

## A new measure of ecological network structure based on node overlap and segregation

Giovanni Strona<sup>1\*</sup> and Joseph A. Veech<sup>2</sup>

<sup>1</sup>European Commission, Joint Research Centre, Institute for Environment and Sustainability, Forest Resources and Climate Unit, 21027 Ispra, Italy; and <sup>2</sup>Department of Biology, Texas State University, San Marcos, TX 78666-4684, USA

### Summary

1. Despite substantial recent progress, ecologists continue to search for methods of measuring the structure of ecological networks. Several studies have focused on nestedness, a pattern reflecting the tendency of network nodes to share interaction partners. Here, we introduce a new statistical procedure to measure both this kind of structure and the opposite one (i.e. species' tendency against sharing interacting partners) that we call 'node segregation'. In addition, our procedure provides also a straightforward measure of modularity, that is, the tendency of a network to be compartmented into separated clusters of interacting nodes.

2. This new analytical measure of network structure assesses the average deviation between the observed number of neighbours shared by any pair of nodes (species), and the expected number, that is computed using a probabilistic approach based on simple combinatorics. The measure can be applied to both bipartite networks (such as plant–pollinators) and unimode networks (such as food webs). We tested our approach on several sets of hypothetical and real-world networks.

3. We demonstrate that our approach makes it possible to identify different kinds of non-random network configurations (nestedness, node segregation and modularity). In addition, we show that nestedness in ecological networks is less common than previously thought, and that most ecological networks (including the majority of mutualistic ones) tend towards patterns of segregated associations.

4. Our analyses show that the new measure of node overlap and segregation can efficiently identify different structural patterns. The results of our analyses conducted on real networks highlight the need to carefully reconsider the assumption that ecological networks are stable due to their inherent nestedness.

**Key-words:** co-occurrence, modularity, nestedness, null model, probability

### Introduction

Representing species interactions in the form of networks has become a common approach to investigate the patterns of coexistence and community structure (Proulx, Promislow & Phillips 2005). Investigating ecological networks makes it possible to summarize complex patterns into simple statistics. This task is particularly compelling in a context where the challenge for ecologists is to progressively move from collecting large amounts of data to analysing them (Hampton *et al.* 2013).

Some properties of ecological networks can be inferred by focusing on individual nodes (or species). For example, one may randomly select species from a food web and count their number of prey and predators in order to obtain the network degree distribution. In this way, one would not have any information on the way nodes are connected one to another (i.e. who eats whom), but would still obtain insight into the mechanisms responsible for network structure (Newman 2003). Other measures require taking into account larger fragments

of the network. Among these, clustering measures (such as the local, global and average clustering coefficients) are of great interest, because they focus on 'triangles' (i.e. two interacting nodes sharing a neighbour) (Watts & Strogatz 1998), which can represent many common real-world interactions. For example, in a trophic web, triangles can indicate several different situations (see Stouffer 2010), such as two predators sharing a prey and also feeding on one another (e.g. intraguild predation; Polis, Myers & Holt 1989) or a parasite using two species that are predator and prey (Strona 2014).

Besides food webs, investigating patterns of node sharing have gained attention in the study of mutualistic networks (Bascompte *et al.* 2003; Johnson, Dominguez-Garcia & Munoz 2013). Researchers have hypothesized that competition between two species (A and B) could be reduced if the number of shared mutualistic partners is maximized (Bastolla *et al.* 2009; Johnson, Dominguez-Garcia & Munoz 2013). Typically, an incremental increase in the population of either species A or B would have a negative effect on the other species. However, if the two species share a mutualistic partner (species C), then growth of A could have a positive effect on C, that, in turn, could positively affect B thereby offsetting some of the direct

\*Correspondence author. E-mail: giovanni.strona@jrc.ec.europa.eu

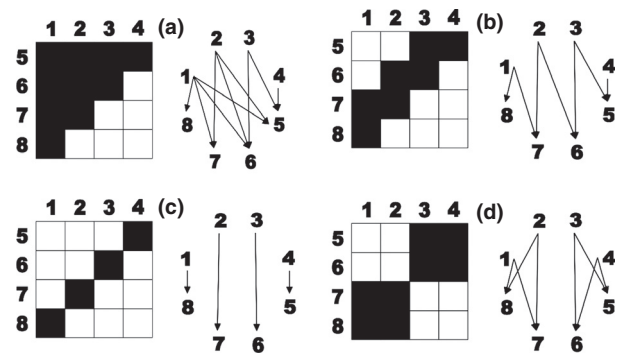
negative effect of A on B. Extended to the whole network, this scenario implies that each node pair could benefit (i.e. have a reduced chance of extinction) from sharing as many neighbouring nodes as possible. This gives the network a nested structure; that is, the set of neighbours connected to any node is a subset of the set of neighbours connected to any node with higher degree (Atmar & Patterson 1993; Kondoh, Kato & Sakato 2010; Johnson, Dominguez-Garcia & Munoz 2013; Staniczenko, Kopp & Allesina 2013).

Several authors have suggested that a nested structure can promote network stability and persistence (Bastolla *et al.* 2009; Thébault & Fontaine 2010). The coexistence of numerous interacting species might be obtained if the species are all part of a highly stable and primarily mutualistic network (Montoya, Pimm & Solé 2006). Although these hypotheses are mostly speculative, they have been used as the basis for a growing corpus of literature that is currently trying to explain why ecological networks are nested (Rohr, Saavedra & Bascompte 2014). Apart from nestedness, there are other network structures that could contribute to network stability. For example, the presence of trophically based subwebs in a food web might confer greater stability by isolating perturbations from the rest of the web (Kondoh, Kato & Sakato 2010).

However, several authors (see, for example, Joppa *et al.* 2010; Johnson, Dominguez-Garcia & Munoz 2013) have shown that the commonness of nestedness in ecological networks could be in part due by methodological issues, which casts some shadows on the robustness of the observed patterns. Here, we provide a new powerful statistical procedure (applicable to both unimode and bipartite networks) that could help clarifying this issue, and we use it to re-examine the structure of a large set of ecological networks. Following previous studies (Almeida-Neto, Guimarães & Lewinsohn 2007; Fortuna *et al.* 2010), we suggest that nestedness represents only one potential outcome, since situations where nodes tend to share few neighbours (herein referred to as segregation), or where high node overlap can be identified within particular groups of nodes (modularity), also constitute meaningful ecological patterns (Podani & Schmera 2012; Podani, Jordán & Schmera 2014).

