T is for Topology

Tanya Strydom Andrew P. Beckerman

2024-03-18

There are many reasons one might want to generate a network and there are many tools on the market that might make that possible. However not all tools are created equally and there is reason to assume that not all networks will suit most purposes. Here the aim is to compare and contrast the different topology generating tools that are on the market and see where they shine and where they fall flat. There probably isn't one model to rule them all but it doesn't mean that we shouldn't be critical when we think about the model we want to use.

1 Introduction

- In order to construct a 'perfect' network *i.e.,* one which *perfectly* captures the dynamics for a specific community one needs to consider and account for many different moving parts (*e.g.,*). So when developing a model it makes sense that you prioritise the aspect of the prediction/construction task that has the most value for your research goal, acknowledging that a model might fall short in others. The thing is that with the growing suite of approaches to generating networks it is important that we don't lose sight of the core philosophy behind the model we use and to ensure that we are using the model best suited to what we want to be accomplishing.
- It is perhaps useful to start with asking why do we want/need models to generate networks. This can be broadly thought of to fall into two categories. Build networks because we want to build concepts vs build networks because we want specificity. Broadly this means that we either want to construct/predict a collection of interactions (generate networks) or a network of interactions (predict interactions).

Arguably the need for methods and tools for constructing interaction networks arises from two different (but still aligned) places of interest within the field of network ecology. On the one side sits the researcher who is interested in generating a set of ecologically plausible but not necessarily realised 'in the field' networks for the purpose of running further simulations (*e.g.,* extinction sim **TODO**) or understanding some higher-level process/concept (*e.g.,* energetics

TODO). This researcher is contrasted by one that is interested in constructing real-world, location specific, interaction data in lieu of inventorying interactions based n observations made in the field (Strydom et al. 2021). Of course these two categories are not two distinct, mutually exclusive, groups but can rather be viewed as operating on a gradient ranging from a need for generality (*i.e.,* creating a network that, when taken in aggregate, the distribution of links (interactions) between species are ecologically plausible) to a need for specificity (local-level predictions between specific species).

- A breakdown of network generators; statement of need and core philosophies
- A breakdown of interaction predictors; statement of need ((Jordano 2016b, 2016a; Poisot et al. 2021)) and core philosophies (trait-matching, coexistence)
- Stands to reason then that we have developed methods specialise in one or the other. Which comes at a cost of 'performance' in other aspects. Knowing how the different model families stack up to each other is thus valuable and that is kinda what we are trying to achieve.

Joel E. Cohen, Newman, and Steele (1985) states that *"[Their] approach is more like gross anatomy than like physiology… that is, the gross anatomy is frozen, rather than in motion."*.

Interestingly Williams and Martinez (2008) also explicitly talk about *structural* food-web models in their introduction… so how I see it that means that there has always been this inherent acknowledgement that models are functioning at a specific 'network level'.

1.1 The history behind the approach

Maybe a brief history of the development of predictive tools/topo generators? Sort of where the theory/body of work was based and how that has changed? IS there a difference between topo generator and predictive tool - I'm inclined to think that it aligns with the whole debate of high level structure vs node-level perfection

Maybe start here with discussing the core mechanistic differences that models will work at — some are really concerned about (and thus constrained by) structure, others are more mechanistic in nature *i.e.,* species a has the capacity to eat species b because traits (read gob size), and then you get Rohr et al. (2010) and Strydom et al. (2022) that sit in the weird liminal latent space…

Here I will probably get on my (newly discovered) soapbox and wax lyrical about how in certain situations structure is enough (and that will probably be for some high-level things like thinking about energy flows etc., I can also see a world in which maybe you want to do some sort of robustness/extinction work - since then you're usually doing 'random' (within limits) extinctions) but there may be use cases where we are really interested in the node-level interactions *i.e.,* species identity is like a thing we need to care about and also be able to retrieve specific interactions at specific nodes correctly. What is the purpose of generating a

network? Is it an element of a bigger question we are asking, *e.g.,* I want to generate a series of networks to do some extinction simulations/bioenergetic stuff OR are we looking for a 'final product' network that is relevant to a specific location? (this can still be broad in geographic scope).

