T is for Topology

Tanya Strydom[1], Andrew P. Beckerman[1]

[1]University of Sheffield,

Corresponding author: Tanya Strydom, `t.strydom@sheffield.ac.uk`

Abstract

Pending…

Plain Language Summary

We want to know a bit more about the different network topology generators (predict tools) and how they differ - *i.e.,* their strengths and weaknesses

1 Introduction

The standard run of the mill that we cannot always feasibly construct networks because 1. hard, 2. time (yay dinosaurs, but also the future and impending doom I guess), and 3. probably something else meaningful that's just slipping my mind at the moment. Some of the usual culprits will come in here like: Jordano (2016b); (Jordano, 2016a); Poisot et al. (2021)

> I'm beginning to think that the core discursive points should revolve around 'old school' vs 'new age' methods but in a way that sort of says in a world of new age tools are the the old school methods able to hole their own. But at the same time making it abundantly clear that there is a stack of context dependence to keep in mind before even thinking about drawing conclusions.

> TODO: standardise language between a topology generator and a topology predictor… I guess they can both be considered models but I'm not quite sure…

1.1 Philosophical contemplation when constructing interaction networks

1.1.1 Why do we want to construct an interaction network?

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 of the spectrum sits the researcher who is interested in generating a set of ecologically plausible but not necessarily realised 'in the field' for the purpose of running further simulations (*e.g.,* extinction sim REF? **TODO**) or understanding some higher-level process (*e.g.,* energetics REF **TODO**). This researcher is contrasted by one that is interested in constructing real-world, location specific interactions networks in lieu of having access to data generated in the field (see Strydom et al. (2021) for a discussion on this). Of course these two categories are not distinct, mutually exclusive, groups but can rather be viewed as operating on a gradient ranging from a need for generality (*something*) to a need for specificity (local-level predictions) when it comes to the quality(?) of the network that is constructed by a specific tool. Of course this research need would also reflected in the model development process itself and thus the idea of what a 'good enough' constructed network will be in the context of assessing the performance of a specific model.

Joel E. Cohen et al. (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 & 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 function at a specific 'network level'.

1.1.2 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 et al. (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
2.1 Overview of models
2.1.1 Structural models
Random model (Erdős & 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)

Cascade model (Joel E. Cohen et al., 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 et al., 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 model (Williams & 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 & Martinez, 2008).

Nested hierarchy model (Cattin et al., 2004):

2.1.2 Mechanistic models

Allometric diet breadth model (ADBM) (Petchey et al., 2008):

Log-ratio (Rohr et al., 2010): Interestingly often used in paleo settings (at least that's what it currently looks like in my mind… (*e.g.,* Yeakel et al. (2014), 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

Model	Core Mechanism	End product	Specificity	Interaction	Data
random	random	network	species agnostic	binary	
cascade	structural	network	species agnostic	binary	
niche	structural	network	species agnostic	binary	
nested hierarchical	structural	network	species agnostic	binary	
ADBM	mechanistic		energetics	quantitative	
log-ratio					
PFIM	mechanistic	metaweb	trait based		
graph embedding	embedding	metaweb	evolutionary	probabilistic	
trait model	mechanistic	metaweb	trait based		
matching					

Model	Core Mechanism	End product	Specificity	Interaction	Data

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.

2.2 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

2.3 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 et al. (2007)
     - Phylogeny stuff: Upham et al. (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 et al. (1985) actually tells us that the cascade model only really works for communities that range from 3-33 species… and Williams & 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
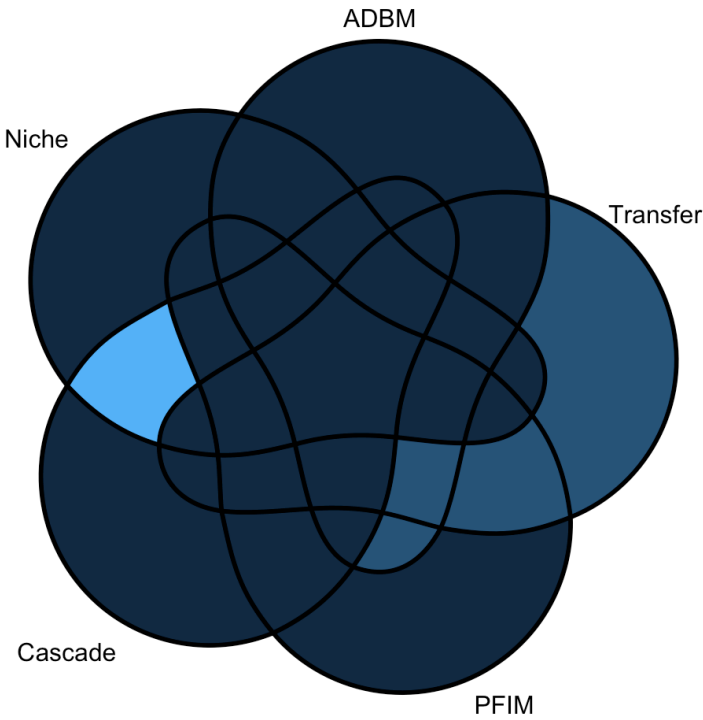
214 3.1 Qualitative stuff

