

THE STRUCTURE OF PROBABILISTIC NETWORKS

T. POISOT, A. CIRTWILL, D. GRAVEL, M.-J. FORTIN, AND D.B. STOUFFER

ABSTRACT

1. There is a growing realization among community ecologists that interactions between species vary in space and time. Yet, our current numerical framework to analyze the structure of interactions, largely based on graph-theoretical approaches, is unsuited to this type of data. Since the variation of species interactions holds much information, there is a need to develop new metrics to exploit it.
2. We present analytical expressions of key network metrics, using a probabilistic framework. Our approach is based on modeling each interaction as a Bernoulli event, and using basic calculus to express the expected value, and when mathematically tractable, its variance. We provide a free and open-source implementation of these measures.
3. We show that our approach allows to overcome limitations of both neglecting the variation of interactions (over-estimation of rare events) and using simulations (extremely high computational demand). We present a few case studies that highlight how these measures can be used.
4. We conclude this contribution by discussing how the sampling and data representation of ecological network can be adapted to better allow the application of a fully probabilistic numerical framework.

Ecological networks are an efficient way to represent biotic interactions between individuals, populations, or species. Historically, their study focused on describing their structure, with a particular attention on food webs (Dunne 2006) and plant-pollinator interactions (Jordano 1987; Bascompte *et al.* 2003). The key result of this line of research was linking this structure to community or ecosystem-level properties such as stability (McCann 2014), coexistence (Bastolla *et al.* 2009; Haerter *et al.* 2014), or ecosystem functioning (Duffy 2002; Thébault & Loreau 2003; Poisot 2012). To a large extent, the description of ecological networks resulted in the emergence of questions about how functions emerged from structure, and this stimulated the development of a rich methodological literature, defining a wide array of structural properties. Given a network as input, measures of network structure return a *property* based on one or several *units* from this network. Some of the properties are *direct* properties (they only require knowledge of the unit on which they are applied), whereas others are *emerging* (they require knowledge of, and describe, higher-order structures). For example, connectance, the realized proportion of potential interactions, is a direct property of a network. The degree of a node (how many interactions it is involved in) is a direct property of the node. The nestedness of a network (that is, the extent to which specialists and generalists overlap), on the other hand, is an emerging property that is not directly predictable from the degree of all nodes. Establishing a difference between direct and emerging properties is important when interpreting their values; direct properties are conceptually equivalent to means, in that they are the first moment of network units, whereas emerging properties are conceptually equivalent to variances or other higher-order moments.

In the recent years, the interpretation of the properties of network structure (as indicators of the action of ecological or evolutionary processes) has been somewhat complicated by the observation that network structure varies through space and time. This happens because, contrary to a long-standing assumption of network studies, species from the same pool do not interact in a consistent way (Poisot *et al.* 2012). Empirical and theoretical studies suggest that the network is not the right unit to understand this variation; rather, network variation is an emerging property of the response of ecological interactions to environmental factors

1 and chance events (Poisot *et al.* 2014). Interactions can vary because of local mismatching in
2 phenology (Olesen *et al.* 2011), populations fluctuations preventing the interaction (Canard
3 *et al.* 2014), or a combination of both (Chamberlain *et al.* 2014; Olito & Fox 2014). Olito &
4 Fox (2014) show that accounting for neutral (population-size driven) and trait-based effects
5 allows the prediction of the cumulative change in network structure, but not of the change at
6 the level of individual interactions. In addition, Carstensen *et al.* (2014) show that within a
7 meta-community, not all interactions are equally variable: some are highly consistent, whereas
8 others are extremely rare.

9 Taken together, these considerations highlight the need to amend our current methodology
10 for the description of ecological networks, in order to give more importance to the variation
11 of individual interactions. Because the methodological corpus available to describe ecolog-
12 ical networks had first been crafted at a time when it was assumed that interactions were
13 invariants, it is unsuited to address the questions that probabilistic networks allow us to ask.
14 In this paper, we show that several direct and emerging core properties of ecological net-
15 works (both bipartite and unipartite) can be re-formulated in a probabilistic context (Yeakel
16 *et al.* 2012; Poisot *et al.* 2014); we conclude by showing how this methodology can be ap-
17 plied to exploit the information contained in the variability of networks, and to reduce the
18 computational burden of current methods in network analysis. We provide a free and open-
19 source implementation of this set of measures in a library for the julia language, available at
20 <http://github.com/PoisotLab/ProbaNet.jl>.

