The macroevolution of bipartite networks

Timothée Poisot [1,2,3]   Daniel B. Stouffer [1]

1: Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand;   **2**: Université de Montréal, Département de Sciences Biologiques;   **3**: Québec Centre for Biodiversity Sciences

Abstract: TODO

Correspondence to Timothée Poisot – `timothee.poisot@umontreal.ca`

The extant structure of biodiversity is the outcome of macro-evolutionary processes, and the modelling of these processes has stimulated a large variety of approaches (Aldous & Popovic 2005; Nee 2006). At their core, most approaches are essentially birth-death processes: they model the rate of speciation and extinction to generate a prediction about both the temporal dynamics of species richness and its predicted current state. Surprisingly, these models tend to consider species as isolated entities; even though they share ancestry, they are not explicitly linked via biotic interactions. This fact is problematic from both an ecological (**???**; Cazelles et al. 2015) and evolutionary (**???, ???, ???, ???**) standpoint since it is widely accepted that interactions serve as an essential scaffold for biodiversity and its emergent properties: they intervene in the coexistence and persistence of species, and drive properties like community persistence or ecosystem function (**???**). After all, predators invariably require prey, hosts require parasites, flowering plants require pollinators, and so on.

Although modern macro-ecological models give an increasingly central role to interactions (Guisan & Thuiller 2005; Thuiller et al. 2013), such models are still unable to predict the structure of complex interacting communities (Jablonski 2008). Nevertheless, there are two key observations upon which solutions to overcome this limitation can be devised. First, extant networks are decidedly non-random with regard to their structure, and their structure is equally non-random with regards to macro-evolutionary processes (**???**). Second, the structure of ecological networks is dynamic over evolutionary timescales (McKane & Drossel 2005). Both these points are strongly suggestive of perpetual and ongoing action of macro-evolutionary processes. It stands to reason then that models of macro-evolution with explicit consideration of species interactions will therefore provide an appropriate theoretical framework to understand how networks evolve. Notably, such a framework enables

the estimation of how much of extant network structure originated through macro-evolution, as opposed to reflecting extant opportunities and constraints (Peralta 2016).

If one assumes that the conservatism of interactions across phylogenies can be explained by the fact that an incipient species inherits its ancestor's interactions upon speciation (???, ???), even a simple model with relatively few parameters can describe the possible evolutionary rules that shape a community's interaction network. Ideally, the parameters of any model such as this—no matter how simple or complex—ought to be calibrated against real-world evolutionary dynamics, similar to how the fossil and molecular record has been used to study species diversification (Benton 2015). Unfortunately, the dearth of well-resolved, long-term time series of species interactions rules out such a comparison to temporal network dynamics. Therefore, we instead addressed the question of network macro-evolution here by using extant ecological networks to calibrate the end points of an interaction-centric birth-death simulation model under the assumption that the best-fitting models will provide insight into the network's likely evolutionary history. Among the variety of ecological networks types, bipartite ones are the most appropriate family to test this model: they have well partitioned interactions between guilds with no complex feedback loops, are present in a variety of systems and types of biological interactions, and there is a wealth of well-studied data available (???). Moreover, taxa from both guilds of a bipartite ecological network are usually tightly evolutionarily linked and require interactions to persist, making them ideal to elucidate evolutionary rules of community structure.

1 ───

Methods

1.1. Model description   We posit that four simple rules govern the evolution of networks. First, every network originally consists of just two species sharing a single interaction; for example, a plant and its herbivore. Second, a speciation event happens at the top level (*e.g.* the herbivore) with probability p, or at the bottom level with probability $1 - p$. Third, the incipient species starts with all interactions of its ancestor. Fourth, some of these interactions are lost with probability $\epsilon(\lambda, k, c)$, which allows interactions—that are gained through speciation—to be lost either at a fixed rate λ or as a function of the incipient species' degree k. The c parameter modulates this relationship further by influencing whether high degree of an ancestor increases, or decreases, the probability of the incipient species losing interactions. We have used the following formulation for ϵ:

