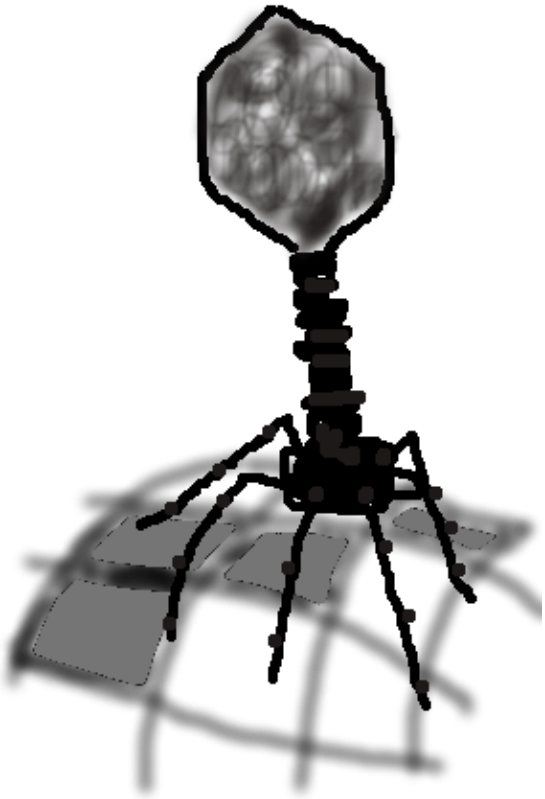
Nestedness and modularity in bipartite networks

Submitted by Stephen James Beckett, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences

In May 2015

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..



“for what art is greater than art that imitates life itself?”

– Robin Hobb, *The Mad Ship*

Abstract

Bipartite networks are a useful way of representing interactions between two sets of entities. Understanding the underlying structures of such networks may give insights into the functionality and behaviour of the systems they represent. Two important structural patterns identified in bipartite networks are nestedness and modularity. Nestedness describes a hierarchical ordering of nodes such that more specialised nodes have interactions with a subset of the partners with which the more generalised nodes interact. Modularity captures the community structure of a network as distinct clusters of interactions, such that there are more connections within communities than between communities. While these network architectures are easy to describe in writing, their quantitative measurement for a given network is a difficult task. Several different methods have been proposed in each case and it is currently unclear which of them should be used in practice. This thesis considers the use, measurement and interpretation of nestedness and modularity in bipartite networks. First, it is shown how bipartite networks can be an effective tool for linking data and theory in community ecology, though use of a coevolutionary model of virus-bacteria interactions. Next, a series of studies is presented that push towards clarification of the best procedures to measure nestedness and modularity in bipartite networks. Robustness of nestedness measures is tested on a synthetic ensemble of networks, showing that apparent nestedness depends strongly on the choice of measure, null model and effect size statistics. Recommendations for performing nestedness are made with relation to individual and cross-network comparisons. Additionally, a new algorithm for identifying weighted modularity is proposed that can be shown to outperform existing methods. Crucially, it is shown that quantitative modular structures differ from traditional binary modular structures with implications for how modularity is reported and used. Improving the way in which nestedness and modularity are measured is a necessary step for integrating data and theory in bipartite networks.

Acknowledgements

First and foremost, my most grateful thanks for the support and patience of my supervisor Hywel Williams who has been a voice of reasoned calm during this (at times) seemingly unrelenting storm.

Similarly a big thank you for the generous moral and emotional support from colleagues in Earth System Science and Exeter Climate Systems. It is good to work in an environment where all are considered peers. My special thanks to Tom, Siobhán, Chris, Nick, Vicky, Sarah, Sarah, Margriet, Anna, Luke, Jim, Ben, Richard, Alastair, Duarte and Tobia. Without your distractions and encouragement I wouldn't have made it!

Thanks to Andy and Lauren for inviting me to work with them on their projects - which have subsequently lead to publications. I especially thank Sean Meadon and Britt Koskella for inviting me down to Cornwall to talk about my theoretical phages, and to see their experiments involving the real thing.

To my housemates over the years: Aindreas, Dori, Rob, Suzanne and Asia - thanks for putting up with me.

I thank the University of Exeter for funding my doctoral studentship. Additional financial support from the London Mathematical Society and the Society for General Microbiology helped me to attend conferences during my studies; and funding provided by the ESF Research Networking Programme FroSpects, the Finnish Doctoral Programme FICS and the Academy of Finland via the Finnish Centre of Excellence in Analysis and Dynamics Research allowed me to attend The Helsinki Summer School on Mathematical Ecology and Evolution.

Finally a heartful thank you to my family and friends without whom I wouldn't be where I am today.

Contents

List of Tables

List of Figures

List of Algorithms

List of Publications

Published

Beckett S.J., Williams H.T.P. 2013. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. Interface Focus **3**(6): 20130033.

http://dx.doi.org/10.1098/rsfs.2013.0033

Watts A.J.R., Lewis C., Goodhead R.M., Beckett S.J., Moger J., Tyler C.R., Galloway T. 2014. Uptake and retention of microplastics by the shore crab *Carcinus maenas*. Environmental Science & Technology **48**(15): 8823-8830.

http://dx.doi.org/10.1021/es501090e

Beckett S.J., Boulton C.A., Williams H.T.P. 2014. FALCON: a software package for analysis of nestedness in bipartite networks [v1; ref status: indexed, http://f1000r.es/3z8] F1000Research **3**: 185.

http://dx.doi.org/10.12688/f1000research.4831.1

Cowley L.A., Beckett S.J., Chase-Toppin M., Perry N., Dallman T.J., Gally D.L., Jenkins C. 2015. Analysis of whole genome sequencing for the *Escherichia coli* O157:H7 typing phages. BMC Genomics **16**: 271.

http://dx.doi.org/10.1186/s12864-015-1470-z

In preparation

Beckett S.J. Improved community detection in weighted bipartite networks. **submitted**

Williams H.T.P., Beckett S.J., O'Neill S., Kurz T. Dynamics of attention and influence in online coverage of the IPCC Fifth Assessment Reports. **in preparation**

Beckett S.J., Williams H.T.P. Brooding on nestedness: nestedness analyses are confounded by sensitivity to measurement choices and network properties. **in preparation**

Software repositories

FALCON : http://github.com/sjbeckett/FALCON

FALCON (Framework for Adaptive ensembLes for the Comparison Of Nestedness) is a library developed for use in MATLAB, Octave and R. It was developed to enable the comparison of nestedness measures and null model interpretations on networks. It is described in Chapter 4 and is used in Chapter 5.

LPAwb+ : http://github.com/sjbeckett/weighted-modularity-LPAwbPLUS

LPAwb+ (label propagation algorithm for weighted bipartite networks using multi-step agglomeration) is an algorithm available for use in the Julia, MATLAB, Octave and R programming languages. The LPAwb+ and Exhaustive LPAwb+ algorithms are community detection algorithms used to attempt to maximise weighted modularity. It is described and used in Chapter 6.

Datasets

Beckett S.J., Williams H.T.P. 2015. Synthetic matrix ensemble for nestedness analysis. figshare.

http://dx.doi.org/10.6084/m9.figshare.1320818

Posters

Beckett S. J., Williams H.T.P. 2013. Towards trait-based models for aquatic virology. figshare.

http://dx.doi.org/10.6084/m9.figshare.714953

Beckett S.J., Williams H.T.P. 2013. Coevolved nestedness and modularity in phage-bacteria infection networks. figshare.

http://dx.doi.org/10.6084/m9.figshare.699871

Authors declaration

Several of the chapters contributing to this thesis were based on joint work, written up for publication. Appropriate acknowledgements and citations are provided at the beginning of each chapter. Here I list author contributions.

Chapter 3 is adapted from the publication:

Beckett S.J., Williams H.T.P. 2013. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. Interface Focus **3**(6): 20130033.

SJB and HTPW designed experiments and wrote the manuscript. SJB performed and analysed experiments.

Chapter 4 is adapted from the publication:

Beckett S.J., Boulton C.A., Williams H.T.P. 2014. FALCON: a software package for analysis of nestedness in bipartite networks [v1; ref status: indexed, http://f1000r.es/3z8] F1000Research **3**: 185.

SJB developed the initial project in MATLAB. SJB, CAB and HTPW enhanced the original package. SJB and CAB ported FALCON into R. SJB and HTPW wrote the manuscript with input from CAB.

Chapter 5 is based on:

Beckett S.J., Williams H.T.P. Brooding on nestedness: nestedness analyses are confounded by sensitivity to measurement choices and network properties. **in preparation**

SJB performed and analysed the experiments and wrote the manuscript. SJB and HTPW designed the experiments and HTPW commented on the manuscript.

Finally, Chapter 6 is based on:

Beckett S.J. Improved community detection in weighted bipartite networks. **submitted**

SJB designed, performed and analysed experiments and wrote the manuscript.

Definitions/abbreviations

BR	Discrepancy measure of nestedness
CC	Cored-cored null model
DD	Degreeprobable-degreeprobable null model
EE	Equiprobable-equiprobable null model
FALCON	Framework for Adaptive ensembLes for the Comparison Of Nestedness - a library for comparing nestedness measures and null models
FF	Fixed-fixed null model
JDM	The nestedness measure of Johnson, Domínguez-García and Muñoz 2013.
LHS	Latin hypercube sampling
LPAb+	Label propagation algorithm for bipartite networks using multi-step agglomeration
LPAwb+	Label propagation algorithm for weighted bipartite networks using multi-step agglomeration
LP-BRIM	Label propagation and bipartite, recursively induced modules algorithm
QuanBiMo	Quantitative Bipartite Modularity algorithm
MD	Manhattan distance, used as a nestedness measure
NODF	Nestedness metric based on Overlap and Decreasing Fill, a nestedness measure
NTC	Nestedness temperature calculator, a nestedness measure (which also involves a reordering rows and columns genetic algorithm)
$P(rewire)$	A rewiring probability applied to each network edge
Q_b	Barber's modularity
Q'_R	Realized modularity
Q^{norm}	Normalised modularity
Q^{max}	Maxmium modularity
SR	Spectral radius, used as a nestedness measure.
SS	Swappable-swappable null model
WNODF	Weighted NODF, a nestedness measure for quantitative networks

Chapter 1

Introduction

Network theory is an interdisciplinary science that offers an approach to understanding the interactions between different entities in complex systems. Consideration of the structural properties of networks may lead to an understanding of how they came to be, the rules that govern them, how they are expected to change over time and how they might respond to perturbations. In particular, bipartite networks depict interactions involving two distinct sets of nodes where connections only occur between, and not within the two sets. Bipartite networks are increasingly being used to understand ecological interactions, for example plant-pollinator [1, 2] and host-parasite [3] interactions. Yet, theoretical understanding of bipartite networks is not as well developed as that for unipartite networks, that involve only one type of node. In this thesis I shall examine two structural properties of bipartite networks, modularity and nestedness, with the aim of improving the way in which these features are identified and quantified.

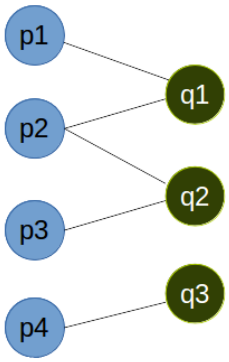
In this chapter, I provide some background knowledge on nestedness, modularity and the methods used to measure these network attributes. First, I give an overview and definitions useful for investigating the properties of bipartite networks in Section 1.1. Then Section 1.2 gives examples of the types of structural patterns that have been identified in bipartite networks including nestedness and modularity. Following this, further details and background of these key network features are given in Section 1.3 (nestedness) and Section 1.4 (modularity) respectively. Some examples of the software used to investigate these features is then provided in Section 1.5. Finally, an overview of the contents of the thesis chapters is presented in Section 1.7.

1.1 Bipartite networks

Networks are composed of several **nodes** that are linked together by **edges** that represent an interaction or an association between nodes. **Bipartite networks** are networks that have two distinct sets of nodes, where edges represent interactions or associations between the two different types of entity. For example Figure 1.1*a* is a bipartite network with four p nodes (p1, p2, p3, p4) and three q nodes (q1, q2, q3) that are linked by five edges (p1 interacts with q1, p2 interacts with q1 and q2, etc.). It is important to note that there are no interactions between nodes of the same type i.e. p nodes can only interact with q nodes and vice-versa. Networks which are described by a single node type (where interactions between nodes of this single type are allowed) are known as **unipartite networks**. The **degree** of each node is the number of edges that each node is connected by. If a node is connected to

lots of other nodes it is said to be a hub, or in the context of ecological networks a **generalist**. On the other hand, nodes that have few edge connections and are more isolated from the rest of the network are said to be **specialists**.

The **adjacency matrix** of a network is a matrix representing whether a node in a network interacts with all the other nodes (where an interaction is shown by a 1 and absence of interaction by a 0). The adjacency matrix for the network in Figure 1.1*a* is shown in Figure 1.1*b*. However, as no interactions between nodes of the same type are allowed the biadjacency matrix is often used for calculations on bipartite networks. The corresponding **biadjacency matrix** is shown in Figure 1.1*c*, where nodes of one type are represented as rows and nodes of the other type are represented as columns in the matrix. Lines shown in Figure 1.1*b* show that the biadjacency matrix appears in the top right of the adjacency matrix and the transposed biadjacency matrix appears in the bottom left.



	p1	p2	p3	p4	q1	q2	q3
p1	0	0	0	0	1	0	0
p2	0	0	0	0	1	1	0
p3	0	0	0	0	0	1	0
p4	0	0	0	0	0	0	1
q1	1	1	0	0	0	0	0
q2	0	1	1	0	0	0	0
q3	0	0	0	1	0	0	0

	q1	q2	q3
p1	1	0	0
p2	1	1	0
p3	0	1	0
p4	0	0	1

Figure 1.1: Different representations of a theoretical bipartite network. The network is represented as (*a*) a graph layout, (*b*) its adjacency matrix and (*c*) its biadjacency matrix.

A biadjacency matrix is **square** if there is an equal number of nodes of each type. **Fill** is given by the total sum of edges in the network. The network's **size** is calculated as the total number of potential edges i.e. the number of rows multiplied by the number of columns in the biadjacency matrix. **Connectance** is the number of actual edges, expressed as a proportion of the total number of potential edges, that exist in the network and can be calculated as fill divided by size.

Binary bipartite networks describe networks on the basis of the **presence** (represented by 1 in the biadjacency matrix) or **absence** (represented by 0) of edges between nodes, as shown in Figure 1.1. On the other hand, in **weighted** bipartite net-

works values in the biadjacency matrix represent the strength of association between two nodes.

1.2 Patterns in bipartite networks

It is generally suspected that ecological networks and other types of real-world networks are not formed from random interactions, but have internal organisation. The way in which a network is organised can give insights into the processes that led to its formation in the real world system it represents. Thus a lot of network science is concerned with the identification of different kinds of topological patterns in networks. Several types of bipartite network topology are shown in Figure 1.2. In this figure each cell represents an interaction between two nodes, one represented on the rows, the other on the columns.

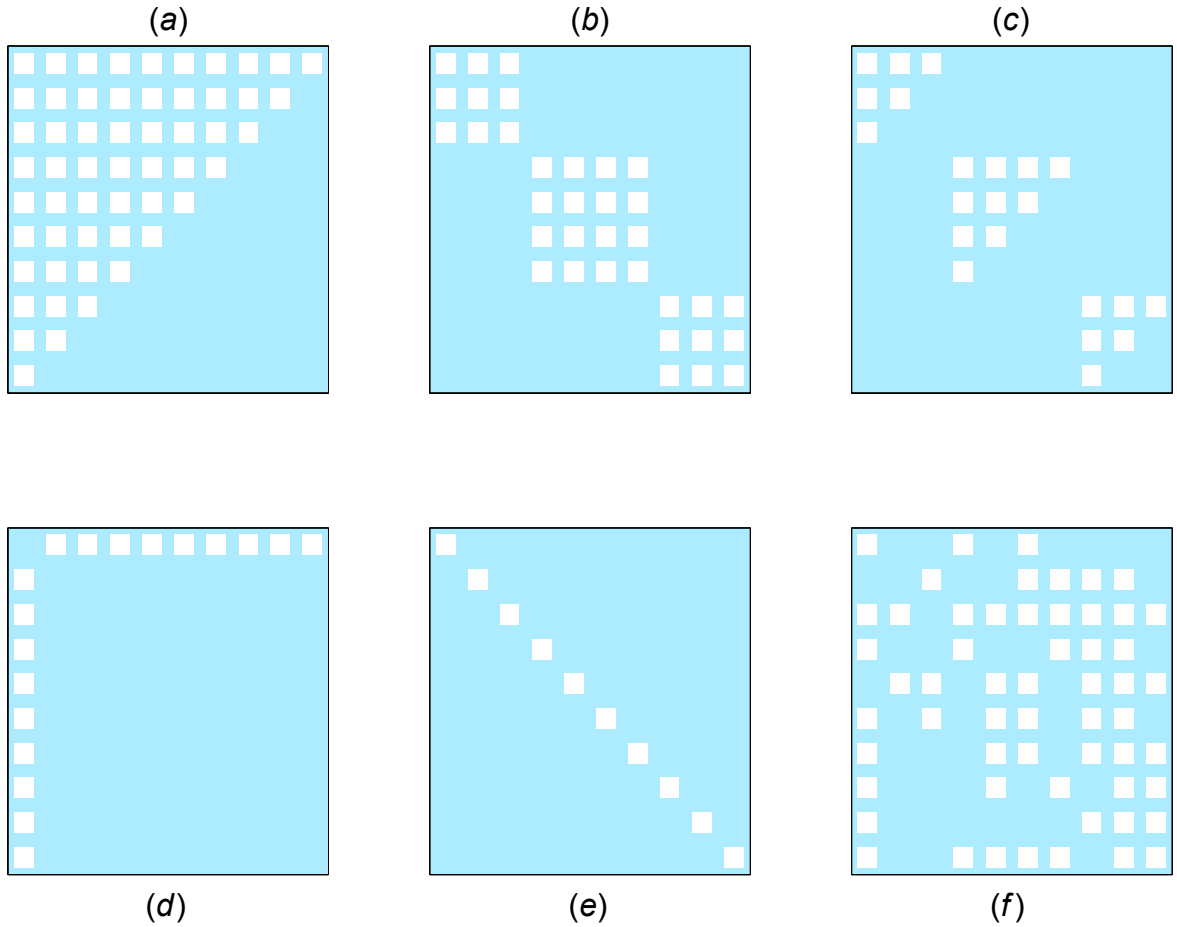


Figure 1.2: Example patterns in bipartite networks. Squares indicate an interaction between row and column nodes. (*a*) nested (*b*) modular (*c*) nested-modular (*d*) dependence asymmetry (*e*) one-to-one (*f*) random.

The two major network patterns that this thesis is concerned with are **nestedness** (Figure 1.2*a*) , a variation in generality between both sets of nodes such that specialist strategies are a subset of the more generalist strategies - this resembles an upper triangular structure in square networks; and **modularity** (Figure 1.2*b*), the tendency for a network to form densely interacting communities, known as **modules**, such that within-community connectivity is high and between-community connectivity is low. Other network representations are also illustrated. Figure 1.2*c* shows the pattern of nested-modularity, which is returned to in Chapter 3, which describes a modular network where each module contains a nested structure. Figure 1.2*d* shows dependence asymmetry [4, 5] and could be construed as an extreme case of nestedness. On the other hand, Figure 1.2*e* shows a one-to-one network structure, which could be construed as an extremal case of modularity where each module contains only a single interaction. Finally, Figure 1.2*f* depicts a random pattern of interactions between the row and column nodes with no discernible structure.

In this thesis, most attention focuses on the patterns of nestedness and modularity that have been shown to exist in empirical phage-bacteria infection networks [3], plant-pollinator networks [2] and elsewhere. In the next sections nestedness (Section 1.3) and modularity (Section 1.4) will be further explored.

1.3 Nestedness

Nestedness describes the tendency for specialist nodes of one type to interact with generalist nodes of the other type, such that more specialist nodes interact with a subset of the nodes that more generalised nodes are connected with. This definition can be satisfied in different ways and several types of nestedness are possible. Figure 1.3 provides three examples of nested networks for which the above conditions are true.

Nestedness has been observed in variety of empirical systems (e.g. [1, 3, 6, 7]) and it is thought that nestedness structure may contribute to the stability of ecological networks [8, 9, 10]. It may also have a role in explaining the formation and evolution of networks [11, 12, 13, 14]. These observations suggest that measuring nestedness is an important consideration when analysing bipartite networks. However, nestedness is a difficult concept to define operationally and many ways of measuring nestedness have been proposed, some of which are given in Table 1.1. However, it is possible to characterise nestedness measures by the underpinning methodology on which they are based. Table 1.1 gives an updated nestedness taxonomy (from [15]) with

six different classifications of nestedness measure. Different approaches to defining nestedness make different assumptions, and therefore have different characteristic strengths and weaknesses. No single method has yet become dominant, whilst new methods are being developed. Deciding on which nestedness measure to use in practical contexts is therefore difficult.

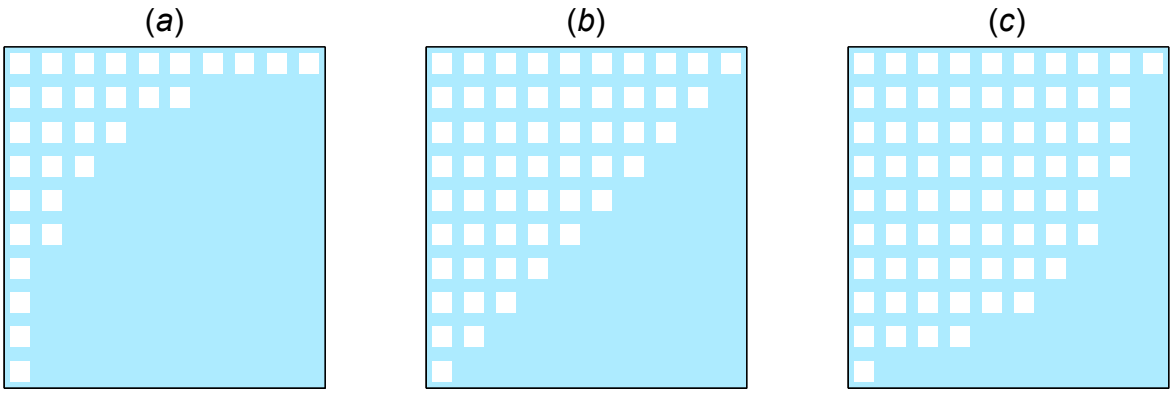


Figure 1.3: Nestedness depends on connectance. Squares indicate an interaction between row and column nodes. (*a*) Concave, (*b*) linear and (*c*) convex nestedness patterns are shown.

Gap metrics measure the way in which interactions are separated from absent interactions and whether the strategies of specialist nodes are subsets of the more generalist nodes [15]. However, they are typically only applied to one of the two sets of nodes in bipartite networks. To account for this it is recommended that the methods are applied to both sets of nodes and that the minimum value is used [15]. Despite this many gap metrics are correlated to network properties, such as size and fill, and sensitive to the way in which network data is presented. Only the discrepancy (BR) measure appears largely insensitive to these properties [19, 25, 15].

Measures characterised as ‘Temperature’ in Table 1.1 are different implementations of a particular description of nestedness. Temperature based methods aim to ‘pack’ the biadjacency matrix such that the majority of associations between rows and columns are on the upper side of the so-called ‘isocline of perfect nestedness’ which is a curve (or in the case of Matrix temperature two meeting linear lines [25]) drawn between opposite corners of the matrix whose curvature is defined by matrix connectance. The temperature is then calculated by counting the “surprises”; the number of absences above the line and the number of presences below the line, which are each penalised by their distance from the line.

Group	Measure	Reference
Gap metrics	Number of species absences from more rich sites (N0)	[16]
	Number of species presences from less rich sites (N1)	[17]
	Number of species at all sites (NC)	[18]
	Unexpected absences (UA)	[17]
	Unexpected precenses (UP)	[17]
	Unexpected total: UA + UP (UT)	[19]
	Number of departures (D)	[20]
	Discrepancy (BR)	[21]
Temperature	Nestedness temperature calculator (NTC0)	[22]
	BINMATNEST	[23]
	ANINHADO	[24]
	Matrix temperature (MT)	[25]
	nestedtemp (NTC)	[26]
Overlap	Number of supersets (HH)	[27]
	Nestedness metric based on overlap and decreasing fill (NODF)	[28]
	Percentage relativized nestedness (PRN)	[29, 30]
	Percentage relativized strict nestedness (PRSN)	[29, 30]
Deg. dist.	Nestedness	[31]
	Nestedness based on common neighbours (CMNB)	[8]
	Nestedness (JDM)	[32]
Placement	**Manhattan distance (MD)**	[33]
	Weighted-Interaction Nestedness Estimator (WINE)	[34]
Spectral	**Spectral radius (SR / $\rho(A)$)**	[35]

Table 1.1: Some of the plethora of nestedness measures for use in binary bipartite networks. Common shorthands and those used in this thesis are shown in brackets. Measures in bold are investigated in this work.

Overlap measures are classified as those based on checking rows and columns of a biadjacency matrix for overlapping similarity. Pairwise comparisons are performed between all pairs of rows and columns to establish the number of overlapping elements and these are used to quantify a nestedness measure. The number of supersets measure (HH) has no standardisation and can be heavily effected by interactions that deviate from a nested pattern and is not recommended [15]. NODF and PRN are similar in their construction, only differing in whether to include equal rows or columns as positive contributions to nestedness or not, which is discussed in [30, 36].

The degree distribution of a network is thought to be an important contributor to nestedness [37] as it describes how much a gradient exists between specialist and generalist nodes. Several measures are starting to attempt to include this network property in calculating nestedness.

Placement measures are calculated by assessing where interactions occur in the network's biadjacency matrix. Using the Manhattan distance (the sum of row and column indexes of existing edges) interactions that are placed further away from the top and left edges of the biadjacency matrix are increasingly penalised when looking for a nested pattern.

Finally the spectral classification includes the spectral radius [35] measure. This uses the maximum dominant eigenvalue of the network.

Both Johnson's measure of nestedness [32] and the spectral radius [35] are order invariant. **Order invariant** measures are calculated using only the topology of a network, such that there is no hierarchical ordering of nodes, meaning that rows and columns can be swapped without changing network topology. However, other nestedness measures are **order dependent** with measurements depending on the ordering of rows and columns in the biadjacency matrix.

If the optimal row and column ordering to maximise nestedness was found by exhaustive search (looking at every possible ordering of both rows and columns), then for a network with R row nodes and C column nodes there exist a total number (T_{nest}) of:

$$T_{nest} = R!C! \tag{1.1}$$

possible network configurations that would need to be evaluated. The larger a network becomes, the larger the number of potential network configurations. However, it is not necessary to check every possible configuration. For pairwise and placement measures (shown in Table 1.1) the most nested configuration is achieved when

rows and columns are ordered by decreasing node degree. On the other hand, for the nestedness measures that depend on calculating departures from the 'isocline of perfect nestedness' in the 'Temperature' category [22, 23, 24, 25, 26] shown in Table 1.1 genetic algorithms are used to converge towards the most nested configuration.

The methods above all considered binary networks. However, there also exist some measures of nestedness that can be applied to weighted networks, these are shown in Table 1.2. Nestedness in weighted networks can be interpreted as the largest edge weights being assigned to the edges between the most generalised nodes i.e. taking account of both the binary topological structure and also the strength of connections. WNODF is a simple extension of NODF that evaluates nestedness by ranking the edge weights of overlapping interactions, whilst WINE and SR use the absolute values of the weights to assess weighted bipartite networks. It is noted that WINE and SR can be applied to both binary and weighted networks, but WNODF can only be applied to weighted networks. Perhaps troublingly, as WNODF is a direct extension of NODF, WNODF interprets a pattern of nestedness as being one that is both binary nested (as in Figure 1.3) as well as quantitatively nested. The WNODF score for a network is no larger than the maximal NODF score. The benefits of the WINE and SR measures is that they can be applied to identify nestedness in networks where all the nodes interact. Even if the weighted edges in such a network were to follow a nested pattern, WNODF would fail to identify it, as this network has a NODF score of 0.

Measure	Description	Reference
Weighted NODF (WNODF)	Weighted NODF - is based on rank orders	[38]
Weighted-Interaction Nestedness Estimator (WINE)	Uses weighted Manhattan distance	[34]
Spectral radius (SR)	Maximum dominant eigenvalue	[35]
Relativized nestedness		[39]

Table 1.2: Nestedness measures for weighted bipartite networks. Measures in bold are described in this work.

Whilst out of the scope of this thesis that focuses on bipartite networks, it is noted that three measures listed in Table 1.1 [31, 35, 32] are also appropriate for use in studying nestedness in unipartite networks.

Due to the number of nestedness measures available it may be difficult to know which nestedness measure is the most suitable. A systematic benchmarking study by Ul-

rich and Gotelli [25] found that gap metrics underperform due to the fact they were originally designed to evaluate nestedness from the perspective of one type of node. However, they found even after making these metrics transpose invariant (so it does not matter which type of nodes are displayed as rows or columns) that these measures are highly sensitive to matrix size and fill [15]. However, of the gap metrics they analysed they found BR was largely insensitive to these network properties. To assess which nestedness measures are being used by researchers, the number of citations made to a paper describing a nestedness measure was recorded. Whilst a citation does not necessarily mean that a nestedness measure was used - using the average number of citations per year within a certain time window can be used as a proxy for the popularity of a nestedness measure. Looking at the average number of citations to some of the papers describing nestedness measures since 2013 (shown in Figure 1.4) it appears that several measures of nestedness are currently being used - and that measures differ in their popularity. Not all measures are assessed in Figure 1.4, but it serves to illustrate that several nestedness measures are being applied in the literature.

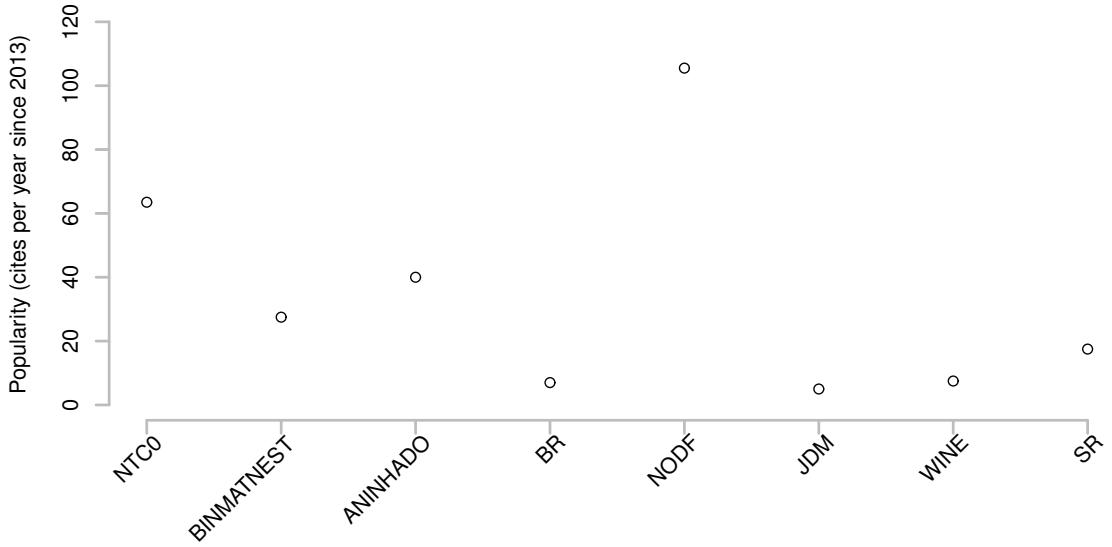


Figure 1.4: Popularity of some of the nestedness measures shown in Table 1.1, judged by the average number of citations per year since 2013 to the paper describing their methodology. Citation data is from google scholar and was retrieved on April 1st 2015. Several different measures of nestedness are being cited (and presumably used) and different measures have different levels of popularity.

Despite the recommendation of the discrepancy measure, by Ulrich and Gotelli 2007

[25], it does not appear to be as widely used as other nestedness measures. Instead it is NODF [28] and different flavours of the nestedness temperature calculator [22, 23, 24] that are most cited in Figure 1.4's analysis. There are at least three plausible reasons for this : (1) nestedness analysis is usually performed using specialist software (e.g. Table 1.3) - the measures used in nestedness analysis are limited by what is made available in the software, (2) researchers may be using nestedness measures on the basis of what other researchers are using, rather than the performance of the technique [40] (3) the amount of time since the method was created may delay popularity of certain measures as not all researchers are aware of the existence of newer techniques.

In this thesis I hope to clarify some of issues surrounding nestedness measure choice. There are a plethora of available nestedness measures which makes it difficult to know which should be used. It is important to know whether these measures are consistent with one another (and therefore the definition of nestedness), whether some measures of nestedness are redundant, and which measures are best able to discriminate nestedness. For these reasons it appeared pertinent to create a piece of nestedness software including several nestedness measures, so that results can be compared against each other, and to make this available to the research community. The measures chosen are highlighted in bold in Table 1.1 and Table 1.2. Further details for calculating these highlighted measures, that are made available in FALCON, are provided in Chapter 2. This software, FALCON, is described in Chapter 4, is open to allow new measures of nestedness that may be developed to be added. In order to evaluate the performance of nestedness measures an ensemble of 30,000 networks is analysed in Chapter 5. No large scale analysis using more than two nestedness measures has been conducted since Ulrich and Gotelli's benchmarking study [25]. In this time several new nestedness measures (including NODF [28], MD [33], SR [35] and JDM [32]) have been proposed.

1.4 Modularity

Modularity was originally described for unipartite networks [41] as a method for identifying communities of nodes within the wider network. Modularity assumes that nodes within a community are more likely to interact amongst themselves than across the rest of the network. As such they seek to identify communities composed of densely clustered edges [42]. Modularity is just one of several ways which have been proposed to identify communities in networks [43, 42]. Even then, several definitions of modularity exist. For bipartite networks several different modularity functions have

been described (e.g. [44, 45, 46]). Here, we focus on the case of **Barber's modularity** [44], that has also been termed bimodularity. Barber's modularity seeks to detect communities composed of nodes of both types simultaneously and does not allow overlapping communities. Other methods such as that of Guimerà [45], search for joint communities on each set of nodes separately. Both Barber's and Guimerà's modularity were recently reviewed [47] in the context of ecological networks. Other methods include that of Murata [46] which detects communities within each type of node simultaneously.

In this thesis only Barber's bipartite modularity [44] and a weighted version [48] based on the same framework are used. The definition of modularity is just a measure of a given community structure in a network - it does not provide a way of finding the best community structure by itself. Modularity is therefore a **goal function** which is used by additional algorithms that seek to maximise this condition. This is not a trivial task. To illustrate the problem, the network depicted in Figure 1.1 is revisited.

Within Barber's modularity framework the maximum number of modules within a network is given as the minimum of the two sets of nodes. For a network with R row nodes and C column nodes this is $F = min(R, C)$. Each module in a network must contain at least one row node and at least one column node. However, there are several different ways that m modules can be assigned across N nodes (of a single type). For example there are 36 ways to assign 3 modules (labelled A, B and C) across 4 nodes (for example the p nodes in Figure 1.1, each character represents the module label of a particular node):

AABC	BAAC	BBAC	ABBC	CCAB	ACCB
AACB	CAAB	BBCA	CBBA	CCBA	BCCA
ABAC	BACA	BABC	ABCB	CACB	ACBC
ACAB	CABA	BCBA	CBAB	CBCA	BCAC
ABCA	BCAA	BACB	ACBB	CABC	ABCC
ACBA	CBAA	BCAB	CABB	CBAC	BACC

whilst there are 6 ways of labelling the q nodes (in Figure 1.1) with 3 modules:

ABC	BAC	CAB
ACB	BCA	CBA

The total number of ways of organising the network in Figure 1.1 into exactly 3 modules is found as $\frac{36\times6}{3!} = 36$. The denominator is required to remove repeated modular configurations. For example in this case the arrangements:

p nodes	q nodes
AABC	ABC
AACB	ACB
BBAC	BAC
BBCA	BCA
CCAB	CAB
CCBA	CBA

all describe the same modular configuration, despite the different node labels. The operation of assigning exactly m modules across N nodes can be described by the function f:

$$f(m,N)=\sum_{i=0}^{m}\left((-1)^{i}\begin{pmatrix} m \\ m-i \end{pmatrix}(m-i)^{N}\right) \tag{1.2}$$

Therefore the total number of modular network configurations (T_{mod}) is given by:

$$T_{mod}=\sum_{a=1}^{F}\left(\frac{f(a,R)\,f(a,C)}{a!}\right) \tag{1.3}$$

For the example given in Figure 1.1 there are a total of 79 possible modular network configurations. Rather than exhaustively checking the modularity of all these network configurations, which would not be tractable in non-trivially sized networks, algorithms are used to attempt to maximise the modularity function. Several different algorithms are in use (including [44, 49, 48]). However, finding the best way to detect communities using modularity is tricky. This is because the landscape of modularity scores from different modular configurations of a network are “glassy” [50]. This implies a rugged landscape of modularity scores with many local modularity maxima. Finding an algorithm that can find the overall global maximum modularity score on such a glassy landscape is tricky. Finding algorithms that can perform both within a reasonable time and detect modules with a high level of accuracy is a current research area.

Most work on modularity in bipartite systems has focussed on networks of binary interactions. However, Dormann and Strauss [48] recently extended the formalism of Barber’s modularity and introduced an algorithm for detecting communities in weighted bipartite networks. Real world networks are not binary, and different nodes interact with each other with different link strengths. Weighted modularity therefore offers a new way of interpreting bipartite networks in modules around the edges with the strongest magnitudes.

In this thesis I seek to investigate how weighted modularity community structures differ from those found using binary modularity and which algorithms are most suited to each case. In Chapter 6 the method of Dormann and Strauss [48] is compared with a modification to an existing algorithm for detecting modules in binary networks [49], which I call LPAwb+. Both methods and the communities they detect are evaluated on weighted and binary ecological networks representing plant-pollinator systems. The LPAwb+ algorithms and resulting workflow are released to the community to aid studies of weighted modularity.

1.5 Software

To quantify patterns in bipartite networks, an observed network pattern is usually compared to reports of the same pattern from an ensemble of networks created using a null model that may conserve some core features of the observed network. Specialised software is usually employed to perform these calculations. Some options already exist for exploring modularity and nestedness and are shown in Table 1.3.

Software/code	O	N	M	W	Language
NESTEDNESS [51]	✗	✓	✗	✗	none (executable only)
Nestedness for Dummies (NeD) [52]	✓	✓	✗	✗	python / web GUI
vegan [26] + bipartite [53]	✓	✓	✓	✓	R
BiMAT (also BiWeb) [54]	✓	✓	✓	✗	MATLAB (python)
MODULAR [55]	✓	✗	✓	✗	C
FALCON (Chapter 2)	✓	✓	✗	✓	MATLAB/Octave/R
LPAwb+ (Chapter 4)	✓	✗	✓	✓	Julia/MATLAB/Octave/R

Table 1.3: Examples of software options for performing bipartite analysis. The top half of the table shows existing solutions, whilst the bottom half shows software introduced in this thesis. O: ‘Open’, denotes whether the source code is made available to potential users, N: ’Nestedness’ and M: ’Modularity’ denote whether these kinds of analysis are available, W: ’Weighted’ denotes whether methods for weighted bipartite networks are available and ’Language’ shows the coding langauge(s) employed to run the software.

Note that other software exists for computing modularity and nestedness, but some

of this has been excluded from Table 1.3 on the basis it focuses on just one particular measure of nestedness or modularity. It is noted that while NESTEDNESS [51] claims to be written in FORTRAN code, only the executable program is available - limiting the user base to those with the correct operating system to run the program, as well as the ability to extend the NESTEDNESS program or to verify its methods. It is for this reason it is marked as not being open. Whilst vegan [26] and bipartite [53] provide many methods for analysing bipartite networks (including nestedness and modularity) there is no built in function for performing and extracting statistics from a null ensemble analysis.

FALCON (Framework for Adaptive ensembLes for the Comparison Of Nestedness) is introduced in Chapter 4 as a tool to calculate and compare nestedness reported by different measures and null models. Due to the large number of nestedness measures present in the literature it is hard to know which should be used. Allowing users to see the effect of methodology on network analysis offers a way to start to bridge this gap. We were not the only research group to notice this potential software gap. Note that the paper describing the independently developed NeD software was published on the same day as the paper describing FALCON. LPAwb+ is introduced in Chapter 6 and proposes two alternative algorithms for investigating modularity in weighted bipartite networks (that can also be applied to binary networks). In both FALCON and LPAwb+ multiple languages were employed to widen the potential user base.

1.6 Overview of significant results

My thesis tackles the methodological considerations surrounding nestedness and modularity, that has lead to the development of two software tools that have been made available to the community: FALCON for evaluating and comparing nestedness in bipartite networks (described in Chapter 4), which uses a new concept of adaptive ensembles to choose the number of null models that a network should be evaluated against; and two algorithms for evaluating modularity in bipartite networks (described in Chapter 6).

This thesis shows that the evaluation of nestedness is confounded by user choice of measures, null models and effect sizes; and is potentially biased by network features, such as connectance - which can lead to different interpretations of nestedness. It is shown that when analysing a single network for nestedness using the

spectral radius with the degreeprobable-degreeprobable null model minimises type I and type II errors; but when comparing nestedness across multiple networks the discrepancy measure evaluated using adjusted normalised temperature scores as effect sizes produces the most useful comparison. Overall the differences between nestedness measure are enshrined in the different approaches to calculating this quantity - and whilst they may agree on what a nested network configuration is, they do not necessarily agree on the types of structures that exist with low nestedness, or how nestedness may degrade between these two states.

In terms of modularity it is shown that LPAwb+ and Exhaustive LPAwb+ offer viable alternatives to the existing QuanBiMo algorithm for detecting modularity in ecologically sized networks. Additionally it is shown these two new algorithms are both more robust and more efficient at detecting communities than the current approach for both weighted and binary networks. Whilst QuanBiMo and Exhaustive LPAwb+ were able to detect the same communities in smaller ecological networks which gives strength to both methods, troublingly it was shown that QuanBiMo produces sub-par solutions using its default settings in larger ecological networks. Interestingly the communities detected in weighted networks may be very different to those found in their corresponding binary networks (when all occurring interactions are set to 1). This raises questions about both weighted and binary modularity and how each should be used in a practical context.

1.7 Chapter summaries

Nestedness and modularity are the core themes that are investigated in this thesis. The following chapters seek to explore how nestedness and modularity can be applied to bipartite networks; and how to move to more rigorous methods.

The following chapter, Chapter 2, provides more details about how the nestedness measures and null models used in this thesis are calculated.

In Chapter 3 nestedness and modularity are used to analyse modelled coevolving communities of bacteria and their viruses, the bacteriophage. We show that a simple coevolutionary model can lead to a nested-modular community structure similar to that observed in natural phage-bacteria communities.

The rest of the thesis is devoted to improving the ways in which nestedness and modularity are measured. In Chapter 4 FALCON, a piece of software designed for

performing nestedness analysis in bipartite networks is introduced. FALCON was produced as a way to cross-examine some of the various techniques that are available in nestedness analysis (see Table 1.1) and encourage moving towards generating more robust approaches to nestedness analysis.

FALCON is then applied to a set of synthetic networks in Chapter 5 in order to gauge how sensitive nestedness analysis is to the choices of measure, null model and effect size. The analysis shows that inconsistencies exist between results obtained using different measures, with various sensitivities found dependent on the choice of null model. Furthermore, it is shown that nestedness is sensitive to network connectance; and that the choice of effect size metric is important if a comparison of the strength of nestedness between different networks is undertaken.

Modularity is returned to in Chapter 6. Various algorithms have been proposed to detect communities in networks by attempting to find the configuration that maximises modularity. However, most of this worked has focussed on binary networks. Recently a measure of weighted modularity has been introduced for bipartite networks [48]. Modification of a promising algorithm used in binary bipartite networks, led to the LPAwb+ algorithm, which shows improvements in searching for communities in weighted bipartite networks.

Finally, Chapter 7 presents the key findings and recommendations of the thesis. I discuss these findings in relation to the wider literature and consider some potential future directions for identifying structural patterns in bipartite networks. Being able to robustly identify structural patterns in networks will provide a firm theoretical basis for addressing questions concerning community interactions and aid the integration of this theory with empirical datasets.

Chapter 2

Calculating nestedness

2.1 Introduction

This chapter provides details about how the nestedness measures and null models used in this thesis are calculated. Section 2.2 details how to calculate some of the various nestedness measures, whilst Section 2.3 gives details about producing some possible null models. The measures and null models described below are available in the FALCON software whose development is reported in Chapter 4.

2.2 Detailed description of nestedness measures in FALCON

2.2.1 Binary measures

NODF

The nestedness measure based on overlap and decreasing fill (NODF) was first described by [28] and has since become one of the most popular methods for describing the nestedness of a matrix. NODF can be found as:

$$NODF = \frac{N_{col} + N_{row}}{\frac{(c(c-1)}{2} + \frac{(r(r-1))}{2}} = \frac{2(N_{col} + N_{row})}{c(c-1) + r(r-1)} \tag{2.1}$$

Here N_{col} and N_{row} are scores found by pairwise comparison of rows and columns, c is the number of columns, and r is the number of rows. N_{col} is found as the sum of scores from pairwise comparisons of each column against all columns to its right. If both columns have the same degree then the score is zero. If they have different degrees, the score is the percentage of elements in the second column which also appear in the first column. N_{rows} is found similarly for pairwise comparisons of each row against all rows below it. The sum of N_{col} and N_{row} is then normalised by the total number of pairwise comparisons. Values for NODF are between 0 (zero nestedness) and 100 (perfect nestedness). If the input matrix is first sorted to maximise nestedness by rank ordering rows and columns by degree, the form of NODF known as $NODF_{MAX}$ is found [30].

τ-Temperature and Manhattan Distance

The τ-Temperature [33] is a nestedness measure based on relative distances between matrix elements. Unlike other distance-based measures (such as NTC [22] and its better described successors BINMATNEST [23] and AININHADO [56]), the τ-Temperature does not use genetic algorithms to sort the data. The τ-Temperature is found by measuring the Manhattan distance D of the network matrix. This is the sum of the row and column indexes of all of the matrix elements A_{ij} that are filled:

$$D = \sum_{A_{ij}>0} (i + j) \tag{2.2}$$

Manhattan distance is lower in more highly nested networks, since rows and columns can be shuffled so that many of the elements appear in upper-left positions where row and column indices are low. Once D is found, a null model is chosen (cf. Section 4.4) and an ensemble of null matrices are created. By finding the mean average Manhattan distance from the ensemble, denoted $< D_{rand} >$, τ-Temperature can be calculated as:

$$\tau = \frac{D}{< D_{rand} >} \tag{2.3}$$

Values $\tau > 1$ imply that D is greater than $< D_{rand} >$ and the network is less nested than expected for a network with the properties defined in the null model. τ is better described as a test statistic of the Manhattan distance, than as a measurement of nestedness itself.

JDM Nestedness

The nestedness measure described in [32], here termed JDM after author initials, treats nestedness as a measure of dissassortativity between the nodes, i.e., negative correlation between row and column degrees for non-zero elements of the input matrix. Their measure calculates the overlap (as the sum of the elements in the squared adjacency matrix which shows the minimum number of length two paths needed to connect any two nodes) of the input matrix and normalises it by the expected nestedness of the configuration model (a random graph with the same empirical degree distribution as the input network) and thus discounts the effect of degree heterogeneity. This nestedness score is unbounded, but when close to 1 it indicates that

the matrix represents an uncorrelated random network. Unadjusted nestedness $\tilde{\eta}$ is calculated using the adjacency matrix a formed from the input bipartite matrix with r rows and c columns, where D is the node degrees in the adjacency matrix:

$$\tilde{\eta} = \frac{1}{(r+c)^2} \sum_{i}^{(r+c)} \sum_{j}^{(r+c)} \left(\frac{a_{ij}^2}{D_i D_j} \right) \tag{2.4}$$

Nestedness of the configuration model η_{conf} can be calculated as:

$$\eta_{conf} = \frac{r \left(\frac{\sum_i^r k_i^2}{c} \right) + c \left(\frac{\sum_j^c d_j^2}{r} \right)}{\left(\frac{\sum_i^r k_i}{c} \right) \left(\frac{\sum_j^c d_j}{r} \right) (r+c)^2} \tag{2.5}$$

which can also be written as:

$$\eta_{conf} = \frac{r < k^2 >_c + c < d^2 >_r}{< k >_c < d >_r (r+c)^2} \tag{2.6}$$

where k are the row degrees and d are column degrees in the bipartite matrix. This leads to the normalised measure of nestedness for bipartite networks defined by [32] as:

$$\eta_{bip} = \frac{\tilde{\eta}}{\eta_{conf}} \tag{2.7}$$

Nestedness Temperature

The original nestedness temperature calculator (NTC) [22] was vaguely described and therefore difficult to re-implement, leading to several subsequent variations utilising similar underlying principles [23, 56, 25, 26]. Here we have recoded the `nestedtemp` function from the R package *vegan* [26]. The nestedness temperature for an input matrix is based on the 'isocline of perfect order', a curve drawn from the lower-left corner of the matrix to the upper-right, with curvature defined by matrix fill (see figure 2.1). Row and column orderings are then permuted using a genetic algorithm to maximise the number of connections above the isocline and minimise connections below the isocline. The number of connections which violate these rules, termed 'surprises', are then counted and normalised to give a score between 0 (highly ordered) and 100 (highly disordered).

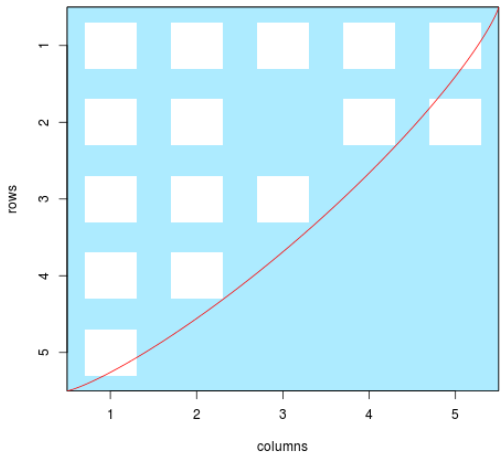


Figure 2.1: The NTC measure is based on an isocline of perfect order. Here there are two 'surprises'.

Discrepancy

Discrepancy [21], here denoted BR, quantifies nestedness as the difference between the input matrix and a perfectly nested matrix of the same dimensions and fill. A duplicate matrix P with the same row degrees as the input matrix, but where the 1s in each row are pushed as far to the left as possible (ignoring the effect this has on underlying network topology). The discrepancy is then found by subtracting the input matrix from this perfectly nested matrix and counting the number of 1s that remain – the number of differences between P and the input matrix. A different discrepancy score can be found by treating columns instead of rows, forming an alternative perfectly nested comparator matrix P' by pushing the 1s in each column to the top.(instead of the row) degrees of the input matrix to form P' a different discrepancy score can be found. Here we modify the original method of [21], which looks at discrepancy only in respect to P, and instead define discrepancy as the minimum of the individual discrepancy scores found from P and P', to remove any bias towards row or column nodes.

2.2.2 Quantitative measures

WNODF

The weighted NODF measure, WNODF [38], uses a similar algorithm to NODF, but is designed for use on quantitative rather than binary networks. In addition to asking which pairs of rows/columns are subsets of one another, WNODF utilises weight information by also requiring that the preceding row/column has greater values in the overlapping elements. In effect, WNODF is a stricter version of NODF; the maximum WNODF score that can be achieved for a quantitative matrix is equal to the NODF score for the binary matrix.

2.2.3 Both binary and quantitative measures

Spectral Radius

The spectral radius (SR) is defined as the absolute value of the maximum real eigenvalue from the adjacency matrix of a given input bipartite matrix. SR was proposed as a nestedness measure by [35] and can be applied to both binary and quantitative matrices.

2.3 Null models available in FALCON

2.3.1 Binary null models

Swappable-Swappable (SS)

The “swappable rows, swappable columns” (SS) null model conserves matrix dimensions (numbers of rows and columns) and fill. It is similar to ‘test one’ in [35], which works by shuffling elements at random within the matrix; however, it differs in that degenerate matrices (those containing rows/columns with no connections) are not permitted.

Fixed-Fixed (FF)

The "fixed rows, fixed columns" (FF) null model conserves dimensions, fill and degree distribution of the original matrix. It is the most strict null model we consider here and is known to suffer from Type II errors (i.e. a failure to detect nestedness) [57]. We use the Curveball algorithm [58] to generate null matrices of this type. It does this by iteratively choosing pairs of rows at random, compiling a list of column indices which contain filled elements in one but not both of the two rows. This list of column indices is then randomly permuted and reassigned to the two rows corresponding to the number of unique positions belonging to each of the original rows. It should be noted that the Manhattan distance is invariant to these permutations.

Cored-Cored (CC)

The "cored rows, cored columns" (CC) null model conserves dimensions and fill as in the SS null model, but also conserves some of the core structure found in the observed input matrix. It is found by performing a total of $M \times N$ trial-swaps on the $M \times N$ input matrix, where two matrix elements are randomly chosen and their values can be swapped only when this does not reduce the corresponding row or column degrees to zero. This ensures that the size structure is conserved and preferentially preserves specialist interactions within the network. The removed elements are then randomly reassigned to the remaining empty spaces to preserve matrix fill.

Degreeprobable-Degreeprobable (DD)

The "degreeprobable rows, degreeprobable columns" (DD) null model first described by [1] has subsequently been a popular choice for application to species-species nested comparisons. Matrix elements are probabilistically determined depending on the degree distribution of the rows and columns of the initial matrix as:

$$p_{ij} = \frac{1}{2}\left(\frac{d_j}{r} + \frac{k_i}{c}\right) \tag{2.8}$$

where p_{ij} is the probability of assigning a 1 to the ith row and jth column of the null matrix, d_j is the column degree of the jth column, k_i is the row degree of the ith row and r and c are the respective number of rows and columns. Due to the stochastic nature of this null model its output matrices will vary in size and fill.

Equiprobable-Equiprobable (EE)

The “equiprobable rows, equiprobable columns” (EE) null model is probabilistic and assumes that the probability of a connection occurring between two nodes is related to the number of total connections in the input matrix. Hence for an input matrix with fill M, r rows and c columns, the probability of a connection being present between two nodes is $p_{ij} = \frac{M}{r \times c}$. Due to the stochastic nature of this null model its output matrices will vary in size and fill. It is the least strictly defined null model we consider here and is known to suffer from Type I errors (a tendency to falsely detect nestedness) [57].

2.3.2 Quantitative null models

Binary Shuffle

This null model was employed by [35] and conserves the entire binary structure of the input matrix and the values of the elements in the matrix, but shuffles the order of these values randomly across the binary structure.

Conserve Row Totals (CRT)

This null model conserves binary structure and the row sum totals, but the values of the elements on each row are changed such that each connection in the row is assigned a random proportion of the row sum total.

Conserve Column Totals (CCT)

This null model conserves binary structure and the column sum totals, but the values of the elements in each column are changed such that each connection in the column is assigned a random proportion of the column sum total.

Row Column Totals Average (RCTA)

Both of the two above null models conserve information related to either the rows or the columns, giving this property precedent over that of the other entity. We also

introduce the Row Column Totals Average (RCTA) null model which uses the average of a single null model made from each of the CCT and CRT null models. As information from both rows and columns is utilised in the creation of this null model it may better fit with the context free ethos of nestedness we persue than either of CRT or CCT alone.

Chapter 3

Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks

Abstract

Phage and their bacterial hosts are the most diverse and abundant biological entities in the oceans, where their interactions have a major impact on marine ecology and ecosystem function. The structure of interaction networks for natural phage-bacteria communities offers insight into their coevolutionary origin. At small phylogenetic scales, observed communities typically show a nested structure, in which both hosts and phage can be ranked by their range of resistance and infectivity respectively. A qualitatively different multiscale structure is seen at larger phylogenetic scales; a natural assemblage sampled from the Atlantic Ocean displays large-scale modularity and local nestedness within each module. Here we show that such "nested-modular" interaction networks can be produced by a simple model of host-phage coevolution in which infection depends on genetic matching. Negative frequency-dependent selection causes diversification of hosts (to escape phage) and phage (to track their evolving hosts). This creates a diverse community of bacteria and phage, maintained by kill-the-winner ecological dynamics. When the resulting communities are visualised as bipartite networks of who-infects-whom, they show the nested-modular structure characteristic of the Atlantic sample. The statistical significance and strength of this observation varies depending on whether the interaction networks take into account the density of the interacting strains, with implications for interpretation of interaction networks constructed by different methods. Our results suggest that the apparently complex community structures associated with marine bacteria and phage may arise from relatively simple coevolutionary origins.

This chapter is based on the publication :

Beckett S.J., Williams H.T.P. 2013. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. Royal Society Interface Focus 3: 20130033. (DOI: 10.1098/rsfs.2013.0033).

We would like to thank Richard Boyle and two anonymous reviewers for their comments that have helped improve this article.

3.1 Introduction

Bacteriophage and their bacterial hosts are the most abundant and diverse replicating entities in the oceans, playing central roles in marine ecology and ecosystem processes [59, 60, 61, 62, 63, 64, 65]. Fast replication and high mutation rates mean that bacteria and phage can evolve – and coevolve – rapidly [66, 67, 68, 69], suggesting that coevolution will influence both ecological dynamics and ecosystem processes. Yet the basic mode of bacteria-phage coevolution is unclear. Experimental studies have demonstrated adaptation of resistance and infectivity ranges over just a few generations of laboratory coevolution [66, 67, 70, 71], often interpreted as a coevolutionary 'arms-race' in which hosts evolve to expand their range of resistance, while phages evolve to expand their host range. Unconstrained arms races are predicted to result in low diversity [72], with a single dominant host/phage strain, or perhaps two dominant host types if there is a trade-off between resistance and resource competition [66]. However, the short time intervals involved mean that the experimental coevolution data can be ambiguous and may sometimes also be consistent with a 'fluctuating selection' mode of coevolution in which infection is highly specific, so that hosts are more resistant to contemporary phage than ancestral or future strains [73]. Fluctuating selection dynamics are consistent with aquatic viral ecology models predicting kill-the winner dynamics [74, 75], whereby the most successful hosts (in terms of resource competition) are prevented from becoming dominant by increased viral predation. In kill-the-winner ecological dynamics, density-dependent predation by specialised viruses imposes negative frequency-dependent selection pressure on hosts, favouring rare phenotypes. Such dynamics are believed to support the maintenance of diverse communities of marine bacteria and phage [76, 77]. Recent genomic studies give empirical support for high natural diversity of marine bacteria and phage, with high specificity of infection and rapid coevolution [76, 78, 69, 79]. Thus despite much progress in experimental coevolution and marine microbial genomics, substantial uncertainties remain about the basic mode of phage-bacteria coevolution.

A complementary source of data about coevolution lies in the structure of natural phage-bacteria communities. Some recent data compilations have represented phage-bacteria communities as bipartite networks representing which phage strains were observed to infect which bacteria strains [3, 80, 81]. Statistical analyses of these 'phage-bacteria infection networks' (most often given in the form of binary matrices of presence(1)-absence(0) of pairwise infection) has so far focused on the matrix metrics of "nestedness" and "modularity" (see inset to Figure 3.1). Nestedness is a

measure of the extent to which the non-zero elements of each row (or column) in the matrix are a subset of the non-zero elements in the subsequent rows (or columns). In a perfectly nested matrix, the entries in each row (column) are a strict subset of the entries in the next row (column); thus each row (column) is nested inside the next row (column). In terms of phage-bacteria interaction, nestedness relates to the differentiation of strains along a gradient from specialist (small range) to generalist (large range). Here "range" refers, for bacteria, to the number of phage strains against which it is resistant, and for phage, to the number of bacterial strains it can infect. A perfectly nested pattern is one where the hosts and phages are each ranked along the specialist-generalist gradient such that the specialist strategies are subsets of the more generalised strategies. A modular network structure occurs when nodes can be partitioned into subsets such that most connections occur within rather than between the different subsets. For bacteria and phages modularity can be interpreted as a specialised interaction structure, without transitivity (i.e. where strains cannot be ranked by increasing range), in which distinct clusters of phage strains preferentially infect distinct clusters of bacterial strains.