21

METRICS

22 Throughout this paper, we use the following notation. \mathbf{A} is a matrix wherein A_{ij} is $P(ij)$,
23 *i.e.* the probability that species i establishes an interaction with species j . If \mathbf{A} represents a
24 unipartite network (*e.g.* a food web), it is a square matrix and contains the probabilities of each
25 species interacting with itself. If \mathbf{A} represents a bipartite network (*e.g.* a pollination network),
26 it will not necessarily be square. We call S the number of species, and R and C respectively the
27 number of rows and columns. $S = R = C$ in unipartite networks, and $S = R + C$ in bipartite
28 networks.

1 Note that all of the measures defined below can be applied on a bipartite network that has
 2 been made unipartite; the unipartite transformation of a bipartite matrix \mathbf{A} is the block matrix

$$(1) \quad \mathbf{B} = \begin{pmatrix} 0_{(R,R)} & \mathbf{A} \\ 0_{(C,R)} & 0_{(C,C)} \end{pmatrix},$$

3 where $0_{(C,R)}$ is a matrix of C rows and R columns filled with 0s, etc.

4 We assume that all interactions are independent (so that $P(ij|kl) = P(ij)P(kl)$ for any species),
 5 and can be represented as Bernoulli trials (so that $0 \leq P(ij) \leq 1$). The later condition allows
 6 to derive estimates for the *variance* ($\text{var}(X) = p(1 - p)$), and expected values ($E(X) = p$).
 7 We can therefore estimate the variance of most properties, using the fact that the variance of
 8 additive independent events is the sum of their individual variances, and that the variance of
 9 multiplicative independent events is

$$(2) \quad \text{var}(X_1 X_2 \dots X_n) = \prod_i \left(\text{var}(X_i) + [E(X_i)]^2 \right) - \prod_i [E(X_i)]^2$$

10 As a final note, all of the measures described below can be applied on the binary (0/1) versions
 11 of the networks and will give the exact value of the non-probabilistic measure. This property is
 12 desirable as it allows our framework to be used on any network, whether they are represented
 13 in a probabilistic or binary way.

14 **Direct properties.**

15 *Connectance and number of interactions.* Connectance (or network density) is the proportion of
 16 possible interactions that are realized, defined as $Co = L / (R \times C)$, where L is the total number
 17 of interactions. As all interactions in a probabilistic network are assumed to be independent,
 18 the expected value of L , is

$$(3) \quad \hat{L} = \sum_4 A_{ij},$$

1 and $\hat{C}_o = \hat{L}/(R \times C)$.

2 The variance of the number of interactions is $\text{var}(\hat{L}) = \sum(A_{ij}(1 - A_{ij}))$.

3 *Node degree.* The degree distribution of a network is the distribution of the number of interac-
4 tions established (number of successors) and received (number of predecessors) by each node.

5 The expected degree of species i is

$$(4) \quad \hat{k}_i = \sum_j (A_{ij} + A_{ji})$$

6 The variance of the degree of each species is $\text{var}(\hat{k}_i) = \sum_j (A_{ij}(1 - A_{ij}) + A_{ji}(1 - A_{ji}))$. Note
7 also that as expected, $\sum \hat{k}_i = 2\hat{L}$.

8 *Generality and vulnerability.* By simplification of the above, generality \hat{g}_i and vulnerability \hat{v}_i are
9 given by, respectively, $\sum_j A_{ij}$ and $\sum_j A_{ji}$, with their variances $\sum_j A_{ij}(1 - A_{ij})$ and $\sum_j A_{ji}(1 - A_{ji})$.

10 **Emerging properties.**

11 *Path length.* Networks can be used to describe indirect interactions between species through
12 the use of paths. The existence of a path of length 2 between species i and j means that they
13 are connected through at least one additional species k . In a probabilistic network, unless
14 some elements are 0, all pairs of species i and j are connected through a path of length 1, with
15 probability A_{ij} . The expected number of paths of length k between species i and j is given by

$$(5) \quad n_{ij}^{(k)} = (\mathbf{A}^k)_{ij},$$

16 where \mathbf{A}^k is the matrix multiplied by itself k times.

17 It is possible to calculate the probability of having at least one path of length k between the
18 two species: this can be done by calculating the probability of having no path of length k ,
19 then multiplying the resulting array of probabilities. For the example of length 2, species i

1 and j are connected through g with probability $A_{ig}A_{gj}$, and so this path does not exist with
2 probability $1 - A_{ig}A_{gj}$. For any pair i, j , let \mathbf{m} be the vector such as $m_g = A_{ig}A_{gj}$ for all
3 $g \notin (i, j)$ (Mirchandani 1976). The probability of not having any path of length 2 is $\prod(1 - \mathbf{m})$.
4 Therefore, the probability of having a path of length 2 between i and j is

$$(6) \quad \hat{p}_{ij}^{(2)} = 1 - \prod(1 - \mathbf{m}).$$

5 In most situations, one would be interested in knowing the probability of having a path of
6 length 2 *without* having a path of length 1; this is simply expressed as $(1 - A_{ij})\hat{p}_{ij}^{(2)}$. One can,
7 by the same logic, generate the expression for having at least one path of length 3:

$$(7) \quad \hat{p}_{ij}^{(3)} = (1 - A_{ij})(1 - \hat{p}_{ij}^{(2)}) (1 - \prod(1 - \mathbf{m})) \prod_{x,y} ((1 - A_{ix})(1 - A_{xy})(1 - A_{yj})),$$

