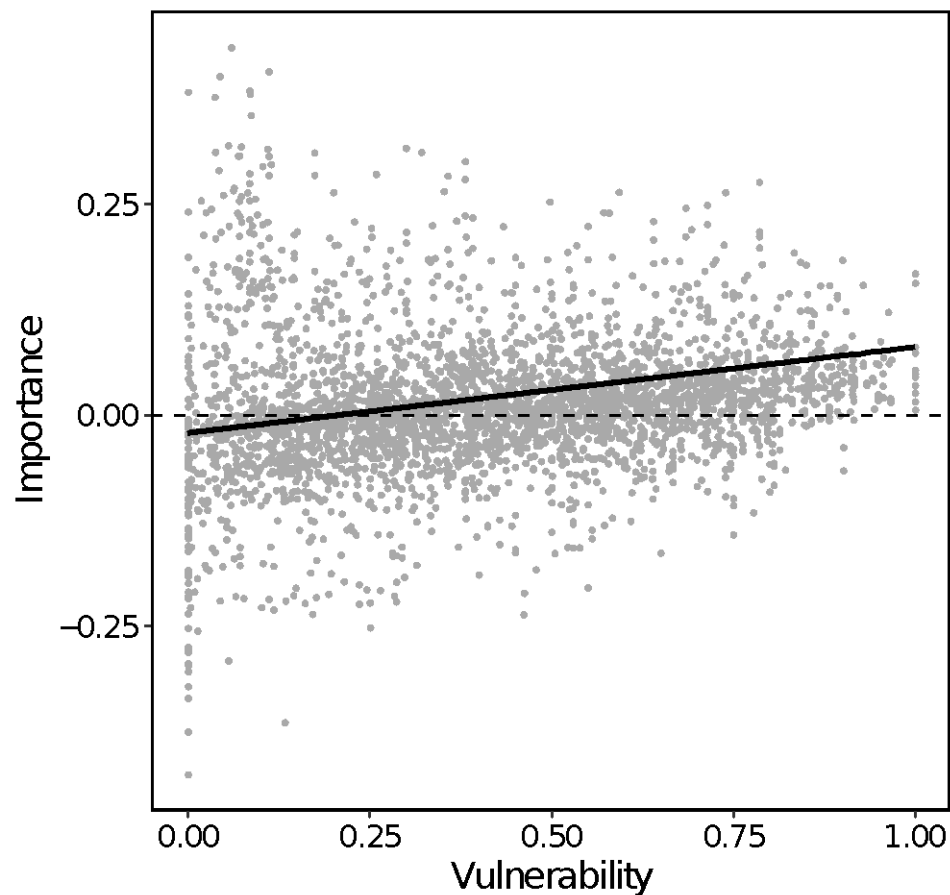

Figure S1: The relationship between vulnerability (the likelihood of a link being lost) and importance (the contribution of a link to a network's structural stability) for all species-species links across 41 mutualistic networks. Best fit line is from a mixed effects model with importance as the response variable, vulnerability as a fixed effect, and network identity as a random effect.

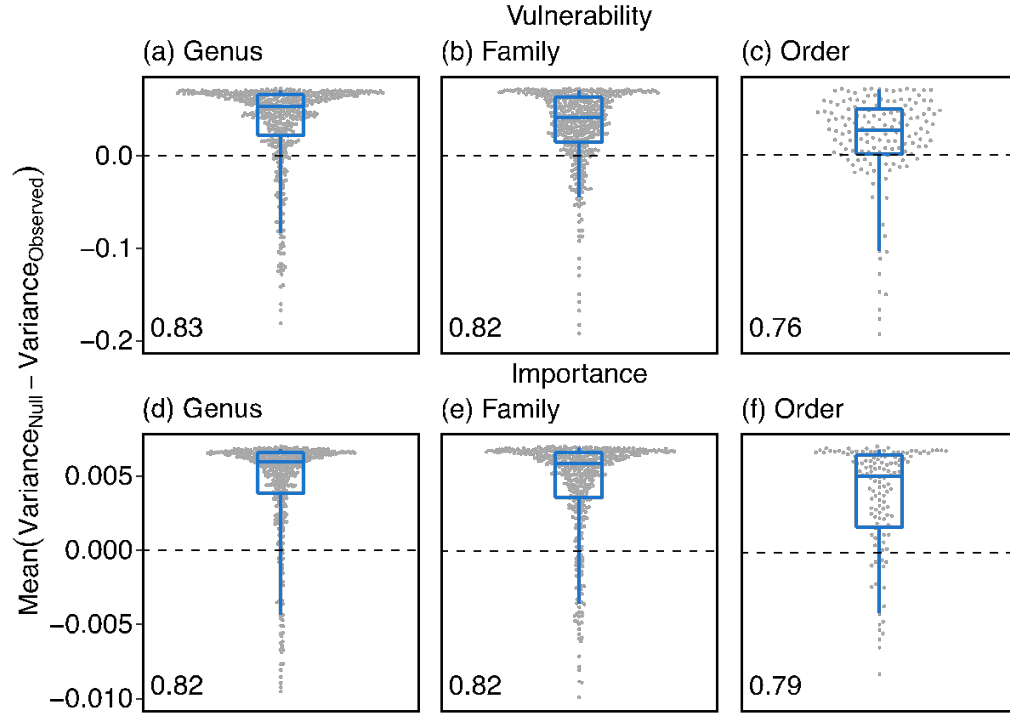# Figure S2: Taxonomic consistency of vulnerability and importance ($\rho = 0.01$)

Figure S2: The degree of taxonomic consistency for each interaction at each taxonomic level, for both vulnerability (likelihood of a link being lost) and importance (contribution of a link to a network's structural stability). Taxonomic consistency is the tendency for properties of an interaction to be more similar across occurrences than expected by chance. Points represent individual interactions. Boxplots represent 5%, 25%, 50%, 75% and 95% quantiles of the same data, moving from the bottom whisker to the top whisker. Number in bottom left of each panel is the proportion of interactions which exhibited positive consistency ($\text{Variance}_{\text{Observed}} < \text{Variance}_{\text{Null}}$). For visualisation, a small number of points with low values were removed. The percentage of points with values lower than the y-axis minimum are as follows for each panel: (a) 1.5%, (b) 1.1%, (d) 7.2%, (e) 6%, (f) 5.3%.

Supplementary materials related to
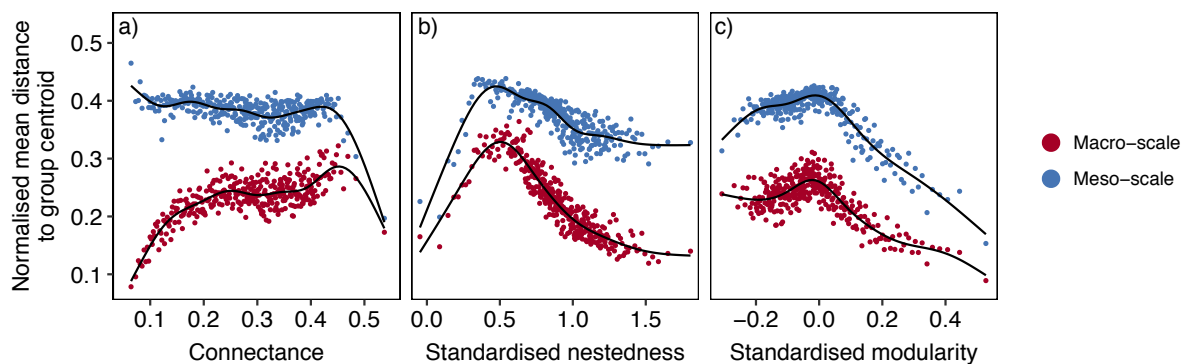
Chapter 4

Appendix 1: Null model version of 'Comparing indices and motifs' analysis

Many network properties, and motif counts, can vary with network size and connectance. We therefore repeated our 'Comparing indices and motifs' analysis, using values for nestedness, modularity and motif counts that were standardised relative to a null expectation. Specifically, we used a null model previously used to study nestedness and modularity (Bascompte et al. 2003, Thébault and Fontaine 2010). In the null model, the probability of two species interacting is the average of the normalised degree of both species (Bascompte et al. 2003, Thébault and Fontaine 2010). This means that the probability of an interaction is proportional to the generalisation of the two species.

For each of the 20,000 networks generated by the bipartite cooperation model, we generated 1000 null networks. We then expressed each structural measure (nestedness, modularity and motif counts) as a relative measure: $M^* = (M - \overline{M_R})/\overline{M_R}$, where M^* is the relative value of a given measure, M is the value of the measure in the original network, and M_R is the mean value of the measure across the 1000 null networks (Bascompte et al. 2003, Thébault and Fontaine 2010).

We found that our results were qualitatively identical to those found when structural measures were not standardised: at all values of connectance, standardised nestedness and standardised modularity, meso-scale variation was greater than macro-scale variation (Supplementary Figure 1).

The median paired differences in variation between macro- and meso-scale subsets were 0.14 when ranked by connectance (Wilcoxon: p < 0.0001), 0.15 when ranked by nestedness (Wilcoxon: p < 0.0001), and 0.15 when ranked by modularity (Wilcoxon: p < 0.0001). For connectance, nestedness and modularity as the ranking variable respectively, the motif framework captured 69%, 65% and 70% more variation in indirect interactions than traditional whole-network indices.
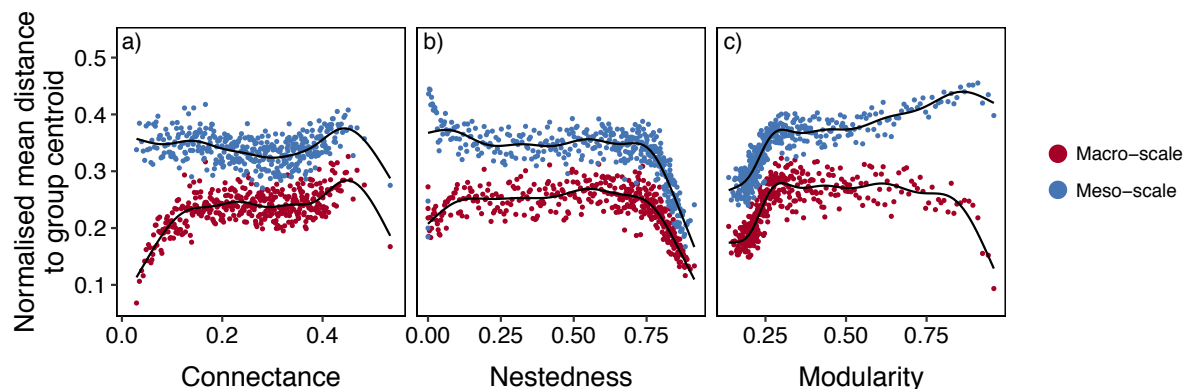

Supplementary Figure 1: : Network variation (normalised mean distance to group centroid) against mean connectance (a), nestedness (b) and modularity (c) for all networks. Points represent subsets of networks.

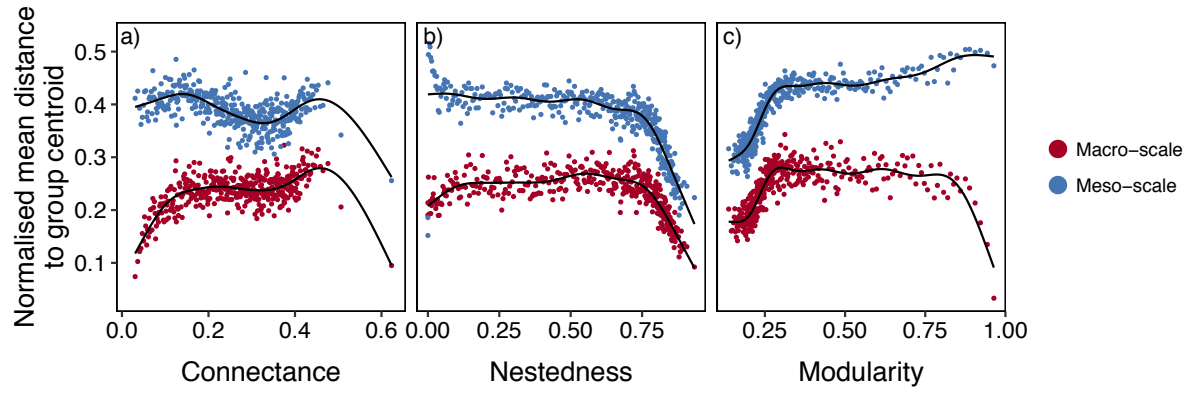# Appendix 2: Comparing indices to motifs from one size class

Motifs can be nested within each other. For example motif 3 is in motif 4, motif 4 is in motif 8, and motif 8 is in motif 18. Motifs nested in other motifs may mean that the same type of information on indirect interactions is described multiple times in a network's motif structural signature. To ensure this effect did not inflate the importance of indirect interactions, we repeated our 'Comparing indices to motifs' analysis, but only used motifs of a given size. Specifically, we compared macro-scale indices to 5-species motifs only (Supplementary Figure 2) and 6-species motifs only (Supplementary Figure 3), in two separate analyses. We found that our results were qualitatively unchanged, and our conclusions identical, by only considering motifs of a single size class, and we therefore conclude that any inflation of the importance of indirect interactions that may have occurred by considering motifs nested within larger motifs does not affect our conclusions.

When comparing indices to 5-species motifs only (Supplementary Figure 2), the median paired differences in variation between macro- and meso-scale subsets were 0.10 when ranked by connectance (Wilcoxon: $p < 0.0001$), 0.09 when ranked by nestedness (Wilcoxon: $p < 0.0001$), and 0.10 when ranked by modularity (Wilcoxon: $p < 0.0001$). For connectance, nestedness and modularity as the ranking variable respectively, the motif framework captured 47%, 41% and 47% more variation in indirect interactions than traditional whole-network indices.

Supplementary Figure 2: : Network variation (normalised mean distance to group centroid) against mean connectance (a), nestedness (b) and modularity (c) for all networks. Points represent subsets of networks. Meso-scale vectors comprised 5-species motifs only.

