

# 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.

14 The extant structure and distribution of biodiversity is the outcome of macro-evolutionary pro-  
15 cesses, and the modeling of these processes stimulated a large variety of approaches<sup>1,2</sup>. At their  
16 core, these approaches are birth-death processes, in that they model the rate of speciation and ex-  
17 tinction to generate a prediction about both the temporal dynamics of species richness, and its pre-  
18 dicted current state. Yet these models often consider species as isolated entities, in that even though  
19 they share ancestry, they are not involved in inter-specific interactions. This is problematic, as from  
20 an ecological<sup>3</sup> and evolutionary<sup>4,5</sup> standpoint, interactions serve as a *scaffold* for community persis-  
21 tence: plant require pollinators, predators require preys, etc.

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

30 Yet the relative lack of well-resolved long-term time series of species interactions rule out the com-  
31 parison of temporal dynamics<sup>9</sup> the (comparatively) extensive fossil and molecular record of species  
32 diversification allows. We addressed the question of network macro-evolution by confronting ex-  
33 tent (bipartite) networks with the outcome of a birth-death simulation models, under the assumption  
34 that the best-fitting models will be a representation of the network's evolutionary history. We posit  
35 that the evolution of networks follows a series of simple rules. First, every network starts as two  
36 species with one interaction. Second, a speciation event happens at the top level with probability  
37  $p$ , and at the bottom level with probability  $1 - p$ . Any incipient species start with all interaction  
38 of the ancestor. Third, interactions of the incipient species are lost with probability  $\epsilon(\lambda, c, k)$  (see  
39 *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 We compared simulation endpoints to 271 bipartite ecological networks (seed dispersal; herbivory;  
43 parasitism; bacteriophagy; pollination – see *Methods summary*) using Approximate Bayesian Com-  
44 putation (ABC). ABC<sup>10,11</sup> gives estimates of posterior distribution of parameters by comparing a  
45 measure of distance between empirical observations and a model, when no analytical expression  
46 of likelihood can be derived. We define the distance between a simulated ( $i$ ) and empirical ( $j$ )  
47 network as  $d(\mathbf{v}_i, \mathbf{v}_j)$ , where  $\mathbf{v}$  is an array of network structural properties, including connectance,  
48 modularity<sup>12</sup>, nestedness<sup>13</sup>, and the distribution of different network motifs<sup>14</sup> (see *Methods sum-*  
49 *mary*). Measures were ranged in  $[0; 1]$ . The posterior distribution of best-fitting parameters, for  
50 each network, is the parameters of the closest 500 simulated models.

51 Posterior distribution of the parameters differ across interaction types (Figure 1). The probability of  
52 speciation at either level ( $p$ ) is the least strongly selected, which suggests that mechanisms pertaining  
53 to the evolution of *interactions* have a stronger impact on extent network structure. There are two  
54 situations for the distribution of  $\lambda$ : herbivory and pollination networks have higher values of this  
55 parameter, meaning that herbivores/pollinators tend to retain the interactions of their ancestors [ref].  
56 All other types of networks were best described by low values of  $\lambda$ ; their interactions appear to  
57 be more labile throughout coevolution. Finally, all systems show a strong bias towards moderately  
58 high values of  $c$ ; this indicates that the effective probability of retaining one's ancestor's interactions  
59 decreases with ancestor's degree. There is an upper bound to the generalism of species over time,  
60 which results in a spectrum of high-degree and low-degree species in networks, as described in  
61 many natural systems<sup>15,16</sup>.

62 The joint distribution of the  $\lambda$  and  $c$  parameters (Figure 2) reveals two possible “states” for networks  
63 to occupy; either  $c$  is close to 0, and  $\lambda$  is large, or  $c$  is close to 1, and  $\lambda$  is low. There exists a con-  
64 tinuum across these two endpoints, alongside which different types of networks fall. Herbivory and  
65 pollination tend to have both low values of  $c$ , and low to high values of  $\lambda$ ; parasitism networks have

low values of  $\lambda$  and low-to-high values of  $c$ . Other types of networks (seed dispersal, bacteriophagy) do not show a strong signal as to their position alongside this gradient.