## Materials and methods

It has been suggested that the average number of shared nodes is a natural measure of nestedness (Bascompte *et al.* 2003; Johnson, Dominguez-Garcia & Munoz 2013). This is consistent with the original concept of nestedness (Patterson & Atmar 1986; Johnson, Dominguez-Garcia & Munoz 2013; Staniczenko, Kopp & Allesina 2013) and with the formulation of one of the most popular nestedness measures, that is NODF (Almeida-Neto *et al.* 2008); in a perfectly nested pattern, each pair of nodes always shares a number of neighbours equal to the degree of the less connected node (Fig. 1a). In other words, the species with lower degree is not linked to any species that the species with higher degree is not also linked to. In any network, there is a finite number of ways that nodes can be connected for a given number of links (connectance) among the nodes. As such, combinatorics can be used to calculate the expected amount of node sharing.



**Fig. 1.** Different network structural patterns. (a) nestedness; (b, c) segregation; (d) modularity. Numbers represent network nodes, while filled cells and arrows represent network edges.

In the context of species co-occurrence, Veech (2013) has shown how the expected number of shared localities between two species can be computed analytically using simple combinatorics. The probabilities given in Veech (2013) can also be obtained through the hypergeometric distribution as described recently in Griffith, Veech & Marsh (2015). Here, we extend the probabilistic species co-occurrence model of Veech (2013) to network node sharing wherein a neighbour node is equivalent to a 'site', and **the task is to determine whether the two focal nodes (or species) 'co-occur' at or share a number of neighbours that is different from the random (probabilistic) expectation.** This allows detection of cases where the number of neighbours shared by two nodes is significantly greater or less than expected (Veech 2013 and Appendix S1). Therefore, *for individual species pairs*, the probabilistic approach can distinguish significant node overlap from significant node segregation (Fig. 1b,c). Moreover, this is accomplished without having to create a randomized network for comparison. In most other approaches (see below) to measuring network structure, it is necessary to construct randomized networks against which to compare the observed network. Furthermore, in these other methods, the amount of node sharing (whether greater or less than expected) is only assessed at the level of the entire network, not individual species pairs. The probabilistic approach that we propose determines whether the observed amount of node overlap (sharing) between a pair of nodes could be due to chance alone (Veech 2014). The probabilistic model of species co-occurrence (Veech 2013; Griffith, Veech & Marsh 2015) has a historical predecessor in the Raup–Crick metric of similarity (Raup & Crick 1979) in that both approaches use combinatorics to analytically derive exact probabilities. The main difference is that the Raup–Crick metric is applied to test for non-random sharing of species among two assemblages (sampling localities), whereas the former tests for non-random sharing of sites by a pair of species. Further, at the time, Raup & Crick (1979) did not recognize the link to the hypergeometric distribution, although their metric has since been linked to that sampling distribution (Chase *et al.* 2011).

## NODE OVERLAP AND SEGREGATION MEASURE ( $\bar{\omega}$ )

Given an undirected network having  $n$  nodes, for each pair of nodes  $V_i$  and  $V_j$  having degree (i.e. number of neighbours), respectively, equal to  $d_i$  and  $d_j$ , relative node overlap can be quantified as:

$$\omega_{ij} = \frac{(S_{ij} - P_{ij})}{\min(d_i, d_j)} \quad \text{eqn 1}$$

where  $S_{ij}$  is the actual number of neighbour nodes shared by  $V_i$  and  $V_j$ , and  $P_{ij}$  is the expected number of shared nodes given by the probabilistic approach, which is computed as:

$$P_{ij} = \sum_{k=1}^{\min(d_i, d_j)} \frac{\binom{n}{k} \times \binom{n-k}{d_j-k} \times \binom{n-d_j}{d_i-k}}{\binom{n}{d_j} \times \binom{n}{d_i}} \times k \quad \text{eqn 2}$$

We note that Equation 2 derives from Veech (2013) which provides a detailed mathematical (probabilistic) description for species co-occurrence among sites, but not node sharing. Because  $S_{ij}$  and  $P_{ij}$  must be less than or equal to  $\min(d_i, d_j)$ ,  $\omega_{ij}$  ranges from  $-1$  to  $1$ , but it is not symmetrical around zero in the sense that a given positive and negative magnitude (e.g.  $\omega_{ij} = 0.5$  and  $\omega_{ij} = -0.5$ ) do not represent an equivalent amount of node sharing and node segregation, respectively. In order to get symmetry and retain scaling from  $-1$  to  $1$ , each  $\omega_{ij}$  is divided by a factor,  $\Omega_{ij}$ , that is the maximum possible value that a given pair of nodes could have.

For any  $V_i$  and  $V_j$  sharing more nodes than expected by chance (i.e. having  $S_{ij} > P_{ij}$ ),  $\Omega_{ij}$  is computed as:

$$\Omega_{ij} = \frac{\min(d_i, d_j) - P_{ij}}{\min(d_i, d_j)} \quad \text{eqn 3}$$

For any  $V_i$  and  $V_j$  sharing fewer nodes than expected by chance (i.e. having  $S_{ij} < P_{ij}$ ), then if:

$$(d_i + d_j - n) < 0 \rightarrow \Omega_{ij} = \frac{P_{ij}}{\min(d_i, d_j)} \quad \text{eqn 4}$$

or else:

$$(d_i + d_j - n) \geq 0 \rightarrow \Omega_{ij} = \frac{P_{ij} - (d_i + d_j - n)}{\min(d_i, d_j)} \quad \text{eqn 5}$$

If the number of observed shared nodes is equal to expected ( $S_{ij} = P_{ij}$ ), then  $\omega_{ij} = 0$  (eqn. 1) and  $\Omega_{ij}$  is set equal to  $1$ . Thus, a symmetric measure of node overlap for  $V_i$  and  $V_j$  is obtained by dividing  $\omega_{ij}$  by  $\Omega_{ij}$ :

$$N_{ij} = \frac{\omega_{ij}}{\Omega_{ij}} \quad \text{eqn 6}$$

Some examples of computation of  $N_{ij}$  values are provided in Appendix S2.  $N_{ij}$  can be computed for each pair of nodes in a network and then averaged to obtain an overall measure of node overlap  $\bar{N}$ . Because  $N_{ij}$  is constrained to be between  $-1$  and  $1$ ,  $\bar{N}$  also varies between  $-1$  and  $1$ .  $\bar{N}$  values near  $-1$  represent a network with less than expected node overlap, whereas  $\bar{N}$  values near  $1$  represent greater than expected node overlap. The expected value of  $\bar{N}$  is zero under a condition of no overall structure of node overlap in the network. A  $\bar{N}$  value around zero can also occur when positive and negative  $N_{ij}$  values cancel one another (which sometimes indicates a situation of network modularity, see next section) or when all  $N_{ij}$  are near zero.