When comparing indices to 6-species motifs only (Supplementary Figure 3), the median paired differences in variation between macro- and meso-scale subsets were 0.14 when ranked by connectance (Wilcoxon: $p < 0.0001$), 0.14 when ranked by nestedness (Wilcoxon: $p < 0.0001$), and 0.15 when ranked by modularity (Wilcoxon: $p < 0.0001$). For connectance, nestedness and modularity as the ranking variable respectively, the motif framework captured 69%, 60% and 69% more variation in indirect interactions than traditional whole-network indices.

Supplementary Figure 3: : Network variation (normalised mean distance to group centroid) against mean connectance (a), nestedness (b) and modularity (c) for all networks. Points represent subsets of networks. Meso-scale vectors comprised 6-species motifs only.

Appendix 3: Supplementary methods for sampling effort analysis

We used simulations to assess the effect of sampling effort on network structural signatures and species role signatures. Sampling simulations were divided into two stages. In the first stage, a proportion, p, of plant species were sampled, with higher degree plant species having a higher probability of being sampled (a supplementary analysis was also conducted where plant species were chosen randomly; see Supplementary Figure 4). The number of sampled plant species was equal to pf where f is the number of plant species. If pf was not an integer, it was rounded to the nearest integer. For the analysis of species roles, it was necessary to ensure that the focal species, s, remained in the network (i.e. that it interacted with one or more partners). If s was a plant, we included it in the sample, then sampled $pf - 1$ plants from all remaining plant species. If s was an animal, we sampled one of s's plant partners either in proportion to their degree for the main analysis, or randomly for the supplementary analysis. We then sampled $pf - 1$ plants from all remaining plant species.

In the second stage, a proportion, q, of each plant species' interactions were sampled with higher strength interactions having a higher probability of being sampled. The number of sampled interactions for a given plant species, w, was equal to ql_w where l_w is the degree of species w. If ql_w was not an integer, it was rounded to the nearest integer. If ql_w was 0, it was set to 1 to ensure all sampled plant species were present in the network. Again, for the analysis of species roles, it was necessary to ensure that s had at least one interaction with another species. If s was a plant species, the above process already ensured this. If s was an animal species, we randomly chose one of its plant partners, a, to have a minimum of one interaction with s. For a, we then sampled $ql_a - 1$ interactions from all remaining animal species. For all other plant species, we sampled ql_w of their interactions as described above.
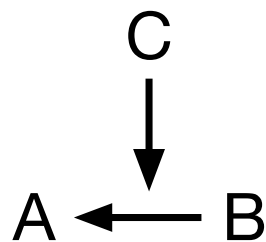
We performed simulations on 40 empirical pollination and seed dispersal networks (Supplementary Table 1) for different values of p and q between 0.5 and 1. We carried out 1000 randomisations for each p and q combination. We measured R^2 between the network structural signature or species role signatures of the sampled network and the network structural signature or species role signatures of the corresponding 'true' network which contained all species and interactions. Network structural signatures were normalised as a proportion of the maximum number of species combinations that could occur in each motif (Poisot and Stouffer 2016), while species role signatures were normalised within each motif size class (Baker et al. 2015).
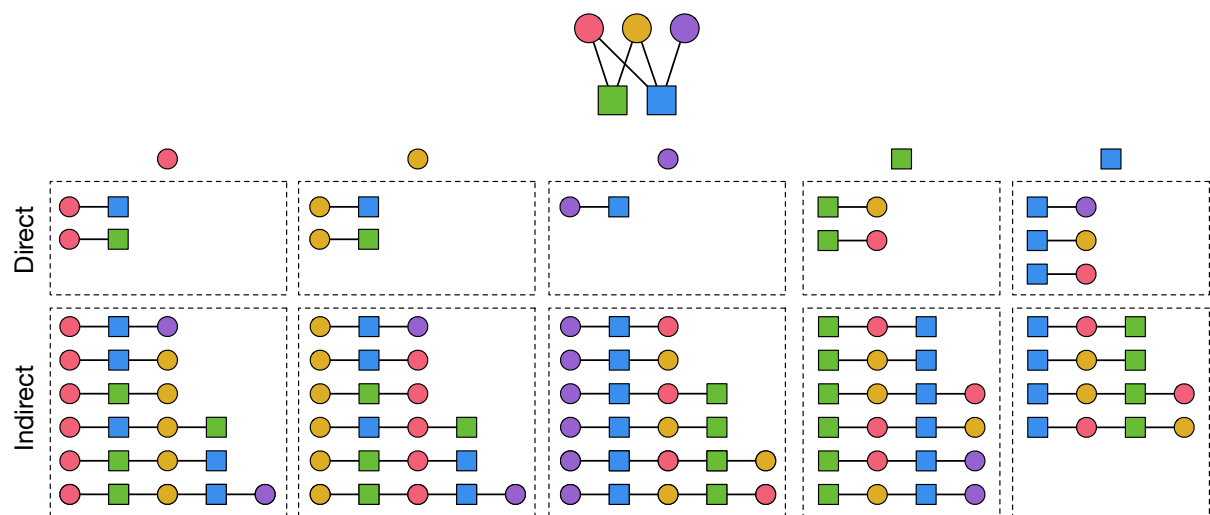
Other supplementary figures and tables

a)

b)

Supplementary Figure 4: Two mechanisms through which species can indirectly influence each other. (a) An interaction chain, where species Z influences species A through a series of direct interactions involving changes in abundance of one or more intermediary species, such as B. (b) An interaction modification, where species C influences species A by changing the interaction between species A and B. Adapted from Wootton (1994).
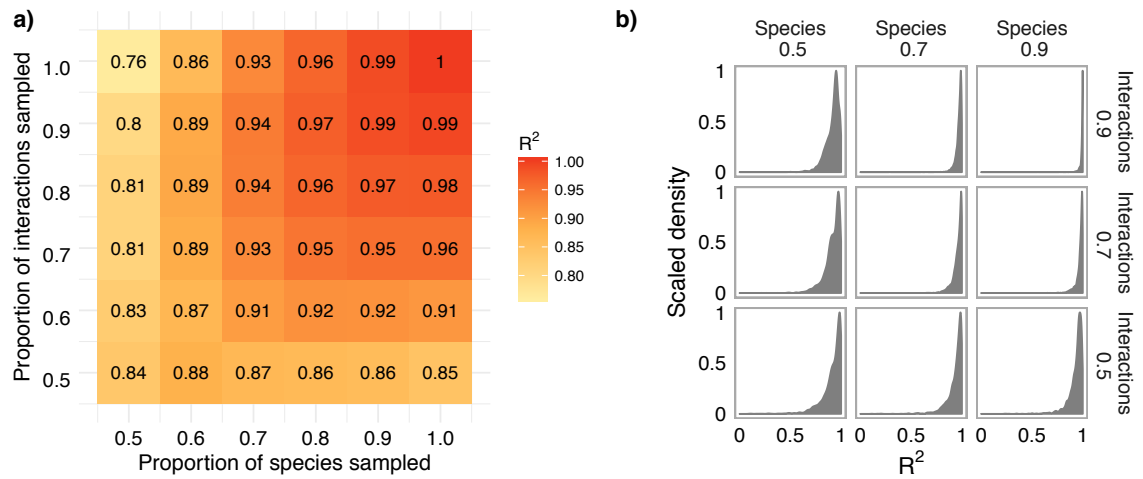

Supplementary Figure 5: A small five-species motif and all the possible direct and indirect interaction chains embedded within it. Each column indicates the donor species.

Supplementary Table 1: Properties of the 40 networks used to assess the effects of sampling on motifs (www.web-of-life.es); NODF (Almeida-Neto et al. 2008)
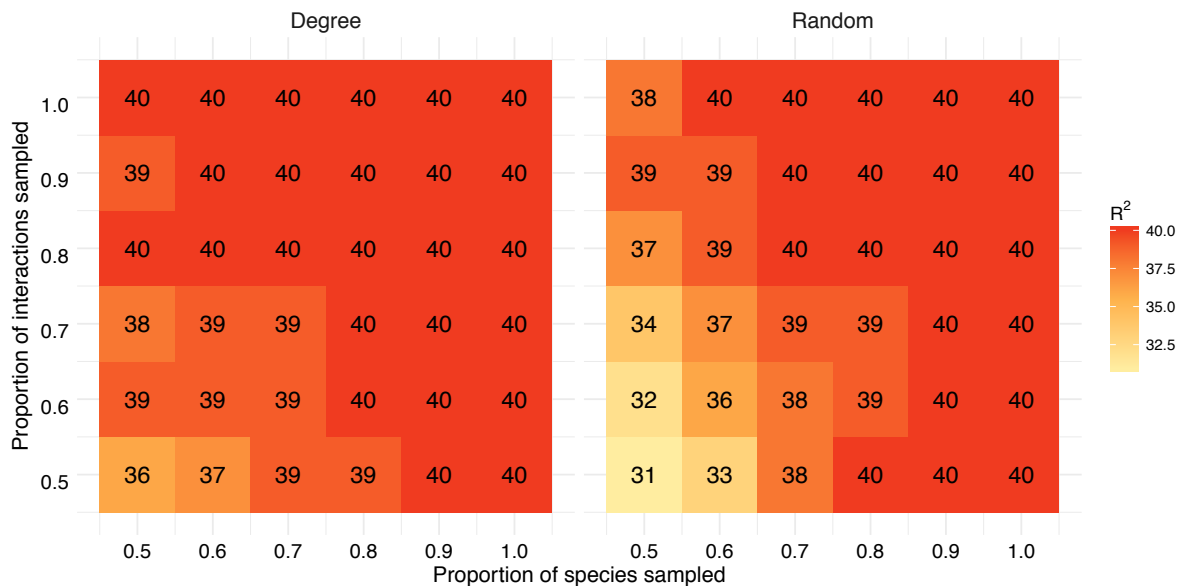
Network	Interaction	Species	Links	Connectance	NODF
M_PL_004	Pollination	114	167	0.14	28.15
M_PL_006	Pollination	78	146	0.14	44.58
M_PL_007	Pollination	52	85	0.15	31.54
M_PL_013	Pollination	65	103	0.2	34.25
M_PL_017	Pollination	104	299	0.15	40.37
M_PL_019	Pollination	125	264	0.08	17.51
M_PL_024	Pollination	29	38	0.19	29.02
M_PL_025	Pollination	57	143	0.25	46.02
M_PL_033	Pollination	47	141	0.32	29.5
M_PL_040	Pollination	72	114	0.09	15.18
M_PL_041	Pollination	74	145	0.11	25.3
M_PL_045	Pollination	43	63	0.14	30.77
M_PL_051	Pollination	104	164	0.13	26.96
M_PL_054	Pollination	431	773	0.02	8.08
M_PL_055	Pollination	259	431	0.03	8.71
M_PL_056	Pollination	456	871	0.03	6.86
M_PL_058	Pollination	113	319	0.12	26.64
M_PL_059	Pollination	26	71	0.42	76.88
M_PL_063	Pollination	59	155	0.38	45.73
M_PL_064	Pollination	22	32	0.29	41.32
M_PL_066	Pollination	37	104	0.56	50.3
M_PL_068	Pollination	40	83	0.3	45.69
M_PL_069_01	Pollination	20	22	0.26	19.18
M_PL_070	Pollination	16	41	0.64	62.23
M_PL_071	Pollination	52	89	0.25	38.53
M_SD_001	Seed dispersal	28	50	0.34	40.77
M_SD_002	Seed dispersal	40	119	0.43	62.16
M_SD_003	Seed dispersal	41	68	0.17	41.09
M_SD_004	Seed dispersal	54	95	0.14	39.82
M_SD_005	Seed dispersal	38	49	0.15	27.93
M_SD_006	Seed dispersal	36	51	0.16	32.79
M_SD_008	Seed dispersal	26	110	0.69	56.33
M_SD_009	Seed dispersal	25	38	0.3	33.02
M_SD_010	Seed dispersal	64	234	0.33	42.13
M_SD_012	Seed dispersal	64	146	0.14	33.04
M_SD_020	Seed dispersal	58	150	0.18	53.55
M_SD_023	Seed dispersal	23	38	0.32	66.8
M_SD_031	Seed dispersal	44	61	0.24	27.7
M_SD_033	Seed dispersal	24	53	0.37	67.45
M_SD_034	Seed dispersal	121	419	0.14	32.35

Supplementary Table 2: Properties of the monthly networks used to compare the roles of *Lassioglossum magense* and *Apis mellifera* in four mountaintop plant-pollinator communities sampled over eight consecutive months between September 2012 and April 2013 in the Seychelles (Kaiser-Bunbury et al. 2017). Nestedness was calculated as NODF (Almeida-Neto et al. 2008).
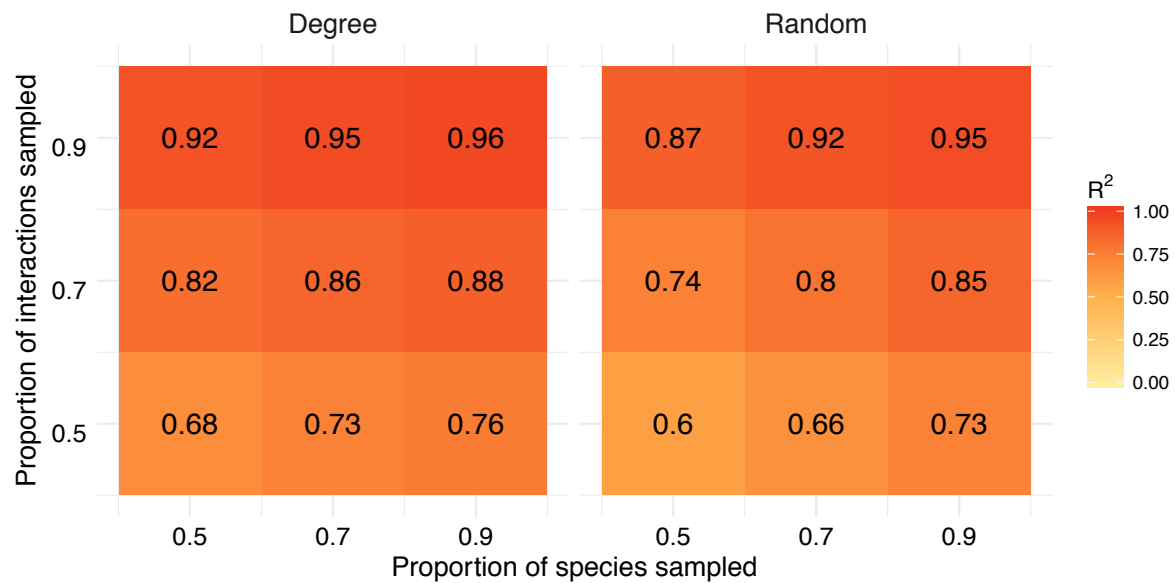
Site	Month	Species	Interactions	Connectance	Nestedness
Bernica	1	25	31	0.27	22.31
Bernica	2	30	31	0.19	11.53
Bernica	3	18	22	0.29	19.3
Bernica	4	31	39	0.2	14.34
Bernica	5	29	41	0.24	20.06
Bernica	6	22	28	0.27	20.63
Bernica	7	27	37	0.24	21.23
Bernica	8	21	27	0.3	22.08
Salazie	1	29	33	0.24	21.3
Salazie	2	23	31	0.3	20.91
Salazie	3	24	29	0.27	20.93
Salazie	4	31	46	0.27	19.48
Salazie	5	33	42	0.21	15.49
Salazie	6	29	36	0.26	14.89
Salazie	7	27	32	0.21	20.7
Salazie	8	14	14	0.42	16.38
Tea Plantation	1	25	33	0.26	21.36
Tea Plantation	2	27	33	0.22	10.89
Tea Plantation	3	43	60	0.2	15.19
Tea Plantation	4	31	42	0.21	15.53
Tea Plantation	5	25	31	0.23	13.72
Tea Plantation	6	35	55	0.25	19.77
Tea Plantation	7	33	45	0.22	14.98
Tea Plantation	8	32	43	0.21	18.11
Trois Freres	1	18	18	0.23	16.12
Trois Freres	2	28	35	0.24	16.2
Trois Freres	3	30	37	0.2	14.02
Trois Freres	4	29	40	0.24	18.35
Trois Freres	5	45	71	0.19	13.6
Trois Freres	6	28	38	0.22	22.95
Trois Freres	7	34	46	0.2	14.27
Trois Freres	8	17	21	0.32	33.57