Figure 1: Venn diagram for qualitative analysis/overview of the fancy maths things
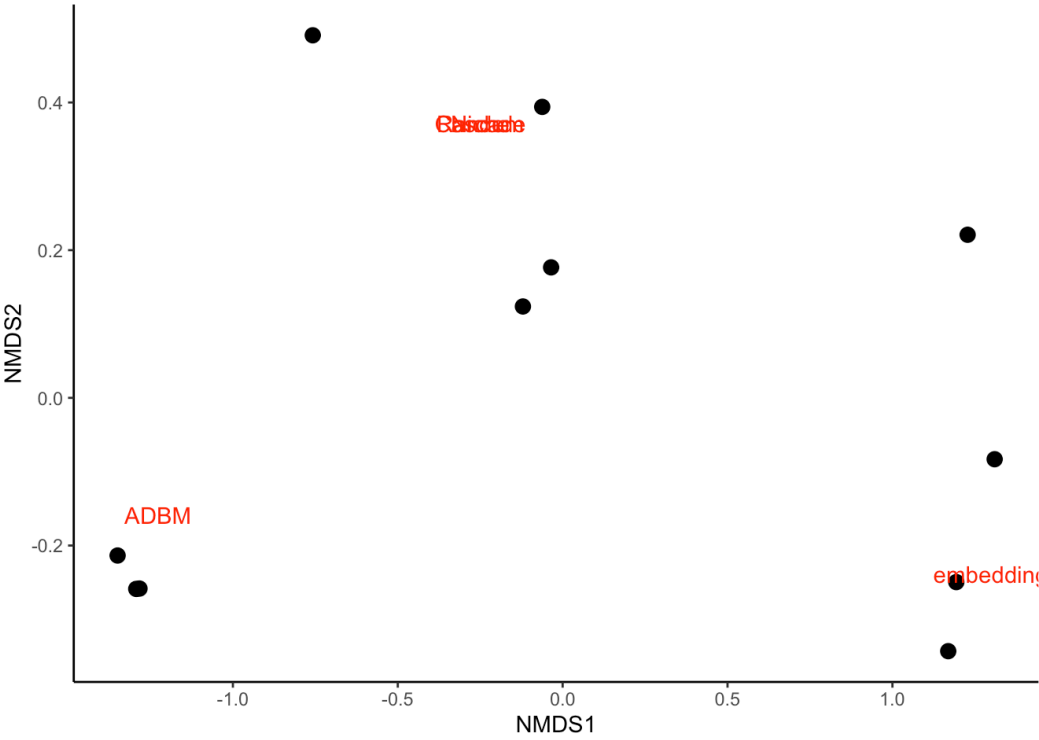
215 Source: Article Notebook

