Supplementary Material for

Untangling the complexity of priority effects in multispecies communities

in *Ecology Letters*

Chuliang Song[1,2,3], Tadashi Fukami[4], Serguei Saavedra[1]

[1]Department of Civil and Environmental Engineering, MIT,
77 Massachusetts Av., 02139 Cambridge, MA, USA

[2]Department of Biology, McGill University,
1205 Dr. Penfield Avenue, Montreal, H3A 1B1 Canada

[3] Department of Ecology and Evolutionary Biology, University of Toronto,
25 Willcocks Street, Toronto, Ontario M5S 3B2 Canada

[4]Department of Biology, Stanford University, Stanford,
California 94305 USA

Contents

A Diversity of assembly dynamics

A.1 First-order estimation of the diversity of assembly dynamics

The diversity of assembly dynamics is defined as the number of topologically different assembly graphs. Here, we derive a first-order approximation of the diversity of priority effects.

By the construction of assembly graphs, the total number of assembly graphs (allowing topological equivalence) is

$$\underbrace{\prod_{n=1}^{S}}_{\substack{\text{focusing on the nodes}\\ \text{with } n \text{ species}}} \left(\underbrace{2^{n+1}-1}_{\substack{\#\text{ of nodes each node with}\\ n \text{ species can connect to}}} \right)^{\underbrace{(S-n)}_{\substack{\#\text{ of}\\ \text{outgoing links}}} \underbrace{\binom{S}{n}}_{\substack{\#\text{of nodes with}\\ n \text{ species}}}} \tag{S1}$$

Since the total number of labelling S species is $n!$, the first-order approximation of the number of topologically different assembly graphs is

$$\frac{\prod_{n=1}^{S}\left(2^{n+1}-1\right)^{(S-n)\binom{S}{n}}}{S!} \tag{S2}$$

Then, we derive the leading asymptotic term of the diversity by taking the logarithm,

$$\log\left(\frac{\prod_{n=1}^{S}\left(2^{n+1}-1\right)^{(S-n)\binom{S}{n}}}{S!}\right) = \sum_{n=1}^{S}(S-n)\log\left(2^{n+1}-1\right)\binom{S}{n} - \log\left(S!\right) \tag{S3}$$

$$\sim \sum_{n=1}^{S}(S-n)\log\left(2^{n+1}\right)\binom{S}{n} - \log\left(S!\right) \tag{S4}$$

$$= 2^{S-2}S(S+1)\log(2) - \log\left(S!\right) \tag{S5}$$

$$\sim 2^{S-2}S^2\log(2) - \sqrt{2\pi}e^{-S}S^{S+\frac{1}{2}} \tag{S6}$$

Thus, we have

$$\frac{\prod_{n=1}^{S}\left(2^{n+1}-1\right)^{(S-n)\binom{S}{n}}}{S!} \sim \exp(2^{S-2}S^2\log(2) - \sqrt{2\pi}e^{-S}S^{S+\frac{1}{2}}) \tag{S7}$$

$$= \frac{2^{2^{S-2}S^2}e^S}{\sqrt{2\pi}S^{S+\frac{1}{2}}}, \tag{S8}$$

which is a super-exponential increase.

We then numerically validate the approximation in Figure S1. Even for 3 species, the approximation works: the diversity of assembly dynamics estimated by Eqn. S8 is 41,674, which is close to the correct diversity 41,979 (the error is less than 1%). Figure S1 provides evidence that this first-order approximation works.