Supplementary Figure 6: Results of simulations assessing the sensitivity of motifs to variation in sampling effort when plant species had a random probability of being sampled. (a) The mean R^2 between the structural signatures of the sampled networks and the structural signatures of their corresponding 'true' networks across all 40 networks for different levels of species and interaction removal. (b) Distribution of mean R^2 between species role signatures in sampled networks and species role signatures in their corresponding 'true' networks for all species across all 40 networks for different levels of species and interaction removal.

Supplementary Figure 7: The number of networks used for each level of species and interaction removal in the sampling analysis of structural signatures. Some combinations of species and interaction removal resulted in sampled versions of some smaller networks which were too small or disconnected to contain larger motifs. For these sampled networks we could not calculate a structural signature which was comparable to that of the 'true' network and therefore these networks were excluded. Each cell shows the number of networks for which we were able to calculate the mean R^2 in the corresponding cells of Figure 5a (when plant species with higher degree had a higher probability of being sampled) and Supplementary Figure 4a (when plant species were sampled randomly).

Supplementary Figure 8: The proportion of species used for each level of species and interaction removal in the sampling analysis of species roles. Some species in sampled versions of smaller networks did not appear in larger motifs, particularly at high levels of species and interaction removal. For these species we could not calculate a species role signature which was comparable to that of the 'true' role and therefore these species were excluded. Each cell shows the proportion of species for which we were able to calculate the mean R^2 in the corresponding panels of Figure 5b (when plant species with higher degree had a higher probability of being sampled) and Supplementary Figure 4b (when plant species were sampled randomly).

References

Almeida-Neto, M. et al. 2008. A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. - Oikos 117: 1227–1239.

Baker, N. J. et al. 2015. Species' roles in food webs show fidelity across a highly variable oak forest. - Ecography (Cop.). 38: 130–139.

Bascompte, J. et al. 2003. The nested assembly of plant--animal mutualistic networks. - Proc. Natl. Acad. Sci. 100: 9383–9387.

Kaiser-Bunbury, C. N. et al. 2017. Ecosystem restoration strengthens pollination network resilience and function. - Nature 542: 223–227.

Poisot, T. and Stouffer, D. 2016. How ecological networks evolve. - bioRxiv in press.

Thébault, E. and Fontaine, C. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. - Science 329: 853–6.

Wootton, J. T. 1994. The Nature and Consequences of Indirect Effects in Ecological Communities. - Annu. Rev. Ecol. Syst. 25: 443–466.

Supplementary materials related to Chapter 5

bmotif: a package for motif analyses of bipartite networks

Benno I. Simmons[1], Michelle J. M. Sweering[1,2], Maybritt Schillinger,[1,2] Lynn V. Dicks[1,3], William J. Sutherland[1], Riccardo Di Clemente[4,5]

[1] Conservation Science Group, Department of Zoology, University of Cambridge, The David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK

[2] Faculty of Mathematics, Wilberforce Road, Cambridge CB3 0WA, UK

[3] School of Biological Sciences, University of East Anglia, Norwich NR4 7TL, UK

[4] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Massachusetts Avenue 77, MA 02139, Cambridge, USA

[5] Centre for Advanced Spatial Analysis (CASA), University College London, Gower Street, London, WC1E 6BT, UK

Corresponding authors:

Benno I. Simmons. Address: Conservation Science Group, Department of Zoology, University of Cambridge, The David Attenborough Building, Pembroke Street, Cambridge, CB2 3QZ, UK. Email: benno.simmons@gmail.com

Riccardo Di Clemente. Address: Centre for Advanced Spatial Analysis (CASA), University College London, Gower Street, London, WC1E 6BT, UK. Email: r.diclemente@ucl.ac.uk

Appendix S1: Formal definition of node positions

There are 44 bipartite motifs with up to six species. Each motif has several nodes and links. Roughly speaking, we say that two nodes are in the same node position if we can swap them without changing the structure of the graph.

To define these notions formally, we first define: An *automorphism* of a graph $G = (V, E)$ is a bijection $f : V \to V$ s.t. $u \sim v \Leftrightarrow f(u) \sim f(v) \; \forall u, v \in V$, i.e. it is a bijection that preserves the structure of links. (Intuitively, a bijection is a 1-1-mapping.) Now we say that two nodes u and v are in the same node position if there exists an automorphism f s.t. $f(u) = v$, $f(v) = u$, i.e. f swaps u and v.
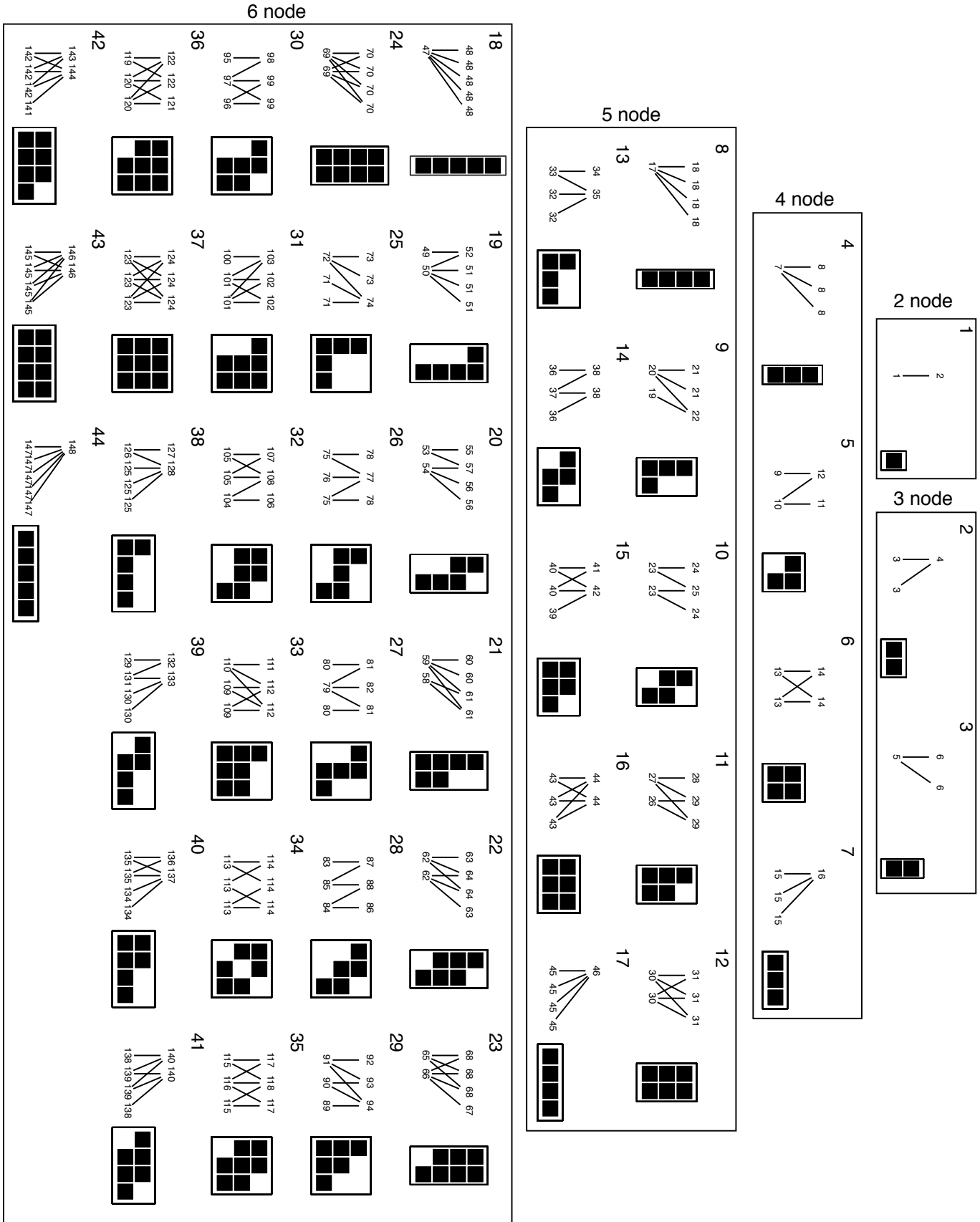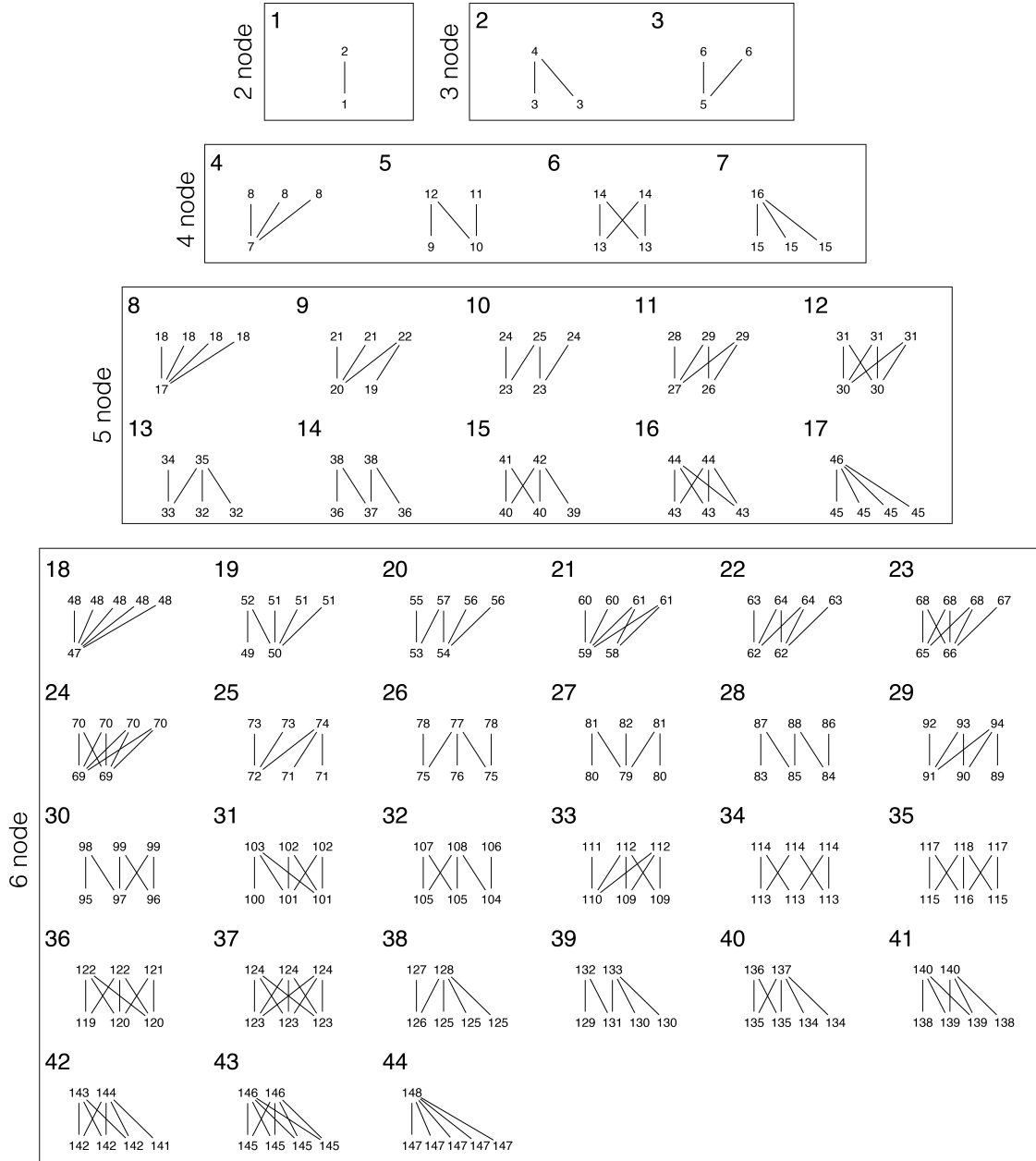
Appendix S2: Matrix representation of motifs

Figure S1: All bipartite motifs containing up to 6 nodes (species) and their corresponding representation as biadjacency matrices. Large numbers identify each motif. Small numbers represent the unique positions species can occupy within motifs, following Baker et al. (2015) Appendix 1. Lines between small numbers indicate undirected species interactions. To the right of each motif is its corresponding biadjacency matrix, **M**: black squares indicate a 1 in the matrix (the presence of an interaction), white squares indicate 0 (the absence of an interaction). There are 44 motifs containing 148 unique positions..