### COMPUTING $\bar{N}$ FOR DIRECTED NETWORKS

In a directed network, we can identify two categories of nodes, that is those having at least one incoming link (heads), and those having at least one outgoing link (tails). Thus, we can compute a separate  $N_{ij}$  value for each set, by evaluating, respectively, the overlap in head nodes for any pair of tail nodes, and the overlap in tail nodes for any pair of head nodes. Obviously, this requires the  $n$  parameter in equations 2, 4

and 5 to be properly adjusted as, respectively, the size of the head set and the size of the tail set. For example, in a food web,  $N_{ij}$  values are computed first focusing on the overlap in food items for each pair of consumers (setting  $n$  equal to the total number of food items), and then by focusing on the overlap in consumers for each pair of food items (setting  $n$  equal to the total number of consumers). Then,  $N_{ij}$  values belonging to the two sets can be averaged together to compute  $\bar{N}$ , or separately, providing two distinct measures of node overlap and segregation ( $\bar{N}_{in}$  and  $\bar{N}_{out}$ ). Differences between  $\bar{N}_{in}$  and  $\bar{N}_{out}$  values (and in the respective distributions of  $N_{ij}$  values) can provide interesting insights into the structure of directed networks, making it possible to detect asymmetric patterns of interactions.

According to the above criterion, node overlap is evaluated within heads and tails but not between them, ensuring, for example, that when evaluating the overlap in the consumers of a pair of plants, we do not include other plants in the computation of  $N_{ij}$ . This derives directly from network structure, however, specific rules can be applied on a per-case basis to improve ecological realism of  $\bar{N}$ , by identifying a particular set of nodes to be used in the computation of each  $N_{ij}$  value (adjusting  $n$  as the size of that set). **In a trophic web, one may take into account trophic levels to discriminate between permitted and forbidden links, setting  $n$  for any  $N_{ij}$  as the actual number of nodes that can be shared by  $V_i$  and  $V_j$  according to their trophic ecology. Similarly, one could decide to exclude cannibalistic interactions by excluding  $V_i$  and  $V_j$  from the computation of  $N_{ij}$ , and thus replacing the parameter  $n$  with  $(n - 2)$ .**

These kinds of adjustments are, in general, not needed for directed bipartite networks because the set of tails is disjoint from the set of heads, that is links are allowed between but not within two different, well-distinguished node categories (e.g. plants and pollinators). However, adjustments can be made to the sets of nodes involved in the computation of each  $N_{ij}$  value, by identifying all nodes that  $V_i$  and  $V_j$  can share according to a specific hypothesis. Since all of these possibilities, however, go well beyond the demonstrative scope of our analyses, we will demonstrate here the use of  $\bar{N}$  in its basic formulation. Thus, for all the analyses on simulated and real bipartite networks, we will compute  $\bar{N}$  as the average of all  $N_{ij}$  values (calculated separately for tail and head nodes). For the analyses on food webs, we will compute  $\bar{N}$  as the average of all  $N_{ij}$  values (again, computed separately for tail and head nodes, that is food items and consumers), with the only exclusion of cannibalistic interactions (see previous paragraph).

### SIGNIFICANCE TESTING

Assessing significance of  $\bar{N}$  is complicated by the fact that the distribution of  $N_{ij}$  values is truncated, often non-normal, and the values themselves are not completely independent of one another. However, we propose a test that is able to overcome these difficulties. First, the truncation and non-normality of the  $N_{ij}$  distribution can be overcome by invoking the central limit theorem (CLT), which has been recently demonstrated to be effective also for truncated distributions (Cha & Cho 2014). According to the CLT, the distribution of  $\bar{N}$  as a random expectation (i.e. a set of  $\bar{N}$  values from random networks) should be approximately normal and centred on zero regardless of the underlying form of each  $N_{ij}$  distribution. Therefore, the significance of  $\bar{N}$  can be assessed with the test statistic,  $z = (\bar{N} - \mu)/SE$ . However, because the  $N_{ij}$  values are not independent, the standard error (SE) is underestimated and the true or effective sample size ( $n$ ) cannot be known. Instead of using the typical  $SE = \sigma/\sqrt{n}$ , where  $n$  = the number of node pairs (or  $N_{ij}$  values), we conservatively use  $n = 1$ . Therefore, our test statistic becomes  $Z_N = (\bar{N} - \mu)/\sigma$  with  $\mu = 0$  and  $\sigma$  = the standard deviation of the

observed  $\bar{N}$ , with the  $P$ -value derived from the standard normal distribution as usual. An observed network (or matrix) has non-random structure if  $Z_N$  is significantly different from zero.

Our test for significance of  $\bar{N}$  is similar to the test typically used in nestedness analyses involving data randomization (e.g. Ulrich, Almeida-Neto & Gotelli 2009; Strona & Fattorini 2014). In those analyses, a  $Z$ -value is calculated as  $Z = [N_{\text{obs}} - \text{mean}(N_{\text{sim}})]/\sigma(N_{\text{sim}})$  where  $N_{\text{obs}}$  is the observed value of a nestedness metric and  $N_{\text{sim}}$  represent values from randomized or null species presence-absence matrices (or networks). In our approach to testing network structure, we do not have randomized networks but rather our test invokes the CLT. A key feature of our analytical approach is the desire to avoid data randomization, given that randomization is often sensitive to the particular algorithm (i.e. null model) employed (Gotelli 2000; Ulrich & Gotelli 2007; Strona & Fattorini 2014).

#### MODULARITY

As mentioned above,  $\bar{N}$  values close to 0 could arise from different scenarios. A case of great interest is that of modularity; that is, a situation where a group(s) of nodes can be identified in a network where node overlap is substantially higher than expected between nodes belonging to the same group, and lower than expected between nodes belonging to different groups (Fig. 1d). Investigating the distribution of  $N_{ij}$  values provides an immediate visual way to distinguish between random and modular networks having  $\bar{N} \sim 0$ , as the first would have a unimodal distribution of  $N_{ij}$  values with a peak centred at around 0, whereas the latter would have a bimodal distribution of  $N_{ij}$  values, with a peak at near  $-1$ , arising from node segregation between different modules, and another peak near 1 arising from node overlap within modules.