$$\text{logit}(\epsilon) = \lambda + c \times (k - 1.0) \tag{1}$$

In this formulation, $\text{logit}^{-1}(\lambda)$ is the probability of the descendant of a species with a single interaction failing to establish, *i.e.* the basal rate of interaction loss, k is the number of interactions of the incipient species, and c is a parameter regulating
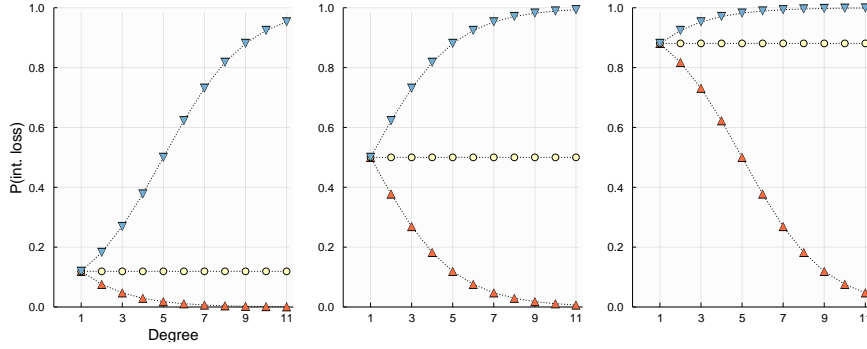
Figure 1 Probability that *each* interaction of the ancestor is lost by the incipient species during speciation. $\text{logit}^{-1}(\lambda) = 0.15, 0.5, 0.85$, and c varied between -1/2 (red), 0 (yellow), and 1/2 (blue).

whether species with more interactions tend to gain or lose interactions over time. Negative values of c imply that *rich get richer*, *i.e.* species with more interactions tend to conserve them more after speciation. The special case of $c = 0$ corresponds to no relationship between the degree of a species and its probability of losing or retaining an interaction over speciation. The resulting probability of interaction loss, and its consequences on degree, is shown in fig. 1. The values of ϵ belong to]0; 1[. Note that, because species are duplicated upon a speciation event, the network still grows over time. If an incipient species should lose all of its interactions, then it fails to establish.

These four rules translate directly into steps for the model: pick a level at random, select a species to duplicate, assess the survival of interactions of the incipient, and add the incipient to the network. These are performed a fixed number of time – we impose an upper limit to the richness at each level, and when this limit is reached, the incipient species replaces one of the resident species at random. An equilibrium for the measures of network structure (see next section) is reached within 1000 timesteps. For all situations, we recorded the network after 5000 iterations.

1.2. Network measures Because there are almost infinitely many measures of network structure used in ecology (Delmas et al. 2017; Antoniazzi et al. 2018), we have focused our analysis on measures that have a number of desirable properties. First; they return a single value, and therefore function as a "summary" of an aspect of network structure. Second; they have been well studied in ecology as having actual relevance on the processes and mechanisms in which networks take part. Third; they are, or can be made, relatively robust to changes in species richness. Finally; they are, or can be made to, return values in [0; 1], to avoid scaling issues when comparing networks.

(1.2.1) *Connectance* Connectance, defined as the ratio of realized interactions on the total number of potential interactions, is one of the most common descriptor of network structure. In a bipartite network with T species at the top, and B at the bottom, having a total of L interactions, it is defined as $Co = L/(T \times B)$. Connectance has a lower bound, as the network cannot have fewer interactions that the number of species in its richer level – the minimal connectance is therefore $c_m = \max(T, B)$. This makes the connectance of networks of different sizes difficult to compare,

especially since bipartite networks tends to have a low connectance. For this reason, we used a corrected version of connectance, defined as

$$Co^\star = \frac{L - c_m}{T \times B - c_m}.$$

(2)

This takes values between 0 (the network has the minimal number of interactions) and 1 (all species are connected), but is robust to variations in species richness.

(1.2.2) Nestedness We measured nestedness, using the η measure of Bastolla et al. (2009), which returns a global nestedness score based on the fact that interactions of relatively specialized species should be a subset of the interactions of more generalized ones. This measure is robust to changes in species richness, and returns values between 0 (not nested) to 1 (perfectly nested) – this measure functions as a summary of the joint degree distribution of the network.

(1.2.3) Modularity We measured modularity using the BRIM approach (Barber 2007) (preliminary analyses revealed no qualitative impact of using other methods to optimize modularity). BRIM returns values close to 1 when there are modules in the network, and values closer to 0 otherwise. The value of modularity for each network is the maximal modularity out of 50 replicates.