Appendix S3: Formulae

A reproduction of Figure 1 from the main text for convenience when reading Appendix S3. Shows all bipartite motifs containing up to 6 nodes (species). Large numbers identify each motif. Small numbers represent the unique positions species can occupy within motifs, following Baker et al. (2015) Appendix 1. Lines between small numbers indicate undirected species interactions. There are 44 motifs containing 148 unique positions..

Introduction

In the following section, we would like to present the formulae that we used in our code. The aim of the `bmotif`-package is to compute motif and position counts quickly, without using subgraph enumeration. Instead, we base the program on combinatorial considerations. We work with the biadjacency matrix and mainly use matrix multiplications to obtain the result, which makes the computation quick. The aim of this appendix is to introduce the user to the mathematical ideas in our implementation. We therefore wrote the formulae in such a way that they match our code as closely as possible. Note that this is not necessarily the most mathematically elegant formalisation; rather it is a way to understand how calculations are implemented in the `bmotif` software. In our calculations, we use several auxiliary matrices, which are defined in the following section.

Notation

In the following, $A \cdot B$ or AB denotes the matrix multiplication of A and B, whereas $A \circ B$ is the Hadamard product, which is defined as elementwise matrix multiplication for matrices of the same size. If we write $A_{ij} * B_{ij}$, we mean normal multiplication of real numbers, we use $*$ to distinguish the multiplication from the matrix multiplication.

M is the biadjacency matrix, i.e.

$$M_{zp} = \begin{cases} 1 & \text{if } z \text{ adjacent to } p \\ 0 & \text{if } z \text{ not adjacent to } p \end{cases}$$

The rows of M correspond to nodes (or species) in the first class, the columns of M correspond to nodes (or species) in the second class.
Z is the number of rows of M, i.e. the number of nodes in the first class.
P is the number of columns of M, i.e. the number of nodes in the second class.
$N_{zp} = 1 - M_{zp}$ is the complement of M, i.e.

$$N_{zp} = \begin{cases} 0 & \text{if } z \text{ adjacent to } p \\ 1 & \text{if } z \text{ not adjacent p} \end{cases}$$

$d_z = \sum_{p=1}^{P} M_{zp}$ defines a vector of length Z, giving the degree of each row
$u_p = \sum_{z=1}^{Z} M_{zp}$ defines a vector of length P, giving the degree of each column

$\mathcal{Z}_{zz'} = \sum_{p=1}^{P} M_{zp} M_{z'p} = (MM^T)_{zz'}$
$\mathcal{Y}_{zz'} = \sum_{p=1}^{P} M_{zp} N_{z'p} = (MN^T)_{zz'}$
$\mathcal{X}_{zz'} = \mathcal{Y}^T = \sum_{p=1}^{P} N_{zp} M_{z'p} = (NM^T)_{zz'}$

Explanation:

$\mathcal{Z}$: This is a matrix of dimension $Z \times Z$. For two rows z, z', i.e. two nodes z, z' in the first class, entry zz' gives the following: it counts the number of columns, or nodes in the second class, which are adjacent to both z and z'.

$\mathcal{Y}$: matrix of dimension $Z \times Z$, entry zz' gives the number of columns which are adjacent to z but not z'

$\mathcal{X}$: the transpose of matrix $\mathcal{Y}$, matrix of dimension $Z \times Z$, entry zz' gives the number of columns which are adjacent to z' but not z

$$\mathcal{P}_{pp'} = \sum_{z=1}^{Z} M_{zp} M_{zp'} = (M^T M)_{pp'}$$
$$\mathcal{Q}_{pp'} = \sum_{z=1}^{Z} M_{zp} N_{zp'} = (M^T N)_{pp'}$$
$$\mathcal{R}_{pp'} = \mathcal{Q}^T = \sum_{z=1}^{Z} N_{zp} M_{zp'} = (N^T M)_{pp'}$$

Explanation:

$\mathcal{P}$: This is a matrix of dimension $P \times P$. For two columns p, p', or two nodes in the second class p, p', the entry pp' gives the number of rows, or nodes in the first class, which are adjacent to both p and p'.

$\mathcal{Q}$: matrix of dimension $P \times P$, entry pp' the number of rows which are adjacent to p but not p'

$\mathcal{R}$: the transpose of matrix $\mathcal{Q}$, matrix of dimension $P \times P$, entry pp' the number of rows which are adjacent to p' but not p

$j_{Zi} = 1$ (defines a vector of 1's of length Z)
$j_{Pi} = 1$ (defines a vector of 1's of length P)

$J_{ij} = 1$ (defines a matrix of dimension $Z \times P$ with all entries $= 1$)
$J_{Zij} = 1$ (defines a matrix of dimension $Z \times Z$ with all entries $= 1$)
$J_{Pij} = 1$ (defines a matrix of dimension $P \times P$ with all entries $= 1$)

$$A_{ijk} = \sum_{p=1}^{P} M_{ip} M_{jp} M_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki})$$
$$B_{ijk} = \sum_{p=1}^{P} M_{ip} M_{jp} N_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki})$$
$$C_{ijk} = \sum_{p=1}^{P} M_{ip} N_{jp} M_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = B_{ikj}$$
$$D_{ijk} = \sum_{p=1}^{P} M_{ip} N_{jp} N_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki})$$
$$E_{ijk} = \sum_{p=1}^{P} N_{ip} M_{jp} M_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = B_{jki}$$
$$F_{ijk} = \sum_{p=1}^{P} N_{ip} M_{jp} N_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = D_{jik}$$
$$G_{ijk} = \sum_{p=1}^{P} N_{ip} N_{jp} M_{kp} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = D_{kij}$$

Explanation:

δ_{ij}: This is the Kronecker delta: $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

A: This is an array of dimension $Z \times Z \times Z$. For three distinct rows, or nodes in the first class, i, j, k, the entry ijk gives the number of possible columns, or nodes in the second class, which are adjacent to all of i, j, k. Note that if i, j, k are not all different, the entry at position ijk will be zero.

B: array of dimension $Z \times Z \times Z$, for distinct rows i, j, k, the entry ijk gives the number of possible columns which are adjacent to i, j, but not adjacent to k; if i, j, k not all distinct, entry is zero

C: array of dimension $Z \times Z \times Z$, for distinct rows i, j, k, the entry ijk gives the number of possible columns which are adjacent to i, k, but not adjacent to j; if i, j, k not all distinct, entry is zero

D: array of dimension $Z \times Z \times Z$, for distinct rows i, j, k, the entry ijk gives the number of possible columns which are adjacent to i, but not adjacent to j, k; if i, j, k not all distinct, entry is zero

E: array of dimension $Z \times Z \times Z$, for distinct rows i, j, k, the entry ijk gives the number of possible columns which are adjacent to j, k, but not adjacent to i; if i, j, k not all distinct, entry is zero

F: array of dimension $Z \times Z \times Z$, for distinct rows i, j, k, the entry ijk gives the number of possible columns which are adjacent to j, but not adjacent to i, k; if i, j, k not all distinct, entry is zero

G: array of dimension $Z \times Z \times Z$, for distinct rows i, j, k, the entry ijk gives the number of possible columns which are adjacent to k, but not adjacent to i, j; if i, j, k not all distinct, entry is zero

$A'_{ijk} = \sum_{z=1}^{Z} M_{zi} M_{zj} M_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki})$
$B'_{ijk} = \sum_{z=1}^{Z} M_{zi} M_{zj} N_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki})$
$C'_{ijk} = \sum_{z=1}^{Z} M_{zi} N_{zj} M_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = B'_{ikj}$
$D'_{ijk} = \sum_{z=1}^{Z} M_{zi} N_{zj} N_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki})$
$E'_{ijk} = \sum_{z=1}^{Z} N_{zi} M_{zj} M_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = B'_{jki}$
$F'_{ijk} = \sum_{z=1}^{Z} N_{zi} M_{zj} N_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = D'_{jik}$
$G'_{ijk} = \sum_{z=1}^{Z} N_{zi} N_{zj} M_{zk} (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ki}) = D'_{kij}$

Explanation:

A': This is an array of dimension $P \times P \times P$. For three distinct columns, or nodes in the second class, i, j, k, the entry ijk gives the number of possible rows, or nodes in the first class, which are adjacent to all of i, j, k. Note that if i, j, k are not all different, the entry at position ijk will be zero.

B': array of dimension $P \times P \times P$, for distinct columns i, j, k, the entry ijk gives the number of possible rows, which are adjacent to i, j, but not adjacent to k; if i, j, k not all distinct, entry is zero

C': array of dimension $P \times P \times P$, for distinct columns i, j, k, the entry ijk gives the number of possible rows, which are adjacent to i, k, but not adjacent to j; if i, j, k not all distinct, entry is zero

D': array of dimension $P \times P \times P$, for distinct columns i, j, k, the entry ijk gives the number of possible rows, which are adjacent to i, but not adjacent to j, k; if i, j, k not all distinct, entry is zero

E': array of dimension $P \times P \times P$, for distinct columns i, j, k, the entry ijk gives the number of possible rows, which are adjacent to j, k, but not adjacent to i; if i, j, k not all distinct, entry is zero

F': array of dimension $P \times P \times P$, for distinct columns i, j, k, the entry ijk gives the number of possible rows, which are adjacent to j, but not adjacent to i, k; if i, j, k not all distinct, entry is zero

G': array of dimension $P \times P \times P$, for distinct columns i, j, k, the entry ijk gives the number of possible rows, which are adjacent to k, but not adjacent to i, j; if i, j, k not all distinct, entry is zero

$f(T)_i = \sum_{j=1}^{Z} \sum_{k=1}^{Z} T_{ijk}$
$g(T)_i = \sum_{j=1}^{P} \sum_{k=1}^{P} T_{ijk}$

Explanation:

$f(T)$: This is a vector with Z entries, where T is a three-dimensional matrix (more formally: tensor) of dimension $Z \times Z \times Z$. At position i, $f(T)$ is the sum of the values of T_{ijk}, summed over all choices of j and k.

$g(T)$: likewise, but T is of dimension $P \times P \times P$ and $g(T)$ has P entries

A general explanation for the formulae

Motif counts

The general approach to compute the motif count for a given motif is as follows. First, consider the class of nodes in the motif which has fewer nodes. (If both classes have the same size, consider either of them.) For example in motif 14, which has two nodes in the first class (rows) and three nodes in the second class (columns), we consider the first class because this has fewer nodes than the second class.

Note that in any motif, the class which has the fewest nodes will have at most three nodes. For the following, assume we consider motifs which have fewer nodes in the first class than in the second (like motif 14). Say we have Z nodes in the first class and P nodes in the second class. Then in these motifs by assumption $Z \leq P$ and also, due to the above, $Z \leq 3$. For other motifs, where we have more nodes in the first class than in the second class (i.e. $P \leq Z$ and then $P \leq 3$), the following description would be very similar. But we would need to swap the two classes throughout the explanation.

Now pick any Z nodes in the network, which are in the first class. Given these nodes, we want to choose the P nodes in the second class such that all nodes together form the required motif. Each of the nodes in the second class needs to satisfy some adjacency conditions. We ask: how many possibilities do we have to pick the nodes in the second class in this way?

To obtain this number, we can use the auxiliary matrices as defined above. The entry of one of these gives the number of possibilities to pick a single node in the second class, satisfying certain conditions. Of course, which of the auxillary matrices we use depends on the adjacency conditions in the given motif. We then have one entry in a certain matrix for each node in the second class. These entries can now be multiplied together if all nodes in the second class have different conditions. Otherwise, we use binomial coefficients. In general, for fixed Z nodes in the first class, we now have a product of matrix entries giving the number of possibilities to pick nodes in the second class that form our given motif.

However, in the end, we are interested in the motif count of the entire network. Therefore, we need to sum over all possible choices of the Z nodes (which are in the first class) in the network. In some motifs, double counting due to symmetry occurs, which we still need to correct. For example, if there are two nodes in the same position in the first class, we need to make sure we do not count degenerated versions of the motif. With degenerated versions we mean versions where two nodes are merged together. (E.g. if the two nodes in the bottom class of motif 2 are merged together, the motif would look like motif 1.) Here we need to subtract the counts that we added wrongly. If there are two nodes in the same position in the second class, we count each motif several times, say k times, so we divide the result by k in the end. The reader is advised to take a look at the examples to see how this is done in detail. But after this, we calculated the desired result.

Example Motif 5: Say we fix 2 nodes in the first class, z and z'. Assume z is the node in position 12 and z' the node in position 11. Which conditions do the nodes in the second class need to satisfy? The node in position 9, call it p, has to be adjacent to z, but shall not be adjacent to z'. How many possibilities are there to pick p? This is given by the entry $\mathcal{Y}_{zz'}$, since $\mathcal{Y}_{zz'}$ gives the number of columns which are adjacent to z but not z' (see above). Meanwhile, the node in position 10, call it p', has to be adjacent to both z, z'. Likewise, the number of possibilities to pick p' is given by $\mathcal{Z}_{zz'}$, since $\mathcal{Z}_{zz'}$ gives the number of columns which are adjacent to both z and z'. We can combine each possible choice of p with each possible choice of p'. Hence, the number of possible choices for p and p' is $\mathcal{Y}_{zz'} * \mathcal{Z}_{zz'}$. Now we need to sum this over all choices of z, z' and
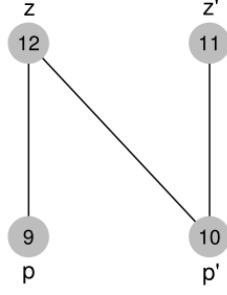
Figure 1: Motif 5

get the following result:

$$\text{motif count for motif 5} = \sum_{z=1}^{Z}\sum_{z'=1}^{Z} \mathcal{Y}_{zz'} * \mathcal{Z}_{zz'}$$

Node position counts

The approach to compute the node position count for a given position is similar. Without loss of generality, assume we consider a position in the first class of some motif. Again, the description for a position in the second class would be very similar, just swap the classes. Say the motif has Z nodes in the first class and P nodes in the second class.