At some point we are going to need to discuss the key differences and implications between predicting a metaweb (*sensu* Jennifer A. Dunne (2006)) and a network realisation. And here I can't help but think about Poisot, Stouffer, and Gravel (2015) (and probably other papers) that discuss how the local factors are going to play a role and even the same pair of species may interact differently in different points in the landscape.

> Do we need to delve into individual-based networks? (*sensu* Tinker 2012, Araújo 2008) I think its probably a step too far and one starts creeping into apples and pears type of comparisons. Especially since these work off of already existing networks (I seem to recall) and its more about about 'tweaking' those - so not so much *de novo* predictions. Although this might be useful to keep in mind when it comes to re-wiring… Also on that note do we opn the re-wiring door here in this ms or wait it out a bit.

2 Data & Methods

Here we only look at families of models that are explicitly developed to construct *de novo* networks, be this in the form of either artificial or synthetic networks.

2.1 Topology Generators

Null models (Erdős and Rényi 1959): Links are assembled randomly, not developed within an ecological framework. But of course could still hold if we assume that communities are randomly assembled in terms of who is interacting with who (I seem to think that's sort of what May was arguing but I would need to remind myself)

Neutral models

Cascade model (Joel E. Cohen, Briand, and Newman 1990): Much like the name suggests the cascade model rests on the idea that species feed on one another in a hierarchical manner. This rests on the assumption that the links within a network are variably distributed across the network; with the proportion of links decreasing as one moves up the trophic levels (*i.e.,* 'many' prey and 'few' predators). This is achieved by assigning all species a random rank, this rank will then determine both the predators and prey of that species. A species will have a particular probability of being fed on by any species with a higher ranking than it, this probability is constrained by the specified connectance of the network. Interestingly here 'species' are treated as any individual that consume and are consumed by the same 'species',

i.e., these are not taxonomical species (Joel E. Cohen, Newman, and Steele 1985). The original cascade model has altered to be more 'generalised' (Stouffer et al. 2005), which altered the probability distribution of the prey that could be consumed by a species.

Niche models (Williams and Martinez 2000): The niche model introduces the idea that species interactions are based on the 'feeding niche' of a species. Broadly, all species are randomly assigned a 'feeding niche' range and all species that fall in this range can be consumed by that species (thereby allowing for cannibalism). The niche of each species is randomly assigned and the range of each species' niche is (in part) constrained by the specified connectance of the network. The niche model has also been modified, although it appears that adding to the 'complexity' of the niche model does not improve on its ability to generate a more ecologically 'correct' network (Williams and Martinez 2008).

Nested hierarchy model (Cattin et al. 2004):

> Gravel et al. (2013) also poses an interesting cross-over between the adbm and niche model.

2.2 Interaction Predictors

Allometric diet breadth model (ADBM) (Petchey et al. 2008): "Our modelling effort therefore asks how well our model can predict the arrangement rather than the number of feeding links in food webs." This is a scaled version of the diet breadth model (Beckerman, Petchey, and Warren 2006)

Log-ratio (Rohr et al. 2010): Interestingly often used in paleo settings (at least that's what it currently looks like in my mind… Pires et al. (2020))

Matching (Rossberg et al. 2006): This one is more of a dynamic model (so BEF) and maybe beyond the scope of this work. I think there is value on only focusing on the 'static' models at this point (probably have said this before elsewhere but yeah)

PFIM (Shaw et al. 2024):

Trait-based (Caron et al. 2022):

Graph embedding (Strydom et al. 2022, 2023): At a high level graph embedding focuses on capturing the structural data of a network as opposed to a list of pairwise (*i.e.,* mechanistic) interactions. Here specifically the embedding is preformed on a known interaction network and captures information as to where species (nodes) are positioned in a network *e.g.,* are they basal prey species or top predators, similar to the log ratio model. In Strydom et al. (2022) the products of the embedding process are fed into a transfer learning framework for novel prediction…

I know tables are awful but in this case they may make more sense. Also I don't think I'm at the point where I can say that the table is complete/comprehensive but it getting there Not

sure about putting in some papers that have used the model - totes happy to drop those I think…

Table 1: Lets make a table that gives an overview of the different topology generators that we will look at. Here I take 'data-driven' to refer to the need for 'real world' data. This can probably be approached in a different way though maybe?