(1.2.4) Motifs Finally, we enumerated six four-species bipartite motifs (Baker et al. 2014). Bipartite motifs are possible conformations of four species spread across two levels, such as for example three consumers sharing one resource, or two consumers both exploiting resources, *etc.*. The five motifs we used are illustrated in {(???)}. Because the number of motifs obviously varies with species richness, we corrected it in the following way. For each network (and each motif), we generated a matrix of probability using the Type II null model of Bascompte et al. (2003), wherein the probability that two species interact is directly proportional to the product of their relative degrees. We then used the probabilistic formulas laid out in Poisot et al. (2016) to measure the expected number of motifs in the probabilistic network (μ_m) and the standard deviation (σ_m). We the measured the actual number of motifs in the network m_c, and measured the z-score with regard to the probabilistic network, $z_m = (m_c - \mu_m)/\sigma_m$. Because this gives real values, we report the motif count as $m = S(z_m)$, where $S(x) = 1/(1 + \exp(-x))$. This gives a motif count in $[0; 1]$, where values above 0.5 mean that the motif is more common than expected by chance, and values below 0.5 means that it is less so.

1.3. Simulations To explore the behaviour of the model, we conducted a series of simulations using $p = 0.5$, varying λ from 10^{-4} to 10^{-1} (every order of magnitude), and c from 0.05 to 2.5 (by increments of 0.05). For every combination of parameters, we performed 500 simulations, using 25 species maximum on every level. The network was returned after 8000 timesteps. The network measures were applied on the endpoint of the simulation.

1.4. Data selection We used empirical data from the *WebOfLife* dataset (`http://web-of-life.es/`). Every network was "cleaned" in the following way. First, species with no interactions (if any) were removed. This yields networks in which all species have at least one interaction. Second, interactions strengths (if present) were removed since our model only requires information about the presence or absence of interactions.

1.5. Parameter selection We used ABC (Approximate Bayesian Computation) to select the parameter values that yielded realistic networks by assessing how closely each replicate of the second numerical experiment resembles empirical communities (Beaumont 2010). For each empirical network, its observed set of summary statistics (all network measures) was compared to each output of the stochastic model. The Euclidean distance between the two arrays was recorded as the score ρ of the parameter set. Because each empirical network is in practice a different optimization problem submitted to the ABC routine, and because ABC requires to set the rejection threshold ρ_{max} on a per-problem basis, setting a global value was not meaningful (Sunnåker et al. 2013). To circumvent this issue, we adopted a two-steps approach. First, we generated 10^4 combinations of priors, and kept in the posterior the top 500 (*i.e.* with the lowest distance). From this, we used maximum likelihood estimation to fit a normal distribution for each parameter. This generated a somewhat more informative prior. We used it to simulate another 10^4 parameters combinations, of which we selected the best 100 to form the posterior. Because the fit of every simulation was generally good, and because there was a low dispersal in the parameters retained for every network, we summarize each empirical network by the average of every posterior value.

1.6. Implementation The model (and all the data analysis code) was written in `Julia` (Bezanson et al. 2017) 0.6.2, using the package `EcologicalNetwork.jl` 1.1.0 – **https://doi.org/10.5281/zenodo.595661**. The code, and copies of the raw data and all intermediate computational artifacts used for this article, is available at OSF.IO