Fix a node in the first class in the network, say z. We want to know how often it is in the given position. Now pick $Z-1$ other nodes in the first class. We now want to find nodes in the second class, such that all nodes together form the motif that the position occurs in. Again, the auxiliary matrices are used to calculate how many possibilities we have to pick these nodes. To obtain the result, we need to sum over all possible choices of the $Z-1$ nodes in the first class. The node position counts are given as a vector, the number of entries of the vector equals the number of nodes in the network which are in the first class. Hence there is one entry for each node in the first class of the network. The entry number z of the vector is the count how often node z occurs in this node position.
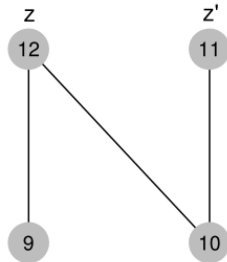

Figure 2: Motif 5

Example of position 12 in motif 5: For some node z in the first class, we fix another node in the first class, say z'. Then, as above, there are $\mathcal{Y}_{zz'} * \mathcal{Z}_{zz'}$ possibilities to pick nodes in the

second class, which form motif 5 together with z, z'. Hence the result is

$$\sum_{z'=1}^{Z} \mathcal{Y}_{zz'} * \mathcal{Z}_{zz'}$$

Now we seek a vector s.t. the above line is the entry at position z of this vector. Consider the matrix $\mathcal{Y} \circ \mathcal{Z}$. Its entry zz' is exactly $\mathcal{Y}_{zz'} * \mathcal{Z}_{zz'}$. Now we want to sum over all possible z', which is the second variable. By definiton of matrix-vector multiplication, this can be realised by multiplying with an all-ones vector, which was above called $\mathbf{j_z}$. Hence, the entry z in the vector $(\mathcal{Y} \circ \mathcal{Z}) \cdot \mathbf{j_z}$ is exactly the result, so this is the wanted vector.
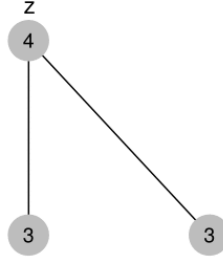
Further worked examples

Motif 2

Figure 3: Motif 2

Fix a node z in the network, which is in the first class. It has d_z neighbours, since d_z is its degree. Out of these d_z neighbours, we need to choose two to complete the motif. Hence we have $\binom{d_z}{2} = \frac{1}{2}d_z(d_z - 1)$ possibilities to pick two nodes in the second class that form motif 2 together with z. To obtain the motif count for the entire network, we sum this over all possible z:

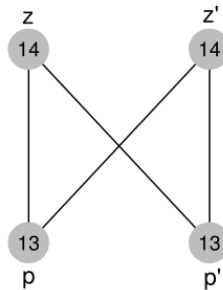$$\frac{1}{2}\sum_{z=1}^{Z} d_z(d_z - 1)$$

Motif 6

Figure 4: Motif 6

This example is introduced to explain the correcting due to symmetry. First, fix nodes z, z' in the network, both in the first class. Now we seek two nodes p, p' in the second class, which both are

adjacent to both z, z'. For each one of p and p', we have $\mathcal{Z}_{zz'}$ possibilities, since $\mathcal{Z}_{zz'}$ gives the number of columns which are adjacent to both z and z'. So for choosing both of them, there are $\binom{\mathcal{Z}_{zz'}}{2} = \frac{1}{2}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'} - 1)$ possibilities. Summing this over all choices of z and z' gives:

$$\sum_{z=1}^{Z}\sum_{z'=1}^{Z} \frac{1}{2}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'} - 1)$$

However, this has excluded the condition $z \neq z'$ so far, which is clearly necessary. If we allow $z = z'$, we would have only one node in the first class and get a degenerated version of motif 6, in this case it would look like motif 2. Hence we need to subtract the terms where $z = z'$. Use that $\mathcal{Z}_{zz} = d_z$ since, by definition, $\mathcal{Z}_{zz}$ gives the number of columns which are adjacent to z and z, so this is just the number of neighbours of z, i.e. its degree.

So the formula becomes:

$$\sum_{z=1}^{Z}\sum_{z'=1}^{Z} \frac{1}{2}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'} - 1) - \sum_{z=1}^{Z} \frac{1}{2}\mathcal{Z}_{zz}(\mathcal{Z}_{zz} - 1) = \sum_{z=1}^{Z}\sum_{z'=1}^{Z} \frac{1}{2}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'} - 1) - \sum_{z=1}^{Z} \frac{1}{2}d_z(d_z - 1)$$

Note that this formula does not account for the symmetry in z and z' yet. The nodes z and z' both satisfy the same conditions. So in each occurrence, they can be swapped and we still have a motif of the same type. Hence every occurrence will be counted twice. Therefore, the entire result needs to be divided by two. This then gives the final formula:

$$\frac{1}{4}\sum_{z=1}^{Z}\sum_{z'=1}^{Z} \mathcal{Z}_{zz'}(\mathcal{Z}_{zz'} - 1) - \frac{1}{4}\sum_{z=1}^{Z} d_z(d_z - 1)$$
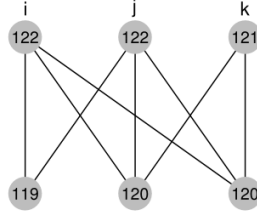
Motif 36

Figure 5: Motif 36

Here we present an example for six-node motifs, where the three-dimensional matrices (more formally known as tensors) are used. Fix nodes i, j, k in the first class in the network. Say node i is the node on the left in position 122, j is the node in the middle also in position 122 and k is the node on the right in position 121 (note that this label order is arbitrary: if the labels are ordered differently this gives a different version of the formula, which is equivalent). There are B_{ijk} possible ways to pick a node in position 119 out of the second class, since B_{ijk} gives the number of columns which are adjacent to i, j, but not adjacent to k (and a node in position 119 is adjacent to the nodes in position 122, which are nodes i, j, but not adjacent to the node in position 121, which is node k). The two bottom nodes in position 120 are both adjacent to all of i, j, k. A_{ijk} gives the number of columns which are adjacent to all of i, j, k. That means, there are A_{ijk} possible choices for one node in position 120. We need two, so we get $\binom{A_{ijk}}{2}$ possibilities to pick such two nodes out of the second class. Hence there are

$$B_{ijk} * \binom{A_{ijk}}{2} = \frac{1}{2}B_{ijk}A_{ijk}(A_{ijk} - 1)$$

possibilities to form motif 36, given nodes i, j, k in the first class. This product then has to be summed over all choices of i, j, k to obtain:

$$\frac{1}{2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z} B_{ijk}A_{ijk}(A_{ijk}-1)$$

However, we have not considered that i and j both are in position 122. In each occurrence, i and j can be swapped and we still have a motif of the same type, so we count each occurrence twice (like in the example of motif 6). Therefore, we need to divide our result by two and get the final formula:

$$\frac{1}{4}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z} B_{ijk}A_{ijk}(A_{ijk}-1)$$
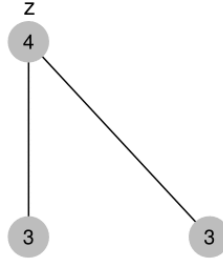
Position 4

Figure 6: Motif 2

Note that position 4 is in motif 2. Consider a node z in the first class. Like in the example of motif 2, we want to pick two nodes in the second class that form motif 2 together with z. That means we need two distinct nodes in the second class which are both adjacent to z. There are d_z neighbours of z and we choose two, so we have are $\binom{d_z}{2} = \frac{1}{2}d_z(d_z-1)$ possibilities to pick such two nodes. Hence z occurs $\frac{1}{2}d_z(d_z-1)$ times in position 4. Note that the entry z of the vector $\mathbf{d}$ is d_z by definition and $\mathbf{j_z}$ is the all-ones-vector, so $\frac{1}{2}d_z(d_z-1)$ is the entry z of the vector $\frac{1}{2}\mathbf{d}\circ(\mathbf{d}-\mathbf{j_z})$.
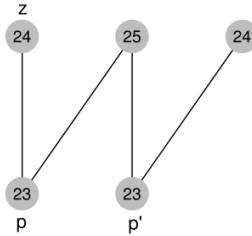
Position 24

Figure 7: Motif 10

This formula is found with a slighty different approach and a few more mathematical ideas. We introduce the notation that $x \sim y$ if x and y are adjacent and $x \nsim y$ if x and y are not adjacent.

The formula could be obtained in a very similar way to the position 87 (see below), but in this case it is more efficient to do it as follows: Fix node z in the first class, assume it is the left node in

position 24. Consider the two nodes in the second class, say the left one is p and the right one is p'. Note that p is adjacent to z ($p \sim z$), but p' is not ($p' \not\sim z$). Following the definition of $\mathcal{P}$, there are $\mathcal{P}_{pp'}$ possibilities to pick the middle node in position 25, since $\mathcal{P}_{pp'}$ gives the number of possible rows adjacent to both p and p' (and the middle node is adjacent to both p, p'). Likewise, there are $\mathcal{R}_{pp'}$ possibilities to pick the right node in position 24. Hence there are $\mathcal{P}_{pp'}\mathcal{R}_{pp'}$ possibilities to pick the middle node in position 25 and the right node in position 24. So we need to sum this quantity over all possible choices of p and p', where we consider that $p \sim z$ and $p' \not\sim z$. In the following, we use that summing over all p which satisfy $p \sim z$ is the same as summing over all p and adding the term M_{zp}, since M_{zp} is one if and only if $p \sim z$ is true. Similarly, summing over all p' which satisfy $p' \not\sim z$ is the same as summing over all p' and adding the term $N_{zp'}$, since $N_{zp'}$ is one if and only if $p' \not\sim z$ is true. Now we repeatedly apply the definition of matrix multiplication to simplify the result:

$$
\begin{aligned}
\text{result} &= \sum_{\substack{p \\ p \sim z}} \sum_{\substack{p' \\ p' \not\sim z}} \mathcal{P}_{pp'}\mathcal{R}_{pp'} \\
&= \sum_{p} \sum_{p'} M_{zp} N_{zp'} \mathcal{P}_{pp'}\mathcal{R}_{pp'} \\
&= \sum_{p} \sum_{p'} M_{zp} N_{zp'} (\mathcal{P} \circ \mathcal{R})_{pp'} \\
&= \sum_{p'} N_{zp'} \sum_{p} M_{zp} (\mathcal{P} \circ \mathcal{R})_{pp'} \\
&= \sum_{p'} N_{zp'} (M \cdot (\mathcal{P} \circ \mathcal{R}))_{zp'}
\end{aligned}
$$

This is the entry z of the vector $(N \circ (M \cdot (\mathcal{P} \circ \mathcal{R}))) \cdot \mathbf{j_P}$ (compare example of position 12, this is again the definition of matrix multiplication).
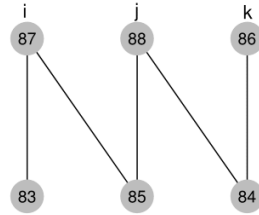
Position 87

Figure 8: Motif 28

One example for six-node motifs. Note that position 87 occurs in motif 28. Consider a node i in the first class. Pick two further nodes j, k in the first class. To find the node position count for the node i, we can argue similarly to the example of motif 28. Assume j is the middle node and k the right node. There are D_{ijk} possible ways to pick a node in position 83 out of the second class, since D_{ijk} gives the number of columns which are adjacent to i, but not adjacent to j, k (and a node in position 83 is adjacent to the node in position 87, which is node i, but not adjacent to any others). Likewise, there are B_{ijk} possible ways to pick a node in position 85, E_{ijk} to pick a node in position 84. Hence there are $D_{ijk}B_{ijk}E_{ijk}$ possibilities to form motif 28 with these i, j, k. This needs to be summed over all choices of j, k. Hence node i occurs

$$
\sum_{j=1}^{Z} \sum_{k=1}^{Z} D_{ijk}B_{ijk}E_{ijk} = \sum_{j=1}^{Z} \sum_{k=1}^{Z} (D \circ B \circ E)_{ijk}
$$

times in position 87. Recall that $f(T)_i = \sum_{j=1}^{Z} \sum_{k=1}^{Z} T_{ijk}$. Now note that $D \circ B \circ E$ is a tensor of dimension $Z \times Z \times Z$, so now plug in $D \circ B \circ E$ for T in the definition of f. Then

$$f(D \circ B \circ E)_i = \sum_{j=1}^{Z} \sum_{k=1}^{Z} (D \circ B \circ E)_{ijk},$$

so this is the wanted result. This is exactly the i-th entry of the vector $\mathbf{f}(D \circ B \circ E)$, which is the required vector.

mcount formulae

Motif 1

$$\sum_{z=1}^{Z}\sum_{p=1}^{P} M_{zp}$$

Motif 2

$$\frac{1}{2}\sum_{z=1}^{Z} d_z(d_z - 1)$$

Motif 3

$$\frac{1}{2}\sum_{p=1}^{P} u_p(u_p - 1)$$

Motif 4

$$\frac{1}{6}\sum_{p=1}^{P} u_p(u_p - 1)(u_p - 2)$$

Motif 5

$$\sum_{z=1}^{Z}\sum_{z'=1}^{Z} \mathcal{Z}_{zz'}\mathcal{Y}_{zz'}$$

Motif 6

$$\frac{1}{4}\sum_{z=1}^{Z}\sum_{z'=1}^{Z} \mathcal{Z}_{zz'}(\mathcal{Z}_{zz'} - 1) - \frac{1}{4}\sum_{z=1}^{Z} d_z(d_z - 1)$$

Motif 7

$$\frac{1}{6}\sum_{z=1}^{Z} d_z(d_z - 1)(d_z - 2)$$

Motif 8

$$\frac{1}{24}\sum_{p=1}^{P} u_p(u_p - 1)(u_p - 2)(u_p - 3)$$

Motif 9

$$\frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P} \mathcal{P}_{pp'}\mathcal{Q}_{pp'}(\mathcal{Q}_{pp'} - 1)$$

Motif 10

$$\frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}\mathcal{Q}_{pp'}\mathcal{R}_{pp'}$$

Motif 11

$$\frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}(\mathcal{P}_{pp'}-1)\mathcal{Q}_{pp'}$$

Motif 12

$$\frac{1}{12}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}(\mathcal{P}_{pp'}-1)(\mathcal{P}_{pp'}-2)-\frac{1}{12}\sum_{p=1}^{P}d_p(d_p-1)(d_p-2)$$

