

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.

Keywords: ecological networks, connectance, degree distribution, nestedness, modularity

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~~[et al. 2013b](#)). 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 *emergent* (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 emergent property that is not directly predictable from the degree of all nodes. Though the difference may appear to be semantics, establishing a difference between direct and emergent properties is important when interpreting their values; direct properties are conceptually equivalent to means, in that they tend to be the first moment of network units, whereas emergent 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 emergent property of the response of ecological interactions to environmental factors and chance events (Poisot *et al.* 2014). Interactions can vary because of local mismatching in phenology (Olesen *et al.* 2011), populations fluctuations preventing the interaction (Canard

47 *et al.* 2014), or a combination of both (Chamberlain *et al.* 2014; Olito & Fox 2014). For example, Olito
48 & Fox (2014) show that accounting for neutral (population-size driven) and trait-based effects allows the
49 prediction of the cumulative change in network structure, but not of the change at the level of individual
50 interactions. In addition, Carstensen *et al.* (2014) show that within a meta-community, not all interactions
51 are equally variable: some are highly consistent, whereas others are extremely rare. These empirical re-
52 sults all point to the fact that species interactions cannot always be adequately modeled as yes-no events;
53 since it is well established that they do vary, it is necessary to represent them as probabilities. To the
54 question of *Do these two species interact?*, we should substitute the question of *How likely is it that they*
55 *will interact?*.

56 The current way of dealing with probabilistic interactions are either to ignore variability entirely or to
57 generate random networks. Probabilistic metrics are a mathematically rigorous alternative to both. When
58 ignoring the probabilistic nature of interactions (henceforth *binary* networks), every non-zero element
59 of the network is assumed to be 1. This leads to over-representation of some rare events, and increases
60 the number of interactions; as a result, this changes the estimated value of different network properties,
61 in a way that is not understood at all. An alternative is to consider only the interactions above a given
62 threshold, which leads to an under-representation of rare events and decreases the effective number of
63 interactions (in addition to the problem that there is no robust criterion to decide on a threshold). More
64 importantly, this introduces the risk of removing species that establish a lot of interactions that each
65 have a low probability. Taken together, these considerations highlight the need to amend our current
66 methodology for the description of ecological networks, in order to give more importance to the variation
67 of individual interactions — current measures neglect the variability of interactions, and are therefore
68 discarding valuable ecological information. Because the methodological corpus available to describe
69 ecological networks had first been crafted at a time when it was assumed that interactions were invariants,
70 it is unsuited to address the questions that probabilistic networks allow us to ask.

71 In this paper, we show that several direct and emergent core properties of ecological networks (both bipar-
72 tite and unipartite) can be re-formulated in a probabilistic context (Yeakel *et al.* 2012; Poisot *et al.* 2014);
73 we conclude by showing how this methodology can be applied to exploit the information contained in the
74 variability of networks, and to reduce the computational burden of current methods in network analysis.

75 We also provide a free and open-source (MIT license) implementation of this suite of measures in a library
 76 for the `julia` language, available at <http://github.com/PoisotLab/ProbabilisticNetwork.jl>.

77 SUITE OF PROBABILISTIC NETWORK METRICS

78 Throughout this paper, we use the following notation. \mathbf{A} is a matrix wherein A_{ij} is $P(ij)$, *i.e.* the proba-
 79 bility that species i establishes an interaction with species j . If \mathbf{A} represents a unipartite network (*e.g.* a
 80 food web), it is a square matrix and contains the probabilities of each species interacting with all others,
 81 including itself. If \mathbf{A} represents a bipartite network (*e.g.* a pollination network), it will not necessarily
 82 be square. We call S the number of species, and R and C respectively the number of rows and columns.
 83 $S = R = C$ in unipartite networks, and $S = R + C$ in bipartite networks.

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

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

86 where $0_{(C,R)}$ is a matrix of C rows and R columns (noted $C \times R$) filled with 0s, etc. Note that for centrality
 87 to be relevant in bipartite networks, this matrix should be made symmetric: $\mathbf{B}_{ij} = \mathbf{B}_{ji}$.