The most straightforward way to quantify the bimodal pattern is evaluating the statistical dispersion of  $N_{ij}$  values around the mean. Thus, we suggest the standard deviation of the  $N_{ij}$  values as a simple measure of modularity (Mod). Although modularity measured in this way provides only a qualitative descriptor (i.e. it is not paired to a significance test), it has the convenient property of being mathematically related to the  $z$ -test for the significance of  $\bar{N}$ . That is, a large standard deviation might prevent an observed  $\bar{N}$  value from being significant; nonetheless, the network might have structure in the form of modules even though the mean amount of node sharing throughout the entire network is not significant.

#### TEST ON SIMULATED NETWORKS

We tested the ability of our approach to correctly identify non-random patterns of network nestedness, segregation and modularity on a set of simulated networks with different amounts and types of structure. To allow comparisons with previous nestedness measures, we created all networks (1000 per type of structure) as bipartite.

To obtain different amounts of order in the simulated networks, we followed Strona & Fattorini (2014). For this, first we generated perfectly nested/segregated/modular networks of various sizes, and then we rewired edges at random, quantifying order based on the percentage of edges rewired. To create nested networks, we constrained the set of neighbours connected to any node to be a subsample of any other set of neighbours connected to a node having the same or a higher degree. To create segregated networks, we constrained each node to share no neighbouring node with any other node in the network (this also forced the in- and out-degree of each node to be equal to 1). For both the nested and the segregated networks, we randomly varied the number of nodes between 20 and 100. To create a different amount of order, for

each network, we reiterated  $r \times E$  times the procedure of extracting two nodes at random and changing their status from connected to disconnected or *vice versa*. The parameter  $E$  indicates the total number of possible edges between all nodes in the network (which, in a bipartite network, corresponds to the size of the corresponding binary matrix). We let  $r$  vary randomly between 0 (maximum order) and 1 (maximum disorder), as we empirically observed that a number of randomization steps equal to  $E$  were enough to approximate the maximum perturbation status of the network (Strona *et al.* 2014).

For each of these nested/segregated random networks, we computed  $\bar{N}$  and, additionally, we quantified nestedness using both *NODF* (Almeida-Neto *et al.* 2008) and the spectral radius of the network adjacency matrix,  $\rho$  (Staniczenko, Kopp & Allesina 2013). To evaluate significance of *NODF* and  $\rho$ , we created two independent sets of null networks using different randomization rules. For  $\rho$ , we used the null model recommended by Staniczenko, Kopp & Allesina (2013), which creates randomly connected networks having the same number of tails, heads and edges of the original one. For *NODF*, we used the proportional null model proposed by Bascompte *et al.* (2003), which starts from the set of nodes of the original network and then connects two nodes with a probability proportional to their in- and out-degree. In all cases, we quantified nestedness using both  $P$  and  $Z$  values (see Strona & Fattorini 2014). Finally, we performed a set of pairwise comparisons between the amount of network structure, network quantitative properties,  $\bar{N}$  values and the different raw and standardized nestedness measures. To make clear the distinction between node overlap and segregation, instead of quantifying order using the parameter  $r$  (that ranges between 0, maximum order, and 1, maximum disorder), we transformed it to  $1-r$  for nested networks and to  $r-1$  for segregated networks. In this way, we obtained a symmetric measure of order ranging from  $-1$  (complete segregation) to 1 (perfect nestedness), with 0 indicating randomness.

We also used simulated networks to conduct a power analysis of our test based on  $Z_N$ . For a specified effect size, we calculated type I error probability as the number of  $\bar{N}$  values that were less than the effect size but still statistically significant ( $\alpha = 0.05$ ) expressed as a proportion of all  $\bar{N}$  values less than the effect size. Type II error probability was calculated as the number of  $\bar{N}$  values that were greater than the effect size but not statistically significant and expressed as a proportion of all  $\bar{N}$  values greater than the effect size. Power was calculated as  $1 - P$  (Type II error).

To evaluate our measure of modularity, we created (bipartite) networks with different degrees of modularity by selecting *a priori* a random number of nodes (varying from 100 to 200) and a random number of modules (varying between 2 and 10). Then, we randomly assigned each node to a module. Finally, for each possible node pair within a module, with a probability  $P$ , we created an edge between the node pair, and with probability  $1-p$ , we created an edge between two nodes picked at random from the whole node set. We let  $p$  vary between 0 (complete randomness) and 1 (maximum modularity). Finally, we assessed modularity in each network as  $Mod = \sigma(N_{ij})$ , and we compared each Mod value with the corresponding  $P$ .

#### TEST ON REAL NETWORKS

We measured  $\bar{N}$  in a large set of real ecological networks featuring different kinds of interactions, and particularly on: (i) all the mutualistic networks available from Web of Life data set ([www.web-of-life.es](http://www.web-of-life.es)), including 59 plant/pollinator networks and 30 plant/seed disperser networks; (ii) two different sets of host/parasite networks, the first including the 50 largest networks compiled by grouping per country all

the fish parasite records included in the Natural History Museum (London, UK) host/parasite data base (<http://www.nhm.ac.uk>), and the second including 26 mammal/flea networks provided by Krasnov *et al.* (2012) (available at <http://datadryad.org/handle/10255/dryad.36193>); (iii) all the community food webs ( $n = 179$ ) available from the ECO-WeB data base (Cohen 2010), as an example of application of  $\bar{N}$  to uni-mode networks. It should be pointed out that many of the food webs in the last data set include ecological functional groups in place of, or together with taxonomic units. Consequently, any interpretation of results requires caveats, since the causes of node overlap and segregation are intimately related to node types, and should therefore be evaluated on a per-case basis.

