

# How ecological networks evolve

Timothée Poisot <sup>1,2,3,\*</sup>, Daniel B. Stouffer <sup>1</sup>

(1) Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

(2) Université de Montréal, Département de Sciences Biologiques, 90 Avenue Vincent d'Indy, Montréal, QC, CAN, H2V3S9

(3) Québec Centre for Biodiversity Sciences, 1205 Dr. Penfield Avenue, Montréal, QC, CAN, H3A1B1

\* e-mail: [tim@poisotlab.io](mailto:tim@poisotlab.io)

**Abstract:** Here be science yo.

**Keywords:** ecological networks      Approximate Bayesian Computation      bipartite networks      macroevolution

**Date:** Work in progress.

The extant structure and distribution of biodiversity is the outcome of macro-evolutionary processes, and the modelling of these processes stimulated a large variety of approaches<sup>1,2</sup>. At their core, these approaches are birth-death processes, in that 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. Yet these models often consider species as isolated entities, in that even though they share ancestry, they are not involved in inter-specific interactions. This is problematic, as from an ecological<sup>3</sup> and evolutionary<sup>4,5</sup> standpoint, interactions serve as a *scaffold* for community persistence: plants require pollinators, predators require preys, etc.

Although modern macro-ecological models give an increasingly central role to interactions<sup>6</sup>, macro-evolution models are still unable to predict the structure of complex interactive communities<sup>7</sup>. Two key observations motivate the overcoming of this limitation. First, extant networks are non-random in their structure, and display a strong macro-evolutionary signal<sup>5</sup>. Second, the structure of ecological networks is dynamic over evolutionary timescales<sup>8</sup>. This is strongly suggestive of ongoing action of macro-evolutionary processes. Models of macro-evolution with explicit consideration of species interactions will provide a theoretical understanding of how networks evolve, and how much of their extant structure originated through evolutionary processes.

Yet the relative lack of well-resolved long-term time series of species interactions rule out the comparison of temporal dynamics<sup>9</sup> the (comparatively) extensive fossil and molecular record of species diversification allows. We addressed the question of network macro-evolution by confronting extant (bipartite) networks with the outcome of a birth-death simulation models, under the assumption that the best-fitting models will be a representation of the network's evolutionary history. We posit that the evolution of networks follows a series of simple rules. First, every network starts as two species with one interaction. Second, a speciation event happens at the top level with probability  $p$ , and at the bottom level with probability  $1 - p$ . Any incipient species start with all interaction of the ancestor. Third, interactions of the incipient species are lost with probability  $\epsilon(\lambda, c, k)$  (see *Methods Summary*); this allows interactions to be lost either at a fixed rate  $\lambda$ , or as a function of the

40 incipient specie's degree  $k$ . These three steps are repeated  $10^4$  times, for  $10^5$  random combinations  
 41 of  $< p, \lambda, c >$ . Whenever either level has more than  $10^2$  species, some are deleted at random.

42 The endpoint of these simulations are compared to 271 bipartite ecological networks (seed disper-  
 43 sal; herbivory; parasitism; bacteriophagy; pollination – see *Methods summary*) using Approximate  
 44 Bayesian Simulation (ABC). ABC<sup>10,11</sup> gives estimates of posterior distribution of parameters by  
 45 comparing a measure of distance between empirical observations and a model, when no analyti-  
 46 cal expression of likelihood can be derived. We define the distance between a simulated ( $i$ ) and  
 47 empirical ( $j$ ) network as  $d(\mathbf{v}_i, \mathbf{v}_j)$ , where  $\mathbf{v}$

## 48 **1 Results and discussions**

### 49 **1.1 Model output**

50 1. Model output

### 51 **1.2 Posterior distribution of parameters**

52 2. Parameters distributions for different types of networks

### 53 **1.3 Accuracy of predictions**

54 3. Z-scores

55 4. Use the method to *generate* realistic networks by sampling the relevant posteriors

## 2 Methods

### 2.1 Data selection

We used empirical data from mutualistic interactions (XX networks), plant-herbivore interactions (XX networks), phage-bacteria networks (XX interactions), plant-dispersers interactions (XX networks), and host-parasite interactions (XX networks). Mutualistic and dispersers interactions were taken from the *WebOfLife* database. Phage-bacteria data are from ????. Host-parasite data are from ????. Plant-herbivore data are from ????. Each network was cleaned in the following way. First, species with no interactions (if any) were removed. Second, interactions strengths (if present) were removed. This yields adjacency matrices in which all species have at least one interaction.

