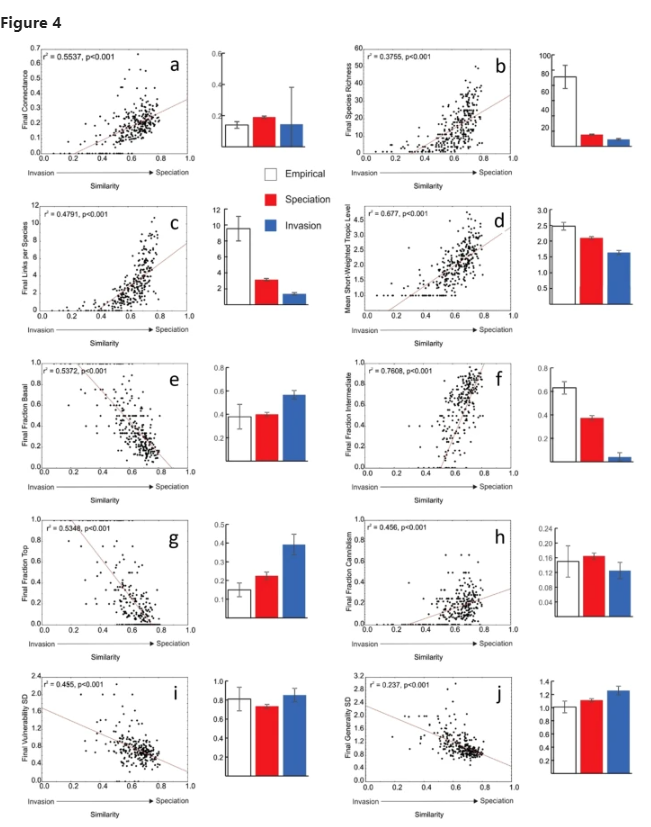
**What methods are used to compare simulated networks with empirical ones?**

We compared the structure of food webs among those primarily assembled by speciation, primarily assembled by invasion, and 19 well-known empirical food webs[57](https://www.nature.com/articles/s41598-019-54443-0#ref-CR57) using 10 measures of network structure. One property is simply the number of species within the food web (*S*). Two other properties are standard measures of food-web trophic interaction richness[45](https://www.nature.com/articles/s41598-019-54443-0#ref-CR45): links per species (*L*/*S*) also referred to as link density; and directed connectance (*C* = *L*/*S*2) which equals the proportion of all possible trophic links that are actually realized. Five more properties indicate the fraction of the following types of species in a food web: top (%*T*, species that have resource species but lack any consumer species[102](https://www.nature.com/articles/s41598-019-54443-0#ref-CR102)), intermediate (%*I*, species that have both resource and consumer species[102](https://www.nature.com/articles/s41598-019-54443-0#ref-CR102)), basal species (%*B*, species that have consumer species but lack resources species e.g., plants[102](https://www.nature.com/articles/s41598-019-54443-0#ref-CR102)); cannibals (*%C*, species that eat themselves)[43](https://www.nature.com/articles/s41598-019-54443-0#ref-CR43); and omnivores (%*Omn*, species that eat species at different trophic levels)[43](https://www.nature.com/articles/s41598-019-54443-0#ref-CR43). Trophic level is calculated as short weighted trophic level (*SWTL*), a measure of trophic level based on mere presence of links that accurately estimates trophic level based on quantitatively weighted links[103](https://www.nature.com/articles/s41598-019-54443-0#ref-CR103). Two additional properties are the standard deviation of mean generality (*GenSD*) and vulnerability (*VulSD*) among species which quantify the variabilities of species’ normalized predator and prey counts respectively[43](https://www.nature.com/articles/s41598-019-54443-0#ref-CR43),[104](https://www.nature.com/articles/s41598-019-54443-0#ref-CR104).



I could do this for each scenario. Measure those metrics and plotting them in a dotplot where the y axis are the 10 metrics, the x axis are the values, and the groups are each scenario (for a visual inspection). Then I could calculate a score that sums the proximity of their values with the empirical networks (empirical value – simulated value) and choose the one with the lowest metric (i.e. more close to the empirical).

Now, how do I measure these metrics?

1. Number of species within the foodweb (S)

2. liks density (L/S)

3. directed connectance (C = L/S2)

4. %T (Top species that have resource species but lack any consumer species)

5. % I (Intermediate species that have both resource and consumer species)

6. % B (Basal species that have consumer species but lack resources species)

7. *%C* (*Cannibal* species that eat themselves)

8. %Omn (species that eat species at different trophic levels)

9. Standard deviation of mean generality (GenSD)

10. Standard deviation of mean vulnerability (VulSD)

\*\*\* With the function motifs(graph, 3) of igraph I see the frequency of motifs for the 16 triad isomorphism classes