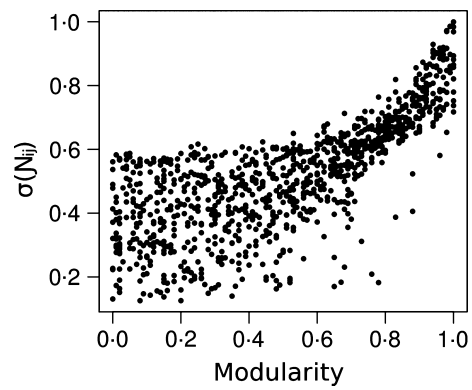
**Results**

COMPARISON WITH SIMULATED NETWORKS

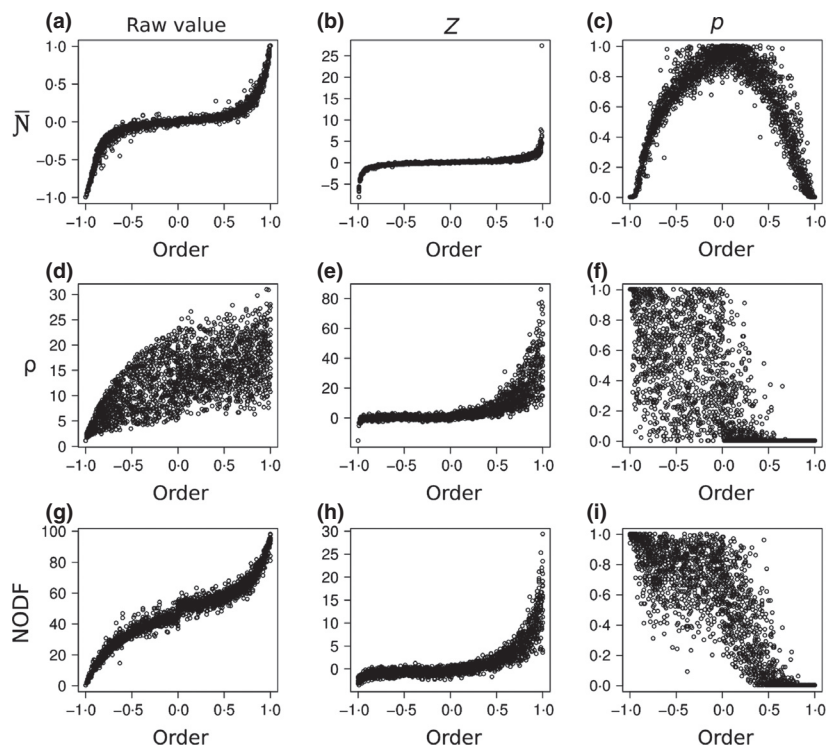
Our new measure,  $\bar{N}$ , performed well in quantifying the amount of order or structure in the simulated networks. As the amount of order increased from  $-0.5$  to  $-1$ ,  $\bar{N}$  rapidly approached  $-1$  indicating increasing node segregation, while when order increased from  $0.5$  to  $1$ ,  $\bar{N}$  rapidly approached  $1$  indicating increasing node overlap (Fig. 2a). As intended,  $\bar{N}$  gradually declined to near zero for networks, in which an increasing number of connections among nodes was random. With 50% or more of the node links being random (order between  $-0.5$  and  $0.5$ ),  $\bar{N}$  was almost always between  $-0.1$  and  $0.1$  (Fig. 2a) indicating that networks with a majority of random links have  $\bar{N}$  values near zero.  $Z_{\bar{N}}$  and  $P$ -values of  $\bar{N}$  effectively distinguished between highly ordered and random networks (Fig. 2b,c). The proposed modularity measure (i.e. the standard deviation of  $N_{ij}$  values) was positively correlated with modularity degree (Spearman's rho =  $0.77$ ,  $P < 0.0001$ )

and made it possible to discriminate between modular and non-modular networks (Fig. 3).

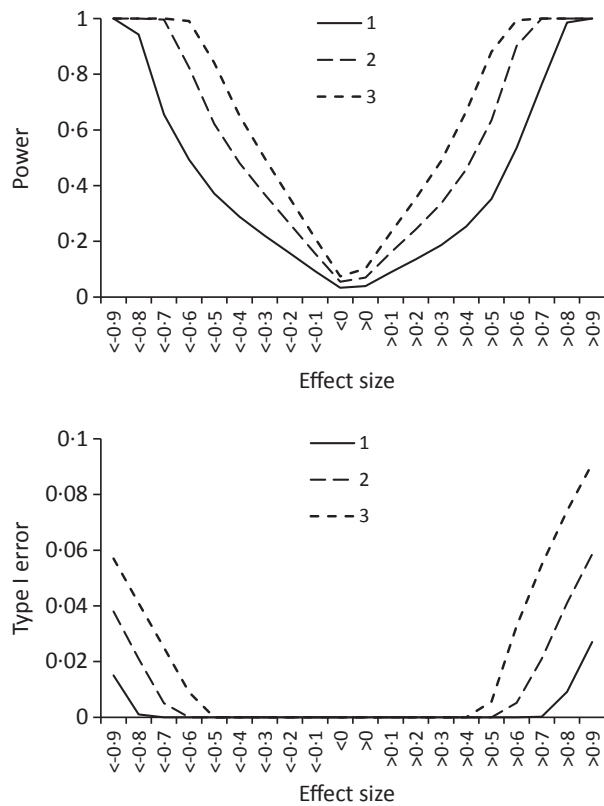
The  $Z$ -test had relatively good power ( $>0.7$ ) when effect size was  $\bar{N} > 0.7$  and  $\bar{N} < -0.8$  for nested and segregated networks respectively (Fig. 4). Our  $Z$ -test is inherently conservative in that we base the SE on an effective sample size of  $n = 1$ . However, the power can be increased by using any number between 1 and 3 (as a power adjustment factor) in the calculation of SE and subsequently  $Z_{\bar{N}}$ . Using a power adjustment factor of  $n = 3$ , the test had good power even for effect sizes as low as  $\bar{N} > 0.5$  and  $\bar{N} < -0.5$  (Fig. 4). In any test of significance, there is generally a trade-off between power and type I error. As power increases, type I error also increases. This trade-off was found for our  $Z$ -test, although type I error proba-



**Fig. 3.** Relationships between the modularity measure based on standard deviation of  $N_{ij}$  values, and the amount of modularity (varying from 0, no modularity, to 1 complete modularity) in 1000 simulated networks.



**Fig. 2.** Relationship between the amount of structure (varying from  $-1$ , complete segregation, to  $1$  perfect nestedness) in a set of 2000 random networks and raw values,  $Z$  and  $P$  values of  $\bar{N}$  (a–c),  $\rho$  (d–f) and  $NODF$  (g–i).



**Fig. 4.** Power and type I error probability for the  $Z_N$  test. The assessment is based on 10 000 simulated networks having  $\bar{N}$  ranging between  $-1$  and  $1$  (effect size). Lines represent different power adjustment settings (see text for details).

bilities generally remained at a reasonably low level typically not exceeding the specified significance level (Fig. 4). We also examined power adjustment = 5 (not shown in Fig. 4) and found that type I error probabilities became intolerably high.

