

How ecological networks evolve

Timothée Poisot^{1,2,3,*}, Daniel B. Stouffer¹

(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

Abstract: Ecological networks represent the backbone of biodiversity. As species diversify over macro-evolutionary time-scales, the structure of these networks changes; this happens because species are gained, and lost, and therefore add or remove interactions in their communities. However, the dynamics of the structure of ecological networks is still poorly understood. Here we show that several types of ecological interactions share common evolutionary mechanisms, that can be parameterized based on extant interaction data. We found that a model mimicking birth-death processes for species interactions describes the structure of extant networks well. Types of ecological interactions (seed dispersal, herbivory, parasitism, bacteriophagy, and poolination), only differed in the position they occupy in the parameters multi-dimensional space. Notably, we found no clustering of parameters values between antagonistic and mutualistic interactions. Our results provide a common modelling framework for the evolution of ecological networks, that can be parameterized, and is explicit about species interactions. We anticipate this result to contribute to the greater consideration of species interactions in models of macro-evolution and adaptive radiations.

Keywords: ecological networks Approximate Bayesian Computation bipartite networks macroevolution

Date: March 30, 2016

The extant structure and distribution of biodiversity is the outcome of macro- evolutionary processes, and the modeling of these processes stimulated a large variety of approaches^{1,2}. At their core, these approaches are all essentially 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. Surprisingly, these models tend to consider species as isolated entities; even though they share ancestry, they are not involved in inter-specific interactions. This is problematic from both an ecological³ and evolutionary^{4,5} standpoint since it is widely accepted that interactions serve as an essential *scaffold* for community persistence. After all, plants require pollinators, predators require prey, hosts require parasites, etc.

Although modern macro-ecological models give an increasingly central role to interactions⁶, such models are still unable to predict the structure of complex interactive communities⁷. 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 non-random with regards to macro-evolutionary processes⁵. Second, the structure of ecological networks is dynamic over evolutionary timescales⁸. This is strongly suggestive of perpetual and ongoing action of macro-evolutionary processes. Models of macro-evolution with explicit consideration of species interactions will therefore give a theoretical framework to understand how networks evolve. Such a framework enables the estimation of how much of their extant structure originated through macro-evolution, as opposed to reflecting extant opportunities and constraints.

If one assumes that an incipient species inherits its ancestor's interactions upon speciation [ref], even a simple model with relatively few parameters can describe the possible evolutionary rules that shape the 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 comparable to that how the extensive fossil and molecular record has been used to study species diversification [ref]. Unfortunately, the dearth of well-resolved long- term time series of species interactions rules out such a comparison to temporal network dynamics⁹. Therefore, we instead

addressed the question of network macro-evolution here by using extant ecological networks to calibrate 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.

We posit that the evolution of networks follows a series of four simple rules. First, every network begins as two species sharing a single interaction; for example, a plant and its herbivore. Second, a speciation event happens at the top level (the herbivore) with probability p , and at the bottom level with probability $1 - p$. Third, all incipient species starts with all interactions of its ancestor. Fourth, some of these interactions are lost with probability $\epsilon(\lambda, c, k)$, 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 whether high degree in the ancestor increases, or decreases, the probability of losing interactions in the incipient). By comparing model-generated networks with different parameters, we can disentangle the key macro-evolutionary drivers of network structure. For example, if the evolutionary dynamics of interactions are critical for the evolutionary dynamics of communities, we expect that the values of any speciation-related parameters will be less important than those of interaction-related one(s).

Following our macro-evolutionary model, we repeated its four steps 10^4 times to generate a large ensemble of model networks whose structure we could compare to those of the empirical networks. We then compared these model-generated networks to a large collection of 271 bipartite ecological networks whose interactions encode seed dispersal, herbivory, parasitism, bacteriophagy or pollination (see *Methods*) using Approximate Bayesian Computation (ABC). ABC^{10,11} gives estimates of posterior distribution of best-fit parameters by comparing a measure of distance between empirical observations and a model, when no analytical expression of likelihood can be derived. We define the distance between a simulated (i) and empirical (j) network as $d(\mathbf{v}_i, \mathbf{v}_j)$, where \mathbf{v} is an array of network structural properties, including connectance, modularity¹², nestedness¹³, and the distribution of different network motifs¹⁴ (see *Methods*). The posterior distribution of best-fitting parameters, for each network, is the parameters of the closest 500 simulated models (1% of the total).