88 We will also assume that all interactions are independent (so that $P(ij|kl) = P(ij)P(kl)$ for any species),
 89 and can be represented as a series of Bernoulli trials (so that $0 \leq P(ij) \leq 1$). The latter condition allows
 90 us to derive estimates for the *variance* ($\text{var}(X) = p(1 - p)$), and expected values ($E(X) = p$). We can
 91 therefore estimate the variance of most properties, using the fact that the variance of additive independent
 92 events is the sum of their individual variances, and that the variance of multiplicative independent events
 93 is

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

94 As a final note, all of the measures described below can be applied on the binary (0/1) versions of the
 95 networks ~~and will give the exact value of the~~ in which case they effectively are the non-probabilistic
 96 version of the measure. This property is particularly desirable as it allows our framework to be used on
 97 any network, whether they are represented in a probabilistic or binary way. Nonetheless, this approach is
 98 different from using *weighted* networks, in that it answers a completely different question. Probabilistic
 99 networks describe the probability that any interaction will happen, whereas weighted networks describe
 100 the effect of the interaction when it happens. Actually, the weight of each interaction is best viewed as a
 101 second modeling step, focusing only on the non-zero cases (*i.e.* the interactions that are realized); this is
 102 similar to the method now frequently used in species distribution models, where the species presence is
 103 modeled first, and its abundance second, using a (possibly) different set of predictors (Boulangéat *et al.*
 104 2012).

105 **Direct properties.**

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

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

109 and $\hat{Co} = \hat{L}/(R \times C)$. Likewise, the variance of the number of interactions is $\text{var}(\hat{L}) = \sum (A_{ij}(1 - A_{ij}))$.

110 *Node degree.* The degree distribution of a network is the distribution of the number of interactions estab-
 111 lished (number of successors) and received (number of predecessors) by each node. The expected degree
 112 of species i is

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

113 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 also that as
 114 expected, $\sum \hat{k}_i = 2\hat{L}$.

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

117 **emergent properties.**

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

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

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

124 It is possible to calculate the probability of having at least one path of length k between the two species:
 125 this can be done by calculating the probability of having no path of length k , then taking the running
 126 product of the resulting array of probabilities. For the example of length 2, species i and j are connected
 127 through g with probability $A_{ig}A_{gj}$, and so this path does not exist with probability $1 - A_{ig}A_{gj}$. For any
 128 pair i, j , let \mathbf{m} be the vector such as $m_g = A_{ig}A_{gj}$ for all $g \notin (i, j)$ (Mirchandani 1976). The probability
 129 of not having any path of length 2 is $\prod(1 - \mathbf{m})$. Therefore, the probability of having a path of length 2
 130 between i and j is

$$(6) \quad \hat{p}_{ij}^{(2)} = 1 - \prod_6 (1 - \mathbf{m}).$$

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

$$(7) \quad \hat{p}_{ij}^{(3)} = (1 - A_{ij})(1 - \hat{p}_{ij}^{(2)}) \left(1 - \prod_{x,y} (1 - \mathbf{m})\right) \prod_{x,y} ((1 - A_{iy})(1 - A_{xj})),$$

134 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 at least one
 135 path from i to j , passing through any pair of nodes x and y , without having any shorter path. In theory,
 136 this approach can be generalized up to an arbitrary path length, but it becomes rapidly untractable.

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

142 *Nestedness.* Nestedness is an important measure of (bipartite) network structure that tells the extent to
 143 which the interactions of specialists and generalists overlap. We use the formula for nestedness proposed
 144 by Bastolla et al. (2009). They define nestedness for each margin of the matrix, as $\eta^{(R)}$ and $\eta^{(C)}$ for,
 145 respectively, rows and columns. As per Almeida-Neto et al. (2008), we define a global statistic for
 146 nestedness as $\eta = (\eta^{(R)} + \eta^{(C)})/2$.

147 Nestedness, in a probabilistic network, is defined as

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

148 where g_i is the expected generality of species i . The reciprocal holds for $\eta^{(C)}$ when using v_i (the vulner-
 149 ability) instead of g_i .

150 The values returned are within $[0; 1]$, with $\eta = 1$ indicating complete nestedness.

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

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

157 , where c is the number of modules,

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

158 , and

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

159 ,

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

162 *Centrality.* Although node degree is a rough first order estimate of centrality, other measures are often
 163 needed. We derive the expected value of centrality according to Katz (1953). This measures general-
 164 izes to directed acyclic graphs (whereas other do not). For example, although eigenvector centrality is
 165 often used in ecology, it cannot be measured on probabilistic graphs. Eigenvector centrality requires the
 166 matrix's largest eigenvalues to be real, which is not the case for all probabilistic matrices. The measure
 167 proposed by Katz is a useful replacement, because it accounts for the paths of all length between two
 168 species instead of focusing on the shortest path.