Assuming that natural interactions between phage and bacteria are ultimately the product of coevolution, phage-bacteria interaction network structures may offer insight into the coevolutionary processes that produced them. At small phylogenetic scales these networks typically show higher-than-expected nestedness and lower-than-expected modularity [3]. High nestedness is consistent with an arms-race mode of coevolution, where hosts and phages evolve to increase their range of resistance/infectivity. However, the largest reported cross-infection assay involved 774 bacterial strains and 298 phage strains isolated from multiple geographically dispersed sites across the Atlantic Ocean [82]. Although no explicit genotyping was conducted in this study, it is likely that this dataset spans a broad phylogenetic scale. Reanalysis of interactions between 286 host strains and 215 phage strains from this dataset [80] found that the resulting network showed large-scale modularity, with local nestedness within each module. This is visualised in Figure 3.1, in which bacteria (rows) and phage (columns) were ordered (following [80]) to maximise statistical modularity and within-module nestedness. Here we highlight the identified modules by colour; interactions falling outside any identified module are shown in white. We also add two inset schematics illustrating perfectly modular and perfectly nested matrix structures. It is difficult to explain this "nested-modular" pattern as the result solely of arms race coevolution; the lack of global nestedness and the presence of distinct modules suggests that some additional mechanism is needed. While models of coevolution based on high specificity of infection often predict diversification [72] – and

might thus be invoked to explain formation of distinct modules – such models do not explain the presence of within-module nestedness.

Here we explore a simple model of coevolution based on genetic matching [83, 84, 85], which we dub the 'relaxed lock-and-key model'. The model is mechanistically justified by reference to coadaptation of (e.g.) phage tail-fibres and host surface receptors [86]. We show that relaxed lock-and-key coevolution is sufficient to produce the core structural features of observed phage-bacteria communities: stable high diversity of bacteria and phage, modularity at large phylogenetic scales, and nestedness at small phylogenetic scales. Furthermore, we show that the strength and statistical significance of the observed nested-modular pattern depends on how the interaction networks are formed. Here we contrast interaction networks based on the potential adsorption rate of each phage strain on each host strain with interaction networks based on actual infection rate measured in an ecological context. Our findings highlight difficulties with comparison of interaction networks constructed by these different methods.

In the next section, we present the relaxed lock-and-key coevolution model in the ecological context of a multi-strain chemostat. This is followed by presentation of results showing the co-diversification of bacteria and phage, the construction of associated adsorption rate and infection rate interaction networks, and analyses of network properties over time. Finally we discuss the relevance of the relaxed lock-and-key model for understanding natural phage-bacteria communities.

3.2 Model

We model bacteria-phage coevolution by adding mutation to numerical simulations of a multi-strain chemostat. Below we describe the ecological model, the coevolutionary model (including how infection rate is calculated for a given pair of bacteria and phage), and methods used to analyse phage-bacteria interaction networks. A description of the parameters used is given in Table 1. The model has been analysed previously [84, 85]; model sensitivity to key parameters is given in [85]. It is derived from a similar model [83] (with the principal difference being in how the evolutionary dynamics are evaluated), which in turn derives from an earlier single-strain ecological model of bacteria-phage growth in a chemostat [87].

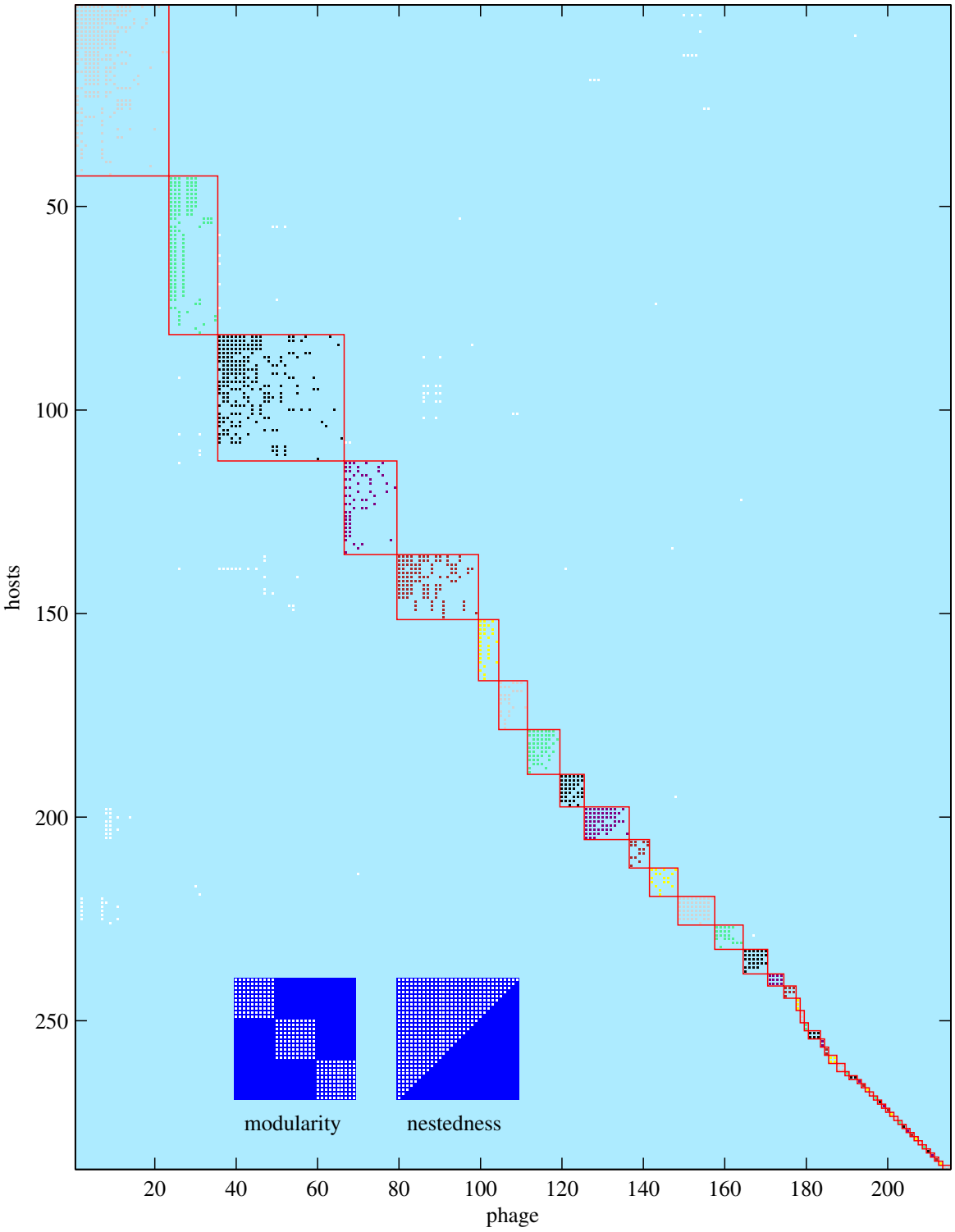


Figure 3.1: Nested-modular interaction structure of 215 phage strains and 286 bacteria strains sampled from the Atlantic Ocean (adapted from [80]). The plot shows which phage strains can infect which host strains from the dataset presented by Moebus and Nattkemper [82]. Flores *et al.* [80] re-sorted the interaction matrix to maximise modularity and within-module nestedness. Here we additionally highlight identified modules by shading; interactions falling outside any module are shown in white. We also add two inset schematics illustrating perfectly modular and perfectly nested matrix structures.

Symbol	Description	Value	Unit
R	Resource concentration	Variable	$\mu g\, ml^{-1}$
N_i	Density of host strain i	Variable	$cells\, ml^{-1}$
V_j	Density of phage strain j	Variable	$virions\, ml^{-1}$
N_{init}	Initial host density	4.6×10^4	$cells\, ml^{-1}$
V_{init}	Initial phage density	8.1×10^5	$virions\, ml^{-1}$
ω	Chemostat dilution rate	0.0033	min^{-1}
R_0	Resource supply concentration	2.2	$\mu g\, ml^{-1}$
ε	Resource conversion rate	2.6×10^{-6}	$\mu g\, cell^{-1}$
γ	Maximum resource uptake rate	0.0123	$\mu g\, min^{-1}$
K	Half-saturation constant	4	$\mu g\, ml^{-1}$
δ_i	Growth scaling for host h_i	Range $[\delta_{min}, \delta_{max}]$	$scalar$
δ_{min}	Min. growth scaling factor	0.8	$scalar$
δ_{max}	Max. growth scaling factor	1.2	$scalar$
ϕ	Maximum adsorption rate	0.104×10^{-8}	$ml(min\, virion)^{-1}$
θ_{ij}	Ads. scaling for v_j on h_i	Range $[0, \phi]$	$scalar$
β	Burst size	71	$virions\, cell^{-1}$
h_i	Genotype of bacteria i	Range $[0, 1]$	$scalar$
v_j	Genotype of phage j	Range $[0, 1]$	$scalar$
$\hat{h}_i$	Resistance phenotype of bacteria i	Range $[0, 1]$	$scalar$
$\hat{v}_j$	Infection phenotype of phage j	Range $[0, 1]$	$scalar$
h_{init}	Initial bacteria genotype	0.2	$scalar$
v_{init}	Initial phage genotype	0.2	$scalar$
S	Specificity of phage	100	$scalar$
μ_N	Host mutation rate	10^{-6}	$cell^{-1}$
μ_V	Phage mutation rate	10^{-5}	$virion^{-1}$
σ_N	Std. dev. of host mut. range	0.01	$scalar$
σ_V	Std. dev. of phage mut. range	0.01	$scalar$
M_N	Bacterial mutation size	Random variable	$scalar$
M_V	Phage mutation size	Random variable	$scalar$
Δt	Integration timestep	10	min
T	Simulation duration	5×10^7	min
ρ	Resolution of genotype diversity	0.001	$scalar$
L	Chemostat volume	1	ml

Table 3.1: Model parameters and variable definitions. Variables can change during a simulation. Numerical values are parameters fixed for the duration of a simulation. Range values are deterministically calculated from other variables/parameters.

3.2.1 Multistrain chemostat model

The ecological model represents the interactions between multiple strains of bacteria and phage in a single-resource chemostat. Resource concentration R and densities of the i^{th} strain of bacteria N_i and the j^{th} strain of phage V_j are governed by the system of equations below:

$$\frac{dR}{dt} = -\omega(R - R_0) - \sum_i \varepsilon \frac{\gamma \delta_i R N_i}{R + K}$$

$$\frac{dN_i}{dt} = -\omega N_i + \frac{\gamma \delta_i R N_i}{R + K} - \sum_j \phi \theta_{ij} N_i V_j$$

$$\frac{dV_j}{dt} = -\omega V_j + \sum_i \beta \phi \theta_{ij} N_i V_j \tag{3.1}$$

Resource concentration is affected by the chemostat washout rate ω, the supply concentration R_0, and uptake by all bacterial strains. Bacterial resource uptake is governed by Monod kinetics [88] with half-saturation rate K and maximum uptake rate γ, adjusted for each bacterial strain i by a genetically encoded scaling coefficient δ_i. Bacterial strain density is a function of washout, population growth and lysis. Resource uptake is converted directly into bacterial population growth via resource conversion constant ε. Each bacterial strain is potentially susceptible to infection by every strain of phage, depending on genetic match. Phage strain density is determined by washout and the sum of production on all available hosts. Adsorption of phage j to host i is the product of the maximum adsorption rate ϕ and a scaling coefficient θ_{ij}. Every adsorption event leads to infection and instantaneous cell lysis (we assume no latent period) creating new phage with burst size β.

3.2.2 Relaxed lock-and-key coevolution model

We model evolution in the multi-strain ecological model by adding a mutation process that introduces new variants of existing bacteria and phage strains. Uncompetitive strains are eventually removed by chemostat dilution. Thus we have a simple model in which bacteria and phage phenotypes can evolve by natural selection. We do not separate the evolutionary and ecological timescales, but instead assume a fixed probability of mutation per new cell/phage and allow evolutionary dynamics to play out

in our numerical simulations. We ran our simulations for 5×10^7 minutes, sufficient for the dynamics to reach a quasi-stable equilibrium. This timescale is chosen to allow a relatively slow evolutionary dynamic in the context of faster ecological dynamics, and is not intended to accurately reflect analogous timescales in natural systems.

Bacterial h and phage v genotypes are modelled as single values in the range $[0, 1]$ (binned at resolution $\rho = 0.001$). Bacteria and phage have mutation rates μ_N and μ_V respectively, applied stochastically for each new cell/phage. Mutation creates a single cell/phage with a genotype created by adding a normal deviate to the parental genotype, with standard deviation (mutation range) of σ_N or σ_V for bacteria and phage respectively. If the density of any population falls below $1\, cell\, ml^{-1}$ or $1\, virion\, ml^{-1}$ (possible due to the continuous nature of the mathematical abstraction), that population is assumed to be lost and is removed from the system. Any simulations where genotypes reached the edges of the permitted range $[0, 1]$ were discarded; however parameters were chosen to ensure this did not occur.

The relaxed lock-and-key coevolution model is created by enforcing a dependence of adsorption rate on the genetic similarity of host and phage. The relative adsorption rate θ_{ij} of phage j on host i is given by:

$$\theta_{ij} = e^{-S\left(\hat{h}_i - \hat{v}_j\right)^2} \tag{3.2}$$

where $\hat{h}_i$ is the bacterial resistance strategy for the i^{th} host encoded by genotype h_i, $\hat{v}_j$ is the phage infectivity strategy for the j^{th} phage encoded by genotype v_j and S represents infection specificity. This function has Gaussian form, with specificity governing the width of the infection curve, i.e. high specificity indicates narrow host range, whilst low specificity indicates wide host range. Hence the closer the numerical values of strategies $\hat{h}_i$ and $\hat{v}_j$ are, the greater the rate of adsorption between host i and phage j.

Bacterial genotype also specifies a growth-rate scaling trait δ ; where each bacterial genotype h_i maps to a growth rate δ_i. Here we impose a simple growth rate fitness landscape with a singular peak of $\delta_{max} = 1.2$ at $h = 0.5$, falling linearly to a minimum of $\delta_{min} = 0.8$ at the edges of the range (i.e. for $h = 0$ and $h = 1$). The growth rate trait δ is a coefficient that affects population growth rate by scaling the rate of resource uptake by bacteria (see Equation 3.1). It is important to note that there are no inherent differences in bacterial resistance (that is, all bacteria have the same size of range of resistance); thus there are no costs of resistance and no explicit trade-offs between growth rate and infection rate. Ecological trade-offs between bacterial growth

and susceptibility to infection can emerge [85], but these depend on the density and composition of the contemporary phage and bacterial communities.

3.2.3 Infection network analysis

Interactions between bacteria and phage strains (i.e. who-infects-whom) can be represented as a network [81]. Such networks are bipartite graphs with interactions between two types of node (bacteria and phage). We use two forms of interaction network to visualise our model phage-bacteria communities. The first type is formed from the adsorption rate $\phi\theta_{ij}$ of phage j on host i , which gives a pairwise interaction matrix for all strains present in the community irrespective of their abundance. The second type of interaction network is formed from the actual infection rates for phage j on host i in the current ecological context, calculated as $\phi\theta_{ij}N_iV_j$ from the adsorption rate and the current densities of each strain, giving a more ecologically relevant measure of interaction.

Since both adsorption rate and infection rate are quantitative metrics, we create binary interaction matrices by applying threshold filters. The resulting binary matrices consist of 1s where the pairwise interaction is strong and 0s elsewhere. The binary matrices can be analysed using standard metrics for nestedness and modularity. They can also be compared to reported interaction networks, which are typically given in binary form. We used the package BiWeb [89], which utilises the LP-BRIM algorithm to find the partition that best maximises Barber's modularity (Q_b) for bipartite networks [44, 90]. Since the LP-BRIM sorting algorithm is sensitive to initialisation, we repeated the modularity assessment 5 times for each measurement, taking the maximum score returned. We measured nestedness using the deterministic NODF (nestedness metric based on overlap and decreasing fill) algorithm [28], which returns a score in the range $[0, 100]$ (where 100 indicates a perfectly nested structure). NODF normalises for matrix size, allowing matrices of differing sizes to be compared.

To quantify the statistical significance of the nestedness (NODF) and modularity (Barber's Qb) scores measured for a particular binary matrix, we calculate statistical significance p as the likelihood of achieving a score greater than or equal to the score for the input matrix in a sample of M matrices from a null distribution of random matrices. Where this method returns $p = 0$ we conservatively assign $p < \frac{1}{M}$. We use a null model in which matrix size (number of rows and columns) and fill (number of 1s) are conserved, but where adjacency (position of 1s) is randomly reassigned. Thus we

maintain the numbers of host and phage strains, and the total number of host-phage interactions, but reassign the pattern of who-infects-whom. We estimate the null distribution using a sample of M matrices generated using a stochastic algorithm with pseudo-random numbers. For significance of NODF scores, we create the sample in two sets, adding matrices to both sets until the mean NODF scores for both sets converge within a tolerance bound of 0.01. The sample used to estimate the null distribution is then formed as the union of both sets. Thus for NODF, the sample size M varies according to how many null matrices are needed to achieve convergence of the means, with a minimum sample of 500 null matrices computed in each case. This method gives reliable estimation of the underlying distribution while retaining computational efficiency. For the modularity scores, we used $M = 100$, assigning each null matrix the highest Barber's modularity score from 5 random initialisations of the LP-BRIM algorithm. Software based on these methods was developed and is described in Chapter 4.

3.3 Results

In previous work [85] we have shown that in the absence of phage, resource limitation leads to competitive exclusion of slow-growing bacteria by fast-growing bacteria. Faster-growing populations draw down resource concentration to a limiting level at which slower-growing populations cannot be sustained against losses from washout and are lost from the community. Thus in the absence of phage, bacteria evolve to the fastest-growing genotype permitted by the simple unimodal growth rate fitness landscape (here located at $h = 0.5$ - see Section 3.2.2). In the presence of phage, host evolution is affected by the additional coevolutionary selection pressures imposed by phage predation.

The relaxed lock-and-key coevolution model robustly produces diversification of bacteria and phage. Figure 3.2 shows a simulation run initialised with a single bacterial strain and perfect-match phage strain (with $h = v = 0.2$). In the first stage (until $t \approx 0.5 \times 10^7 min$) bacteria evolve to increase growth rates, while phage evolve to track their hosts through genotypic space. Once the bacteria reach the maximum growth rate genotype at $h = 0.5$, they can no longer improve fitness by increasing growth rate (and would remain at this fitness peak indefinitely in the absence of phage [85]). However, phage create a strong selective pressure for host diversification due to density-dependent predation, which favours host mutants with a lower genetic match

to dominant phage strains. This causes an evolutionary branching event to occur between $t \approx 0.6 \times 10^7 min$ and $t \approx 1.1 \times 10^7 min$. Further evolutionary branching events occur until $t \approx 2.5 \times 10^7 min$. After this period of coevolutionary diversification the distribution of strains settles down to a quasi-stable state for the remainder of the simulation. At this stage there are 5 clearly identifiable clusters of similar genotypes for both bacteria and phage, where each cluster (hereafter "species") represents an ecologically similar (but genetically diverse) sub-population. Each phage species is attracted towards its two flanking host species, while each host species is repelled by its flanking phage species; this process of attraction and repulsion sometimes results in transient oscillatory dynamics (this effect is clearly seen for $t \approx 2 - 3 \times 10^7 mins$).

Figure 3.3 visualises community structure for a timeslice from the simulation taken at $t = 3.5 \times 10^7 min$. There are 64 bacterial strains and 97 phage strains present in the community. Densities of the different strains vary widely and are unevenly distributed, but 5 clearly identifiable species of similar genotypes in both the bacteria (lower left) and phage (lower right) populations are visible. For ease of reference we label the host species H1-H5 and the phage species P1-P5.

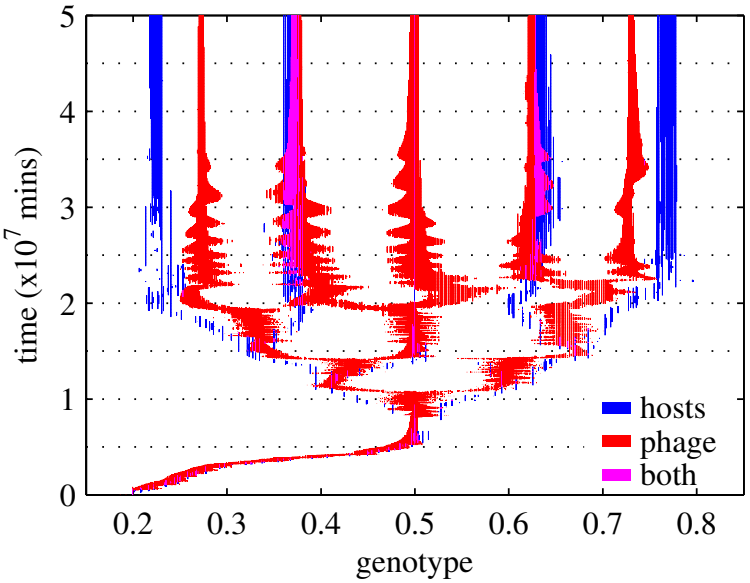


Figure 3.2: Coevolutionary diversification of bacteria and phage strains with relaxed lock-and-key model. Plot shows phage-bacteria community composition over time for a simulation run initialised with a single host strain and perfectly matched phage (initial genotypes $h = v = 0.2$). Genotypes with non-zero density for bacteria, phage, or both bacteria and phage are highlighted. The fastest-growing bacteria genotype is located at $h = 0.5$. The cross-section of the community at time $t = 3.5 \times 10^7$ is examined in Figure 3.3. Horizontal dotted lines indicate time points for which interaction matrices are presented in figures 3.4 and 3.5.

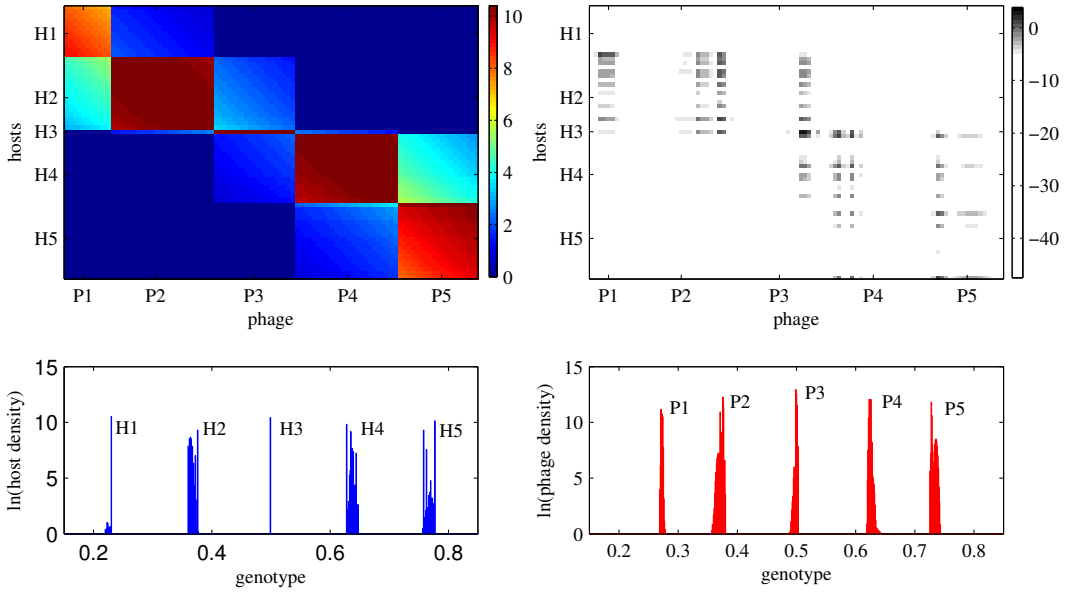


Figure 3.3: Snapshot of phage-bacteria community structure and interaction during the case study simulation (measured at $t = 3.5 \times 10^7 min$). *Upper left:* Adsorption rate matrix plot showing pairwise adsorption rate ($\phi\theta_{ij}$, values scaled$\times 10^{-10}$) for each host (i) and phage (j) strain present in the current community. *Upper right:* Infection rate matrix plot showing actual infection rate in the context of the community (ln $\mid \phi\theta_{ij} N_i V_j \mid$) for each host-phage pair. *Lower left:* Current density of all bacterial genotypes (ln $\mid N_i \mid$). *Lower right:* Current density of all phage genotypes (ln $\mid V_j \mid$). For ease of reference, labels H1-H5 and P1-P5 are manually attached to identify bacteria and phage "species".

We visualise interactions between phage and bacterial strains in Figure 3.3 by plotting the (density-independent) adsorption rate matrix (upper left) and the (density-weighted) infection rate matrix (upper right). These matrices represent a snapshot of the coevolving interactions between phage and bacteria - note that the matrices include all strains that are present, but do not cover the whole genetic space (i.e. absent strains are not plotted). Each host species interacts strongly with a single phage species, as indicated by the modular structure apparent in the adsorption rate matrix (e.g. P1 is specialised on H1, P2 on H2, and so on). However, there are also weaker interactions with adjacent phage species (e.g. H2 is also weakly affected by P1 and P3). The infection rate matrix gives a different view of the community; whereas the adsorption rate shows a potential interaction, the infection rate shows the interaction in terms of actual mortality of the host caused by the phage in ecological interaction. The infection rate matrix shows that most of the potential interactions shown in the adsorption rate matrix are ecologically insignificant, since only a few strains in each species are present at sufficient density for a strong interaction to occur. Also there

is a many-many interaction structure, with each bacterial strain infected by multiple phage strains, and each phage infecting multiple bacteria. Interestingly, the infection rate matrix also suggests a degree of modularity, though module membership appears different to that seen in the adsorption rate matrix.

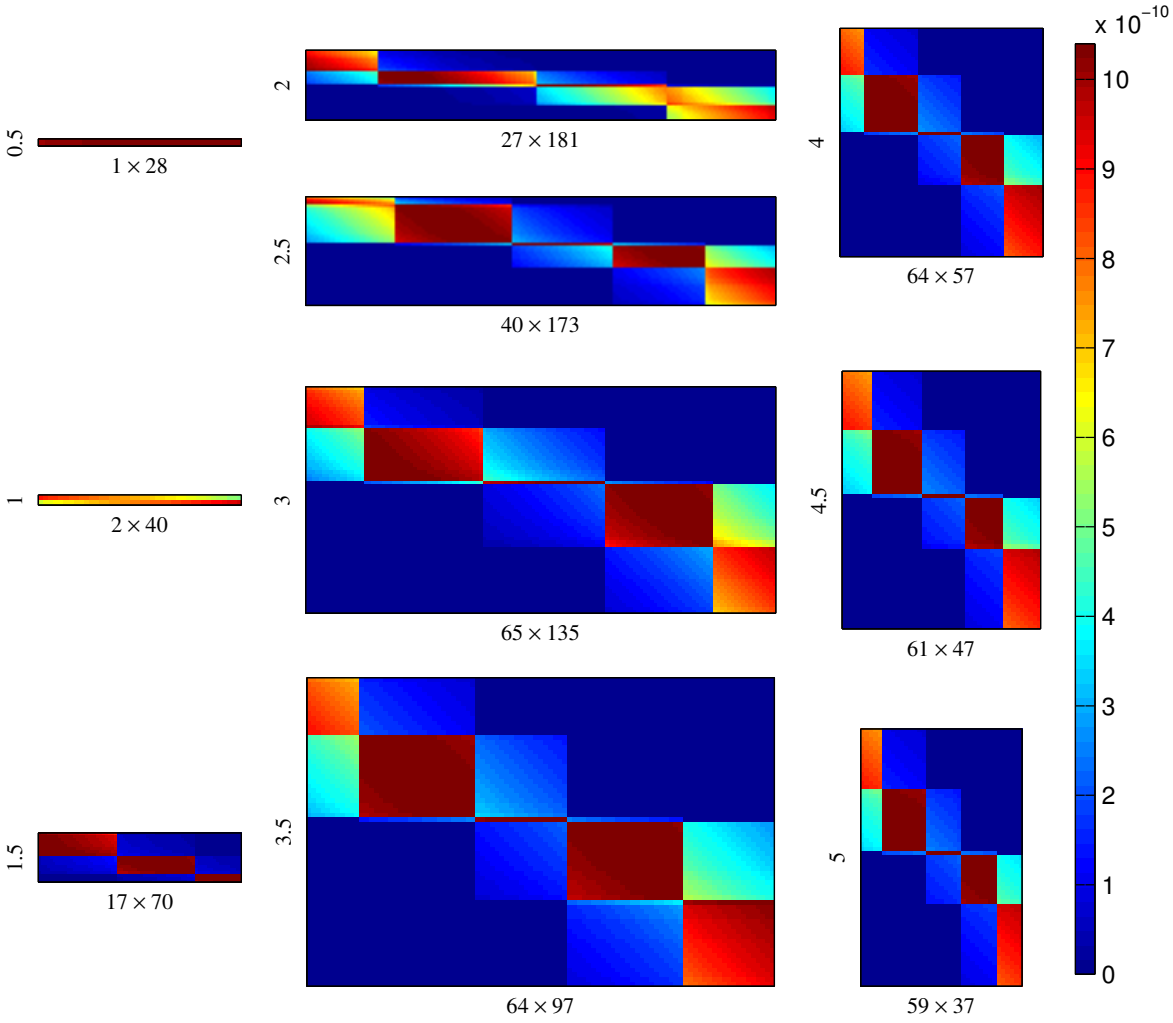


Figure 3.4: Coevolutionary formation of the adsorption rate interaction network over time. Colour shows pairwise interaction strength ($\phi\theta_{ij}$) for bacterial strains (rows) and phage strains (columns). Network size (number of bacterial strains x number of phage strains) is shown as the x-axis label for each plot, timepoint ($\times 10^7 min$) is shown as the y-axis label. Plots show timepoints corresponding to the dotted gridlines shown in Figure 3.2, with time increasing down each column, from top-left to bottom-right.

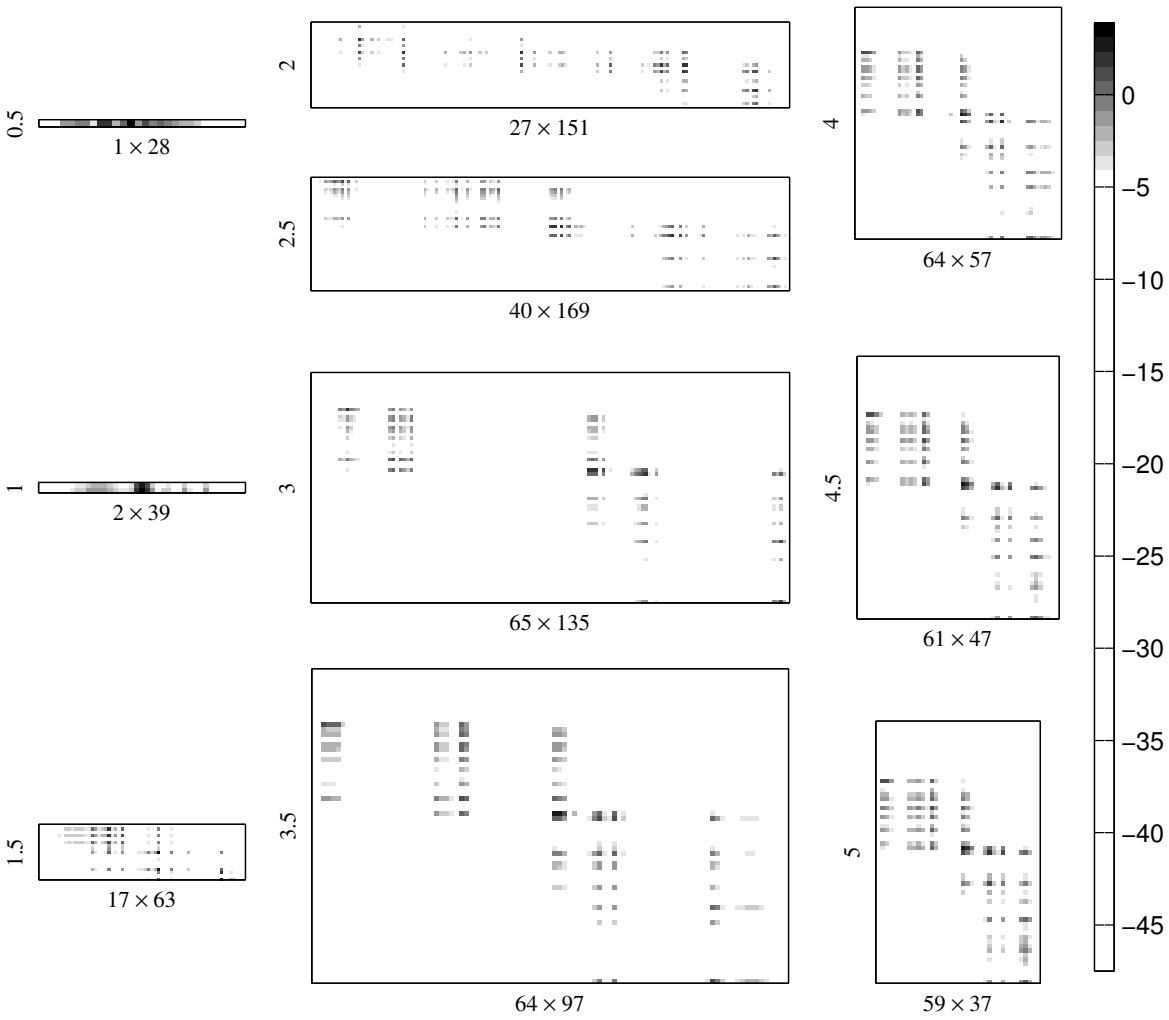


Figure 3.5: Coevolutionary formation of the infection rate interaction network over time. Grayscale shows pairwise interaction strength ($\ln \mid \phi\theta_{ij}N_iV_j \mid$) for bacterial strains (rows) and phage strains (columns). Network size (number of bacterial strains x number of phage strains) is shown as the x-axis label for each plot, timepoint ($\times 10^7 min$) is shown as the y-axis label. Plots show timepoints corresponding to the dotted gridlines shown in Figure 3.2, with time increasing down each column, from top-left to bottom-right.

Figures 3.4 and 3.5 show how the adsorption rate and infection rate interaction networks respectively change over time as bacteria and phage coevolve. Initially there is low diversity and the matrices are small, but over time matrix size increases as hosts and phage diversify. At later timepoints, matrix size reduces somewhat, reflecting an overall drop in diversity as competition excludes weaker strains and the system converges to a quasi-stable state. Matrix sizes show trends in strain diversity, with host diversity rising to a stable level around 60-65 strains and phage diversity initially rising, then falling, during the course of the simulation. This trend is reflected at smaller scale in the sizes of the modules in the adsorption rate matrix. Modular inter-

action structure is always apparent in the adsorption rate matrix, though the number of modules varies over time. Structure in the infection rate matrix is harder to discern visually, though some modularity seems to be apparent at later timepoints.

We converted our quantitative interaction matrices into binary matrices using a threshold filter (see Section 3.2.3), to give matrices suitable for comparison with reported phage-bacteria interaction networks [3, 80]. The resulting binary networks are a coarse-grained representation of the underlying data, but can be used with the LP-BRIM and NODF algorithms to quantify modularity and nestedness respectively. Figure 3.6 shows the effect of different thresholds. The size and fill of the resulting binary matrix depends on the threshold used. The importance of choosing an appropriate threshold is well illustrated by the adsorption rate matrices, where choosing too low or too high a threshold results in binary matrices that do not capture the modular structure apparent in the raw data; setting too low a threshold gives overlapping modules, setting too high a threshold may result in loss of some modules. For the remainder of the analysis, we use an adsorption rate threshold of 0.8ϕ and an infection rate threshold of $0.0083\,cells\,(ml\,min)^{-1}$, which typically gave good agreement with visual interpretation of the raw data. The results presented below for adsorption rate matrices are weakly sensitive (but qualitatively robust) to the choice of threshold (data not shown). Results for infection rate matrices are robust to choice of threshold.

The order of rows and columns in the binary matrices can be permuted without changing the underlying network structure. Figure 3.7 shows binary networks formed from the adsorption rate and infection rate matrices for the timeslice ($t = 3.5 \times 10^7 min$) shown in Figure 3.3. Different row and column re-orderings are applied in each panel. The upper binary matrix shows a random re-ordering that removes the phylogenetic ordering that arises from model formulation (whereby hosts and phages are ordered by genotype), thus representing how unsorted results of an experimental infection assay might appear. The middle binary matrix is sorted to maximise modularity using the LP-BRIM algorithm, with identified modules highlighted in colour. The lower binary matrix is sorted to maximise nestedness using the NODF algorithm.

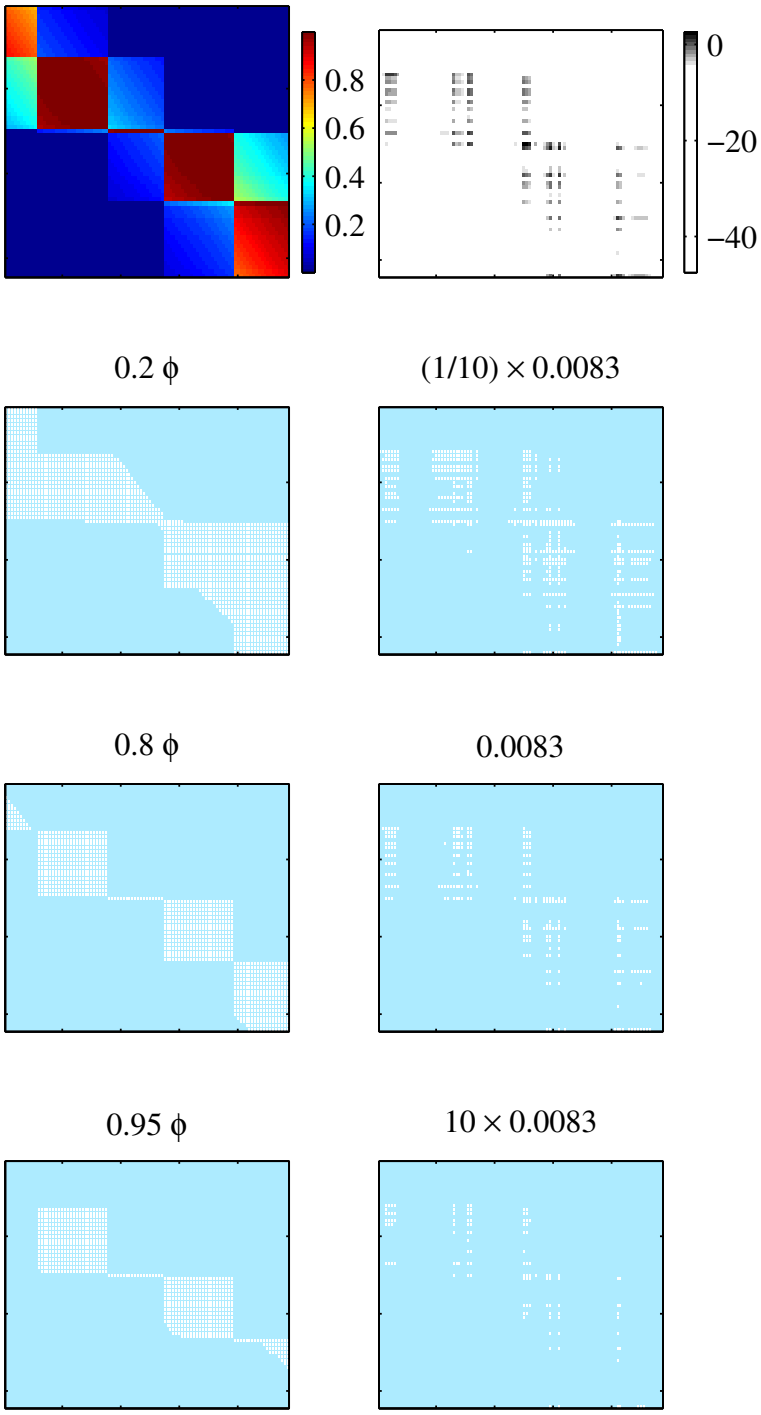


Figure 3.6: Effect of choice of filter threshold on binary representation of interaction matrices. Top plot in each column shows the raw (unfiltered) matrix data for pairwise adsorption rates (left column) and infection rates (right column). Lower plots show binary matrices formed by filtering raw data with different thresholds. Threshold values of 0.8ϕ for adsorption rates and $0.0083\,cells\,(ml\,min)^{-1}$ for infection rates were used for all subsequent analyses.

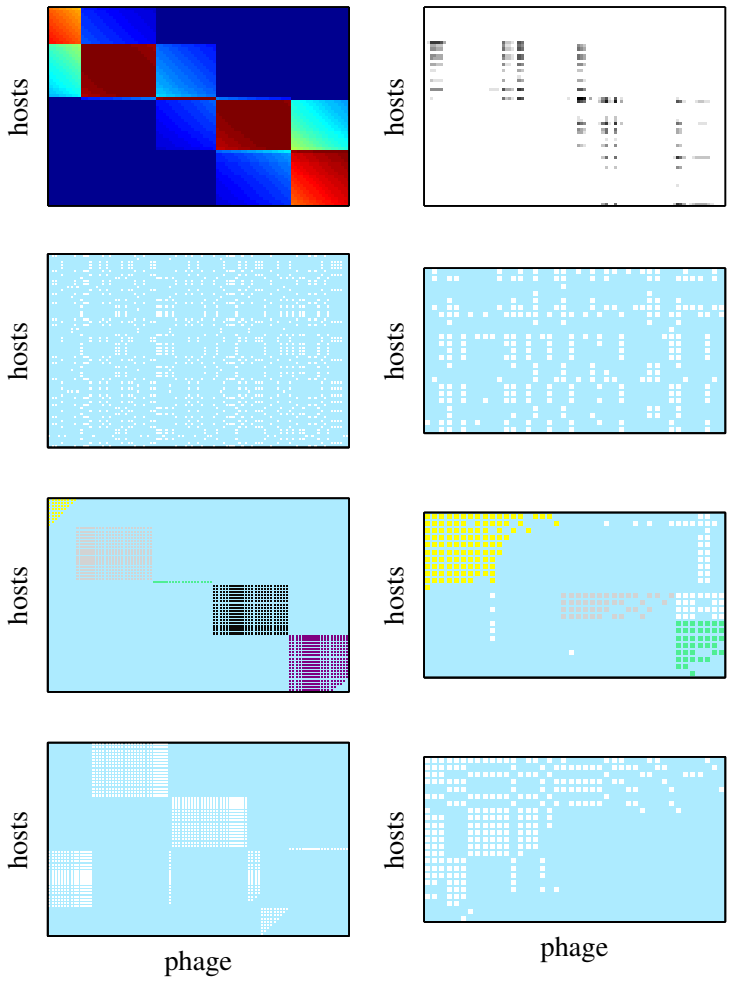


Figure 3.7: Comparison of binary interaction networks formed by applying a threshold filter to (left column) adsorption rate and (right column) infection rate matrices, for a snapshot of the system at $t = 3.5 \times 10^7 min$, arranged with host strains on rows and phage strains in columns. *Top panel:* quantitative interaction data. *Lower 3 panels:* binary matrices formed using a threshold filter of the quantitative data, sorted for (upper) random permutation, (middle) maximisation of modularity, (lower) maximisation of nestedness.

For the same case study, we used binary interaction matrices to study the temporal dynamics of nestedness (from the NODF algorithm) and modularity (using Barber's modularity score returned by the LP-BRIM algorithm). Figure 3.8 shows mean values across an ensemble of ten simulations with identical parameters. The ensemble runs differ only in the pseudo-random numbers used in the stochastic mutation process. This stochastic variation leads to different timing and order of evolutionary branching events, but the quasi-steady state reached is similar in all simulations (data not shown). Host strain diversity rises over time before levelling off, while phage strain diversity is highest during the diversification phase and decreases as quasi-steady

state is approached.

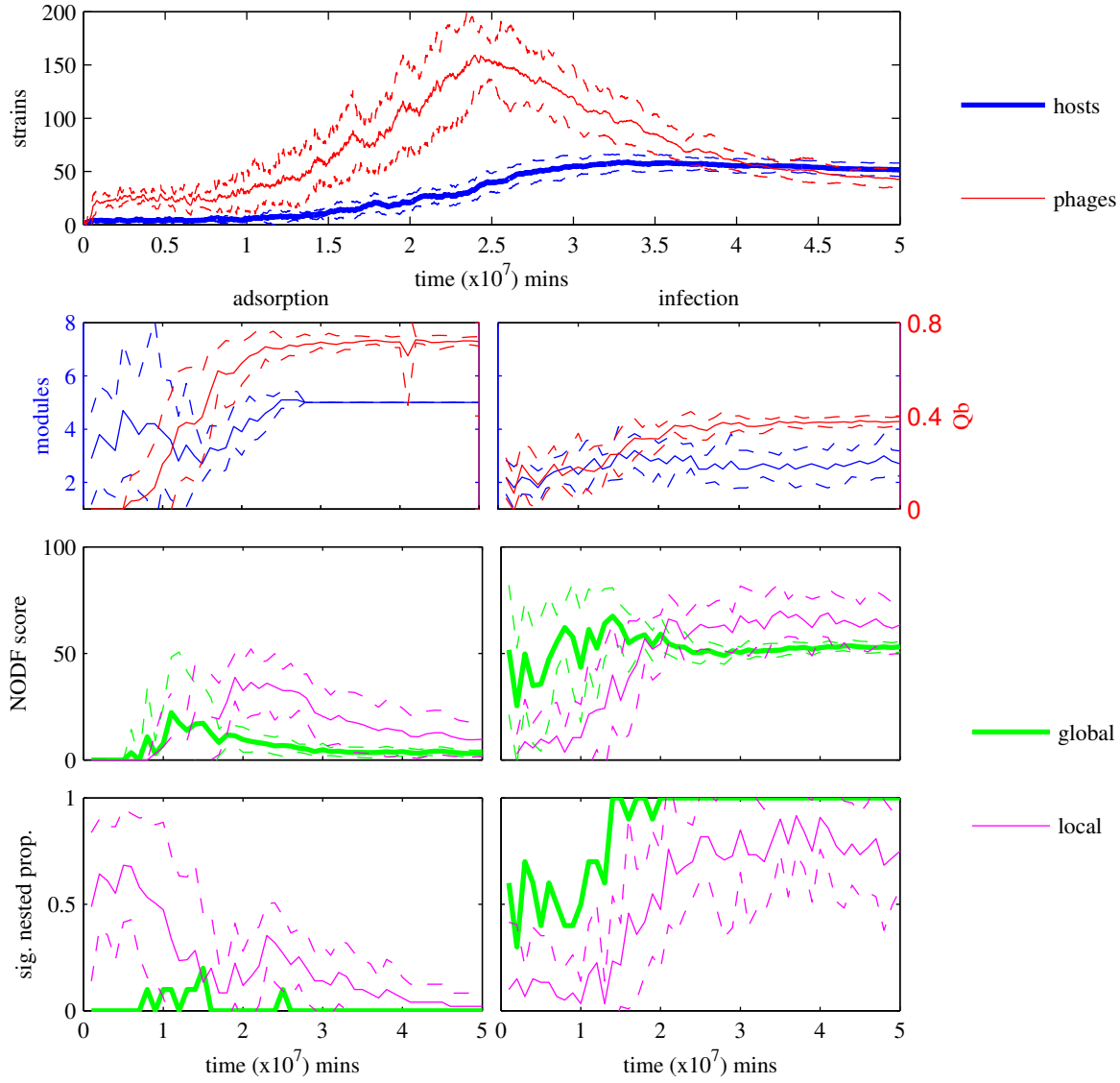


Figure 3.8: Timeseries of matrix metrics for nestedness and modularity of binary interaction networks during coevolutionary simulations. Data shown are mean values (solid lines) $\pm$ 1 std. dev. (dashed lines) from an ensemble of ten simulation runs with identical parameters but different stochastic mutations. Top: Number of bacterial and phage strains. The following panels show metrics for adsorption rate matrices (left column) and infection rate matrices (right column). Upper-middle: Number of modules detected (left axis) and Barber's modularity score Qb (right axis). Lower-middle: NODF nestedness score of the whole network (global) and mean within-module NODF nestedness score (local) in each simulation. Lower: Proportion of NODF nestedness scores in each simulation that were statistically significant at a level of $p < 0.05$ for the whole network (global) and within-module (local).

Figure 3.8 (upper-middle row) shows that the interaction structure shown by adsorp-

tion rate matrices converges on five clearly identifiable modules that are well-detected by the LP-BRIM algorithm. Module identification is less reliable early in the simulation, when the algorithm sometimes produces false positives (e.g. by identifying "modules within modules"), giving large variances in module number early in the simulations. Fewer modules are detected in the infection rate matrices and there is greater variance in the number detected throughout the simulations. Barber's modularity metric is significantly higher for the adsorption rate matrices than for the infection rate matrices throughout the simulations, confirming the visual suggestion (e.g. Figure 3.3) of a stronger modular structure for this form of interaction. We also performed significance tests against Barber's modularity for the adsorption rate and infection rate matrices for all simulations at $t = 2.5 \times 10^7 min$ and $t = 5 \times 10^7 min$, finding that modularity for all matrices tested was significant at a level of $p < 0.01$.

The lower half of Figure 3.8 shows timeseries for nestedness (measured by NODF) and statistical significance of nestedness (here we assume statistical significance at a level of $p < 0.05$). Global nestedness of the whole adsorption rate matrix is typically low and rarely statistically significant. In contrast, global nestedness of the whole infection rate matrix is relatively higher and (after the initial diversification phase of the coevolutionary dynamics) almost always statistically significant.

For each simulation, we also calculated nestedness for the modules identified by the LP-BRIM algorithm. Figure 3.8 (lower four panels) shows the mean within-module NODF score across all modules detected in a particular simulation, as well as the proportion of modules which were statistically nested. Mean within-module nestedness is typically higher than global nestedness for both forms of interaction matrix. Mean within-module nestedness of adsorption rate matrices is typically low and statistically insignificant for most of the timeseries, but rises and shows a higher frequency of statistical significance approaching the mid-point of the simulation runs, when phage diversity is highest. Visual inspection of the adsorption rate matrices showed that a large proportion of identified modules were completely filled (all 1s). These cases give NODF$= 0$ at a significance level of $p = 1$. For the remaining minority of non-filled modules, mean within-module nestedness was typically strong and statisticially significant. Mean within-module nestedness of infection rate matrices is typically high and in most cases statistically significant.

Overall there is a mixed signal from our statistical comparison of binary interaction networks formed from adsorption rates and from infection rates. Both forms of interaction network show a multiscale nested-modular interaction structure to some extent. For adsorption rate matrices, global modularity is very strong while within-

module nestedness is often weak and statistically insignificant; however, there is a minority of modules which are strongly and significantly nested during the middle section of the simulations when phage diversity is highest. For infection rate matrices, global modularity is weaker, but identified modules are typically significantly nested. Interestingly, the infection rate matrices are also typically significantly globally nested.

3.4 Discussion

Here we have shown that a parsimonious 'relaxed lock-and-key' coevolution model based on genetic matching is sufficient to reproduce several core structural features of observed natural communities of bacteria and phage. Negative frequency-dependent selection from phage drives host diversification, which is then mirrored by phage diversification to track their hosts. At steady state, a diverse community of hosts and phage is maintained by kill-the-winner ecological dynamics. Two forms of phage-bacteria interaction network representing the coevolved communities show similar multiscale structural patterns to those observed for natural communities; modularity at large phylogenetic scales and nestedness at smaller scales.

We have aimed with our theoretical study to show that a very simple coevolutionary model can produce nested-modular structures and have selected the chemostat formalism as one of the simplest models in which coevolutionary dynamics can be studied. We do not rule out the possibility that other processes, including spatial structure or multiple resources, might affect observed natural patterns. However, such processes lie beyond the scope of our current study, which aims to show what might occur due to coevolution alone, in the absence of any other (possibly confounding) additional processes. We believe that the most useful models are often the simplest and argue that the use of the chemostat formalism does not affect the generality of our results.

An important caveat is that there is currently only a single dataset [82, 80] showing the macroscale nested-modular interaction structure that the relaxed lock-and-key model produces. Thus it is possible that we are over-estimating the importance of this dataset and hence misjudging the capability of the relaxed lock-and-key model to explain natural community structures. However, there are multiple observations of stable high diversity in natural phage-bacteria communities (e.g. [61, 63, 76, 78]),

as well as many examples of nested interaction networks at smaller phylogenetic scales [3]. At larger phylogenetic scales, there are good arguments to support the specificity of infection needed to produce a modular network structure; for example, phage target particular receptors which may only be present in a small number of bacterial lineages, limiting their potential host range [76, 86]. Thus we cautiously suggest that the nested-modular structure should be robust at large phylogenetic scales and propose that additional large-scale cross-infection studies would be a fruitful area for further research.

In the adsorption rate matrices, the nested-modular pattern is most clearly observed when the system is still in the transient diversification phase, i.e. before the system reaches an evolutionary steady state and while the species in the system are still adapting. This section of the coevolutionary dynamics is also where the phage diversity is highest. Since the adsorption rate network includes interactions irrespective of the density of the bacteria/phage strains involved, it likely includes many interactions between low-density strains which are in process of being out-competed and excluded from the system by fitter mutants. This may have implications for observations of natural communities, where overlapping ecological and evolutionary timescales for bacteria and phage [68, 69] imply that many natural communities may not be at evolutionary steady state. We hypothesise that nested-modular structures in density-independent interaction networks will be most obvious when overall adaptation rates within the community are high, for example, in communities adapting to a changed or dynamic environment.

A note of caution must be raised about the methodological grounding for comparisons of model output to empirical observations. The two forms of interaction network that we studied (based either on adsorption rates or on infection rates) showed different statistical features. In this study, adsorption rate networks showed higher global modularity and lower within-module nestedness, while infection rate networks showed lower global modularity and higher within-module nestedness. While both forms of interaction network studied here showed broadly similar multiscale structure, in general the fit to empirical data will depend both on which form of model network was chosen and also on the method by which the empirical network was produced. With theoretical models, all information is accessible - thus our model networks accurately reflect the full diversity of the model community. However, methods for the more difficult task of constructing interaction networks for natural communities inevitably introduce different kinds of bias into the network structure that is output. A common experimental method for determining interaction networks for natural phage-bacteria communities appears to be to collect a sample from the natural environment, isolate

as many strains as possible, then use a plaque assay to test for infection of each potential host by each phage. However, sampling inevitably carries a bias towards collecting only the more numerous strains, of which only a small fraction will be cultivable [91, 92]. Thus while empirical studies reporting natural interaction networks may aim to present a complete record of all strains present in the community (i.e. to produce something similar to our model adsorption rate networks), they may actually (without any failure of the experimental method) be presenting networks more akin to our infection rate networks, which only include interactions between abundant strains. Thus it is not clear which of our binary interaction networks should be compared to the networks reported for natural communities.

Another methodological issue surrounds the comparison of weighted and binary interaction networks; here we have converted quantitative networks to binary networks to aid comparison with empirical data. Infectivity assays commonly only measure presence-absence of infection, rather than the messier rate/affinity data that indicate the quantitative strength of interactions, which are harder to measure [93, 94, 81]. Thus empirical data is often given in binary form and analysed using statistical tools developed for binary matrices. Conversion of quantitative data into binary form loses information and inevitably introduces bias that will accentuate some features and mask others. A challenge for empirical researchers is to develop methods for measuring interaction strengths between phage and bacterial strains, rather than just presence-absence of interaction [94]. A challenge for theoretical researchers is to develop better statistical tools for analysing the weighted phage-bacteria interaction networks that will thereby be produced.

The original kill-the-winner model of aquatic virus ecology [74] describes one-to-one interactions between viruses and bacteria, such that no cross-infections occur. In that scenario, specialised infection leads to negative density-dependent predation from viruses, which favours rare bacteria phenotypes and acts to maintain diversity. This contradicts available data for real phage-bacteria systems [3, 80]. The relaxed lock-and-key model allows for cross-infection based on genetic similarity, while producing a stable diverse community maintained by kill-the-winner dynamics. Observations of natural kill-the-winner dynamics [77] are thus consistent with the relaxed lock-and-key model. We note that several alternative coevolution models are unlikely to capture nested-modular interaction structures and kill-the-winner ecological dynamics. Gene-for-gene genetic models are consistent with arms race dynamics [95] and transitive range expansion, but do not permit stable high diversity without the addition of explicit trade-offs [72] (and even then diversity is typically limited to dimorphism [66]). Matching-alleles genetic models can drive diversification [72, 95],

but do not produce the modular structure or within-module nestedness.

There remain other mechanisms that will affect the interactions between bacteria and phage and several of these could potentially produce high diversity and nested-modular network structures. The relaxed lock-and-key model is perhaps most easily interpreted as representing coevolution of phage tail-fibres and bacterial cell-surface receptors. However, the real infection process is multi-stepped and coevolution may occur at different stages, including initial adsorption to extracellular surface receptors [96] and also subsequent intracellular defence mechanisms [97, 86, 98, 99, 100]. Additionally our model does not take into account non-mutational processes of genetic variation, such as gene loss and horizontal gene transfer [94, 76], or the possible role of environmental heterogeneity and/or spatial localisation [80]. Any of these factors can affect natural community structures and could in principle have produced the patterns created here by coevolution. The strength of the relaxed lock-and-key model is by appeal to Occam's razor; it offers a parsimonious and sufficient explanation of multiple observed phenomena.

Chapter 4

FALCON: a software package for analysis of nestedness in bipartite networks

Abstract

Nestedness is a statistical measure used to interpret bipartite interaction data in several ecological and evolutionary contexts, e.g. biogeography (species-site relationships) and species interactions (plant-pollinator and host-parasite networks). Multiple methods have been used to evaluate nestedness, which differ in how the metrics for nestedness are determined. Furthermore, several different null models have been used to calculate statistical significance of nestedness scores. The profusion of measures and null models, many of which give conflicting results, is problematic for comparison of nestedness across different studies. We developed the FALCON software package to allow easy and efficient comparison of nestedness scores and statistical significances for a given input network, using a selection of the more popular measures and null models from the current literature. FALCON currently includes six measures and five null models for nestedness in binary networks, and two measures and four null models for nestedness in weighted networks. The FALCON software is designed to be efficient and easy to use. FALCON code is offered in three languages (R, MATLAB, Octave) and is designed to be modular and extensible, enabling users to easily expand its functionality by adding further measures and null models. FALCON provides a robust methodology for comparing the strength and significance of nestedness in a given bipartite network using multiple measures and null models. It includes an "adaptive ensemble" method to reduce under-sampling of the null distribution when calculating statistical significance. It can work with binary or weighted input networks. FALCON is a response to the proliferation of different nestedness measures and associated null models in the literature. It allows easy and efficient calculation of nestedness scores and statistical significances using different methods, enabling comparison of results from different studies and thereby supporting theoretical study of the causes and implications of nestedness in different biological contexts.