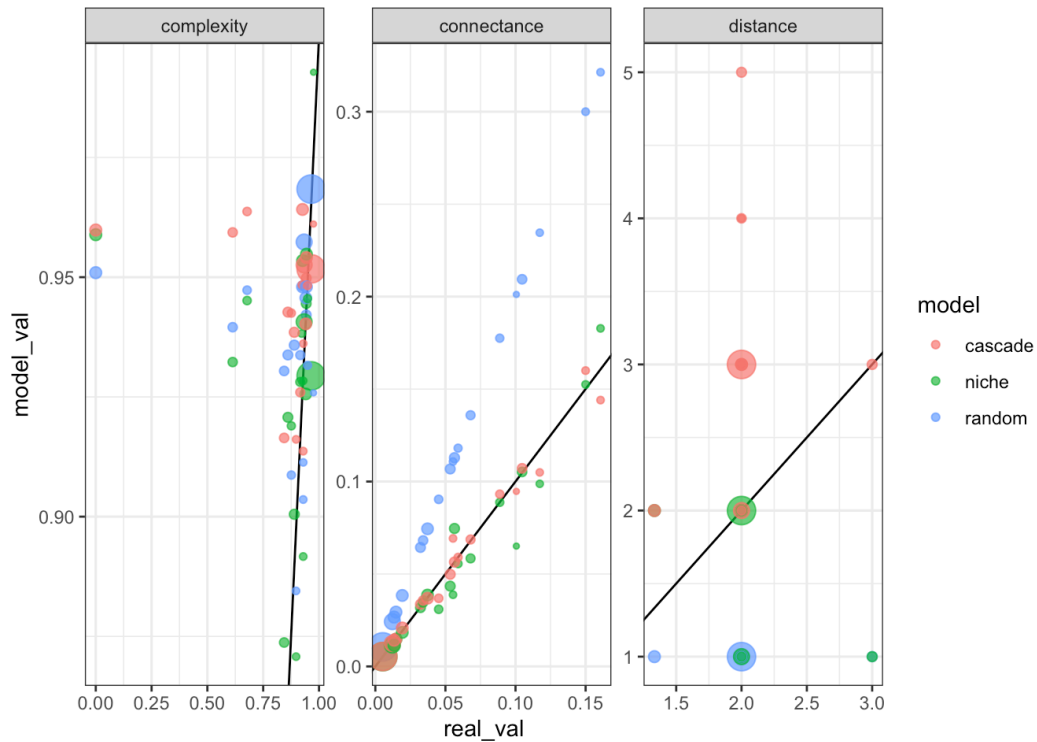216 See also

Figure 2: NMDS of things

3.2 Quantitative stuff

Figure 3: Real vs observed values for network summary statistics

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 et al. (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

239   to build an initial network but maybe a second one to constrain it a bit

240   better. I blame this thought on the over-connected PFIM food webs…

References

242   Source: Article Notebook

243 Bambach, R. K., Bush, A. M., & Erwin, D. H. (2007). Autecology and the Filling
244   of Ecospace: Key Metazoan Radiations. *Palaeontology*, *50*(1), 1–22. https://
245   doi.org/10.1111/j.1475-4983.2006.00611.x
246 Berlow, E. L., Brose, U., & Martinez, N. D. (2008). The "Goldilocks factor" in food
247   webs. *Proceedings of the National Academy of Sciences*, *105*(11), 4079–4080.
248   https://doi.org/10.1073/pnas.0800967105
249 Caron, D., Maiorano, L., Thuiller, W., & Pollock, L. J. (2022). Addressing the
250   Eltonian shortfall with trait-based interaction models. *Ecology Letters*, *25*(4),
251   889–899. https://doi.org/10.1111/ele.13966
252 Caron, D., Brose, U., Lurgi, M., Blanchet, G., Gravel, D., & Pollock, L. J. (2023,
253   May). Trophic interaction models predict interactions across space, not food
254   webs. EcoEvoRxiv. https://doi.org/10.32942/X29K55
255 Cattin, M.-F., Bersier, L.-F., Banašek-Richter, C., Baltensperger, R., & Gabriel,
256   J.-P. (2004). Phylogenetic constraints and adaptation explain food-web structure.
257   *Nature*, *427*(6977), 835–839. https://doi.org/10.1038/nature02327
258 Cohen, Joel E., Newman, C. M., & Steele, J. H. (1985). A stochastic theory of
259   community food webs I. Models and aggregated data. *Proceedings of the Royal
260   Society of London. Series B. Biological Sciences*, *224*(1237), 421–448. https://
261   doi.org/10.1098/rspb.1985.0042
262 Cohen, Joel E., Briand, F., & Newman, C. (1990). *Community Food Webs: Data
263   and Theory.* Berlin Heidelberg: Springer-Verlag.
264 Dunne, Jennifer A. (2006). The Network Structure of Food Webs. In Jennifer A.
265   Dunne & M. Pascual (Eds.), *Ecological networks: Linking structure and dynam-
266   ics* (pp. 27–86). Oxford University Press.
267 Dunne, Jennifer A., Williams, R. J., Martinez, N. D., Wood, R. A., & Erwin, D.
268   H. (2008). Compilation and Network Analyses of Cambrian Food Webs. *PLOS
269   Biology*, *6*(4), e102. https://doi.org/10.1371/journal.pbio.0060102
270 Erdős, P., & Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae*.
271   https://doi.org/10.5486/PMD.1959.6.3-4.12
272 Herberstein, M. E., McLean, D. J., Lowe, E., Wolff, J. O., Khan, M. K., Smith, K.,
273   et al. (2022). AnimalTraits - a curated animal trait database for body mass,
274   metabolic rate and brain size. *Scientific Data*, *9*(1), 265. https://doi.org/
275   10.1038/s41597-022-01364-9
276 Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., et
277   al. (2009). PanTHERIA: A species-level database of life history, ecology, and
278   geography of extant and recently extinct mammals. *Ecology*, *90*(9), 2648–2648.
279   https://doi.org/10.1890/08-1494.1
280 Jordano, P. (2016a). Chasing Ecological Interactions. *PLOS Biology*, *14*(9), e1002559.
281   https://doi.org/10.1371/journal.pbio.1002559
282 Jordano, P. (2016b). Sampling networks of ecological interactions. *Functional Ecol-
283   ogy*. https://doi.org/10.1111/1365-2435.12763
284 Petchey, O. L., Beckerman, A. P., Riede, J. O., & Warren, P. H. (2008). Size, for-
285   aging, and food web structure. *Proceedings of the National Academy of Sciences*,
286   *105*(11), 4191–4196. https://doi.org/10.1073/pnas.0710672105
287 Petchey, O. L., Beckerman, A. P., Riede, J. O., & Warren, P. H. (2011). Fit,
288   efficiency, and biology: Some thoughts on judging food web models. *Jour-
289   nal of Theoretical Biology*, *279*(1), 169–171. https://doi.org/10.1016/
290   j.jtbi.2011.03.019