8 where \mathbf{m} is the vector of all $A_{ix}A_{xy}A_{yj}$ for $x \notin (i, j), y \neq x$. This gives the probability of having
9 at least one path from i to j , passing through any pair of nodes x and y , without having any
10 shorter path. In theory, this approach can be generalized up to an arbitrary path length, but it
11 becomes rapidly untractable.

12 *Unipartite projection of bipartite networks.* The unipartite projection of a bipartite network is
13 obtained by linking any two nodes of one mode that are connected through at least one node
14 of the other mode; for example, to plants are connected if they share at least one pollinator. It is
15 readily obtained using the formula in the *Path length* section. This yields either the probability
16 of an edge in the unipartite projection (of the upper or lower nodes), or if using the matrix
17 multiplication, the expected number of such nodes.

18 *Nestedness.* Nestedness is an important measure of (bipartite) network structure that tells the
19 extent to which the interactions of specialists and generalists overlap. We use the formula
20 for nestedness proposed by Bastolla et al. (2009). They define nestedness for each margin of

1 the matrix, as $\eta^{(R)}$ and $\eta^{(C)}$ for, respectively, rows and columns. As per Almeida-Neto et al.
 2 (2008), we define a global statistic for nestedness as $\eta = (\eta^{(R)} + \eta^{(C)})/2$.
 3 Nestedness, in a probabilistic network, is defined as

$$(8) \quad \eta^{(R)} = \sum_{i < j} \frac{\sum_k A_{ik} A_{jk}}{\min(g_i, g_j)},$$

4 where g_i is the expected generality of species i . The reciprocal holds for $\eta^{(C)}$ when using v_i
 5 (the vulnerability) instead of g_i .
 6 The values returned are within $[0; 1]$, with $\eta = 1$ indicating complete nestedness.

7 *Modularity*. Modularity represents the extent to which networks are compartmentalized, *i.e.*
 8 the tendency for subsets of species to be strongly connected together, while they are weakly
 9 connected to the rest of the network (Stouffer & Bascompte 2011). Modularity is measured as
 10 the proportion of interactions between nodes of an arbitrary number of modules, as opposed
 11 to the random expectation. Assuming a vector \mathbf{s} which, for each node in the network, holds
 12 the value of the module it belongs to (an integer in $[1, c]$), Newman (2004) proposed a general
 13 measure of modularity, which is

$$Q = \sum_{m=1}^c (e_{mm} - a_m^2)$$

14 , where c is the number of modules,

$$e_{mm} = \sum_{ij} \frac{A_{ij}}{2c} \delta(\mathbf{s}_i, \mathbf{s}_j)$$

15 , and

$$a_m = \sum_n e_{mn}$$

16 ,

- 1 with δ being Kronecker's function, returning 1 if its arguments are equal, and 0 otherwise.
- 2 This formula can be *directly* applied to probabilistic networks.

3 *Centrality.* Although node degree is a rough first order estimate of centrality, other measures
 4 are often needed. We derive the expected value of centrality according to Katz (1953). This
 5 measure generalizes to directed acyclic graphs (whereas other do not). Although eigenvector
 6 centrality is often used in ecology, it cannot be measured on probabilistic graphs. Eigenvector
 7 centrality requires the matrix's largest eigenvalues to be real, which is not the case for all
 8 probabilistic matrices. The measure proposed by Katz is a useful replacement, because it uses
 9 the paths of all length between two species instead of focusing on the shortest path.

10 The expected number of paths of length k between i and j is $(\mathbf{A}^k)_{ij}$. Based on this, the expected
 11 centrality of species i is

$$(9) \quad C_i = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (\mathbf{A}^k)_{ji}.$$

12 The parameter $\alpha \in [0; 1]$ regulates how important long paths are. When $\alpha = 0$, only first-order
 13 paths are accounted for (and the centrality is equal to generality). %DG: to the degree or
 14 generality? When $\alpha = 1$, paths of all length are equally important. As C_i is sensitive to the
 15 size of the matrix, we suggest normalizing by $\mathbf{C} = \sum C$, so that

$$(10) \quad C_i = \frac{C_i}{\mathbf{C}}.$$

16 This results in the *expected relative centrality* of each node in the probabilistic network.

17 *Species with no outgoing links.* Estimating the number of species with no outgoing links (succes-
 18 sors) can be useful when predicting whether, *e.g.*, predators will go extinct. A species has no