#### COMPARISON BETWEEN $\bar{N}$ AND OTHER NESTEDNESS MEASURES

The spectral radii ( $\rho$ ) computed for the set of simulated networks were poorly correlated with the corresponding amount of order (Fig. 2d). Conversely, comparison of *NODF* values with order of the simulated networks produced results similar to those obtained using  $\bar{N}$ , that is *NODF* was able to detect quite well variation in network structure, both for nested (as in

Strona & Fattorini 2014) and for segregated networks (Fig. 2g). However, the correlation between order and  $\bar{N}$  was more tight than that between order and *NODF*, that is a given amount of structure was typically indexed by a smaller range in  $\bar{N}$  values than in *NODF* values (cf. Fig. 2a,g).

$Z$ - and  $P$ -values of  $\rho$  performed better than the spectral radius in assessing network structure (Fig. 2e,f). Conversely,  $Z$ - and  $P$ -values of *NODF* were less correlated with the amount of structure than the raw *NODF* value (Fig. 2h,i). In general,  $Z$ - and  $P$ -values of  $\bar{N}$  (Fig. 2b,c) were more efficient than those of  $\rho$  (Fig. 2e,f) and *NODF* (Fig. 2h,i) in measuring segregation.

#### INDEPENDENCE OF $\bar{N}$ FROM NETWORK PROPERTIES

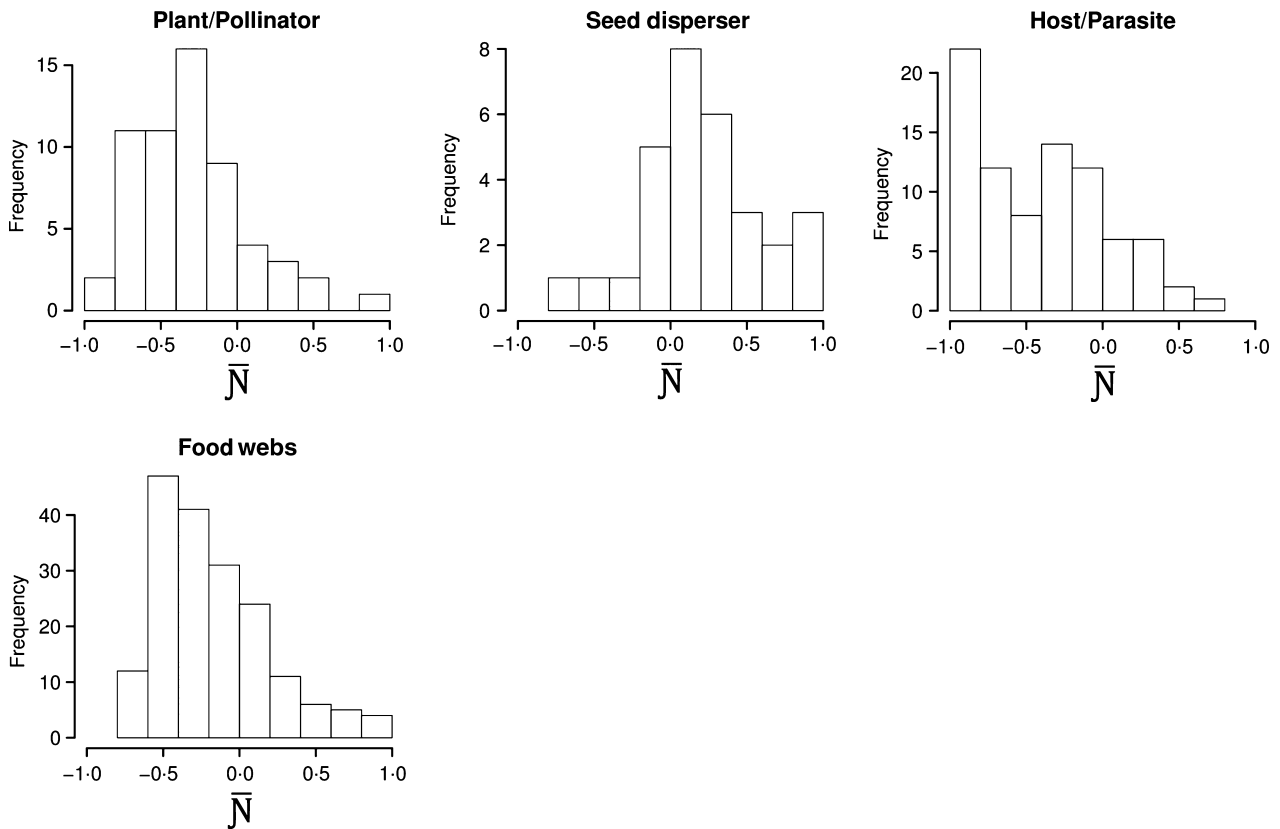
The simulated networks were much variable in their structure and symmetry, with number of edges ranging between 10 and 1375 (mean =  $246.6 \pm 283.4$ ), connectance ranging between 0.02 and 0.75 (mean =  $0.28 \pm 0.24$ ) and symmetry (ratio between tails and heads) varying between 0.2 and 4.9 (mean =  $0.95 \pm 0.70$ ). Neither  $\bar{N}$  nor the other investigated nestedness measures (i.e. *NODF* and the spectral radius,  $\rho$ ) were completely independent from network properties (Table 1). The number of network nodes was strongly correlated only with spectral radius  $\rho$ , but  $\bar{N}$ , *NODF* and  $\rho$  were all correlated with the number of edges and network connectance, that is the fraction of existing edges over the possible total number. However, both edge number and connectance were strongly correlated with the amount of simulated order (with  $r_s$  respectively equal to 0.751 and 0.752). Therefore, to ‘factor out’ this intermediary effect of order, we obtained the residuals from regressions of number of edges vs. order and connectance vs. order. When using the residuals, none of the measures were correlated with edge number and connectance (Table 1), with the exception of  $\rho$ , which was still moderately correlated with the number of edges ( $r_s = 0.28$ ).

