Navigating food web prediction; assumptions, rationale, and methods

Tanya Strydom [1]; Jennifer A. Dunne [2]; Timothée Poisot [3,4]; Andrew P. Beckerman [1]

Abstract: TODO

Keywords: food web, network construction, scientific ignorance

At the heart of modern biodiversity science are a set of concepts about biodiversity, community structure, productivity, and asynchrony, and how they define the stability, resilience, and dynamics of complex communities. The use of species interaction networks provides a powerful abstraction that one can use to help quantify, conceptualise, and understand these concepts. However, network ecology has its own nuance and idiosyncrasies that not only provide a barrier to entry but causes dissonance even within the field (Dormann, 2023). This is perhaps particularly pervasive within the space of network prediction…

One of the fundamental challenges that we are faced when working with and studying interaction networks (and, within the context of this manuscript, specifically food webs) is that there is a scarcity of 'real world' interaction data (Hortal et al., 2015; Poisot et al., 2021). The difficulty of recording interactions in the field (Jordano, 2016a, 2016b) has necessitated that researchers find and develop alternative means to construct and build food webs using **models** (Morales-Castilla et al., 2015; Strydom et al., 2021). Over the past decade, there has been a proliferation of tools and processes for characterising food webs, these models span a wide range of philosophies that rely on different approaches, data, and definitions, which ultimately determine how the food web is constructed and coded. Although the development of these different models have carved out the path for constructing either synthetic, ecologically plausible networks (Poisot, Gravel, et al., 2016), or providing 'first draft' networks that can be utilised in real world settings (Strydom et al., 2022) we are still lacking in discussions that are explicitly comparing and contrasting how the way one chooses to approach the task of constructing a food web is introducing (and ultimately embedding) specific assumptions and hypotheses (Petchey et al., 2008). Most attempts that focus on comparing and contrasting models are focused on the same group of **model families** (Pichler et al., 2020; Williams & Martinez, 2008) and only benchmark the different models as opposed to contextualising them within the bigger framework of understanding the data needs of the different models, as well as how the resulting network is defined and structured. As food webs become a more integrated part of some of the broader fields of ecology (Bhatia et al., 2023; Thuiller et al., 2024) it is critical that we review these different model families as a whole (not only in isolation), and move away from simply benchmarking the performance of these different model families. This is important because different models impose different constraints upon themselves and will not only delimit and dictate the potential questions one will be able to ask (Petchey et al., 2011) but also determine the appropriate research setting for which the model (and resulting network) can be used. For example the use of 'structural food webs' are useful for developing additional theory such as re-wiring of networks (Staniczenko et al., 2010) but would be meaningless if one's intention is to produce a location-specific network [do we need an *e.g.,* ref??]. This will allow us to ensure the right models are being used to answer the right questions, particularly within the context of trying to accelerate cross-cutting research in the face of global change.

When navigating the seas of using and constructing food webs the researcher needs to be able to clearly articulate and define the parameters that are used to define their food web(s) of interest. This will aid them in being able to select the correct model to help them to reach their goal. In order to be able to make informed decisions it is important that one has a strong grasp of exactly what it means to 'code'/define a food web (Section 1), a clear understanding of why one wants to predict a food web (Section 2), and ultimately one needs to be able to asses and evaluate which model family is going to best match up with the goal of network prediction (Section 3). Here we specifically aim to look at not look at only the performance of the different models but also initiate a (thus far lacking) discussion around how the interplay between the language used to define networks and the underlying theory/philosophy should also be a part of the broader discussion when it comes to the task of 'model selection'.

[Figure 1 about here.]

1 The anatomy of a food web

Defining a food web seems simple, it is the representation of the interactions (edges) between species (nodes), however the definition of 'edges' and 'nodes', as well as the scale at which they are aggregated can take many forms. As highlighted in Poisot, Stouffer, et al. (2016) networks can be constructed at the population (the links among individuals), community (the links between species), or metacommunity (fluxes between locations) level. Even if one were to limit their scope to thinking of interaction networks only in terms of food webs at the community-level there are still many ways to define the various components of the network Panel A of 1, one needs to understand the different intentions/assumptions that are made when a food web is constructed. Although the main intention of a food web is to capture and represent the feeding links between species there are many ways to define the nodes (*e.g.,* species or taxonomic group), edges (*e.g.,* **potential** or **realised feeding links**), the magnitude of the edges (*e.g.,* binary vs probabilistic), and even how the network itself is delimited (does it represent an aggregation of interactions over time?).