Motif 13

$$\frac{1}{2}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}\mathcal{Y}_{zz'}(\mathcal{Y}_{zz'}-1)$$

Motif 14

$$\frac{1}{2}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}\mathcal{Y}_{zz'}\mathcal{X}_{zz'}$$

Motif 15

$$\frac{1}{2}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'}-1)\mathcal{Y}_{zz'}$$

Motif 16

$$\frac{1}{12}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'}-1)(\mathcal{Z}_{zz'}-2)-\frac{1}{12}\sum_{z=1}^{Z}d_z(d_z-1)(d_z-2)$$

Motif 17

$$\frac{1}{24}\sum_{z=1}^{Z}d_z(d_z-1)(d_z-2)(d_z-3)$$

Motif 18

$$\frac{1}{120}\sum_{p=1}^{P}u_p(u_p-1)(u_p-2)(u_p-3)(u_p-4)$$

Motif 19

$$\frac{1}{6}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}\mathcal{Q}_{pp'}(\mathcal{Q}_{pp'}-1)(\mathcal{Q}_{pp'}-2)$$

Motif 20

$$\frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}\mathcal{Q}_{pp'}(\mathcal{Q}_{pp'}-1)\mathcal{R}_{pp'}$$

Motif 21

$$\frac{1}{4}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}(\mathcal{P}_{pp'}-1)\mathcal{Q}_{pp'}(\mathcal{Q}_{pp'}-1)$$

Motif 22

$$\frac{1}{4}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}(\mathcal{P}_{pp'}-1)\mathcal{Q}_{pp'}\mathcal{R}_{pp'}$$

Motif 23

$$\frac{1}{6}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}(\mathcal{P}_{pp'}-1)(\mathcal{P}_{pp'}-2)\mathcal{Q}_{pp'}$$

Motif 24

$$\frac{1}{48}\sum_{p=1}^{P}\sum_{p'=1}^{P}\mathcal{P}_{pp'}(\mathcal{P}_{pp'}-1)(\mathcal{P}_{pp'}-2)(\mathcal{P}_{pp'}-3)-\frac{1}{48}\sum_{p=1}^{P}u_p(u_p-1)(u_p-2)(u_p-3)$$

Motif 25

$$\frac{1}{4}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}F_{ijk}(F_{ijk}-1)\ =\ \frac{1}{4}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}F'_{ijk}(F'_{ijk}-1)$$

Motif 26

$$\frac{1}{2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}D_{ijk}B_{ijk}C_{ijk}\qquad=\ \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}F'_{ijk}G'_{ijk}A'_{ijk}$$

Motif 27

$$\frac{1}{2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}F_{ijk}G_{ijk}A_{ijk}\qquad=\ \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}D'_{ijk}B'_{ijk}C'_{ijk}$$

Motif 28

$$\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}B_{ijk}C_{ijk}G_{ijk}\qquad=\ \sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}B'_{ijk}C'_{ijk}G'_{ijk}$$

Motif 29

$$\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}B_{ijk}D_{ijk}\qquad=\ \sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}B'_{ijk}D'_{ijk}$$

Motif 30

$$\frac{1}{2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}B_{ijk}G_{ijk} \qquad = \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}B'_{ijk}(B'_{ijk}-1)C'_{ijk}$$

Motif 31

$$\frac{1}{4}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}F_{ijk}(A_{ijk}-1) = \frac{1}{4}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}B'_{ijk}(B'_{ijk}-1)A'_{ijk}$$

Motif 32

$$\frac{1}{2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}B_{ijk}(B_{ijk}-1)C_{ijk} = \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}B'_{ijk}G'_{ijk}$$

Motif 33

$$\frac{1}{4}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}B_{ijk}(B_{ijk}-1)A_{ijk} = \frac{1}{4}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}F'_{ijk}(A'_{ijk}-1)$$

Motif 34

$$\frac{1}{6}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}E_{ijk}B_{ijk}C_{ijk} \qquad = \frac{1}{6}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}E'_{ijk}B'_{ijk}C'_{ijk}$$

Motif 35

$$\frac{1}{2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}B_{ijk}C_{ijk} \qquad = \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}B'_{ijk}C'_{ijk}$$

Motif 36

$$\frac{1}{4}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}(A_{ijk}-1)B_{ijk} = \frac{1}{4}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}(A'_{ijk}-1)B'_{ijk}$$

Motif 37

$$\frac{1}{36}\sum_{i=1}^{Z}\sum_{j=1}^{Z}\sum_{k=1}^{Z}A_{ijk}(A_{ijk}-1)(A_{ijk}-2) = \frac{1}{36}\sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{k=1}^{P}A'_{ijk}(A'_{ijk}-1)(A'_{ijk}-2)$$

Motif 38

$$\frac{1}{6}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}\mathcal{Y}_{zz'}(\mathcal{Y}_{zz'}-1)(\mathcal{Y}_{zz'}-2)$$

Motif 39

$$\frac{1}{2}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}\mathcal{Y}_{zz'}(\mathcal{Y}_{zz'}-1)\mathcal{X}_{zz'}$$

Motif 40

$$\frac{1}{4}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'}-1)\mathcal{Y}_{zz'}(\mathcal{Y}_{zz'}-1)$$

Motif 41

$$\frac{1}{4}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'}-1)\mathcal{Y}_{zz'}\mathcal{X}_{zz'}$$

Motif 42

$$\frac{1}{6}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'}-1)(\mathcal{Z}_{zz'}-2)\mathcal{Y}_{zz'}$$

Motif 43

$$\frac{1}{48}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\mathcal{Z}_{zz'}(\mathcal{Z}_{zz'}-1)(\mathcal{Z}_{zz'}-2)(\mathcal{Z}_{zz'}-3)-\frac{1}{48}\sum_{z=1}^{Z}d_z(d_z-1)(d_z-2)(d_z-3)$$

Motif 44

$$\frac{1}{120}\sum_{z=1}^{Z}d_z(d_z-1)(d_z-2)(d_z-3)(d_z-4)$$

positions formulae

Position 1

$$\mathbf{u}$$

Position 2

$$\mathbf{d}$$

Position 3

$$\mathcal{P} \cdot \mathbf{j_P} - \mathbf{u}$$

Position 4

$$\frac{1}{2}\mathbf{d} \circ (\mathbf{d} - \mathbf{j_Z})$$

Position 5

$$\frac{1}{2}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P})$$

Position 6

$$\mathcal{Z} \cdot \mathbf{j_Z} - \mathbf{d}$$

Position 7

$$\frac{1}{6}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P})$$

Position 8

$$\frac{1}{2}M \cdot ((\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}))$$

Position 9

$$(\mathcal{P} \circ \mathcal{R}) \cdot \mathbf{j_P}$$

Position 10

$$(\mathcal{P} \circ \mathcal{Q}) \cdot \mathbf{j_P}$$

Position 11

$$(\mathcal{X} \circ \mathcal{Z}) \cdot \mathbf{j_Z}$$

Position 12

$$(\mathcal{Y} \circ \mathcal{Z}) \cdot \mathbf{j_Z}$$

Position 13

$$\frac{1}{2}(\mathcal{P} \circ (\mathcal{P} - J_P)) \cdot \mathbf{j_P} - \frac{1}{2}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P})$$

Position 14

$$\frac{1}{2}(\mathcal{Z} \circ (\mathcal{Z} - J_Z)) \cdot \mathbf{j_Z} - \frac{1}{2}\mathbf{d} \circ (\mathbf{d} - \mathbf{j_Z})$$

Position 15

$$\frac{1}{2}M^T \cdot ((\mathbf{d} - \mathbf{j_Z}) \circ (\mathbf{d} - 2\mathbf{j_Z}))$$

Position 16

$$\frac{1}{6}\mathbf{d} \circ (\mathbf{d} - \mathbf{j_Z}) \circ (\mathbf{d} - 2\mathbf{j_Z})$$

Position 17

$$\frac{1}{24}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}) \circ (\mathbf{u} - 3\mathbf{j_P})$$

Position 18

$$\frac{1}{6}M \cdot ((\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}) \circ (\mathbf{u} - 3\mathbf{j_P}))$$

Position 19

$$\frac{1}{2}(\mathcal{P} \circ \mathcal{R} \circ (\mathcal{R} - J_P)) \cdot \mathbf{j_P}$$

Position 20

$$\frac{1}{2}(\mathcal{P} \circ \mathcal{Q} \circ (\mathcal{Q} - J_P)) \cdot \mathbf{j_P}$$

Position 21

$$(N \circ (M \cdot ((\mathcal{Q} - J_P) \circ \mathcal{P}))) \cdot \mathbf{j_P}$$

Position 22

$$\frac{1}{2}(M \circ (M \cdot (\mathcal{Q} \circ (\mathcal{Q} - J_P)))) \cdot \mathbf{j_P}$$

Position 23

$$(\mathcal{P} \circ \mathcal{Q} \circ \mathcal{R}) \cdot \mathbf{j_P}$$

Position 24

$$(N \circ (M \cdot (\mathcal{P} \circ \mathcal{R}))) \cdot \mathbf{j_P}$$

Position 25

$$\frac{1}{2}(M \circ (M \cdot (\mathcal{Q} \circ \mathcal{R}))) \cdot \mathbf{j_P}$$

Position 26

$$\frac{1}{2}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ \mathcal{R}) \cdot \mathbf{j_P}$$

Position 27

$$\frac{1}{2}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ \mathcal{Q}) \cdot \mathbf{j_P}$$

Position 28

$$\frac{1}{2}(N \circ (M \cdot (\mathcal{P} \circ (\mathcal{P} - J_P)))) \cdot \mathbf{j_P}$$

Position 29

$$(M \circ (M \cdot (\mathcal{Q} \circ (\mathcal{P} - J_P)))) \cdot \mathbf{j_P}$$

Position 30

$$\frac{1}{6}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P)) \cdot \mathbf{j_P} - \frac{1}{6}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P})$$

Position 31

$$\frac{1}{4}(M \circ M \cdot ((\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P))) \cdot \mathbf{j_P} - \frac{1}{4}M \cdot ((\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}))$$

Position 32

$$(N^T \circ (M^T \cdot ((\mathcal{Y} - J_Z) \circ \mathcal{Z}))) \cdot \mathbf{j_Z}$$

Position 33

$$\frac{1}{2}(M^T \circ (M^T \cdot (\mathcal{Y} \circ (\mathcal{Y} - J_Z)))) \cdot \mathbf{j_Z}$$

Position 34

$$\frac{1}{2}(\mathcal{Z} \circ \mathcal{X} \circ (\mathcal{X} - J_Z)) \cdot \mathbf{j_Z}$$

Position 35

$$\frac{1}{2}(\mathcal{Z} \circ \mathcal{Y} \circ (\mathcal{Y} - J_Z)) \cdot \mathbf{j_Z}$$

Position 36

$$(N^T \circ (M^T \cdot (Z \circ X))) \cdot \mathbf{j_Z}$$

Position 37

$$\frac{1}{2}(M^T \circ (M^T \cdot (\mathcal{Y} \circ \mathcal{X}))) \cdot \mathbf{j_z}$$

Position 38

$$(\mathcal{Z} \circ \mathcal{Y} \circ \mathcal{X}) \cdot \mathbf{j_z}$$

Position 39

$$\frac{1}{2}(N^T \circ (M^T \cdot (\mathcal{Z} \circ (\mathcal{Z} - J_Z)))) \cdot \mathbf{j_z}$$

Position 40

$$(M^T \circ (M^T \cdot (\mathcal{Y} \circ (\mathcal{Z} - J_Z)))) \cdot \mathbf{j_z}$$

Position 41

$$\frac{1}{2}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ \mathcal{X}) \cdot \mathbf{j_z}$$

Position 42

$$\frac{1}{2}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ \mathcal{Y}) \cdot \mathbf{j_z}$$

Position 43

$$\frac{1}{4}(M^T \circ (M^T \cdot ((\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z)))) \cdot \mathbf{j_z} - \frac{1}{4}M^T \cdot ((\mathbf{d} - \mathbf{j_z}) \circ (\mathbf{d} - 2\mathbf{j_z}))$$

Position 44

$$\frac{1}{6}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z)) \cdot \mathbf{j_z} - \frac{1}{6}\mathbf{d} \circ (\mathbf{d} - \mathbf{j_z}) \circ (\mathbf{d} - 2\mathbf{j_z})$$

Position 45

$$\frac{1}{6}M^T \cdot ((\mathbf{d} - \mathbf{j_z}) \circ (\mathbf{d} - 2\mathbf{j_z}) \circ (\mathbf{d} - 3\mathbf{j_z}))$$

Position 46

$$\frac{1}{24}\mathbf{d} \circ (\mathbf{d} - \mathbf{j_z}) \circ (\mathbf{d} - 2\mathbf{j_z}) \circ (\mathbf{d} - 3\mathbf{j_z})$$

Position 47

$$\frac{1}{120}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}) \circ (\mathbf{u} - 3\mathbf{j_P}) \circ (\mathbf{u} - 4 \circ \mathbf{j_P})$$

Position 48

$$\frac{1}{24}M \cdot ((\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}) \circ (\mathbf{u} - 3\mathbf{j_P}) \circ (\mathbf{u} - 4 \circ \mathbf{j_P}))$$

Position 49

$$\frac{1}{6}(\mathcal{P} \circ \mathcal{R} \circ (\mathcal{R} - J_P) \circ (\mathcal{R} - 2\mathbf{j_P})) \cdot \mathbf{j_P}$$

Position 50

$$\frac{1}{6}(\mathcal{P} \circ \mathcal{Q} \circ (\mathcal{Q} - J_P) \circ (\mathcal{Q} - 2\mathbf{j_P})) \cdot \mathbf{j_P}$$

Position 51

$$\frac{1}{2}(N \circ (M \cdot (\mathcal{P} \circ (\mathcal{Q} - J_P) \circ (\mathcal{Q} - 2\mathbf{j_P})))) \cdot \mathbf{j_P}$$

Position 52

$$\frac{1}{6}(M \circ (M \cdot (\mathcal{Q} \circ (\mathcal{Q} - J_P) \circ (\mathcal{Q} - 2\mathbf{j_P})))) \cdot \mathbf{j_P}$$