For each network, we next calculated the average distance to all its best matching simulation outputs, and used the z-score of this value to determine which type of networks was best predicted using our model (Figure 3). The best predicted networks were herbivory and pollination; this suggest that these networks have a strong evolutionary signal. By contrast, other types of networks can be less accurately predicted because they are more sensitive to random chance or ecological mechanisms.

## 1 Methods summary

### 1.1 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 come from the *WebOfLife* database. Phage-bacteria (which are functionally equivalent to host-parasitoid) data are from 17. Host-parasite data are from 18. Plant-herbivore data are from 19. 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.

### 1.2 Simulations

We conducted the following two numerical experiments. First, we conducted a systematic exploration of the model's behavior using evenly spaced parameter values. Each combination of parameters was simulated 1000 times. This allowed us to ensure that the model could return networks

with all possible configurations, and that the output covered a range of network structure larger than what was observed in nature. Second, we sampled the parameter space uniformly, by drawing  $10^5$  parameters sets at random from within the aforementioned bounds. These outputs were used in the parameter selection experiment described below.

### 1.3 Network measures

We measured four key families of bipartite network structure indices. First, connectance, which is the  $\frac{L}{T \times B}$ , with  $L$  the number of interactions, and  $T$  and  $B$  the number of species in the top and bottom groups. Second, nestedness<sup>20</sup>, using the NODF measure, which **todo**. Third, modularity, using LP-BRIM<sup>21,22</sup>, which gives values close to 1 when there are modules in the network, and values closer to 0 otherwise. Finally, we measured the proportion of **xx** bipartites motifs<sup>23</sup>. [\[more\]](#)

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

### 1.4 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 be-

tween 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).

## 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).
6. Thuiller, W. *et al.* A road map for integrating eco-evolutionary processes into biodiversity models. *Ecol. Lett.* **16**, 94–105 (2013).
7. Jablonski, D. BIOTIC INTERACTIONS AND MACROEVOLUTION: EXTENSIONS AND MISMATCHES ACROSS SCALES AND LEVELS. *Evolution* **62**, 715–739 (2008).
8. Roopnarine, P. D. & Angielczyk, K. D. The evolutionary palaeoecology of species and the tragedy of the commons. *Biology Letters* **8**, 147–50 (2012).

- 129 9. Donoghue, P. C. & Benton, M. J. Rocks and clocks: calibrating the Tree of Life using fossils and  
130 molecules. *Trends in Ecology & Evolution* **22**, 424–431 (2007).
- 131 10. Csilléry, K., Blum, M. G., Gaggiotti, O. E. & François, O. Approximate Bayesian computation  
132 (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418 (2010).
- 133 11. Wilkinson, R. D. Approximate Bayesian computation (ABC) gives exact results under the as-  
134 sumption of model error. *Stat. Appl. Genet. Mol. Biol.* **12**, 129–141 (2013).
- 135 12. Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination net-  
136 works. *PNAS* **104**, 19891–19896 (2007).
- 137 13. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases  
138 biodiversity. *Nature* **458**, 1018–1020 (2009).
- 139 14. Stouffer, D. B., Camacho, J., Jiang, W. & Amaral, L. A. N. Evidence for the existence of a  
140 robust pattern of prey selection in food webs. *Proc. R. Soc. B Biol. Sci.* **274**, 1931–40 (2007).
- 141 15. Poisot, T., Bever, J. D., Nemri, A., Thrall, P. H. & Hochberg, M. E. A conceptual framework  
142 for the evolution of ecological specialisation. *Ecol. Lett.* **14**, 841–851 (2011).
- 143 16. Williams, R. J. Biology, Methodology or Chance? The Degree Distributions of Bipartite Eco-  
144 logical Networks. *PLoS One* **6**, e17645 (2011).
- 145 17. Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host-phage  
146 interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E288–297 (2011).
- 147 18. Stanko, M., Miklisová, D., Goüy de Bellocq, J. & Morand, S. Mammal density and patterns of  
148 ectoparasite species richness and abundance. *Oecologia* **131**, 289–295 (2002).
- 149 19. Thebault, E. & Fontaine, C. Does asymmetric specialization differ between mutualistic and  
150 trophic networks? *Oikos* **117**, 555–563 (2008).