This chapter is based on the publication:

Beckett S.J., Boulton C.A., Williams H.T.P. 2014. FALCON: a software package for analysis of nestedness in bipartite networks [v1; ref status: indexed, http://f1000r.es/3z8] F1000Research **3**: 185. (DOI: 10.12688/f1000research.4831.1).

We thank Virginia Domínguez-García with implementation of JDM nestedness.

4.1 Introduction

Nestedness is a statistical property of systems where two kinds of entity interact, which can be represented as bipartite networks. Originally used as a metric for species-site distributions [22, 21], nestedness has recently gathered much attention as a metric for bipartite species interaction networks, e.g. plant-pollinator mutualisms [1, 101] and host-virus interactions [3, 80, 102]. Various discussions have considered the sources of nestedness in such systems and its potential implications for ecological dynamics [9, 10, 11, 101, 35, 103, 104]. However, it is unclear how to systematically compare results for different ecological datasets. Furthermore, nestedness is not restricted to ecological datasets, but is a generic property of any bipartite network. Thus there is a need for measures of nestedness that are context-independent and do not depend on any particular (ecological) interpretation. Multiple methods for measuring nestedness have been used in different studies, along with multiple approaches to calculating statistical significance of the measured values. This provides a large number of ways in which nestedness could be evaluated [105, 25, 106]. Before theoretical investigation of the mechanisms of nestedness can be properly investigated, robust measures and statistical tests for nestedness are required to allow comparison of results from different studies.

Here we present FALCON – a free software package that allows the user to easily compute several measures of nestedness and associated statistical significances based on a selection of null models. FALCON stands for “Framework for Adaptive ensembLes for the Comparison Of Nestedness”. FALCON operates on any form of bipartite interaction data represented as a matrix of associations and is set up to be deliberately ’blind’ to the source and interpretation of input data. FALCON is based on the assumption that nestedness is a general statistical property of matrices and therefore its measurement should be independent of context or interpretation. FALCON calculates nestedness as a statistical property of a matrix, by returning the nestedness score for the most-nested configuration of the input matrix. Since calculating statistical significance of nestedness scores can be computationally demanding, involving generation of a large ensemble of matrices from a null distribution, FALCON uses a novel “adaptive ensemble” method to improve efficiency by using the minimal ensemble size sufficient to give robust statistics.

Several software packages for calculating nestedness already exist, but these are subject to various factors which make direct comparison of different nestedness measures, and statistical interpretation of returned values difficult to achieve. Several

nestedness measures are handled by packages which deliver a single measure, making comparison difficult. Some are specific to a particular operating system. Some do not make source code available for reimplementation, reducing confidence in their outputs and prevent future extensions. Two packages for the R statistical programming language, bipartite [107] and vegan [26], together contain functions for several nestedness measures and associated null models, as well as many other tools for analysis of bipartite ecological networks. However, these packages offer no obvious implementation of significance testing (the principal method for reporting results of nestedness analyses) and they also lack several nestedness measures which have been recently developed. FALCON is designed to address these deficiencies, enabling calculation of nestedness and statistical significance using a variety of measures and null models, with open source code provided for several platforms.

The FALCON package is available for three commonly used numerical analysis platforms: MATLAB, Octave and R. MATLAB (http://www.mathworks.co.uk/products/matlab/) is a commercial software platform, while Octave (https://www.gnu.org/software/octave/) and R (http://www.r-project.org/) are both freely available open source platforms. FALCON can be freely downloaded on Github (http://github.com/sjbeckett/FALCON) or figshare [108] and all code is open and accessible. A guide to downloading, installing and running FALCON accompanies the code. This document describes the assumptions on which FALCON is based, how it calculates nestedness and statistical significance, and gives details of the adaptive ensemble method used to improve computational efficiency and provides a case study to demonstrate it's usage and outputs.

4.2 What is nestedness?

Nestedness is a statistical property of bipartite interaction data presented in matrix form. In a perfectly nested matrix, the entries in each successive row are a strict subset of those in the previous row, while the entries in each successive column are a strict subset of those in the previous column (Figure 4.1). Interpretation of nestedness depends on context.

The concept of nestedness was first described in studies of how species distributions varied between sites [109, 110, 111], and later defined quantitatively as measuring the 'amount of order/disorder' in matrices representing presence/absence of species in island communities [22]. Used in this way, nestedness is calculated from a matrix of presence-absence data where rows are species and columns are sampling sites

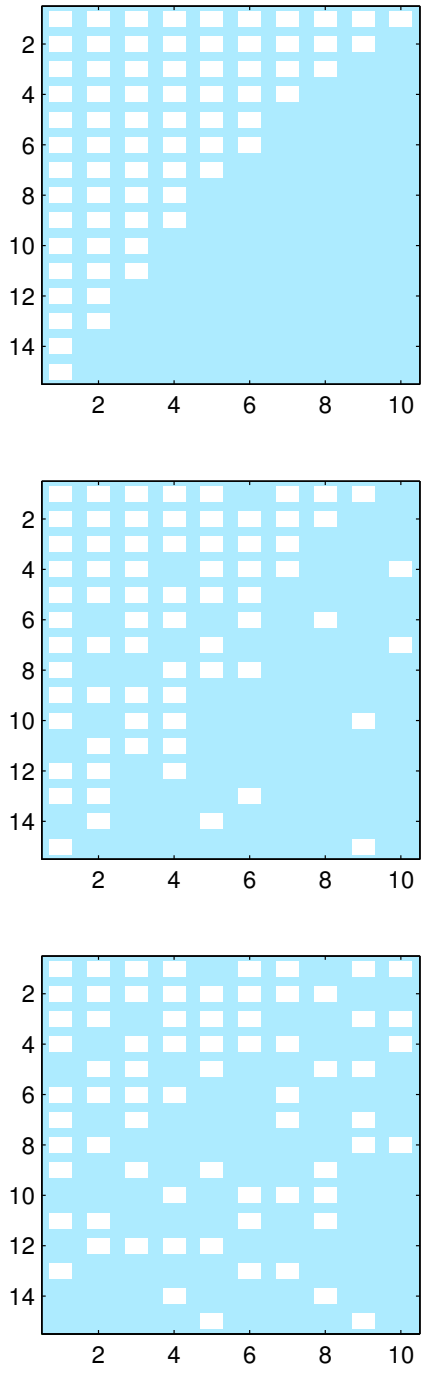


Figure 4.1: Perfectly nested, weakly nested, and randomly connected matrices. White squares indicate connections between two kinds of entity arranged in rows and columns.

along some environmental or spatial gradient. A perfectly "nested" matrix (see Figure 4.1) would be achieved when the set of species present at each site along the gradient is a subset of the species present at the previous site. Since then the concept of nestedness has been extended in various directions; see [15] for an historical overview of the nestedness concept. Nestedness has continued to be applied to spatial patterning (e.g. [112]) and has been linked with β-diversity [113], but has also been applied to study mutualistic or antagonistic species-species interactions [114, 57], species-time relationships for a single site [115], and several other types of bipartite networks [10, 116, 11, 6, 117, 118]. For pairwise interactions (e.g. plant-pollinator or host-parasite systems), nestedness has been interpreted as placing species along a gradient of generalism-specialism in the number of partners they interact with; in this context, perfect nestedness is achieved when species within each class are ordered such that the interaction set (set of partners) for each species is a strict subset of that of the next species, and the most generalised species of one class interact with the most specialised species of the other class.

Nestedness is calculated from a biadjacency matrix representing pairwise interactions between two kinds of entity (one represented by rows, the other by columns). The order of rows and columns for a biadjacency matrix is arbitrary with respect to connectivity; rows and columns can be permuted without affecting the underlying topology of the interaction network. Any non-arbitrary ordering of rows and columns in the matrix representation necessitates the use of additional contextual information to specify which order rows and columns should take. While some datasets may suggest a "natural" ordering to rows and columns in the matrix representation of data (e.g. when one of the dimensions represents an environmental/spatial/temporal gradient), for many applications of nestedness there is no natural ordering (e.g. species interactions).

As stated above, we consider that nestedness should be a context-free metric, so that it can be applied to data without requiring any supplemental information on row/column ordering. This assumption implies that the ordering of rows and columns should not affect the measurement of nestedness. While some nestedness measures are insensitive to row/column ordering, several of the most commonly used measures are highly sensitive to ordering, introducing indeterminacy to the quantification of nestedness when rows/columns are ordered arbitrarily. To avoid this indeterminacy and return a single robust nestedness score for a given input matrix, FALCON can sort the rows and columns such that nestedness (however calculated) is maximised. Since re-ordering rows/columns in a matrix representation does not alter the structural information (node adjacency) of the underlying data, this re-ordering is a reasonable

approach and makes measurement of nestedness more consistent.

4.3 Measures of nestedness in FALCON

Nestedness is most commonly calculated for binary data representing presence/absence of an interaction between two entities, but nestedness can also be calculated for weighted data that also indicate the strength of the interaction. The methods used to calculate nestedness vary depending on whether binary or weighted interaction data is provided. The nestedness measures available in FALCON are shown and briefly described below and in Table 4.1; further details are given in Section 2.2.

The nestedness measures considered here are not trivial variations upon each other, but differ significantly in their derivations. However, some similarities can be drawn. Spectral radius (SR) [35] and the measure of Johnson, Domínguez-García, & Muñoz [32] (JDM) are invariant to the ordering of rows and columns in the network and are calculated using the adjacency matrix of the network. On the other hand, discrepancy (BR) [21], Manhattan distance (MD) [33] and nestedness based on overlap and decreasing fill (NODF) [28] are all sensitive to row/column ordering and maximised when rows/columns are ranked by degree. The nestedness temperature calculator (NTC) [22, 26] involves sorting of rows and columns against the 'isocline of perfect order' (see Figure 2.1) such that it maximises connections above the isocline and minimises connections below the isocline. BR is similarly calculated relative to an idealised 'maximally packed' matrix. NODF is found through pairwise comparisons of overlap between subsequent rows and columns, whilst MD is found by assigning a weight to each connection as a sum of it's row and column indexes. The measures also differ in how nestedness is scored; the degree of nestedness in a network increases with increasing measure score for JDM, NODF and SR, but with decreasing measure score for BR, MD and NTC.

4.4 Comparison of nestedness scores

Nestedness is strongly sensitive to the size (number of rows and columns) and fill (number of non-zero entries) of the input matrix [23]. This is problematic in practical terms, since we often wish to compare nestedness of matrices that differ in these basic properties; in fact, cases where we compare empirically derived matrices with

Name	Shorthand	Binary/Weighted	Brief description	Reference
Nestedness based on overlap and decreasing fill	NODF	Binary	Pairwise row and column comparisons	[28]
Manhattan distance	MD	Binary	Sum of row and column indexes of connections	[33]
Nestedness temperature calculator	NTC	Binary	Difference from an 'isocline of perfect order'	[22] [26]
Jonhson, Domínguez-García & Muñoz	JDM	Binary	Measure of dissassortivity using configuration model	[32]
Discrepancy	BR	Binary	Difference from a 'maximally packed' matrix	[21]
Weighted NODF	WNODF	Weighted	Weighted version of NODF	[38]
Spectral radius	SR	Both	Maximum real eigenvalue of adjacency matrix	[35]

Table 4.1: Nestedness measures available in FALCON.

identical size and fill are an exception. Thus comparison of absolute values of nestedness metrics are not informative and may be misleading. To compare nestedness of matrices with differing size and fill, observed nestedness should always be interpreted in the context of a null distribution of matrices with similar properties. Measuring observed nestedness relative to expected nestedness derived from a null distribution of similar matrices allows determination of both effect size (e.g. as a z-score, which is commonly used to compare different nestedness schemes [15, 36])) and statistical significance (e.g. as a p-value giving the expected frequency of the observed score in the null distribution). This approach necessitates choice of a suitable null model and generation of a distribution of random matrices drawn from it.

In the present context, a null model is a method for creating a distribution of matrices that conserve some properties of the input matrix while varying other properties at random [119]. We continue the "context-free" approach in our treatment of null models; to allow comparison of nestedness across different scenarios, a good null model should not make assumptions about the mechanisms by which data were generated, but treat the matrix as an independent data structure. However, to be comparable to the input matrix, null matrices must conserve some key matrix properties (such as size and fill) on which nestedness depends. The null models available in FALCON are given in Table 4.2; further detail is given in Section 2.3. FALCON includes some of the more popular null models from the literature, alongside some additional null models that we feel can be useful. Null models vary in whether the original data is binary or quantitative, and in which properties of the original input matrix are preserved.

4.5 How FALCON works

4.5.1 Inputs and outputs

FALCON requires several inputs:

- an input network in the form of a bipartite matrix
- whether binary or quantitative nestedness should be investigated (quantitative matrices can be analysed using binary measures)
- whether to sort rows and columns to maximise nestedness score

Name	Description	Binary/weighted	Conserved features	Reference
SS	Shuffles positions randomly	Binary	Shape, Fill	[35]
FF	Permutations of structure with same node degrees	Binary	Shape, Fill, Degree	[58]
CC	Some structure is preserved, rest is shuffled	Binary	Shape, Fill	
DD	Determined probabilistically by node degree	Binary	Probabilistic Degree	[1]
EE	Determined probabilistically by fill	Binary	Probabilistic Fill	
Binary Shuffle	Order of weighted links is swapped	Weighted	Binary positions, weights	[35]
CRT	Random weights where row totals conserved	Weighted	Binary positions, row totals	
CCT	Random weights where column totals conserved	Weighted	Binary positions, column totals	
RCTA	Average of conserve row totals and column totals	Weghted	Binary positions	

Table 4.2: Null models in FALCON.

- which nestedness measures should be used
- which null models nestedness should be tested under
- whether the ensemble of null models should be created with a fixed number or adaptively chosen
- whether or not to plot the distributions of nestedness scores

Output is returned to the user in the form of:

- the most nested configuration of the input matrix
- the nestedness measure(s) of the input matrix
- the expected value of nestedness under the null model(s) (as the mean measure of matrices created in the ensemble)
- the number of ensemble members used to calculate significance in each null model
- the statistical significance of the nestedness of the input matrix against each null model as a p-value
- the standard deviation and sample *z*-scores of the measure in the ensemble as well as other properties.

4.5.2 What FALCON does

FALCON follows the process shown in Figure 4.2. First it sorts the user input matrix into a maximally nested configuration and removes any empty rows/columns before finding the nestedness of this matrix using the users chosen measures. Then FALCON goes through each of the user specified null models one by one, creating an ensemble of null matrices according to the rules of each null model. Each null matrix is then sorted and measured by each of the chosen nestedness measures. Thus, for each null model, nestedness measures are calculated for each of the null matrices in a single null ensemble, enabling direct comparison of results. The size of the null

ensemble is determined by the input choice of using either the fixed or adaptive ensemble size (see Section 4.5.5). Statistics are computed from the measures found in the null ensemble (and the direction in which that nestedness measure is calculated), before the next null model ensemble is instantiated. Once all null models have been computed the results are returned to the user.

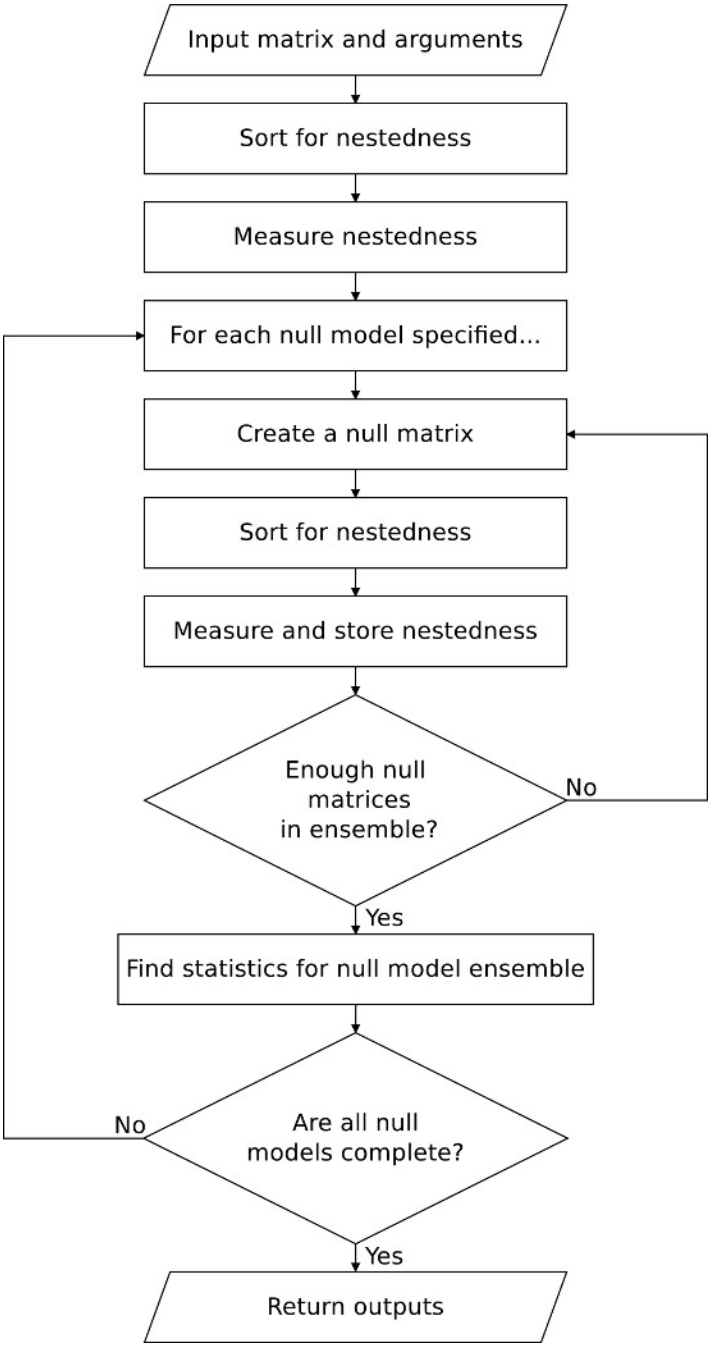


Figure 4.2: FALCON algorithmic procedure

4.5.3 Direction of increasing nestedness

For different nestedness measures, increasing scores can represent either increasing or decreasing nestedness as discussed in Section 4.3. FALCON initially determines whether a higher measure score is related to greater nestedness (or vice versa) in the chosen measure by comparing the scores returned for a highly nested network (see Figure 4.3*a*) and a highly non-nested network (a weighted checkerboard configuration; Figure 4.3*b*), for which the fill (number of non-zero elements) and element sums are equal. The direction of increasing nestedness for a given measure is used during calculation of statistical significance. This method of determining direction each time the algorithm runs is included to allow easy extensibility; if a new measure is added, FALCON will automatically determine which direction indicates increasing nestedness.

(*a*)

$$\begin{pmatrix} 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 0 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 \\ 6 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(*b*)

$$\begin{pmatrix} 4 & 4 & 4 & 0 & 4 & 0 & 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 4 & 4 & 0 & 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 & 4 & 4 & 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 & 4 & 0 & 4 & 4 & 4 & 0 \\ 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 & 4 \\ 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 & 0 & 4 \end{pmatrix}$$

Figure 4.3: (*a*) Weighted nested matrix. (*b*) Weighted non-nested matrix (checkerboard). Both (*a*) and (*b*) have 55 non-zero elements that sum to 220.

4.5.4 Initial sort

For efficiency, FALCON is set up to initially sort the input matrix by row and column degrees for calculation of BR, MD and NODF, retains this sorted configuration for calculation of JDM and SR, and subsequently re-sorts for NTC in order to find the maximal nestedness of a binary matrix. For quantitative data, FALCON uses the same methods as for binary interactions, but also utilises weight data to break symmetry when two rows (columns) have the same degree; in this case, the row (column) which has greater values for most overlapping elements is ranked highest. Where two or more rows (columns) share the same degree and most overlapping elements, the rows (columns) are ranked according to the total sum of row (column) elements. This sorting does not affect the underlying topology or the relationships in

the data. FALCON also allows the user to decide if any sorting is performed, enabling the "context free" assumption to be relaxed (e.g. for investigation of gradient-based nestedness [36]).

4.5.5 Size of null ensemble

FALCON uses a bootstrap method to calculate the statistical significance of a given nestedness score, since the true null distributions of the test statistics are not known. The ensemble size used for this calculation can either be fixed or calculated adaptively by FALCON to improve computational efficiency and reduce undersampling effects. Note that the strongest significance that can be assigned is $p < \frac{1}{N}$ where N is the ensemble size.

Fixed

The number of null matrices used to make up the ensemble is fixed by the user. This method is effective providing that the ensemble is large enough to have statistical power; the larger the ensemble, the more power the test has and the closer the answer will be to the *p*-value for the (unknown) true null distribution. However, it is not obvious how large the ensemble needs to be; in the literature, amongst others, [57] use 1,000 null models in their ensembles, whilst [103] use 10,000, and [80] use 100,000. A large number of different null matrix configurations are possible for a given input matrix and we may wish to avoid undersampling [120]; however, at the same time very large ensembles can make the calculation of significance computationally intractable.

Adaptive

FALCON includes a mechanism for adaptive determination of ensemble size. This is intended to ensure robust statistics are achieved, avoiding concerns about undersampling or oversampling [120], while minimising computational load. The adaptive method works by creating two ensembles in parallel using the same null model. Starting with a minimum ensemble size of 500 in each group, the ensembles are expanded until they show similar statistical properties. This condition is met when the null hypothesis (both ensembles come from the same distribution) of a Mann-Whitney U-test cannot be rejected at 10% significance. When this occurs, it suggests each group represents a good sample of the underlying distribution, and the two groups are combined to form a single null ensemble used to calculate final statistics. The expansion

of the size of the ensemble has an upper limit of 100,000 members in case the null hypothesis is always rejected. The adaptive ensemble methods balances statistical precision with computational efficiency; we conservatively use 1,000 as a minimum final ensemble size such that a *p*-value as low as 0.001 can be assigned.

4.5.6 Output Statistics

p-value

The *p*-value is the probability that a matrix drawn from the null distribution will be more nested than the input matrix. Low values ($p \to 0$) indicate that the input matrix is highly nested relative to the null distribution; commonly a threshold of $p \leq 0.05$ or $p \leq 0.01$ is used to denote a statistically significant level of nestedness.Here p is calculated by counting the frequency of matrices in the null ensemble that are more nested than the input matrix; for cases where no member of the null ensemble is more nested than the input matrix we conservatively assign $p < \frac{1}{N}$ where N is the ensemble size.

Normalised Temperature

The normalised temperature is inspired by the τ-Temperature [33]. It describes the relationship between the nestedness measure found for the input matrix and the expected nestedness measure derived from the null model ensemble. It is described as:

$$T = \frac{Measure}{< Measure >} \tag{4.1}$$

where $< Measure >$ denotes the expected value. In simple terms, the normalised temperature indicates whether the input matrix is more or less nested than the expectation for a null distribution of similar matrices. Where the measure gives increasing scores with increasing nestedness, $T > 1$ indicates greater-than-expected nestedness. Where the measure gives decreasing scores with increasing nestedness, $T < 1$ indicates greater-than-expected nestedness.

Mean

The mean average of the set of nestedness measure found for each of the ensemble members is returned.

Standard Deviation

The standard deviation (σ) of the set of nestedness measures found for each of the ensemble members is returned.

Sample z-score

The *z*-score, or standard score, is calculated as the difference between the nestedness measure and its expected value divided by the standard deviation of the sample:

$$z = \frac{Measure - < Measure >}{\sigma} \tag{4.2}$$

It is a measure of the number of standard deviations the nestedness measure of the input matrix is above the expected value. Hence, the way it should be interpreted, as with the normalised temperature, depends on whether nestedness increases with increasing measure score.

4.6 FALCON usage - case study

To demonstrate FALCON we analyse nestedness analysis in a bipartite network representing the hashtags used by a sample of Twitter users. Data were collected using the Twitter API [121] by searching for tweets including the hashtag "#IPCC" in the time period 21st September 2013 - 5th October 2013. A list of all hashtags used by all users found in the search dataset was then used to create a binary bipartite adjacency matrix for users and hashtags. This was then sampled to create a smaller matrix used for this case study by including each row/column with probability of 0.1 and removing any empty rows/columns. The resulting matrix was stored in a comma-separated file called 'IPCC_HTuse_10_10_1_53x27.csv'.

```
1 >> UserHashtag = importdata('IPCC_HTuse_10_10_1_53x27.csv')
2 >> data = UserHashtag.data;
3 >> output = PERFORM_NESTED_TEST(data,1,1,{'NODF','SPECTRAL_RADIUS'},[2 3],[],1)
4 >> MATRIXPLOT(output.NestedConfig.DegreeMatrix)
```

The box above shows the command sequence used to perform a binary nestedness analysis using FALCON in MATLAB. The first line reads in the ".csv" datafile, which includes row and column headers. The second line extracts the adjacency matrix from the imported data. The third line runs FALCON, using two binary nestedness

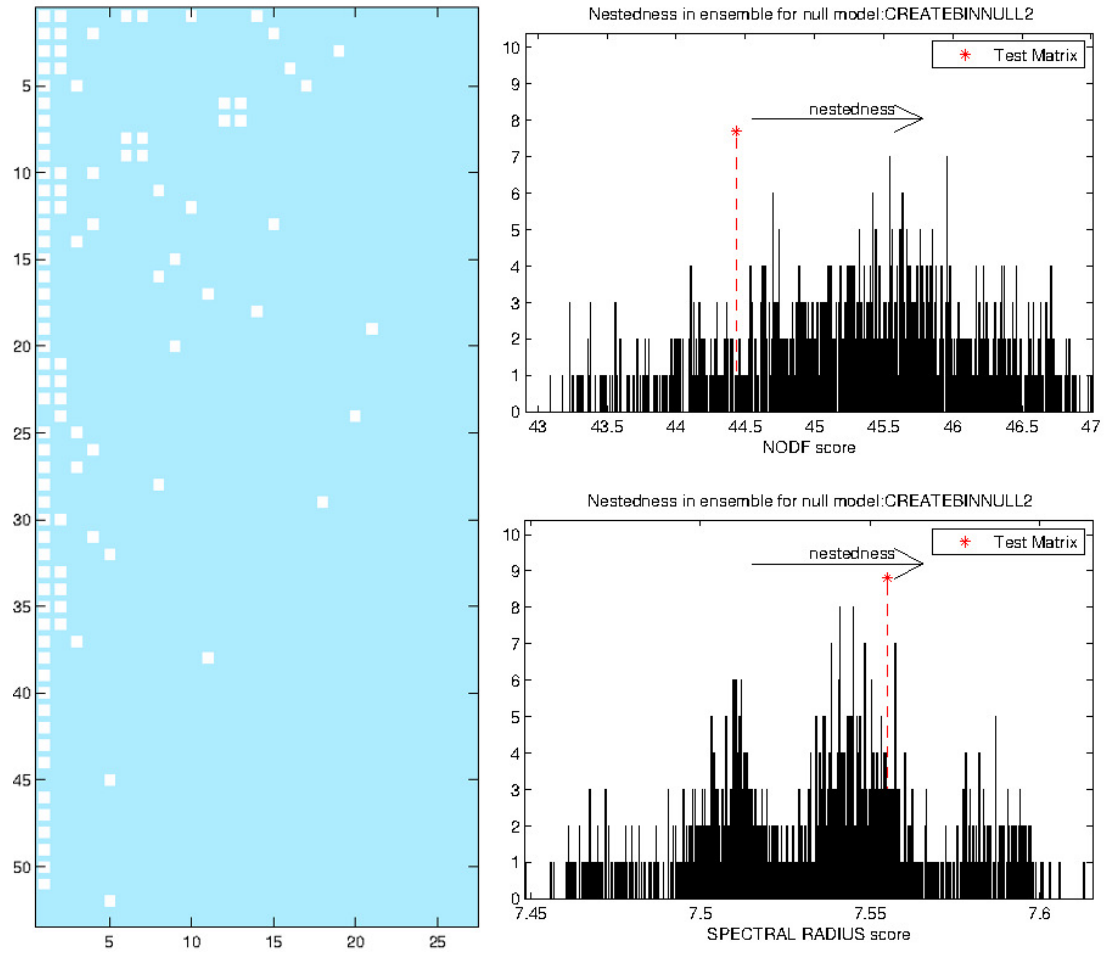


Figure 4.4: Example output from FALCON. Left: the nested UserHashtag data. Top right: Distribution of NODF scores found for an ensemble of FF null models generated for the UserHashtag network. Bottom right: distribution of spectral radius scores found from the same null matrix ensemble. The asterix marks the nestedness score of the input network.

Null model	FF		CC	
Nestedness measure	NODF	SR	NODF	SR
Measure	44.4339	7.5552	44.4339	7.5552
Ensemble Size	1000	1000	1000	1000
Mean	45.3207	7.5346	29.9048	6.4755
Standard Deviation	0.8357	0.0336	3.5600	0.2136
z-score	-1.0612	0.6123	4.0812	5.0550
p-value	0.8490	0.2480	<0.001	<0.001
Normalised Temperature	0.9804	1.0027	1.4858	1.1667

Table 4.3: Example output from FALCON showing sample statistics for the User-Hashtag network using the operations in the box above. Nestedness statistics were computed for the FF and CC null models using the NODF and spectral radius measures.

measures (NODF and SR) and three null models (CC and FF, i.e. nulls 2 and 3) using the adaptive solver and displaying histogram plots. The fourth line plots the input matrix in its most nested configuration, as determined by FALCON. The nested configuration of the matrix and output histograms from significance testing are shown in Figure 4.4, whilst Table 4.3 shows example output from the significance testing. Further examples for use of FALCON in R are given in supporting information accompanying the software.

4.7 Summary

In this paper we have presented FALCON, a software tool for reliable and efficient calculation of nestedness (and associated effect size and statistical significance) based on a selection of popular nestedness measures and null models used in the literature. FALCON treats nestedness purely as a statistical property of a bipartite matrix and removes any form of interpretation or contextual information from the analysis. This enables FALCON to be used to compare nestedness across a wide variety of application areas, noting that the concept of nestedness has already spread from its origin in island biogeography to include species-species interactions and other scenarios, and is likely to find further application in other domains. The contribution of FALCON is to enable easy cross-comparison of observed nestedness using different nestedness measures and null models. We hope that this functionality will allow greater methodological uniformity and comparability of studies of nestedness. We are in the process of performing a large comparison study of nestedness metrics using FALCON (Beckett and Williams., in prep.). Uniformity of measurement and comparability of empirical results is an important preliminary step that must be achieved to enable understanding of the mechanistic basis and ecological (and otherwise) implications of nestedness. We hope that FALCON will be of use to other researchers and help illuminate this intriguing property of bipartite networks in many natural systems.

Chapter 5

Brooding on nestedness: nestedness analyses are confounded by sensitivity to measurement choices and network properties

Abstract

Nestedness describes a network topology in which the interactions of the more specialised nodes are subsets of the interactions made by the more generalised nodes. Many measures and null models have been proposed to determine the nestedness of bipartite networks. This raises questions about how these methods differ; if they reach the same conclusions and whether some methods are preferable to others. Using an ensemble of synthetic networks, with known levels of nestedness, we examine how the choice of nestedness measures, choice of null models and choice of effect size metric may effect the outcomes of nestedness analysis in bipartite networks. We find that the detection of nestedness is sensitive to measure–null–effect size choices. Different measure-null model combinations are susceptible to different levels of type I and type II errors; we identify the most robust combinations as those providing the best tradeoff between the two error types. Furthermore, nestedness calculations are systematically biased by network connectance, which makes comparing nestedness between networks difficult. However, we show that some measure–null–effect size metric combinations can give robust comparisons for networks for a limited range of connectances. Our analysis highlights some of the challenges in nestedness analysis and has implications for the reporting of nestedness analyses in the literature.

This chapter is being prepared for submission:

Beckett S.J., Williams H.T.P. Brooding on nestedness: nestedness analyses are confounded by sensitivity to measurement choices and network properties.

This work has made use of the resources provided by the University of Exeter Science Strategy and resulting Systems Biology initiative. Primarily these include HPC facilities managed by Konrad Paszkiewicz of the College of Environmental and Life Sciences and Pete Leggett of the University of Exeter Academics services unit.

5.1 Introduction

Nestedness is a structural feature of bipartite networks that has been identified in a variety of biological [1, 3, 7], economical [10, 11] and social (Williams et al., in prep. [122], [123]) systems. Disassortative networks are those where the more specialised nodes are more likely to interact with the more generalist nodes. In bipartite networks nestedness measures the variation in generality and disassortativity between two types of entity, such that the more specialist nodes interactions are a subset of the interactions of the more generalised nodes. It is noted that nestedness can be applied to unipartite networks (e.g. [32, 31]) and weighted networks (e.g. [35]). But, in this paper we focus on the case of binary bipartite networks - representing systems on the basis of the presence or absence of interactions between nodes. Nestedness was first conceived in this context to describe site-species relationships in island biogeography theory [109, 110, 111]. Since, it has become a more general concept applicable to all types of bipartite networks. Bipartite networks can be represented as matrices where each matrix element represents the type of interaction between one node type (on the rows) and the other node type (on the columns). This is the biadjacency matrix. In this framework nested matrices are those where proceeding rows are proper subsets of the preceding rows; and proceeding columns are proper subsets of the preceding columns. In occurrence monitoring networks, that describe which species occur at which sites, nestedness is related to migration and disturbance processes and may be useful in designing conservation strategy [7]. Nestedness is also important in networks of species interactions e.g. describing mutualistic or antagonistic relationships, where it is supposed that a nested architecture is related to system stability [10, 11, 124, 104]. However, the link between nestedness and stability is equivocal e.g. it does not appear to hold for networks of mixed interactions [125]. Research into the causes and implications of nestedness is ongoing.

Such research depends on robust methods to identify and measure nestedness, yet the nestedness concept is difficult to define operationally. A plethora of different measures of nestedness have been proposed, based on different underpinning assumptions.

Whichever measure is used, the strength of nestedness within a network is usually reported by the significance of the result (i.e. as a p-value) and by its effect size (e.g. as a z-score , though other choices of effect size are sometimes used [15]). Calculating these statistics necessitates the use of a null model to generate a sample

of random networks. Several different null models are available to perform nestedness analysis, which together with the variety of measures and effect size statistics creates a large possible space of possible methodologies for analysing nestedness. This raises questions for researchers: How should we measure nestedness? Do all measures of nestedness correspond with one another? Do methodological choices affect the conclusions drawn?

Moving beyond analysis of a single network, many users of nestedness also wish to compare nestedness between multiple networks. Examples include meta-studies [1, 114, 3, 7] to question whether certain natural systems are generally nested or not; and comparing networks through time [6, 115, 102] or space [6, 7]. However, in order to do this it is necessary to calculate statistics that are unbiased by network properties [126] so that it is the underlying nested structure that is being compared, rather than any artificial detail connected to the method. In bipartite networks this may be difficult to achieve as there are far fewer topological architectures a network can take when connectance is very high or very low [120] which may bias measures of nestedness and other network characteristics [126]. By evaluating how nestedness measures and effect sizes calculated under them behave with relation to network properties we address the question of whether there is systematic bias of nestedness measures with respect to network connectance, which would have implications for comparing nestedness across multiple networks.

To answer these questions we want to compare the results for different nestedness measures and null models on the same set of networks. Performing benchmarking tests for nestedness is not a new exercise, for example [19, 25, 127, 57, 128]. However, most of these types of study are based on a relatively small number of real world networks - in comparison to the total number of possible network configurations. Strona et al. [128] use a synthetic ensemble comprising 10,000 networks of various sizes (with up to 50 nodes of both types) along a nestedness gradient to compare results for two nestedness measures. However, their nestedness gradient scheme required all considered networks to be square (an equal number of nodes of each type). Here, employing a different nestedness gradient scheme we use an ensemble of synthetic networks of sizes relevant to ecology to investigate nestedness - including non-square networks. By creating networks which are ‘perfectly nested’ and then adding varying levels of noise, we are able to investigate the ability to discriminate and measure nestedness across a set of networks along a gradient from nested to random.

In this paper we want to find out if it matters which of the nestedness measures

are chosen - and if so, whether a dependable nestedness measure can be found for a chosen purpose. Here, we performed nestedness analysis on 30,000 synthetic networks using six nestedness measures in five different null models. In total 773,230 statistical tests for nestedness were performed using the synthetic network ensemble. We show that the type of null model employed in nestedness analysis may effect how likely different measures of nestedness are to reach the same conclusions and that measures differ in their abilities to distinguish nestedness. We offer general suggestions to researchers interested in nestedness analysis - regardless of their study system(s) on the basis of our results.

We separate the results into two parts. In the first part we focus on differences that exist between different nestedness measures and null models applied on individual networks. Ideally a dependable nestedness measure would be one that can accurately distinguish the difference between networks that are and those that are not nested. This can be assessed by evaluating their ability to minimise the number of random networks reported as being significantly nested (type I error) and to minimise the number of networks with a nested architecture as being reported as not significantly nested (type II error). Study I (Section 5.3) deals with correspondences between the chosen nestedness measures when used to analyse individual networks, whether any of these are redundant and whether it makes a difference which method is used. Then, Study II (Section 5.4) goes beyond looking at nestedness within an individual network and seeks to find the best way to compare nestedness among different networks. We investigate how network properties effect the results of traditional nestedness analysis and offer a potential solution when comparisons of nestedness across multiple networks is necessary.

5.2 Methods

Here we give an overview of the methods used in both Study I and II. In Section 5.2.1 we describe how nestedness was measured; including the nestedness measures and null models tested as well as how p-values and effect sizes were calculated. Section 5.2.2 deals with the creation of the synthetic network ensemble that nestedness was calculated for. This includes how perfectly nested networks were created; how noise was added to these through rewiring and how we chose the networks that were evaluated. Section 5.2.3 details the availability of data and methods associated with this study. The methods used to evaluate reported nestedness in the synthetic ensemble are given in Section 5.3.1, which deals with variation of nestedness calculated on

individual networks (the focus of Study I), and later Section 5.4.1 details the methods used to evaluate comparisons of nestedness across multiple networks (the focus of Study II).

5.2.1 Measuring nestedness

FALCON [129] is software that has a statistical framework for comparing nestedness measures against the same null matrices in a null model ensemble using an adaptive ensemble. The number of null networks used in each null model ensemble is calculated by iteratively testing whether two sampling groups appear to have come from the same underlying discrete distribution using a Mann-Whitney U test. This bootstrapping methodology results in null ensembles with between 1,000 and 100,000 null matrices.

Nestedness measures Six measures which were used to calculate nestedness in binary bipartite networks:

- a nestedness temperature calculator [26] (**NTC**)

  One of the first nestedness measures was the nestedness temperature calculator [22] that shuffles rows and columns in a biadjacency matrix against a curve known as the 'isocline of perfect order', to minimise the number of unexpected presences on one side of the curve and absences on the other side - these are penalised more heavily the further away they fall from the isocline. We use a later formulation of the nestedness temperature calculator [26] as the original methods were not well explained and have been critically reexamined since [23, 26].

- discrepancy [21, 129] (**BR**)

  Discrepancy [21] is the sum of differences between the focal network and a perfectly nested network formed using the row sums of the focal networks biadjacency matrix. We modified this measure [129] (following Ulrich et al. [15]) so that operations are performed on both rows and columns and the overall minimum is used.

- nestedness metric based on overlap and decreasing fill [28] (**NODF**)

  Nestedness can also be evaluated by performing comparisons between pairs of rows (and columns) and calculating whether one row (column) is a subset of the other. These operations form the basis of the nestedness metric based on overlap and decreasing fill [28].

- Manhattan distance [33] (**MD**)

  The Manhattan distance, which is the sum of row and column matrix indices of presences, can also be considered as a measure of nestedness as the score is minimised in a nested structure [33].

- spectral radius [35] (**SR**)

  The dominant eigenvalue of a bipartite network, which mathematically is related to stability, has been shown to be related to nestedness [35] and is denoted as the spectral radius.

- Johnson et al.'s nestedness measure [32] (**JDM**)

  A further measure proposed by Johnson et al. [32] proposes that nestedness can be assessed by quantifying row degree-column degree disassorsitivity; that is the amount by which low degree row nodes are connected to high degree column nodes and vice versa.

Null models Five binary null models were used, which each preserve some feature of the targetted network by:

- randomly swapping the interacting nodes (**SS**)

  Creates networks that have the same number of edges and nodes as the target network, but where the edges are placed randomly.

- fixing the row and column degree distribution [58] (**FF**)

Creates networks that have the same node degree distributions as the target network.

- conserving the specialists interaction structure [129] (**CC**)

  Creates networks that have the same number of edges and nodes as the target network, but where specialist edges are kept fixed, whilst remaining edges are placed randomly.

- probabilistically conserving row and column degree distributions [1] (**DD**)

  Creates randomised networks that have edge setting probabilities proportional to the corresponding nodes degree distributions in the target network.

- probabilistically conserving connectance (**EE**)

  Creates randomised networks of the Erdős–Rényi type that have an edge setting probability set to the connectance of the target network.

Details for all measures and null models are provided in FALCON [129] and in Section 2.2 and Section 2.3.

Significance of nestedness To evaluate whether a detected nestedness pattern is significant or not p-values were calculated. This is done by finding the proportional placement in the ranked distribution of nestedness measures scores within a null model ensemble. To be more precise it is the proportion of the total number of null networks that have greater nestedness than that in the observation network. In cases when all null model networks were less nested than that in the observed network we conservatively assigned a p-value score of $p = 1/N$ where N is the number of networks that made up the null ensemble.

Nestedness effect size We elected to investigate two effect size measures in this study: the sample z-score and the adjusted normalised temperature score. The sample z-score finds the number of standard deviations (sd) that observed nestedness (Obs) is away from the expected nestedness (Exp) and is calculated as:

$$z = \frac{Obs - Exp}{sd} \tag{5.1}$$

using the nestedness values found in a null model ensemble. Adjusted normalised temperature is a reworking of normalised temperature [129] which is similar to both relative nestedness [1] and τ-temperature [33] (see Appendix D: Robustness of nestedness discrimination ability), but includes a correction factor (the adjustment) for the direction that nestedness is measured in by a particular nestedness measure. Adjusted normalised temperature is calculated as:

$$A = \left(\frac{Obs}{Exp}\right)^{Dir} \tag{5.2}$$

where:

$$Dir = \begin{cases} -1 & \text{when nestedness increases with increasing measure score (NODF, SR, JDM)} \\ +1 & \text{when nestedness increases with decreasing measure score (MD, BR, NTC)} \end{cases}$$

5.2.2 Creating a synthetic network ensemble

Perfectly nested networks As stated above nested matrices are those where proceeding rows are proper subsets of preceding rows; and proceeding columns are proper subsets of preceding columns. They are typically thought of as an upper triangular structure - but in fact by the previous definition there exists many such matrices with varying curvatures associated with the positions of presences and absences in a nested matrix. To create perfectly nested matrices we define the curve for the unit square:

$$y = \left(1 - (1 - x)^t\right)^{1/t} \tag{5.3}$$

This curve is drawn on the unit square for a given curvature parameter t. When $t = 1$ the curve is the linear diagonal line traditionally associated with nestedness.

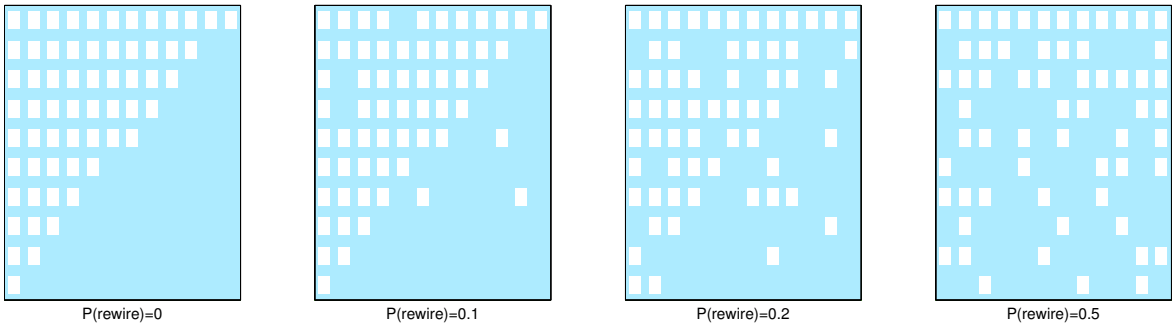


Figure 5.1: Rewiring of a 'perfectly nested' bipartite network with 10 rows, 12 columns and curvature of 1. (a) the 'perfectly nested' network, (b-d) increasing levels of $P(rewire)$ applied to the network in (a).

When $t > 1$ and $t \to \infty$ the defined curve becomes increasingly concave, whilst when $t < 1$ and $t \to 0$ the defined curve becomes increasingly convex. We then transform this curve defined in the unit square such that it starts at the co-ordinates $(1, 1)$ and ends at $(r - 1, c - 1)$ where r is the number of rows and c is the number of columns of the nested matrix being built. We define the elements along the top row and first column as all presences. Then in the rectangle defined by the co-ordinates $(1, 1), (r - 1, 1), (r - 1, c - 1), (1, c - 1)$ we need to determine which elements should be presences (above the curve) and absences (below the curve). We do this by using the midpoint of the cell and evaluating whether this lies below or above the curve. If the midpoint lies on the curve we decide to elect that it also counts as a presence.

Rewiring of networks In order to create samples that show varying degrees of nestedness we define rewiring; the process of switching the state of a given connection in a matrix from on to off (presence to absence) and in response changing a random connection in the opposite state from off to on (absence to presence). We can then define the concept of a probability of rewiring ($P(rewire)$ such that for every element (connection) in a matrix there is a probability $P(rewire) = q$ that rewiring will take place on that element. As this process is repeated iteratively through each of the elements in the matrix there is a chance that elements may change state several times within a single rewiring operation. All rewired matrices retain the same number of elements as the starting matrix, though there exists a possibility that the number of rows or columns may be reduced (as they may contain rows or columns with no presences). The lower q is the more likely the rewired matrix will resemble the starting matrix, whilst matrices rewired with a larger q are less likely to resemble the starting matrix and are more likely to appear random. An example of rewiring is given in Figure 5.1.

Sampling strategy In order to assess how the impact of nestedness measure-null model combination is affected across a range of matrices we elected to use Latin hypercube sampling to determine the number of rows, the number of columns and the curvatures that the starting perfectly nested matrices should have. The number of rows and columns were chosen as integers in the range [5 , 100], which are comparable sizes of presence-absence bipartite networks used in ecology, and curvature parameters were mapped onto the range [$^1/_4$, 4] such that there were representative number of both convex and concave matrices. We used this method to define 500 initial perfectly nested matrices. We also chose to use six values for $P(rewire)$ to define a gradient between nested and random matrices. We created 10 replicate matrices at each level of $P(rewire)$ on each of the 500 initial matrices giving (number of initial matrices) $\times$ (number of $P(rewire)$ levels) $\times$ (number of replicates) $= 500 \times 6 \times 10 = 30,000$ matrices to analyse for nestedness in different measurement-null model combinations using FALCON.

5.2.3 Data availability

FALCON is available from http://github.com/sjbeckett/FALCON . The 30,000 networks evaluated in the synthetic ensemble are available as csv files on figshare [130], where each file represents one of the networks as a biadjacency matrix listing the initial configuration network, the $P(rewire)$ applied and the replicate. This dataset [130] also includes code for creating perfectly nested networks and the noise induced rewired networks. In addition the resulting p-values, measures and effect sizes from each nestedness analysis are available stored as csv files.

5.3 Study I: Measuring nestedness in individual networks

5.3.1 Methods for evaluating nestedness in individual networks

Calculating agreement Consistency between different nestedness measures was evaluated using three different values. We assessed the direct correlation between the measure scores (independent of null models). If two measures were able to

detect similar levels of nestedness, then a monotonic (but not necessarily linear) correlation is expected. In order to measure this we used Spearman's rank correlation coefficient, which assesses correlations of rank order between the nestedness measures.

We also evaluated the consistency of nestedness significance between the measures. To do this we first classified p-values into three categories: $p <= 0.05$, $0.05 < p < 0.95$ and $p >= 0.95$. Then we found the proportional agreement between two nestedness measures by counting the number of times that the same categorisation was reached by both measures and dividing by the number of networks being evaluated. This process was independently repeated for each null model.

Finally we chose to evaluate differences in effect size found using different nestedness measures. The effect size is used to find a standardised number for the strength of nestedness in a particular network. Thus, unlike for the 'raw' measure scores, it is expected that correlations of effect size between two measures should be both monotonic and linear. Therefore we evaluated the correlation between effect sizes using the Pearson correlation coefficient. In addition we expect that using nestedness effect size, it should be possible to recover the amount of rewiring ($P(rewire)$) applied to each network. This latter property would be useful and necessary for comparing nestedness between different networks and is covered in Section 5.4.1.

Discrimination ability To evaluate how well each measure was able to discriminate nestedness by significance, we calculated the proportion of networks at each $P(rewire)$ level that were judged to be significantly nested at the $p <= 0.05$ level. This was repeated for each measure in each null model. A good nestedness measure - null model combination is one where the majority of highly structural nested networks ($P(rewire) = 0.01$) are judged as being significantly nested, avoiding type II error, and the majority of more randomly structured networks ($P(rewire) = 0.5$) are not judged to be significantly nested, avoiding type I error.

5.3.2 Do measures of nestedness agree?

To quantify the direct agreement between measures we plotted the pairwise correlation between measures over the full ensemble of 30,000 networks. If measures were in direct agreement we should observe a linear, or a least monotonic relationship.

Visual inspection shows that nestedness measures do not directly agree with one another (Figure 5.2). There is no linear relationship and in most cases the scatterplots only show loose agreement. Calculation of Spearman's rank correlation coefficient indicates weak, but significant correlations for most pairs, with the exception of the strong relationship between MD and SR. All correlations except of that between JDM and NODF were judged to be statistically significant. Despite the significance of the correlations, the overall finding here is that nestedness measures cannot be directly compared and shows the need for standardisation through the use of null models.

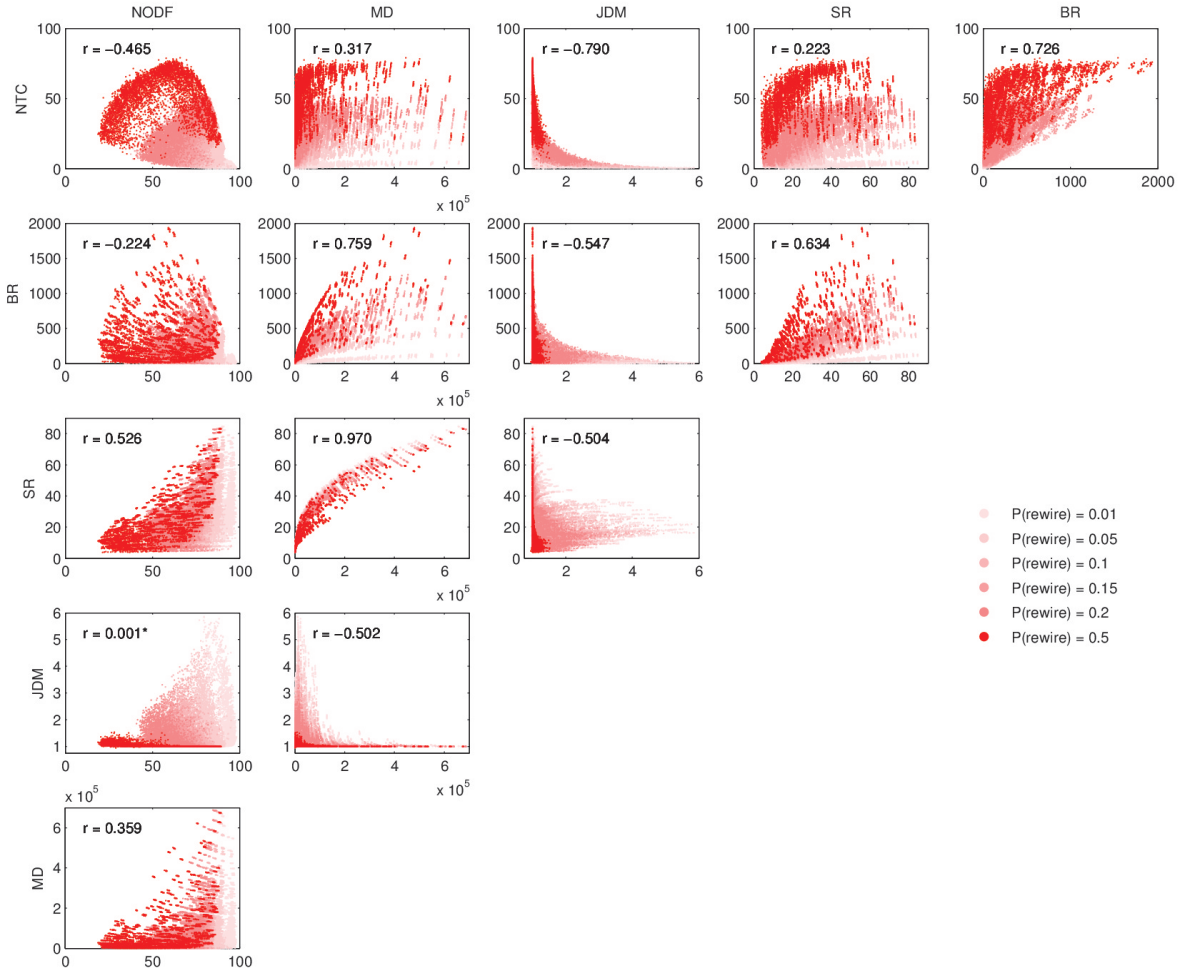


Figure 5.2: Scatterplots of the correlation between each of the considered nestedness measures for the synthetic ensemble of 30,000 networks. $P(rewire)$ of each network is highlighted in each plot. Inset numbers denoted by r show Spearman's rank correlation coefficient between two measures - the correlation was judged to be significant ($p < 2.2 \times 10^{16}$) in all cases, except for the comparison of JDM and NODF ($p = 0.82$) denoted by $\times$.

While we were able to calculate the NTC metric for each of the 30,000 networks (shown in Figure 5.2), we lacked the computational capacity to perform statistical

significance tests with the NTC on every network. Details of the 4,646 networks upon which the NTC was used to calculate statistical significance tests are provided in Appendix B: Measuring the NTC.

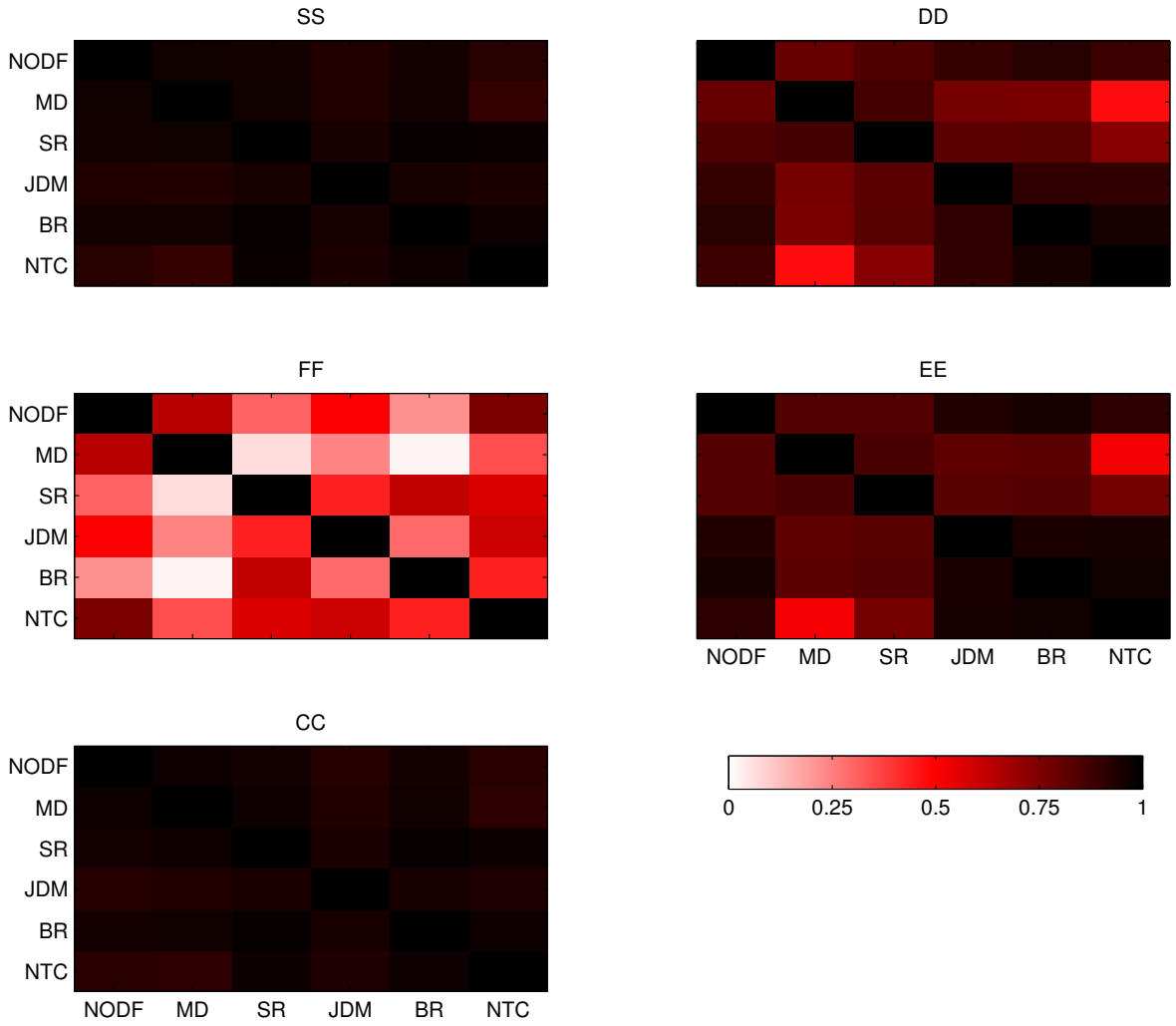


Figure 5.3: Proportion of pair-wise agreement of 'significance' of nestedness between different measures across the synthetic network ensemble. Each subplot shows the results under a different null model. Three categories of significance were chosen : $p <= 0.05$, $0.05 < p < 0.95$ and $p >= 0.95$. An agreement between two measures is reached when both measures classify a network in the same category as one another. This process was repeated for each combination of measures on each of the five null models. Proportion of agreement was made on 4,646 networks where NTC involved; and 30,000 in other cases.

Nestedness is commonly reported in terms of statistical significance. In order to determine the agreement between nestedness measures, pair-wise comparisons of p-values were computed. We used a null ensemble to estimate the probability of a network with greater nestedness being found within the null distribution associated

with each null model (a p-value) for each of the networks studied, for each nestedness measure. Since p-values are typically used to determine whether a network is ($p <= 0.05$) or is not ($p > 0.05$) nested, we classified the output p-values into three categories $p <= 0.05$, $0.05 < p < 0.95$ and $p >= 0.95$ to check for agreement between nestedness measures in different null models.

Figure 5.3 shows, for each of the five null models, the proportion of networks under which the same categorisation of the p-value was reached by two nestedness measures. If two measures report the same p-value classification as each other in every network studied, the proportional pair-wise agreement of significance is equal to 1. We find that the SS and CC null models report strong agreement between all nestedness measures i.e. the same conclusions about network nestedness are reached regardless of the nestedness measure used. SS and CC also show strong similarity with one another. Similarly we find that DD and EE share similar patterns of pair-wise agreement. Within DD and EE we see two clusters of measures with strong agreement: MD and SR share high agreement scores, as does the grouping of NODF, JDM, BR and NTC. This suggests that two types of nestedness may exist in these null models. The FF null model, on the other hand showed no strong pattern of pair-wise agreement between any combination of nestedness measures. In the FF null model, the nestedness found may be entirely dependent on the particular nestedness measure chosen.