Position 53

$$\frac{1}{2}(\mathcal{P} \circ \mathcal{Q} \circ \mathcal{R} \circ (\mathcal{R} - J_P)) \cdot \mathbf{j_P}$$

Position 54

$$\frac{1}{2}(\mathcal{P} \circ \mathcal{Q} \circ (\mathcal{Q} - J_P) \circ \mathcal{R}) \cdot \mathbf{j_P}$$

Position 55

$$\frac{1}{2}(M \circ (N \cdot (\mathcal{P} \circ \mathcal{Q} \circ (\mathcal{Q} - J_P)))) \cdot \mathbf{j_P}$$

Position 56

$$(M \circ (N \cdot (\mathcal{P} \circ \mathcal{Q} \circ (\mathcal{R} - J_P)))) \cdot \mathbf{j_P}$$

Position 57

$$\frac{1}{2}(M \circ (M \cdot (\mathcal{Q} \circ (\mathcal{Q} - J_P) \circ \mathcal{R}))) \cdot \mathbf{j_P}$$

Position 58

$$\frac{1}{4}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ \mathcal{R} \circ (\mathcal{R} - J_P)) \cdot \mathbf{j_P}$$

Position 59

$$\frac{1}{4}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ \mathcal{Q} \circ (\mathcal{Q} - J_P)) \cdot \mathbf{j_P}$$

Position 60

$$\frac{1}{2}(N \circ (M \cdot ((\mathcal{Q} - J_P) \circ \mathcal{P} \circ (\mathcal{P} - J_P)))) \cdot \mathbf{j_P}$$

Position 61

$$\frac{1}{2}(M \circ (M \cdot ((\mathcal{P} - J_P) \circ \mathcal{Q} \circ (\mathcal{Q} - J_P)))) \cdot \mathbf{j_P}$$

Position 62

$$\frac{1}{2}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ \mathcal{Q} \circ \mathcal{R}) \cdot \mathbf{j_P}$$

Position 63

$$\frac{1}{2}(N \circ (M \cdot (\mathcal{P} \circ (\mathcal{P} - J_P) \circ \mathcal{R}))) \cdot \mathbf{j_P}$$

Position 64

$$\frac{1}{2}(M \circ (M \cdot ((\mathcal{P} - J_P) \circ \mathcal{Q} \circ \mathcal{R}))) \cdot \mathbf{j_P}$$

Position 65

$$\frac{1}{6}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P) \circ \mathcal{R}) \cdot \mathbf{j_P}$$

Position 66

$$\frac{1}{6}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P) \circ \mathcal{Q}) \cdot \mathbf{j_P}$$

Position 67

$$\frac{1}{6}(N \circ (M \cdot (\mathcal{P} \circ (\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P)))) \cdot \mathbf{j_P}$$

Position 68

$$\frac{1}{2}(M \circ (M \cdot ((\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P) \circ \mathcal{Q}))) \cdot \mathbf{j_P}$$

Position 69

$$\frac{1}{24}(\mathcal{P} \circ (\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P) \circ (\mathcal{P} - 3J_P)) \cdot \mathbf{j_P} - \frac{1}{24}\mathbf{u} \circ (\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}) \circ (\mathbf{u} - 3\mathbf{j_P})$$

Position 70

$$\frac{1}{12}(M \circ (M \cdot ((\mathcal{P} - J_P) \circ (\mathcal{P} - 2J_P) \circ (\mathcal{P} - 3J_P)))) \cdot \mathbf{j_P} - \frac{1}{12}M \cdot ((\mathbf{u} - \mathbf{j_P}) \circ (\mathbf{u} - 2\mathbf{j_P}) \circ (\mathbf{u} - 3\mathbf{j_P}))$$

Position 71

$$\frac{1}{2}\mathbf{g}(F' \circ (F' - J_{P3}) \circ A')$$

Position 72

$$\frac{1}{4}\mathbf{g}(D' \circ (D' - J_{P3}) \circ A')$$

Position 73

$$\frac{1}{2}\mathbf{f}(F \circ (F - J_{Z3}) \circ A)$$

Position 74

$$\frac{1}{4}\mathbf{f}(D \circ (D - J_{Z3}) \circ A)$$

Position 75

$$\mathbf{g}(D' \circ A' \circ F')$$

Position 76

$$\frac{1}{2}\mathbf{g}(A' \circ F' \circ G')$$

Position 77

$$\frac{1}{2}\mathbf{f}(D \circ B \circ C)$$

Position 78

$$\mathbf{f}(B \circ F \circ E)$$

Position 79

$$\frac{1}{2}\mathbf{g}(B' \circ C' \circ D')$$

Position 80

$$\mathbf{g}(B' \circ F' \circ E')$$

Position 81

$$\mathbf{f}(A \circ D \circ F)$$

Position 82

$$\frac{1}{2}\mathbf{f}(A \circ F \circ G)$$

Position 83

$$\mathbf{g}(B' \circ E' \circ G')$$

Position 84

$$\mathbf{g}(E' \circ B' \circ D')$$

Position 85

$$\mathbf{g}(B' \circ C' \circ G')$$

Position 86

$$\mathbf{f}(B \circ E \circ G)$$

Position 87

$$\mathbf{f}(D \circ B \circ E)$$

Position 88

$$\mathbf{f}(F \circ B \circ C)$$

Position 89

$$\mathbf{g}(A' \circ E' \circ F')$$

Position 90

$$\mathbf{g}(A' \circ B' \circ F')$$

Position 91

$$\mathbf{g}(A' \circ B' \circ D')$$

Position 92

$$\mathbf{f}(A \circ E \circ F)$$

Position 93

$$\mathbf{f}(B \circ F \circ A)$$

Position 94

$$\mathbf{f}(A \circ B \circ D)$$

Position 95

$$\frac{1}{2}\mathbf{g}(B' \circ E' \circ (E' - J_{P3}))$$

Position 96

$$\frac{1}{2}\mathbf{g}(B' \circ (B' - J_{P3}) \circ E')$$

Position 97

$$\frac{1}{2}\mathbf{g}(B' \circ C' \circ (C' - J_{P3}))$$

Position 98

$$\frac{1}{2}\mathbf{f}(D \circ A \circ E)$$

Position 99

$$\mathbf{f}(A \circ F \circ C)$$

Position 100

$$\frac{1}{4}\mathbf{g}(A' \circ E' \circ (E' - J_{P3}))$$

Position 101

$$\frac{1}{2}\mathbf{g}(A' \circ B' \circ (B' - J_{P3}))$$

Position 102

$$\frac{1}{2}\mathbf{f}(A \circ (A - J_{Z3}) \circ F)$$

Position 103

$$\frac{1}{4}\mathbf{f}(D \circ A \circ (A - J_{Z3}))$$

Position 104

$$\frac{1}{2}\mathbf{g}(D' \circ A' \circ E')$$

Position 105

$$\mathbf{g}(B' \circ A' \circ G')$$

Position 106

$$\frac{1}{2}\mathbf{f}(B \circ E \circ (E - J_{Z3}))$$

Position 107

$$\frac{1}{2}\mathbf{f}(B \circ (B - J_{Z3}) \circ E)$$

Position 108

$$\frac{1}{2}\mathbf{f}(B \circ (B - J_{Z3}) \circ C)$$

Position 109

$$\frac{1}{2}\mathbf{g}(A' \circ (A' - J_{P3}) \circ F')$$

Position 110

$$\frac{1}{4}\mathbf{g}(A' \circ (A' - J_{P3}) \circ D')$$

Position 111

$$\frac{1}{4}\mathbf{f}(A \circ E \circ (E - J_{Z3}))$$

Position 112

$$\frac{1}{2}\mathbf{f}(A \circ B \circ (B - J_{Z3}))$$

Position 113

$$\frac{1}{2}\mathbf{g}(B' \circ C' \circ E')$$

Position 114

$$\frac{1}{2}\mathbf{f}(B \circ C \circ E)$$

Position 115

$$\mathbf{g}(A' \circ B' \circ E')$$

Position 116

$$\frac{1}{2}\mathbf{g}(A' \circ B' \circ C')$$

Position 117

$$\mathbf{f}(A \circ E \circ B)$$

Position 118

$$\frac{1}{2}\mathbf{f}(A \circ B \circ C)$$

Position 119

$$\frac{1}{4}\mathbf{g}(A' \circ (A' - J_{P3}) \circ E')$$

Position 120

$$\frac{1}{2}\mathbf{g}(A' \circ (A' - J_{P3}) \circ B')$$

Position 121

$$\frac{1}{4}\mathbf{f}(A \circ (A - J_{Z3}) \circ E)$$

Position 122

$$\frac{1}{2}\mathbf{f}(B \circ A \circ (A - J_{Z3}))$$

Position 123

$$\frac{1}{12}\mathbf{g}(A' \circ (A' - J_{P3}) \circ (A' - 2J_{P3}))$$

Position 124

$$\frac{1}{12}\mathbf{f}(A \circ (A - J_{Z3}) \circ (A - 2J_{Z3}))$$

Position 125

$$\frac{1}{2}(N^T \circ (M^T \cdot (\mathcal{Z} \circ (\mathcal{Y} - J_Z) \circ (\mathcal{Y} - 2J_Z)))) \cdot \mathbf{jz}$$

Position 126

$$\frac{1}{6}(M^T \circ (M^T \cdot (\mathcal{Y} \circ (\mathcal{Y} - J_Z) \circ (\mathcal{Y} - 2J_Z)))) \cdot \mathbf{jz}$$

Position 127

$$\frac{1}{6}(\mathcal{Z} \circ \mathcal{X} \circ (\mathcal{X} - J_Z) \circ (\mathcal{X} - 2J_Z)) \cdot \mathbf{jz}$$

Position 128

$$\frac{1}{6}(\mathcal{Z} \circ \mathcal{Y} \circ (\mathcal{Y} - J_Z) \circ (\mathcal{Y} - 2J_Z)) \cdot \mathbf{jz}$$

Position 129

$$\frac{1}{2}(M^T \circ (N^T \cdot (\mathcal{Z} \circ \mathcal{Y} \circ (\mathcal{Y} - J_Z)))) \cdot \mathbf{jz}$$

Position 130

$$(M^T \circ (N^T \cdot (\mathcal{Z} \circ \mathcal{Y} \circ (\mathcal{X} - J_Z)))) \cdot \mathbf{jz}$$

Position 131

$$\frac{1}{2}(M^T \circ (M^T \cdot (\mathcal{Y} \circ (\mathcal{Y} - J_Z) \circ \mathcal{X}))) \cdot \mathbf{jz}$$

Position 132

$$\frac{1}{2}(\mathcal{Z} \circ \mathcal{Y} \circ \mathcal{X} \circ (\mathcal{X} - J_Z)) \cdot \mathbf{jz}$$

Position 133

$$\frac{1}{2}(\mathcal{Z} \circ \mathcal{Y} \circ (\mathcal{Y} - J_Z) \circ \mathcal{X}) \cdot \mathbf{jz}$$

Position 134

$$\frac{1}{2}(N^T \circ (M^T \cdot ((\mathcal{Y} - J_Z) \circ \mathcal{Z} \circ (\mathcal{Z} - J_Z)))) \cdot \mathbf{jz}$$

Position 135

$$\frac{1}{2}(M^T \circ (M^T \cdot ((\mathcal{Z} - J_Z) \circ \mathcal{Y} \circ (\mathcal{Y} - J_Z)))) \cdot \mathbf{jz}$$

Position 136

$$\frac{1}{4}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ \mathcal{X} \circ (\mathcal{X} - J_Z)) \cdot \mathbf{jz}$$

Position 137

$$\frac{1}{4}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ \mathcal{Y} \circ (\mathcal{Y} - J_Z)) \cdot \mathbf{jz}$$

Position 138

$$\frac{1}{2}(N^T \circ (M^T \cdot (\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ \mathcal{X}))) \cdot \mathbf{jz}$$

Position 139

$$\frac{1}{2}(M^T \circ (M^T \cdot ((\mathcal{Z} - J_Z) \circ \mathcal{Y} \circ \mathcal{X}))) \cdot \mathbf{jz}$$

Position 140

$$\frac{1}{2}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ \mathcal{Y} \circ \mathcal{X}) \cdot \mathbf{jz}$$

Position 141

$$\frac{1}{6}(N^T \circ (M^T \cdot (\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z)))) \cdot \mathbf{jz}$$

Position 142

$$\frac{1}{2}(M^T \circ (M^T \cdot ((\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z) \circ \mathcal{Y}))) \cdot \mathbf{jz}$$

Position 143

$$\frac{1}{6}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z) \circ \mathcal{X}) \cdot \mathbf{jz}$$

Position 144

$$\frac{1}{6}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z) \circ \mathcal{Y}) \cdot \mathbf{jz}$$

Position 145

$$\frac{1}{12}(M^T \circ (M^T \cdot ((\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z) \circ (\mathcal{Z} - 3J_Z)))) \cdot \mathbf{jz} - \frac{1}{12}M^T \cdot ((\mathbf{d} - \mathbf{jz}) \circ (\mathbf{d} - 2\mathbf{jz}) \circ (\mathbf{d} - 3\mathbf{jz}))$$

Position 146

$$\frac{1}{24}(\mathcal{Z} \circ (\mathcal{Z} - J_Z) \circ (\mathcal{Z} - 2J_Z) \circ (\mathcal{Z} - 3J_Z)) \cdot \mathbf{jz} - \frac{1}{24}\mathbf{d} \circ (\mathbf{d} - \mathbf{jz}) \circ (\mathbf{d} - 2\mathbf{jz}) \circ (\mathbf{d} - 3\mathbf{jz})$$

Position 147

$$\frac{1}{24}M^T \cdot ((\mathbf{d} - \mathbf{jz}) \circ (\mathbf{d} - 2\mathbf{jz}) \circ (\mathbf{d} - 3\mathbf{jz}) \circ (\mathbf{d} - 4\mathbf{jz}))$$

Position 148

$$\frac{1}{120}\mathbf{d} \circ (\mathbf{d} - \mathbf{jz}) \circ (\mathbf{d} - 2\mathbf{jz}) \circ (\mathbf{d} - 3\mathbf{jz}) \circ (\mathbf{d} - 4\mathbf{jz})$$