- 151 20. Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D. & Ulrich, W. A consistent  
152 metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*  
153 **117**, 1227–1239 (2008).
- 154 21. Liu, X. & Murata, T. Community detection in large-scale bipartite networks. in *Transactions*  
155 *of the Japanese Society for Artificial Intelligence* **5**, 184–192 (Institute of Electrical & Electronics  
156 Engineers (IEEE), 2009).
- 157 22. Barber, M. & Clark, J. Detecting network communities by propagating labels under constraints.  
158 *Phys. Rev. E* **80**, 1–11 (2009).
- 159 23. Baker, N. J., Kaartinen, R., Roslin, T. & Stouffer, D. B. Species' roles in food webs show fidelity  
160 across a highly variable oak forest. *Ecography* (2014). doi:10.1111/ecog.00913



161 **List of Figures**

162	1	Posterior distributions of parameters $p$ , $\lambda$ and $c$ . . . . .	10
163	2	Relationships between parameters $\lambda$ and $c$ . . . . .	11
164	3	Predictive power of the model across different types of networks. . . . .	12

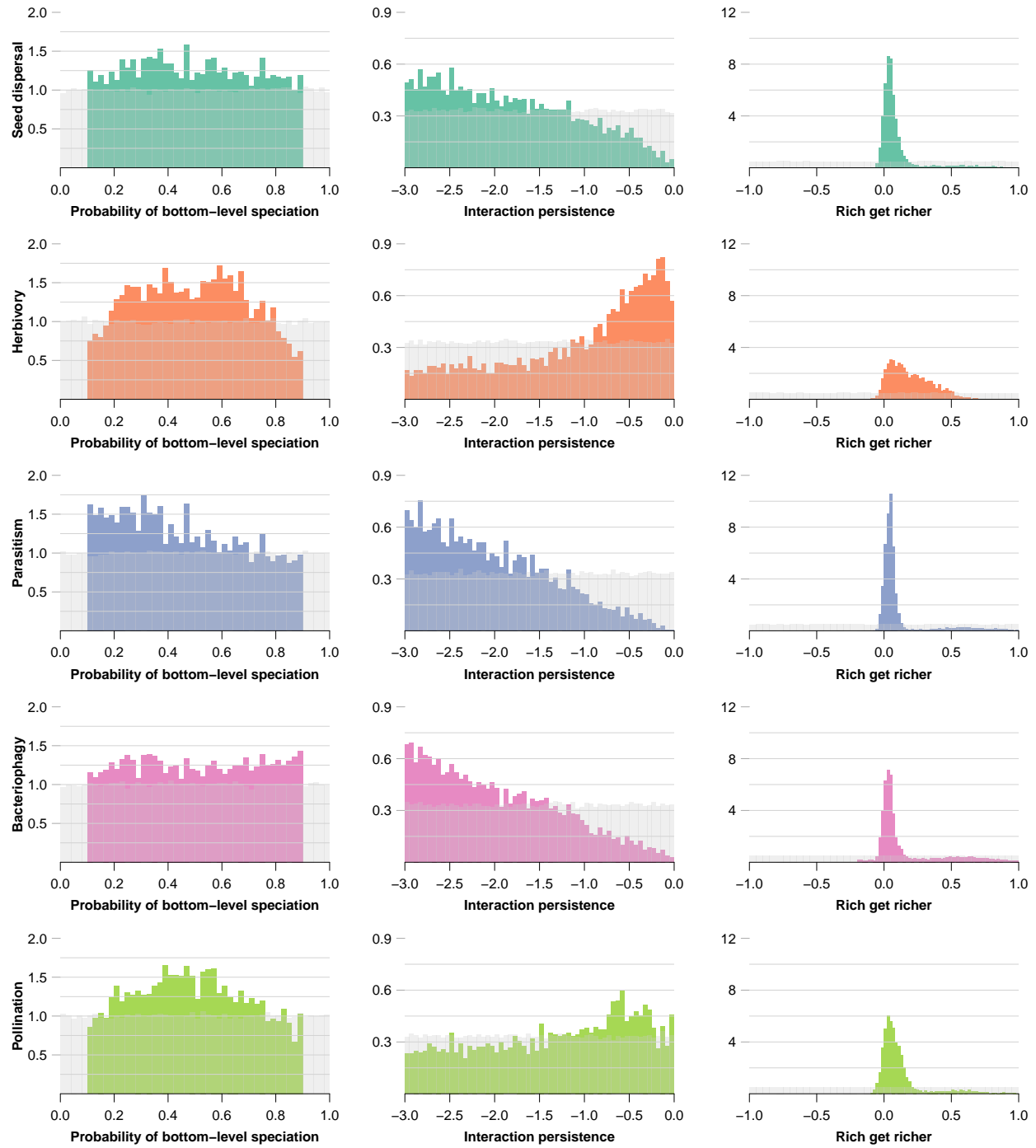


Figure 1: Posterior distributions of parameters  $p$ ,  $\log_{10}\lambda$  and  $\log_{10}c$ . The grey shaded area is a representation of the uniform prior distribution. Although there is no strong selections on the values of  $p$ , networks do differ strongly both from the prior, and from one another, on  $\lambda$  and  $c$ .