We first observed that the posterior distribution of the parameters differs across interaction types (Figure 1). The probability of speciation at either level (p) is the least strongly selected, which suggests that mechanisms pertaining to the evolution of interactions have a stronger impact on extant network structure than does the distribution of speciation rates. We also encountered two situations for the distribution of the interaction rate λ : herbivory and pollination networks have higher values of this parameter, implying that herbivores and pollinators tend to retain the interactions of their ancestors more than herbivores did [ref]. All other types of networks were best described by low values of λ ; their interactions consequently appear to be more labile throughout the course of evolution. 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 upper bound, a fact that results in the very spectrum of high-degree and low-degree species that is ubiquitous empirically^{15,16}.

The optimal values of λ and c , however, are not independent since they ultimately affect the same process: the probability of the incipient species losing its ancestor's interactions. A more thorough understanding of the dynamics of interactions throughout evolution can therefore be obtained by examining these parameters' joint distribution. Doing so reveals two possible "states" for networks to occupy based on the results of our model (Figure 2); either c is close to 0 and λ is large or c is close to 1 and λ is low. Notably, different types of networks fall in a specific place within this continuum. Herbivory and pollination tend to have both low values of c and low to high values of λ —implying that foobar, foo, and bar—whereas parasitism networks have low values of λ and low-to-high values of c —implying that varza, barsom, and Zoroaster himself. The two remaining network types, seed dispersal and 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.

Finally, we applied a classification tree to the parameter values describing each empirical network (Figure 4). The tree had a misclassification rate of 35.4%, meaning that knowing only the value of parameters λ and c , the correct type of ecological interaction can be estimated in around 65% of cases. The structure of tree also reveals that antagonistic and mutualistic interactions *do not* form different clusters¹⁷.

In conclusion it works well, please send this paper for review

1 Methods

1.1 Data selection

We used empirical data of plant-pollinator interactions (59 networks), plant-herbivore interactions (23 networks), phage-bacteria networks (38 interactions), plant-dispersers interactions (30 networks), and host-parasite interactions (121 networks). Pollination and seed-dispersal interactions come from the *WebOfLife* database [ref]. Phage-bacteria (which are functionally equivalent to host-parasitoid) data are from Ref 18. Host-parasite data are from Ref 19. Plant-herbivore data are from Ref 20. Every 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 behaviour 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 structures larger than what was observed in nature. Second, we sampled the parameter space uniformly, by drawing 10^5 parameter sets at random from within the aforementioned bounds. These outputs were used in the parameter selection experiment described below.

These three steps are repeated 10^4 times, for 10^5 random combinations of $\langle p, \lambda, c \rangle$. Whenever either level has more than 10^2 species, some are deleted at random within this level. This ensure that the network is at most composed of 200 species. Preliminary analyses revealed that this threshold had no impact on the results presented as long as it was reasonably large (≥ 50).

1.3 Network measures

We measured four key families of bipartite network structure indices. Measures were ranged in $[0, 1]$. 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, nestedness²¹, using the NODF measure, which **todo**. Third, modularity, using LP-BRIM^{22,23}, 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 motifs²⁴. [\[more\]](#)

So that the motif statistics would also fall in the range $[0, 1]$, we corrected the raw number of motifs to account for the number of species in each layer of the bipartite network. For example, the maximum number of motifs with 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 $\binom{T}{2}$ and $\binom{B}{2}$, respectively). This gives a total number of sets of species that *could* be involved in a 2×2 motif. Note that this implies that all values represent 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 observed set of 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 then selected the posterior distribution as the 500 parameters sets that gave the best scores (i.e. above the 95th percentile). [\[ref\]](#)?

1.5 Decision tree

We used a classification tree to separate the networks along the continuum of values of c and λ . The response was the type of networks, and the classifiers were the \log_{10} of c and λ . We used the implementation from the `tree` package (v. 1.0.36) for R (v. 3.2.2). Splits were decided according to Gini ratio. [\[weights?\]](#)

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).
9. Donoghue, P. C. & Benton, M. J. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution* **22**, 424–431 (2007).
10. Csilléry, K., Blum, M. G., Gaggiotti, O. E. & François, O. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418 (2010).
11. Wilkinson, R. D. Approximate Bayesian computation (ABC) gives exact results under the as-

sumption of model error. *Stat. Appl. Genet. Mol. Biol.* **12**, 129–141 (2013).

12. Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *PNAS* **104**, 19891–19896 (2007).

13. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458**, 1018–1020 (2009).

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

15. Poisot, T., Bever, J. D., Nemri, A., Thrall, P. H. & Hochberg, M. E. A conceptual framework for the evolution of ecological specialisation. *Ecol. Lett.* **14**, 841–851 (2011).

16. Williams, R. J. Biology, Methodology or Chance? The Degree Distributions of Bipartite Ecological Networks. *PLoS One* **6**, e17645 (2011).

17. Thébault, E. & Fontaine, C. Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. *Science* **329**, 853–856 (2010).

18. Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host-phage interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E288–297 (2011).

19. Stanko, M., Miklisová, D., Göüy de Bellocq, J. & Morand, S. Mammal density and patterns of ectoparasite species richness and abundance. *Oecologia* **131**, 289–295 (2002).

20. Thebault, E. & Fontaine, C. Does asymmetric specialization differ between mutualistic and trophic networks? *Oikos* **117**, 555–563 (2008).

21. Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*

203 **117**, 1227–1239 (2008).

204 22. Liu, X. & Murata, T. Community detection in large-scale bipartite networks. in *Transactions*
205 *of the Japanese Society for Artificial Intelligence* **5**, 184–192 (Institute of Electrical & Electronics
206 Engineers (IEEE), 2009).

207 23. Barber, M. & Clark, J. Detecting network communities by propagating labels under constraints.
208 *Phys. Rev. E* **80**, 1–11 (2009).

209 24. Baker, N. J., Kaartinen, R., Roslin, T. & Stouffer, D. B. Species' roles in food webs show fidelity
210 across a highly variable oak forest. *Ecography* (2014). doi:10.1111/ecog.00913

211 **List of Figures**

212	1	Posterior distributions of parameters p , λ and c	12
213	2	Relationships between parameters λ and c	13
214	3	Predictive power of the model across different types of networks.	14
215	4	Classification tree of the networks as a function of best parameters values.	15