#### COMPARISON WITH REAL NETWORKS

We found that situations of overall node segregation and strong differences in overlap between node categories (e.g. plants vs. pollinators, host vs. parasites, etc.) are common in both mutualistic and antagonistic networks. In most of the investigated networks (72%), node pairs tended to share fewer neighbours than expected by chance, leading to negative

**Table 1.** Relationships (expressed as Spearman’s rank correlation coefficients) between  $\bar{N}$ ,  $\rho$ , *NODF* and their respective  $Z$  values, and network properties measured in a set of 1000 nested and 1000 segregated simulated networks. Correlations for the residuals are indicated by ‘ $\epsilon$ ’

	Edges	Nodes	Connectance	$\epsilon$ (Edges vs. Order)	$\epsilon$ (Connectance vs. Order)	Order
$\bar{N}$	0.74	0.00	0.73	-0.12	-0.02	0.98
$Z_{\bar{N}}$	0.74	-0.01	0.75	-0.08	0.01	0.98
$\rho$	0.82	0.63	0.33	0.28	-0.08	0.67
$Z(\rho)$	0.73	0.09	0.67	0.11	0.08	0.85
<i>NODF</i>	0.74	-0.03	0.78	-0.10	0.03	0.98
$Z(\text{NODF})$	0.68	0.00	0.69	0.02	0.03	0.87



**Fig. 5.** Distribution of  $\bar{N}$  values computed on the different categories of ecological networks, and particularly plant–pollinator, seed–disperser and host–parasite networks and food webs.

$\bar{N}$  values (Fig. 5). However, among all the investigated ecological networks (353), only 44 showed significant structure ( $P < 0.05$ , with  $Z_N$  computed using an adjustment factor  $n = 3$ ), 33 of which had  $\bar{N} < -0.7$ . These findings contrast with the fact that, consistent with previous studies (Bascompte *et al.* 2003; Staniczenko, Kopp & Allesina 2013), most of these networks are significantly nested when examined with other procedures such as the spectral radius approach and *NODF* (see Table S1).

As mentioned,  $\bar{N}$  values near zero could indicate modularity, that is situations where many highly positive and negative  $N_{ij}$  values cancel one another in the averaging procedure. We found that most networks (including those not showing strong patterns as revealed by  $\bar{N}$ ) were highly modular (average  $\text{Mod} = 0.78 \pm 0.18$ ), thus indicating that overall randomness is indeed rare in ecological networks.

## Discussion

By using a new approach to measuring network structure, our results indicate that nestedness is probably not as widespread as currently thought. Among the network typologies we investigated, patterns of substantial node overlap were moderately common only in seed–disperser networks (Fig. 5). This is consistent with the fact that most granivores and frugivores (particularly mammals and birds) tend to be dietary generalists. (Wheelwright 1985). In the other kinds of net-

works, including the plant–pollinator ones, we found an overall tendency towards node segregation (i.e. against sharing interacting partners), which supports the idea that specialization may be the key to species coexistence (Pauw 2013).

Investigating how the taxonomic, biogeographical and ecological relatedness between species pairs affects their respective  $N_{ij}$  values could provide fundamental insights into these issues, possibly leading to a better understanding of the determinants of ecological network structure. Although this kind of detailed analysis was far beyond the scope of this paper, we hope that this approach will be pursued by future researchers.

Our findings are consistent with the possibility that nested networks are no more likely to be stable than are unstructured ones, and that the strength of interactions may play a fundamental role in system stability (Allesina & Tang 2014). Thus, the controversial notion that nestedness promotes the stability of ecological networks (Bastolla *et al.* 2009; Thébaud & Fontaine 2010; James, Pitchford & Plank 2012; Rohr, Saavedra & Bascompte 2014) should not be taken without questioning. Although our study is not the first casting doubt on the ubiquity of nestedness (see Joppa *et al.* 2010; Johnson, Dominguez-Garcia & Munoz 2013), it is the first providing analytical evidence that most species tend to share fewer partners than expected by chance.

By offering a synthetic measure capable of simultaneously quantifying node overlap and segregation,  $\bar{N}$  provides various advances to the field of ecological network analysis.

A fundamental aspect of our method is that the deviation (positive or negative) between the observed and the expected overlap for any pair of nodes in a network is computed by using a probabilistic approach based on combinatorics (as in Veech 2013). This makes it possible to obtain analytically the expected number of shared partners between two nodes while controlling their respective number of edges and the total number of nodes in the network. In this sense, our approach is substantially different from other nestedness measures that assess the deviation between expected and observed patterns of node overlap *a posteriori* (using null model analysis) and only at the level of the whole network (Ulrich & Gotelli 2007).

We have proposed  $\bar{N}$  as a unique measure to assess the overall tendency of a network towards nestedness or segregation. Although we have shown with our simulations that this measure is a faithful representation of the amount of structure of a network, it should be emphasized that the core information provided by our analytical approach lays in the distribution of  $N_{ij}$  values that can offer a detailed picture of how the tendency towards nestedness and/or segregation is distributed across the network. Researchers can rely on  $\bar{N}$  or on the distribution of  $N_{ij}$  values (or both) for their analysis depending on their particular purposes. Moreover, substantial information can be obtained by comparing the distributions of expected and observed node overlap values; we hope this will be investigated in more detail in future studies of real-world networks.

Similarly to *NODF* and spectral radius, our measure was not completely independent from connectance, with weakly connected networks tending towards low  $\bar{N}$  values, due to the fact that networks of this kind are constrained towards a segregated edge configuration. The effect of connectance on nestedness measures is often considered undesirable. A possible solution to this issue is using null model analysis, that is comparing the observed nestedness with that simulated in a set of null networks having the same (or comparable) connectance of the target network (Ulrich & Gotelli 2007; Almeida-Neto *et al.* 2008; Staniczenko, Kopp & Allesina 2013).

Our approach, in principle, does not prevent the use of data randomization that is common to most null models; it is permissible to compare  $\bar{N}$  of a given network to that of any other network built according to any criteria. Nevertheless, we have provided an alternative test of significance that has sufficient power, and that is free from the potential issues related to traditional null model approaches and data randomization (Gotelli 2000; Ulrich & Gotelli 2007; Strona & Fattorini 2014). However, we are confident that users might find even better ways to use  $N_{ij}$  values (as well as the distributions of observed and expected node overlap values) to analyse ecological networks. Our hope is that the probabilistic approach will lead to a more thorough and comprehensive analysis and thereby a more in-depth understanding of ecological network structure.

## Acknowledgements

We are indebted with four anonymous Reviewers for their useful comments and suggestions on a previous version of the manuscript.