Appendix S4: Computational performance

Empirical networks

To assess the speed of bmotif functions, we used *mcount* and *node_positions* to calculate the complete motif profiles of 175 empirical pollination and seed dispersal networks and the positions of all their constituent species. Networks were obtained from the Web of Life dataset (www.web-of-life.es). The networks varied in size from 6 to 797 species (mean: 77.1; standard deviation: 117.8). Analyses were carried out on a computer with a 4.0 GHz processor and 32 GB of memory. Functions were timed using the R package 'microbenchmark' (Mersmann, 2015). Results are shown in Fig. S2.
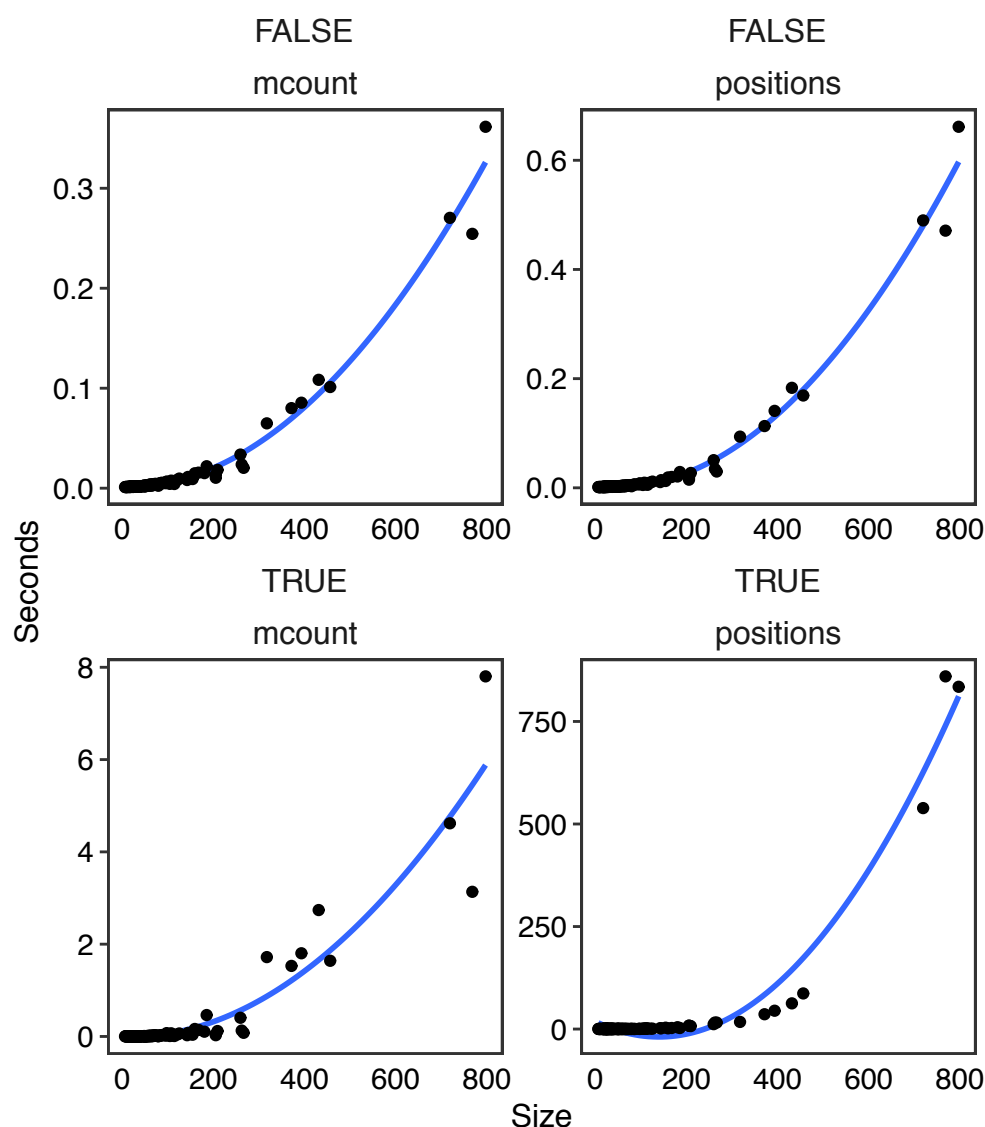

Figure S2: Relationship between network size and computational performance for *mcount* and *node_positions* for motifs containing up to five nodes (FALSE) and six nodes (TRUE). Functions were timed on 175 empirical networks. Lines are best fit polynomial curves of degree 2.

As expected, the time taken for a function to run increases monotonically with the size of the network (number of species). When six-node motifs were excluded, *mcount* and *node_positions* took 0.36 and 0.66 seconds, respectively, to complete for the largest network

in our dataset (797 species). For smaller networks which are more typical of the communities analysed by ecologists, both functions completed in substantially less than one second. This speed is possible as all formulae involved in calculations of motifs up to five-nodes use relatively simple operations, such as matrix multiplication or the binomial coefficient. When six-node motifs were included, for a network with 78 species (close to the mean network size of 77.1 species), *mcount* completed in 0.01 seconds, while *node_positions* completed in 0.32 seconds. For the largest network, *mcount* completed in 7.8 seconds, while *node_positions* took 13.9 minutes. Six-node motifs slow down calculations as, unlike five-node motifs, their algorithms require the use of the tensor product. Overall, the speed of bmotif makes motif analyses compatible with the permutational approaches frequently used in network ecology, particularly for analyses with motifs up to five-nodes and for six-node analyses of all but the largest networks. For example, using bmotif it would be feasible to calculate motif frequency distributions across thousands of null networks, which is a widely-used approach to disentangle the mechanisms responsible for network structure (Bascompte, Jordano, Melián, & Olesen, 2003; Dormann, Frund, Bluthgen, & Gruber, 2009).

Random networks

We carried out two analyses using randomly-generated networks to examine the effects of network size (number of species) and connectance on the computational performance of individual motif and motif position calculations. For the first analysis, we generated random networks with a fixed size, varying the connectance between 0.2 and 1. We generated 1000 networks for each value of connectance. For each of these sets of 1000 networks, we recorded the mean time for our code to calculate the frequency of five motifs (motifs 1, 2, 5, 10 and 28; one from each of the five motif size classes) and the number of times each species occurred in five motif positions (positions 1, 3, 9, 23 and 85; one from each motif size class). The dimensions of the generated networks were set as the median number of rows and columns of 230 empirical ecological bipartite networks (22 rows, 13 columns) obtained from the Web of Life repository (www.web-of-life.es). For the second analysis, we generated random networks of a fixed connectance, varying the size between 10 and 200 species. We generated 1000 networks for each value of size and recorded the mean time for our code to calculate the frequency of the same five motifs and positions. The connectance of the generated networks was the median connectance of the empirical network dataset (0.243) and the row:column ratio (ratio of number of species in one level, such as hosts, to the number of species in the other level, such as parasitoids) was also set as the empirical median (2). Functions were timed using the R package 'microbenchmark' (Mersmann, 2015).

We found that connectance had little effect on the performance of individual motif and position calculations (Fig. S3), while a polynomial of degree two explained the increase in time with network size ($R^2 > 0.99$ for all motifs and positions) (Fig. S4).
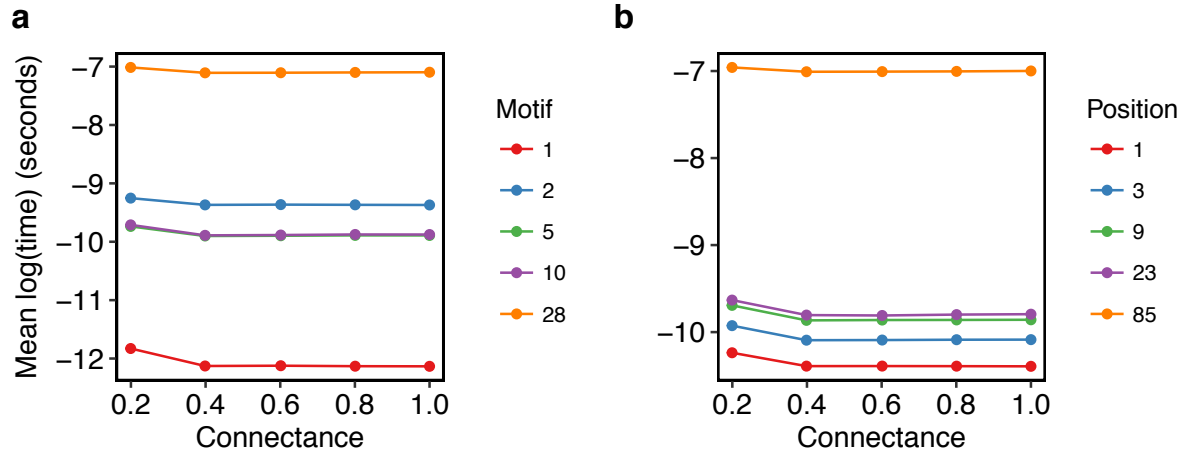
Figure S3: Relationship between connectance and computational time taken to calculate the frequency of (a) five motifs, one from each motif size class, and (b) five motif positions, one from each motif size class. Functions were run on randomly-generated networks of a given connectance. For each level of connectance, we generated 1000 random networks and record the mean time for the functions to complete. Lines connecting each point are shown for visualisation.
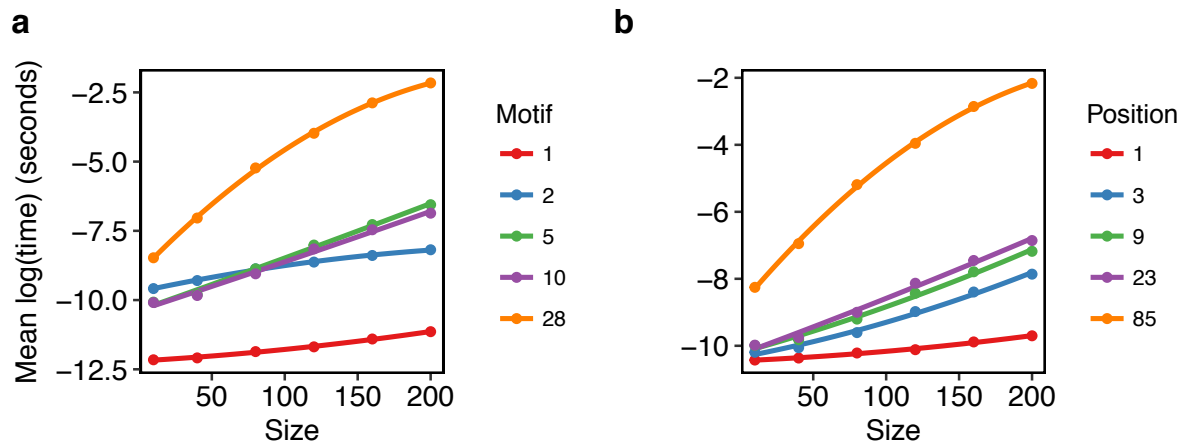

Figure S4: Relationship between size and computational time taken to calculate the frequency of (a) five motifs, one from each motif size class, and (b) five motif positions, one from each motif size class. Functions were run on randomly-generated networks of a given size. For each level of size, we generated 1000 random networks and record the mean time for the functions to complete. Lines are best fit polynomials of degree two.

Appendix S5: Description of *mcount* and *node_positions* outputs

mcount takes a network as input and returns a data frame with one row for each motif (17 or 44 rows depending on whether motifs up to five or six nodes are requested, respectively) and three columns. The first column is the motif identity as in Fig. 1; the second column is the motif size class (number of nodes each motif contains); and the third column is the frequency with which each motif occurs in the network (a network's motif profile). For comparing multiple networks it is important to normalise motif frequencies. Therefore, if the 'normalisation' argument is TRUE, three columns are added to the data frame, each corresponding to a different method for normalising motif frequencies. The first column ('normalise_sum') expresses the frequency of each motif as a proportion of the total number of motifs in the network. The second column ('normalise_sizeclass') expresses the frequency of each motif as a proportion of the total number of motifs within its size class. The final column ('normalise_nodesets') expresses the frequency of each motif as the number of species combinations that occur in a motif as a proportion of the number of species combinations that could occur in that motif. For example, in motifs 9, 10, 11 and 12, there are three species in the top set (*A*) and two species in the lower set (*B*) (Fig. 1). Therefore, the maximum number of species combinations that could occur in these motifs is given by the product of binomial coefficients, choosing three species from *A* and two from *P*: $\binom{A}{3}\binom{B}{2}$ (Poisot & Stouffer, 2016). The most appropriate normalisation depends on the question being asked. For example, 'normalise_sum' allows for consideration of whether species are more involved in smaller or larger motifs. Conversely, 'normalise_sizeclass' focuses the analysis on how species form their interactions among different arrangements of *n* nodes.

node_positions takes a network as input and returns a data frame, **W**, with one row for each species and one column for each node position (46 or 148 columns, depending on whether motifs up to five or six nodes are requested, respectively; Fig. 1). w_{rc} gives the number of times species *r* occurs in position *c*. Each row thus represents the structural role or 'interaction niche' of a species. The 'level' argument allows positions to be requested for all species, species in set *A* only or species in set *B* only, returning a data frame with *A* + *B* rows, *A* rows or *B* rows, respectively. Two types of normalisation are provided: 'sum' normalisation expresses a species' position frequencies as a proportion of the total number of times that species appears in any position; 'size class' normalisation uses the same approach, but normalises frequencies within each motif size class. Again, the most appropriate normalisation depends on the question being asked: if movements between motif size classes are of interest, 'sum' normalisation is most appropriate; if the focus is on how species form interactions among a given number of nodes, then 'size class' normalisation should be chosen.

References

Bascompte, J., Jordano, P., Melián, C. J., & Olesen, J. M. (2003). The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387.

Dormann, C. F., Frund, J., Bluthgen, N., & Gruber, B. (2009). Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. *The Open Ecology Journal*, *2*(1), 7–24. doi:10.2174/1874213000902010007

Mersmann, O. (2015). microbenchmark: Accurate Timing Functions. R package version 1.4-

2.1.

Poisot, T., & Stouffer, D. (2016). How ecological networks evolve. *BioRxiv*. Retrieved from http://biorxiv.org/content/early/2016/08/29/071993.abstract