2 ——————————————————————————————

Results

2.1. Model behavior By varying the parameters λ (base probability of losing and interactions) and c (effect of the degree on ϵ, eq. 1), we are able to generate a range of scenarios using the model. In fig. 2, we report the values of all network measures. As expected, low values of λ and values of c below unity result in the largest connectance. Most network measures respond in a non-linear way to either parameters, because the ability to speciate is ultimately having an impact on the number of species at each level, which in turn constrains measures such as motif occurrences, nestedness, etc. It should be further noted that the "network space" is not entirely covered, *i.e.* no matter what the parameters are, some combinations of
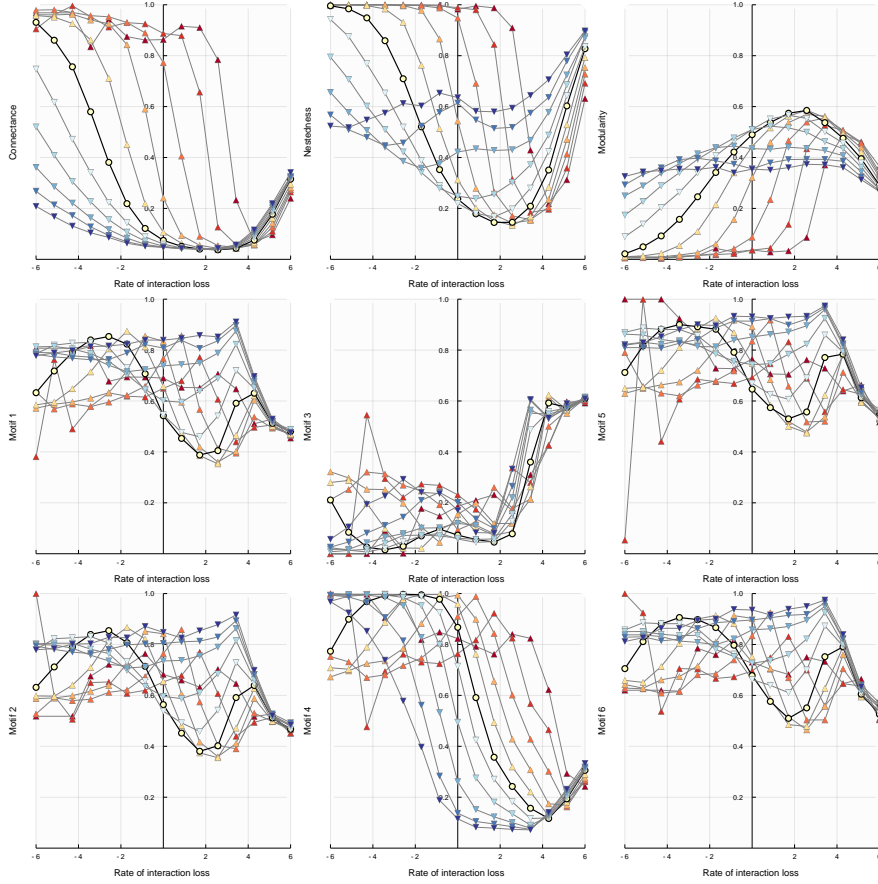
Figure 2 Exploration of the model's behavior with a range of values for λ (x-axis) and c (blue: tendency to decrease degree relative to the ancestor; red: tendency to increase degree relative to the ancestor). Each point represents the average value for up to 10^4 simulations for the parameters combination. The bold line corresponds to $c = 0$, *i.e.* no effect of the degree of the ancestor. Colors are the same as in fig. 1.

structure are never observed. This is most likely to indicate the impossibility of these combinations to exist due to the presence of covariance between all network structure measures.

2.2. Evolutionary parameters by network type We first observed that the posterior distribution of the parameters differs across interaction types (fig. 3). There is no obvious distribution of p by network type, which is expected since the value of p primarily ties into the ratio of top-level to total species, and this not affected by the type of interaction (nor is it involved in measures of network structure). We will focus on the two parameters governing the rate of interaction loss, λ and c. All networks seem to exist on a diagonal going from low λ but high c (interactions are hard to lose, but it is hard to increase in degree; mostly host-parasite networks), to low λ but high c (interactions are hard to maintain, but it is easier to increase in degree; mostly pollination networks).