## Data accessibility

Mutualistic networks are available from Web of Life data set ([www.web-of-life.es](http://www.web-of-life.es)). Data necessary to build fish host/parasite networks grouped by country can be obtained from the Natural History Museum (London, UK) host/parasite data base (<http://www.nhm.ac.uk>). The 26 mammal/flea networks provided by Krasnov *et al.* (2012) are available at <http://datadryad.org/handle/10255/dryad.36193>. All the food webs used in this study are included in the ECO-WeB data base (Cohen 2010), which can be downloaded at <http://hdl.handle.net/10209/306>.

## References

- Allesina, S. & Tang, S. (2014) Stability criteria for complex ecosystems. *Nature*, **483**, 205–208.
- Almeida-Neto, M., Guimarães Jr, P.R. & Lewinsohn, T.M. (2007) On nestedness analyses: rethinking matrix temperature and anti-nestedness. *Oikos*, **116**, 716–722.
- Almeida-Neto, M., Guimaraes, P., Guimarães, P.R., Loyola, R.D. & Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**, 1227–1239.
- Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, **96**, 373–382.
- Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003) The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences USA*, **100**, 9383–9387.
- Bastolla, U., Fortuna, M.A., Pascual-García, A., Ferrera, A., Luque, B. & Bascompte, J. (2009) The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, **458**, 1018–1020.
- Cha, J. & Cho, B.R. (2014) Classical statistical inference extended to truncated populations for continuous process improvement: test statistics, p-values, and confidence intervals. *Quality and Reliability Engineering International*, doi: 10.1002/qre.1719
- Chase, J.M., Kraft, N.J., Smith, K.G., Vellend, M. & Inouye, B.D. (2011) Using null models to disentangle variation in community dissimilarity from variation in  $\alpha$ -diversity. *Ecosphere*, **2**, art24.
- Cohen, J.E. (2010) *Ecologists' Co-Operative Web Bank. Version 1.1. Machine-readable data base of food webs*. The Rockefeller University, New York. URL <http://hdl.handle.net/10209/306> [accessed 10 December 2014]
- Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Moullot, D., Krasnov, B.R., Poulin, R. & Bascompte, J. (2010) Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology*, **79**, 811–817.
- Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.
- Griffith, D.M., Veech, J.A. & Marsh, C.J. (2015) Cooccur: probabilistic Species Co-occurrence Analysis in R. *Journal of Statistical Software*, in press (accepted).
- Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S. & Porter, J.H. (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment*, **11**, 156–162.
- James, A., Pitchford, J.W. & Plank, M.J. (2012) Disentangling nestedness from models of ecological complexity. *Nature*, **487**, 227–230.
- Johnson, S., Dominguez-García, V. & Munoz, M.A. (2013) Factors determining nestedness in complex networks. *PLoS ONE*, **8**, e74025.
- Joppa, L.N., Montoya, J.M., Solé, R., Sanderson, J. & Pimm, S.L. (2010) On nestedness in ecological networks. *Evolutionary Ecology Research*, **12**, 35–46.
- Kondoh, M., Kato, S. & Sakato, Y. (2010) Food webs are built up with nested subwebs. *Ecology*, **91**, 3123–3130.
- Krasnov, B.R., Fortuna, M.A., Moullot, D., Khokhlova, I.S., Shenbrot, G.I. & Poulin, R. (2012) Phylogenetic signal in module composition and species connectivity in compartmentalized host–parasite networks. *American Naturalist*, **179**, 501–511.
- Montoya, J.M., Pimm, S.L. & Solé, R.V. (2006) Ecological networks and their fragility. *Nature*, **442**, 259–264.
- Newman, M.E. (2003) The structure and function of complex networks. *SIAM Review*, **45**, 167–256.
- Patterson, B.D. & Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, **28**, 65–82.
- Pauw, A. (2013) Can pollination niches facilitate plant coexistence? *Trends in Ecology and Evolution*, **28**, 30–37.
- Podani, J., Jordán, F. & Schmera, D. (2014) A new approach to exploring architecture of bipartite (interaction) ecological networks. *Journal of Complex Networks*, **2**, 168–186.



- Podani, J. & Schmera, D. (2012) A comparative evaluation of pairwise nestedness measures. *Ecography*, **35**, 889–900.
- Polis, G.A., Myers, C.A. & Holt, R.D. (1989) The ecology and evolution of intra-guild predation – potential competitors that eat each other. *Annual Review in Ecology and Systematics*, **20**, 297–330.
- Proulx, S.R., Promislow, D.E. & Phillips, P.C. (2005) Network thinking in ecology and evolution. *Trends in Ecology and Evolution*, **20**, 345–353.
- Raup, D.M. & Crick, R.E. (1979) Measurement of faunal similarity in paleontology. *Journal of Paleontology*, **53**, 1213–1227.
- Rohr, R.P., Saavedra, S. & Bascompte, J. (2014) On the structural stability of mutualistic systems. *Science*, **345**, 1253497.
- Staniczenko, P.P., Kopp, J.C. & Allesina, S. (2013) The ghost of nestedness in ecological networks. *Nature Communications*, **4**, 1391.
- Stouffer, D.B. (2010) Scaling from individuals to networks in food webs. *Functional ecology*, **24**, 44–51.
- Strona, G. (2014) The underrated importance of predation in transmission ecology of direct lifecycle parasites. *Oikos*, doi: 10.1111/oik.01850.
- Strona, G. & Fattorini, S. (2014) On the methods to assess significance in nestedness analyses. *Theory in Biosciences*, **133**, 179–186.
- Strona, G., Nappo, D., Boccacci, F., Fattorini, S. & San-Miguel-Ayanz, J. (2014) A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*, **5**, 4114.
- Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, **329**, 853–856.
- Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3–17.
- Ulrich, W. & Gotelli, N.J. (2007) Null model analysis of species nestedness patterns. *Ecology*, **88**, 1824–1831.
- Veech, J.A. (2013) A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, **22**, 252–260.
- Veech, J.A. (2014) The pairwise approach to analysing species co-occurrence. *Journal of Biogeography*, **41**, 1029–1035.
- Watts, D.J. & Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.
- Wheelwright, N.T. (1985) Fruit-size, gape width, and the diets of fruit-eating birds. *Ecology*, **80**, 8–818.

Received 24 February 2015; accepted 2 April 2015

Handling Editor: Timothée Poisot

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Calculation of the *P*-value for an observed number of shared nodes.

**Appendix S2.** Examples of computation of  $N_{ij}$  values in an undirected network.

**Table S1.**  $\bar{N}$ ,  $\rho$ , *NODF* (and their respective *Z* and *P* values), *Mod* and quantitative properties of all the ecological networks included in the analyses.