5.3.3 How does performance of nestedness differ?

Agreements between measures does not indicate good performance in terms of accurate discrimination between nested from non-nested networks. To test the ability of each measure-null combination to discriminate nested from non-nested networks, we looked at each measures capacity to detect nestedness on networks at each $P(rewire)$ level using a significance threshold of $p <= 0.05$. Figure 5.4 shows the results for the entire dataset (excluding the NTC measure), while Figure S10 shows the results found from a sample of the 2,334 networks where NTC was measured at each $P(rewire)$ (includes the NTC measure). A strong performing measure-null combination should indicate a high proportion of significantly nested networks when $P(rewire)$ is low, to minimise type II error, and a low proportion of significantly nested networks when $P(rewire)$ is high, to minimise type I errors. A secondary performance indicator is that the proportion of significantly nested networks will decline monotonically as $P(rewire)$ increases.

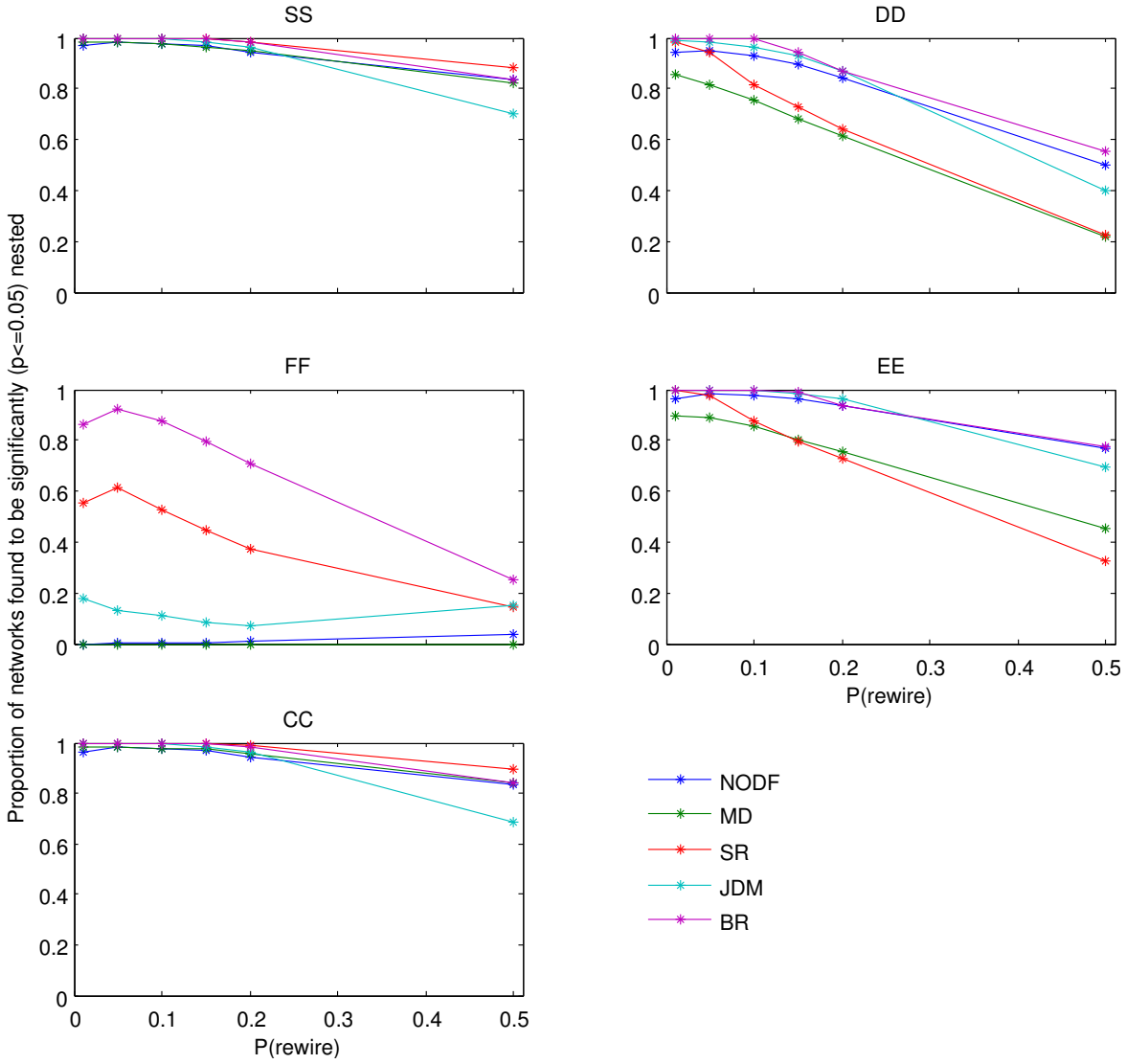


Figure 5.4: The ability of different measures to discriminate significant nestedness under each of the five null models in the full ensemble of 30,000 synthetic networks. Each subplot show the results for a different null model. Each line shows the proportion of networks at each level of $P(rewire)$ that were found to be significantly nested ($p <= 0.05$) using a specific nestedness measure.

These results show the close agreement between the SS and CC null models; and the similar arrangements of measures between DD and EE null models. However, they also reveal that SS and CC null models are susceptible to high type I error - they detect many of the networks with the highest $P(rewire)$ as being significantly nested. The FF null model alternatively shows evidence of high type II error under all measures other than BR and SR - they fail to detect highly nested networks (low $P(rewire)$) as being significantly nested. In both samples BR is better able to discriminate between high and low $P(rewire)$ networks than SR. The DD and EE null models show similar detection ability between the NODF, BR, JDM (and NTC) mea-

sures; whilst MD and SR behave differently. For all measures DD appears better able to discriminate nestedness for high $P(rewire)$ networks than EE.

We note that differences exist between the two samples used (Figure 5.4 and Figure S10). There appears to be an increased ability to discriminate nestedness significance in high $P(rewire)$ networks across the measures in the NTC sample (Figure S10). Additionally in the NTC sample we note that the ability of BR and SR to discriminate nestedness in low $P(rewire)$ networks is greatly reduced in FF, and this also occurs for MD in the DD and EE null models. These differences suggest that the sample of networks upon which we were able to calculate statistical significance with the NTC is biased.

As assigning the threshold p-value at which networks can be judged significantly nested or not is a somewhat arbitrary user choice we also repeated the analysis shown in figures 5.4 and S10 with a significance threshold of $p <= 0.001$ instead of $p <= 0.05$. These results (figures S9 and S11) show similar trends and support the robustness of our analysis.

In all 4 cases and across all tested null models BR exhibited the smallest type II errors (highest proportion of low $P(rewire)$ networks that were significantly nested), whilst MD exhibited the highest type II errors. Type II error was mostly small, except in the FF null model, where it was high for all except the BR measure in the full network ensemble. In the NTC sample of networks BR also suffered high type II errors. This suggests that the NTC sample is a biased sample of networks from the whole ensemble.

Type I errors differed between null models and across the 4 cases considered; in SS and CC null models JDM and MD had lowest type I errors, in FF NODF (MD is as well but this is inconsequential due to it being insensitive to this null model), and in DD and EE MD and SR had the lowest type I errors.

Whilst the combination of SR in the DD null model looks like the best in the whole network ensemble(Figure 5.4); this result does not hold in the NTC network sample. In the smaller sample SR with EE looks best when using $p <= 0.05$ (Figure S9) , but this result is sensitive to the choice of p-value threshold. At $p <= 0.001$ BR with DD appears to be the best choice (Figure S11).

5.3.4 Effect size agreement

Statistical significance offers only a partial indication of the level of nestedness in a given network. Therefore we also evaluated whether effect sizes were comparable between different nestedness measures. Within each null model we plotted effect sizes for each nestedness measure against one another. If effect sizes are comparable between different nestedness measures we expect that pairwise correlations should show a linear relationship. For direct comparability it would also be necessary that effect sizes from both measures were of the same strength i.e. the relationship between effect sizes of two measures is linear and lies on the $y = x$ line. Below we show the pairwise measure correlations between z-scores (Figure 5.5) and adjusted normalised temperature scores (Figure 5.6) in the DD null model. The results for other null models are shown in Appendix E: Comparing effect sizes.

Figure 5.5 shows the pairwise measure correlations of sample z-scores calculated using the DD null model. There appears to be general agreement to z-scores which show strong relationships. Pearson's product moment correlation coefficient was computed to quantify these relationships. All correlations were statistically significantly different from 0 ($p < 2.2 \times 10^{-16}$) and the relationship between BR and NTC was strongest. However, the relationship between z-scores appear to display non-linearity and different nestedness measures have different scalings. This shows that effect sizes are not directly comparable between different nestedness measures.

Using the adjusted normalised temperature as a measure of effect size provides a different viewpoint. Whilst there is still a positive relationship between adjusted normalised temperatures found by different measures the correlations are generally weaker than those found using z-scores (except for NODF-JDM, MD-SR, MD-JDM and JDM-SR). The strongest relationship found was between NODF and SR. Scalings still differed between measures, even if the range of values is more consistent between measures than that found by z-scores. Again, the more random networks (high $P(rewire)$) had scores that were more similar than those networks that had a more nested structure (low $P(rewire)$). It is also noticeable that adjusted normalised temperature scores calculated for NTC and BR appear to show banding of the different $P(rewire)$ levels.

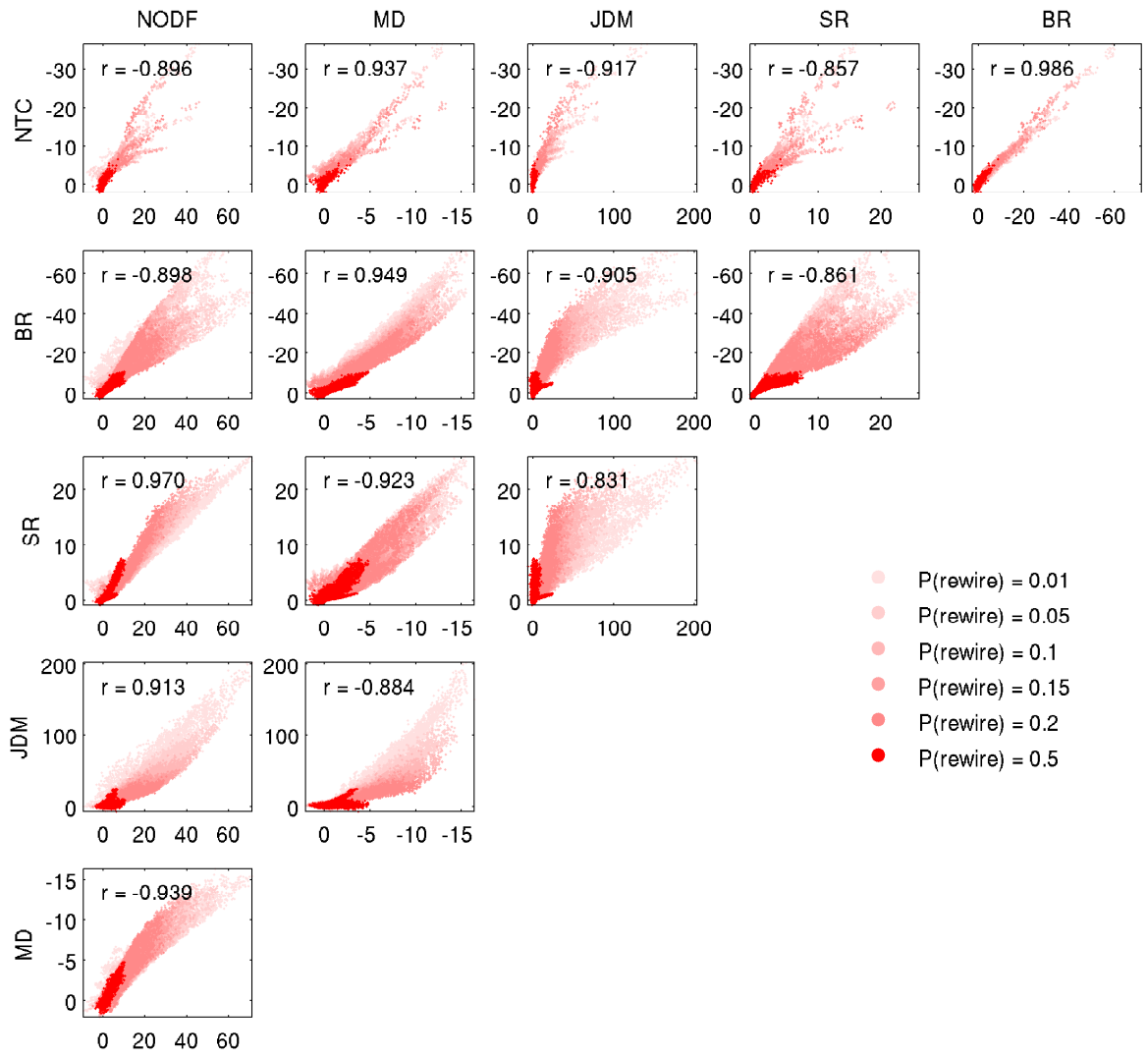


Figure 5.5: Sample z-score measure-measure scatterplots for the DD null model. The axes of MD, BR and NTC measures are inverted, such that the direction of increasingly nested z-scores is in the same direction. Inset are Pearson's moment correlation coefficient for each pair of z-score distributions. Each correlation is statistically significantly different from 0 ($p < 2.2 \times 10^{-16}$).

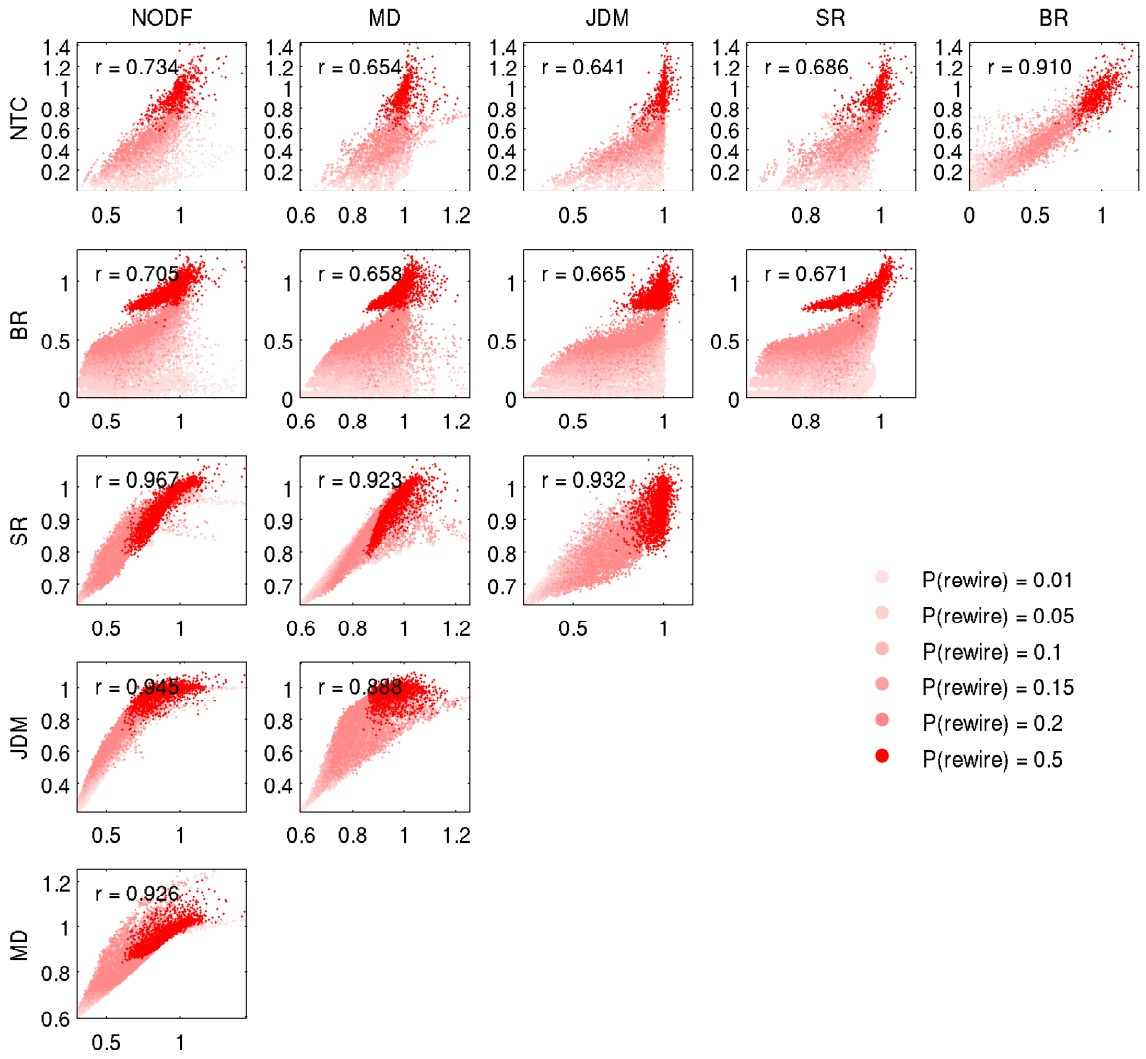


Figure 5.6: Adjusted normalised temperature scatterplots for DD null model. Inset are Pearson's moment correlation coefficient for each pair of adjusted normalised temperature score distributions. Each correlation is statistically significantly different from 0 ($p < 2.2 \times 10^{-16}$).

These patterns for adjusted normalised temperature are repeated across the SS, CC and EE null models which all appear very similar (figures S17, S19 and S21). Only FF (Figure S18) appears different, where there is poor agreement between all nestedness measures and only BR shows signs of banding $P(rewire)$'s.

There was also poor agreement across nestedness measures in FF for z-scores, and no obvious banding (Figure S13). The DD and EE (figures 5.5 and S16) exhibit similar patterns in z-scores, as do SS (Figure S12) and CC (Figure S14), where SS and EE respectively have larger scalings than their counterparts.

5.3.5 Conclusions on evaluating nestedness in individual networks

Within the SS and CC null models all measures displayed similar efficacy at detecting nestedness. Within the FF null model, there was weak agreement between all nestedness measures; and in the DD and EE null models, two groups of nestedness detecting measures appeared.

We find that the SS and CC null models performed in almost the exact same fashion in terms of statistical significance, though effect sizes differed in magnitude. In SS edges are randomly shuffled around the network whilst preserving the number of nodes, whilst in CC edges between periphery nodes (those with low node degree) are protected and edges connected to core nodes (with high node degree) are shuffled through the network. Though these null models share different philosophies, the results generated are extremely similar. On this basis we would recommend disregarding the CC null model in its current implementation in favour of the simpler rules of the SS null model. However, it is also noted that neither SS or CC were able to accurately resolve the differences between low and high $P(rewire)$ networks, experiencing high type I errors, which meant that the majority of networks tested showed significant nestedness, regardless of their $P(rewire)$. They are unable to discriminate between nested and non-nested networks.

The FF null model, on the other hand, suffers from high type II errors under the majority of the nestedness measures we tested. In addition there was poor agreement on statistical significance between all measures (Figure 5.3). Indeed it is noted that due to the formulation of MD, this measure finds the same nestedness score in every FF null ensemble network and cannot be used. However, BR, appeared less susceptible to this problem and actually appears to be able to discriminate nestedness between different $P(rewire)$ networks. If the FF null model is used we recommend using the BR measure.

The probabilistic null models DD and EE suffered less from type I and type II errors than the mechanistic null models (SS, CC, FF). Discrimination between low and high $P(rewire)$ networks was best in the DD null model, which uses information about each node degree, rather than just the total number of edges in the network. Within both these null models we found two tentatively different groups of measures that agreed more with each other than other measures. These groupings suggest that MD and SR may be discovering a different kind of nestedness to that found by the NTC, NODF, BR and JDM in these null models. As NTC and NODF are widely used

nestedness measures whose meaning can be more easily interpreted in a visual context than a networks eigenvalue, it may be tempting to oust SR as a brood parasite of nestedness! However, we actually find that SR may be able to better discriminate between low and high $P(rewire)$ networks for nestedness than any of the other tested measures. Therefore using SR in the DD null model appears the best available measure-null model combination as assessed by significance levels. However, this was not true in the NTC sample (shown in Appendix D: Robustness of nestedness discrimination ability).

5.4 Study II: Comparison of nestedness between networks

5.4.1 Methods for evaluating nestedness across multiple networks

Effect size and rewiring level To assess the ability of effect sizes to recover the $P(rewire)$ applied to the initial networks we rank ordered the effect sizes from weakest to strongest effect and found the corresponding $P(rewire)$'s associated with each. We call this list Y. This was compared with list X , composed of the list of $P(rewire)$'s that would be expected if effect size were able to recover the level of rewiring applied to each network i.e. the smallest effect sizes are when $P(rewire)$ is greatest and the largest effect sizes are found when $P(rewire)$ is lowest. Average absolute displacement, M , between lists Y and X was calculated for each nestedness measure in each null model as:

$$M = \frac{1}{N} \sum_{i=1}^{N} (|Y_i - X_i|) \tag{5.4}$$

where N is the number of networks that the test was calculated for. Then we normalised M by the maximum amount of average absolute displacement M_{max} , calculated by replacing Y with the list X in reverse in Equation 5.4. Normalised average absolute displacement ($NAAD$) was then found as:

$$NAAD = \frac{M}{M_{max}} \tag{5.5}$$

Network properties Ideally nestedness analysis should be unbiased by network properties [126]. To investigate this we compared the computed nestedness effect sizes to three network properties: **size** (number of rows multiplied by the number of columns in the networks biadjacency matrix), **fill** (the number of network edges - the total number of interactions between the two node types) and **connectance** (the proportion of total interactions calculated as fill divided by size).

Standardising effect sizes Effect sizes were standardised for each of the network properties. This was achieved by a) finding all effect sizes (E) for each unique network property value b) finding the minimum (min) and maximum (max) effect size for this network property value and c) transforming each of the effect sizes (E_i) at this network property value onto a new range (F) between 0 and 1 as:

$$F_i = (E_i - min)/(max - min) \tag{5.6}$$

5.4.2 Can we compare nestedness between networks using effect sizes?

In order to evaluate whether nestedness effect sizes were able to discriminate between networks of different rewiring levels the normalised average absolute displacement was computed. The normalised average absolute displacement between expected rewiring level and the observed rewiring level, ranked by two different effect size measures, was calculated for each measure in each null ensemble. The smaller that normalised average absolute displacement is, the better the agreement between the observed and expected ordering of $P(rewire)$ levels and the more useful this combination may be at comparing nestedness between different networks. Figure 5.7 shows the mean average displacement for all measure-null-effect size combinations in this study across the synthetic ensemble. The BR measure finds the best agreement between observed and expected $P(rewire)$ levels in all of the null models using the adjusted normalised temperature as an effect size. Results for the sample of networks from the ensemble upon which the NTC was calculated were similar (Figure S22) where again BR using the adjusted normalised temperature came out on top. Whilst scores in the FF null model were worse, the results between the other null models were close. However, in both the NTC sample and that of the whole synthetic ensemble the best results were found using the CC null model.

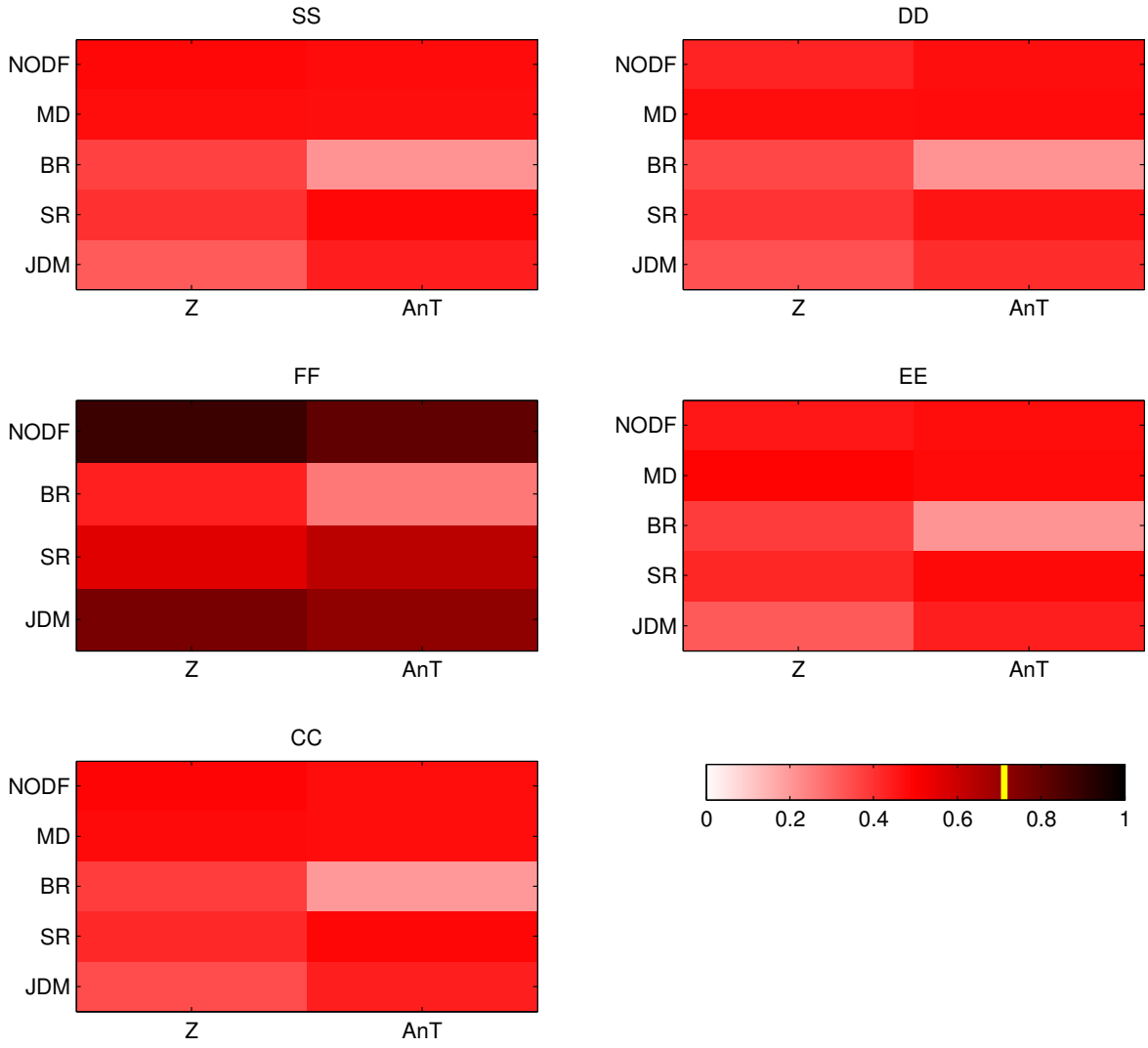


Figure 5.7: Comparative ability of effect sizes calculated from each null model for each nestedness measure by computing normalised average absolute displacement (NAAD) scores between observed $P(rewire)$ and expected $P(rewire)$. Each subplot shows results under each null model, where nestedness measures are across the rows and different effect sizes are made across the columns. Effect sizes used were sample z-scores (Z) and adjusted normalised temperature scores (AnT). As an effect size for Manhattan distance (MD) cannot be calculated using the FF null model, so this was not shown in the figure. The mean expected NAAD score for a random effect size is marked on the colourbar(10,000 random samples gave a distribution of NAAD scores with mean=0.7126 and standard deviation=0.0029).

5.4.3 Why does nestedness differ?

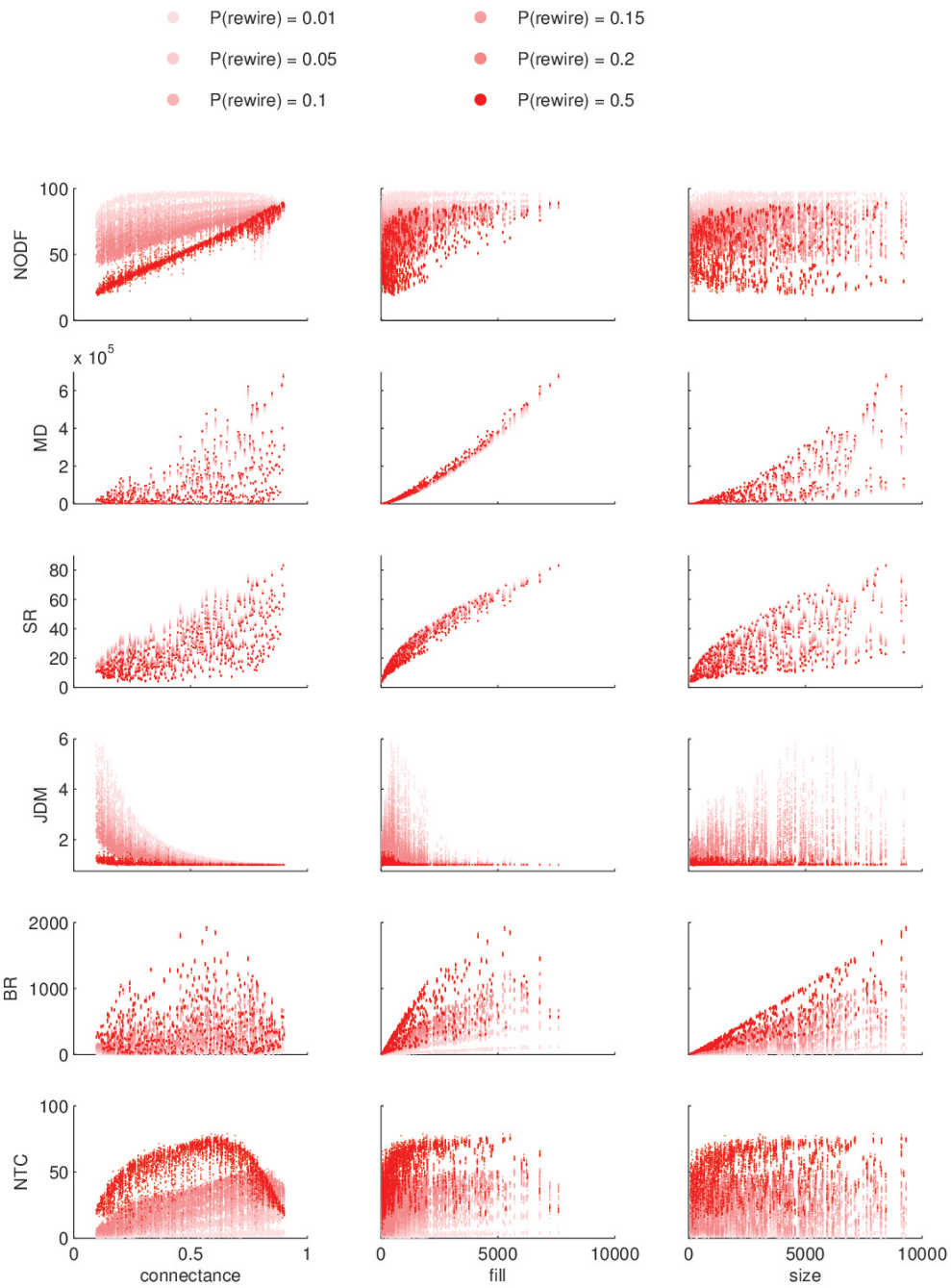


Figure 5.8: How nestedness measures relate to network properties of connectance, fill and size. $P(rewire)$ is highlighted for each of the 30,000 networks in the synthetic ensemble.

Another question is what causes nestedness measures to differ. As shown in Figure 5.2 differences exist between the measures, in Figure 5.8 we plot measures against three network features: connectance, the proportion of network edges, fill, the abso-

lute number of network edges, and size, the potential number of interactions. Preferably, networks with the same $P(rewire)$ level should have similar nestedness, independent of these network properties, so that horizontal bands corresponding to each $P(rewire)$ are evident. While in networks with low connectance it is possible to distinguish between rewiring levels of networks measured with NODF or JDM; we find networks with high connectance all appear to be comparably nested using these measures. The NTC appears to be able to distinguish different rewiring levels for most connectances - but the relationship is non-linear (especially noticeable in networks where $P(rewire) = 0.5$). The network fill strongly effects the range of possible values that MD and SR can take. This shows another reason for standardisation using a null model - to account for the effects of these network properties. However, if the recorded effect sizes do not standardise for these observed differences these network features may introduce bias to analysis, undermining the ability to relate information about the underlying network structure.

5.4.4 Network properties affect effect sizes

The way in which nestedness measures varies with network properties also influences the range of effect sizes that are recorded. Below we show how connectance covaries with z-scores (Figure 5.9) and adjusted normalised temperature scores (Figure 5.10). Connectance in our synthetic ensemble varied between 0.0972 and 0.9012, and at least 50 networks were analysed at each unique network property value. Covariation of effect size with network size and fill are shown in the Appendix F: Effect size and network properties. For z-scores we found that increasing network size and fill increased the range of possible z-score values for all null model and nestedness measure combinations (figures S23 and S24). We found adjusted normalised temperature scores were also effected by network size and fill (figures S25 and S26); except in the cases when the BR nestedness measure (Figure 5.11) and the NTC nestedness measure was used. However due to the paucity of data collected for the NTC measure (see Appendix B: Measuring the NTC) in comparison to other nestedness measures this is harder to evaluate.

As one of the reasons for performing standardisation testing to find an effect size is to remove statistical biases related to network properties - the majority of nestedness measure - null model combinations do not look appealing. We find low fill and small sized networks have excluded z-score space across all nestedness measures. Similarly there are strong relationships between connectance and the majority of nestedness measures evaluated using the adjusted normalised temperature. Even BR,

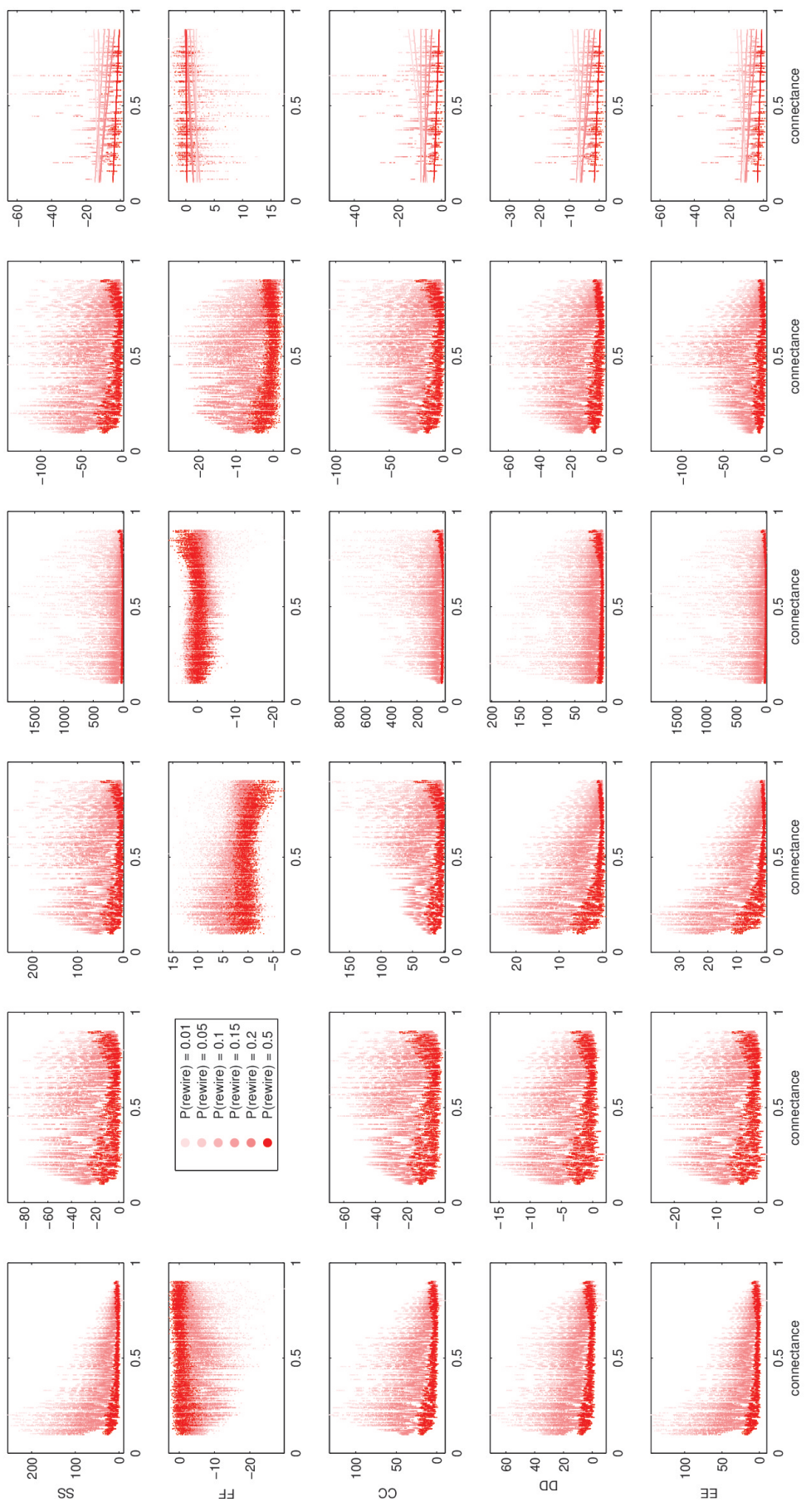


Figure 5.9: Z-score covariation with connectance. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure. Y-axes for the MD, BR and NTC measures are reversed, such that increasing strength of nestedness effect is moving upwards within each subplot. Linear regression lines for each $P(rewire)$ are drawn for the NTC measure to aid comparison.

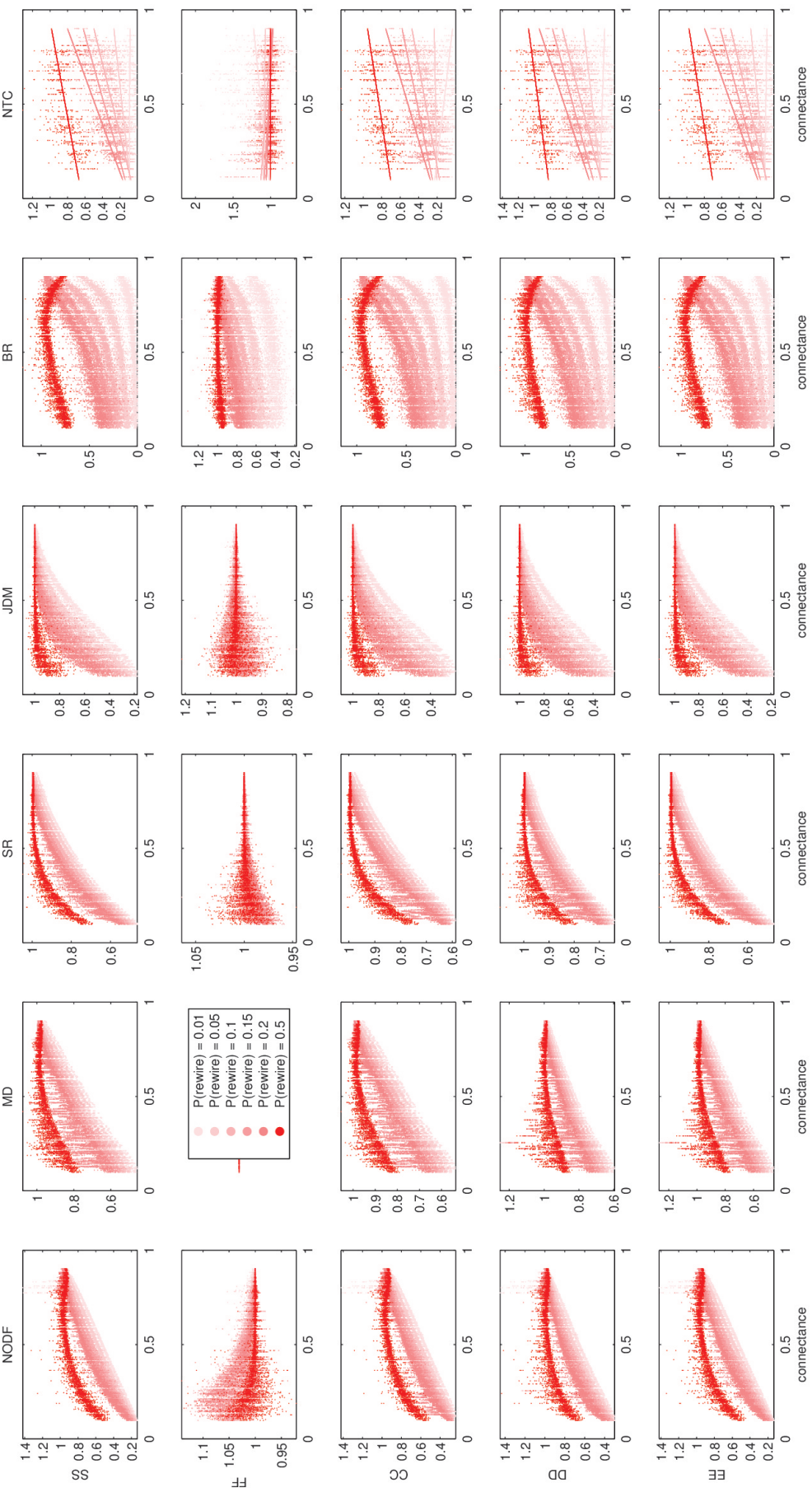


Figure 5.10: Adjusted normalised temperature covariation with connectance. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure. Linear regression lines for each $P(rewire)$ are drawn for the NTC measure to aid comparison.

which appears to be less biased against network fill and size (Figure 5.11) using the adjusted normalised temperature is effected at higher levels of connectance.

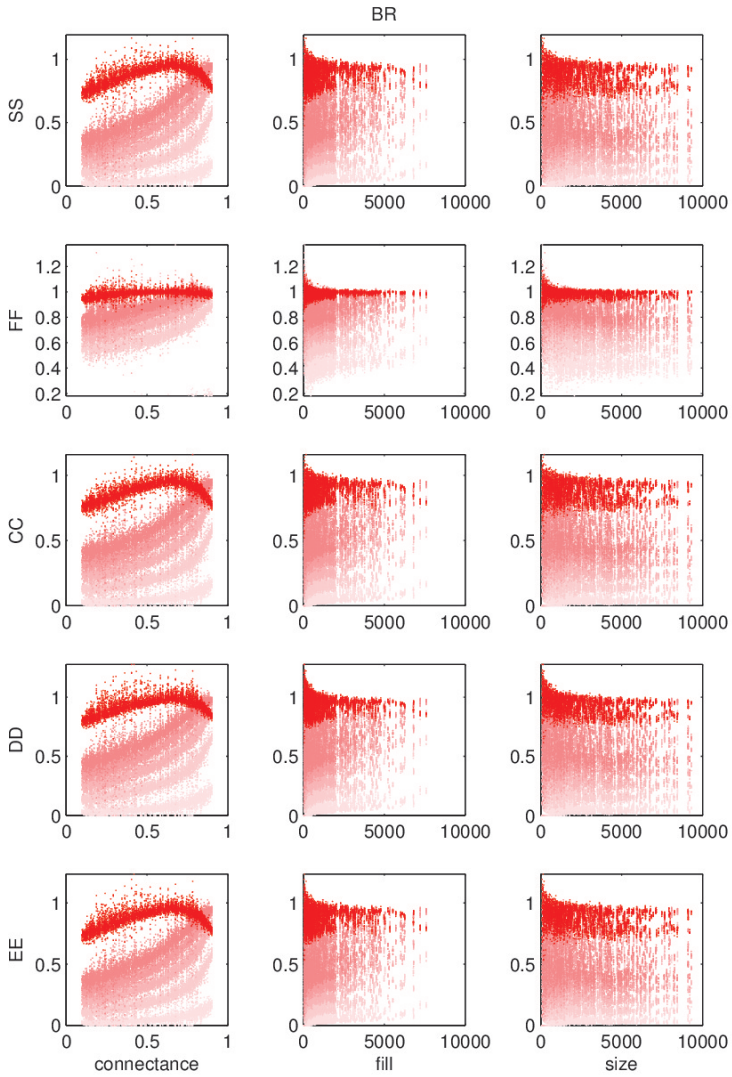


Figure 5.11: Adjusted normalised temperature scores found using discrepancy (BR) under each null model (each row of plots). Plotting these against network properties shows bias in highly connected networks; and that bias appears less related to size and fill except in networks with very few edges or that are very small.

5.4.5 Standardising effect sizes

For each network property value there exist several effect size calculations from different networks at different $P(rewire)$'s within our ensemble of synthetic networks. By transforming the range of effect sizes calculated at each network property value onto the range [0,1] we were able to assess the ranking of effect sizes against their

associated $P(rewire)$ values. Figures in Appendix G: Standardising the standardisation show that performing this transformation reduces bias associated with network size and fill for both z-scores and adjusted normalised temperatures. However, when comparing the transformed values against connectance (figures 5.12 and 5.13) we found that the ability to distinguish between $P(rewire)$ levels is reduced for networks with high connectance as shown in Figure 5.10 for BR. This suggests that network connectance leads to a systematic bias in the way that nestedness is calculated.

Qualitatively figures 5.12 and 5.13 show the same patterns:- that only BR can noticeably detect changes in $P(rewire)$ using the FF null model; in general when connectance is low $P(rewire)$ levels are detectable, but as connectance increases the ability to distinguish between different $P(rewire)$ levels becomes more difficult. The way in which this detection ability decays as connectance increases does differ between nestedness measures, effect sizes and (to a lesser degree) null models. Expected bias in networks with low connectance [120] was harder to spot, but at lower connectances it is harder to differentiate between $P(rewire) <= 0.2$ networks.

5.4.6 Conclusions on comparing nestedness between networks

As effect size measurements are made through evaluation on a null model they report where a network falls in relation to the distribution of null networks. But, if we want to compare nestedness across different networks we need to use metrics that allow us to posit that network A is more nested than network B. As a proxy to answering this question we ranked the reported effect sizes within each null model by rewiring level and measured the average distance from the expected rewiring level. Our results highlight the choice of effect size measure as a further complicating factor in nestedness analysis - such that measure-null-effect choices may aid/hamper cross-examination of nestedness between different networks. Indeed, we show that the best way to compare nestedness between networks is using the BR measure with the adjusted normalised temperature as an effect size metric. Our results show that this works best in the CC null model; but similar scores were obtained in SS, DD and EE null models.

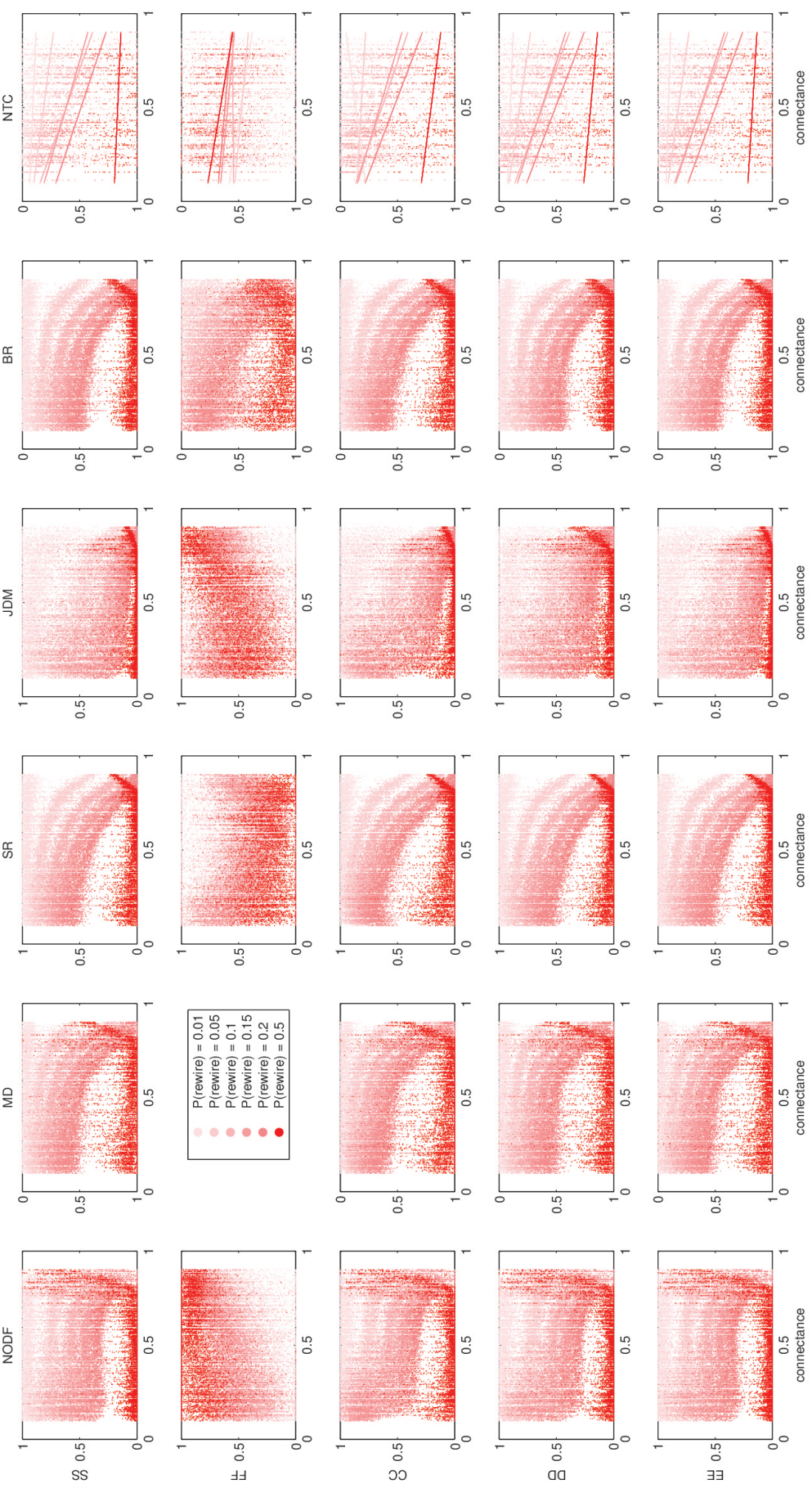


Figure 5.12: Standardised z-scores shown against connectance calculated for each nestedness measure (columns) within each null model (rows). $P(rewire)$ levels are highlighted.

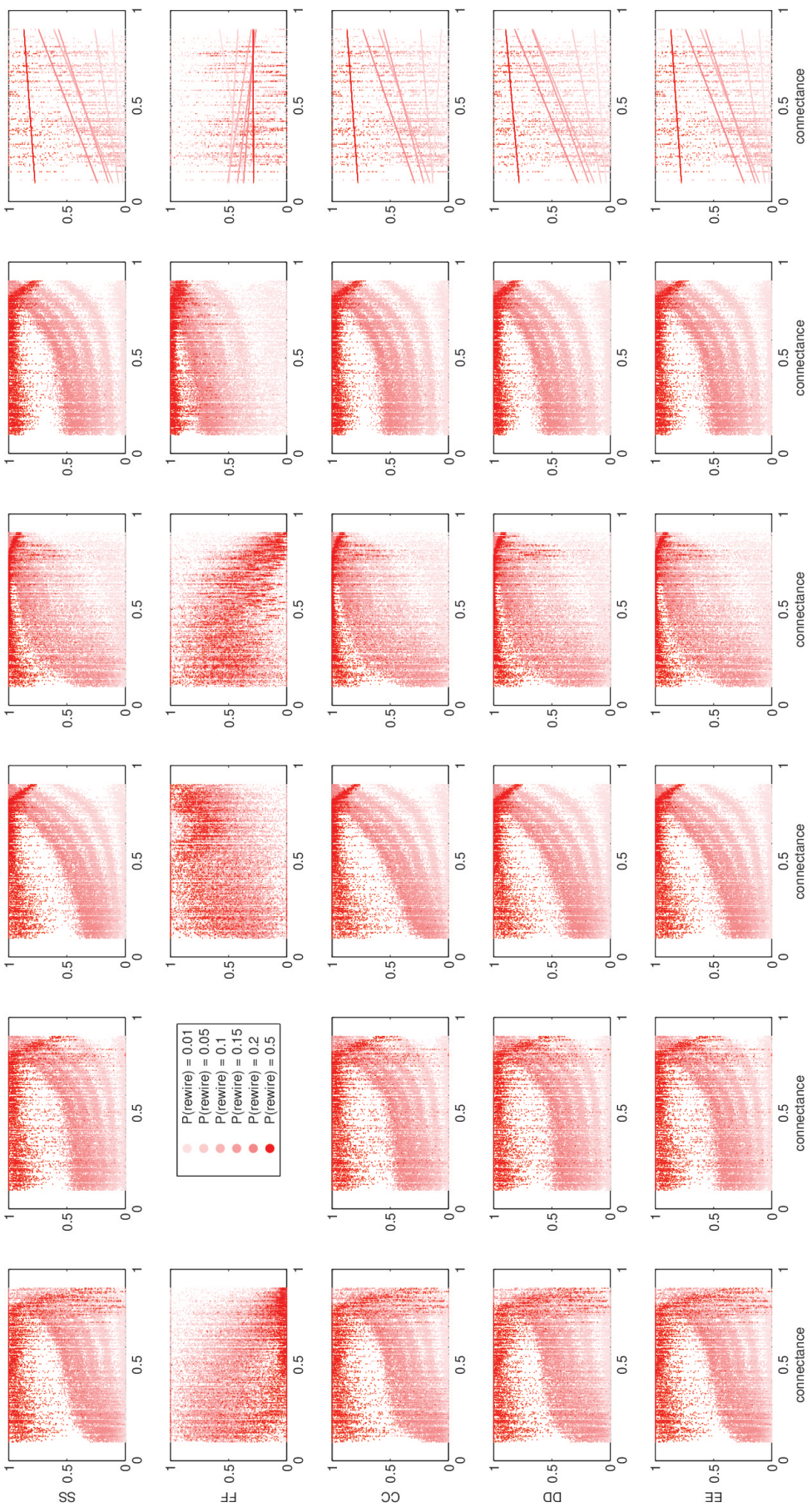


Figure 5.13: Standardised adjusted normalised temperature scores shown against connectance calculated for each nestedness measure (columns) within each null model (rows). $P(rewire)$ levels are highlighted.

5.5 Discussion

In order to perform nestedness analysis a large number of choices need to be made by a user. The key decisions investigated here are the choice of nestedness measure, the choice of null model, and the choice of statistics to summarise that null model. Whilst the choice of null model is context dependent (i.e. it is dependent on the questions you wish to ask of the network), the choice of null model statistics and nestedness measures is context free - and we would hope there is a rational way to choose between the available methods. However, we find that the choices made by a user have important implications for detection and interpretation of nestedness within a particular network.

A good nestedness measure should be symmetric, so that the same answer will be gained regardless of which type of nodes are represented by rows or columns [15, 36]. If nestedness is to be tested against extrinsic variables then it is useful to take measurements using measures that explicitly depend on the order of rows and columns within a matrix [15, 36]. However, gradient free or order invariant measures such as SR and JDM, can be applied when questioning the existence of nestedness structure - the subject this study focuses on. In addition we feel that network structure will appear as a result of complex interplay between several extrinsic and intrinsic factors, which may be difficult to disentangle without use of more mechanistic, system motivated models.

There exists a large number of measures that claim to measure nestedness. One of the key questions we wanted to ask with this study was which measure(s) should fledge for the nestedness concept; and whether there are any that should be avoided. Our results from 773,230 statistical tests show there is no simple answer to this question as it explicitly depends on the null model choice used to investigate nestedness.

In Section 5.4.3 it was seen that nestedness measures are related to network features of connectance, fill and size in different ways – and this may be one reason for differences between the measures. Another plausible explanation of the observed differences between nestedness measures relates to the way in which nestedness measures are in general defined. A network structure with high nestedness is acknowledged as one in which specialist nodes act upon a subset of the the nodes that a more generalist node interacts with. However, this definition excludes what kind of structuring a network with low nestedness would correspond to. It may be that different measures of nestedness have different definitions of a low nestedness network

configuration; and furthermore the way in which the level of nestedness declines between high and low nestedness configurations may vary between the measures. This exercise is left for future work, but hints toward being cautious in choice of nestedness measure and the interpretation of its result.

Taking these two explanations together suggests that measures of nestedness cannot be treated as being equivalent to one another; indeed, they may be measuring different quantities of a network. Furthermore, if nestedness is to be considered a useful concept in network theory it should be necessary to provide details about how the gradient that nestedness is measured on is defined and what makes any particular network more or less nested than any other particular network.

However, by evaluating the distribution of effect sizes by the underlying network properties we confirm that there exists a systematic bias to the connectance of a network [120, 126] - particularly to more highly connected networks. There may exist a 'safe zone' within which BR with the adjusted normalised temperature can be used to compare different networks. Whilst we may expect the majority of real networks to have connectances within the 'safe zone' there will be exceptions. Comparing adjusted normalised temperatures of networks inside and outside the 'safe zone' is not a fair comparison.

Despite the numerous nestedness measures in existence, finding measure-null-effect size combinations that are not sensitive to connectance, or other network properties should become a fundamental research goal for those who wish to compare nestedness between two or more networks to elucidate the real-world significance of any discernible patterns, rather than some superfluous statistical artifact.

5.6 Conclusions

Our analysis using a large synthetic ensemble of bipartite networks reveals that analysis methodology can impact the results of nestedness analysis. Noise based testing on the basis of rewiring suggests that using SR with the DD null model is the best compromise of minimising type I and type II errors in significance testing. However, there was typically a tradeoff between measures that were good at reducing type I errors, but not so good at reducing type II errors or vice versa. It may be prudent to use two measures of nestedness in analysis of bipartite networks - one which is better at avoiding each type of error, to make a more informed judgement about the

significance of nestedness in a network. However, in the FF null model only BR was able to discriminate between nested and less nested architectures with any efficacy. In the SS and CC null models, no measure was insusceptible to type I errors and on this basis we recommend avoiding these null models.

Null models are used in order to say how likely a pattern is to appear, in comparison to some given set of rules. The most popular way is by comparing against a randomised set of networks that conserve some of the core features of the system under observation. This statistical top down approach may be useful for discerning whether a pattern exists or not, but is not based on realistic rules of community composition. Mechanistic approaches [131], where can study how strength of different processes may impact the likelihood of interaction structure, may be a more informative approach - but this requires system specific knowledge.

Sample z-scores are easy to interpret in the context of a single network and a single null model. However, as basic network properties can bias the range of attainable effect sizes it is in general not recommended to attempt to compare effect sizes between different networks. Of the methods we tested BR measured using the adjusted normalised temperature as an effect size comes out as the best option. This combination works well across null models (though worst in the FF null model); and while it does not suffer from bias due to size or fill, it does from connectance. However, it may serve for comparing networks in the regime where connectance is between the lower limit we tested (0.0972) and 0.5. In this 'safe zone' the relationship between effect size and connectance of different $P(rewire)$ was relatively flat. When connectance is greater than 0.5, it became harder to discriminate between different $P(rewire)$ levels. We also expect this may occur at connectances lower than we tested - in both cases due to the smaller number of possible states that a network can take [120]. In order to fairly compare nestedness across networks, these biases need addressing.

Chapter 6

Improved community detection in weighted bipartite networks

Abstract

Real-world complex networks are composed of non-random quantitative interactions, yet many community detection algorithms only use the presence or absence of interactions between nodes. Weighted modularity is a potential method for evaluating the quality of communities in quantitative networks. Modularity optimisation is a method for finding communities in a network. QuanBiMo has been proposed to maximise weighted modularity in bipartite networks. This paper introduces two new algorithms, LPAwb+ and Exhaustive LPAwb+, for maximising weighted modularity in bipartite networks. These algorithms robustly identify partitions with high modularity scores. Exhaustive LPAwb+ consistently matched or outperformed QuanBiMo, whilst the speed of LPAwb+ makes it an attractive choice for detecting the modularity of larger networks. Searching for modules using weighted data (rather than binary data) provides a different and potentially insightful method for evaluating network partitions.