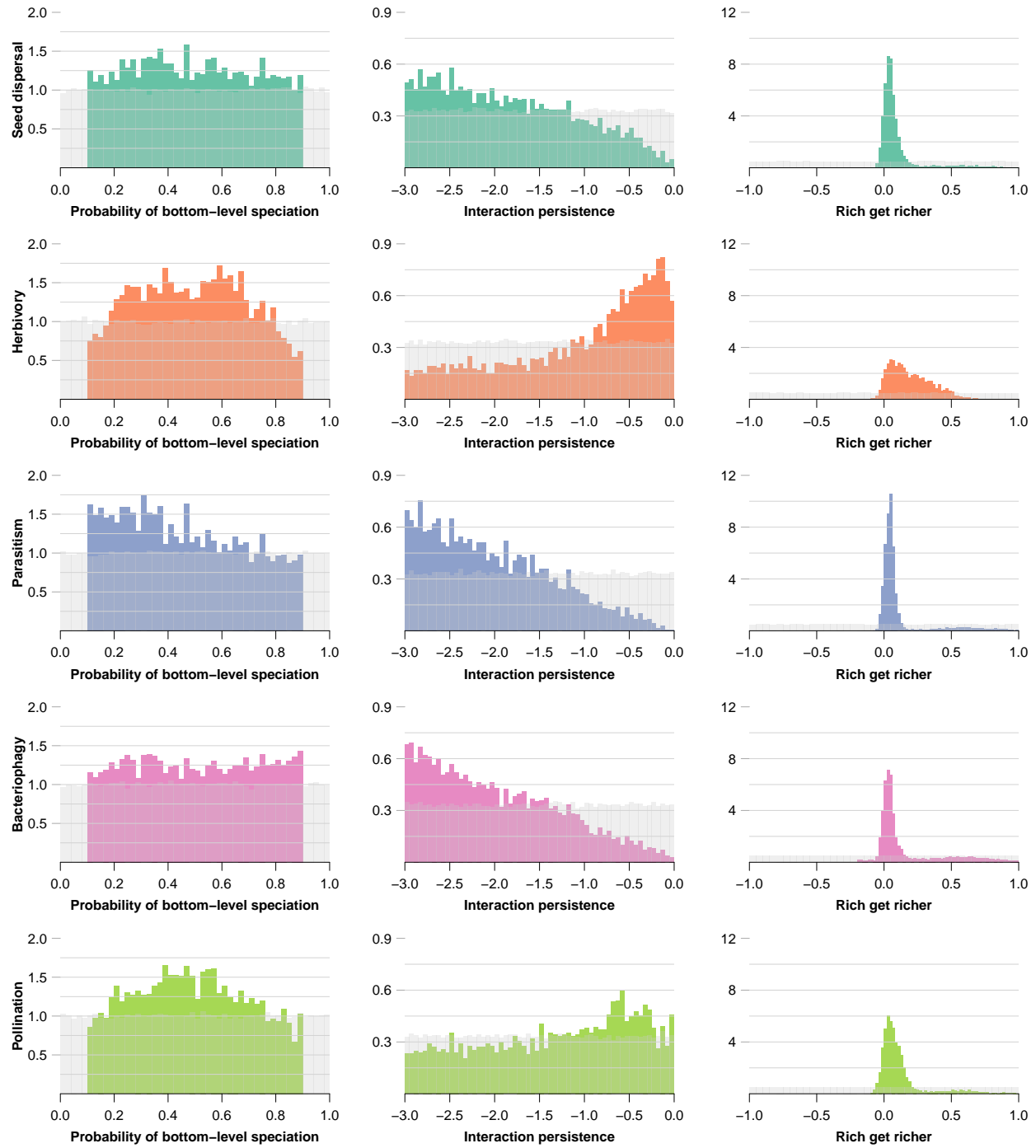


Figure 1: Posterior distributions of parameters p , $\log_{10}\lambda$ and $\log_{10}c$. The grey shaded area is a representation of the uniform prior distribution. Although there is no strong selections on the values of p , networks do differ strongly both from the prior, and from one another, on λ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