Model	Core Mechanism	Predicts	Specificity	Interaction	Data-driven
random	random	networks	species agnostic	binary	no
cascade	structural	networks	species agnostic	binary	no
niche	structural	networks	species agnostic	binary	no
nested hierarchical	structural	networks	species agnostic	binary	no
ADBM	mechanistic	interactions	energetics	quantitative	
log-ratio		interactions			
PFIM	mechanistic	interactions	trait based		
graph embedding	embedding	interactions	evolutionary	probabilistic	yes
trait model matching	mechanistic	interactions	trait based		yes

> Might be nice to have a little appendix/supp mat that breaks down the models in detail so that they are all in one place so that someone (grad student being told they need to build networks) some day can go and educate themselves with slightly lower effort. This will also be useful for me should I end up having to do some actual coding - think of this as step one in the pseudo code process.

2.3 Datasets used

Here I think we need to span a variety of domains, at minimum aquatic and terrestrial but maybe there should be a 'scale' element as well *i.e.,* a regional and local network. I think there is going to be a 'turning point' where structural will take over from mechanistic in terms of performance. More specifically at local scales bioenergetic constraints (and co-occurrence) may play a bigger role in structuring a network whereas at the metaweb level then mechanistic may make more (since by default its about who can potentially interact and obviously not constrained by real-world scenarios) *sensu* Caron et al. (2023). Although having said that I feel that contradicts the idea of backbones (*sensu* Bramon Mora (sp?) et al & Stouffer et al) But that might be where we get the idea of core *structure* vs something like linkage density.

So core things like trophic level/chain length will be conserved but connectance might not (I think I understand what I'm trying to say here)

I think we should also use the Dunne (I think) Cambrian (also think) network (I was correct and its this one Jennifer A. Dunne et al. (2008)). Because 1) it gives the paleo-centric methods their moment in the sun and 2) I think it also brings up the interesting question of can we use modern structure to predict past ones? Here one might expect a more mechanistic approach to shine.

Draw the other datasets from `Mangal` because they will be nicely formatted and essentially at point and shoot level

2.4 Comparing different models

For now the (still essentially pending) workflow/associated code can be found at the following repository BecksLab/topology_generators

1. Shortlist/finalise the different topo generators
2. collate/translate into `Julia`
    - *e.g.,* some models wil be in SpeciesInteractionNetworks.jl (new EcoNet); I know (parts of) the transfer learning stuff is and the niche model
    - others will need to be coded out (the more simpler models should be easier)
    - can also consider `R` but then it becomes a case of porting things left and right depending on how we decide to do the post analyses
3. Curate networks for the different datasets/scenarios we select - I feel like there might be some scenarios that we can't do all models for all datasets but maybe I'm being a pessimist.
    - Need to also think about where one might find the additional data for some of the models…
        - Body size: Herberstein et al. (2022) - Although maybe Andrew has strong thotsTM RE the one true body size database to rule them all…
        - Other trait sources: Wilman et al. (2014) and Jones et al. (2009)
        - This is where we'll get the paleo traits from if I'm correct Bambach, Bush, and Erwin (2007)
        - Phylogeny stuff: Upham, Esselstyn, and Jetz (2019) (what we used for TL but its only mammals…) but I'm sure there will be others
    - Also limitation of scope… *e.g.,* do we even dare to think about including plants/basal producers (see *e.g.,* Valdovinos et al. (2023))
    - Taxonomic harmonisation - something to think about and check
4. compare model performance based on the ideas currently listed in the results section.