1.1 How do we define a node?

Although this may seem an elementary question in the context of food webs — a node *should* represent a (taxonomic) species, the reality is that nodes can often represent an aggregation of different species - so called 'trophic species' or segregation of species by life stages. Representing nodes as non-taxonomic species can be useful in certain contexts (Williams & Martinez, 2000) and in cases where the adult and larval stages of a species have different diets it may make ecological sense (Clegg et al., 2018) meaning that it is not uncommon that networks often have nodes that have different definitions of a 'species' *e.g.* consisting of both taxonomic and trophic species. Practical implications of how we are aggregating the nodes is that the resolution may not always be 'pixel perfect' *i.e.,* we may be unable to assess the co-extinction risk of a species pair, however there is value in having nodes that represent an aggregation of species, as these convey a much more general overview of how the links are distributed within the community.

1.2 What is meant by an edge?

As discussed earlier there are many ways to define the links between species — even feeding links. At its core links within food webs can be thought of as a representation of either the flow of a resource [ref], realised (Pringle, 2020) or potential (Dunne, 2006) feeding links, or energy transfer and material flow (Lindeman, 1942). How we specify links will influence the resulting structure of the network - and the inferences we will make thereof. For example taking a food web that consists of links representing *potential* feeding links between species will be meaningless if you are interested in understanding *e.g.,* the flow of energy through the system as the links within the network are over overrepresented. In addition to the various ways of defining the links between species pairs there are also a myriad of ways in which the links themselves can be quantified. Links between species are often treated as being present or absent (*i.e.,* binary) but it is also possible to use probabilities (which quantifies how likely an interaction is to occur, Poisot, Cirtwill, et al., 2016) or continuous measurements (which quantifies the strength of of an interaction, Berlow et al., 2004). Moving away from a purely binary way of representing allows us to quantify a level of (un)certainty of our knowledge of interactions (*i.e.,* moving from being able to ask if are they occurring to quantifying how likely they are to occur) does add an additional level of 'complexity' to the construction and interpretation of networks, but ultimately it allows us to capture more information at different scales (Banville, in prep).

1.3 Putting the parts together; what does it mean?

The ingredients one uses to construct networks from nodes and edges generates a unique representation of the mechanisms (see Box 1 - Mechanisms that determine feeding links) that allow inference and reasoning about the structure, aspects of dynamics (*e.g.,* stability), and potentially the function of communities (*e.g.,* flux). It is thus beneficial to keep in mind that in the process of 'codifying' a network one is already embedding some sort of hypothesis as to the nature of the feeding links between species (Brimacombe et al., 2023; Proulx et al., 2005). Here it may be meaningful to contextualise the different 'types' of food webs within the larger research programmes (or even practical needs) that have been driving the construction of them.

Box 1 - Mechanisms that determine feeding links

There are many ideas as to what are the underlying mechanisms that determine the links between species. The way one chooses to encode a network will most likely also be reflective of (or only be able to encapsulate) one or a few of the different mechanisms. There is probably even an argument to be had that depending on how we define a network we will probably expect some of the 'hypotheses' of the different mechanisms to hold. *e.g.,* I think most people will agree that the feasibility of interactions between specific species pairs is not random (there needs to be some sort of trait/form complementarity) but how/if they interact within the environment (*i.e.,* the realisation of the interactions) *might* as well be (also probably even more relevant if one thinks about/works with trophic species…)

Proximity

We are co-occurring in space and in time and thus we can interact (Barberán et al., 2012)

Mass-effect

Our (respective and instantaneous) abundance in that time and space is going to influence how we interact. *Sensu* Hubbell (2001) Neutral Theory

Complementarity

We have a set of 'traits' that means we can interact including:

- You as a prey item fit in my gob (I can eat you, ~~even if its small bites~~) [ref]
- You as a prey item are energetically 'worth it' ~~and allowing me to hit all the right macros~~ [ref foraging ecology]
- As a predator I have the required traits that allow me to ~~kill~~ unalive and eat you (*sensu* forbidden links Jordano, 2016b)
- As predator and prey we have been co-occurring for a long time and I have found ways to eat you (trying to capture the idea of evolutionary time)

'Structural'

The 'energy budget' for the environment means that only y links are possible between us x number of species and so our interactions reflect that. Or is it more the only way we can all access the energy resource is by arranging ourselves into trophic units…

None

We are therefore we interact. This is random.