2.3. Predictive ability In fig. 4, we report the distance between the empirical network and the top 100 simulations, as $1 - \sum |n_0 - x|/|\mathbf{x}|$, where n_0 is the value of every measure on the empirical network, and x is the vector containing the average values in the networks retained as part of the posterior distribution. For the three datasets with the most networks (parasitism, seed dispersal, and pollination), all
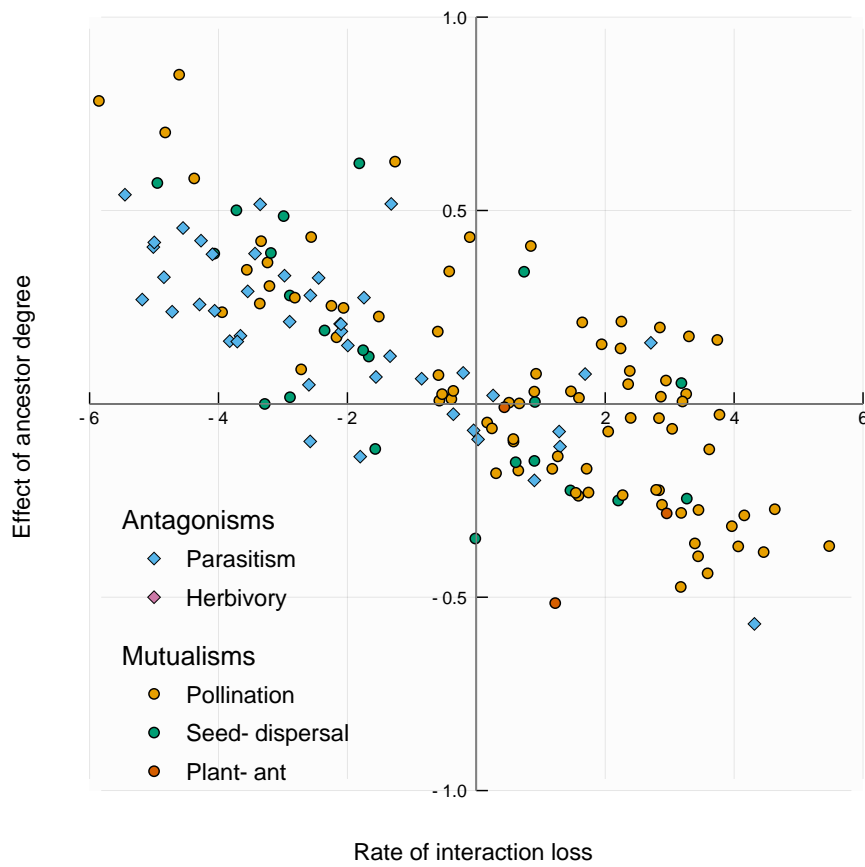
Figure 3 Scatterplots of the weighted averages for λ and c by network type. Each point represents a network. All networks distribute alongside a continuum of values. In the upper-left corner, interactions are controlled by the degree of the ancestor species; in the lower-right corner, there is a trend to increase the degree, but the rate of interaction loss is higer.
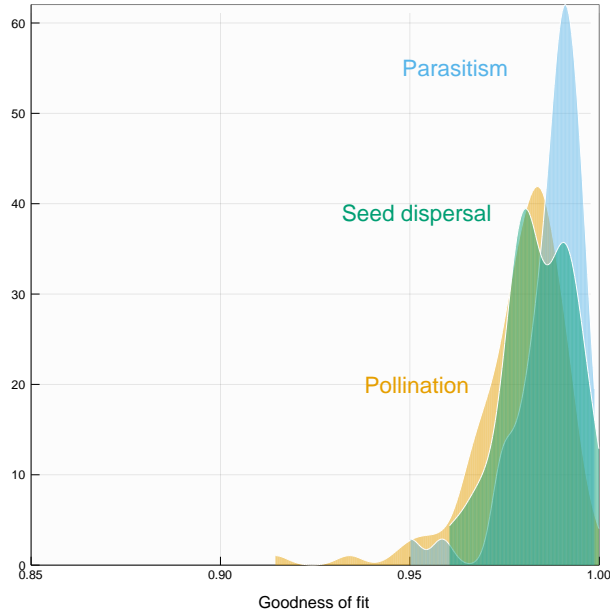
Figure 4 Goodness of fit (measured as the Euclidean distance between the measures on the empirical network and the best 100 simulations) for three types of interactions. All values are above 90%, which indicates that our evolutionary model is able to capture aspect of the structure of extant networks.
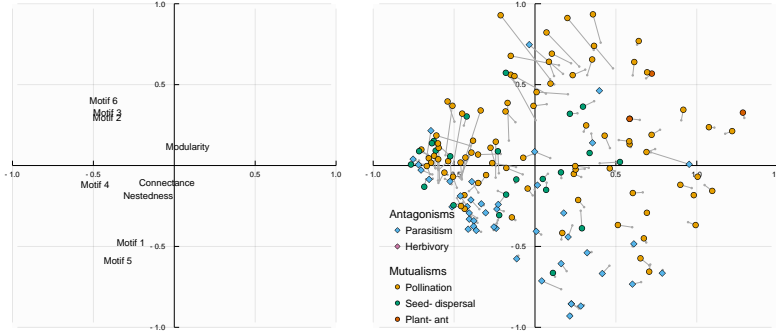

Figure 5 Left panel: ordination (Principal Component Analysis) of the network variables based on the empirical networks. Right panel: projection of empirical networks, as well as of the centroid of their simulated counterparts (grey dots) in the same space.

of these values are above 0.9, which indicates an extremely good fit between the predictions and the empirical network.

We expand this result by looking at the match between the empirical and simulated networks in the multi-variate space of network measures (fig. 5). First, a lot of inter-network variation is driven by motif composition (which is not surprising, as connectance, nestedness, and modularity, are known to have strong covariance). Second, empirical networks are close to the centroid of their best simulated matches, indicating a high fit of the model.