- 1 successots if it manages *not* to establish any outgoing interaction, which for species i happens
 2 with probability

$$(11) \quad \prod_j (1 - A_{ij}).$$

- 3 The number of expected such species is therefore the sum of the above across all species:

$$(12) \quad \hat{P}P = \sum_i \left(\prod_j (1 - A_{ij}) \right).$$

- 4 and its variance is

$$(13) \quad \text{var}(\hat{P}P) = \sum_i \left(\prod_j (1 - A_{ij}^2) - \left(\prod_j (1 - A_{ij}) \right)^2 \right)$$

- 5 *Species with no incoming links.* Using the same approach as for the number of species with no
 6 outgoing links, the expected number of species with no incoming links is therefore

$$(14) \quad \hat{T}P = \sum_i \left(\prod_{j \neq i} (1 - A_{ji}) \right)$$

- 7 Note that we exclude self-interactions, as top-predators can, and often do, engage in cannibal-
 8 ism.

- 9 *Number of species with no interactions.* Predicting the number of species with no interactions
 10 (or whether any species will have at least one interaction) is useful when predicting whether
 11 species will be able to integrate into an existing network, for example. Note that from a
 12 methodological point of view, this can be a helpful *a priori* measure to determine whether null
 13 models of networks will have a lot of species with no interactions, and so will require intensive
 14 sampling.

1 A species has no interactions with probability

$$(15) \quad \prod_{j \neq i} (1 - A_{ij})(1 - A_{ji})$$

2 As for the above, the expected number of species with no interactions (*free species*) is the sum
3 of this quantity across all i :

$$(16) \quad \hat{F}S = \sum_i \prod_{j \neq i} (1 - A_{ij})(1 - A_{ji})$$

4 The variance of the number of species with no interactions is

$$(17) \quad \text{var}(\hat{F}S) = \sum_i \left(A_{ij}(1 - A_{ij})A_{ji}(1 - A_{ji}) + A_{ij}(1 - A_{ij})A_{ji}^2 + A_{ji}(1 - A_{ji})A_{ij}^2 \right)$$

5 *Self-loops.* Self-loops (the existence of an interaction of a species onto itself) is only meaningful
6 in unipartite networks. The expected proportion of species with self-loops is very simply
7 defined as $\text{Tr}(\mathbf{A})$, that is, the sum of all diagonal elements. The variance is $\text{Tr}(\mathbf{A} \diamond (1 - \mathbf{A}))$,
8 where \diamond is the element-wise product operation.

9 *Motifs.* Motifs are sets of pre-determined interactions between a fixed number of species (Milo
10 *et al.* 2002), such as for example one predator sharing two preys. As there are an arbitrarily
11 large number of motifs, we will illustrate the formulae with only two examples.

12 The probability that three species form an apparent competition motif (one predator, two
13 preys) where i is the predator, j and k are the preys, is

$$(18) \quad P(i, j, k \in \text{app. comp}) = A_{ij}(1 - A_{ji})A_{ik}(1 - A_{ki})(1 - A_{jk})(1 - A_{kj})$$

1 Similarly, the probability that these three species form an omnivory motif, in which i and j
 2 consume k , and i consumes j , is

$$(19) \quad P(i, j, k \in \text{omniv.}) = A_{ij}(1 - A_{ji})A_{ik}(1 - A_{ki})A_{jk}(1 - A_{kj})$$

3 The probability of the number of *any* motif m with three species in a network is given by

$$(20) \quad \hat{N}_m = \sum_i \sum_{j \neq i} \sum_{k \neq j} P(i, j, k \in m)$$

4 It is indeed possible to have an expression of the variance of this value, or of the variance of
 5 any three species forming a given motif, but their expressions become rapidly untractable and
 6 are better computed than written.

7 **Network comparison.** The dissimilarity of a pair of (ecological) networks can be measured
 8 using the framework set forth by Koleff et al. (2003). Measures of β -diversity compute the
 9 dissimilarity between two networks based on the cardinality of three sets, a , c , and b , which
 10 are respectively the shared items, items unique to superset (network) 1, and items unique to
 11 superset 2 (the identity of which network is 1 or 2 matters for asymmetric measures). Supersets
 12 can be the species within each network, or the interactions between species. Following Poisot
 13 et al. (2012), the dissimilarity of two networks can be measured as either β_{WN} (all interactions),
 14 or β_{OS} (interactions involving only common species), with $\beta_{OS} \leq \beta_{WN}$.