5. Make a pretty picture that summarises things - maybe overlapping Venn circles that showcase which models do well in the different spheres/aspects of life

3 Results

How we want to compare and contrast. I think there won't be a 'winner' and thus we need to think of 'tests' that are going to measure performance in different situations/settings. With that in mind I think some valuable points to consider would be:

- Structural vs pairwise link predictions (graph vs node level)
  - % of links correctly retrieved
  - connectance
  - trophic level
  - generalism vs specialism
  - something related to false positives/negatives
  - intervality
- Data 'cost' (some methods might need a lot lot of supporting data vs something very light weight)
- I think it would be remiss to not also take into consideration computational cost
- something about the network output - I'm acknowledging my biases and saying that probabilistic (or *maybe* weighted) links are the way

Joel E. Cohen, Newman, and Steele (1985) actually tells us that the cascade model only really works for communities that range from 3-33 species… and Williams and Martinez (2008) also highlights how structural models really only work for small communities

> maybe we can put these into broader categories - if we do start doing the venn overlap thing. *E.g.,* local scale predictions, regional scale predictions, pairwise interactions, structural (energetics), computationally cheap, low cost data

3.1 Quantitative stuff

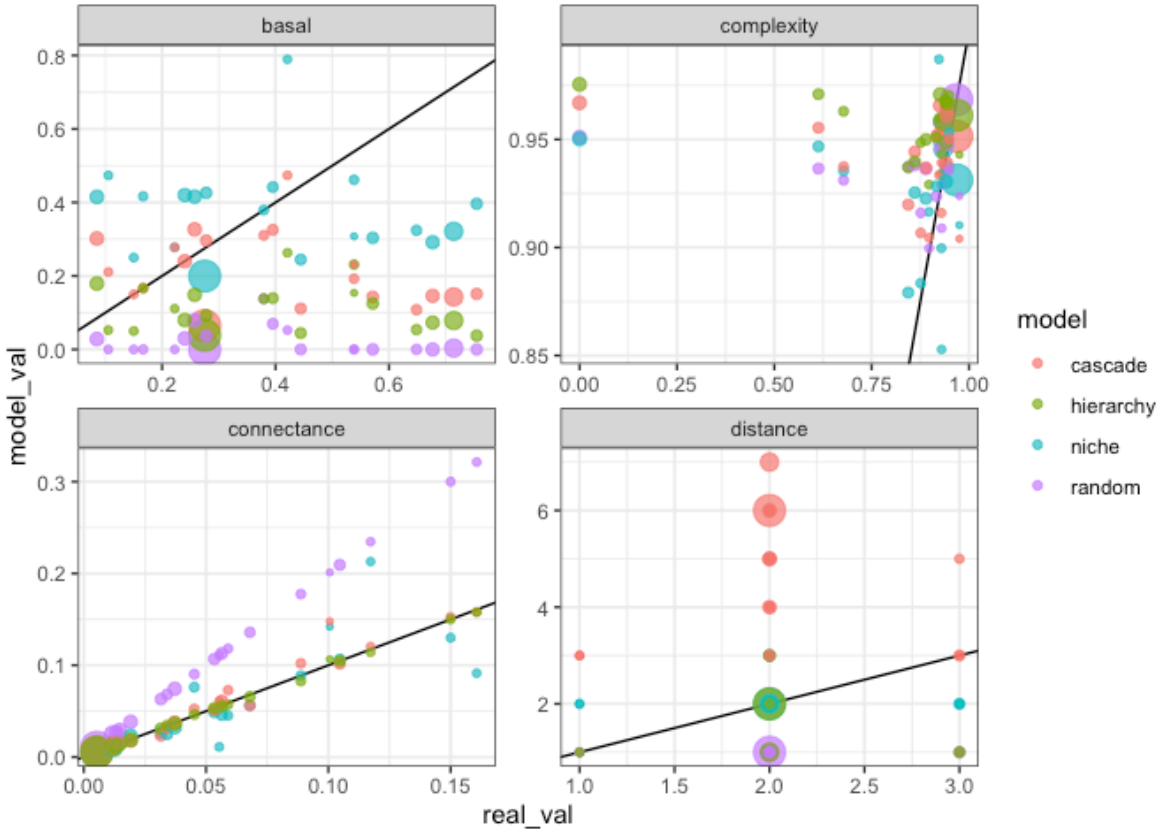

Figure 1: Real vs observed values for network summary statistics. Note here that 'basal' is calculated as the proportion of species that have a generality value of zero *i.e.*, are basal AFAIK