3

Discussion

Finally, all systems show a strong bias towards moderately high values of c; this indicates that the effective probability of a species retaining its ancestor's interactions decreases with its ancestor's degree. That is, the generalism of species over time has

an emergent upper bound, a fact that results in the very spectrum of high-degree and low-degree species that is ubiquitous empirically (**???**).

Emergence of generalists / specialists /// parasite data

Ponisio & M'Gonigle (2017) / Poisot & Stouffer (2017)

Because some structural properties of networks emerge in the absence of rules on evolutionary traits, Valverde et al. (2018) have suggested that they represent evolutionary spandrels. Maynard et al. (2018) show that these spandrels tie into ecological rules on networks, specifically the feasibility of assemblies history. Our results show that, while it is true that a simple rule-based model can generate almost any network configuration (fig. 2), extent bipartite empirical networks are compatible with only a narrow, non-random subset of all possible parameters (fig. 3). Therefore, conserved networks structures need not be evolutionary spandrels, in that they can represent a state towards which most ecological communities eventually tend.

Our results show that the structure of extant bipartite networks can be adequately reproduced by a speciation/extinction model that accounts for biotic interactions. The selection on parameters related to interaction diversification and persistence was stronger than on the parameter related to the rate of speciation, suggesting that the importance of biotic interactions in macro-evolution may have been understated. Our results also highlight that, while the evolutionary persistence of interactions is undeniably important in the macro-evolution of community structure, different type of ecological interactions respond in different ways. This offers a very stimulating possibility – namely, that because the mode of coevolution *within* the interaction between two species differ as a function of their ecological interactions (Thompson 1994), this can cascade up to the macro-evolutionary scale in the form of a signal of long-term interaction persistence.

References

Aldous & Popovic. (2005). A critical branching process model for biodiversity. *Adv Appl Probab.* 37:1094–115.

Antoniazzi et al. (2018). A Useful Guide of Main Indices and Software Used for Ecological Networks Studies. *Ecological Networks in the Tropics.* Springer, Cham; pp. 185–96.

Baker et al. (2014). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography.* 38:130–9.

Barber. (2007). Modularity and community detection in bipartite networks. *Phys Rev E.* 76:066102.

Bascompte et al. (2003). The nested assembly of plant-animal mutualistic networks. *PNAS.* 100:9383–7.

Bastolla et al. (2009). The architecture of mutualistic networks minimizes compe-

tition and increases biodiversity. *Nature.* 458:1018–20.

Beaumont. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annu Rev Ecol Evol Syst.* 41:379–406.

Benton. (2015). Exploring macroevolution using modern and fossil data. *Proc Biol Sci.* 282.

Bezanson et al. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* 59:65–98.

Cazelles et al. (2015). A theory for species co-occurrence in interaction networks. *Theor Ecol.* 9:39–48.

Delmas et al. (2017). Analyzing ecological networks of species interactions. *bioRxiv.*:112540.

Guisan & Thuiller. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol Lett.* 8:993–1009.

Jablonski. (2008). Biotic Interactions and Macroevolution: Extensions and Mismatches Across Scales and Levels. *Evolution.* 62:715–39.

Maynard et al. (2018). Network spandrels reflect ecological assembly. *Ecol Lett.*:n/a–a.

McKane & Drossel. (2005). Models of food web evolution. *Ecological networks: Linking structure to dynamics in food webs.* p. 223.

Nee. (2006). Birth-death models in macroevolution. *Annu Rev Ecol Evol Syst.* 37:1–17.

Peralta. (2016). Merging evolutionary history into species interaction networks. *Funct Ecol.*

Poisot & Stouffer. (2017). Interactions retain the co-phylogenetic matching that communities lost. *Oikos.*:n/a–a.

Poisot et al. (2016). The structure of probabilistic networks. Vamosi, ed. *Methods Ecol Evol.* 7:303–12.

Ponisio & M'Gonigle. (2017). Coevolution leaves a weak signal on ecological networks. *Ecosphere.* 8:n/a–a.

Sunnåker et al. (2013). Approximate Bayesian Computation. *PLoS Comput Biol.* 9.

Thompson. (1994). The Coevolutionary Process. Chicago: University of Chicago Press;

Thuiller et al. (2013). A road map for integrating eco-evolutionary processes into biodiversity models. *Ecol Lett.* 16:94–105.

Valverde et al. (2018). The architecture of mutualistic networks as an evolutionary spandrel. *Nat Ecol Evol.* 2:94.