15 Within our framework, these measures can be applied to probabilistic networks. The expected
 16 values of \bar{a} , \bar{c} , and \bar{b} are, respectively, $\sum \mathbf{A}_1 \diamond \mathbf{A}_2$, $\sum \mathbf{A}_1 \diamond (1 - \mathbf{A}_2)$, and $\sum (1 - \mathbf{A}_1) \diamond \mathbf{A}_2$. Whether
 17 β_{OS} or β_{WN} is measured requires to alter the matrices \mathbf{A}_1 and \mathbf{A}_2 . To measure β_{OS} , one must
 18 remove all unique species; to measure β_{WN} , one must expand the two matrices so that they
 19 have the same species at the same place, and give a weight of 0 to the added interactions.

In this section, we will provide an overview of the applications of probabilistic network measures. The current way of dealing with probabilistic interactions are either to ignore variability entirely or to generate random networks. Probabilistic metrics are a mathematically rigorous alternative to both. When ignoring the probabilistic nature of interactions, what we call *Binary* from here on, every non-zero element of the network is assumed to be 1. This leads to over-representation of some rare events, and increases the number of interactions. An alternative is to consider only the interactions above a given threshold, which leads to an under-representation of rare events and decreases the number of interactions.

When generating random networks, what we call *Bernoulli trials* from here on, a binary network is generated by doing a Bernoulli trial with probability A_{ij} , for each element of the matrix. This is problematic because higher order structures involving rare events will be under-represented in the sample, and because most naive approaches are likely to generate free species, especially in sparsely connected networks frequently encountered in ecology (Milo *et al.* 2003; Poisot & Gravel 2014) – on the other hand, non-naive approaches break the assumption of independence between interactions.

Comparison of probabilistic networks. In this sub-section, we apply the above measures to a bacteria–phage interaction network. Poullain *et al.* (2008) have measured the probability that 24 phages can infect 24 strains of bacteria of the *Pseudomonas fluorescens* species (group SBW25). Each probability has been observed through independent infection assays, and can take values of 0, 0.5, and 1.0.

Measuring the structure of the Binary, Bernoulli trials, and Probabilistic network gives the following result:

| Measure | Binary | Bernoulli trials | Probabilistic |
|---------|--------|--------------------|--------------------|
| links | 336 | 221.58 ± 57.57 | 221.52 ± 57.25 |
| η | 0.73 | 0.528 | 0.512 |

| Measure | Binary | Bernoulli trials | Probabilistic |
|--------------|--------|------------------|---------------|
| $\eta^{(R)}$ | 0.72 | 0.525 | 0.507 |
| $\eta^{(C)}$ | 0.75 | 0.531 | 0.518 |

As these results show, transforming the probabilistic matrix into a binary one (i) overestimates nestedness by ≈ 0.2 , and (ii) overestimates the number of links by 115. For the number of links, both the probabilistic measures and the average and variance of 10^4 Bernoulli trials were in strong agreement (they differ only by the second decimal place).

Using Bernoulli trials had the effect of slightly over-estimating nestedness. The overestimation is statistically significant from a purely frequentist point of view, but significance testing is rather meaningless when the number of replicates is this large and can be increased arbitrarily; what is important is that the relative value of the error is small enough that Bernoulli trials are able to adequately reproduce the probabilistic structure of the network. It is not unexpected that Bernoulli trials are this close to the analytical expression of the measures; due to the experimental design of the Poullain et al. (2008) study, probabilities of interactions are bound to be high, and so variance is minimal (most elements of \mathbf{A} have a value of either 0 or 1, and so their individual variance is 0). Still, despite overall low variance, the binary approach severely mis-represents the structure of the network.

Null-model based hypothesis testing. In this section, we analyse 59 pollination networks from the literature using two usual null models of network structure, and two models with intermediate constraints. These data cover a wide range of situations, from small to large, and from densely to sparsely connected networks. They provide a good demonstration of the performance of probabilistic metrics. Data come from the *InteractionWeb Database*, and were queried on Nov. 2014.

We use the following null models. First (Type I, Fortuna & Bascompte (2006)), any interaction between plant and animals happens with the fixed probability $P = C_0$. This model controls for connectance, but removes the effect of degree distribution. Second, (Type II, Bascompte et al.

(2003)), the probability of an interaction between animal i and plant j is $(k_i/R + k_j/C)/2$, the average of the richness-standardized degree of both species. In addition, we use the models called Type III in and out (Poisot *et al.* 2013), that uses respectively the row-wise and column-wise probability of an interaction, as a way to understand the impact of the degree distribution of upper and lower level species.

Note that these null models will take a binary network, and through some rules turn it into a probabilistic one. Typically, this probabilistic network is used as a template to generate Bernoulli trials and measure some of their properties, the distribution of which is compared to the empirical network. This approach is computationally inefficient (Poisot & Gravel 2014), especially using naive models (Milo *et al.* 2003), and as we show in the previous section, can yield biased estimates of the true average of nestedness (and presumably other properties).