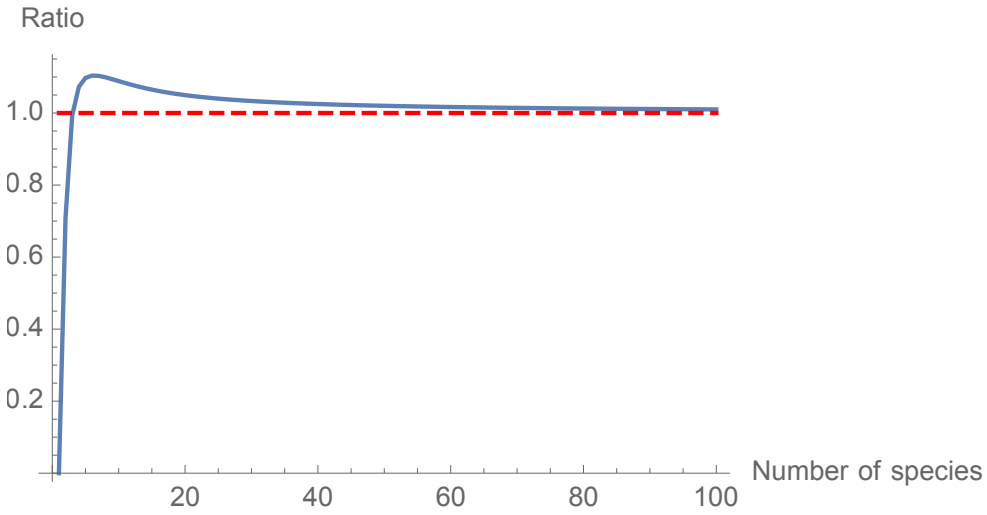

Figure S1: This figure shows the analytic first-order estimation versus the asymptotic approximation of the first-order estimation. The x-axis represents the number of species while the y-axis represents the ratio between analytic versus asymptotic. The ratio quickly approaches to 1.

A.2 Comparing the diversity of assembly dynamics and that of binary interaction matrix

We compare the diversity of assembly dynamics with the diversity of binary interaction matrix. For S species, a first-order approximation shows that the diversity of topologically different binary interaction matrix scales as

$$\frac{2^{(S^2)}}{S!} \sim \frac{2^{S^2-\frac{1}{2}} e^S}{\sqrt{\pi} S^{S+\frac{1}{2}}}, \tag{S9}$$

which is an exponential increase.

Figure S2 plots how the diversity of assembly dynamics and of binary interaction matrix scales with community size.

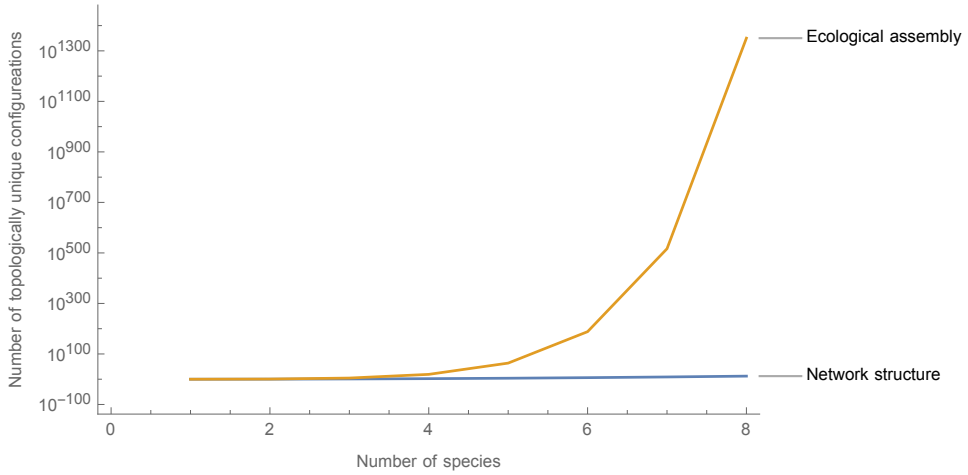

Figure S2: Comparing the diversity of assembly dynamics with that of binary interaction matrix. As expected, the diversity of assembly dynamics is much richer than that of binary interaction matrix.

B Pool of assembly histories

In the main text, we have considered equal invasion times and random arrival times for all species. However, our theoretical framework is general and flexible to modify this assumption to account for more realistic ecological patterns. Here we consider two types of modifications.

One type of modification is to consider different invasion times. An ecological example is unequal dispersal rates among species. In general, if we allow species i invades m_i times, the pool consists of $(\sum_{j=1}^{S} m_j)!/(m_1! \cdots m_S!)$ histories. Another type of modification is to consider nonrandom arrival times among species. An ecological example is seasonal phenological events, where some species always arrive earlier than some other species during a season. Given the specific constraints, we can use combinatorics to enumerate all the possible histories. Regardless of the modifications, we can follow the same procedure outlined in Figure 2A to quantify the predictability under the given pool of assembly histories.

As a proof of concept, we here present a case where species have different invasion times. For example, suppose we have a 3-species community with 6 invasions in total. What was studied in the main text are the cases where each species invades twice. However, we can modify it by considering two other possibilities with unequal invasion times: (1) one species invades 4 times, and the two other species each invades 1 time; (2) one species invades 3 times, another one invades 2 times, and the other one invades 1 time. Figure S3 shows all cases of species invasion times in a 3-species community with 6, 8, and 10 total invasions, respectively. We found that the explanatory power of neural networks is highest with equal number of invasions and decreases as the number of invasions becomes more unequal.