169 As described above, the expected number of paths of length k between i and j is $(\mathbf{A}^k)_{ij}$. Based on this,
 170 the expected centrality of species i is

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

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

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

175 This results in the *expected relative centrality* of each node in the probabilistic network, which sums to
 176 unity.

177 *Species with no outgoing links.* Estimating the number of species with no outgoing links (successors)
 178 can be useful when predicting whether, *e.g.*, predators will go extinct. A species has no successors if it
 179 manages *not* to establish any outgoing interaction, which for species i happens with probability

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

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

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

181 and its variance is

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

182 Note that in a non-probabilistic context, species with no outgoing links would be considered primary
183 producers. This is not the case here: if interactions are probabilistic events, then *e.g.* a top predator may
184 have no preys, which do not mean it will not become a primary producer. For this reason, the trophic
185 position of the species may better be measured on the binary version of the matrix.

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

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

188 Note that we exclude self-interactions, as top-predators can, and often do, engage in cannibalism.

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

194 A species has no interactions with probability

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

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

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

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

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

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

202 *Motifs.* Motifs are sets of pre-determined interactions between a fixed number of species (Milo *et al.*
 203 2002; Stouffer *et al.* 2007), such as for example one predator sharing two preys. As there are an arbitrarily
 204 large number of motifs, we will illustrate the approach with only two examples.

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

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

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

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

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

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

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

213 **Network comparison.** The dissimilarity of a pair of (ecological) networks can be measured using the
 214 framework set forth by Koleff et al. (2003). Measures of β -diversity compute the dissimilarity between
 215 two networks based on the cardinality of three sets, a , c , and b , which are respectively the shared items,
 216 items unique to superset (network) 1, and items unique to superset 2 (the identity of which network is 1 or
 217 2 matters for asymmetric measures). Supersets can be the species within each network, or the interactions
 218 between species. Following Poisot et al. (2012), the dissimilarity of two networks can be measured as
 219 either β_{WN} (all interactions), or β_{OS} (interactions involving only common species), with $\beta_{OS} \leq \beta_{WN}$.

220 Within our framework, these measures can be applied to probabilistic networks. The expected values of
 221 \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 β_{OS} or β_{WN} is
 222 measured requires to alter the matrices \mathbf{A}_1 and \mathbf{A}_2 . To measure β_{OS} , one must remove all unique species;
 223 to measure β_{WN} , one must expand the two matrices so that they have the same species at the same place,
 224 and give a weight of 0 to the added interactions.

225 APPLICATIONS

226 In this section, we contrast the use of probabilistic measures to the current approaches of either using
 227 binary networks, or working with null models through simulations. When generating random networks,
 228 what we call *Bernoulli trials* from here on, a binary network is generated by doing a Bernoulli trial
 229 with probability A_{ij} , for each element of the matrix. This is problematic because higher order struc-
 230 tures involving rare events will be under-represented in the sample, and because most naive approaches
 231 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 (interaction is variable), 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
$\eta^{(R)}$	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 – though their confidence interval varies as a

function of the number of observations from which the probability is derived). 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 $P = Co$. 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, 2013a), that use the row-wise and column-wise probability of an interaction respectively, 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).

We measured the nestedness of the 59 (binary) networks, then generated the random networks under the four null models, and calculated the expected nestedness using the probabilistic measure. 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 ??.

There are two striking results. First, empirical data are consistently *more* nested than the null expectation, as evidenced by the fact that all Δ_N values are strictly positive. Second, this underestimation is *linear* between null models I and II (in that it does not depend on how nested the empirical network is), although null model II is always closer to the nestedness of the empirical network (which makes sense, since null

281 model II incorporates the higher order constraint of respecting the degree distribution of both levels).
 282 That the nestedness of the null model probability matrix is so strongly determined by the nestedness
 283 of the empirical networks calls for a closer evaluation of how the results of null models are interpreted
 284 (especially since Bernoulli simulations revealed a very low variance in the simulated nestedness).
 285 There is a strong, and previously unaccounted for, circularity in this approach: empirical networks are
 286 compared to a null model which, as we show, has a systematic bias *and* a low variance (in simulations),
 287 meaning that differences in nestedness that are small (thus potentially ecologically irrelevant) have a
 288 good chance of being reported as significant. Interestingly, models III in and III out made overall *fewer*
 289 mistakes at estimating nestedness – resp. 0.129 and 0.123, compared to resp. 0.219 and 0.156 for
 290 model I and II. Although the error is overall sensitive to model type (Kruskal-Wallis $\chi^2 = 35.80$, d.f. =
 291 3, $p \leq 10^{-4}$), the three pairs of models that were significantly different after controlling for multiple
 292 comparisons are I and II, I and III in, and I and III out (model II is not different from either models III in
 293 or out).
 294 In short, this analysis reveals that (i) the null expectation of a network property under randomization
 295 scenarios can be obtained through the analysis of the probabilistic matrix, instead of the analysis of
 296 simulated Bernoulli networks; (ii) Different models have different systematic biases, with models of the