This chapter is based on the submitted paper:

Beckett S.J. Improved community detection in weighted bipartite networks.

I thank Xin Liu for the hints and nudges that helped me setup the LPAb+ algorithm - leading to the weighted counterpart presented here. I thank Timothée Poisot and Hywel Williams for comments that improved the manuscript.

6.1 Introduction

Bipartite networks are the representation of interactions between two distinct classes of nodes. Identifying structure within these networks is useful in explaining their formation, function and behaviour. Modularity is an evaluation of the partitioning of nodes into separate subsets, forming modules. These are also known as groups, compartments, communities or subgraphs. Determining functional groups of networks is an important challenge for a diverse set of fields including sociology, ecology and the physical sciences.

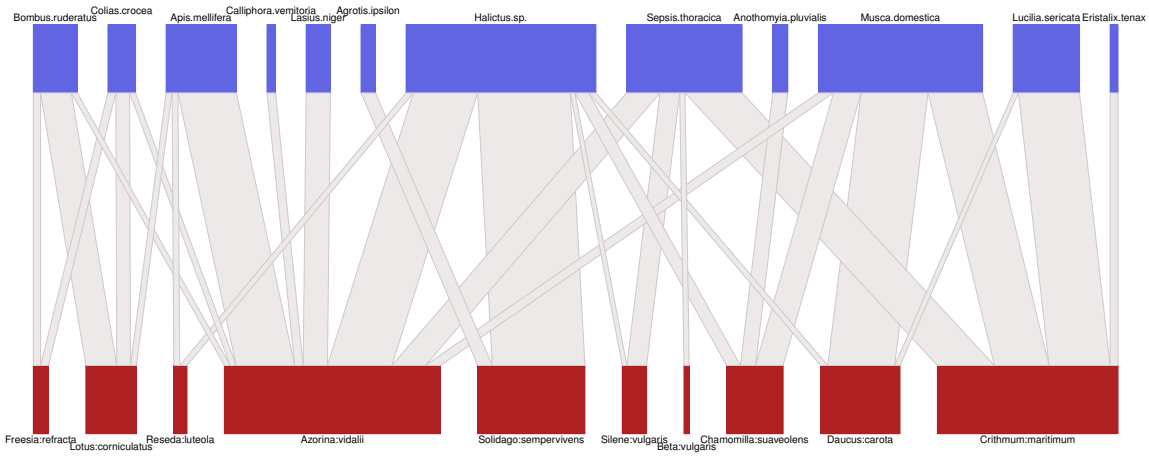
Maximising the modularity of a network is one method for detecting communities originally developed for unipartite (in which all nodes are allowed to interact with one another) networks [41]. Modularity is highest when each module appears isolated from the rest of the network. This occurs when nodes interact often with nodes in the same module, but there are few between module interactions. Negative modularity scores imply fewer interactions occur within modules than expected in a random network. But, positive modularity indicates that within module connectivity is higher than expected. The smallest and largest possible modularity scores that can be found are network dependent [132].

There are several definitions of modularity used in bipartite networks. Guimerà's modularity [45] and Barber's modularity [44] were recently reviewed [47] in the context of ecological networks. Guimerà's modularity uses weighted projections to identify separate communities within each node type. In contrast, Barber's modularity identifies joint communities composed of both types of node. In this paper I concentrate on the modularity definition proposed by Barber and its extension to weighted networks [48] to search for communities composed of both node types, which in the context of this study are communities of plants and their respective pollinators.

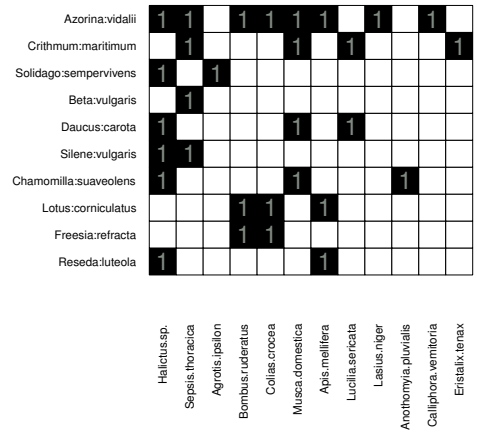
Modularity is a major feature of plant-pollinator networks [2] and may contribute to network stability in these systems. They can be represented as bipartite networks with interactions between pollinators and plants. Pollinating species cannot pollinate other pollinating species, while plants cannot visit each one another – the only allowed interactions are between different plants and pollinators (an example network is shown in Figure 6.1).

The majority of approaches to community detection only focus on whether two nodes have an association, regardless of the strength of those associations [43]. However, there are some exceptions in unipartite networks [133]. QuanBiMo [48] is the

(*a*)



(*b*)



(c)

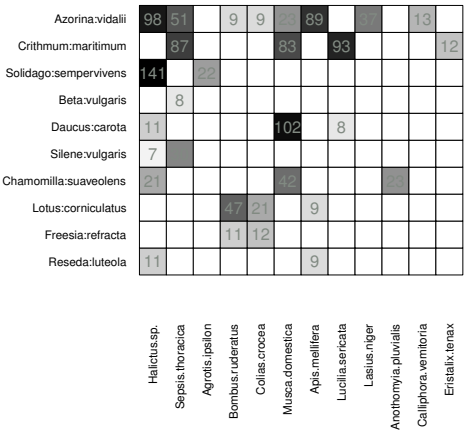


Figure 6.1: (*a*) The olesen2002flores bipartite network of 12 species of pollinators (blue nodes (top)) visiting 10 plant species (red nodes (bottom)). The width of the edges linking the nodes represents the number of pollinator-plant visitations, whilst the width of the nodes represents the marginal total of visits made by a pollinator species or received by a plant species. (*b*) The same network represented by the incidence matrix denoted $\tilde{A}$ in the text, where the plant species are represented as rows and the pollinator species as columns and the presence of visitations between a pollinator and plant species is represented by a 1. (*c*) The incidence matrix $\tilde{A}$ is the binary equivalent of $\tilde{W}$, the weighted interaction matrix shown here. The cell numbers correspond to the number of observed pollinator-plant visitations that occurred (where there is no number in a square there were 0 visitations)

first algorithm to maximise weighted modularity in bipartite networks with quantitative data.

It may be possible to adapt some of the methods available for binary data to deal with quantitative information, rather than having to discard this important data dimension. The LPAb+ algorithm [49] for maximising modularity in binary bipartite networks has been shown to outperform seven other available methods for binary networks [49, 134] whilst retaining fast time complexity. These qualities make it a good candidate for extension to the case of weighted networks.

The definitions of binary and weighted modularity are presented. I show how to alter the LPAb+ algorithm so it can detect weighted modularity and denote this algorithm LPAwb+. A further modification allowing a more thorough search of modularity space is also presented. I call this Exhaustive LPAwb+. The three algorithms for maximising weighted modularity are compared on a dataset of 23 plant-pollinator networks. I find that QuanBiMo is highly sensitive to its input parameters, which may lead to reporting of modularity far below the optimal value in a given network. QuanBiMo reported less consistent modularity scores than either LPAwb+ or Exhaustive LPAwb+. These experiments show that Exhaustive LPAwb+ and QuanBiMo performed well on smaller networks, whilst the speed of LPAwb+ makes it particularly suitable for use on larger datasets. The inclusion of quantitative information in networks alters the structure of detected modules.

6.2 Methods

6.2.1 Modularity

Barber's modularity

Bipartite or two-mode networks are made of two disjoint sets of nodes such that interactions only occur between nodes of opposite types. To generalise we say there are two node types: red and blue - and that interactions are only allowed between red and blue nodes. If there are r nodes of the red type and c nodes of the blue type, the adjacency matrix A is given in block diagonal form as:

$$A = \begin{pmatrix} 0_{r\times r} & \tilde{A}_{r\times c} \\ \tilde{A}^T_{c\times r} & 0_{c\times c} \end{pmatrix}$$

where $\tilde{A}$ is the incidence matrix describing the connections between the different types of nodes (here T indicates the matrix transpose). This formulation allows bipartite modularity to be written as [44]:

$$Q_B = \frac{1}{m}\sum_{u=1}^{r}\sum_{v=1}^{c}\left(\tilde{A}_{uv} - \frac{k_u d_v}{m}\right)\delta\left(g_u, h_v\right) \tag{6.1}$$

where m is matrix fill - the number of edges in $\tilde{A}$, k describes the node degree for red nodes (the number of blue nodes each red node interacts with) and d describes the node degree for blue nodes (the number of red nodes each blue node associates with). Red node labels are denoted g , whilst h are the labels for blue nodes and the Kronecker delta function $\delta\left(g_u, h_v\right)$ is equal to one when nodes u and v are classified as being in the same module (i.e. they have the same label value) or zero otherwise.

Weighted bipartite modularity

Weighted bipartite modularity, Q_W , can be defined as [48]:

$$Q_W = \frac{1}{M}\sum_{u=1}^{r}\sum_{v=1}^{c}\left(\tilde{W}_{uv} - \tilde{E}_{uv}\right)\delta\left(g_u, h_v\right)$$

$$= \frac{1}{M}\sum_{u=1}^{r}\sum_{v=1}^{c}\left(\tilde{W}_{uv} - \frac{y_u z_v}{M}\right)\delta\left(g_u, h_v\right) \tag{6.2}$$

$\tilde{E}$ is the matrix of the null expectations of interaction between two nodes, where y is the row marginal totals and z is the column marginal totals of $\tilde{W}$, the weighted incidence matrix. In a binary network $\tilde{W}$ is equivalent to the binary incidence matrix $\tilde{A}$, the marginal totals will equal the node degrees ($y = k$ and $z = d$) and M, the sum of edge weights will equal m, the fill. Thus Equation 6.2 will reduce to Equation 6.1 for a binary network. Furthermore Equation 6.2 can be reformulated into its matrical form [44, 54] to allow for vectorised computation as:

$$Q_W = \frac{1}{M} tr\left(R\left(\tilde{W} - \tilde{E}\right)C\right) \tag{6.3}$$

where for a network with F communities, R is the F×r red label matrix and C is the c×F blue label matrix. R (and C) are binary matrices with a single 1 in each row (column) indicating which community each red (blue) node belongs to (this information is held by the red and blue labels). These definitions of weighted bipartite modularity can now be used in the modified framework of the LPAb+ algorithm.

6.2.2 Weighted modularity maximising algorithms

QuanBiMo

The quantitative bipartite modularity algorithm (QuanBiMo) of [48], based on the hierarchical random graph algorithm [135], uses a simulated annealing method to attempt to maximise weighted bipartite modularity. It is a C++ routine that is available in the R package bipartite [136] through the function `computeModules`. The default settings available in bipartite version 2.04 were used (steps$= 10^6$, tolerance$= 1^{-10}$).

LPAwb+

A key feature of the LPAwb+ algorithm is that it simplifies to the LPAb+ algorithm when a binary network is used as input. The algorithm is made from two stages - a 'bottom up' step that maximises modularity on a node-by-node basis using label propagation; and a 'top down' step that joins modules together when it results in increased network modularity. First the dimensions of the network are used to decide how to run the algorithm (whose pseudocode is given in Algorithm 6.1); this is because a bipartite network can have at the most $F = min(r, c)$ communities with our chosen definition of modularity. The LPAwb+ algorithm is initialised by giving a unique label to each of the nodes in the smallest of the two sets. The LPAwb+ algorithm is sensitive to the initial labelling of nodes - this can lead to different values of modularity being reported. To combat this issue the initial node labels are randomly assigned and it is suggested that the LPAb+ algorithm is run multiple times on a given network to find the greatest modularity score [49].

Stage 1 - label propagation stage - bottom up

Asynchronous updating of red, then blue labels on the network is performed to locally maximise modularity (Equation 6.2). For a particular red node x this can be written as choosing a new label g_x by trying to maximise the condition:

$$\begin{aligned} g_x &= \left(\sum_{v=1}^{c}\left(\tilde{W}_{xv} - \tfrac{y_x z_v}{M}\right)\right)\delta\left(g, h_v\right) \\ &= \left(\sum_{v=1}^{c}\tilde{W}_{xv}\delta\left(g, h_v\right) - \sum_{v=1}^{c}\left(\tfrac{y_x z_v}{M}\right)\delta\left(g, h_v\right)\right) \end{aligned} \tag{6.4}$$

Red nodes only use information about the blue nodes to update their labels (g) and similarly blue node labels (h) are updated only using information about the red nodes. Simplifying Equation 6.4 and creating an analogue for the updating rules for blue node labels leads to the following set of conditions:

Algorithm 6.1 LPAwb+ pseudo-code

```
Inputs : an incidence matrix
Output : row module labels, column module labels, modularity score

1    start
2
3        Find the smallest of the matrix dimensions and make these the red nodes
4        Initialise and randomly assign a unique label to each red node
5        Initialise the blue labels
6        run Stage1: Repeatedly update labels to locally maximise modularity
7
8        find the number of communities
9
10        while joining communities will result in increased modularity: {
11            run Stage2: Merge two communities that will increase modularity most
12            run Stage1: Repeatedly update labels to locally maximise modularity
13            find the number of communities
14        }
15
16        Assign red and blue labels to row and column labels (see line 3)
17
18    return row labels, column labels and modularity
```

$$\begin{cases} g_x^{new} = \arg\max_g \left(N_{xg} - \frac{y_x Z_g}{M} \right) \\ h_x^{new} = \arg\max_h \left(N_{xh} - \frac{Y_h z_x}{M} \right) \end{cases} \tag{6.5}$$

where the new label assigned to node x of type g (red) or h (blue) is that which maximises g or h on the right-hand side (if more than one solution exists, one is chosen at random). Here N_{xg} is the number of nodes connecting to x labelled g, while Z_g is the sum of blue node degrees labelled g and Y_h is the sum of red node degrees labelled h. As these 'bottom-up' updating rules (Equation 6.5) are mutually exclusive of one another they are applied asynchronously such that blue labels are updated, then red nodes are updated, then blue nodes are updated and so on until modularity (Equation 6.2) can no longer be increased.

Stage 2 - agglomeration stage - top down

When modularity can no longer be increased via stage 1's 'bottom-up' steps, a localised maximum of modularity for the network is reached, however this may not be the global maximum. The second stage seeks to prevent the algorithm getting stuck at local maxima by merging groups of communities together. Each identified community t is composed of blue and red nodes that share the same label i.e. when $g_u = h_v$. If there are F communities in total, then the merging of two different communities t_i and t_j can only occur if this would result in an increase in network modularity and

if there is no third community t_k ($1 \leq k \leq F$, $i \neq j \neq k$) whose merger with t_i or t_j would result in a larger increase to modularity.

Once this merger of communities is completed, stages 1 and then stage 2 are repetitively performed until it is no longer possible to increase network modularity by merging any of the possible communities together. These modules (communities) and the modularity of this partition are the solution provided by the LPAwb+ algorithm.

Exhaustive LPAwb+

Exploratory research with QuanBiMo and LPAwb+ revealed LPAwb+ often got stuck in a suboptimal solution with a larger number of modules, when compared with QuanBiMo, as LPAwb+ starts by identifying the largest possible number of modules, then iteratively merges them until modularity cannot be increased.

Knowing that LPAwb+ is sensitive to node label initialisation [49] and that it performs faster than QuanBiMo I designed a new algorithm, Exhaustive LPAwb+ (see Algorithm 6.2). Exhaustive LPAwb+ computes LPAwb+ multiple times with different random initialisations of node labels chosen from μ unique possible labels; and returns the solution which finds the greatest modularity score.

Algorithm 6.2 Pseudo-code for Exhaustive LPAwb+

```
Inputs : an incidence matrix, minimum number of modules, repititions
Output : row module labels, column module labels, modularity score

1    start
2
3        Sol1 = run LPAwb+
4        M = number of modules found in Sol1
5
6        for each value A from minimum number of modules up to M: {
7            for every repetition: {
8                Sol2 = run LPAwb+ with A initial modules
9                if Sol2 has greater modularity than Sol1:
10                    Sol1 = Sol2
11           }
12       }
13
14    return row labels, column labels and modularity from Sol1
```

Exhaustive LPAwb+ takes three inputs; the incidence matrix for the network of interest, the number of times that LPAwb+ should be run for each value of μ, and the

minimum number of unique labels (modules) to start running LPAwb+ with. Therefore μ ranges between this minimum value and the number of modules returned by a single execution of the LPAwb+ algorithm (when each node is initialised with a unique label) which is used as an upper limit.

Setting the minimum number of modules to search for small, and the number of repetitions high will increase the chance of detecting the global modularity optimum for a network; but is likely to be computationally costly. I chose to give Exhaustive LPAwb+ default settings of ten repetitions for each value of μ, starting from a minimum of four modules (note this does not preclude solutions with fewer modules being identified due to the merging process in LPAwb+) as the speed taken to perform these calculations appeared favourable to QuanBiMo for the test datasets.

6.2.3 Comparing Modularity

Normalised Modularity

The modularity values of Q_B and Q_W found above are network specific - properties such as the size and number of links in a network affect the magnitude of modularity that can be found [132, 47, 48]. In order to compare the strength of assortative mixing across different network studies it is necessary to account for the possibility of these effects. [48] recommend using a null model to generate an ensemble of networks from which the standardised effect size of modularity can be assessed as a z-score. However, it is unclear what would make an appropriate null model for weighted networks. An alternative method is to normalise the modularity values by the maximum value that modularity can take, found in the 'perfectly mixed' network, in which all edges are assigned to a module and there are no links between different modules [132]. Extending this for weighted bipartite networks gives:

$$Q^{max} = \frac{1}{M}\left(M - \sum_{u=1}^{r}\sum_{v=1}^{c}\frac{y_u z_v}{M}\right)\delta\left(g_u, h_v\right) \tag{6.6}$$

where as before M is the sum of the edges in the incidence matrix with marginal row totals, y and marginal column totals z. Then normalised modularity is found as:

$$Q^{norm} = \frac{Q}{Q^{max}} \tag{6.7}$$

Realised Modularity

Realised modularity [137] has been suggested as a posterior measure of modularity that classifies the proportion of links in a network that are within, rather than between modules. Here I extend this measure so it can be applied to weighted as well as binary networks. If M is the sum of all edge weights in a network and H is the sum of all within-module edge weights, then realised weighted modularity is expressed as:

$$Q_R^{'} = 2\left(\frac{H}{M}\right) - 1 \tag{6.8}$$

$Q_R^{'}$ takes values between -1, indicating that no edges exist between nodes in the same module, and 1, when all edges are interactions within-modules. If $Q_R^{'} = 0$ half of the edge weights in the network are found connecting nodes within the same module and the remaining edge weights are node connections between different modules. Note that in a weighted network $Q_R^{'}$ says nothing about the actual number of edges between or within modules, only the strength of the connecting edges.

Normalised Mutual Information

The normalised mutual information criterion is used as a way to compare the similarity of network structures found by different community detection methods [138, 47]. For two different partitions A and B of the same network with a total of n nodes (red and blue), with C_A and C_B modules respectively, the normalised mutual information is:

$$NMI(A;B) = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} N_{ij}\log\left(\frac{N_{ij}n}{N_i N_j}\right)}{\sum_{i=1}^{C_A} N_i \log\left(\frac{N_i}{n}\right) + \sum_{j=1}^{C_B} N_j \log\left(\frac{N_j}{n}\right)} \tag{6.9}$$

where N is the confusion matrix with elements N_{ij} which indicate the number of nodes that appear in the ith module of partition A and the jth module of partition B; N_i is the number of nodes in module i of partition A and N_j is the number of nodes in module j of partition B. If $NMI(A;B) = 0$ there is no shared information between partitions A and B - they each have identified very different community structures; whilst if $NMI(A;B) = 1$ the information given by partitions A and B is identical - the same community structure has been found by A and B.

6.2.4 Data

I used the 23 plant-pollinator networks available in the bipartite R package (22 of which were used in [48] and the additional junker2013 network) taken from the NCEAS dataset (https://www.nceas.ucsb.edu/interactionweb/resources.html). These networks show the number of observed visitations by each recorded pollinator species to each recorded plant species at different field sites across the world. Some network properties are shown in Table S2.

6.2.5 Computing Modularity

I computed the binary and quantitative networks for each of the datasets, removing rows and columns that contained no interaction data from the analysis. QuanBiMo, LPAwb+ and Exhaustive LPAwb+ were run 100 times for each binary and each weighted network in order to assess the modular structures found and the fidelity of the algorithms. I then quantified the differences between the modular structures found by the binary and weighted algorithms using the normalised mutual information criterion and investigated the differences in normalised and realised modularity.

Code implementations for the LPAwb+ algorithm are currently available online for the Julia, MATLAB/Octave and R programming languages. This and the R code used to create the figures and perform the analysis presented in this paper is available online (https://github.com/sjbeckett/weighted-modularity-LPAwbPLUS). For fair comparison in timing the algorithms all computations were performed in R version 3.1.1 using version 2.04 of the bipartite package on an Intel(R) Core(TM) i5-4570 CPU @ 3.20GHz desktop computer.

6.3 Results

Figure 6.2*a-b* shows the maximum modularity scores detected by each algorithm (from 100 replicates) for each of the networks. Full details are shown in Table 6.1 for binary networks and Table 6.2 for weighted networks. As expected (by definition) Exhaustive LPAwb+ scores were always equal or greater than those detected by

LPAwb+. Each algorithm detected similar maximum modularity scores for each network, with the exception of the datasets of kato1990, junker2013, barrett1987 and elberling 1999 in binary networks (Figure 6.2*a*) and kato1990, junker2013, elberling1999, kevan1970 and barrett1987 for weighted networks (Figure 6.2*b*) in which LPAwb+ and Exhaustive LPAwb+ detected much greater modularity scores than QuanBiMo.

Table 6.1 shows the greatest modularity scores detected by each algorithm, the number of modules in these partitions and the average execution time for each algorithm in the analysis of binary networks. The same partition was found by all three algorithms in only the schemske1978 network; both QuanBiMo and Exhaustive LPAwb+ found the same partitions for another 15 networks; whilst Exhaustive LPAwb+ found the greatest modularity score for 6 networks and QuanBiMo found the best modularity score in the inouye1988 network. LPAwb+ was by far the algorithm with the quickest execution time. Exhaustive LPAwb+ performs faster on small networks than QuanBiMo and more slowly on larger networks, however it generally found a much greater modularity score than QuanBiMo for these networks. The partitions found by LPAwb+ had more modules than those found by the solution with the greatest modularity.

For weighted networks Table 6.2 shows there were 5 networks for which the same maximum modularity was detected by all three algorithms, 10 networks in which QuanBiMo and Exhaustive LPAwb+ found the greatest modularity, 7 networks for which Exhaustive LPAwb+ found the greatest modularity and a single network, small1976, that was maximised by QuanBiMo. QuanBiMo had a similar average performance time to the binary networks, with LPAwb+ finding modularity more quickly in weighted than in binary networks. Exhaustive LPAwb+ has a similar performance time for smaller networks as under binary conditions and performs faster for the larger networks - which can be ascribed to the lower number of modules detected by LPAwb+ for the weighted networks. LPAwb+ detects partitions which generally have more modules than that with the greatest modularity, while QuanBiMo generally finds partitions with fewer modules than the solution found with greatest modularity.

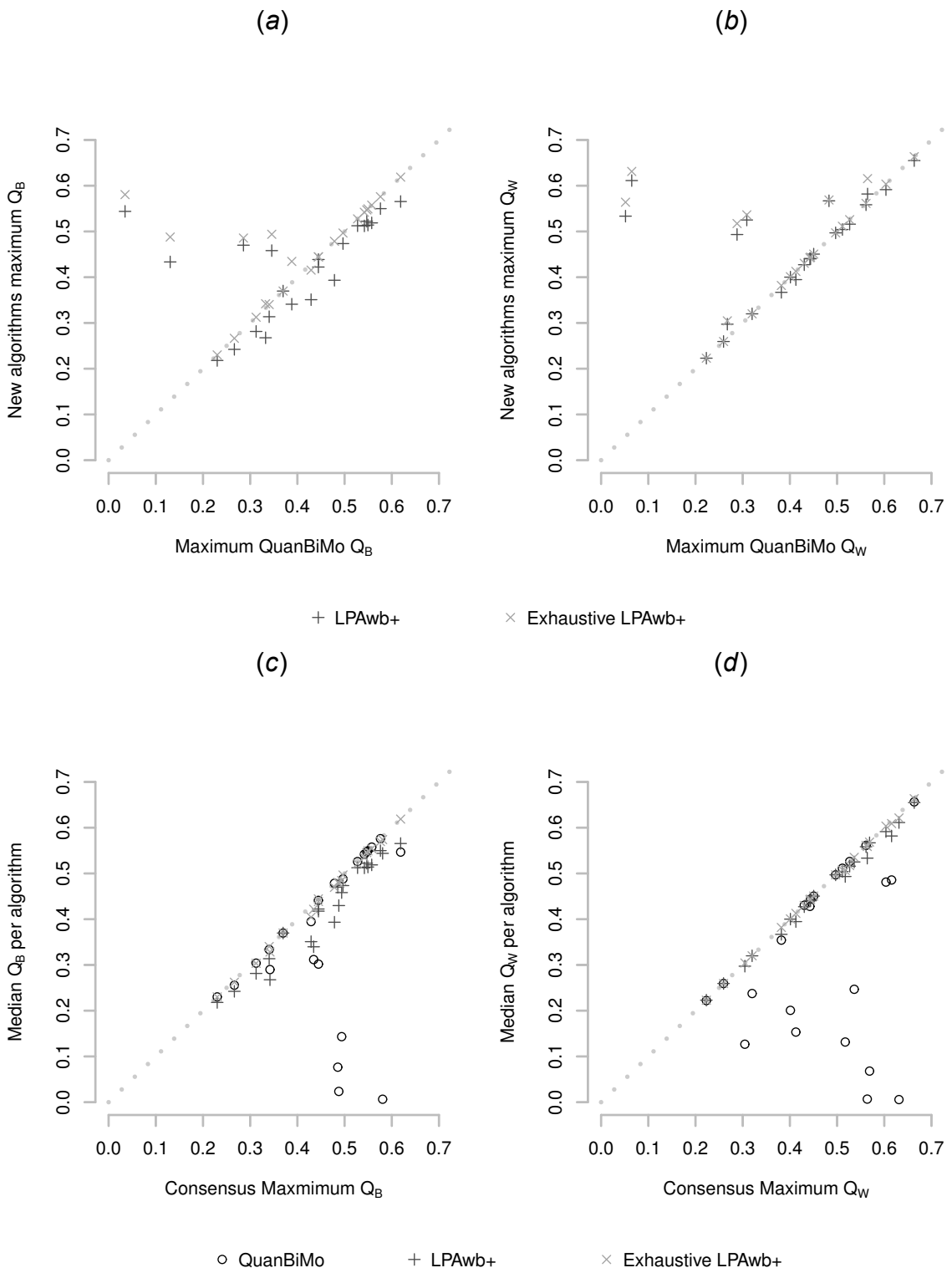


Figure 6.2: Comparisons of detected modularity scores by each algorithm (from 100 repetitions on each of the 23 plant-pollinator networks). (*a-b*) comparison of maximum detected scores for each algorithm. The dotted line indicates the performance of QuanBiMo. (*c-d*) comparison of median detected modularity scores to the maximum of the modularity scores found across the algorithms - the consensus maximum modularity. The dotted line represents algorithm efficacy, where median modularity score is equal to the maximum consensus modularity score that was detected. (*a,c*) shows a comparison of binary modularity scores,Q_B, whilst (*b,d*) shows the weighted modularity scores, Q_W.

Figure 6.2*c-d* shows the median detected modularity scores for each algorithm against the overall maximum modularity score for each network. Figure 6.2*c* shows that Exhaustive LPAwb+ consistently finds modularity scores closest to the maximal value, that LPAwb+ scores were close, but not so close and that whilst QuanBiMo could achieve consistency as good as the Exhaustive LPAwb+, for several networks QuanBiMo had a median value much lower than the maximum modularity detected. Similarly in Figure 6.2*d* Exhaustive LPAwb+ shows high consistency as does LPAwb+ (more so than for binary networks), whilst QuanBiMo in general performs less consistently for weighted networks than binary networks.

	QuanBiMo			LPAwb+			Exhaustive LPAwb+		
Network	Q_B	M	t	Q_B	M	t	Q_B	M	t
Safariland	**0.558**	6	1.067	0.519	9	0.014	**0.558**	6	0.641
barrett1987	0.286	4	11.811	0.470	11	0.070	**0.486**	8	3.667
bezerra2009	**0.230**	3	1.106	0.218	5	0.008	**0.230**	3	0.734
elberling1999	0.346	6	24.470	0.458	22	0.382	**0.494**	8	23.353
inouye1988	**0.429**	9	18.532	0.351	31	0.710	0.415	11	74.624
junker2013	0.130	5	46.076	0.433	55	2.762	**0.488**	19	405.596
kato1990	0.035	5	1551.827	0.544	74	14.196	**0.581**	20	3441.840
kevan1970	0.388	6	29.303	0.341	23	0.279	**0.434**	5	43.059
memmott1999	0.333	5	10.598	0.268	19	0.151	**0.342**	5	19.302
mosquin1967	**0.479**	6	0.916	0.393	11	0.014	**0.479**	6	0.819
motten1982	**0.313**	6	2.763	0.281	10	0.032	**0.313**	6	2.512
olesen2002aigrettes	**0.340**	4	1.149	0.314	7	0.011	**0.340**	4	1.254
olesen2002flores	**0.444**	4	0.949	0.422	7	0.008	**0.444**	4	0.533
ollerton2003	**0.445**	6	5.334	0.439	8	0.026	**0.445**	6	1.179
schemske1978	**0.370**	6	1.869	**0.370**	6	0.009	**0.370**	6	0.359
small1976	**0.266**	5	1.803	0.242	8	0.021	**0.266**	5	2.103
vazarr	**0.542**	7	1.431	0.512	9	0.016	**0.542**	7	0.865
vazcer	**0.619**	6	2.000	0.565	9	0.015	**0.619**	6	0.744
vazllao	**0.576**	6	1.129	0.550	8	0.016	**0.576**	6	0.915
vazmasc	**0.547**	6	1.340	0.522	8	0.011	**0.547**	6	0.486
vazmasnc	**0.527**	6	1.969	0.512	8	0.014	**0.527**	6	0.533
vazquec	**0.497**	4	1.529	0.474	7	0.013	**0.497**	4	0.479
vazquenc	**0.549**	5	0.834	0.514	7	0.009	**0.549**	5	0.300

Table 6.1: Comparison of QuanBiMo, LPAwb+ and Exhaustive LPAwb+ algorithms on binary ecological interaction networks. Q_B is the greatest value of binary modularity from 100 replicates on the network, M is the corresponding number of modules found in this partition and t is the mean time taken to compute each algorithm once. Numbers have been rounded to 3 d.p. Numbers shown in bold are those with the highest Q_B score.

Network	QuanBiMo Q_W	QuanBiMo M	QuanBiMo t	LPAwb+ Q_W	LPAwb+ M	LPAwb+ t	Exhaustive LPAwb+ Q_W	Exhaustive LPAwb+ M	Exhaustive LPAwb+ t
Safariland	**0.430**	5	1.258	0.427	7	0.014	**0.430**	5	0.721
barrett1987	0.483	5	10.107	0.567	9	0.057	**0.569**	7	3.577
bezerra2009	**0.223**	5	1.178	**0.223**	5	0.008	**0.223**	5	0.645
elberling1999	0.288	6	25.416	0.493	18	0.190	**0.517**	10	21.288
inouye1988	0.565	11	25.368	0.582	22	0.413	**0.615**	9	54.377
junker2013	0.052	5	83.548	0.533	33	1.133	**0.564**	17	287.774
kato1990	0.065	5	2355.046	0.611	48	6.000	**0.631**	23	2425.382
kevan1970	0.309	2	35.164	0.525	10	0.096	**0.536**	5	26.133
memmott1999	0.267	4	12.333	0.297	10	0.065	**0.305**	7	11.660
mosquin1967	**0.444**	6	0.970	0.440	7	0.009	**0.444**	6	0.669
motten1982	**0.382**	4	3.292	0.367	6	0.020	**0.382**	4	1.902
olesen2002aigrettes	**0.259**	5	1.181	**0.259**	5	0.008	**0.259**	5	0.905
olesen2002flores	**0.497**	5	0.989	**0.497**	5	0.006	**0.497**	5	0.415
ollerton2003	**0.413**	6	6.023	0.395	7	0.024	**0.413**	6	1.243
schemske1978	**0.320**	4	1.792	**0.320**	4	0.009	**0.320**	4	0.392
small1976	**0.527**	8	1.984	0.516	11	0.026	0.526	9	1.909
vazarr	**0.442**	6	1.733	0.441	7	0.014	**0.442**	6	0.883
vazcer	**0.604**	6	2.317	0.591	7	0.015	**0.604**	6	0.725
vazllao	**0.561**	6	1.386	0.558	8	0.013	**0.561**	6	0.839
vazmasc	**0.663**	6	1.436	0.655	7	0.010	**0.663**	6	0.456
vazmasnc	**0.401**	6	2.291	0.400	7	0.012	**0.401**	6	0.565
vazquec	**0.511**	6	1.835	0.504	7	0.013	**0.511**	6	0.474
vazquenc	**0.450**	4	0.815	**0.450**	4	0.007	**0.450**	4	0.265

Table 6.2: Comparison of QuanBiMo, LPAwb+ and Exhaustive LPAwb+ algorithms on weighted ecological interaction networks. Q_W is the greatest value of weighted modularity from 100 replicates on the network, M is the corresponding number of modules found in this partition and t is the mean time taken to compute each algorithm once. Numbers have been rounded to 3 d.p. Numbers shown in bold are those with the highest Q_W score.

The average time to run each algorithm is shown in Figure 6.3. Performance time is network dependent; where it takes longer to report modularity for larger networks. LPAwb+ performed quickest on all networks by roughly 2 orders of magnitude. Performance on the binary (Figure 6.3*a*) and quantitative (Figure 6.3*b*) network representations was similar. However, QuanBiMo performed faster for binary (rather than quantitative) inputs on 20 of the 23 networks. On the other hand, LPAwb+ ran quicker with quantitative network representations (20 out of 23), as did Exhaustive LPAwb+ (18 out of 23). For the ten cases where Exhaustive LPAwb+ took longer than QuanBiMo, Exhaustive LPAwb+ found a partition with greater modularity seven times, both

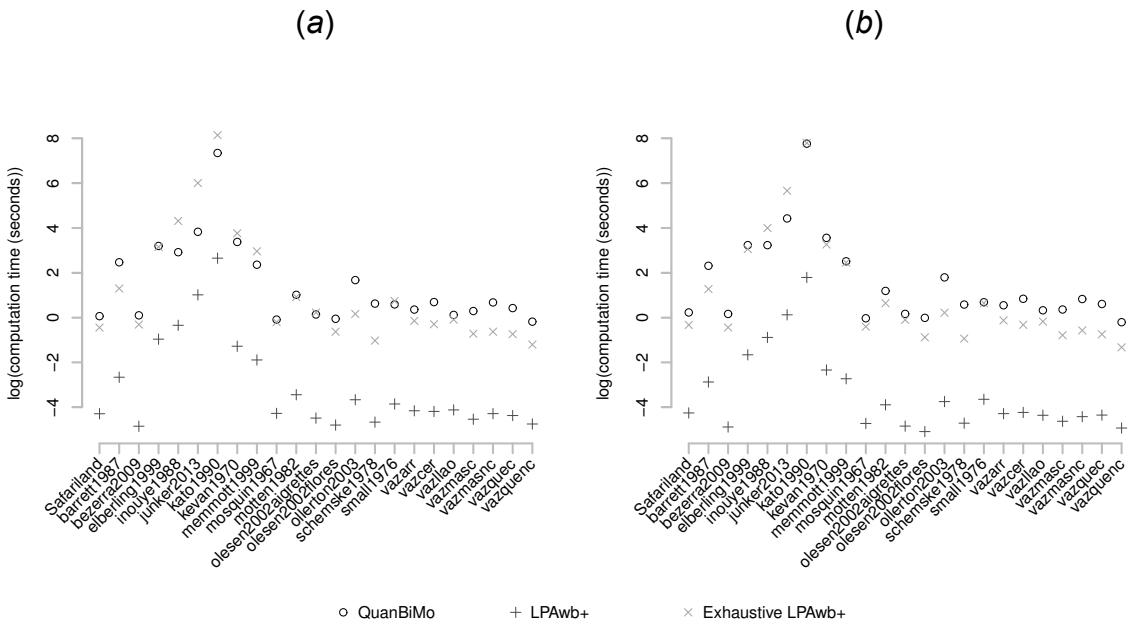


Figure 6.3: Average computational time for each algorithm (measured over 100 replicates) on the (*a*) binary and (*b*) quantitative representations of each plant-pollinator network.

QuanBiMo and Exhaustive LPAwb+ found greatest modularity twice and QuanBiMo found the greatest modularity score once (the binary representation of inouye1988).

6.3.1 Contrasting Binary and Quantitative Modular Structure

Maximising binary modularity and maximising weighted modularity results in different identified modular structures. Figure 6.4*a* shows the partition with the greatest binary modularity for the olesen2002flores network, whilst Figure 6.4*b* shows the partition with the greatest weighted modularity. The same dataset has qualitatively different structure between its weighted and binary representations. The shared normalised mutual information for these two partitions is $NMI = 0.619$, quantifying this difference.

Figure 6.5*a* shows the differences in normalised modularity and normalised mutual information between the binary and weighted network representations. Only 3 of the networks (vazquenc, vazmasnc and vazcer) have a normalised mutual information greater than 0.8 - indicating major differences in identified binary and quantitative modular structures. The strength of assortative mixing, measured by normalised modularity, was generally greater in weighted than binary networks. However, 4 net-

works (olesen2002aigrettes, vazarr, bezerra2009, vazmasnc) showed greater assortative mixing in their binary representations and for 2 networks (olesen2002flores, vazllao) the assortative mixing strength was nearly the same in both binary and weighted networks - though the community partitions are very different.

Not only were the detected modularity scores different between the binary and weighted networks - but the number of modules found in each partition of these networks also differed. Only 8 of the networks had the same number of modules under binary and weighted conditions; whilst 8 had more modules in the weighted networks and 7 had more modules in the binary network representation (Table 6.1, Table 6.2).

There appears to be a weak positive relationship between realised modularity and modularity (Figure 6.6*a*), however normalised and realised modularity appear to be much more strongly correlated (Figure6.6*b*). There does not appear to be a relationship between the binary and quantitative measures for each network.

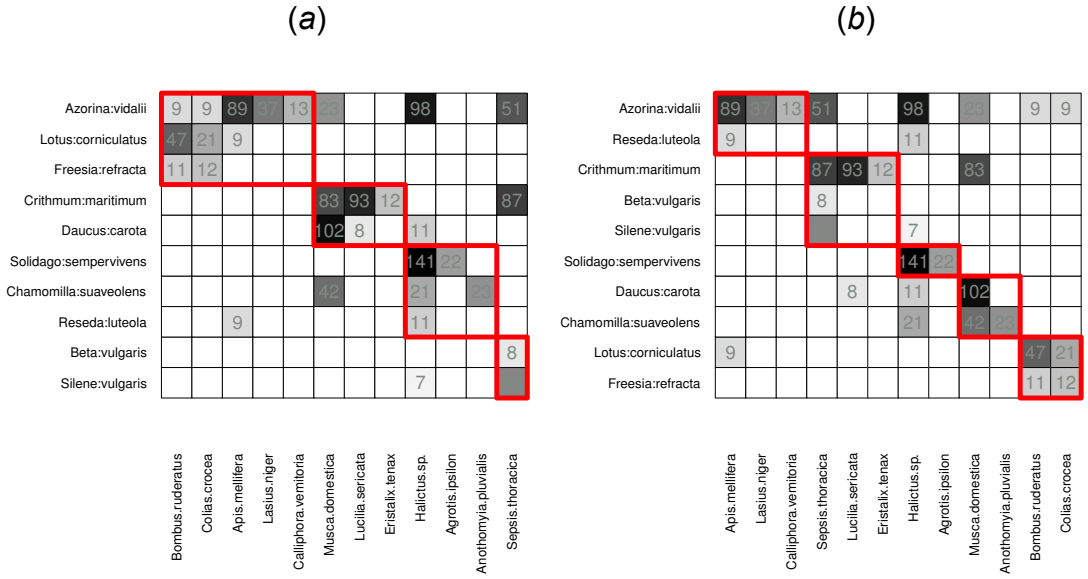


Figure 6.4: A visual comparison of the modular structures identified for the olesen2002flores dataset of plant-pollinator visitations as a (*a*) binary ($Q_B = 0.444$, 4 modules, $Q_B^{norm} = 0.625$) and (*b*) quantitative ($Q_w = 0.497$, 5 modules, $Q_W^{norm} = 0.625$) network. Modules are identified in red. The normalised mutual information shared between these two modular compositions is $NMI = 0.619$ indicating a qualitative difference in the revealed modular structure.

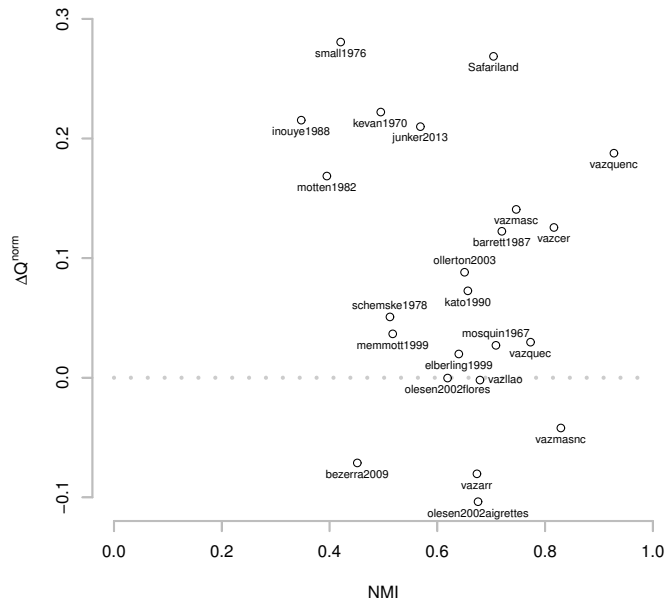


Figure 6.5: The change in normalised modularity scores found between the weighted and binary networks ($\Delta Q^{norm} = Q_W^{norm} - Q_B^{norm}$) against the normalised mutual information between the weighted and binary partitions for each network.

(*a*) (*b*)

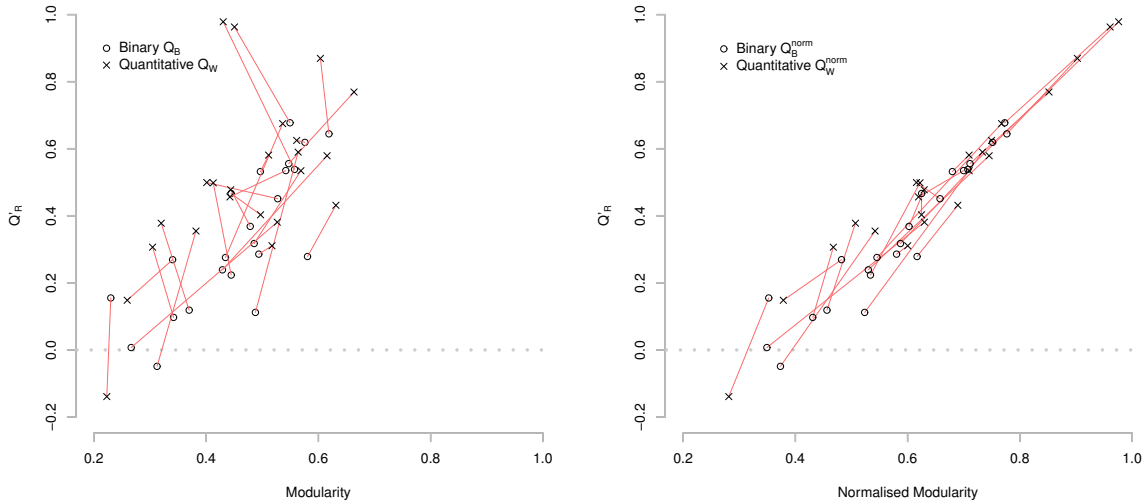


Figure 6.6: (*a*) The greatest modularity scores (Q_B and Q_W) for each network and their corresponding realised modularity scores (Q'_R). (*b*) The normalised modularity scores (Q_B^{norm} and Q_w^{norm}) calculated using the partitions with greatest modularity scores plotted against their corresponding realised modularity scores. Each red line joins together the binary and quantitative scores of the same network.

6.4 Discussion

I tested the efficacy of three algorithms maximising Dormann's bipartite modularity in plant-pollinator networks. LPAwb+ and Exhaustive LPAwb+ gave more consistent modularity scores than those found by QuanBiMo across the test networks. The robustness of modularity maximisation algorithms is important when considering the reproducibility of results. QuanBiMo struggled to report "good" modularity scores in the larger datasets. All three algorithms were able to detect greater modularity than previously reported (Figure 6 in [48]) and were generally performed well on the binary and quantitative test networks (a binary network can be seen as a special case of a quantitative network). But, QuanBiMo has the potential to fall into below par solutions and there is no diagnostic to show when this occurs.

Different modular structures were found for each of the binary and weighted representations of plant-pollinator networks. In binary networks modules are formed by attempting the maximise the density of edges; whilst in quantitative networks modules are formed that maximise the density of edge weights. In the former, strongly interacting nodes are just as important as nodes that only rarely interact; whilst in the latter modules are likely to form around the strongest node-node interactions.

Normalised modularity measures the strength of assortative mixing and is a useful network index that can be used as a comparison indicator across different network studies. Modularity by itself is often used as a network indicator - but this is not appropriate when comparing different networks whose theoretical modularity maximums may differ. I find normalised modularity is strongly correlated with the proportion of within module interactions (realised modularity) which is an intuitive way for understanding modularity.

LPAwb+ was not able to maximise modularity so well as Exhaustive LPAwb+ or QuanBiMo on the majority of datasets (though the modularity found was near the maximal value found here), but its fast performance makes it an ideal algorithm for exploratory research and for investigating modularity in larger networks, where parallelisation of the algorithm [49] may become useful.

There is no guarantee that the greatest possible modularity was found in any of the test networks here; indeed maximising bipartite modularity is an NP-hard problem [139] and it may be difficult to find an algorithm which performs well on this problem for any possible network.

The QuanBiMo algorithm takes two input values; the number of algorithmic steps that should be performed to attempt to find greater modularity than the current partitions modularity; and the tolerance threshold for greater modularity scores. Clearly the default values were not appropriate for some of the networks assessed here; where much greater modularity was detected by the new algorithms. However, there is no diagnostic to tell that QuanBiMo has returned a sub-par modularity value without comparisons (which may be a lengthy process); or what suitable input parameters may be for a particular network. There is a strong tradeoff between computational effort and the accuracy of the returned modularity. On the other hand LPAwb+ takes no input parameters and was able to quickly find modularity scores near to the consensus maximum modularity. Exhaustive LPAwb+ has two input parameters; the minimum number of modules to search for and the number of times that LPAwb+ should be initialised for each module number. Unlike QuanBiMo, these parameters have physical meaning in the context of the network – and the time complexity of this algorithm can be estimated from the number of calls that will be made to the LPAwb+ algorithm (as LPAb+'s time complexity is known [49]).

There are four challenges to address when attempting to maximise modularity [50] which are also relevant to weighted modularity. Any modularity maximisation algorithm only uses information within the incidence matrix and is thus agnostic to hierarchies within the dataset - the algorithm will find communities at the resolution that has the greatest modularity it can compute; which may be different to the resolution which corresponds best with any additional information known about the network. This is further complicated as several hierarchical levels may exist within an individual network. Some work has started to address this problem in terms of visualising the network as a multiscale structure [80, 54, 48], but this requires finding a suitable starting resolution. As found with QuanBiMo, the ability of algorithms to maximise modularity can be highly dependent on network properties such as size. Finally it is recognised that the modularity landscape is "glassy" - there are many local modularity maxima; but detecting the global peak is extremely difficult and finding an algorithm that can capably traverse this "glassy" landscape is a challenge.

A further challenge will be to find appropriate null models to test weighted modularity against in order to standardise the effect size of modularity in different networks [48]. In principle it would be good to test against a null ensemble in which both the allowed interactions and the strength of these interactions are allowed to vary. However, in this paper I have only focussed on the optimisation of weighted modularity.

Another limitation of the weighted modularity definition explored here is that it is only

valid on networks where all connections are positive. However, methods have been created to search for modules in weighted networks with positive and negative link strengths in unipartite networks that could easily be extended for bipartite networks [140].

I focussed on a specific definition of modularity in this paper - but note that others do exist [45, 46]. Thébault [47] compared two binary bipartite modularity based measures that have been applied in ecology and concluded that different forms of modularity may be useful in different contexts; but that the form of modularity used here [44, 48] corresponded well with that for unipartite networks [41, 133] - and is well suited for identifying densely connected modules. Other modularity measures [45, 46] do not identify joint communities made of both types of nodes – but rather identify communities within each type of node, though neither of these approaches has yet been extended to weighted networks to my knowledge.

The major advantage in a definition of weighted modularity is that it allows for much more information about a network to be used to detect communities. Both binary and weighted measurements contain different information about a network and may be useful - though I expect weighted measurements may in general contain more relevance for the analysis of real world networks – the strength of interactions is undoubtedly an important component of network structure. Other modularity definitions and their weighted extensions are also in need of further investigation to consider communities within each type of node and how these may overlap with the joint communities considered here.

6.5 Conclusions

Real world networks are not formed of binary interactions. I encourage researchers to apply weighted modularity measures to their datasets and evaluate the community partitions that are identified.

LPAwb+ is an algorithm that would be well suited for exploratory analysis and use on large networks - as it is fast and, whilst it did not return the best modularity values of the methods tested here, the solutions it did find were consistently high. Care has to be taken with both QuanBiMo and Exhaustive LPAwb+ in setting appropriate input parameter settings such that the analysis is not computationally infeasible. I would

recommend using Exhaustive LPAwb+ over QuanBiMo; as Exhaustive LPAwb+ has more meaningful input parameters, can perform no worse than LPAwb+ and its performance was less variable than QuanBiMo on the networks tested in this study.

I have made the code for the LPAwb+ and Exhaustive LPAwb+ algorithms; as well as the analysis performed in this paper available online (https://github.com/sjbeckett/weighted-modularity-LPAwbPLUS) to allow researchers to replicate my findings and encourage those with access to potentially interesting weighted bipartite datasets to analyse them using these methods.

Chapter 7

Discussion

7.1 Overview

In this chapter I bring together some of the central themes and discussions from the previous chapters and discuss potential future directions for studying nestedness and modularity in bipartite networks. First I will give a summary of the previous chapters (Section 7.2) and some of the other projects I have been involved in during my thesis (Section 7.3). Next, I provide considerations for releasing scientific software based on my experiences with FALCON and LPAwb+ (Section 7.4). Following this, I discuss the implications of knowledge gained in this thesis to the study of nestedness and modularity (Section 7.5). Then further points are raised considering best practices and future directions for nestedness (Section 7.6) and modularity (Section 7.7) analysis. Finally, I offer general considerations and recommendations for examining structural patterns in bipartite networks (Section 7.8).

7.2 Chapter summary

To briefly summarise the previous chapters:

The concepts of nestedness and modularity in bipartite networks were introduced in Chapter 1 as well as some of the considerations required to evaluate these quantities. Specific details for calculating the nestedness measures used in this thesis as well as the null models that have been used to evaluate them are described in Chapter 2.

In Chapter 3 nestedness and modularity were used as indicators of community composition in an coevolutionary ecological model of phage and bacteria. By creating timeseries of these indicators it was possible to use nestedness and modularity as signals of how community composition evolved. Analysis showed that nested-modular structures such as those observed in natural phage-bacteria systems can arise from the simple coevolutionary rules of the lock-and-key model.

Chapter 4 is a description of the software package FALCON that originated from the analysis of phage-bacterial infection networks described in Chapter 3. FALCON is software designed to encourage and assist nestedness analysis - by allowing users to compare nestedness statistics calculated by multiple nestedness measures using multiple null models. The concept of adaptive null ensembles was introduced to internally test null ensembles for robustness and encourage robust reporting of statistical tests.

Using FALCON, a dataset of 30,000 synthetic bipartite networks was analysed for nestedness in Chapter 5. These networks were created along a gradient between nested and random configurations and have a range of different sizes and connectances. By computing nestedness on each of these networks, using the measures and null models in FALCON, it was possible to make comparisons between the outcomes of different measure, null model and effect size combinations. This enabled recommendations to be made for analysing nestedness in bipartite networks.

In Chapter 6 attention was turned from nestedness to modularity. A new algorithm for maximising modularity in weighted bipartite networks was introduced. A comparison between three algorithms: QuanBiMo, LPAwb+ and Exhaustive LPAwb+ was performed using 23 networks describing the mutualistic interactions between plant species and the pollinators that visited them. Weighted networks may have considerably different modules to those identified in the binary configurations of the same network.

7.3 Additional research projects

There are a number of projects that I undertook during my studies, but are not presented here. Below are three projects that have been completed. Some are related directly to concepts discussed in this thesis, but I shall start with the one which is not.

Microplastic ingestion by crabs:

Microplastics (plastic particles <5mm in length) constitute a recently recognised form of pollutant, prevalent in the oceans, that are of a size that makes them available to a variety of marine biota. They are a product of plastic degradation, but also have several direct anthropogenic sources such as microbeads in cosmetic products. Understanding if and how microplastics may effect the ecology and functioning of individual species and marine ecosystems is an active research area in marine ecology [141]. Using a species of shore crab, Watts et al. [142] demonstrate that microplastics can be retained in the crabs foregut for over 14 days after ingesting food laden with microplastics, or in the gills for over 21 days after exposure to water containing microplastics. I built simple mathematical models to describe microplastics passing through the crabs gills and guts, which were parameterised using data collected during the experiments [142]. Although it was not possible to fully parameterise the

model, it was shown to exhibit qualitatively similar behaviour to that observed during the experiments. This model provides a conceptual basis for the direction of future studies, both in terms of a mechanistic description of the passage and retention of microplastics and the kinds of data that may be required to better understand this process - and ultimately its effects.

Clinical phage typing:

Escherichia coli O157:H7 infections have adverse effects on human health worldwide. It is estimated a third of reported cases of this foodborne pathogen lead to hospitalisations in the UK. Different strains of *Escherichia coli O157:H7* are associated with different at-risk demographics and levels of virulence. Cowley et al. [143] seek to gain a better understanding of these differences by performing genomic analysis on 16 phages commonly used to classify *Escherichia coli O157:H7* strains through analysis of a who-infects-whom profile. My contribution to this work was the analysis of the resulting phage-bacteria infection network for nestedness and modularity. This network exhibited a large number of cross infections between strains and showed both significant nestedness and modularity. Modularity was shown to be associated with genetic similarity, while nestedness supports the choice of this set of 16 phages to distinguish between the bacterial strains.

Networks of climate change discussion:

Social media platforms, such as Twitter, offer a global forum for communication, dissemination and discussion of a diverse range of topics. Learning how attention is focussed on certain topics, and the form of information transfer, may be useful for designing communication strategies. Williams et al. (in prep. [122]) studied Twitter data related to climate-change related topic hashtags surrounding the release of the three working group contributions that constituted the Intergovernmental Panel on Climate Change's Fifth Assessment Report. Attention for each contribution initially focussed on the released IPCC report, but this was quickly superseded by secondary sources including news media, commentary and campaigning. Analysing bipartite networks of hashtag-domain associations (formed by finding which hashtags and web domains appear in the same tweet) a nested architecture was revealed. Some hashtags were covered by many domains, whilst others were more domain specific indicating variation in the breadth of climate change coverage; and that there was also variation in

the depth of climate change coverage offered by different domains (number of hashtags covered). Additionally by comparing these networks through time it was found that hashtag-domain networks appeared more nested during the period around the release of each working group contribution.

7.4 Software and code availability

Two pieces of software have been created and made available as a result of work in this thesis. FALCON [129] was released to enable and encourage comparison of different nestedness indicators and null models, whilst LPAwb+ was released as an algorithm to improve the detection of modular communities in weighted bipartite networks. Both FALCON (in [143, 144]) and LPAwb+ (in [143]) have already been used in publications to analyse scientific datasets.

Here I outline some general recommendations from my experiences in making these works available. Both FALCON and LPAwb+ were made available in multiple programming languages. There is no single pervasive coding language used by the research community (either in general and by those interested in modularity and nestedness in bipartite networks - see Table 1.3). On a personal note, I have had formal introductions to six programming languages during my postgraduate studies and have assisted in teaching introductory courses to biology undergraduates using MATLAB, Python and R. Releasing code in multiple languages may increase the number of potential users who can use, interact, extend and be inspired by the produced methodology, though it could take substantial time to perform these translations. When thinking of potential users, the availability of software is also of concern. FALCON was originally developed in MATLAB, but we chose to translate it into R as, unlike MATLAB, it can be run for free.

It is also worth discussing how research software is made available to potential users. Writing code can be a messy process, as it is done for the primary benefit of the programmer, but releasing it could be beneficial. Most scientists are not trained as software engineers, but software programming is a core skill behind an increasing number of scientific advances, whether they be theory or data driven. Working messy code is preferable to no code [145]. The incentive to release code may indeed improve the quality of code being released [146]; and an increasing amount of advice is being offered for the best practises in releasing software (e.g. [147]). Releasing software as an executable file alone may not be good enough - especially as these

may be operating system dependent. Having access to the source code is extremely useful for checking and reproducing methodology; and for extending or embedding these methods for use in other applications. If methods used by a software application are not clearly described or easily reproducible, the software could be treated as a “blackbox” - as the assumptions and limitations of the methods are not clear [148] - which is one of the reasons so many temperature-based measures of nestedness [22, 23, 24, 25, 26] exist. The release of the source code itself allows the program to be read and understood, even if a user does not have the necessary software to run it.

On a similar theme, sometimes the statement: “code is available on request from the authors” is attached to the end of a research paper. I would argue that this isn’t good enough. If the code is not sufficiently archived by the authors, it could be lost forever. If the original authors change or leave their institution, it cannot be assumed they will take their code with them. The long term must also be considered, as papers can continue to be cited decades or more after they were originally published. Eventually the authors will no longer be around to provide code on request. Similar arguments can be made against those making software available on a private or institutional website - which may not be supported in the long term. Putting software into a public data repository or making it available with a paper where it is assigned a digital object identifier (DOI) and supported by an arrangement such as CLOCKSS (Controlled Lots of Copies Keep Stuff Safe) [149] would appear to be a good solution. Referring to an object’s DOI is more stable than linking to the website address as it is a persistent identifier to the online material: if the content hosting organisation changes the website address, the same DOI will be redirected to this new address. Additionally CLOCKSS support ensures that even if the content hosting organisation goes out of business, the documents it was hosting will be rehosted elsewhere using resources of other CLOCKSS members. Again the original DOI will redirect to the new website address. These systems will be useful for providing long term stable access to software.

Academic journals can do more to ensure scientific software is released alongside publications in a responsible manner, such that it is accessible with long term support. Additionally further recognition that software may be improved upon after a publication is released, and is therefore not necessarily a static object is required. For example, FALCON could be extended to include other nestedness measures and null models, including methods not available at the time of original publication. Publishing teams should be aware of the issue of such ‘living software’ and create suitable publication routes that allow authors to update publications. This may be

a complex issue as the number of contributing authors, in addition to the software content, may change through time.