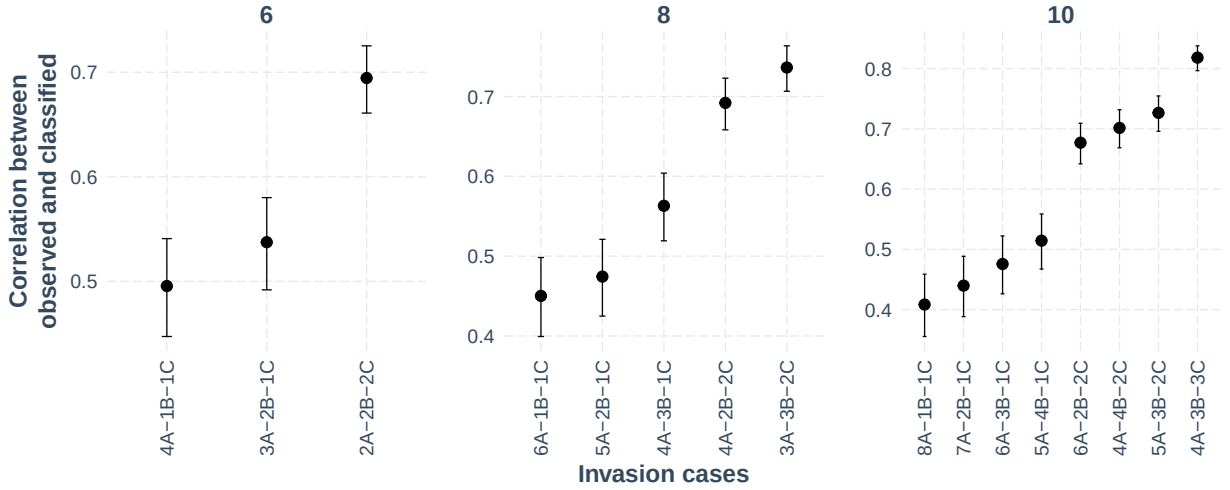

Figure S3: **Effects of unequal invasion times.** Here we focus on 3-species communities. Each panel shows different number of total invasions (6, 8, and 10, respectively). The x axis denotes different cases with the same number of total invasions. For example, 4A-1B-1C denotes that species A invades 4 times while species B and C each invade 1 time. The y axis denotes the correlation between observed predictability and classified predictability (following the same procedure of the neural network in Appendix C). The correlation is higher when species have similar numbers of invasion times.

C Implementation of neural network

We employed a neural network to explain the predictability of community assembly from the four topological sources in the assembly dynamics (Murphy, 2012). The architecture is visualized in Figure C. The input layer is a four-dimensional vector, which encodes the four topological features of an assembly graph. The five hidden layers all have ReLU activation. The output layer is the explained predictability from the input. 80% of data is used as the training and validation sets, while the remaining 20% of data is used as the test set. We used the R package `keras` (Allaire & Chollet, 2021), which provides an R interface to Keras (Chollet *et al.*, 2015).