%DG: there is something I don't get in the logic here. You mean that you randomize % the network according to some rules, do it many times, compute a probabilistic network % and then do a Bernoulli trial out of it? Why are these last two steps necessary? I % figure it's not what you meant, but some rewording would help clarifying your thoughts % After a couple of readings I realize that you need to better explain how to i) generate \$ the null expectation analytically instead of trials and ii) test for hypotheses \$ It is critical, because otherwise the comparison you are doing below is confusing.

We measured the nestedness of the 59 networks, then generated the random networks under the four null models, and calculated the expected nestedness. For each null model i , the difference $\Delta_N^{(i)}$ in nestedness N is expressed as $\Delta_N^{(i)} = N - \mathcal{N}^{(i)}(N)$, where $\mathcal{N}^{(i)}(N)$ is the nestedness of null model i . Our results are presented in [Figure 1](#).

There are two striking results. %DG: strange way to report the difference between the null model and the data. % Makes more sense to say the data is consistently more nested than the null expectation First, null models consistently *underestimate* the nestedness of the 59 polination networks, as evidenced by the fact that all Δ_N values are strictly positive. %DG: not clear what you mean by linear Second, this underestimation is *linear* between null models I and II, although null model II is always closer to the *actual* nestedness value. %DG: what do you mean by non-random value of the null model??? The markedly non-random value of

1 the null nestednesses when compared to the empirical values calls for a closer evaluation of
 2 how the results of null models are interpreted (especially since Bernoulli simulations revealed
 3 a very low variance in the simulated nestedness). %DG: here the mistake comes from the
 4 comparison of random trials and the analytical % null expectation? Interestingly, models III
 5 in and III out made overall *fewer* mistakes at estimating nestedness – resp. 0.129 and 0.123,
 6 compared to resp. 0.219 and 0.156 for model I and II. Although the error is overall sensitive
 7 to model type (Kruskal-Wallis $\chi^2 = 35.80$, d.f. = 3, $p \leq 10^{-4}$), the three pairs of models that
 8 where significantly different after controlling for multiple comparisons are I and II, I and III
 9 in, and I and III out (model II is not different from either models III in or out).
 10 In short, this analysis reveals that (i) the null expectation of a network property under ran-
 11 domisation scenarios can be obtained through the analysis of the probabilistic matrix, instead
 12 of the analysis of simulated Bernoulli networks; (ii) Different models have different systematic
 13 biases, with models of the type III performing overall better for nestedness than any other
 14 models. This can be explained by the fact that nestedness of a network, as expressed by Bas-
 15 tolla et al. (2009), is the average of a row-wise and column-wise nestedness. These depend on
 16 the species degree, and as such should be well predicted by models III.