7.5 Implications for interpreting nestedness and modularity in ecological systems

7.5.1 Comparing bipartite indicators in different networks

Once it is possible to generate a network metric, such as nestedness or modularity, a particular application is to see how these metrics vary between similar networks. In Chapter 3, for example, we were interested in how nestedness and modularity of the communities of phage and bacteria evolved through time. Macroecological studies have focussed on how these network properties differ through time and space [6, 150]. However, in Chapter 5 it was shown that nestedness measures were sensitive to network properties, and that standardisation of nestedness metrics through use of a null model did not resolve this issue completely. This may be due to the number of possible network permutations being related to network properties such as connectance [120, 126] and also methodological problems with calculating nestedness e.g. NODF can be extremely sensitive to small changes in network structure [151]. In Chapter 5, we showed empirically that the best nestedness indicator for comparisons across different networks is the discrepancy measure (BR) combined with the adjusted normalised temperature. This recommendation is used in analysis of Twitter hashtag-domain networks relating to climate change [122]. It is not clear why this combination works so much better than other combinations of measures and effect sizes across all the null models that were tested. An improved understanding of this topic may lead to the design of better nestedness measure, null model and effect size combinations and is an open topic for future research.

7.5.2 Bipartite network patterns and ecosystem stability

Several recent high profile papers on nestedness (e.g. [8, 9, 101, 152]) have focussed on the structure-stability debate - the potential of network structure to determine the stability of a network. Both nested [8, 9] and modular [153] network topological structures are thought to have implications for network stability. The literature

has focussed on the implications of nestedness being detected in plant-pollinator networks and how this relates to stability [101, 152, 14]. However, plant-pollinator systems are not only nested, but also modular [2] (also see Chapter 6). Additionally (and maybe surprisingly [154]) antagonistic relationships between phages and bacteria are also found to exhibit modular and nested architectures [94, 154]. In the eco-evolutionary phage-bacteria model in Chapter 3 the analysis of infection networks revealed structure that was both nested and modular at the size of the network, with additional nested structures within the modules. Focussing on just nestedness or just modularity as driving conditions for stability may be misleading. Though modularity and nestedness have been shown to be negatively correlated [114], both of these network properties working in tandem may be important to network stability as both patterns are identifiable in both mutualistic and antagonistic systems.

In mutualistic networks some work has placed connectance as a major driver of both nestedness and modularity [9]. James et al. [101] suggest that the size of the network and connectance are better predictors of persistence than nestedness (however, there are concerns about the assumptions of the model used [152]). Additionally the degree-distribution has been implicated as an important factor [37] - though this itself is driven by network connectance [120]. It is not clear that nestedness is the key driver of ecological stability. Furthermore, work in Chapter 5 suggests the methodology chosen to assess nestedness may alter the conclusions drawn. Finding methodology that best assesses nestedness is a primary goal to support this interesting theory driven work.

Moving towards understanding network patterns in light of quantitative information is also a welcome direction. Indeed recent work suggests edge weight heterogeneity may be most important [155] with nestedness as an additional factor. When considering stability it is also important to question how these bipartite networks fit within the larger ecosystems they are a part of - it is suggested that interaction strength correlations drive stability at the large scale [156]. More work in these directions is encouraged. The collection of quantitative data to characterise both the physiology of individual species as well as species interactions may lead to more integrated insights between interaction networks, life history traits and stability.

7.5.3 Application to phage-bacteria networks

In the classic “kill-the-winner” model of phage-bacteria ecology coexistence between types is controlled by top-down selection on bacteria, such that the abundances of

the fastest growing bacterial strains are regulated by higher levels of phage infection and lysis [74]. Each bacterial strain is infected by a unique phage strain – the infection strategy is specialised. Theoretical work on phage-bacterial infection networks has begun to investigate multistrain communities with overlapping host ranges - principally looking at nested architectures [157, 93, 158]. Coexistence is sustained in these communities by the introduction of tradeoffs. A cost of resistance tradeoff between host growth rate and the hosts ability to defend from phage predation [157] as well as tradeoffs between virus infectivity and host range [93] are required for stability [159] when the infection network is nested. However, phage-bacterial communities are not perfectly nested [3] and these types of models only consider ecological stability in the context of a given infection pattern - they do not explicitly consider how such patterns are generated in these bipartite networks.

Finding the conceptual rules that allow for the generating and sustaining of diverse communities of phages and their bacterial hosts is a core challenge. The lock-and-key model in Chapter 3 that is conceptually grounded in phage tail fibres fitting bacterial surface receptors is a step towards this goal. The lock-and-key model produced nested-modular networks similar to that observed in the Atlantic Ocean [82, 80] and shows that simple coevolutionary rules allow for the generation of complex network structures. Whilst theoretical arguments suggest adapting the “kill-the-winner” model to focus on strain, rather than species specific interactions will lead to nested-modular architectures [160] more coevolutionary driven theory needs to be developed. Being able to relate model rules back to empirical datasets will be a useful way towards understanding the evolutionary history (and potential trajectories) of interaction networks. As such comparing the structure of bipartite networks generated in models with those from empirical data is a useful way of linking theory to data.

Despite their small size, the interplay of phage and bacteria can be complex [161]. Ecological and evolutionary dynamics occur at similar timescales and a variety of different biological strategies exist to exploit one another [86, 98]. For instance, phage can adopt a temperate lifestyle strategy that allows them to integrate their genes with those of their hosts as a prophage - the prophage lies dormant within the host until some later time, perhaps due to environmental stress, when the lytic lifestyle is again triggered: causing the bacterial cell to create new phages that eventually burst free in a lysis event. On the other hand, there is evidence that some bacteria can adopt a suicidal strategy on recognising a phage infection event that may protect the larger population. There are many future directions for theoretical phage-bacteria community dynamics. Some suggestions for future research include investigating community dynamics in situations where a variety of different strategies can be adopted,

incorporating events such as horizontal gene transfer and analysing spatial effects. It is likely that the type of mechanisms acting in communities will effect both the dynamics and the structures displayed in these networks [162].

Data for studying phage-bacterial infection networks typically comes from infection assays. Bacterial and phage strains are first isolated, each bacteria-phage pairing is assigned a testing area (typically a well or plate) where the bacteria is allowed to grow in media, before the phage is added. If a plaque forms on the bacterial lawn (areas left clear where bacterial cells are destroyed) it is assumed bacterial lysis has occurred via phage infection and burst. If a plaque forms it is assumed that that bacteria-phage pair interact, else they do not. As highlighted in Chapter 3, there are multiple potential problems with this methodology e.g. it may be difficult to isolate all strains (especially low abundance strains), some strains may be unculturable [163, 91, 92], it may not account for lysogenic phage-bacterial interactions. Advances in single-celled and genomics approaches (reviewed in [164, 165]) may provide useful advances in finding phage-bacterial infection networks in environmental samples and metagenome datasets. Additionally fluorescent gene marking methods (such as phageFISH [166]) may allow identification between, lytic, chronic and lysogenic infection cycles [164]. These approaches could yield data that helps to better understand and to model phage-bacterial systems, at both small (clinical) and large (environmental) scales.

Better ways of measuring the quantitative parameters of viral infection are also much needed [93]. In the context of phage-bacterial infection networks these may be more important when trying to understand geographically localised strains i.e. trying to understand how colocalised phage and bacteria pairs interact [154], than in a clinical context where phage-bacterial datasets may not be colocalised and the purpose may be more about discriminating between different phage or bacterial strains [143]. However, understanding the quantitative effects of multi-phage and multi-bacteria communities in situ will become especially important in determining, for example, the appropriate phage ingredients in producing “cocktails” for use in phage-therapy (e.g. to treat bacterial infections), or how use of antibiotics may alter microbiome diversity [167]. Appropriate methods for approaching these quantitative bipartite networks are therefore needed.

7.6 Recommendations for improving nestedness analysis

7.6.1 Best practice

The strength of nestedness analysis is dependent on the intertwining of measure, null model and effect size choices. In Chapter 5 we show that the swappable-swappable [35] (SS) and cored-cored [129] (CC) null models are poor at discriminating nested from non-nested patterns, with high type I errors (falsely accepting nestedness) across all tested nestedness measures. By contrast, the fixed-fixed [58] FF null model exhibited high type II errors (falsely rejecting nestedness) for all measures other than discrepancy [21] (BR) and spectral radius [35] (SR). If the FF null model is used, it is recommended that this is done in conjunction with BR. The degreeprobable-degreeprobable [1] (DD) null model offered more discrimination of the less nested networks than the equiprobable-equiprobable (EE) null model, where SR with the DD null model demonstrated the best tradeoff between type I and type II errors. For considering the strength of nestedness within a single network, z-scores offer a good statistic regardless of measure choice. However, z-scores are not directly comparable between different nestedness measures. Additionally, if comparing the strength of nestedness across different networks is considered, we show that using BR with the adjusted normalised temperature is the best procedure for all the tested null models. To summarise:

- For single network assessments using SR with the DD null model and z-scores is encouraged.
- If it is necessary to use the FF null model, then the BR measure should be used.
- If comparing multiple networks for nestedness, do this using BR and the adjusted normalised temperature.

7.6.2 Order matters (sometimes)

In the analysis presented in Chapter 5 it was assumed that the only information available about a network was the presence or absence of interactions between nodes and that the aim was to detect the most nested configuration possible. In the absence

of such information the order of rows and columns in the networks biadjacency matrix is arbitrary. To this goal, both the focal network and the networks in the null ensemble were all sorted for maximum nestedness prior to taking measurements. This was the goal of our analysis and is an appropriate focus of nestedness analysis, especially for networks where other information is difficult to measure, such as in phage-bacteria interaction networks where there is no natural hierarchical ordering of the nodes. But, this is not the only way in which nestedness analysis is conducted.

Another aim of nestedness analysis can be to focus on how environmental gradients effect nestedness within a network [15]. However, in general several intra-node, inter-node and external factors are likely to jointly contribute to network structure and function. It is unlikely a single mechanism will be responsible for overall network structure. In this kind of analysis, where the ordering of the node sets is important to attempt to understand network nestedness, neither the focal or the null ensemble networks should be sorted to maximise nestedness. This may lead to results that differ from that presented in Chapter 5 for the NODF, MD and BR nestedness measures. However, JDM and SR nestedness measures are order invariant – this makes them inappropriate for considering how factors that vary across nodes influence network nestedness.

7.6.3 Future directions

One of the reasons Rodríguez-Gironés and Santamaría [23] recreated the nestedness temperature calculator [22] was due to the poor performance of the packing algorithm (used to shuffle rows and columns to minimise the number of “surprises” - see Figure 2.1). The lack of details about the methodology used and availability of code lead several authors to redevelop the original methods. Packing the matrix is an important procedure when calculating nestedness temperature. Domínguez-García and Muñoz have recently compared several potential new packing algorithms [168] and recommend a new algorithm MusRank as it is able to produce highly visual and more intuitive displays of nestedness - than may be produced by packing algorithms that are based on minimising surprises to the isocline of perfect curvature. Furthermore, it is closely associated with the robustness of ecological networks to extinction events and therefore may provide information beneficial for conservation. The interpretation of nestedness that this study offers has no quantitative measure as yet - so it is difficult to compare with its contemporaries.

However, a new method that can be compared against other contemporary nestedness measures is of interest. By equating nestedness with the average amount of overlap between nodes Strona and Veech [169] developed a combinatorial approach to network structure that is able to distinguish segregated and nestedness network patterns and appears to outperform NODF and SR. A key strength of their approach is that it does not require the use of randomised null models – null expectation is analytically derived. Applying their approach to empirical datasets suggests that nestedness is less prevalent than found in other analyses e.g. [1]. This implies that previous observations of the widespread occurrence of nestedness is biased by the methodology used. This finding resonates with the work presented in Chapter 5, which shows strong sensitivity to the combination of measure and null model used in nestedness analysis.

These methods are not included in FALCON currently, but as it is open source, new approaches such as these could be included. As the field of nestedness analysis continues to grow such "living code depositories" will be a useful way of making technical advances available as well as demonstrating their worth.

7.7 Recommendations for improving modularity analysis

7.7.1 Best Practice

A variety of modularity maximising algorithms exist in the literature. Four algorithms have been examined in this thesis: LP-BRIM was used in Chapter 3 to investigate phage-bacteria community composition and LPAwb+, Exhaustive LPAwb+ and QuanBiMo [48] were compared against each other for plant-pollinator communities in Chapter 6. While many modularity maximisation algorithms have been compared for use in binary bipartite networks, I believe Chapter 6 provides the first comparison of algorithms for detecting modules in weighted bipartite networks. As with nestedness analysis, using null models is a useful way of determining the strength of any identified modular structure. It was demonstrated that normalised modularity (Q^{norm}) and realised modularity (Q'_R) are highly correlated and present a more meaningful statistic than the raw modularity score. They provide an assessment of detected community structure in comparison to what would be found in a perfectly modular network, where all interactions are assigned to modules. These two measures give scores

that provide more information than given by the modularity score itself and help with the interpretation of detected community structure. Different modularity maximising algorithms may have differing performance related to the network of investigation. LPAwb+ is shown to outperform LP-BRIM and a variety of other algorithms on binary bipartite networks [134]. However, we show that QuanBiMo and Exhaustive LPAwb+ are both able to detect greater levels of modularity than LPAwb+ in both binary and weighted bipartite networks. I would recommend the use of the Exhaustive LPAwb+ algorithm as it performed well on both binary and weighted networks, and is less susceptible than QuanBiMo to falling into low modularity solutions, especially on larger networks. However, Exhaustive LPAwb+ is the slowest of these algorithms on larger networks: in these cases LPAwb+ may provide the best balance between speed and accuracy. To summarise the recommendations:

- 1) Report normalised modularity or realized modularity
- 2) Use LPAwb+ in the first instance – this appears the most suitable choice for large networks
- 3) Considering the time it took to run LPAwb+ and the number of modules found, it is possible to establish sensible parameters to use with the Exhaustive LPAwb+ algorithm (if it is not too computationally costly)
- 4) Repeat several times to account for algorithmic stochasticity (see 7.7.2)

7.7.2 Algorithmic stochasticity

An understated issue in modularity analysis is accounting for algorithmic stochasticity. Stochasticity is manifested either by defining some random initial configuration (such as with LPAwb+), or by making random choices through iterations (such as with QuanBiMo). This means that different iterations of a modularity algorithm may return different identified community structure with different modularity scores. In Chapter 3 when analysing phage-bacteria community structures the LP-BRIM modularity algorithm was used multiple times on each null network and only the greatest modularity score for each null network was used in the statistical analysis. In Chapter 6 it was found that QuanBiMo could be highly sensitive to stochasticity, in comparison to either the LPAwb+ or Exhaustive LPAwb+ algorithms. Using a highly sensitive algorithm in significance testing could produce results which are unreproducible and

may artificially inflate the chance of detecting a statistically significant result and artificially increase the effect size of such a result. Accounting for algorithm stochasticity when performing analysis involving modularity, for example by using an ensemble approach as in Chapter 3, should be encouraged.

7.7.3 Other community detection methods

Modularity is not the only method for detecting communities. Leger et al. [42, 170] make a distinction between different community detection methods as either searching for clusters of densely interacting nodes that they call communities (such as modularity based methods) and sets of nodes that display similar interaction patterns which they define as structurally homogeneous subsets. They find that in binary networks edge-betweenness [171] may provide the best method for detecting communities, whilst in weighted networks a stochastic block model [172] technique is preferred. They also favour modularity maximisation [173] as an alternative to the stochastic block model technique, as this can be very slow. Note that a different modularity maximisation algorithm was used to those described in this thesis. Despite using bipartite network data, the algorithm used [173] and corresponding modularity goal function is designed for use with unipartite networks. Additionally, Chapter 6 and previous work [49, 134] suggests that algorithm choice can have a significant effect on the communities that are identified and their corresponding modularity scores (as well as computational resources required to perform detection). One of the problems with the approach to modularity maximisation used in Leger et al.'s analysis [170] is that the number of retrieved subgroups is overestimated - however, this is specific to the algorithm used, rather than to the method of maximising modularity. For example, the QuanBiMo algorithm starts with 1 module (including all nodes) and new modules are iterated in when it improves modularity. On the other hand, the LPAwb+ algorithm uses a top-down approach; starting in a configuration where each of the nodes in the smallest node set is assigned its own unique module - modules are then joined when it leads to an improvement to the modularity score. It was possible to find large discrepancies between the number of modular groups identified by the different algorithms assessed in Chapter 6. There was no time to evaluate our methods against this analysis [170], however, the data and methodology for the 6,400 evaluated networks have been made available and we plan to evaluate our methods against this dataset in the future. While ecological studies have mainly focussed on using modularity maximising techniques as a way of identifying modules [47, 170], other techniques for identifying communities exist [43, 42] some of which, such as

stochastic block model methods in weighted networks and edge-betweenness methods in binary networks [170], are worth investigating further.

7.7.4 Reporting of modular community structure

Returning to modularity, the way in which modular communities are reported is another issue worthy of pursuit. Why should the best division of communities be found by maximising a goal function (e.g. Barber's modularity - as used in Chapter 3 and Chapter 6) - and does the best mathematical division of nodes represent real communities? Modularity landscapes are glassy – characterised by a highly disordered and rugged terrain with many local maxima [50]. Additionally the network may contain information at different hierarchical levels, but there is no way to tune the resolution of the modularity algorithms [50]. It is typical that there may be several community classifications that achieve modularity scores near the maximum - which is one of the reasons that modularity algorithms frequently give stochastic output. Perhaps a more meaningful guide to a networks community structure might be given by showing the probability that each node has of being classified in the same community as other nodes. I began preliminary work in this direction during my studies using the output from multiple runs of the LPAwb+ algorithm on a single network and calculating the proportion of times nodes were classified as belonging to the same modules. As real ecological networks are often formed of overlapping communities this approach may be useful as it does not rely on a single fixed definition of a community; as noted in Chapter 6 other modularity definitions than Barber's exist for bipartite networks. However, this approach I propose may be sensitive to the way in which the modularity algorithm employed is able to traverse the modularity landscape. On the other hand, it would provide more information than given by the traditional fixed community model that would be beneficial for those wishing to apply community detection in a practical context, for example to help design conservation strategy [170]. Additionally methods for testing how the outcome of such a community probability model differs from random community formation through use of a null model would need to be developed.

7.8 General recommendations for bipartite network analysis

7.8.1 Call for standardised benchmarking tests

More robust methods for evaluating nestedness and modularity are required. There is a need to discover measure, null model and effect size combinations that are insensitive to basic network properties such as size, fill and connectance – or at least create a robust guide for choosing the best method of detection for use in a particular network. In nestedness analysis lots of measures are being created, whilst in modularity analysis several algorithms have been proposed to maximise bipartite modularity. Evaluating the performance of these methods is computationally intensive and not a trivial task. It would be good to create some core benchmarking schemes that new measures and algorithms (as well as null models and effect sizes) can be compared against, rather than new benchmarking schemes needing to be created each time. Making such a scheme open and accessible to all within a research community will be key to the success of such a scheme. In addition, reaching agreement on the methods used to assess these network properties will be necessary if such a scheme is to become accepted and used. Despite these challenges, such a goal is worth pursuing. Being able to compare the available options will be a valuable tool, both to theorists interested in developing better methods as well as to those wishing to apply these techniques to their empirical networks.

To this purpose the underlying networks (as well as output statistics) used for benchmarking in Chapter 5 have been released to the research community [130]. A similar dataset using networks on a gradient between highly modular and random could be created for modularity. Such a method has been used in unipartite networks [171] and could be generalised to the case of bipartite networks. I also note that Leger et al. [170] provided a comparison of clustering algorithms using 6,400 synthetic networks based on properties of 47 ecological networks in the Interaction Web Database. They have made their work available online and this constitutes an important dataset for comparing different modularity algorithms, as well as other community detection methods in weighted and binary networks. These datasets represent key steps towards benchmarking schemes for community detection and nestedness.

7.8.2 Making use of quantitative information

A great deal of work in ecological systems has dealt with binary bipartite networks. However, representing interactions as either on or off is a grand simplification of the systems being described and of the data available. As more and more weighted datasets and methods to analyse them become available, it will be of increasing importance to account for differences between measures. Already there are at least four measures of nestedness that can be applied to weighted networks [34, 38, 35, 39]. Benchmarking these measures in a similar way to that performed for nestedness in binary networks in Chapter 5 would be a useful advance to the understanding of weighted nestedness. Comparing the structural patterns identified in binary networks with those identified in weighted networks is also of great interest. Staniczenko et al. [35] found models of mutualistic communities showed a strong nestedness signal in binary networks, but it was not present in weighted networks. A similar story may be apparent for modularity, which has only recently been applied to weighted bipartite networks. In Chapter 6 it was shown that different modular structures were identified in plant-pollinator networks when changing between binary and weighted network representations that may or may not be associated with a change in the strength of observed modularity. The ecological relevance of these differences is currently unknown, but it suggests that patterns easily identifiable in binary networks may not be generalisable to weighted networks.

7.8.3 Identification of multi-scaling topologies

Nested-modular network structures such as that identified in Moebus and Nattkemper's phage-bacteria dataset collected in the Atlantic Ocean [82, 80, 102] as seen in Chapter 3 are a form of multiscale network structure. Multiscale structures are those in which network features are found at different hierarchical levels, which may indicate for example geographical location [80], or species level selection [102]. Two software packages: BiMAT [54] and bipartite [53] (which implements the QuanBiMo [48] algorithm) offer ways for searching for multiscale structure. BiMAT can identify nested-modularity by searching for nested structures within identified modules (as shown in [80]), whilst QuanBiMo is able to identify modules within modules.

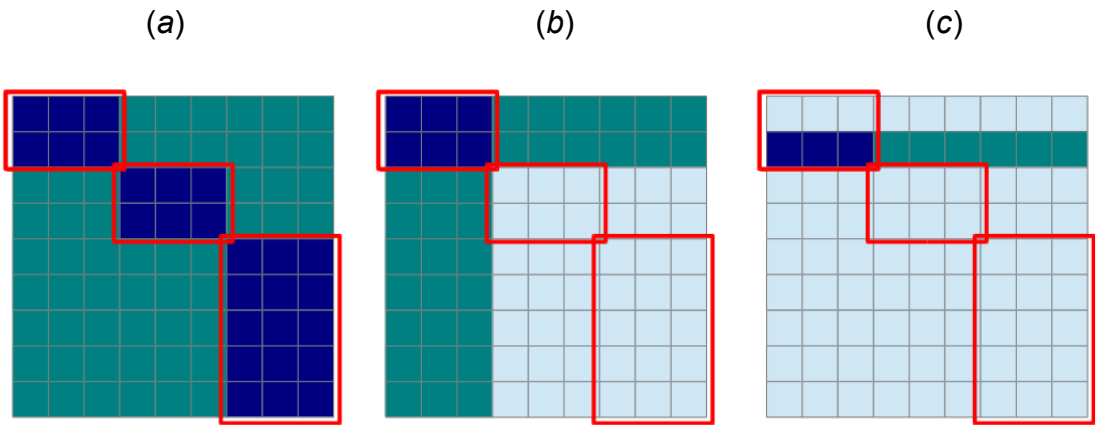


Figure 7.1: Suggestion of how realised modularity (Q'_R) (see Section 6.2.3) may be used to assess modular structure (*a*) at network level, (*b*) at module level (evaluation for top left module shown) and (*c*) at node level (evaluation for second row node shown) within an example network. Red rectangles highlight an identified modular structure. Dark blue cells are those that should be summed to find H and turquoise cells and blue cells should be added to find M in each case.

Nestedness and modularity are usually classified at the scale of the whole network. This raises the question of how should these subscale features be related back to the multiscale network they are part of? Creating appropriate multiscale statistics that test how individual nodes, within-module connections and between-module properties relate could be useful for understanding multiscale topologies. An example of how this may be performed for a metric such as realised modularity (see Chapter 6) is shown in Figure 7.1.

In Chapter 3 nested-modularity was quantified by looking at the average nestedness from each of the modules. Is there a better way to quantify nested-modular structure at the scale of the whole network? What are the suitable statistics for use in multiscale structures? These questions are left for future researchers.

7.8.4 A new framework for describing bipartite network topology?

It has been suggested that modularity is the opposite of nestedness [114]. This description seems to hold in the work of Podani and Schmera [174] who introduce a framework to describe interaction networks by comparing all possible pairs of rows (and then columns) and evaluating them on a simplex. In this framework nestedness

is shown to be the opposite of species replacement - which is maximised in a perfectly modular pattern. This approach has also been extended for use in weighted networks [39] and looks a promising order invariant technique for evaluating the structural features of networks. Additionally there may be undesirable correlations between measures for different individual networks patterns [175] that may make interpretation of the network difficult. By simultaneously evaluating the network in terms of overlap between nodes, rather than any individual network pattern, this framework becomes a useful guide to understanding how nodes contribute to network topology. This approach potentially offers more information about a network than can be gained using traditional measures of modularity and nestedness, however comparisons between these methods and this new framework would be advantageous in determining how well they relate and whether this framework is suitable for addressing the questions that nestedness and modularity analysis were developed to tackle.

7.8.5 Null model hypothesis testing

Finding the nestedness or modularity of a network alone is not in itself enough. Rather, we want to know how this compares to some expectation about the network [57]. In the absence of analytical forms for null model distributions, we are left with a dependence on randomisation techniques (though see “node segregation” [169] in Section 7.6.3). This can create a problem as there exist a different number of network reconfigurations, dependent on network properties. Typically null model analyses involve a large number of instantiations of the null model rule in order to compare the results from these to the focal network. However, the chosen number of null models used is more often tied to some aesthetically pleasing number, than the number of possible reconfigurations and the ability of the null model to simulate this distribution. This may reduce our ability to perform robust statistical analysis in general and is something that needs to be considered in practical applications.

Until recently [58] the FF null model was generated by making sequential permutations on the focal network. Not only was this method slow, but was subject to bias as a suitable number of “burn-in” swaps need to be performed to generate sufficiently randomised solutions that are independent of the previous solution [176, 177, 178]. It is argued that a number can be placed on the number of steps between samples in such a solution. However, I find it much more likely that this number is network dependent. However, the introduction of the Curveball algorithm [58] allows the FF null model to be generated without any sequential dependence - each null network is independent. For this reason we implemented the Curveball algorithm in FALCON.

However, both these points raised another question - what number of null networks need to be used to sufficiently simulate the null model distribution and generate robust statistics? Rather that being some manually decreed arbitrary user choice, maybe an algorithm can be used to decide the suitable number of networks to be used for the simulated distribution to represent the theoretical null model distribution. This is the premise of the adaptive null ensemble method used by FALCON. Crucially this method can perform no worse than that achieved by using 1,000 null models (a common ensemble size in the context of nestedness analysis). Checking whether the simulated null model distributions are internally consistent offers a way towards more robust significance testing in nestedness and modularity network analysis where the sample space of null models in network dependent. However, this principle may also be useful in applications outside the realm of bipartite network analysis.

7.9 Conclusions

Bipartite networks are increasingly useful in many research fields. Two new tools FALCON and LPAwb+ have been introduced to improve the identification of nestedness and modularity in bipartite networks. I have shown that methods for modularity and nestedness can be made more robust by judging their performances using benchmarking schemes. This research shows a way to move towards a consensus on robust methods for describing these interesting properties of bipartite networks and facilitate their use in identifying and explaining their existence in natural phenomena.

Appendices

Appendix A: Network sampling for synthetic network ensemble

We used a Latin Hyercube sampling (LHS) design to choose the characteristics of the initial 500 'perfectly nested' networks in this study. Figure S1 shows the set of parameters chosen by the LHS design which were used.

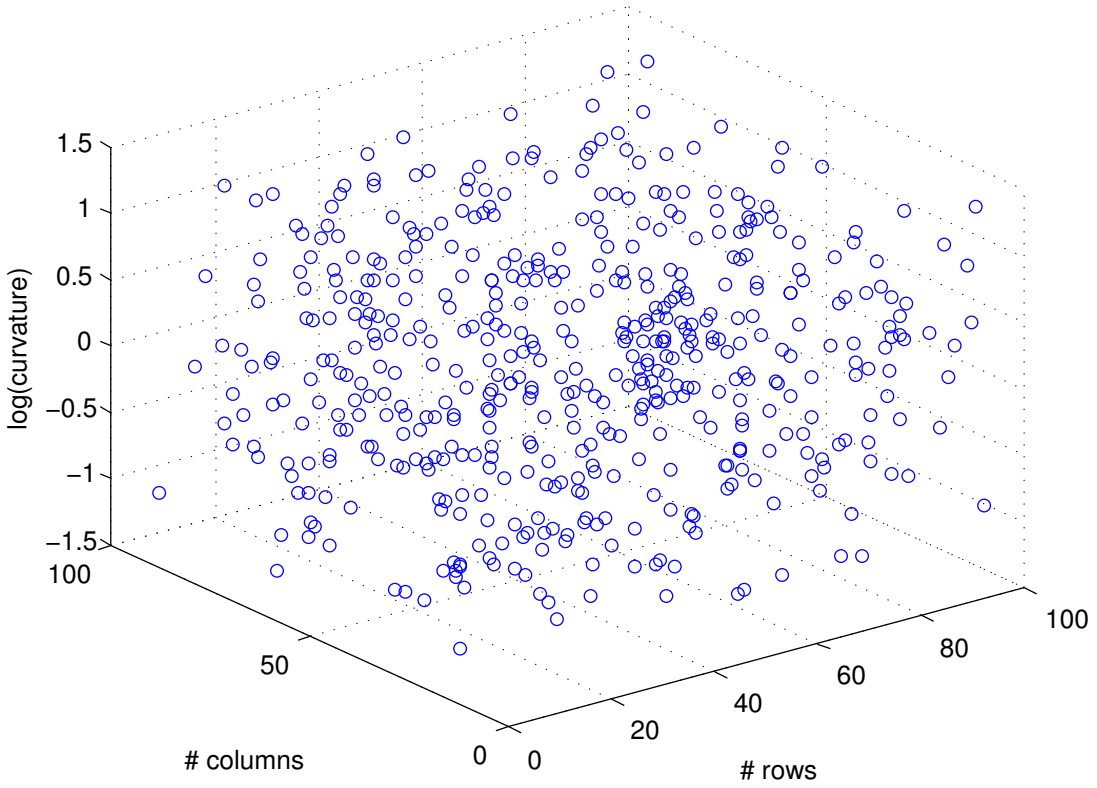


Figure S1: Row, column and curvature parameters chosen for the 500 initially 'perfectly nested' networks using Latin hypercube sampling design to create representative networks of the entire sampling space.

Appendix B: Measuring the NTC

Computational performance and nestedness temperature calculations

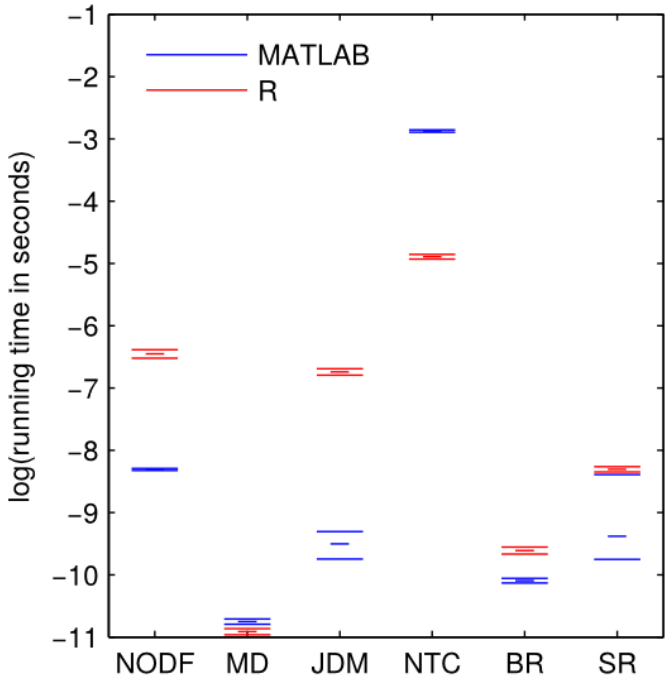


Figure S2: Average computational timings of measures and null models in FALCON on 10x10 networks, from 100 evaluations.

Of the six nestedness measures considered here, the nestedness temperature calculator (NTC) is considerably the slowest to compute (Figure S2). Whilst we were able to compute NTC scores for each of the 30,000 networks in our synthetic ensemble; due to computational limitations we were unable to calculate the statistical significance of these results under all networks.

$P(rewire)$	#Networks NTC measured in statistical ensembles
0.01	847
0.05	751
0.10	1,216
0.15	593
0.20	671
0.50	568
TOTAL	4,646
TOTAL SHARED	2,334

Table S1: Number of networks for which statistical significance of the Nestedness temperature calculator (NTC) was evalauted.

We used the time taken to perform the NTC on each of the networks in the synthetic ensemble as a proxy for the time it would take to compute NTC in the corresponding null network ensembles. When the NTC was calculated in less than 0.12 seconds (corresponding to roughly 0.12 (seconds to measure a network) $\times$ 5(null models) $\times$ 1000(minimum null ensemble size) $\simeq$ 600 seconds of extra performance time per network); the NTC was used to measure the null network ensembles. If the time taken to calculate the NTC was greater than 0.12, then statistical significance was not calculated. Note that while 0.12 seconds does not sound a very large amount of time; if the NTC took 0.12 seconds to be computed on all 30,000 networks - and statistical significance was calculated - this would correspond to at least an extra 18,000,000 seconds ($>$208 days) computation time! Table S1 shows the number of networks for each $P(rewire)$ level for which statistical significance was calculated. Statistical significance for NTC was calculated for 4,646 of the 30,000 networks in the synthetic ensemble - this sample is used to calculate the proportional agreement between measures in Figure 5.3 and Figure S3; and also in the measure-measure scatterplots of p-values and in all figures showing effect sizes. We also created a sample of 2,334 networks for which NTC was measured across all six $P(rewire)$ levels (389 at each level) - which is used in figures S10, S11 and S22.

Appendix C: Robustness of measure agreement results

Distinguishing nestedness at p $<=$ 0.001

In Figure S3 we recalculate the results shown in Figure 5.3 using a smaller p-value threshold of significance of $p <= 0.001$. Qualitatively similar patterns of agreement are found, indicating the robustness of the result.

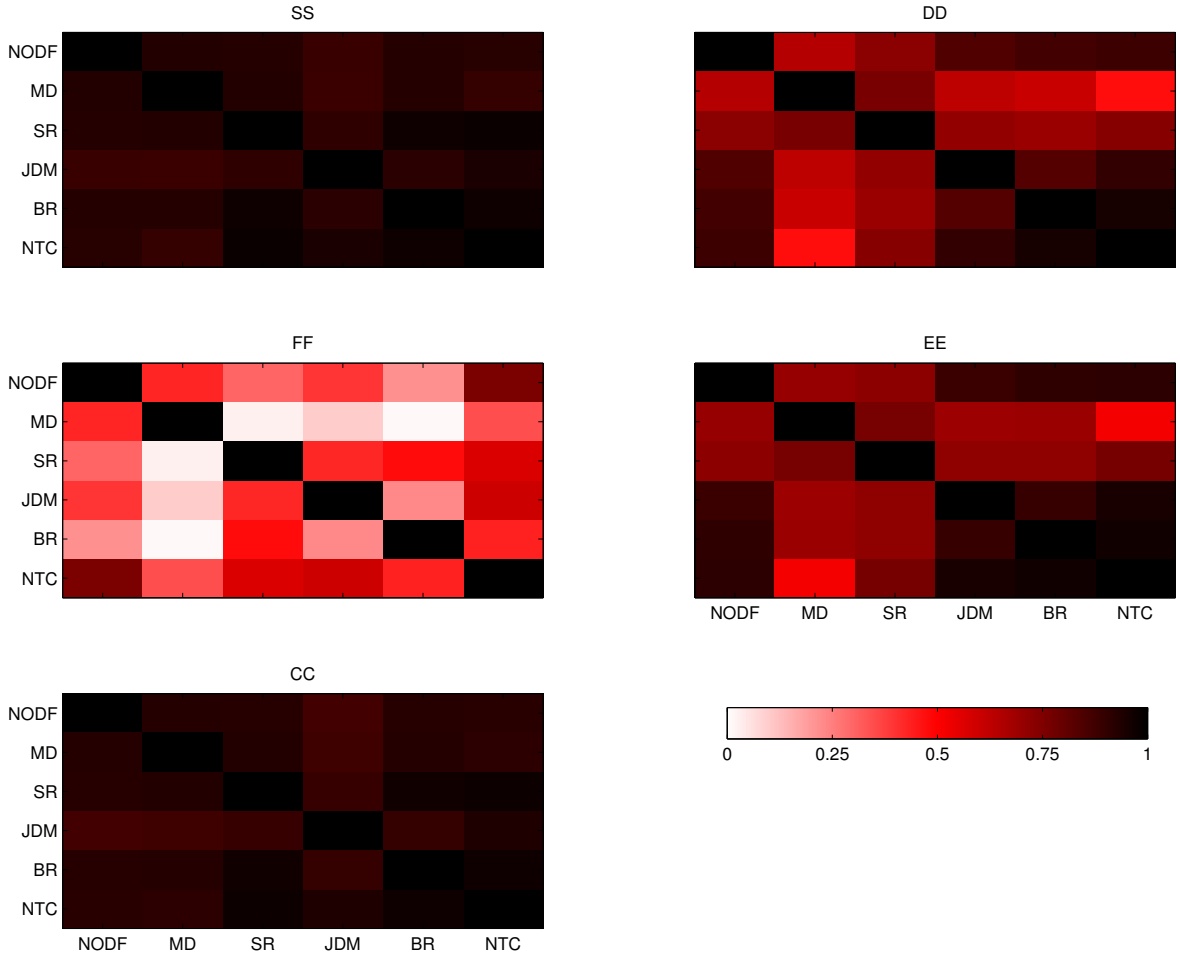


Figure S3: Proportion of pair-wise agreement of 'significance' of nestedness between different measures across the synthetic network ensemble. Each subplot shows the results under a different null model. Three categories of significance were chosen : $p <= 0.001$, $0.001 < p < 0.999$ and $p >= 0.999$. An agreement between two measures is reached when both measures classify a network in the same category as one another. This process was repeated for each combination of measures on each of the five null models. Proportion of agreement was made on 4,646 networks where NTC was calculated; and 30,000 in other cases.

Distribution of p-values

The following plots (figures S4 - S8) show scatterplots of the statistical significance of nestedness obtained by each of the measures, using p-values. Each figure shows the results obtained by a different null model. Within each null model, statistical significance was calculated on the same set of null matrices so the results are directly comparable. Above each measure name am histogram (with bins of width 0.05) shows the distribution of observed p-values for that measure. The histogram's height is scaled by the number of networks for which statistical significance was calculated (4,646 networks for NTC; 30,000 networks for all other measures). This scale is given by the extent of the lines in the top-left subplot of each figure. It is noted that under the FF null model (Figure S5) all p-values found using the Manhattan distance measure (MD) are found as $p = 1$. This result is obtained as MD is insensitive to the swaps made in the FF null model.

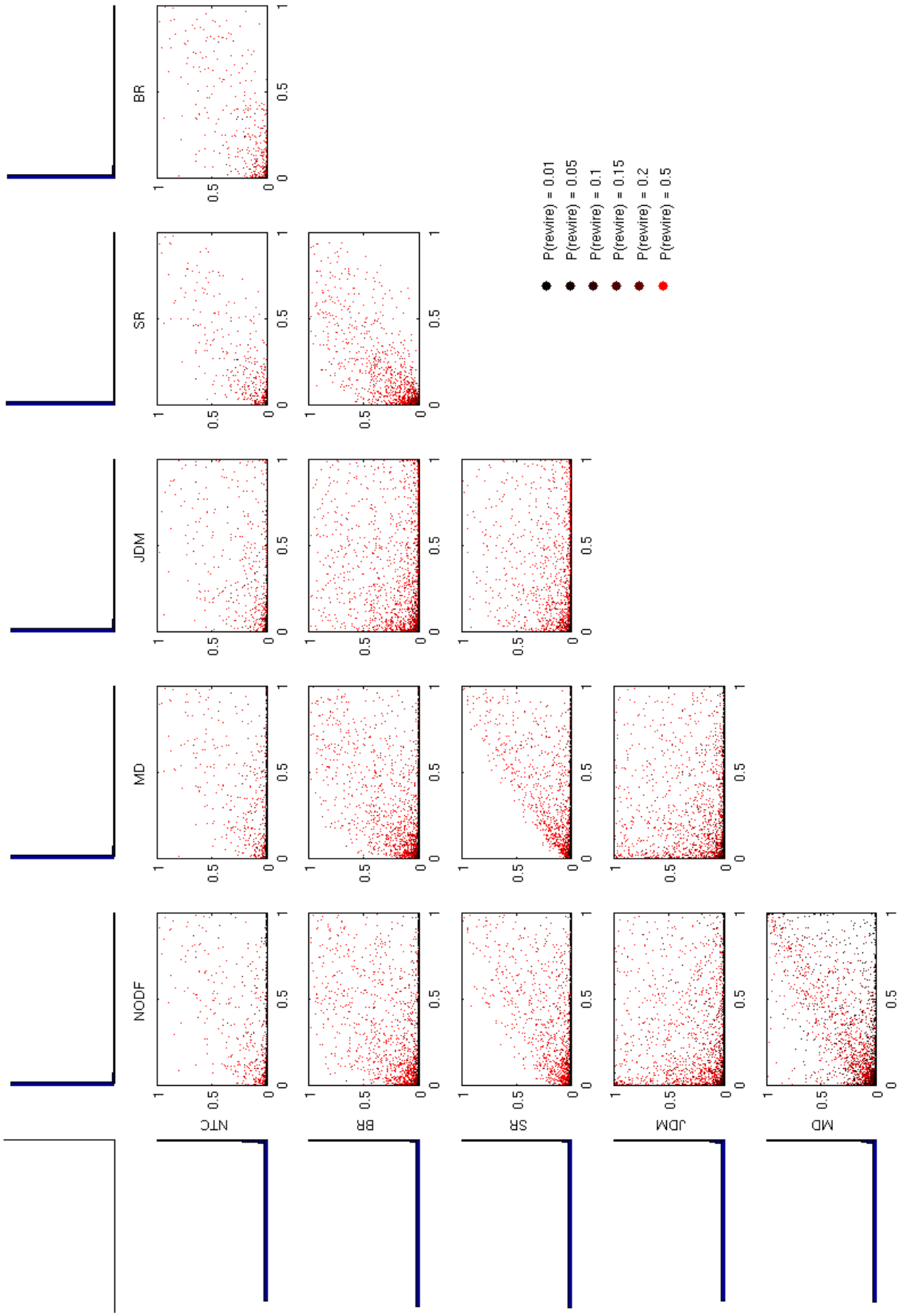


Figure S4: P-value measure-measure scatterplots for the SS null model.

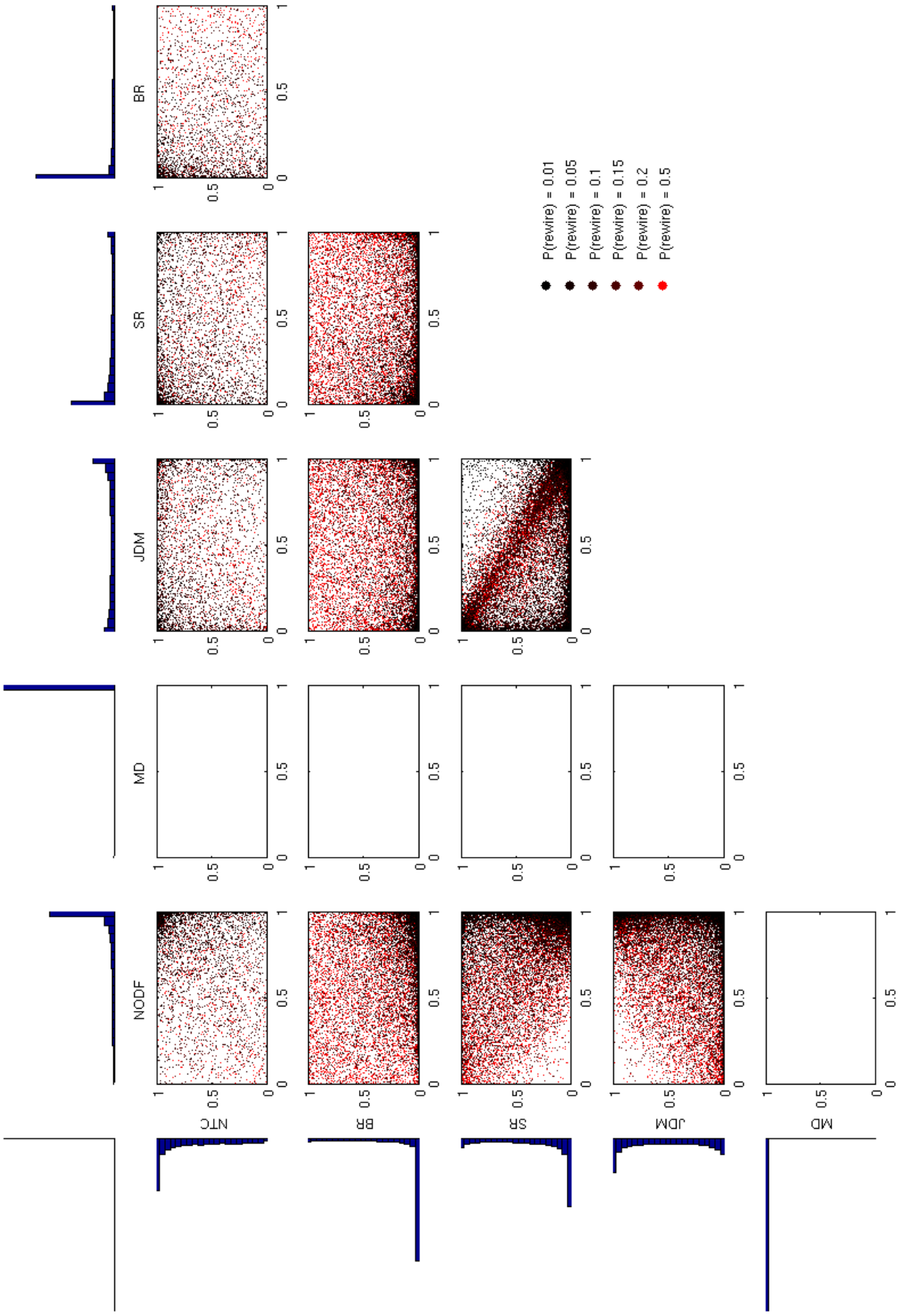


Figure S5: P-value measure-measure scatterplots for the FF null model.

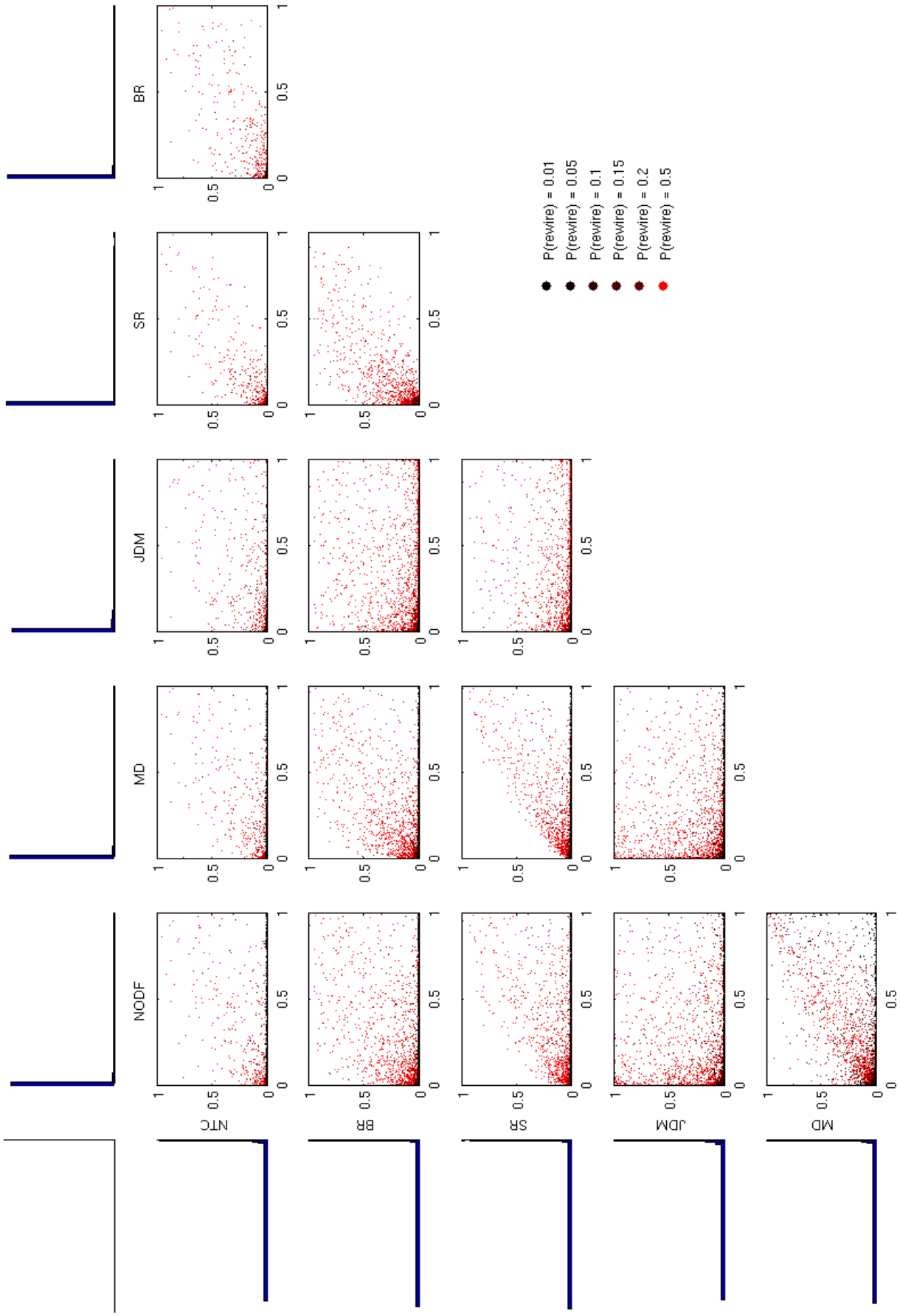


Figure S6: P-value measure-measure scatterplots for the CC null model.

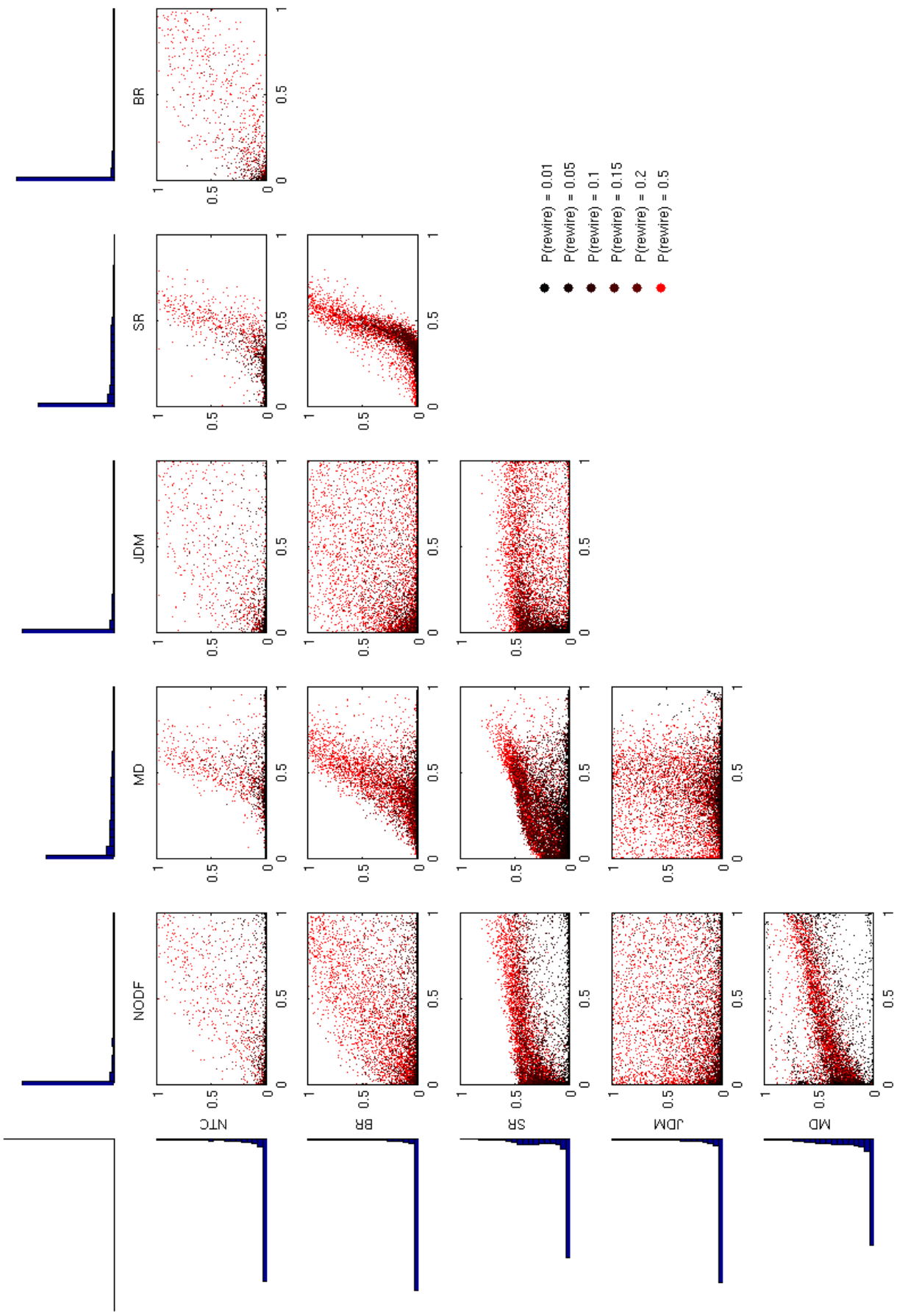


Figure S7: P-value measure-measure scatterplots for the DD null model.

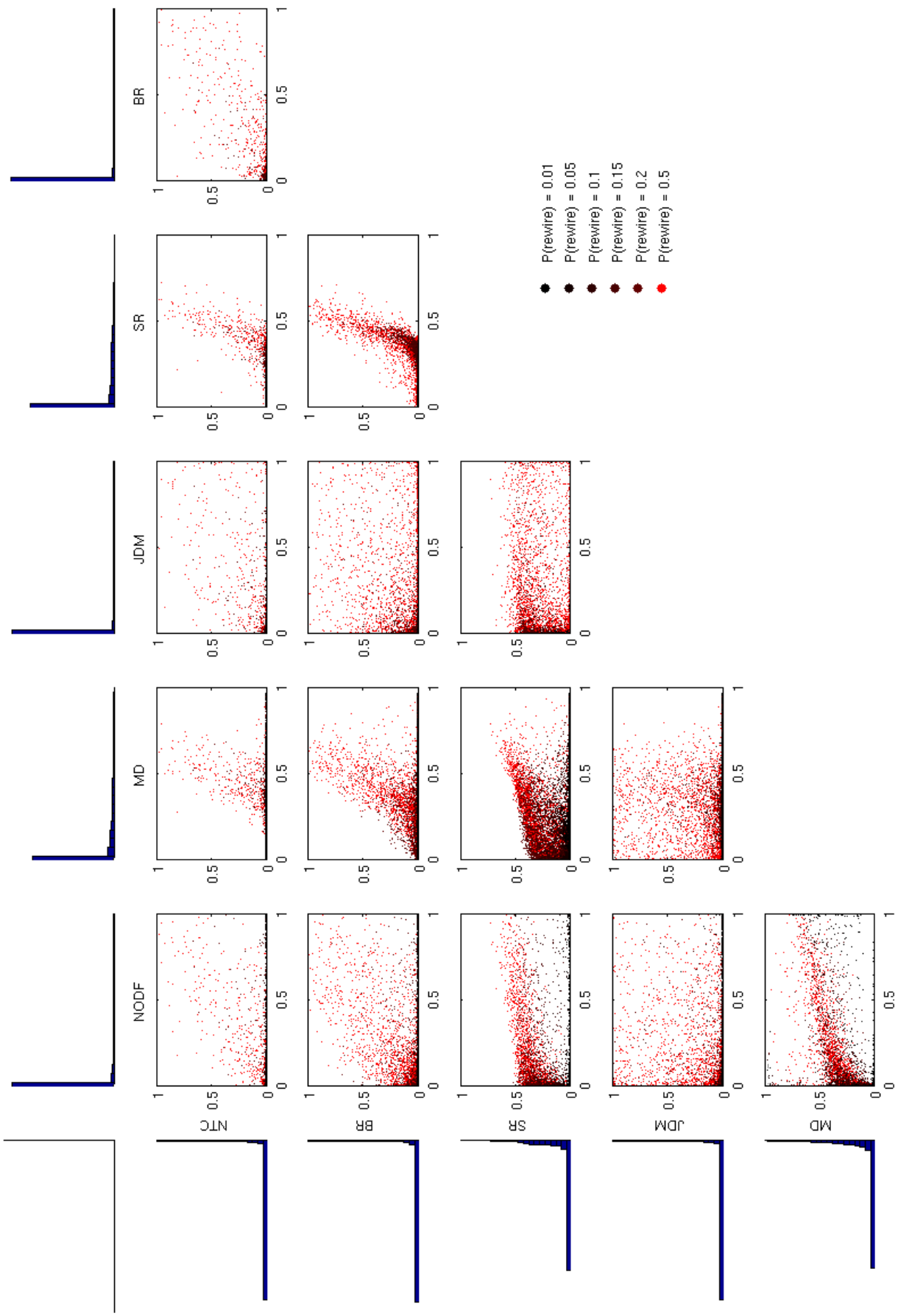


Figure S8: P-value measure-measure scatterplots for the EE null model.

Appendix D: Robustness of nestedness discrimination ability

Figure S9 shows a recalculation of Figure 5.4 where a significance threshold of $p <= 0.001$ is used instead of $p <= 0.05$. Qualitatively our results are robust, though the stricter threshold reduces the amount of type I error observed.