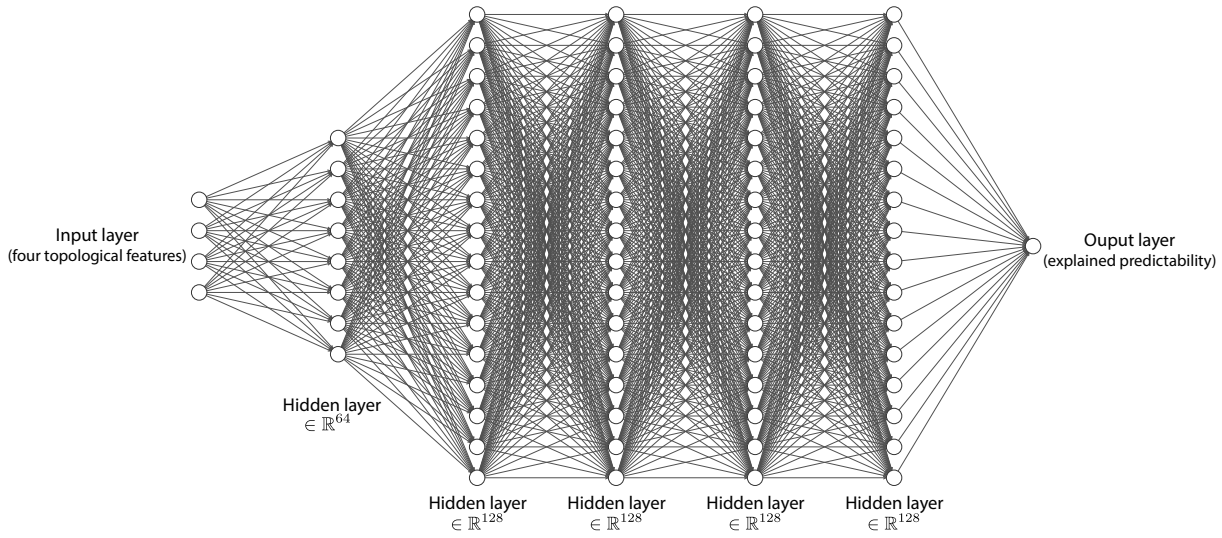

Figure S4: Architecture of neural network. It has four hidden layers. The first hidden layer has 64 nodes while all other hidden layers have 128 nodes. For visualization, not all nodes in the hidden layers were shown.

D Classification of priority effects in species-rich communities

We have presented the results for 3-species communities in the main text. Here we show that our key results stand for 4- and 5-species communities as well. As Figure 1 and Appendix A show, it is computationally infeasible to enumerate all possible assembly dynamics for more than 3 species. Thus, we randomly draw 5,000 assembly graphs for 4- and 5-species communities, respectively. Figures S5 and S6 show the results of classification for 4- and 5-species communities.

We find that the explanatory power (i.e., the correlation between observed and classified predictability) is around 0.9 when repeated invasions are allowed. It is slightly lower than the explanatory power in 3-species communities. The decrease in explanatory power is likely due to the small sampling size (n = 5,000) for 4- and 5-species communities. However, designing an unbiased sampling is challenging. For example, the unbiased sampling of species abundance under constraints has only been resolved recently (Diaz *et al.*, 2021; Locey & White, 2013). Thus, an unbiased, effective sampling of assembly graphs warrants further investigation.

We also find that, with more species (from 3 to 4 to 5), the contribution of the number of stable states to community predictability increases. We expect that this trend continues for even more species. However, given the exponential computational complexity in computing the graph properties of interest (stable states, transient paths, etc.), we were unable to verify this hypothesis via simulation. If the hypothesis is correct, laboratory experiments should focus on stable states in large communities, and on transient paths and compositional cycles in small communities. Note that most direct experimental tests of stable states have focused on small communities (Schröder *et al.*, 2005).

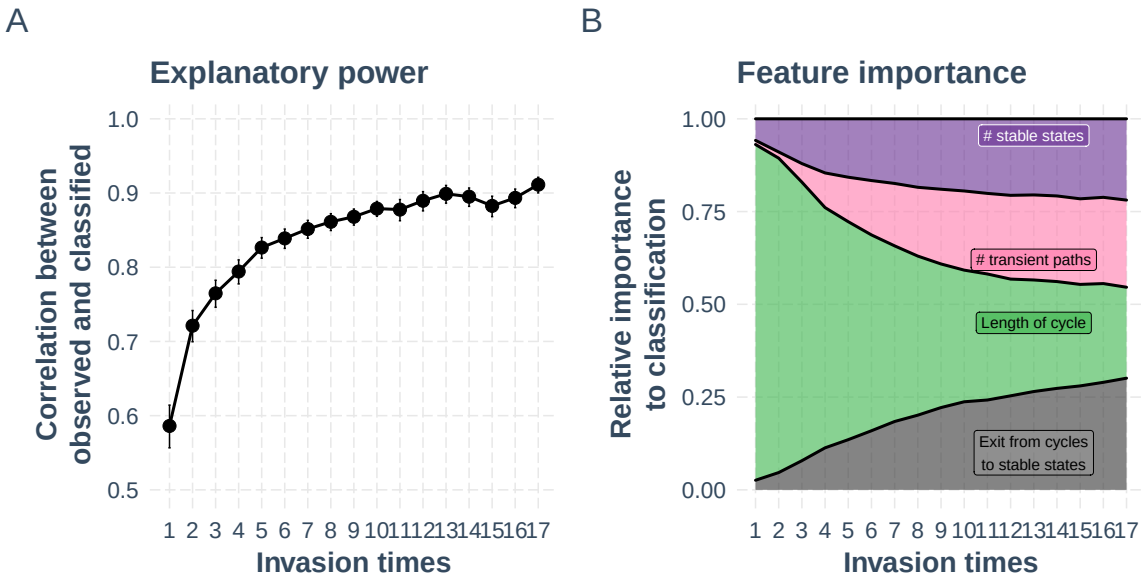

Figure S5: Same as Figure 5 except we consider 4 species.