Source: Article Notebook

This is actually an awful way to try and summarise the data but rolling with it for now...

4 Discussion

I think a big take home will (hopefully) be how different approaches do better in different situations and so you as an end user need to take this into consideration and pick accordingly. I think Petchey et al. (2011) might have (and share) some thoughts on this (thanks Andrew). I feel like I need to look at Berlow, Brose, and Martinez (2008) but maybe not exactly in this context but vaguely adjacent.

An interesting thing to also think about (and arguably it will be addressed based on some of the other thoughts and ideas) is data dependant and data independent 'parametrisation' of the models…

I probably think about this point too much but a point of discussion that I think will be interesting to bring up the idea that if a model is missing a specific pairwise link but doing well at the structural level then when does it matter? I think this is covered with the whole node vs graph level performance but I kind of just want to bring it up here again because also one of those things that I think about a bit too much probably…

> Thinking very long term here and maybe a bit beyond the scope but also thinking about a multi- model approach? So in other words using one model to build an initial network but maybe a second one to constrain it a bit better. I blame this thought on the over-connected PFIM food webs…

References

Source: Article Notebook

Bambach, Richard K., Andrew M. Bush, and Douglas H. Erwin. 2007. "Autecology and the Filling of Ecospace: Key Metazoan Radiations." *Palaeontology* 50 (1): 1–22. https://doi.org/10.1111/j.1475-4983.2006.00611.x.

Beckerman, Andrew P., Owen L. Petchey, and Philip H. Warren. 2006. "Foraging Biology Predicts Food Web Complexity." *Proceedings of the National Academy of Sciences* 103 (37): 13745–49. https://doi.org/10.1073/pnas.0603039103.

Berlow, Eric L., Ulrich Brose, and Neo D. Martinez. 2008. "The 'Goldilocks Factor' in Food Webs." *Proceedings of the National Academy of Sciences* 105 (11): 4079–80. https://doi.org/10.1073/pnas.0800967105.

Caron, Dominique, Ulrich Brose, Miguel Lurgi, Guillaume Blanchet, Dominique Gravel, and Laura J. Pollock. 2023. "Trophic Interaction Models Predict Interactions Across Space, Not Food Webs." EcoEvoRxiv. https://doi.org/10.32942/X29K55.

Caron, Dominique, Luigi Maiorano, Wilfried Thuiller, and Laura J. Pollock. 2022. "Addressing the Eltonian Shortfall with Trait-Based Interaction Models." *Ecology Letters* 25 (4): 889–99. https://doi.org/10.1111/ele.13966.

Cattin, Marie-France, Louis-Félix Bersier, Carolin Banašek-Richter, Richard Baltensperger, and Jean-Pierre Gabriel. 2004. "Phylogenetic Constraints and Adaptation Explain Food-Web Structure." *Nature* 427 (6977): 835–39. https://doi.org/10.1038/nature02327.

Cohen, Joel E, Frederic Briand, and Charles Newman. 1990. *Community Food Webs: Data and Theory*. Biomathematics. Berlin Heidelberg: Springer-Verlag.

Cohen, Joel E., C. M. Newman, and John Hyslop Steele. 1985. "A Stochastic Theory of Community Food Webs I. Models and Aggregated Data." *Proceedings of the Royal Society*

of London. Series B. Biological Sciences 224 (1237): 421–48. https://doi.org/10.1098/rspb.1985.0042.