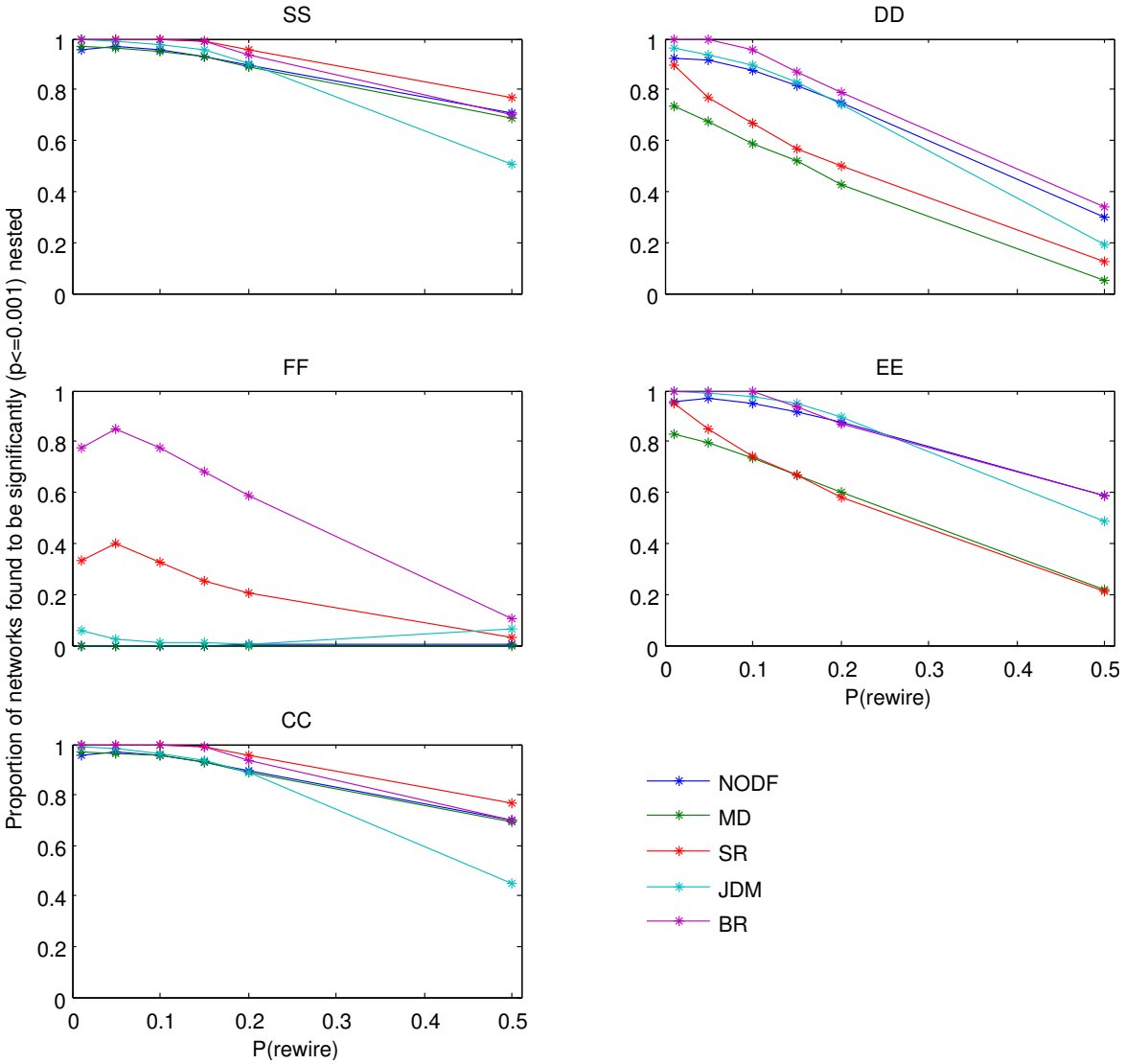


Figure S9: The ability of different measures to discriminate significant nestedness under each of the five null models in the full ensemble of 30,000 synthetic networks. Each subplot show the results for a different null model. Each line shows the proportion of networks at each level of $P(rewire)$ that were found to be significantly nested ($p <= 0.001$) using a specific nestedness measure.

In figures S10-S11 we recalculate the results shown in Figure 5.4 and Figure S9 using the sample of networks from the ensemble that the NTC was measured on (see Appendix B: Measuring the NTC).

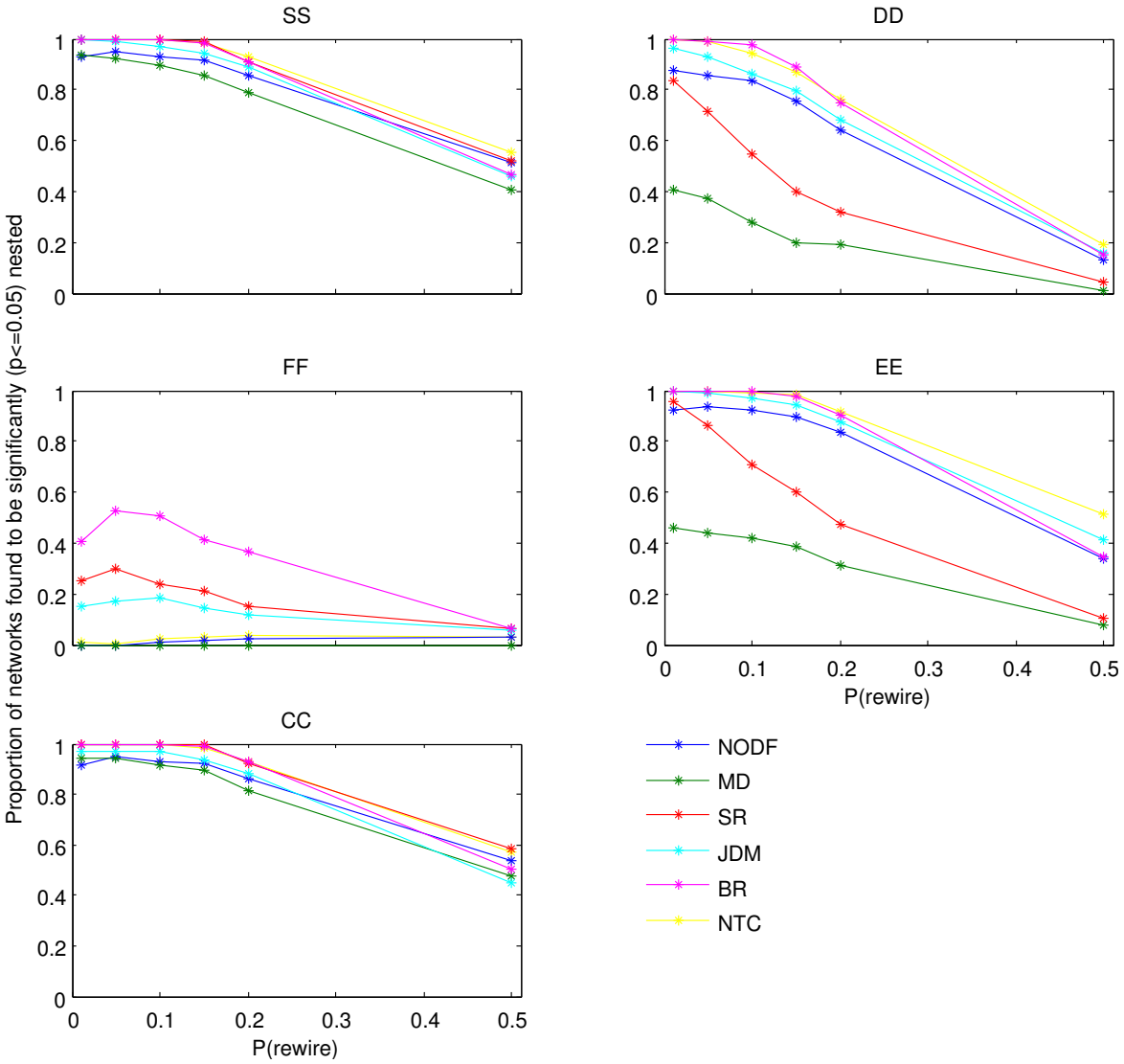


Figure S10: The ability of different measures to discriminate significant nestedness under each of the five null models for 2,334 networks from the synthetic null ensemble where the nestedness temperature calculator (NTC) was measured (389 for each $P(rewire)$ level). Each subplot show the results for a different null model. Each line shows the proportion of networks at each level of $P(rewire)$ that were found to be significantly nested ($p <= 0.05$) using a specific nestedness measure.

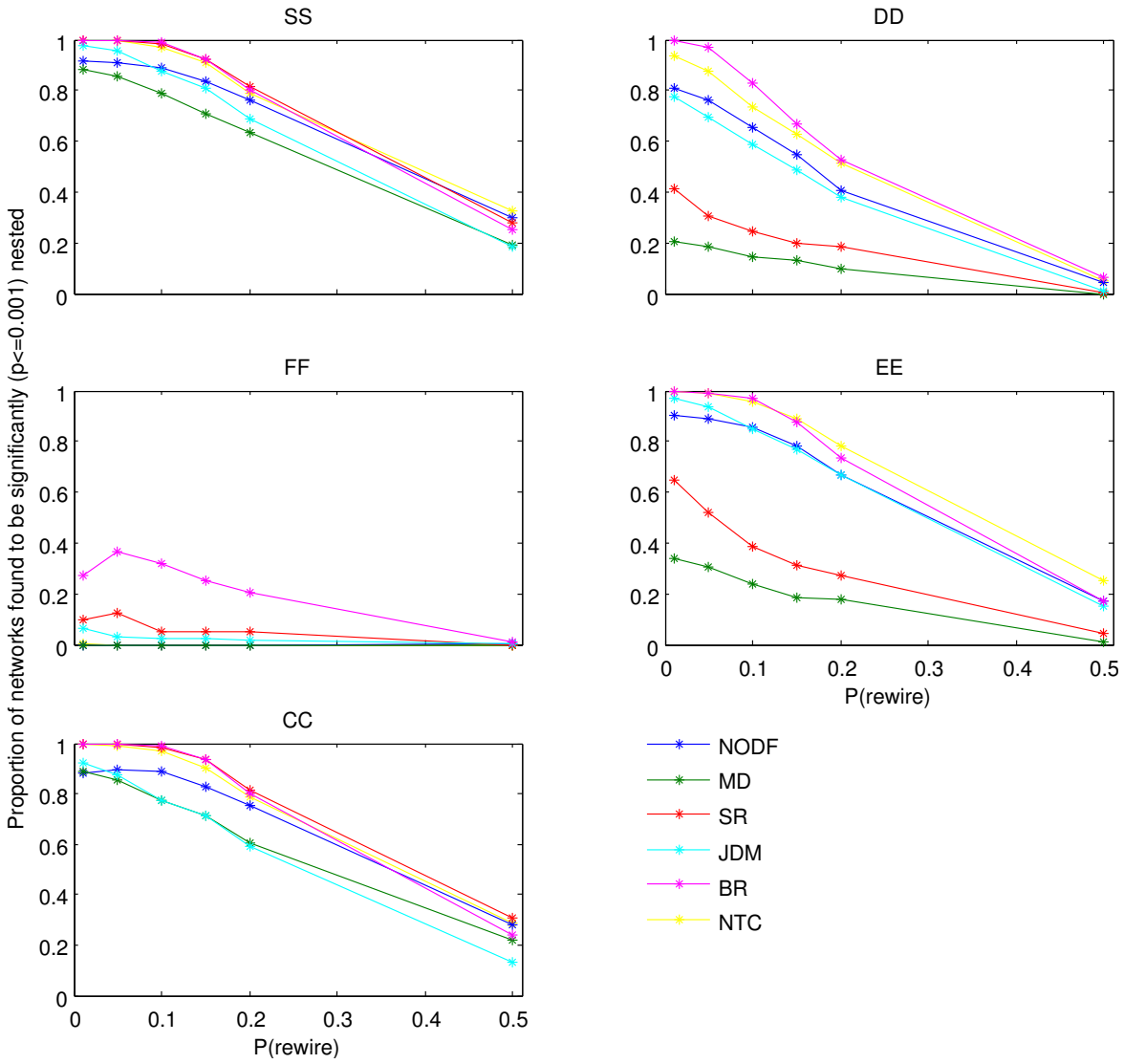


Figure S11: The ability of different measures to discriminate significant nestedness under each of the five null models for 2,334 networks from the synthetic null ensemble where the nestedness temperature calculator (NTC) was measured (389 for each $P(rewire)$ level). Each subplot show the results for a different null model. Each line shows the proportion of networks at each level of $P(rewire)$ that were found to be significantly nested ($p <= 0.001$) using a specific nestedness measure.

In all 4 cases and across all tested null models BR exhibited the smallest type II errors (highest proportion of low $P(rewire)$ networks that were significantly nested), whilst MD exhibited the highest type II errors. Type II error was mostly small, except in the FF null model, where it was high for all except the BR measure in the full network ensemble. In the NTC sample of networks BR also suffered high type II errors. This suggests that the NTC sample is a biased sample of networks from the whole ensemble.

Type I errors differed between null models and across the 4 cases considered; in SS and CC null models JDM and MD had lowest type I errors, in FF NODF (MD is as well but this is inconsequential due to it being insensitive to this null model), and in DD and EE MD and SR had the lowest type I errors.

Whilst the combination of SR in the DD null model looks like the best in the whole network ensemble; this result does not hold in the NTC network sample. In the smaller sample SR with EE looks best when using $p <= 0.05$, but this result is sensitive to the choice of p-value threshold. At $p <= 0.001$ BR with DD appears to be the best choice.

Appendix E: Comparing effect sizes

In the following we define x as the observed nestedness of the network of interest, $\bar{x}$ as the mean average nestedness found in a null model ensemble and σ as the standard deviation of nestedness within a null model ensemble.

Distribution of sample z-scores

The following plots (figures S12 - S16) show scatterplots of nestedness effect size obtained by each of the measures, using sample z-scores. Sample z-scores are calculated as:

$$z = \frac{x - \overline{x}}{\sigma}$$

In order to more easily evaluate these plots we have reversed the axes of the BR, MD and NTC measures. By doing so, we transform the plots so that the direction of increasing nestedness (left to right on the x-axis and bottom to top on the y-axis) is the same in every subplot. As such z-scores on the right (above) of 0 indicate a network was more nested than expected under the null model, whilst z-scores on the left (underneath) of 0 indicate a network was less nested than expected under the null model. In Figure S13 dashed lines showing 0 effect size are provided. In addition no effect size is plotted for MD as this measure was insensitive to the swaps made in the FF null model.

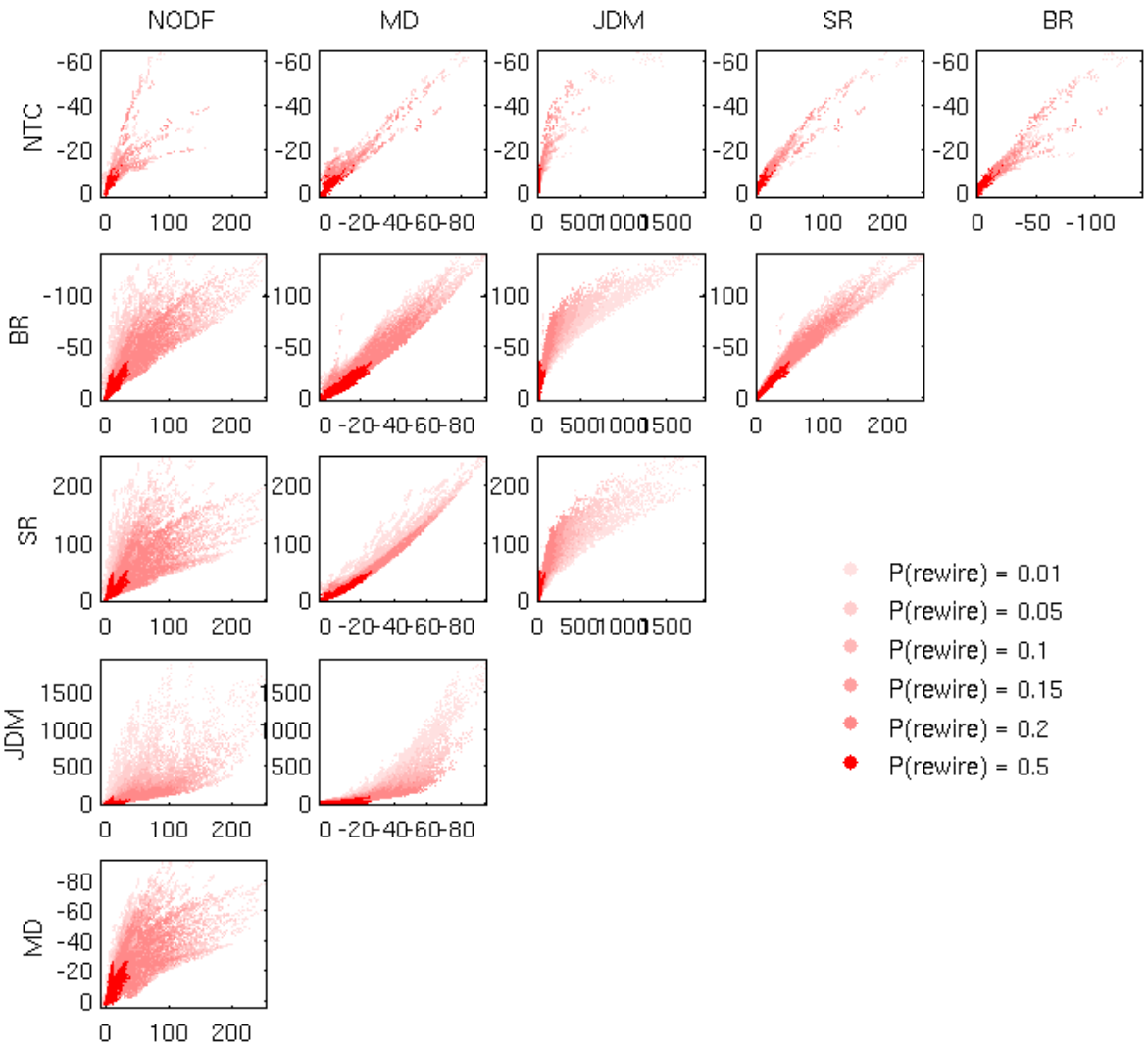


Figure S12: Z-score measure-measure scatterplots for the SS null model.

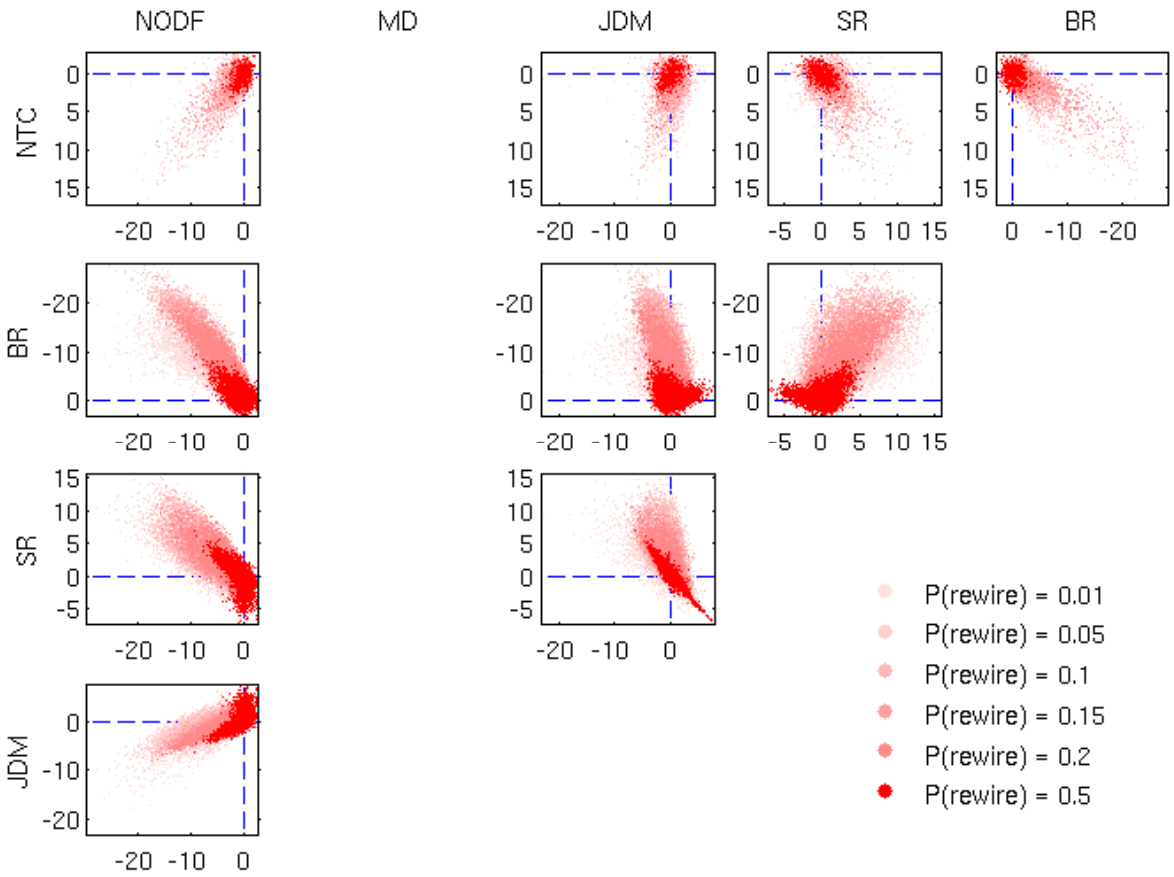


Figure S13: Z-score measure-measure scatterplots for the FF null model.

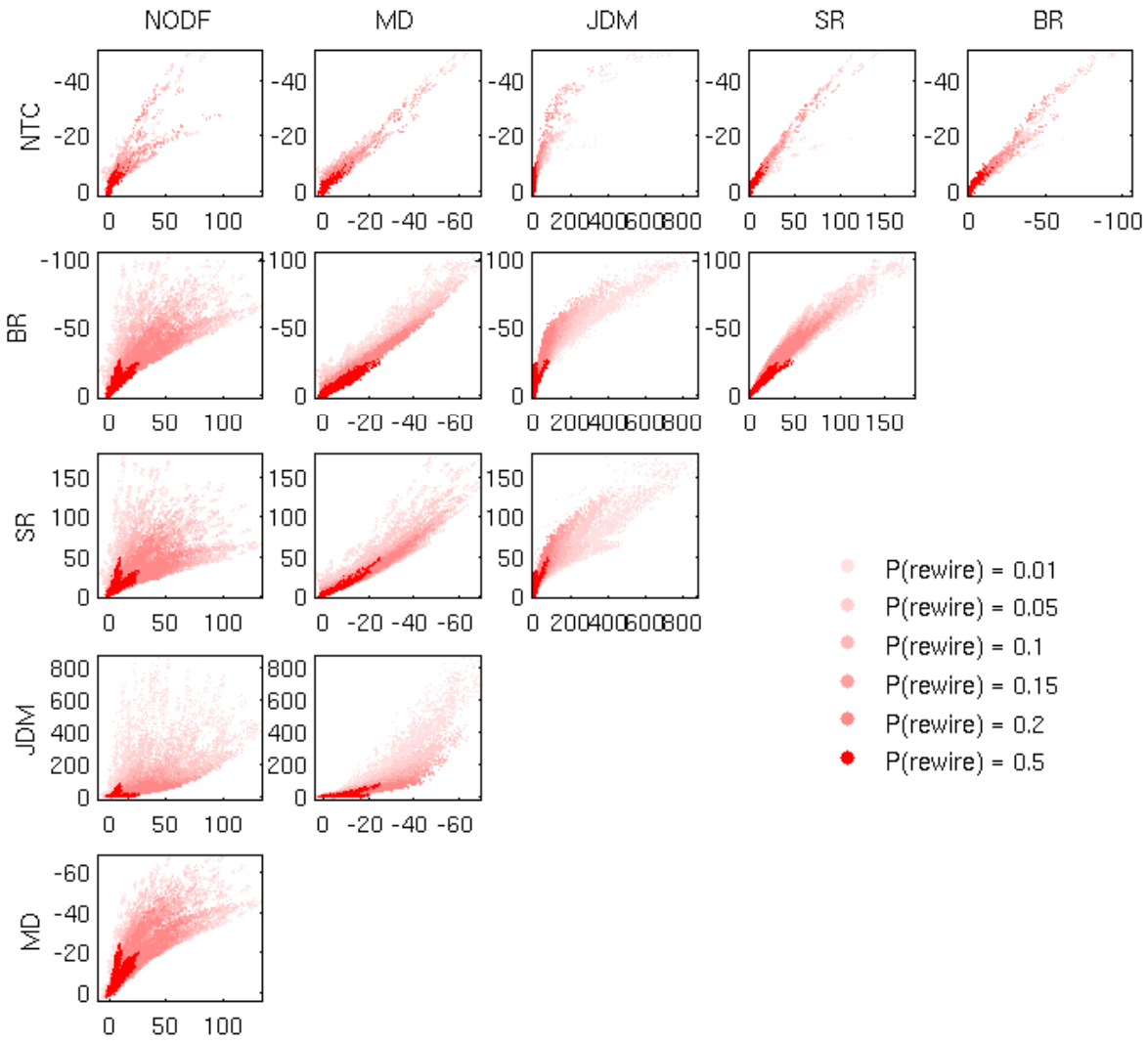


Figure S14: Z-score measure-measure scatterplots for the CC null model.

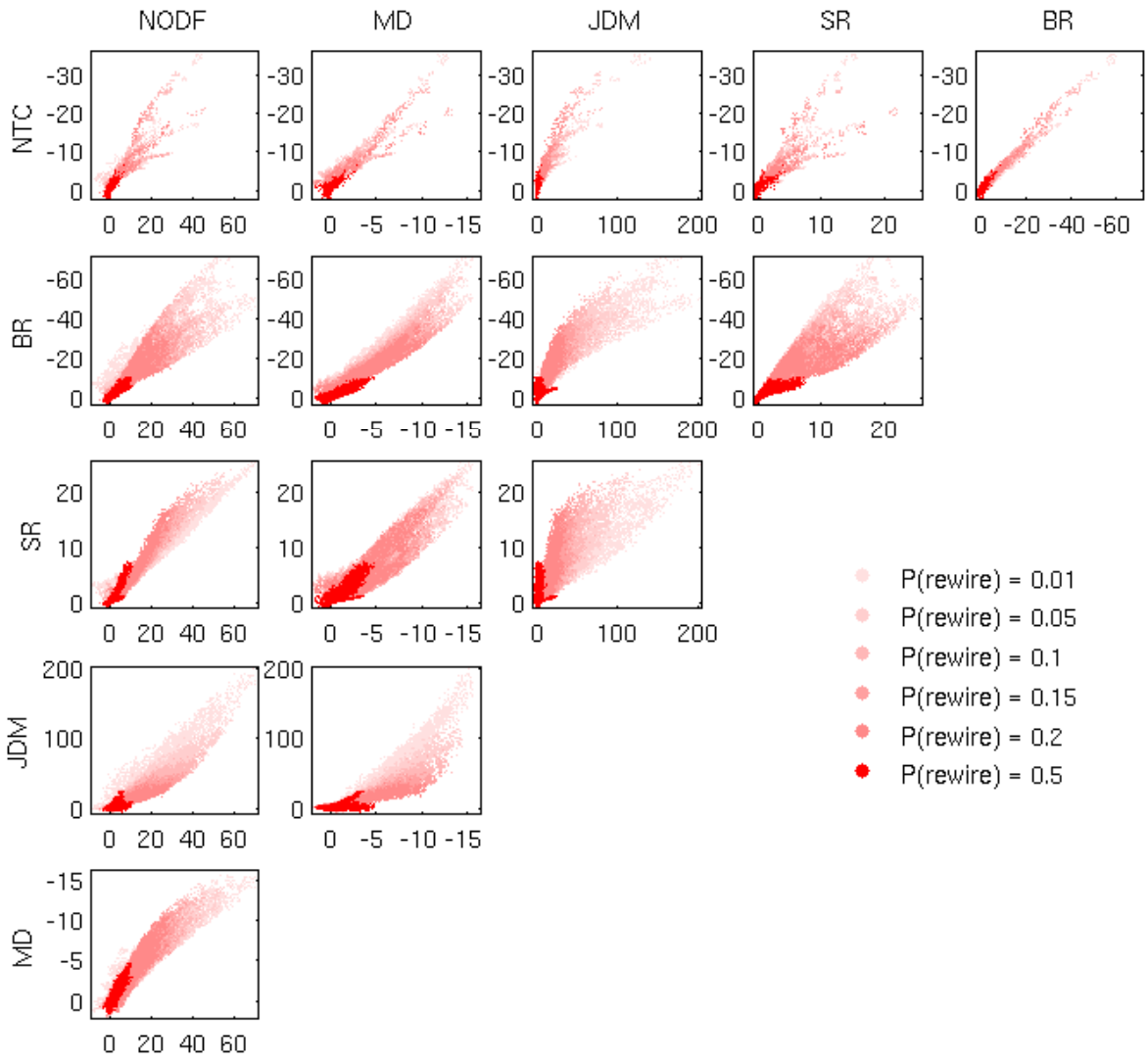


Figure S15: Z-score measure-measure scatterplots for the DD null model (also shown in Figure 5.5).

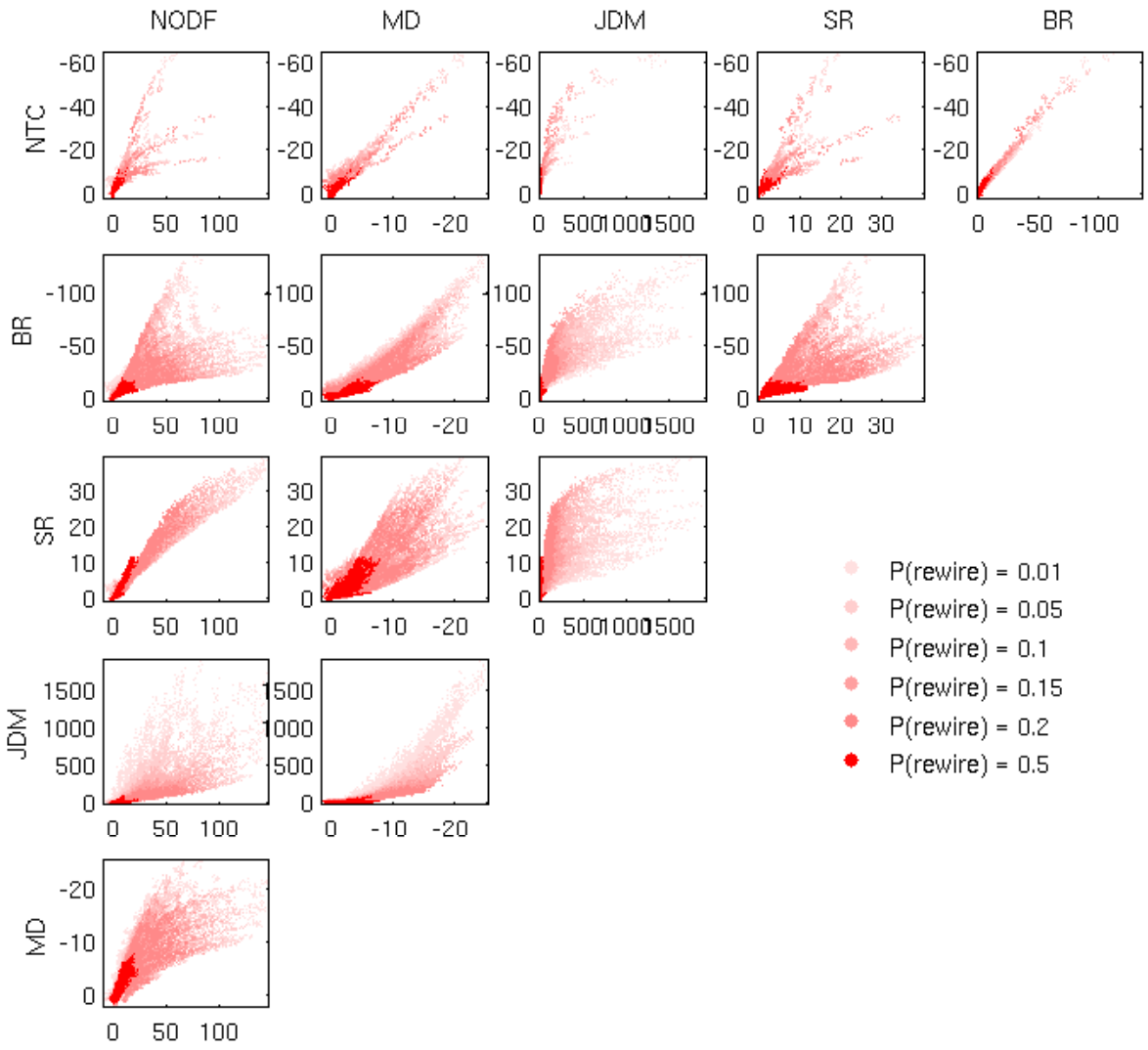


Figure S16: Z-score measure-measure scatterplots for the EE null model.

Distribution of adjusted normalised temperature scores

The following plots (figures S17-S21) show the distribution of adjusted normalised temperature scores for the synthetic networks.We formalised the normalised temperature as a test statistic [129] based on the proposed τ-temperature nestedness measure [33] as:

$$T = \frac{x}{\overline{x}}$$

This statistic is just a shift of the relative nestedness (RN) effect size such that:

$$RN = \frac{x - \bar{x}}{\bar{x}} = \frac{x}{\bar{x}} - 1 = T - 1$$

However, neither relative nestedness or normalised temperature take into account the direction in which nestedness is measured. This may have an effect on the way test statistics can be compared. To account for this we propose the adjusted normalised temperature:

$$A = \left(\frac{x}{\bar{x}}\right)^{U}$$

where the value of U depends on the direction in which nestedness is measured:

$$U = \begin{cases} -1 & \text{nestedness increases with increasing measure score} \\ +1 & \text{nestedness increases with decreasing measure score} \end{cases}$$

so for example using the NODF measure (which has a maximum of 100) $U = -1$, and for BR (which has a minimum of 0) $U = 1$. In this framework for networks that are more nested than expected $A \to 0$, whilst networks which are less nested than expected then $A \to +\infty$. In networks where observed nestedness is similar to that which is expected $A \to 1$.

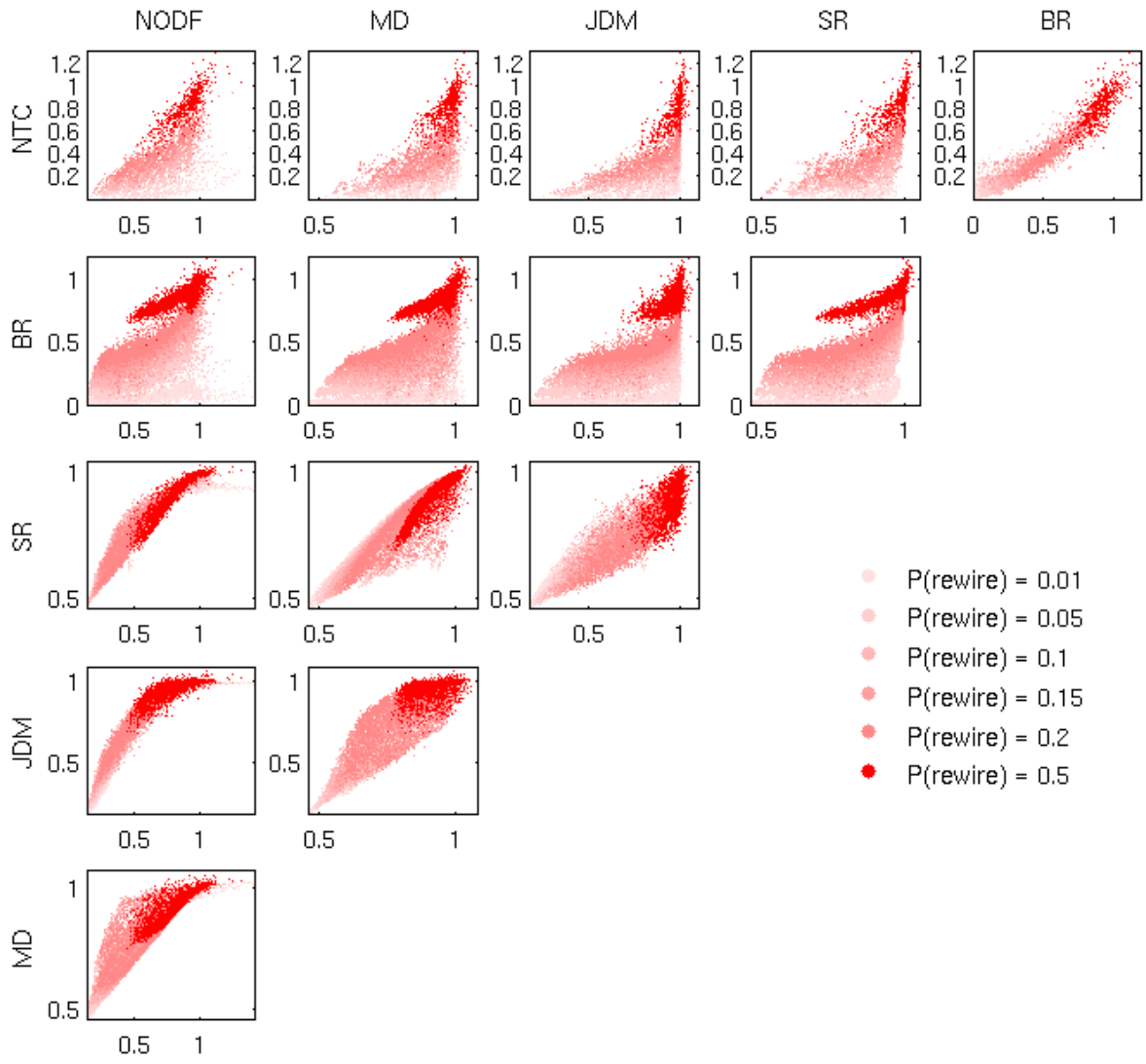


Figure S17: Distribution of adjusted normalised temperature scores in the SS null model

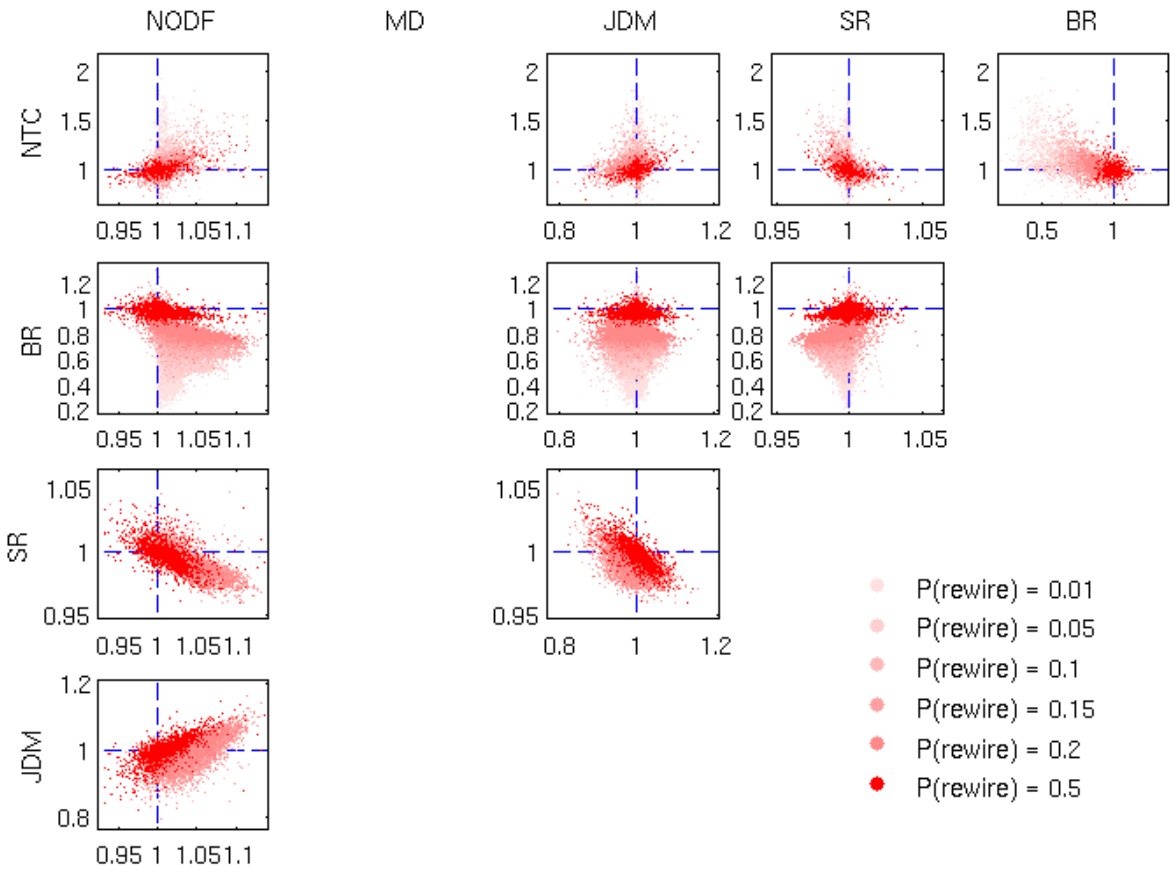


Figure S18: Distribution of adjusted normalised temperature scores in the FF null model

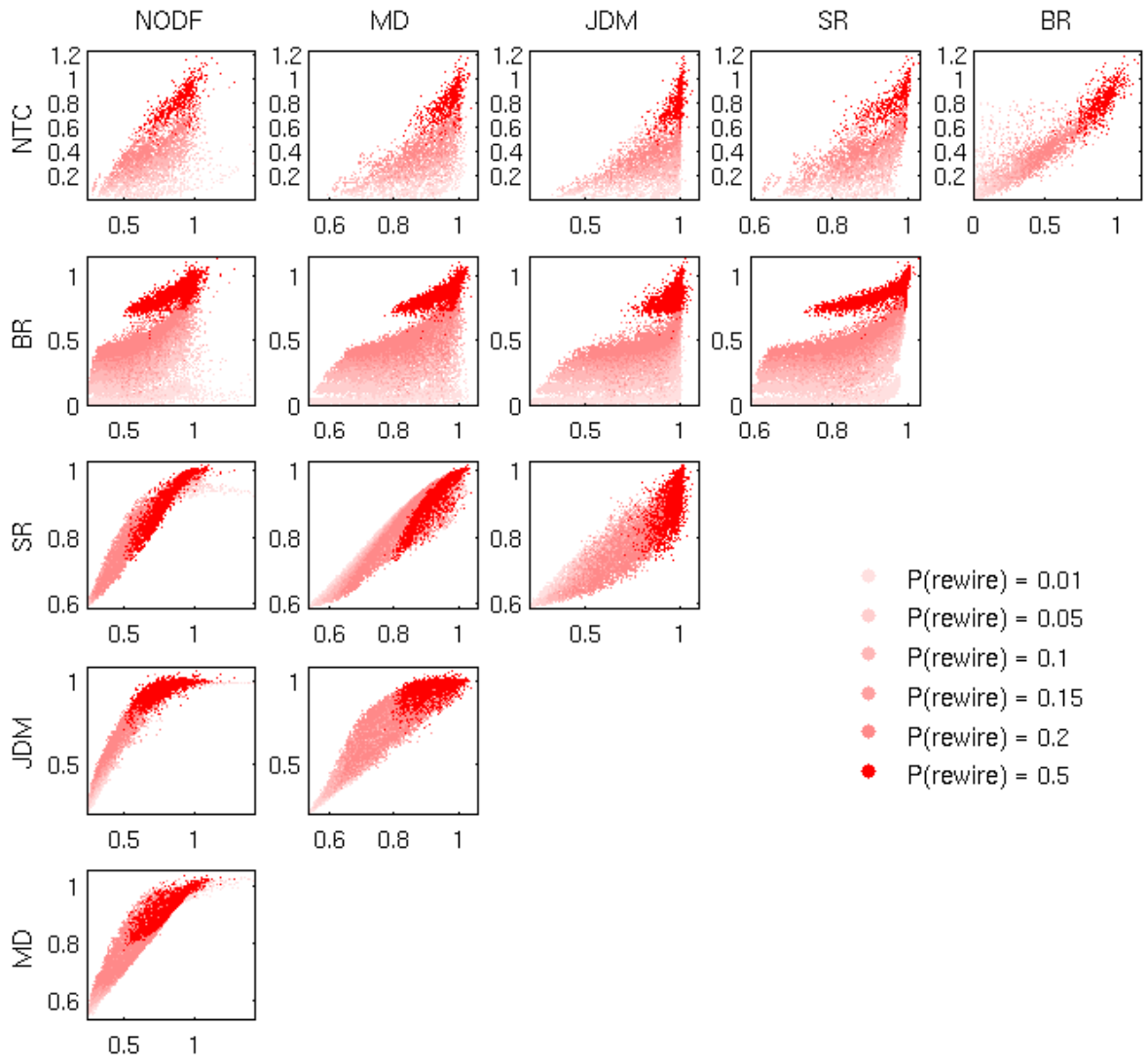


Figure S19: Distribution of adjusted normalised temperature scores in the CC null model

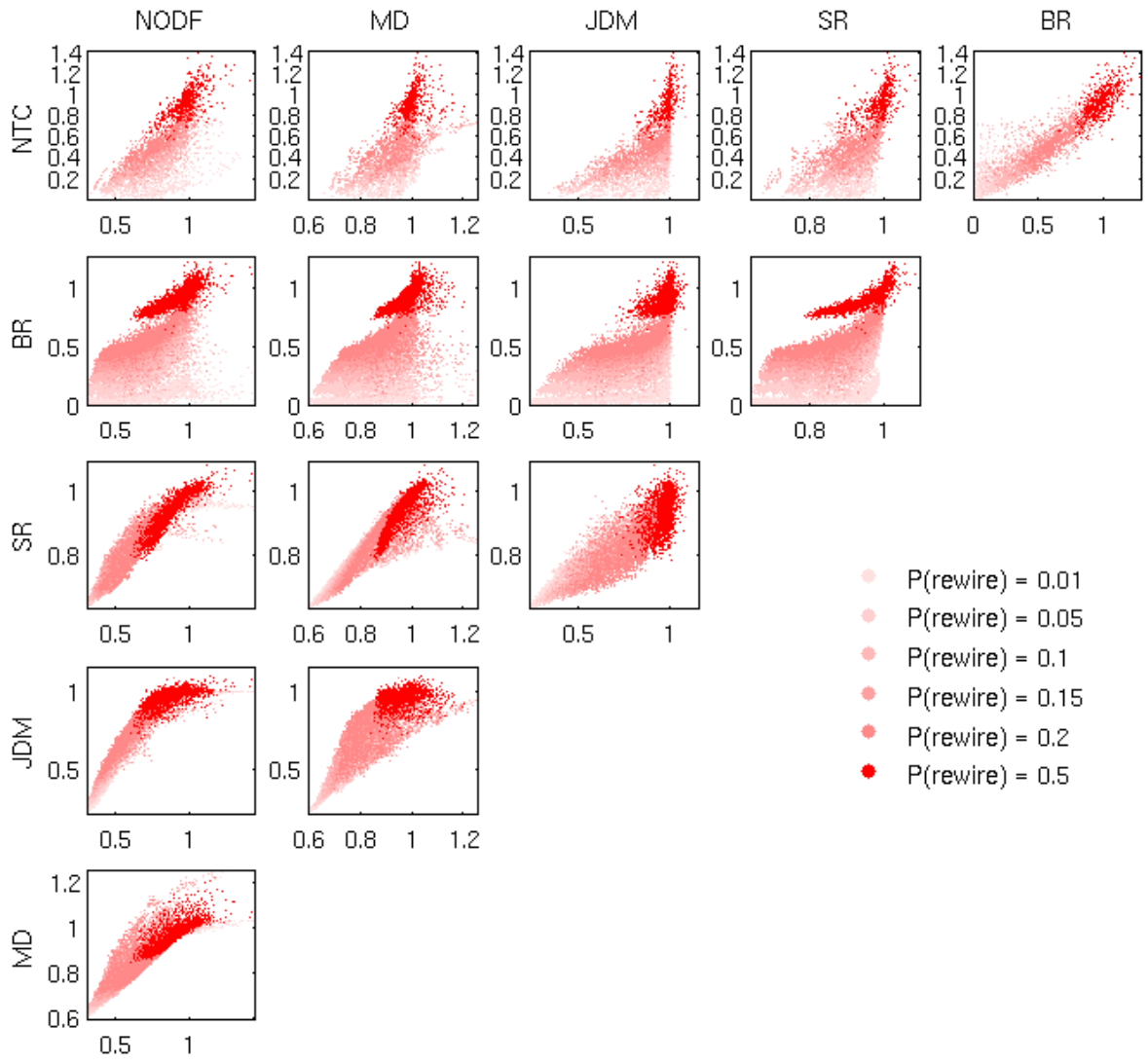


Figure S20: Distribution of adjusted normalised temperature scores in the DD null model (Also shown in Figure 5.6).

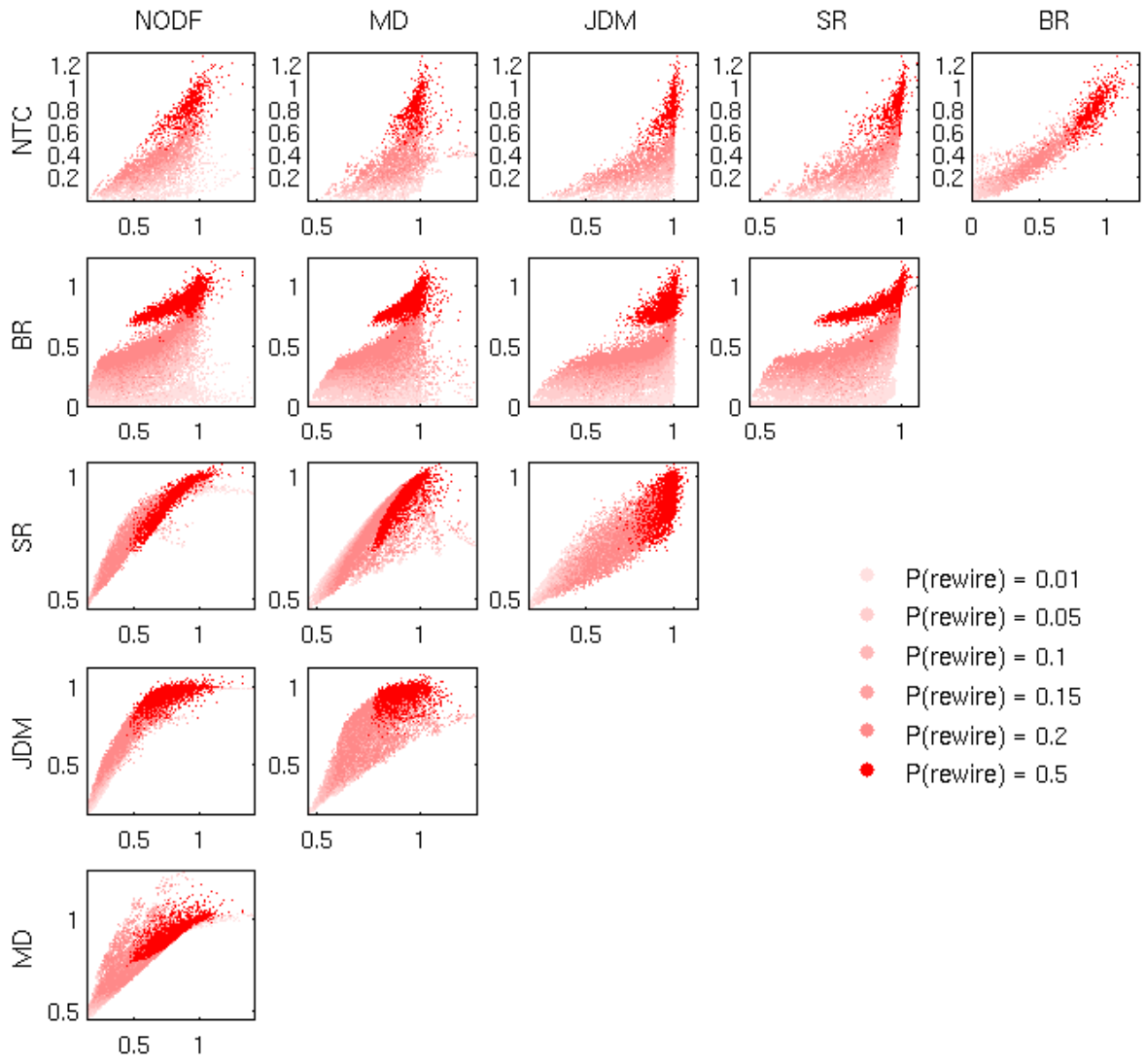


Figure S21: Distribution of adjusted normalised temperature scores in the EE null model

Metrics for Effect Size comparisons

Figure S22 shows a recalculation of Figure 5.7 on the NTC network sample to show the robustness of our result and where the NTC measure appears to fall.

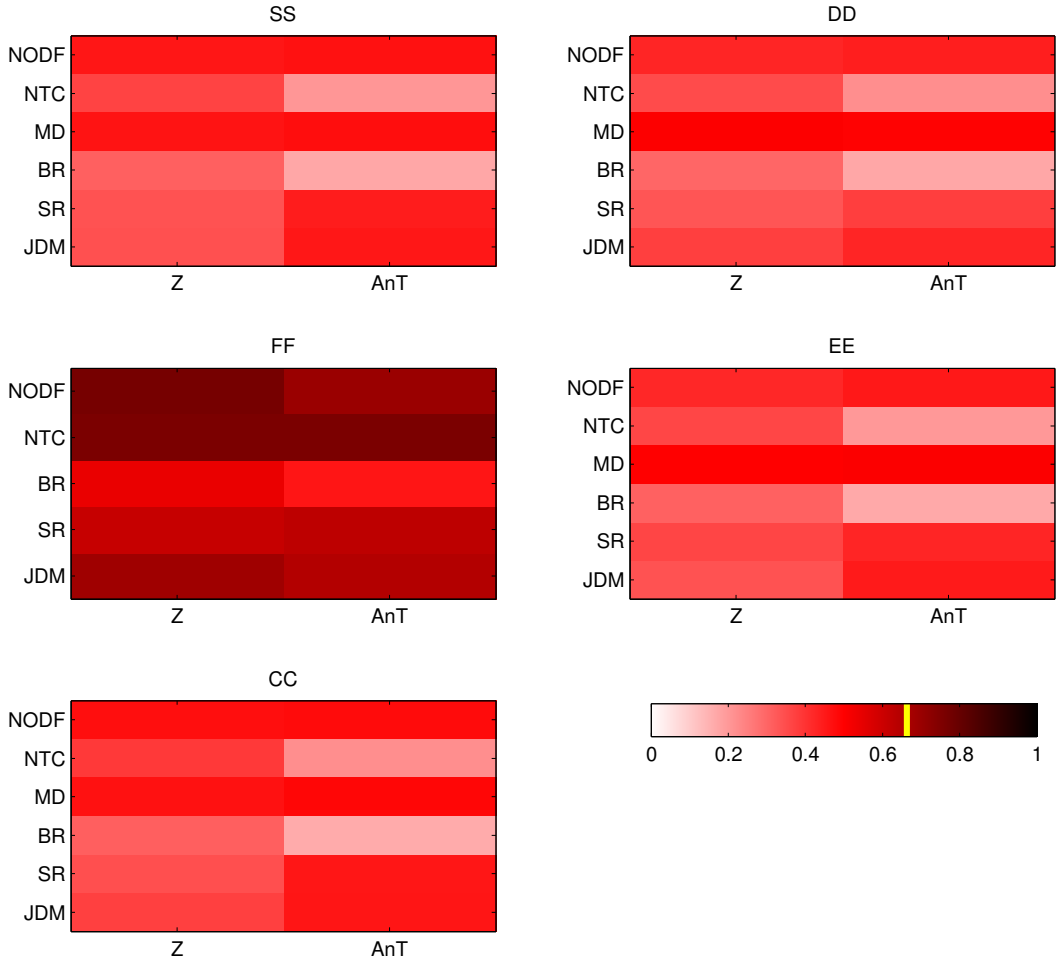


Figure S22: Comparative ability of effect sizes calculated from each null model for each nestedness measure by computing normalised average absolute displacement (NAAD) scores between observed $P(rewire)$ and expected $P(rewire)$ using the NTC network sample. Each subplot shows results under each null model, where nestedness measures are across the rows and different effect sizes are made across the columns. Effect sizes used were sample z-scores (Z) and adjusted normalised temperature scores (AnT). As an effect size for Manhattan distance (MD) cannot be calculated using the FF null model, so this was not shown in the figure. The expected NAAD score for a random effect size is marked on the colourbar(10,000 random samples gave a distribution of NAAD scores with mean=0.663 and standard deviation=0.0065).

Appendix F: Effect size and network properties

Three network properties; connectance, fill and size; were investigated in terms of how they may effect the results for a chosen effect size, using a particular nestedness measure in a particular null model. The graphical results for these comparisons are shown for sample z-scores (figures S23-S24), and for adjusted normalised temperature scores (figures S25-S26).