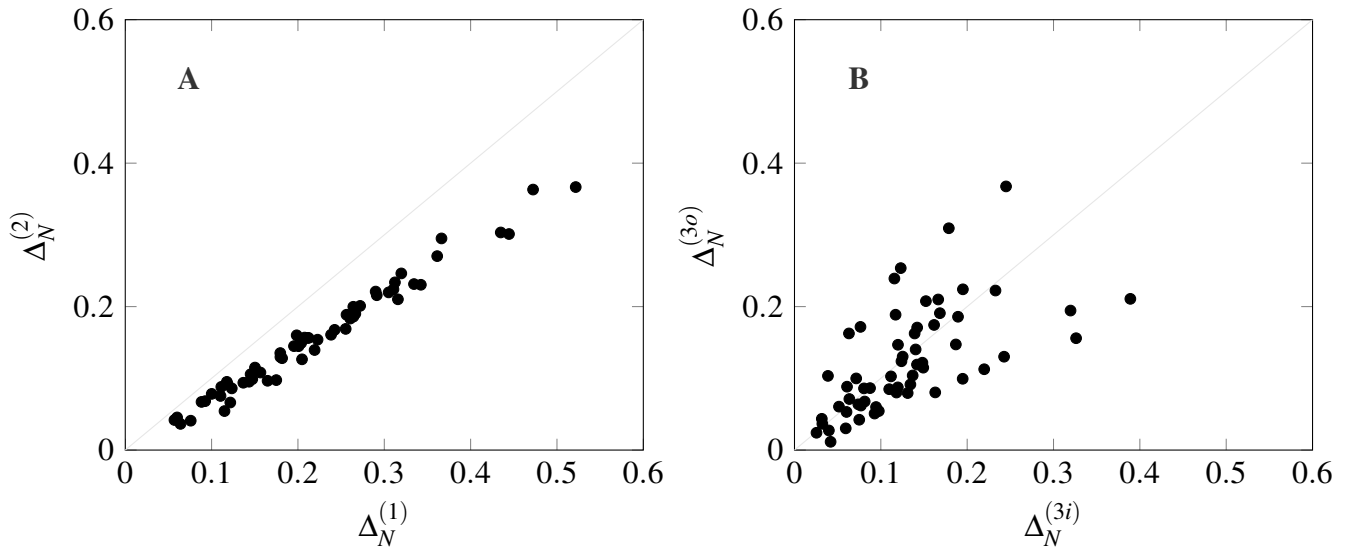


FIGURE 1. Results of the null model analysis of 59 plant-pollination networks. **A.** There is a consistent tendency for (i) both models I and II to estimate less nestedness than in the empirical network, although null model II yields more accurate estimates. **B.** Models III in and III out also estimate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

297 type III performing overall better for nestedness than any other models. This can be explained by the
298 fact that nestedness of a network, as expressed by Bastolla et al. (2009), is the average of a row-wise and
299 column-wise nestedness. These depend on the species degree, and as such should be well predicted by
300 models III.

301 ~~IMPLICATIONS FOR DATA COLLECTION~~DISCUSSION

302 Understanding the structure of ecological networks, and whether it relates to ecosystem properties, is
303 emergent as a key challenge for community ecology. A proper estimation of this structure requires
304 tools that address all forms of complexity, the most oft-neglected yet pervasive of which is the fact that
305 interactions are variable. By developing these metrics, we allow future analyses of network structure
306 to account for this phenomenon. There are two main considerations highlighted by this methodological
307 development. First, in what way are probabilistic data independent; second, what are the implications for
308 data collection.

309 **Non-independance of interactions.** We developed and presented a set of measures to quantify the ex-
310 pected network structure, using the probability that each interaction is observed or happens, in a way that
311 do not require time-consuming simulations. Our framework is set up in such a way that the probabilities
312 of interactions are considered to be independent.