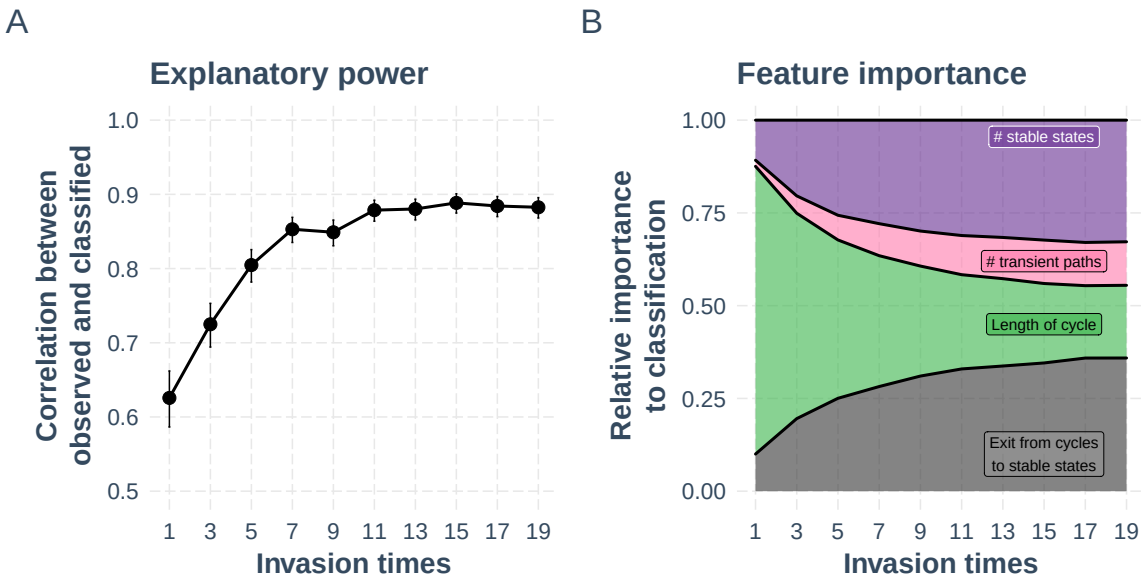

Figure S6: Same as Figure 5 except we consider 5 species.

E Unreachable nodes and links

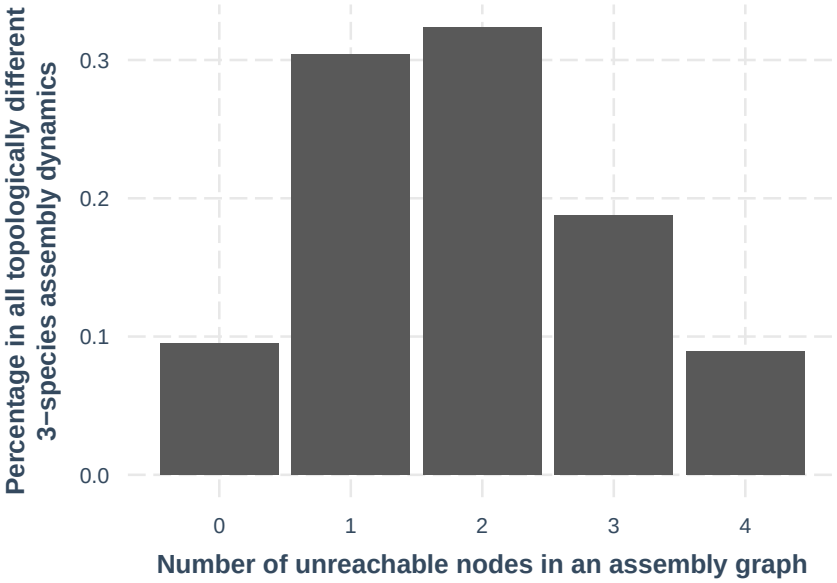

Figure S7: We define unreachable nodes as links that cannot be possibly reached via bottom-up assembly. The horizontal axis denotes the number of unreachable nodes in an assembly graph, while the vertical axis denotes the percentage of reachable nodes among all topologically different 3-species dynamics.