Z-scores

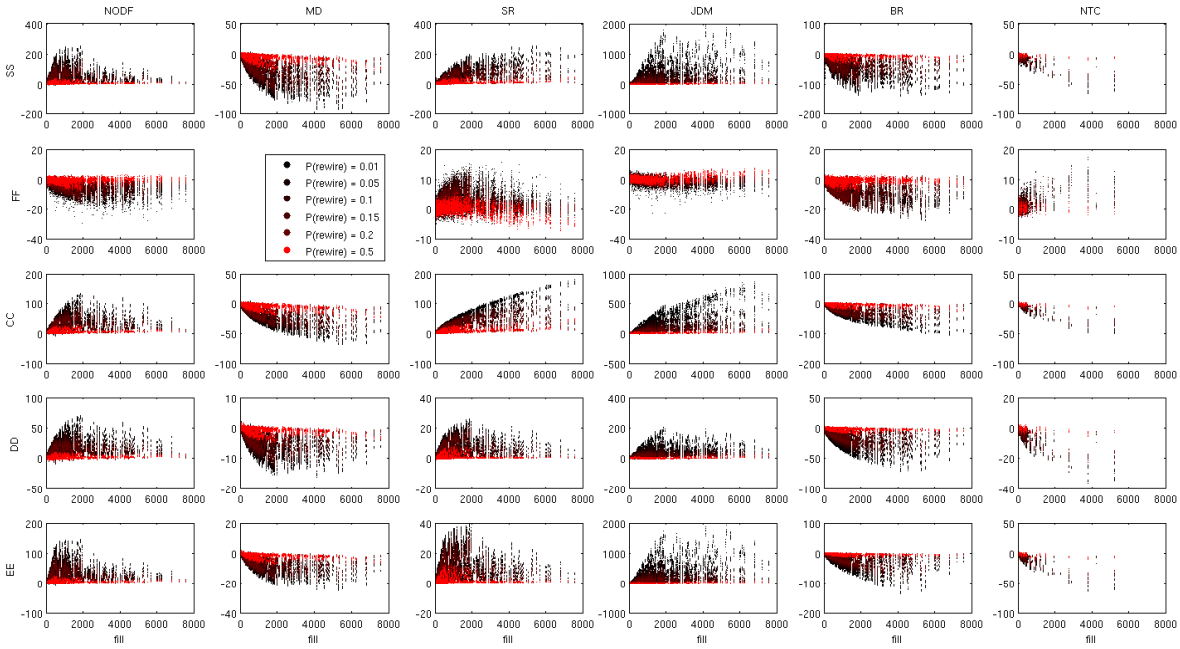


Figure S23: Z-score variation with fill. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

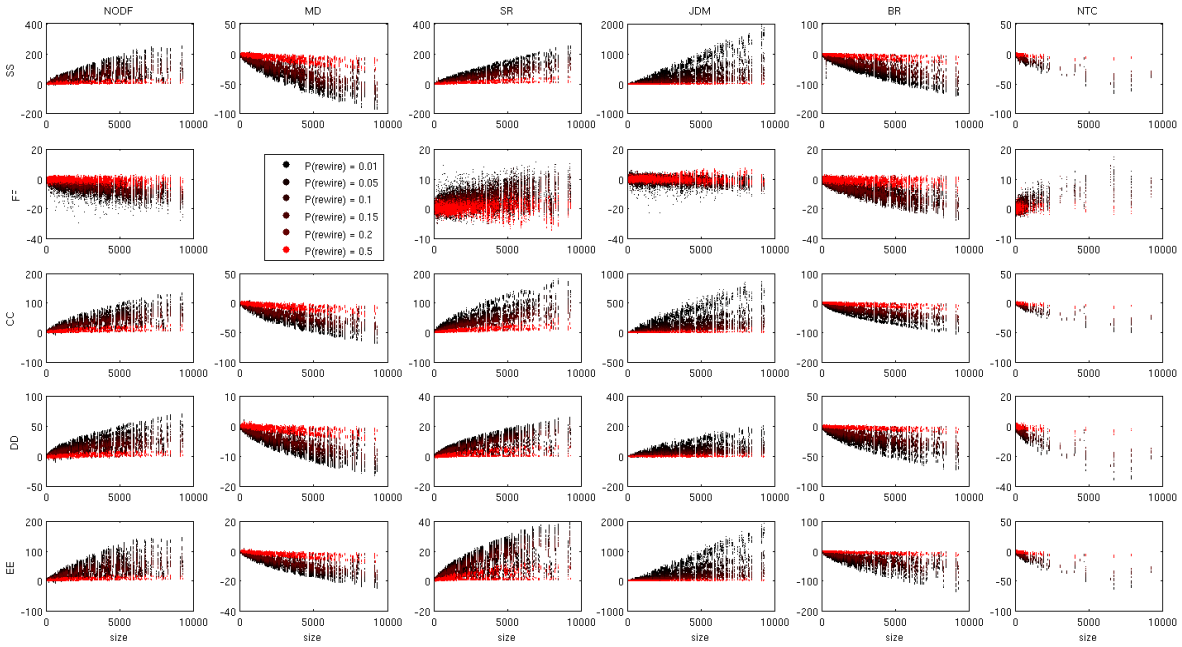


Figure S24: Z-score variation with size. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

Adjusted normalised temperature

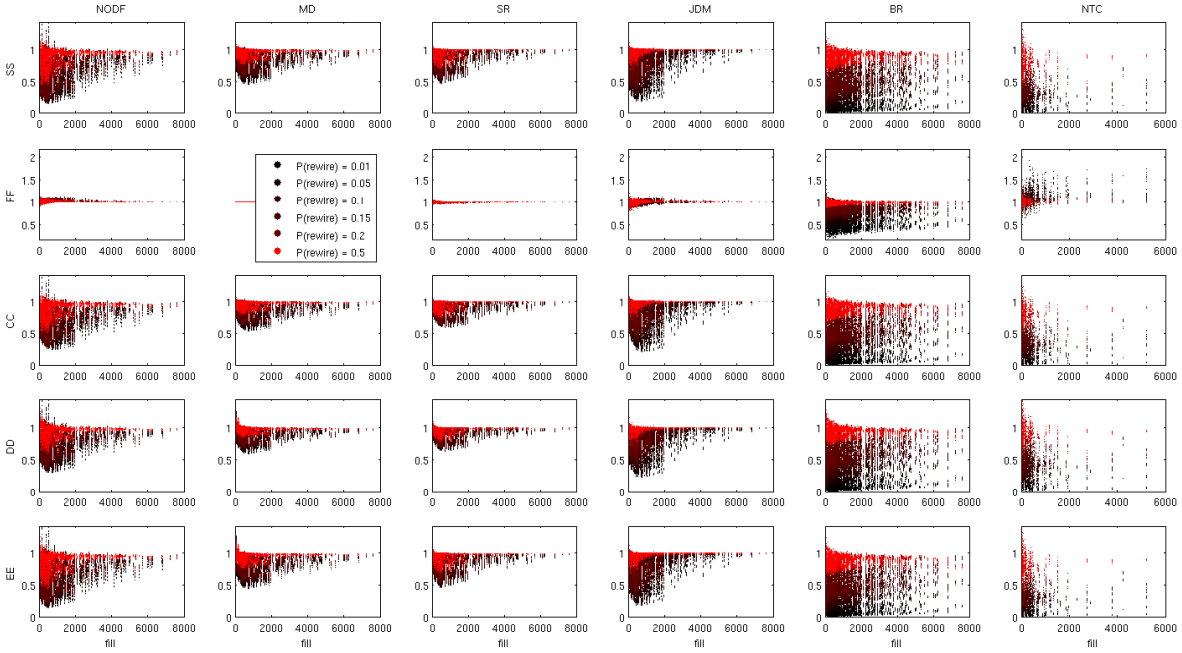


Figure S25: Adjusted normalised temperature variation with fill. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

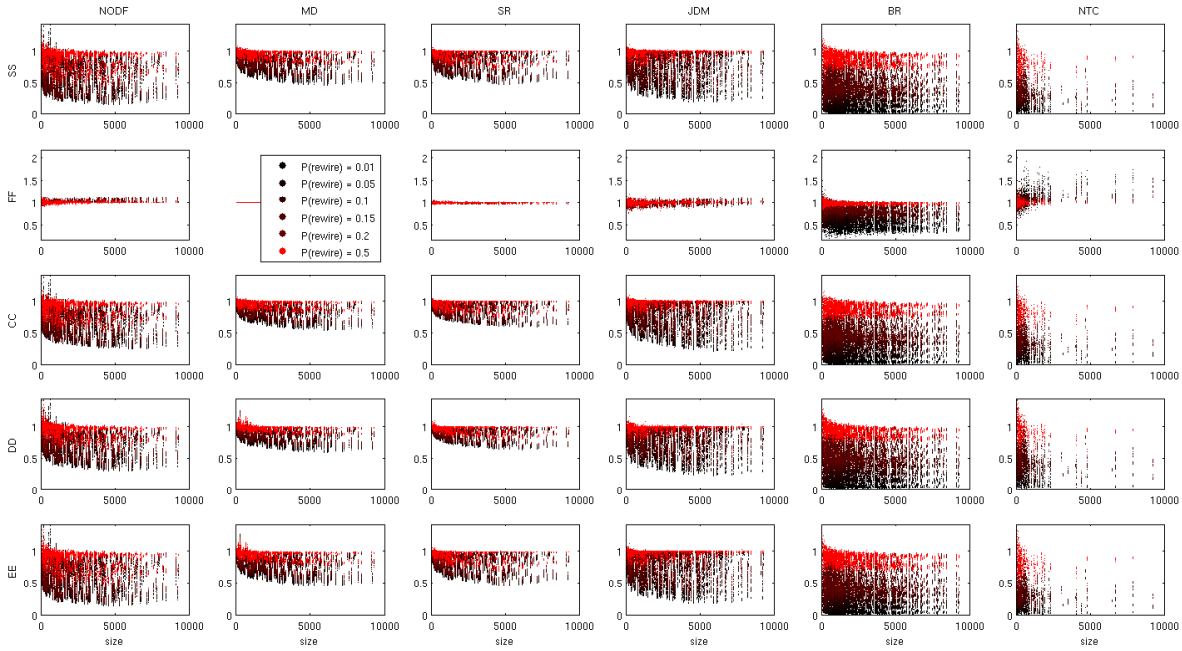


Figure S26: Adjusted normalised temperature variation with size. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

Appendix G: Standardising the standardisation

Figures in Appendix F: Effect size and network properties showed that the effect sizes from both z-scores and adjusted normalised temperature scores were sensitive to (and hence biased by) network properties. We ask is there a way to account for this? Is there a way to standardise the effect size scores?

For each unique network property value (fill,connectance,size) we have effect sizes collated from a range of different networks, which have different rewiring levels. We can transform the range of these scores at each unique network property value onto the range [0,1] by first subtracting the minimum effect size score, then dividing this value by the range of effect size scores (maximum minus the minimum effect size score) found at this network property value. Performing this operation across each of the unique property values gives us the information shown in figures S27-S30. These figures show that measures all exhibit similar qualitative patterns of sensitivity across network properties. None of the nestedness measures in the null models that we have tested are able to avoid bias from network properties and this appears to be a systematic bias that future studies will need to address.

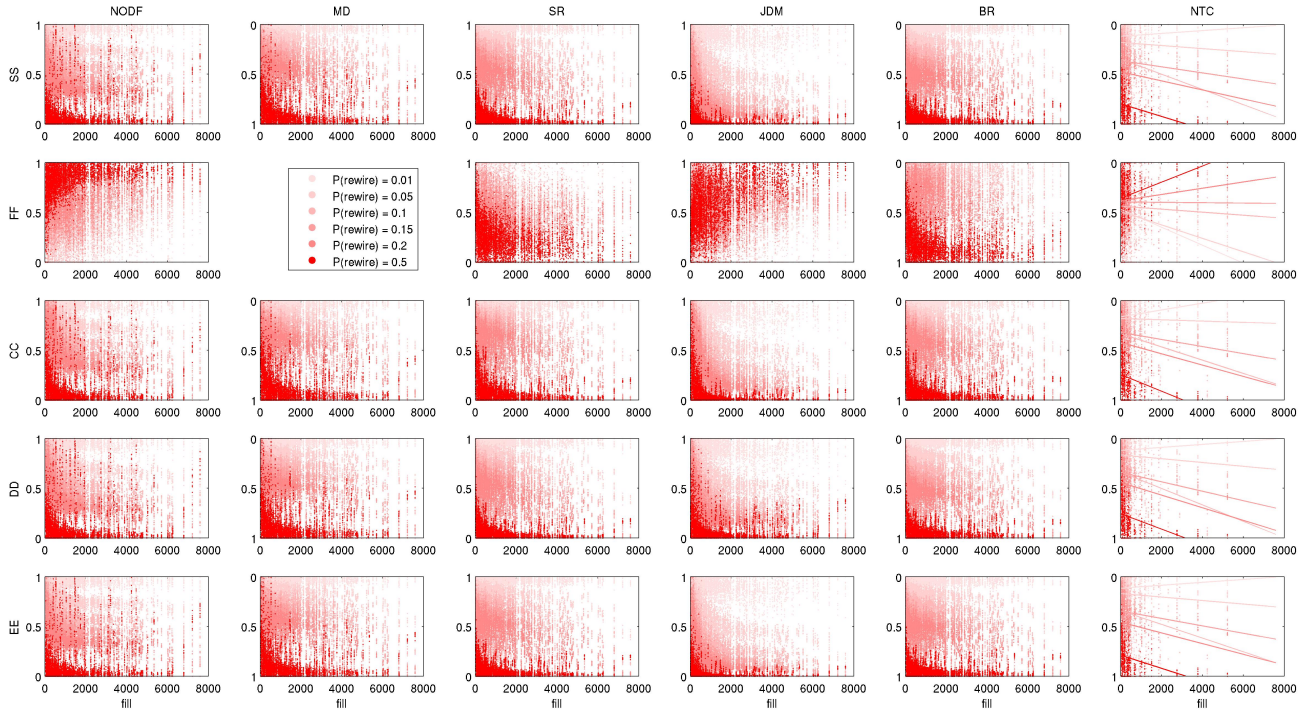


Figure S27: Standardised z-scores variation with fill. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

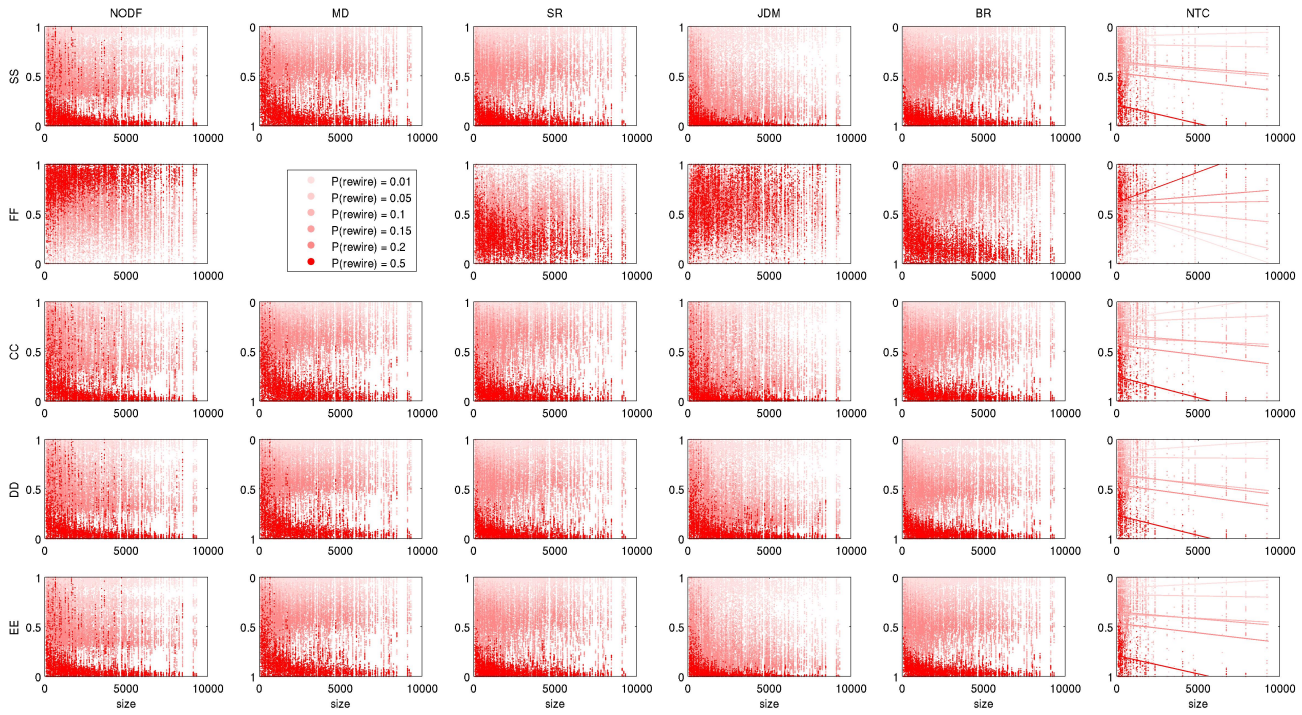


Figure S28: Standardised z-scores variation with size. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

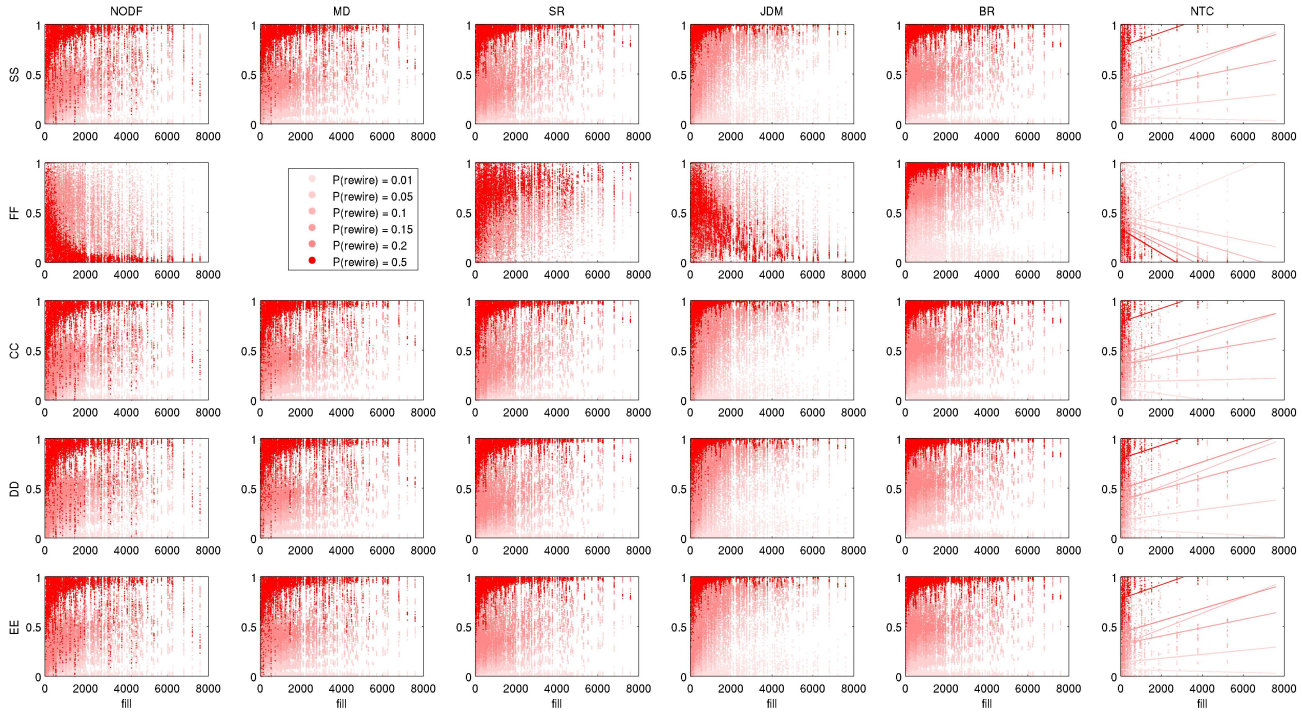


Figure S29: Standardised adjusted normalised temperature variation with fill. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

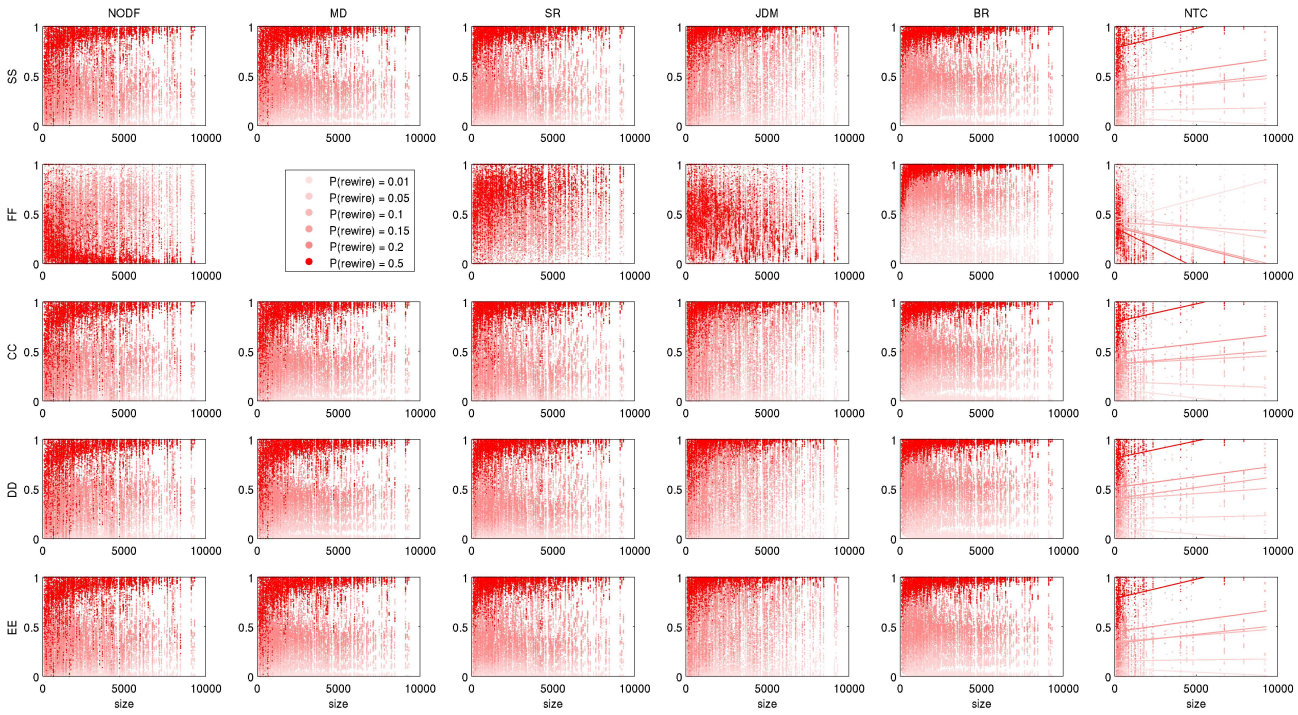


Figure S30: Standardised adjusted normalised temperature variation with size. Each of the rows represents a different null model, whilst each column of subplots represents a different nestedness measure.

Appendix H: Properties of weighted plant-pollinator networks

Network	Rows	Columns	Edges	Fill	Reference
Safariland	9	27	39	1130	[179, 180, 181]
barrett1987	12	102	167	550	[182]
bezerra2009	13	13	71	28224	[183]
elberling1999	23	118	238	383	[184]
inouye1988	41	83	268	1459	[185]
junker2013	56	257	572	3053	[186]
kato1990	91	679	1206	2392	[187]
kevan1970	30	114	312	2523	[188]
memmott1999	25	79	299	2183	[189]
mosquin1967	11	18	38	134	[190]
motten1982	13	44	143	2225	[191]
olesen2002aigrettes	14	13	52	1512	[192]
olesen2002flores	10	12	30	1139	[192]
ollerton2003	9	56	103	594	[193]
schemske1978	7	32	59	299	[194]
small1976	13	34	141	992	[195]
vazarr	10	29	43	515	[179, 180, 181]
vazcer	9	33	45	613	[179, 180, 181]
vazllao	10	29	42	677	[179, 180, 181]
vazmasc	8	26	36	286	[179, 180, 181]
vazmasnc	8	35	51	719	[179, 180, 181]
vazquec	8	27	47	592	[179, 180, 181]
vazquenc	7	24	31	761	[179, 180, 181]

Table S2: Network properties of the datasets used in Chapter 6

Appendix I: Comparing weighted modularity algorithm robustness

Network	QuanBiMo R	$\tilde{x}$	U	F	Q'_R	LPAwb+ R	$\tilde{x}$	U	F	Q'_R	Exhaustive LPAwb+ R	$\tilde{x}$	U	F	Q'_R
Safariland	89	0.558	1	0	0.538	100	0.519	2	0	0.519	40	0.554	1	0	0.538
barrett1987	1	0.077	1	0	0.593	98	0.470	1	0	0.470	1	0.481	1	0	0.317
bezerra2009	73	0.230	1	0	0.155	100	0.218	1	0	0.218	52	0.230	1	0	0.155
elberling1999	1	0.143	1	5	0.311	100	0.458	2	0	0.458	1	0.484	1	0	0.286
inouye1988	1	0.395	1	0	0.239	51	0.351	1	0	0.351	1	0.404	1	0	0.082
junker2013	1	0.024	1	0	0.619	44	0.430	27	0	0.433	1	0.479	1	0	0.112
kato1990	1	0.006	1	0	0.945	92	0.544	85	0	0.544	1	0.574	1	0	0.279
kevan1970	1	0.312	1	1	0.276	6	0.340	4	0	0.341	1	0.422	1	0	0.276
memmott1999	1	0.290	1	0	0.124	57	0.268	8	0	0.268	1	0.328	1	0	0.097
mosquin1967	64	0.479	1	0	0.368	100	0.393	1	0	0.393	25	0.470	1	0	0.368
motten1982	6	0.304	1	0	-0.049	100	0.281	1	0	0.281	8	0.304	1	0	-0.049
olesen2002aigrettes	19	0.334	1	0	0.269	98	0.314	1	0	0.314	80	0.340	1	0	0.269
olesen2002flores	24	0.441	1	0	0.467	98	0.422	2	0	0.422	61	0.444	1	0	0.467
ollerton2003	1	0.302	1	3	0.223	43	0.418	1	0	0.439	9	0.439	1	0	0.223
schemske1978	53	0.370	1	0	0.119	100	0.370	1	0	0.370	100	0.370	1	0	0.119
small1976	9	0.256	1	0	0.007	100	0.242	1	0	0.242	13	0.262	1	0	0.007
vazarr	100	0.542	1	0	0.535	100	0.512	1	0	0.512	17	0.535	1	0	0.535
vazcer	28	0.547	1	0	0.644	100	0.565	1	0	0.565	73	0.619	1	0	0.644
vazllao	100	0.576	1	0	0.619	82	0.550	2	0	0.550	39	0.570	1	0	0.619
vazmasc	100	0.547	2	0	0.556	100	0.522	1	0	0.522	48	0.546	2	0	0.556
vazmasnc	14	0.526	1	0	0.451	100	0.512	2	0	0.512	8	0.521	1	0	0.451
vazquec	26	0.488	1	0	0.532	100	0.474	1	0	0.474	73	0.497	1	0	0.532
vazquenc	100	0.549	1	0	0.677	100	0.514	1	0	0.514	74	0.549	1	0	0.677

Table S3: Extra results from the evaluations of the binary version of these networks. R is the number of times that the best partitions (with highest Q_B) were found from the 100 tests, $\tilde{x}$ is the median Q_B score, U is the number of unique configurations found with the maxmium Q_B score (for each method) judged by comparing the normalised mutual information of partitions sharing this value, F is number of times that the algorithms reported a failure (from the 100 runs) and Q'_R is the realised modularity of the partition with highest Q_B score (for each method). Numbers have been rounded to 3 d.p.

Network	QuanBiMo R	$\tilde{x}$	U	F	Q'_R	LPAwb+ R	$\tilde{x}$	U	F	Q'_R	Exhaustive LPAwb+ R	$\tilde{x}$	U	F	Q'_R
Safariland	91	0.430	1	0	0.979	100	0.427	1	0	0.963	42	0.430	1	0	0.979
barrett1987	1	0.068	1	0	0.836	100	0.567	1	0	0.560	11	0.568	1	0	0.535
bezerra2009	21	0.222	1	0	-0.139	100	0.223	1	0	-0.139	100	0.223	1	0	-0.139
elberling1999	1	0.131	1	3	0.530	100	0.493	4	0	0.180	1	0.507	1	0	0.311
inouye1988	1	0.486	1	0	0.565	100	0.582	1	0	0.406	1	0.609	1	0	0.579
junker2013	1	0.007	1	0	0.743	100	0.533	1	0	0.452	1	0.559	1	0	0.590
kato1990	1	0.006	1	0	0.903	100	0.611	1	0	0.355	1	0.621	1	0	0.431
kevan1970	1	0.247	1	0	0.739	100	0.525	1	0	0.583	7	0.535	1	0	0.675
memmott1999	1	0.127	1	0	0.532	100	0.297	1	0	0.132	2	0.304	1	0	0.306
mosquin1967	78	0.444	1	0	0.478	100	0.440	1	0	0.403	89	0.444	1	0	0.478
motten1982	16	0.354	1	0	0.355	100	0.367	1	0	0.212	100	0.382	1	0	0.355
olesen2002aigrettes	96	0.259	1	0	0.148	100	0.259	1	0	0.148	100	0.259	1	0	0.148
olesen2002flores	67	0.497	1	0	0.403	100	0.497	1	0	0.403	100	0.497	1	0	0.403
ollerton2003	1	0.153	1	2	0.498	100	0.395	1	0	0.431	98	0.413	1	0	0.498
schemske1978	5	0.238	1	0	0.378	100	0.320	1	0	0.378	100	0.320	1	0	0.378
small1976	33	0.526	1	0	0.381	100	0.516	1	0	0.260	1	0.517	1	0	0.337
vazarr	21	0.428	1	0	0.456	100	0.441	1	0	0.449	93	0.442	1	0	0.456
vazcer	30	0.481	1	0	0.869	100	0.591	1	0	0.830	80	0.604	1	0	0.869
vazllao	100	0.561	1	0	0.625	100	0.558	1	0	0.586	61	0.561	1	0	0.635
vazmasc	31	0.656	1	0	0.769	100	0.655	1	0	0.727	80	0.663	1	0	0.769
vazmasnc	26	0.201	1	0	0.499	100	0.400	1	0	0.497	31	0.401	1	0	0.499
vazquec	56	0.511	1	0	0.581	100	0.504	1	0	0.544	22	0.508	1	0	0.581
vazquenc	100	0.450	1	0	0.963	100	0.450	1	0	0.963	100	0.450	1	0	0.963

Table S4: Extra results from the evaluations of the weighted version of these networks. R is the number of times that the best partitions (with highest Q_W) were found from the 100 tests, $\tilde{x}$ is the median Q_W score, U is the number of unique configurations found with the maxmium Q_W score (for each method) judged by comparing the normalised mutual information of partitions sharing this value, F is number of times that the algorithms reported a failure (from the 100 runs) and Q'_R is the realised modularity of the partition with highest Q_W score (for each method). Numbers have been rounded to 3 d.p.

Bibliography

[1] 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), 9383–9387.

[2] Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences* **104**(50), 19891–19896.

[3] Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., and Weitz, J. S. 2011. Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences* **108**(28), E288–E297.

[4] Vázquez, D. P. and Aizen, M. A. 2004. Asymmetric specialization: a pervasive feature of plant-pollinator interactions. *Ecology* **85**(5), 1251–1257.

[5] Corso, G. and Britton, N. 2014. The puzzling affinity between modularity and dependence asymmetry. *Ecological Complexity* **20**, 195–200.

[6] Dalsgaard, B., Trøjelsgaard, K., Martín González, A. M., Nogués-Bravo, D., Ollerton, J., Petanidou, T., Sandel, B., Schleuning, M., Wang, Z., Rahbek, C., Sutherland, W. J., Svenning, J.-C., and Olesen, J. M. 2013. Historical climate-change influences modularity and nestedness of pollination networks. *Ecography* **36**(12), 1331–1340.

[7] Matthews, T. J., Cottee-Jones, H. E. W., and Whittaker, R. J. 2015. Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets. *Diversity and Distributions* **21**(4), 392–404.

[8] Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B., and Bascompte, J. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458**(7241), 1018–1020.

[9] Thébault, E. and Fontaine, C. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**(5993), 853–856.

[10] Saavedra, S., Stouffer, D. B., Uzzi, B., and Bascompte, J. 2011. Strong contributors to network persistence are the most vulnerable to extinction. *Nature* **478**(7368), 233–235.

[11] Bustos, S., Gomez, C., Hausmann, R., and Hidalgo, C. A. 2012. The dynamics of nestedness predicts the evolution of industrial ecosystems. *PloS one* **7**(11), e49393.

[12] Suweis, S., Simini, F., Banavar, J. R., and Maritan, A. 2013. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature* **500**(7463), 449–452.

[13] Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B., and Hochberg, M. E. 2013. Phage–bacteria infection networks. *Trends in Microbiology* **21**(2), 82–91.

[14] Pawar, S. 2014. Why are plant-pollinator networks nested? *Science* **345**(6195), 383–383.

[15] Ulrich, W., Almeida-Neto, M., and Gotelli, N. J. 2009. A consumer's guide to nestedness analysis. *Oikos* **118**(1), 3–17.

[16] Patterson, B. D. and Atmar, W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society* **28**(1-2), 65–82.

[17] Cutler, A. 1991. Nested faunas and extinction in fragmented habitats. *Conservation Biology* **5**(4), 496–504.

[18] Wright, D. H. and Reeves, J. H. 1992. On the meaning and measurement of nestedness of species assemblages. *Oecologia* **92**(3), 416–428.

[19] Wright, D. H., Patterson, B. D., Mikkelson, G. M., Cutler, A., and Atmar, W. 1998. A comparative analysis of nested subset patterns of species composition. *Oecologia* **113**(1), 1–20.

[20] Lomolino, M. V. 1996. Investigating causality of nestedness of insular communities: selective immigrations or extinctions? *Journal of Biogeography* **23**(5), 699–703.

[21] Brualdi, R. A. and Sanderson, J. G. 1999. Nested species subsets, gaps, and discrepancy. *Oecologia* **119**(2), 256–264.

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

[23] Rodríguez-Gironés, M. A. and Santamaría, L. 2006. A new algorithm to calculate the nestedness temperature of presence–absence matrices. *Journal of Biogeography* **33**(5), 924–935.

[24] Guimarães, P. R. and Guimaraes, P. 2006. Improving the analyses of nestedness for large sets of matrices. *Environmental Modelling & Software* **21**(10), 1512–1513.

[25] Ulrich, W. and Gotelli, N. J. 2007. Null model analysis of species nestedness patterns. *Ecology* **88**(7), 1824–1831.

[26] Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. *vegan: Community Ecology Package*, 2013. R package version 2.0-10.

[27] Hausdorf, B. and Hennig, C. 2003. Nestedness of north-west european land snail ranges as a consequence of differential immigration from pleistocene glacial refuges. *Oecologia* **135**(1), 102–109.

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

[29] Podani, J. and Schmera, D. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos* **120**(11), 1625–1638.

[30] Podani, J. and Schmera, D. 2012. A comparative evaluation of pairwise nestedness measures. *Ecography* **35**(10), 889–900.

[31] Lee, D.-S., Maeng, S. E., and Lee, J. W. 2012. Scaling of nestedness in complex networks. *Journal of the Korean Physical Society* **60**(4), 648–656.

[32] Johnson, S., Domínguez-García, V., and Muñoz, M. A. 2013. Factors determining nestedness in complex networks. *PloS One* **8**(9), e74025.

[33] Corso, G. and Britton, N. F. 2012. Nestedness and τ-temperature in ecological networks. *Ecological Complexity* **11**, 137–143.

[34] Galeano, J., Pastor, J. M., and Iriondo, J. M. 2009. Weighted-interaction nestedness estimator (WINE): a new estimator to calculate over frequency matrices. *Environmental Modelling & Software* **24**(11), 1342–1346.

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

[36] Ulrich, W. and Almeida-Neto, M. 2012. On the meanings of nestedness: back to the basics. *Ecography* **35**(10), 865–871.

[37] Saavedra, S. and Stouffer, D. B. 2013. "Disentangling nestedness" disentangled. *Nature* **500**(7463), E1–E2.

[38] Almeida-Neto, M. and Ulrich, W. 2011. A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling & Software* **26**(2), 173–178.

[39] Podani, J., Ricotta, C., and Schmera, D. 2013. A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. *Ecological Complexity* **15**, 52–61.

[40] Joppa, L. N., McInerny, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., Gavaghan, D., and Emmott, S. 2013. Troubling trends in scientific software use. *Science* **340**(6134), 814–815.

[41] Newman, M. E. J. and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical review E* **69**(2), 026113.

[42] Leger, J.-B., Vacher, C., and Daudin, J.-J. 2014. Detection of structurally homogeneous subsets in graphs. *Statistics and computing* **24**(5), 675–692.

[43] Fortunato, S. 2010. Community detection in graphs. *Physics Reports* **486**(3), 75–174.

[44] Barber, M. J. 2007. Modularity and community detection in bipartite networks. *Physical Review E* **76**(6), 066102.

[45] Guimerà, R., Sales-Pardo, M., and Amaral, L. A. N. 2007. Module identification in bipartite and directed networks. *Physical Review E* **76**(3), 036102.

[46] Murata, T. Modularity for bipartite networks. In Data Mining for Social Network Data, Memon, N., Xu, J. J., Hicks, D. L., and Chen, H., editors, volume 12 of *Annals of Information Systems*, 109–123. Springer US 2010.

[47] Thébault, E. 2013. Identifying compartments in presence–absence matrices and bipartite networks: insights into modularity measures. *Journal of Biogeography* **40**(4), 759–768.

[48] Dormann, C. F. and Strauss, R. 2014. A method for detecting modules in quantitative bipartite networks. *Methods in Ecology and Evolution* **5**(1), 90–98.

[49] Liu, X. and Murata, T. 2010. An efficient algorithm for optimizing bipartite modularity in bipartite networks. *JACIII* **14**(4), 408–415.

[50] Good, B. H., de Montjoye, Y.-A., and Clauset, A. 2010. Performance of modularity maximization in practical contexts. *Physical Review E* **81**(4), 046106.

[51] Ulrich, W. Nestedness–a FORTRAN program for measuring order and disorder in ecological communities. .

[52] Strona, G., Galli, P., Seveso, D., Montano, S., and Fattorini, S. 2014. Nestedness for dummies (ned): a user friendly web interface for exploratory nestedness analysis. *Journal of Statistical Software* **59**, 1–9.

[53] Dormann, C. F., Gruber, B., and Fründ, J. 2008. Introducing the bipartite package: analysing ecological networks. *R News* **8**(2), 8–11.

[54] Flores, C. O., Poisot, T., and Weitz, J. S. 2014. BiMAT: a MATLAB ® package to facilitate the analysis and visualization of bipartite networks. *arXiv preprint arXiv:1406.6732* .

[55] Marquitti, F. M. D., Guimarães, P. R., Pires, M. M., and Bittencourt, L. F. 2014. MODULAR: software for the autonomous computation of modularity in large network sets. *Ecography* **37**(3), 221–224.

[56] Guimarães, P. R. and Guimarães, P. 2006. Improving the analyses of nestedness for large sets of matrices. *Environmental Modelling & Software* **21**(10), 1512–1513.

[57] Joppa, L. N., Montoya, J. M., Solé, R., Sanderson, J., and Pimm, S. L. 2010. On nestedness in ecological networks. *Evolutionary Ecology Research* **12**, 35–46.

[58] Strona, G., Nappo, D., Boccacci, F., Fattorini, S., and San-Miguel-Ayanz, J. 2014. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications* **5**.

[59] Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**(6736), 541–548.

[60] Wilhelm, S. W. and Suttle, C. A. 1999. Viruses and nutrient cycles in the sea. *Bioscience* **49**(10), 781–788.

[61] Weinbauer, M. G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28**(2), 127–181.

[62] Suttle, C. A. 2005. Viruses in the sea. *Nature* **437**(7057), 356–361.

[63] Suttle, C. A. 2007. Marine viruses – major players in the global ecosystem. *Nature Reviews Microbiology* **5**(10), 801–812.

[64] Brussaard, C. P. D., Wilhelm, S. W., Thingstad, F., Weinbauer, M. G., Bratbak, G., Heldal, M., Kimmance, S. A., Middelboe, M., Nagasaki, K., Paul, J. H., et al. 2008. Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME Journal* **2**(6), 575–578.

[65] Weitz, J. S. and Wilhelm, S. W. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports* **4**(17).

[66] Bohannan, B. J. M. and Lenski, R. E. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecology Letters* **3**, 362–377.

[67] Buckling, A. and Rainey, P. B. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**(1494), 931–936.

[68] Lennon, J. T. and Martiny, J. B. H. 2008. Rapid evolution buffers ecosystem impacts of viruses in a microbial food web. *Ecology Letters* **11**(11), 1178–1188.

[69] Marston, M. F., Pierciey, F. J., Shepard, A., Gearin, G., Qi, J., Yandava, C., Schuster, S. C., Henn, M. R., and Martiny, J. B. H. 2012. Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences* **109**(12), 4544–4549.

[70] Brockhurst, M. A., Morgan, A. D., Fenton, A., and Buckling, A. 2007. Experimental coevolution with bacteria and phage: The *Pseudomonas fluorescens*—$\phi 2$ model system. *Infection, Genetics and Evolution* **7**, 547–552.

[71] Forde, S. E., Thompson, J. N., Holt, R. D., and Bohannan, B. J. M. 2008. Coevolution drives temporal changes in fitness and diversity across environments in a bacteria–bacteriophage interaction. *Evolution* **62**(8), 1830–1839.

[72] Agrawal, A. and Lively, C. M. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research* **4**(1), 79–90.

[73] Gandon, S., Buckling, A., Decaestecker, E., and Day, T. 2008. Host–parasite coevolution and patterns of adaptation across time and space. *Journal of Evolutionary Biology* **21**(6), 1861–1866.

[74] Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45**(6), 1320–1328.

[75] Winter, C., Bouvier, T., Weinbauer, M. G., and Thingstad, T. F. 2010. Trade–offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. *Microbiology and Molecular Biology Reviews* **74**(1), 42–57.

[76] Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Beltran Rodriguez-Brito, L. P., Thingstad, T. F., and Forest Rohwer, A. M. 2009. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7**(11), 828–836.

[77] Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy, J., Martin-Cuadrado, A. B., Mira, A., Nulton, J., Pasic, L., Rayhawk, S., Rodriguez-Mueller, J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T. F., Tran, T., Thurber, R. V., Willner, D., Youle, M., and Rohwer, F. 2010. Viral and microbial community dynamics in four aquatic environments. *ISME Journal* **4**(6), 739–751.

[78] Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D. 2011. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**(7353), 604–608.

[79] Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., Ellisman, M., Deerinck, T., Sullivan, M. B., and Giovannoni, S. J. 2013. Abundant SAR11 viruses in the ocean. *Nature* **494**(7437), 357–360.

[80] Flores, C. O., Valverde, S., and Weitz, J. S. 2013. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *ISME Journal* **7**, 520–532.

[81] Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B., and Hochberg, M. E. 2013. Phage-bacteria infection networks. *Trends in Microbiology* **21**(2), 82 – 91.

[82] Moebus, K. and Nattkemper, H. 1981. Bacteriophage sensitivity patterns among bacteria isolated from marine waters. *Helgoländer Meeresuntersuchungen* **34**(3), 375–385.

[83] Weitz, J. S., Hartman, H., and Levin, S. A. 2005. Coevolutionary arms races between bacteria and bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* **102**(27), 9535–9540.

[84] Williams, H. T. P. Coevolving parasites improve host evolutionary search on structured landscapes. In Artificial Life 13: Proceedings of the Thirteenth International Conference on the Simulation and Synthesis of Living Systems, Adami, C., Bryson, D. M., Ofria, C., and Pennock, R. T., editors, 129–136 (MIT Press, Cambridge, MA, 2012).

[85] Williams, H. T. P. 2013. Phage-induced diversification improves host evolvability. *BMC Evolutionary Biology* **13**, 17.

[86] Labrie, S. J., Samson, J. E., and Moineau, S. 2010. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**(5), 317–327.

[87] Levin, B. R., Stewart, F. M., and Chao, L. 1977. Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. *American Naturalist* **111**, 3–24.

[88] Monod, J. 1949. The growth of bacterial cultures. *Annual Reviews in Microbiology* **3**(1), 371–394.

[89] Poisot, T. and Flores, C. 2012, BiWeb: https://github.com/tpoisot/biweb. Technical report.

[90] Liu, X. and Murata, T. Community detection in large-scale bipartite networks. In *Web Intelligence and Intelligent Agent Technologies 2009 (WI-IAT'09)*, volume 1, 50–57. IEEE, 2009.

[91] Edwards, R. A. and Rohwer, F. 2005. Viral metagenomics. *Nature Reviews Microbiology* **3**(6), 504–510.

[92] Stewart, E. J. 2012. Growing unculturable bacteria. *Journal of Bacteriology* **194**(16), 4151–4160.

[93] Jover, L. F., Cortez, M. H., and Weitz, J. S. 2013. Mechanisms of multi–strain coexistence in host–phage systems with nested infection networks. *Journal of Theoretical Biology* **332**, 65 – 77.

[94] Koskella, B. and Meaden, S. 2013. Understanding bacteriophage specificity in natural microbial communities. *Viruses* **5**(3), 806–823.

[95] Quigley, B. J. Z., López, D. G., Buckling, A., McKane, A. J., and Brown, S. P. 2012. The mode of host–parasite interaction shapes coevolutionary dynamics and the fate of host cooperation. *Proceedings of the Royal Society B: Biological Sciences* **279**(1743), 3742–3748.

[96] Scanlan, P. D., Hall, A. R., Lopez-Pascua, L. D. C., and Buckling, A. 2011. Genetic basis of infectivity evolution in a bacteriophage. *Molecular Ecology* **20**(5), 981–989.

[97] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**(5819), 1709–1712.

[98] Stern, A. and Sorek, R. 2011. The phage–host arms race: Shaping the evolution of microbes. *Bioessays* **33**(1), 43–51.

[99] Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S., and Koonin, E. V. 2012. Viral diversity threshold for adaptive immunity in prokaryotes. *mBio* **3**(6), e00456–12.

[100] Seed, K. D., Lazinski, D. W., Calderwood, S. B., and Camilli, A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**(7438), 489–491.

[101] James, A., Pitchford, J. W., and Plank, M. J. 2012. Disentangling nestedness from models of ecological complexity. *Nature* **487**(7406), 227–230.

[102] Beckett, S. J. and Williams, H. T. P. 2013. Coevolutionary diversification creates nested-modular structure in phage–bacteria interaction networks. *Interface Focus* **3**(6), 20130033.

[103] McQuaid, C. F. and Britton, N. F. 2013. Host-parasite nestedness: A result of co-evolving trait-values. *Ecological Complexity* **13**, 53 – 59.

[104] Lever, J. J., Nes, E. H., Scheffer, M., and Bascompte, J. 2014. The sudden collapse of pollinator communities. *Ecology Letters* **17**(3), 350–359.

[105] Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* **81**(9), 2606–2621.

[106] Csermely, P., London, A., Wu, L., and Uzzi, B. 2013. Structure and dynamics of core/periphery networks. *Journal of Complex Networks* **1**(2), 93–123.

[107] Dormann, C. F., Gruber, B., and Fruend, J. 2008. Introducing the bipartite package: Analysing ecological networks. *R News* **8**(2), 8–11.

[108] Beckett, S. J., Boulton, C. A., and Williams, H. T. P. 2014. FALCON: nestedness statistics for bipartite networks. *figshare* .

[109] Hultén, E. *Outline of the history of arctic and boreal biota during the Quaternary period: their evolution during and after the glacial period as indicated by the equiformal progressive areas of present plant species*, volume 1. Bokförlags Aktiebolaget Thule, 1937.

[110] Darlington, P. J. *Zoogeography: the geographical distribution of animals*. Wiley, New York, 1957.

[111] Daubenmire, R. 1975. Floristic plant geography of eastern washington and northern idaho. *Journal of Biogeography* **2**(1), 1–18.

[112] Fraser, C. I., Terauds, A., Smellie, J., Convey, P., and Chown, S. L. 2014. Geothermal activity helps life survive glacial cycles. *Proceedings of the National Academy of Sciences* **111**(15), 201321437.

[113] Baselga, A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography* **21**(12), 1223–1232.

[114] Fortuna, M. A., Stouffer, D. B., Olesen, J. M., Jordano, P., Mouillot, D., Krasnov, B. R., Poulin, R., and Bascompte, J. 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology* **79**(4), 811–817.

[115] Chávez, V. A., Doncaster, C. P., Dearing, J. A., Wang, R., Huang, J., and Dyke, J. G. Detecting regime shifts in artificial ecosystems. In *Advances in Artificial Life, ECAL*, volume 12, 625–632, 2013.

[116] Tinker, M., Guimarães, P. R., Novak, M., Marquitti, F. M. D., Bodkin, J. L., Staedler, M., Bentall, G., and Estes, J. A. 2012. Structure and mechanism of diet specialisation: testing models of individual variation in resource use with sea otters. *Ecology Letters* **15**(5), 475–483.

[117] Piepenbrink, A. and Gaur, A. S. 2013. Methodological advances in the analysis of bipartite networks an illustration using board interlocks in indian firms. *Organizational Research Methods* **16**(3), 474–496.

[118] Melo, A. S., Cianciaruso, M. V., and Almeida-Neto, M. 2014. treenodf: nestedness to phylogenetic, functional and other tree-based diversity metrics. *Methods in Ecology and Evolution* **5**(6), 563–572.

[119] Gotelli, N. J. 2001. Research frontiers in null model analysis. *Global Ecology and Biogeography* **10**(4), 337–343.

[120] Poisot, T. and Gravel, D. 2014. When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ* **2**, e251.

[121] Twitter. Twitter REST API v1.1 Documentation, 2014. Accessed: 30th May 2014.

[122] Williams, H. T. P., Beckett, S. J., O'Neill, S., and Kurz, T. Dynamics of attention and influence in online coverage of the IPCC Fifth Assessment Reports. *in prepartion* .

[123] Borge-Holthoefer, J., Baños, R. A., Gracia-Lázaro, C., and Moreno, Y. 2015. The nested assembly of collective attention in online social systems. *arXiv preprint* , arXiv:1501.06809.

[124] Jover, L. F., Cortez, M. H., and Weitz, J. S. 2013. Mechanisms of multi-strain coexistence in host–phage systems with nested infection networks. *Journal of Theoretical Biology* **332**, 65–77.

[125] Sauve, A., Fontaine, C., and Thébault, E. 2014. Structure–stability relationships in networks combining mutualistic and antagonistic interactions. *Oikos* **123**(3), 378–384.

[126] Chagnon, P.-L. 2015. Characterizing topology of ecological networks along gradients: The limits of metrics' standardization. *Ecological Complexity* **22**, 36–39.

[127] Timi, J. T. and Poulin, R. 2008. Different methods, different results: temporal trends in the study of nested subset patterns in parasite communities. *Parasitology* **135**(01), 131–138.

[128] Strona, G. and Fattorini, S. 2014. On the methods to assess significance in nestedness analyses. *Theory in Biosciences* **133**(3-4), 179–186.

[129] Beckett, S. J., Boulton, C. A., and Williams, H. T. P. 2014. FALCON: a software package for analysis of nestedness in bipartite networks [v1; ref status: indexed, http://f1000r.es/3z8]. *F1000Research* **3**(185).

[130] Beckett, S. J. and Williams, H. T. P. Synthetic matrix ensemble for nestedness analysis. *figshare* , http://dx.doi.org/10.6084/m9.figshare.1320818.

[131] Moore, J. E. and Swihart, R. K. 2007. Toward ecologically explicit null models of nestedness. *Oecologia* **152**(4), 763–777.

[132] Newman, M. E. J. *Networks: an introduction*. Oxford University Press, 2010.

[133] Newman, M. E. J. 2004. Analysis of weighted networks. *Physical Review E* **70**(5), 056131.

[134] Costa, A. and Hansen, P. 2014. A locally optimal hierarchical divisive heuristic for bipartite modularity maximization. *Optimization Letters* **8**(3), 903–917.

[135] Clauset, A., Moore, C., and Newman, M. E. J. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**(7191), 98–101.

[136] Dormann, C., Gruber, B., and Fruend, J. 2008. Introducing the bipartite package: Analysing ecological networks. *R news* **8/2**, 8–11.

[137] Poisot, T. 2013. An a posteriori measure of network modularity [v3; ref status: indexed, http://f1000r.es/2ju]. *F1000Research* **2**(130).

[138] Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. 2005. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment* **2005**(09), P09008.

[139] Miyauchi, A. and Sukegawa, N. 2014. Maximizing Barber's bipartite modularity is also hard. *Optimization Letters* , doi: 10.1007/s11590–014–0818–7.

[140] Gómez, S., Jensen, P., and Arenas, A. 2009. Analysis of community structure in networks of correlated data. *Physical Review E* **80**(1), 016114.

[141] Wright, S. L., Thompson, R. C., and Galloway, T. S. 2013. The physical impacts of microplastics on marine organisms: a review. *Environmental Pollution* **178**, 483–492.

[142] Watts, A. J. R., Lewis, C., Goodhead, R. M., Beckett, S. J., Moger, J., Tyler, C. R., and Galloway, T. S. 2014. Uptake and retention of microplastics by the shore crab *Carcinus maenas*. *Environmental Science & Technology* **48**(15), 8823–8830.

[143] Cowley, L. A., Beckett, S. J., Chase-Toppin, M., Perry, N., Dallman, T. J., Gally, D. L., and Jenkins, C. 2015. Analysis of whole genome sequencing for the *Escherichia coli O157:H7* typing phages. *BMC Genomics* **16**(271).

[144] Zhao, M., Geekiyanage, N., Xu, J., Khin, M. M., Nurdiana, D. R., Paudel, E., and Harrison, R. D. 2015. Structure of the epiphyte community in a tropical montane forest in sw china. *PloS one* **10**(4).

[145] Barnes, N. 2010. Publish your computer code: it is good enough. *Nature* **467**(7317), 753–753.

[146] Easterbrook, S. M. 2014. Open code for open science? *Nature Geoscience* **7**(11), 779–781.

[147] Wilson, G., Aruliah, D., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H., Huff, K. D., Mitchell, I. M., Plumbley, M. D., et al. 2014. Best practices for scientific computing. *PLoS Biology* **12**(1), e1001745.

[148] Fischer, J. and Lindenmayer, D. B. 2002. Treating the nestedness temperature calculator as a “black box” can lead to false conclusions. *Oikos* **99**(1), 193–199.

[149] Reich, V. 2008. CLOCKSS—it takes a community. *The Serials Librarian* **54**(1-2), 135–139.

[150] Sebastián-González, E., Dalsgaard, B., Sandel, B., and Guimarães, P. R. 2015. Macroecological trends in nestedness and modularity of seed-dispersal networks: human impact matters. *Global Ecology and Biogeography* **24**(3), 293–303.

[151] Joppa, L. N. and Williams, R. 2011. The influence of single elements on nested community structure. *Methods in Ecology and Evolution* **2**(5), 541–549.

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

[153] Stouffer, D. B. and Bascompte, J. 2011. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences* **108**(9), 3648–3652.

[154] Flores García, C. O. *Phage–Bacteria Infection networks: from nestedness to modularity and back again*. PhD thesis, Georgia Institute of Technology, USA, 2014.

[155] Feng, W. and Takemoto, K. 2014. Heterogeneity in ecological mutualistic networks dominantly determines community stability. *Scientific Reports* **4**, 5912.

[156] Tang, S., Pawar, S., and Allesina, S. 2014. Correlation between interaction strengths drives stability in large ecological networks. *Ecology Letters* **17**(9), 1094–1100.

[157] Våge, S., Storesund, J. E., and Thingstad, T. F. 2013. Adding a cost of resistance description extends the ability of virus–host model to explain observed patterns in structure and function of pelagic microbial communities. *Environmental Microbiology* **15**(6), 1842–1852.

[158] Haerter, J. O., Mitarai, N., and Sneppen, K. 2014. Phage and bacteria support mutual diversity in a narrowing staircase of coexistence. *ISME Journal* **8**, 2317–2326.

[159] Korytowski, D. A. and Smith, H. L. 2015. How nested and monogamous infection networks in host-phage communities come to be. *Theoretical Ecology* **8**(1), 111–120.

[160] Thingstad, T. F., Pree, B., Giske, J., and Våge, S. 2015. What difference does it make if viruses are strain-, rather than species-specific? *Frontiers in Microbiology* **6**, 320.

[161] Koskella, B. and Brockhurst, M. A. 2014. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS microbiology reviews* **38**(5), 916–931.

[162] Meaden, S., Paszkiewicz, K., and Koskella, B. 2015. The cost of phage resistance in a plant pathogenic bacterium is context-dependent. *Evolution* , doi: 10.1111/evo.12652.

[163] Rappé, M. S. and Giovannoni, S. J. 2003. The uncultured microbial majority. *Annual Reviews in Microbiology* **57**(1), 369–394.

[164] Dang, V. T. and Sullivan, M. B. 2014. Emerging methods to study bacteriophage infection at the single-cell level. *Frontiers in Microbiology* **5**, 724.

[165] Brum, J. R. and Sullivan, M. B. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* **13**(3), 147–159.

[166] Allers, E., Moraru, C., Duhaime, M. B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R., and Sullivan, M. B. 2013. Single-cell and population level viral infection dynamics revealed by phagefish, a method to visualize intracellular and free viruses. *Environmental Microbiology* **15**(8), 2306–2318.

[167] Tazzyman, S. J. and Hall, A. R. 2015. Lytic phages obscure the cost of antibiotic resistance in *Escherichia coli*. *ISME Journal* **9**, 809–820.

[168] Domínguez-García, V. and Muñoz, M. A. 2015. Ranking species in mutualistic networks. *Scientific Reports* **5**, 8182.

[169] Strona, G. and Veech, J. A. 2015. A new measure of ecological network structure based on node overlap and segregation. *Methods in Ecology and Evolution* , doi: 10.1111/2041–210X.12395.

[170] Leger, J.-B., Daudin, J.-J., and Vacher, C. 2015. Clustering methods differ in their ability to detect patterns in ecological networks. *Methods in Ecology and Evolution* **6**(4), 474–481.

[171] Girvan, M. and Newman, M. E. J. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99**(12), 7821–7826.

[172] Mariadassou, M., Robin, S., and Vacher, C. 2010. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics* **4**(2), 715–742.

[173] Newman, M. E. J. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103**(23), 8577–8582.

[174] Podani, J. and Schmera, D. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence – absence data. *Oikos* **120**(11), 1625–1638.

[175] Ulrich, W. and Gotelli, N. J. 2013. Pattern detection in null model analysis. *Oikos* **122**(1), 2–18.

[176] Miklós, I. and Podani, J. 2004. Randomization of presence-absence matrices: comments and new algorithms. *Ecology* **85**(1), 86–92.

[177] Gotelli, N. J. and Ulrich, W. 2011. Over-reporting bias in null model analysis: a response to Fayle and Manica (2010). *Ecological Modelling* **222**(7), 1337–1339.

[178] Fayle, T. M. and Manica, A. 2011. Bias in null model analyses of species co-occurrence: A response to Gotelli and Ulrich (2011). *Ecological Modelling* **222**(7), 1340–1341.

[179] Vázquez, D. P. and Simberloff, D. 2002. Ecological specialization and susceptibility to disturbance: conjectures and refutations. *The American Naturalist* **159**(6), 606–623.

[180] Vázquez, D. P. *Interactions among introduced ungulates, plants, and pollinators: a field study in the temperate forest of the southern Andes*. PhD thesis, University of Tennessee, Knoxville, Tennessee, USA., 2002.

[181] Vázquez, D. P. and Simberloff, D. 2003. Changes in interaction biodiversity induced by an introduced ungulate. *Ecology Letters* **6**(12), 1077–1083.

[182] Barrett, S. C. and Helenurm, K. 1987. The reproductive biology of boreal forest herbs. i. breeding systems and pollination. *Canadian Journal of Botany* **65**(10), 2036–2046.

[183] Bezerra, E. L., Machado, I. C., and Mello, M. A. 2009. Pollination networks of oil-flowers: a tiny world within the smallest of all worlds. *Journal of Animal Ecology* **78**(5), 1096–1101.

[184] Elberling, H. and Olesen, J. M. 1999. The structure of a high latitude plant-flower visitor system: the dominance of flies. *Ecography* **22**(3), 314–323.

[185] Inouye, D. W. and Pyke, G. H. 1988. Pollination biology in the Snowy Mountains of Australia: comparisons with montane Colorado, USA. *Australian Journal of Ecology* **13**(2), 191–205.

[186] Junker, R. R., Blüthgen, N., Brehm, T., Binkenstein, J., Paulus, J., Martin Schaefer, H., and Stang, M. 2013. Specialization on traits as basis for the niche-breadth of flower visitors and as structuring mechanism of ecological networks. *Functional Ecology* **27**(2), 329–341.

[187] Kato, M., Kakutani, T., Inoue, T., and Itino, T. 1990. Insect-flower relationship in the primary beech forest of Ashu, Kyoto: an overview of the flowering phenology and the seasonal pattern of insect visits. *Contributions from the Biological Laboratory, Kyoto University* **27**(4), 309–375.

[188] Kevan, P. G. *High arctic insect-flower visitor relations: the inter-relationships of arthropods and flowers at Lake Hazen, Ellesmere Island, Northwest Territories, Canada*. PhD thesis, University of Alberta, Alberta, Canada., 1970.

[189] Memmott, J. 1999. The structure of a plant-pollinator food web. *Ecology Letters* **2**(5), 276–280.

[190] Mosquin, T. and Martin, J. 1967. Observations on the pollination biology of plants on Melville Island, NWT, Canada. *Canadian Field Naturalist* **81**, 201–205.

[191] Motten, A. F. 1986. Pollination ecology of the spring wildflower community of a temperate deciduous forest. *Ecological Monographs* **56**(1), 21–42.

[192] Olesen, J. M., Eskildsen, L. I., and Venkatasamy, S. 2002. Invasion of pollination networks on oceanic islands: importance of invader complexes and endemic super generalists. *Diversity and Distributions* **8**(3), 181–192.

[193] Ollerton, J., Johnson, S. D., Cranmer, L., and Kellie, S. 2003. The pollination ecology of an assemblage of grassland asclepiads in south africa. *Annals of Botany* **92**(6), 807–834.

[194] Schemske, D. W., Willson, M. F., Melampy, M. N., Miller, L. J., Verner, L., Schemske, K. M., and Best, L. B. 1978. Flowering ecology of some spring woodland herbs. *Ecology* **59**(2), 351–366.

[195] Small, E. 1976. Insect pollinators of the Mer Bleue peat bog of Ottawa. *Canadian field-naturalist* **90**(1), 22–28.