

## Supplementary Material

Appendix 1 to *Deciphering probabilistic species interaction networks*

Francis Banville<sup>1,2,3</sup> Tanya Strydom<sup>1,3</sup> Penelope S. A. Blyth<sup>4</sup> Chris Brimacombe<sup>5</sup> Michael Catchen<sup>3,6</sup>  
Gabriel Dansereau<sup>1,3</sup> Gracielle Higino<sup>2</sup> Thomas Malpas<sup>4</sup> Hana Mayall<sup>4</sup> Kari Norman<sup>1</sup>  
Dominique Gravel<sup>2,3</sup> Timothée Poisot<sup>1,3</sup>

<sup>1</sup> Université de Montréal <sup>2</sup> Université de Sherbrooke <sup>3</sup> Quebec Centre for Biodiversity Science <sup>4</sup> University of Sheffield <sup>5</sup> University of Toronto <sup>6</sup> McGill University

## Host-parasite network data

**1.1. Data description** We use the collection of tripartite host-parasite networks sampled across Europe of Kopelke *et al.* (2017). This dataset contains well-resolved binary local interactions between willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Out of a total of 374 local networks, we retained those containing at least 5 species, resulting in a set of 233 georeferenced local networks (networks sampled within areas of 0.1 to 0.3 km<sup>2</sup> during June and/or July spanning 29 years). Given its replicated networks spanning large spatiotemporal scales, this dataset is well-suited for analyzing network variability.

We built a metaweb of binary interactions by aggregating all local interactions, which gave us a regional network composed of 274 species and 1080 interactions.

**1.2. Metawebs of probabilistic interactions** We converted these binary regional interactions into probabilistic ones using simple assumptions. Our aim is not to estimate precise probability values, but to create plausible metawebs of probabilistic interactions for our illustrative examples.

We created two metawebs of probabilistic interactions by employing constant false positive and false negative rates for all regional interactions. In the first metaweb, we set both false positive and false negative rates to zero to prevent artificially inflating the total number of interactions, enabling a more accurate comparison with binary interaction networks. This gave us a probability of regional interaction of 1 when at least one interaction has been observed locally and of 0 in the absence of any observed interaction between a given pair of species. This metaweb was used in Box 2.

In the second metaweb, we introduced a 5% false positive rate to account for spurious interactions and a 10% false negative rate to address the elevated occurrence of missing interactions in ecological networks (Catchen *et al.* 2023). We believe these rates represent reasonable estimates of missing and spurious potential interactions, but confirming their accuracy is challenging due to the unavailability of data on the actual feasibility of interactions. Observed interactions were thus given a probability of regional interaction of 95%, whereas unobserved ones were assigned a probability of 10%. This metaweb was used in Boxes 3 and 5.

**1.3. Local networks of probabilistic interactions** We built local networks of probabilistic interactions using the taxa found in the empirical local networks and attributing pairwise interaction probabilities based on the metawebs of probabilistic interactions  $P(M_{i,j})$  and a constant value of  $P(L_{i,j,k}|M_{i,j})$  across interactions:

$$P(L_{i,j,k}) = P(L_{i,j,k}|M_{i,j}) \times P(M_{i,j}). \quad (1)$$

We set all values of  $P(L_{i,j,k}|M_{i,j})$  to 0.5, 0.75, or 1.0 depending on the simulation. Intermediate values of  $P(L_{i,j,k}|M_{i,j})$  around 50% indicate considerable spatiotemporal variability, while higher values close to 1.0 indicate that regional interactions are nearly always realized locally.

## Methods for Box 2

**2.1. Methods for Figure 2 (a-b)** To measure the dissimilarity between the metaweb of binary interactions and aggregated local networks, we randomly selected one local network of binary interactions and sequentially

sampled additional networks while aggregating both their species and interactions. We repeated this sampling process one hundred times and highlighted the median dissimilarity values across simulations, as well as the 50% and 95% percentile intervals.

We compared the metaweb and the aggregated local networks using the dissimilarity in species composition  $\beta_S$ , and the dissimilarity of interactions between common species  $\beta_{OS}$  indices. Both dissimilarity indices were calculated based on the number of items shared by the two networks ( $c_{LM}$ ) and the number of items unique to the metaweb ( $u_M$ ) and the aggregated local network ( $u_L$ ). The  $\beta_S$  dissimilarity index uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions between shared species. Both indices were calculated following the  $\beta_W$  index of Whittaker (1960):

$$\beta_W = \frac{c_{LM} + u_L + u_M}{(2c_{LM} + u_L + u_M)/2} - 1.$$

**2.2. Methods for Figure 2 (c-d)** We investigate how the number of local interactions and connectance scale with the number of sampled (aggregated) local networks of probabilistic or binary interactions. We built probabilistic local networks following Eq. 1, and set  $P(M_{i,j})$  to 1 when the interaction was observed at least once, and to 0 otherwise (i.e., no false positive or false negative in the metaweb). We used a constant value of  $P(L_{i,j,k}|M_{i,j})$  for all non-aggregated interactions.

When aggregating local networks of probabilistic interactions, the constancy of the probability of regional interaction across the entire study area means that any rise in the probability of local interaction is solely attributable to an increase in  $P(L_{i,j,k}|M_{i,j})$ . For example, let  $L_1$  and  $L_2$  be two local networks and  $L_{1,2}$  the aggregated network. If  $P(L_{i,j,1}|M_{i,j})$  and  $P(L_{i,j,2}|M_{i,j})$  are the probabilities that two potentially interacting taxa interact respectively in  $L_1$  and  $L_2$ , the probability  $P(L_{i,j,1,2}|M_{i,j})$  that these taxa interact in the aggregated network  $L_{1,2}$  is obtained by:

$$P(L_{i,j,1,2}|M_{i,j}) = 1 - [1 - P(L_{i,j,1}|M_{i,j})] \times [1 - P(L_{i,j,2}|M_{i,j})], \quad (2)$$

assuming independence between the interaction of the two taxa in different networks. This equation represents the probability that the interaction is realized in either (1) exclusively the local network  $L_1$ , (2) exclusively the local network  $L_2$  or (3) both, given that the two taxa have the biological capacity to interact.

Aggregated local networks were obtained by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions, with the value of  $P(L_{i,j,k}|M_{i,j})$  adjusting according to Eq. 2.

---

## References

- Catchen, M.D., Poisot, T., Pollock, L.J. & Gonzalez, A. (2023). The missing link: Discerning true from false negatives when sampling species interaction networks.
- Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). [Food-web structure of willow-galling sawflies and their natural enemies across Europe](#). *Ecology*, 98, 1730–1730.
- Whittaker, R.H. (1960). [Vegetation of the Siskiyou Mountains, Oregon and California](#). *Ecological Monographs*, 30, 279–338.