313 Estimating interaction probabilities based on species abundances (Olito & Fox 2014; Canard *et al.* 2014)
314 do not, for example, yield independent probabilities: changing the abundance of one species changes all
315 probabilities in the network. They are not Bernoulli events either, as the sum of all probabilities derived
316 this way sums to unity. On the other hand, “cafeteria experiments” (in which two species are directly
317 exposed to one another to observe whether or not an interaction occurs) give truly independent probab-
318 ities of interactions; even a simple criteria, such as the frequency of interactions when the two species
319 are put together, is a way of estimating probability. Using the approach outline by Poisot et al. (2014),
320 both sources of information (species abundance and the outcome of experiments) can be combined to
321 estimate the probability that interactions will happen in empirical communities. This effort requires
322 improved communications between scientists collecting data and scientists developing methodology to
323 analyze them.

324 Another way to obtain approximation of the probability of interactions is to use spatially replicated sam-
325 pling. Some studies (Tylianakis *et al.* 2007; Olito & Fox 2014; Carstensen *et al.* 2014; Trøjelsgaard *et*
326 *al.* 2015) surveyed the existence of interactions at different locations, and a simple approach of divid-
327 ing the number of observations of an interaction by the number of co-occurrence of the species involved
328 will provide a (somewhat crude) estimate of the probability of this interaction. This approach requires
329 extensive sampling, especially since interactions are harder to observe than species (Poisot *et al.* 2012;
330 Gilarranz *et al.* 2014), yet it enables the re-analysis of existing datasets in a probabilistic context.

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

336 Implications for data collection. **Acknowledgements:** This work was funded by a CIEE working
337 group grant to TP, DG, and DBS. TP is funded by a starting grant from the Université de Montréal.

338 REFERENCES

- 339 Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D. & Ulrich, W. (2008). A consistent
340 metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**,
341 1227–1239. Retrieved October 10, 2014,
- 342 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003). The nested assembly of plant-animal
343 mutualistic networks. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 9383–9387.
- 344 Bastolla, U., Fortuna, M.A., Pascual-García, A., Ferrera, A., Luque, B. & Bascompte, J. (2009). The
345 architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, **458**,
346 1018–1020. Retrieved October 10, 2014,
- 347 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to
348 disentangle the drivers of species distributions and their abundances. *Ecol. Lett.*, **15**, 584–593.
- 349 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). Empirical
350 evaluation of neutral interactions in host-parasite networks. *The American Naturalist*, **183**, 468–479.

351 Carstensen, D.W., Sabatino, M., Trøjelsgaard, K. & Morellato, L.P.C. (2014). Beta Diversity of Plant-
 352 Pollinator Networks and the Spatial Turnover of Pairwise Interactions. *PLoS ONE*, **9**, e112903. Retrieved
 353 November 19, 2014,

354 Chamberlain, S.A., Cartar, R.V., Worley, A.C., Semmler, S.J., Gielens, G., Elwell, S., Evans, M.E.,
 355 Vamosi, J.C. & Elle, E. (2014). Traits and phylogenetic history contribute to network structure across
 356 Canadian plantpollinator communities. *Oecologia*, 1–12. Retrieved September 11, 2014, from <http://link.springer.com/article/10.1007/s00442-014-3035-2>
 357

358 Duffy, J.E. (2002). Biodiversity and ecosystem function: the consumer connection. *Oikos*, **99**, 201–219.

359 Dunne, J.A. (2006). The Network Structure of Food Webs. *Ecological networks: Linking structure and*
 360 *dynamics* (eds J.A. Dunne & M. Pascual), pp. 27–86. Oxford University Press.

361 Fortuna, M.A. & Bascompte, J. (2006). Habitat loss and the structure of plantanimal mutualistic net-
 362 works. *Ecol. Lett.*, **9**, 281–286. Retrieved January 21, 2015,

363 Gilarranz, L.J., Sabatino, M., Aizen, M.A. & Bascompte, J. (2014). Hot spots of mutualistic networks. *J*
 364 *Anim Ecol*, n/a–n/a. Retrieved November 17, 2014,

365 Haerter, J.O., Mitarai, N. & Sneppen, K. (2014). Phage and bacteria support mutual diversity in a nar-
 366 rowing staircase of coexistence. *ISME Journal*.

367 Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: connectance,
 368 dependence asymmetries, and coevolution. *Am. Nat.*, **129**, 657–677.