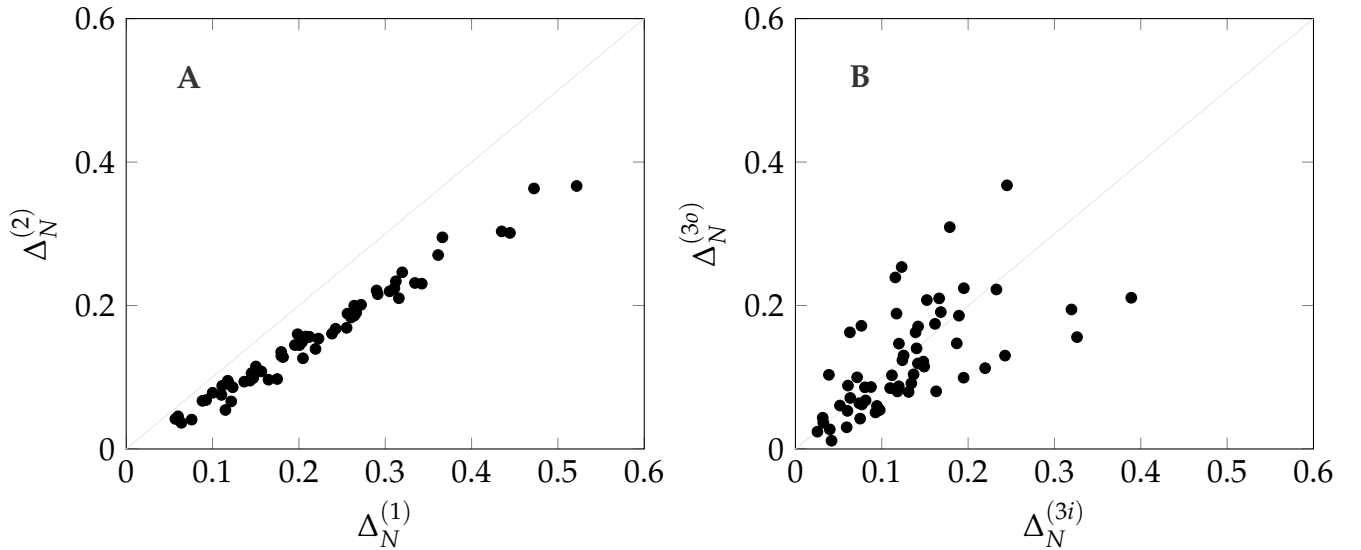


FIGURE 1. Results of the null model analysis of 59 plant-pollination networks.
A. There is a consistent tendency for (i) both models I and II to estimate less
 nestedness than in the empirical network, although null model II yields more
 accurate estimates. **B.** Models III in and III out also estimate less nestedness than
 the empirical network, but neither has a systematic bias.

We developed and presented a set of measures to quantify the expected network structure, using the probability that each interaction is observed or happens, in a way that do not require time-consuming simulations. Our framework is set up in such a way that the probabilities of interactions are considered to be independent. Estimating interaction probabilities based on species abundances (Olito & Fox 2014; Canard *et al.* 2014) do not, for example, yield non-independant probabilities: changing the abundance of one species changes all probabilities in the network. They are not Bernoulli events either, as the sum of all probabilities derived this way sums to unity. On the other hand, “cafeteria experiments” give truly independant probabilities of interactions; even a simple criteria, such as the frequency of interactions when the two species are put together, is a way of estimating probability. Using the approach outline by (???), both sources of information (species abundance and the outcome of experiments) can be combined to estimate the probability that interactions will happen in empirical communities. This effort requires improved communications between scientists collecting data and scientists developing methodology to analyse them.

Another way to obtain approximation of the probability of interactions is to use spatially replicated sampling. Some studies (Tylianakis *et al.* 2007; Olito & Fox 2014; Carstensen *et al.* 2014; Trøjelsgaard *et al.* 2015) surveyed the existence of interactions at different locations, and a simple approach of dividing the number of observations of an interaction by the number of co-occurrence of the species involved will provide a (somewhat crude) estimate of the probability of this interaction. This approach requires extensive sampling, especially since interactions are harder to observe than species (???), yet it enables the re-analysis of existing datasets in a probabilistic context.

Understanding the structure of ecological networks, and whether it relates to ecosystem properties, is emerging as a key challenge for community ecology. A proper estimation of this structure requires tools that address all forms of complexity, the most oft-neglected yet pervasive of which is the fact that interactions are variable. By developing these metrics, we allow future analyses of network structure to account for this phenomenon.

1 **Acknowledgements:** This work was funded by a CIEE working group grant to TP, DG, and
2 DBS. TP is funded by a starting grant from the Université de Montréal.

3 REFERENCES

- 4 Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D. & Ulrich, W. (2008). A con-
5 sistent metric for nestedness analysis in ecological systems: reconciling concept and measure-
6 ment. *Oikos*, **117**, 1227–1239. Retrieved October 10, 2014,
- 7 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plant-
8 animal mutualistic networks. *Proceedings of the National Academy of Sciences of the United States*
9 *of America*, **100**, 9383–9387.
- 10 Bastolla, U., Fortuna, M.A., Pascual-García, A., Ferrera, A., Luque, B. & Bascompte, J. (2009).
11 The architecture of mutualistic networks minimizes competition and increases biodiversity.
12 *Nature*, **458**, 1018–1020. Retrieved October 10, 2014,
- 13 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). Empir-
14 ical evaluation of neutral interactions in host-parasite networks. *The American Naturalist*, **183**,
15 468–479.
- 16 Carstensen, D.W., Sabatino, M., Trøjelsgaard, K. & Morellato, L.P.C. (2014). Beta Diversity of
17 Plant-Pollinator Networks and the Spatial Turnover of Pairwise Interactions. *PLoS ONE*, **9**,
18 e112903. Retrieved November 19, 2014,
- 19 Chamberlain, S.A., Cartar, R.V., Worley, A.C., Semmler, S.J., Gielens, G., Elwell, S., Evans, M.E.,
20 Vamosi, J.C. & Elle, E. (2014). Traits and phylogenetic history contribute to network structure
21 across Canadian plant–pollinator communities. *Oecologia*, 1–12. Retrieved September 11, 2014,
22 from <http://link.springer.com/article/10.1007/s00442-014-3035-2>
- 23 Duffy, J.E. (2002). Biodiversity and ecosystem function: the consumer connection. *Oikos*, **99**,
24 201–219.
- 25 Dunne, J.A. (2006). The Network Structure of Food Webs. *Ecological networks: Linking structure*
26 *and dynamics* (eds J.A. Dunne & M. Pascual), pp. 27–86. Oxford University Press.