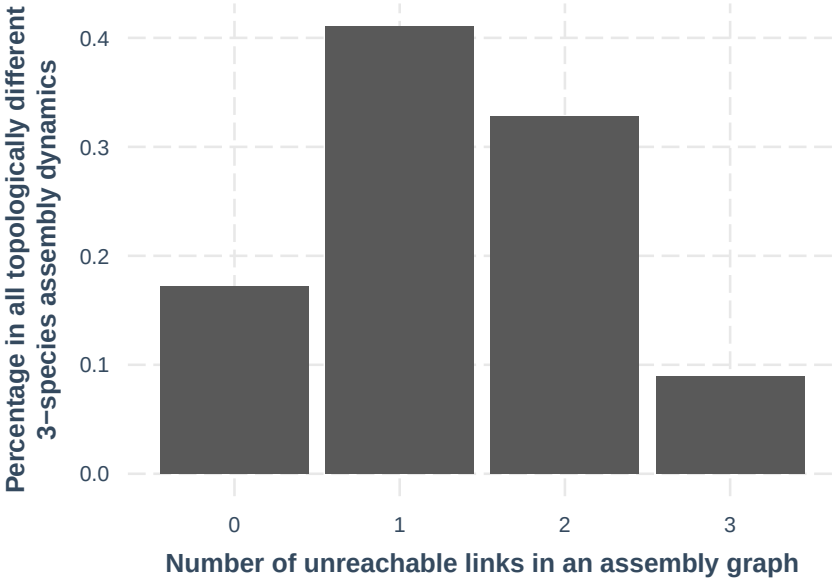

Figure S8: We define unreachable links as links that cannot be possibly reached via bottom-up assembly. The horizontal axis denotes the number of unreachable links in an assembly graph, while the vertical axis denotes the percentage of reachable links among all topologically different 3-species dynamics.

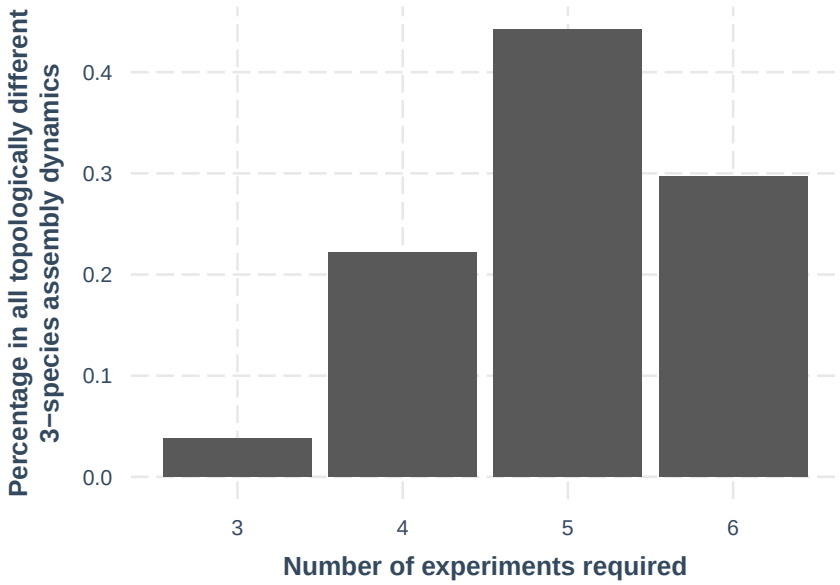

Figure S9: One experiment refers to a scenario where each species is introduced once into the community. In this sense, 3-species communities can perform 6 different experiments. However, not all experiments are always needed to map the assembly graph. The horizontal axis denotes the number of such experiments required to fully infer the assembly graph, while the vertical axis denotes the percentage of reachable nodes among all topologically different 3-species dynamics.

References

Allaire, J. & Chollet, F. (2021). *keras: R Interface to 'Keras'*. URL `https://CRAN.R-project.org/package=keras`. R package version 2.4.0.

Chollet, F. *et al.* (2015). Keras. URL `https://github.com/fchollet/keras`.

Diaz, R. M., Ye, H. & Ernest, S. M. (2021). Empirical abundance distributions are more uneven than expected given their statistical baseline. *Ecology Letters*.

Locey, K. J. & White, E. P. (2013). How species richness and total abundance constrain the distribution of abundance. *Ecology letters*, 16, 1177–1185.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Schröder, A., Persson, L. & De Roos, A. M. (2005). Direct experimental evidence for alternative stable states: a review. *Oikos*, 110, 3–19.