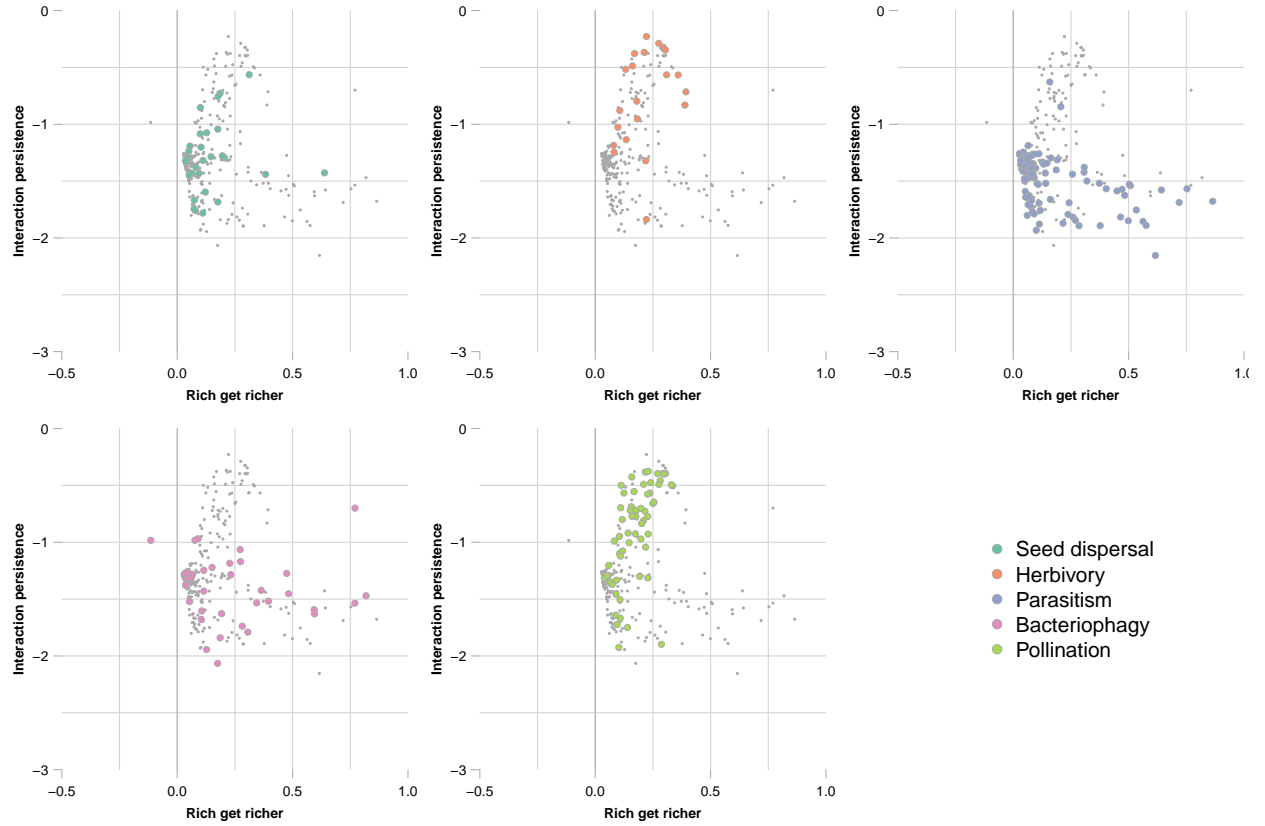


Figure 2: Relationships between parameters  $\lambda$  and  $c$  in the five different types of networks. The grey points indicate all networks in the dataset. Different types of ecological interactions occupy different positions along the  $\lambda$ - $c$  continuum.

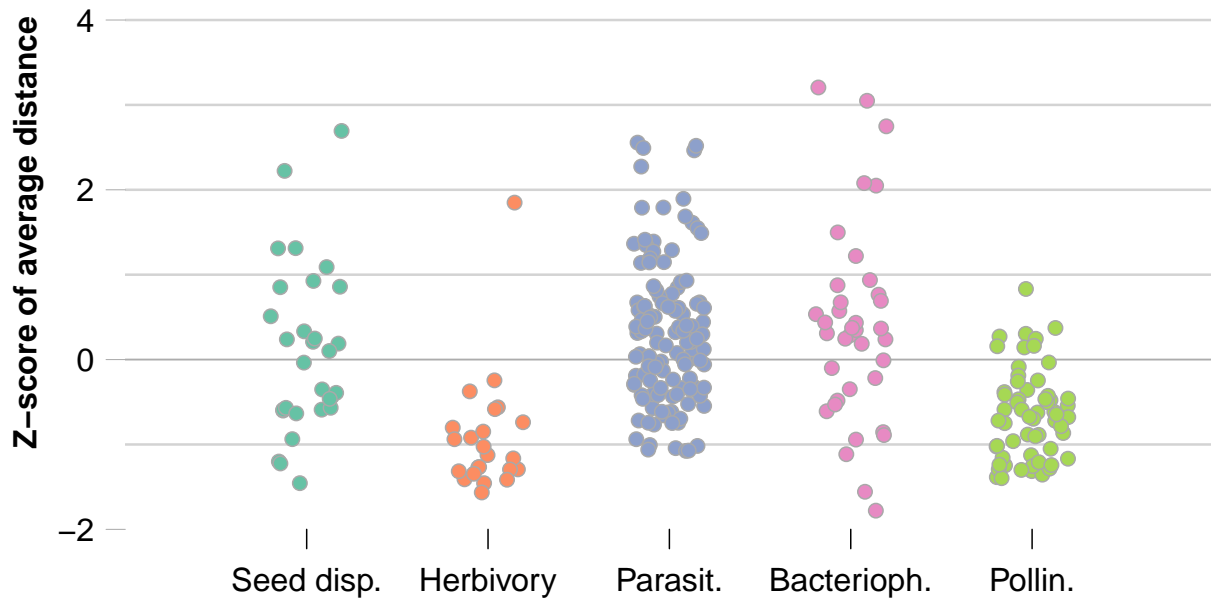


Figure 3: Z-score of *average* distances for the best-matching simulations. Herbivory and pollination networks are better predicted by this model, while z-scores for seed dispersal, parasitism, and bacteriophagy, are centered around 0. The differences in z-scores may come from the fact that evolutionary processes have a stronger footprint on the extant structure of some types of interactions (herbivory, pollination).