Dunne, Jennifer A. 2006. "The Network Structure of Food Webs." In *Ecological Networks: Linking Structure and Dynamics*, edited by Jennifer A Dunne and Mercedes Pascual, 27–86. Oxford University Press.

Dunne, Jennifer A., Richard J. Williams, Neo D. Martinez, Rachel A. Wood, and Douglas H. Erwin. 2008. "Compilation and Network Analyses of Cambrian Food Webs." *PLOS Biology* 6 (4): e102. https://doi.org/10.1371/journal.pbio.0060102.

Erdős, Paul, and Alfréd Rényi. 1959. "On Random Graphs I." *Publicationes Mathematicae*. https://doi.org/10.5486/PMD.1959.6.3-4.12.

Gravel, Dominique, Timothée Poisot, Camille Albouy, Laure Velez, and David Mouillot. 2013. "Inferring Food Web Structure from Predator–Prey Body Size Relationships." *Methods in Ecology and Evolution* 4 (11): 1083–90. https://doi.org/10.1111/2041-210X.12103.

Herberstein, Marie E., Donald James McLean, Elizabeth Lowe, Jonas O. Wolff, Md Kawsar Khan, Kaitlyn Smith, Andrew P. Allen, et al. 2022. "AnimalTraits - a Curated Animal Trait Database for Body Mass, Metabolic Rate and Brain Size." *Scientific Data* 9 (1): 265. https://doi.org/10.1038/s41597-022-01364-9.

Jones, Kate E., Jon Bielby, Marcel Cardillo, Susanne A. Fritz, Justin O'Dell, C. David L. Orme, Kamran Safi, et al. 2009. "PanTHERIA: A Species-Level Database of Life History, Ecology, and Geography of Extant and Recently Extinct Mammals." *Ecology* 90 (9): 2648–48. https://doi.org/10.1890/08-1494.1.

Jordano, Pedro. 2016a. "Chasing Ecological Interactions." *PLOS Biology* 14 (9): e1002559. https://doi.org/10.1371/journal.pbio.1002559.

———. 2016b. "Sampling Networks of Ecological Interactions." *Functional Ecology*, September. https://doi.org/10.1111/1365-2435.12763.

Petchey, Owen L., Andrew P. Beckerman, Jens O. Riede, and Philip H. Warren. 2008. "Size, Foraging, and Food Web Structure." *Proceedings of the National Academy of Sciences* 105 (11): 4191–96. https://doi.org/10.1073/pnas.0710672105.

———. 2011. "Fit, Efficiency, and Biology: Some Thoughts on Judging Food Web Models." *Journal of Theoretical Biology* 279 (1): 169–71. https://doi.org/10.1016/j.jtbi.2011.03.019.

Pires, Mathias M., Diego Rindel, Bruno Moscardi, Livia R. Cruz, Paulo R. Guimarães, Sergio F. dos Reis, and S. Ivan Perez. 2020. "Before, During and After Megafaunal Extinctions: Human Impact on Pleistocene-Holocene Trophic Networks in South Patagonia." *Quaternary Science Reviews* 250 (December): 106696. https://doi.org/10.1016/j.quascirev.2020.106696.

Poisot, Timothée, Gabriel Bergeron, Kevin Cazelles, Tad Dallas, Dominique Gravel, Andrew MacDonald, Benjamin Mercier, Clément Violet, and Steve Vissault. 2021. "Global Knowledge Gaps in Species Interaction Networks Data." *Journal of Biogeography* n/a (n/a). https://doi.org/10.1111/jbi.14127.

Poisot, Timothée, Daniel B. Stouffer, and Dominique Gravel. 2015. "Beyond Species: Why Ecological Interaction Networks Vary Through Space and Time." *Oikos* 124 (3): 243–51. https://doi.org/10.1111/oik.01719.

Rohr, Rudolf Philippe, Heike Scherer, Patrik Kehrli, Christian Mazza, and Louis-Félix Bersier.

2010. "Modeling Food Webs: Exploring Unexplained Structure Using Latent Traits." *The American Naturalist* 176 (2): 170–77. https://doi.org/10.1086/653667.