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

371 Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence-absence data. *J.*
 372 *Anim. Ecol.*, **72**, 367–382.

373 McCann, K.S. (2014). Diversity and Destructive Oscillations: Camerano, Elton, and May. *Bulletin of the*
 374 *Ecological Society of America*, **95**, 337–340. Retrieved October 7, 2014,

375 Milo, R., Kashtan, N., Itzkovitz, S., Newman, M.E.J. & Alon, U. (2003). On the uniform generation of
 376 random graphs with prescribed degree sequences. *ArXivcond-Mat0312028*. Retrieved October 9, 2014,
 377 from <http://arxiv.org/abs/cond-mat/0312028>

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

380 Mirchandani, P.B. (1976). Shortest distance and reliability of probabilistic networks. *Comput. Oper.*
 381 *Res.*, **3**, 347–355. Retrieved October 12, 2014, from [http://www.sciencedirect.com/science/](http://www.sciencedirect.com/science/article/pii/0305054876900174)
 382 [article/pii/0305054876900174](http://www.sciencedirect.com/science/article/pii/0305054876900174)

383 Newman, M.E.J. (2004). Fast algorithm for detecting community structure in networks. *Phys. Rev. E*,
 384 **69**, 066133. Retrieved February 24, 2015,

385 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2011). Missing
 386 and forbidden links in mutualistic networks. *Proc. R. Soc. B*, **278**, 725–732. Retrieved October 7, 2014,

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

389 Poisot, T. ~~(2012). L’ABC de la spécialisation: apparition, biodiversité, conservation. *Prisme À Idées*, **4**,~~
 390 ~~49–52.~~

391 ~~Poisot, T.~~ & Gravel, D. (2014). When is an ecological network complex? Connectance drives degree
 392 distribution and emerging network properties. *PeerJ*, **2**, e251. Retrieved September 13, 2014,

393 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species
 394 interaction networks. *Ecol Lett*, **15**, 1353–1361. Retrieved November 3, 2014,

395 Poisot, T., Lounnas, M. & Hochberg, M.E. (~~2013~~[2013a](#)). The structure of natural microbial enemy-victim
 396 networks. *Ecol. Process.*, **2**, 13. Retrieved March 27, 2015,

397 [Poisot, T., Mouquet, N. & Gravel, D. \(2013b\). Trophic complementarity drives the biodiversityecosystem](#)
 398 [functioning relationship in food webs. *Ecol. Lett.*, n/a–n/a. Retrieved May 22, 2013,](#)

399 Poisot, T., Stouffer, D.B. & Gravel, D. (2014). Beyond species: why ecological interaction networks vary
 400 through space and time. *Oikos*, n/a–n/a. Retrieved November 5, 2014,

401 Poullain, V., Gandon, S., Brockhurst, M.A., Buckling, A. & Hochberg, M.E. (2008). The Evolution
 402 of Specificity in Evolving and Coevolving Antagonistic Interactions Between a Bacteria and Its Phage.
 403 *Evolution*, **62**, 1–11. Retrieved October 10, 2014,

404 Stouffer, D.B. & Bascompte, J. (2011). Compartmentalization increases food-web persistence. *Proc.*
 405 *Natl. Acad. Sci. U.S.A.*, **108**, 3648–3652.

406 Stouffer, D.B., Camacho, J., Jiang, W. & Amaral, L.A.N. (2007). Evidence for the existence of a robust
 407 pattern of prey selection in food webs. *Proc. R. Soc. B Biol. Sci.*, **274**, 1931–40.

408 Thébault, E. & Loreau, M. (2003). Food-web constraints on biodiversityecosystem functioning relation-
 409 ships. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 14949–14954.

410 Trøjelsgaard, K., Jordano, P., Carstensen, D.W. & Olesen, J.M. (2015). Geographical variation in mu-
 411 tualistic networks: similarity, turnover and partner fidelity. *Proc. R. Soc. B*, **282**, 20142925. Retrieved
 412 February 1, 2015,

413 Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). Habitat modification alters the structure of
 414 tropical hostparasitoid food webs. *Nature*, **445**, 202–205. Retrieved January 28, 2013, from [http:](http://www.nature.com/nature/journal/v445/n7124/abs/nature05429.html)
 415 [//www.nature.com/nature/journal/v445/n7124/abs/nature05429.html](http://www.nature.com/nature/journal/v445/n7124/abs/nature05429.html)

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