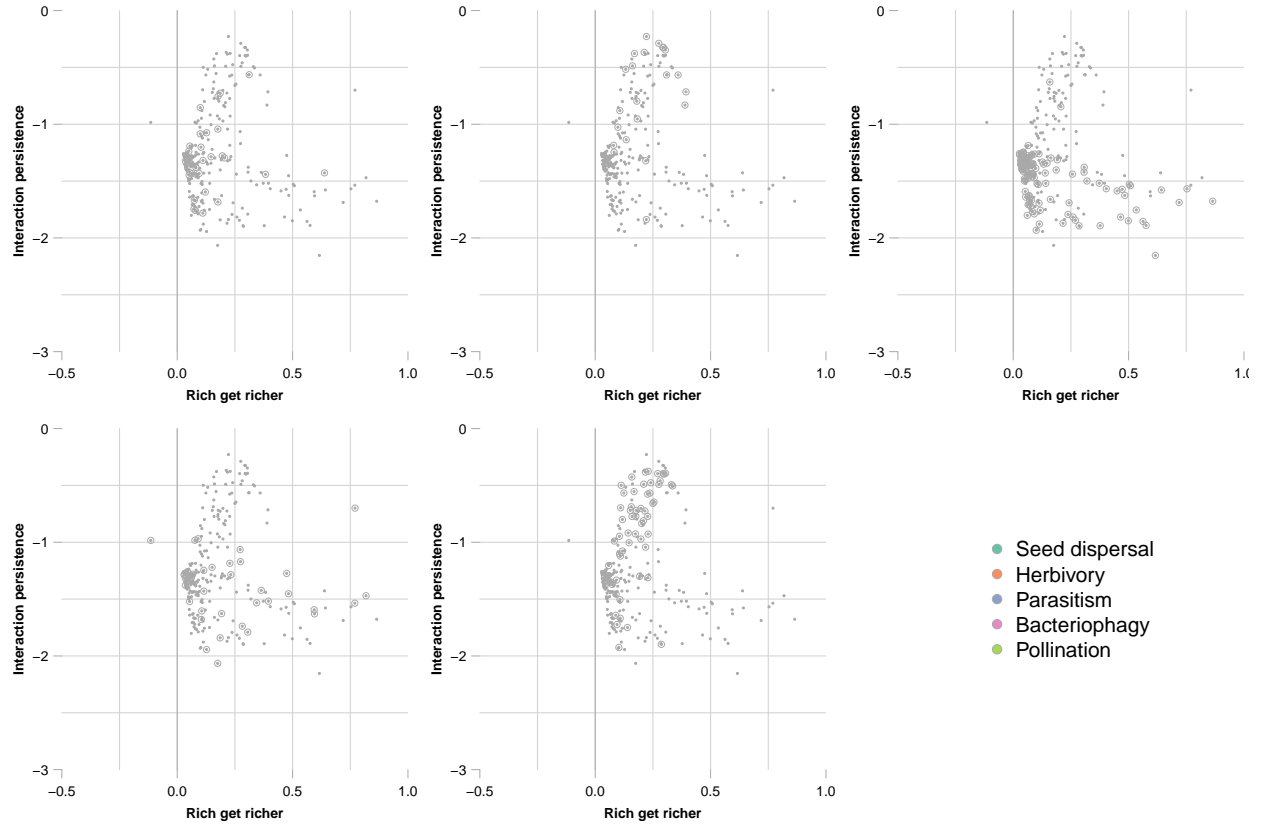


Figure 2: Relationships between parameters λ and c in the five different types of networks. The grey points indicate all networks in the dataset. Different types of ecological interactions occupy different positions along the λ - c continuum.

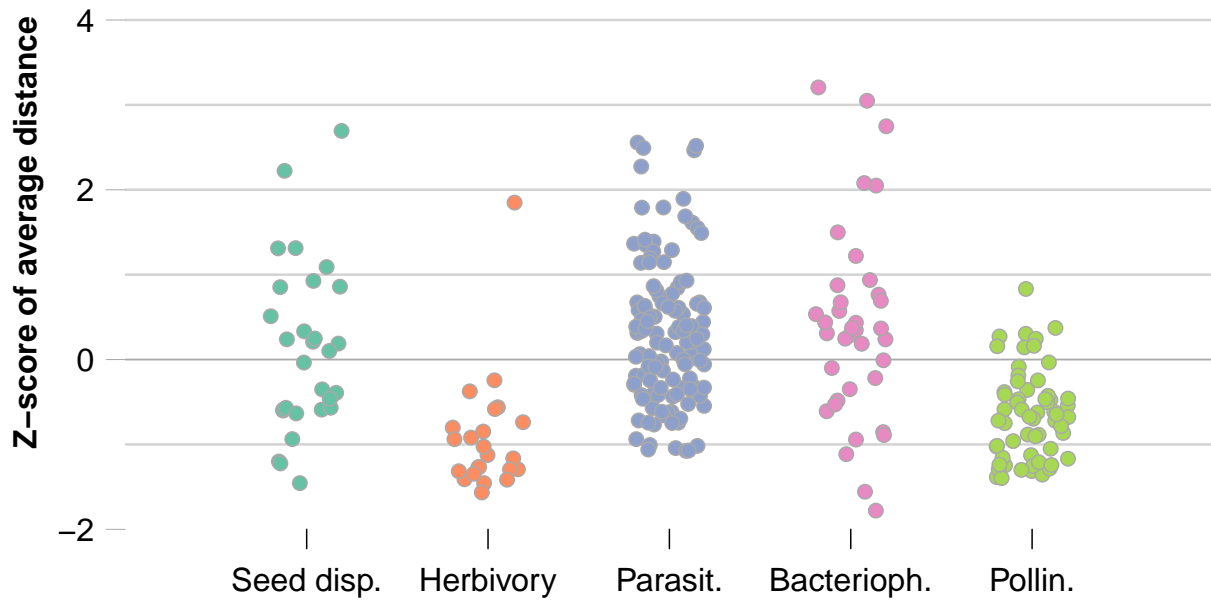


Figure 3: Z-score of *average* distances for the best-matching simulations. Herbivory and pollination networks are better predicted by this model, while z-scores for seed dispersal, parasitism, and bacteriophagy, are centered around 0. The differences in z-scores may come from the fact that macro-evolutionary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