Pires, M. M., Rindel, D., Moscardi, B., Cruz, L. R., Guimarães, P. R., dos Reis, S. F., & Perez, S. I. (2020). Before, during and after megafaunal extinctions: Human impact on Pleistocene-Holocene trophic networks in South Patagonia. *Quaternary Science Reviews*, *250*, 106696. https://doi.org/10.1016/j.quascirev.2020.106696

Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, *124*(3), 243–251. https://doi.org/10.1111/oik.01719

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., et al. (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography*, *n/a*(n/a). https://doi.org/10.1111/jbi.14127

Rohr, R. P., Scherer, H., Kehrli, P., Mazza, C., & Bersier, L.-F. (2010). Modeling Food Webs: Exploring Unexplained Structure Using Latent Traits. *The American Naturalist*, *176*(2), 170–177. https://doi.org/10.1086/653667

Rossberg, A. G., Matsuda, H., Amemiya, T., & Itoh, K. (2006). Food webs: Experts consuming families of experts. *Journal of Theoretical Biology*, *241*(3), 552–563. https://doi.org/10.1016/j.jtbi.2005.12.021

Shaw, J. O., Dunhill, A. M., Beckerman, A. P., Dunne, J. A., & Hull, P. M. (2024, January). A framework for reconstructing ancient food webs using functional trait data. bioRxiv. https://doi.org/10.1101/2024.01.30.578036

Stouffer, D. B., Camacho, J., Guimerà, R., Ng, C. A., & Nunes Amaral, L. A. (2005). Quantitative Patterns in the Structure of Model and Empirical Food Webs. *Ecology*, *86*(5), 1301–1311. https://doi.org/10.1890/04-0957

Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., 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

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., et al. (2022). Food web reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution*, *13*(12), 2838–2849. https://doi.org/10.1111/2041-210X.13835

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., 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), 2917–2930. https://doi.org/10.1111/2041-210X.14228

Upham, N. S., Esselstyn, J. A., & Jetz, W. (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, F. S., Hale, K. R. S., Dritz, S., Glaum, P. R., McCann, K. S., Simon, S. M., et al. (2023). A bioenergetic framework for aboveground terrestrial food webs. *Trends in Ecology & Evolution*, *38*(3), 301–312. https://doi.org/10.1016/j.tree.2022.11.004

Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, *404*(6774), 180–183. https://doi.org/10.1038/35004572

Williams, R. J., & Martinez, N. D. (2008). Success and its limits among structural models of complex food webs. *Journal of Animal Ecology*, *77*(3), 512–519. https://doi.org/10.1111/j.1365-2656.2008.01362.x

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, *95*(7), 2027–2027. https://doi.org/10.1890/13-1917.1

344 Yeakel, J. D., Pires, M. M., Rudolf, L., Dominy, N. J., Koch, P. L., Guimarães,
345   P. R., & Gross, T. (2014). Collapse of an ecological network in Ancient Egypt.
346   *PNAS*, *111*(40), 14472–14477. https://doi.org/10.1073/pnas.1408471111