Rossberg, A. G., H. Matsuda, T. Amemiya, and K. Itoh. 2006. "Food Webs: Experts Consuming Families of Experts." *Journal of Theoretical Biology* 241 (3): 552–63. https://doi.org/10.1016/j.jtbi.2005.12.021.

Shaw, Jack O., Alexander M. Dunhill, Andrew P. Beckerman, Jennifer A. Dunne, and Pincelli M. Hull. 2024. "A Framework for Reconstructing Ancient Food Webs Using Functional Trait Data." bioRxiv. https://doi.org/10.1101/2024.01.30.578036.

Stouffer, D. B., J. Camacho, R. Guimerà, C. A. Ng, and L. A. Nunes Amaral. 2005. "Quantitative Patterns in the Structure of Model and Empirical Food Webs." *Ecology* 86 (5): 1301–11. https://doi.org/10.1890/04-0957.

Strydom, Tanya, Salomé Bouskila, Francis Banville, Ceres Barros, Dominique Caron, Maxwell J. Farrell, Marie-Josée Fortin, et al. 2022. "Food Web Reconstruction Through Phylogenetic Transfer of Low-Rank Network Representation." *Methods in Ecology and Evolution* 13 (12): 2838–49. https://doi.org/10.1111/2041-210X.13835.

Strydom, Tanya, Salomé Bouskila, Francis Banville, Ceres Barros, Dominique Caron, Maxwell J. Farrell, Marie-Josée Fortin, et al. 2023. "Graph Embedding and Transfer Learning Can Help Predict Potential Species Interaction Networks Despite Data Limitations." *Methods in Ecology and Evolution* 14 (12): 2917–30. https://doi.org/10.1111/2041-210X.14228.

Strydom, Tanya, Michael D. Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau, Philippe Desjardins-Proulx, Norma R. Forero-Muñoz, et al. 2021. "A Roadmap Towards Predicting Species Interaction Networks (Across Space and Time)." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1837): 20210063. https://doi.org/10.1098/rstb.2021.0063.

Upham, Nathan S., Jacob A. Esselstyn, and Walter Jetz. 2019. "Inferring the Mammal Tree: Species-level Sets of Phylogenies for Questions in Ecology, Evolution, and Conservation." *PLOS Biology* 17 (12): e3000494. https://doi.org/10.1371/journal.pbio.3000494.

Valdovinos, Fernanda S., Kayla R. S. Hale, Sabine Dritz, Paul R. Glaum, Kevin S. McCann, Sophia M. Simon, Elisa Thébault, William C. Wetzel, Kate L. Wootton, and Justin D. Yeakel. 2023. "A Bioenergetic Framework for Aboveground Terrestrial Food Webs." *Trends in Ecology & Evolution* 38 (3): 301–12. https://doi.org/10.1016/j.tree.2022.11.004.

Williams, Richard J., and Neo D. Martinez. 2000. "Simple Rules Yield Complex Food Webs." *Nature* 404 (6774): 180–83. https://doi.org/10.1038/35004572.

———. 2008. "Success and Its Limits Among Structural Models of Complex Food Webs." *Journal of Animal Ecology* 77 (3): 512–19. https://doi.org/10.1111/j.1365-2656.2008.01362.x.

Wilman, Hamish, Jonathan Belmaker, Jennifer Simpson, Carolina de la Rosa, Marcelo M. Rivadeneira, and Walter Jetz. 2014. "EltonTraits 1.0: Species-level Foraging Attributes of the World's Birds and Mammals." *Ecology* 95 (7): 2027–27. https://doi.org/10.1890/13-1917.1.

Yeakel, Justin D., Mathias M. Pires, Lars Rudolf, Nathaniel J. Dominy, Paul L. Koch, Paulo R. Guimarães, and Thilo Gross. 2014. "Collapse of an Ecological Network in Ancient Egypt." *PNAS* 111 (40): 14472–77. https://doi.org/10.1073/pnas.1408471111.