### 2.2 Stochastic model

We model the diversification of bipartite networks. Bipartite networks are a useful way of representing several types of ecological interactions in which a group of species (*blue*) interacts with a second group of species (*red*; examples include pollination, frugivory, seed dispersal, parasitism, ...). The starting point of every simulation is the simplest possible network: one blue and one red species, with a single interaction between them.

Our model is structurally close to speciation/extinction models – at each time step, there is a probability  $p$  that one of the blue species (taken at random) will undergo a speciation event (and conversely, a probability  $1 - p$  that a red species will speciate). Upon speciation, the incipient species starts with *all* interactions of its ancestor; each of these interactions is lost with probability  $\epsilon$ ,

$$\epsilon = \epsilon_0 \times \text{whatever} \tag{1}$$

75 , where .

- 76 • parameters ranges (from preliminary simulations)

## 77 **2.3 Simulations**

78 We conducted the following two numerical experiments. First, we conducted a systematic explo-  
79 ration of the model’s behavior using evenly spaced parameter values. Each combination of param-  
80 eters was simulated 1000 times. Second, we sampled the parameter space uniformly, by drawing  
81  $10^5$  parameters sets at random from within the aforementioned bounds. These outputs were used in  
82 the parameter selection experiment described below.

## 83 **2.4 Network measures**

84 We measure th

- 85 • connectance
- 86 • nestedness
- 87 • motifs (explain the correction)

88 The raw number of motifs was corrected to account for the number of species in each layer of the  
89 bipartite network. The maximum number of a motif with (e.g.) 2 species at the top and 2 species  
90 at the bottom is the product of the number of combinations of 2 species in the top layer, and of 2  
91 species in the bottom layer (evaluated by their binomial coefficients). This gives a total number of  
92 sets of species that *could* be involved in a 2x2 motif; the raw number of this motif is divided by  
93 this maximum number. This yields values in the 0-1 range, representing the proportion of sets of  
94 species that *do* form a given motif out of the sets of species that *could*.

- modularity

## 2.5 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. For each empirical network, its array of ranged summary statistics was compared to each output of the stochastic model. The inverse of the Euclidean distance between the two arrays was recorded as the score of the parameter set. As every empirical network had different absolute values of scores, fixing a common threshold was not feasible. We selected the posterior distribution as the 500 parameters sets that gave the best scores (i.e. above the 95th percentile).

## 3 References

1. Maruvka, Y. E., Shnerb, N. M., Kessler, D. A. & Ricklefs, R. E. Model for macroevolutionary dynamics. *Proceedings of the National Academy of Sciences* **110**, E2460–E2469 (2013).
2. Nee, S. Birth-Death Models in Macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **37**, 1–17 (2006).
3. Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. Trophic theory of island biogeography. *Ecol. Lett.* **14**, 1010–1016 (2011).
4. Eklof, A., Helmus, M. R., Moore, M., Allesina, S. & Eklöf, A. Relevance of evolutionary history for food web structure. *Proc. R. Soc. B Biol. Sci.* **279**, 1588–1596 (2011).
5. Stouffer, D. B., Sales-Pardo, M., Sirer, M. I. & Bascompte, J. Evolutionary Conservation of Species' Roles in Food Webs. *Science* **335**, 1489–1492 (2012).

- 115 6. Thuiller, W. *et al.* A road map for integrating eco-evolutionary processes into biodiversity mod-  
116 els. *Ecol. Lett.* **16**, 94–105 (2013).
- 117 7. Jablonski, D. BIOTIC INTERACTIONS AND MACROEVOLUTION: EXTENSIONS AND  
118 MISMATCHES ACROSS SCALES AND LEVELS. *Evolution* **62**, 715–739 (2008).
- 119 8. Roopnarine, P. D. & Angielczyk, K. D. The evolutionary palaeoecology of species and the  
120 tragedy of the commons. *Biology Letters* **8**, 147–50 (2012).
- 121 9. Donoghue, P. C. & Benton, M. J. Rocks and clocks: calibrating the Tree of Life using fossils and  
122 molecules. *Trends in Ecology & Evolution* **22**, 424–431 (2007).
- 123 10. Csilléry, K., Blum, M. G., Gaggiotti, O. E. & François, O. Approximate Bayesian computation  
124 (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418 (2010).
- 125 11. Wilkinson, R. D. Approximate Bayesian computation (ABC) gives exact results under the as-  
126 sumption of model error. *Stat. Appl. Genet. Mol. Biol.* **12**, 129–141 (2013).