1 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plant–animal mutual-
2 istic networks. *Ecology Letters*, **9**, 281–286. Retrieved January 21, 2015,

3 Haerter, J.O., Mitarai, N. & Sneppen, K. (2014). Phage and bacteria support mutual diversity
4 in a narrowing staircase of coexistence. *The ISME journal*.

5 Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: con-
6 nectance, dependence asymmetries, and coevolution. *The American Naturalist*, **129**, 657–677.

7 Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, **18**, 39–43.
8 Retrieved October 9, 2014,

9 Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence-absence
10 data. *Journal of Animal Ecology*, **72**, 367–382.

11 McCann, K.S. (2014). Diversity and Destructive Oscillations: Camerano, Elton, and May. *Bul-*
12 *letin of the Ecological Society of America*, **95**, 337–340. Retrieved October 7, 2014,

13 Milo, R., Kashtan, N., Itzkovitz, S., Newman, M.E.J. & Alon, U. (2003). On the uniform gener-
14 ation of random graphs with prescribed degree sequences. *arXiv:cond-mat/0312028*. Retrieved
15 October 9, 2014, from <http://arxiv.org/abs/cond-mat/0312028>

16 Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002). Network
17 motifs: simple building blocks of complex networks. *Science (New York, N.Y.)*, **298**, 824–7.

18 Mirchandani, P.B. (1976). Shortest distance and reliability of probabilistic networks. *Computers*
19 *& Operations Research*, **3**, 347–355. Retrieved October 12, 2014, from [http://www.sciencedirect.](http://www.sciencedirect.com/science/article/pii/0305054876900174)
20 [com/science/article/pii/0305054876900174](http://www.sciencedirect.com/science/article/pii/0305054876900174)

21 Newman, M.E.J. (2004). Fast algorithm for detecting community structure in networks. *Physi-*
22 *cal Review E*, **69**, 066133. Retrieved February 24, 2015,

23 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2011).
24 Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biologi-*
25 *cal Sciences*, **278**, 725–732. Retrieved October 7, 2014,

26 Olito, C. & Fox, J.W. (2014). Species traits and abundances predict metrics of plant–pollinator
27 network structure, but not pairwise interactions. *Oikos*, n/a–n/a. Retrieved January 19, 2015,

- 1 PoisoT, T. (2012). L'ABC de la spécialisation: apparition, biodiversité, conservation. *Le Prisme*
2 à Idées, **4**, 49–52.
- 3 PoisoT, T. & Gravel, D. (2014). When is an ecological network complex? Connectance drives
4 degree distribution and emerging network properties. *PeerJ*, **2**, e251. Retrieved September 13,
5 2014,
- 6 PoisoT, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of
7 species interaction networks. *Ecology Letters*, **15**, 1353–1361. Retrieved November 3, 2014,
- 8 PoisoT, T., Lounnas, M. & Hochberg, M.E. (2013). The structure of natural microbial enemy-
9 victim networks. *Ecological Processes*.
- 10 PoisoT, T., Stouffer, D.B. & Gravel, D. (2014). Beyond species: why ecological interaction
11 networks vary through space and time. *Oikos*, n/a–n/a. Retrieved November 5, 2014,
- 12 Poullain, V., Gandon, S., Brockhurst, M.A., Buckling, A. & Hochberg, M.E. (2008). The evolu-
13 tion of specificity in evolving and coevolving antagonistic interactions between a bacteria and
14 its phage. *Evolution*, **62**, 1–11. Retrieved October 10, 2014,
- 15 Stouffer, D.B. & Bascompte, J. (2011). Compartmentalization increases food-web persistence.
16 *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3648–3652.
- 17 Thébault, E. & Loreau, M. (2003). Food-web constraints on biodiversity–ecosystem functioning
18 relationships. *Proceedings of the National Academy of Sciences of the United States of America*, **100**,
19 14949–14954.
- 20 Trøjelsgaard, K., Jordano, P., Carstensen, D.W. & Olesen, J.M. (2015). Geographical variation in
21 mutualistic networks: similarity, turnover and partner fidelity. *Proceedings of the Royal Society*
22 *B: Biological Sciences*, **282**, 20142925. Retrieved February 1, 2015,
- 23 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). Habitat modification alters the structure
24 of tropical host–parasitoid food webs. *Nature*, **445**, 202–205. Retrieved January 28, 2013, from
25 <http://www.nature.com/nature/journal/v445/n7124/abs/nature05429.html>

- 1 Yeakel, J.D., Guimarães, P.R., Novak, M., Fox-Dobbs, K. & Koch, P.L. (2012). Probabilistic
- 2 patterns of interaction: the effects of link-strength variability on food web structure. *Journal of*
- 3 *The Royal Society Interface*, rsif20120481. Retrieved November 3, 2014,