2 Why do we want to predict food webs?

As discussed in Section 1 there are many ways to define a food web, meaning that there are equally as many reasons one might be interested in predicting a food web. However we may think of two primary drivers for wanting to predict networks (Panel B Figure 1), namely an interest in generating a set of ecologically plausible networks (*i.e.,* being able to describe networks using a model) or being able to recover (predict) location specific, 'realised', interactions for a specific species community (*i.e.,* being able to predict/infer the interactions between species). Of course these two categories are not distinct, mutually exclusive, groups but can rather be viewed as operating on a continuum ranging from a need for generality (*i.e.,* creating a network that, when taken in aggregate, the distribution of links (interactions) between nodes (species) are ecologically plausible) to a need for specificity (*i.e.,* local-level predictions between specific species pairs). Although the ability to predict 'real-world' interactions (and the resulting food webs) can have more intuitive 'real world' applications *e.g.,* being able to 'recover' food webs that have since gone extinct (Dunne et al., 2008; Yeakel et al., 2014), using pairwise interactions to understand species distributions (Pollock et al., 2014) or even co-extinction risk (Dunn et al., 2009), a more structural approach to network construction affords one an opportunity to interrogate some of the more high-level mechanisms that are structuring networks (Box 1).

It is perhaps more important that when one is talking about 'why' they want to predict networks to articulate exactly what anatomical part of the food web we are interested in scrutinising.

3 How do we predict food webs?

Selecting a model for the task of network prediction should come down to two things; what *aspect* of a food web one is interested in predicting, and what data are available, necessary, and sufficient. As shown in panel B of Figure 1 the interest in a network is (usually) at either the 'structural' or 'interaction' level and the development of models for the task of network prediction often focus on high fidelity (performance) at one of these scales. With this in mind it is beneficial to think of the different model families relative to these two different goals; here we refer to models that are used to predict the structure of a network as **topology generators** and models developed to infer the interactions for a given species pool as **interaction predictors**. It is meaningful to make this distinction because although it is possible to construct a food web given using an *interaction predictor* the models themselves lack any sort of parametrisation of the network structure and so the resulting network is a poor reflection of the actual network structure (Caron et al., 2024). This is primarily because *interaction predictors* are models that evaluate the feasibility of an interaction between species pairs and not in the context of feasibility at the community level. Models themselves are a reflection of the different goals and intentions of the research program from which they are developed and are often 'described' by a specific mechanism that will determine the resulting structure or interactions (Box 1). Models such as the niche (Williams & Martinez, 2000) or cascade (Cohen et al., 1990) were developed with the intent of being used to understand the *structural* aspects of food webs, specifically how links are distributed amongst species in the community, whereas bayesian (Cirtwill et al., 2019) or trait hierarchy (Shaw et al., 2024) models have been developed on the basis that the traits of a species are the underlying mechanism in determining the feasibility of interactions (*i.e.,* species a has the capacity to eat species b). Along with predicting different anatomical parts of a food web the different models have varying degrees of data that are needed to 'parametrise' the network. Once these two limitations are assessed and addressed it is then possible to select the model (or model family) that will best be able to capture food web feature that the researcher is most interested in (see Box 2 - Assessing model outputs). It is thus clear that (realistically) there will probably never be a 'best fit' tool that is able to construct a food web that will span the entire range of needs, and rather the responsibility lies with the researcher to be aware of not only the underlying philosophy of the specific toolset (as this could have knock-on effects when using those networks for downstream analyses/simulations; pers. comms. Beckerman, 2024), but also how well the tool is able to retrieve the specific network or interaction properties that is of interest.

> In order for a model to formalise a 'complete' food web it is necessary to formalise two aspects of the network, 'who eats whom' (to determine the links between nodes) as well as the structure of the network (to limit the distribution of links), however most models are inclined to focus on one of the two aspects panel B of 1.

> Crucially most topology generators lack some key data on the interaction between species (this can be because of how the model itself defines species or the way in which links are assigned in the network) and interaction predictors lack some sort of parametrisation of network structure (just because two species can interact it does not mean that they will, Poisot et al., 2015).

> 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).

3.1 Model families

As there are many food web models to choose from it is perhaps useful to think about the models in terms of model families, a summary of these families is presented in Table 1 and along with Figure 2 highlights the differences and similarities of the philosophies and assumptions that determine a network. A more extensive overview of the different models that fall with in the different model families can be found in SuppMat 1 and for a more detailed breakdown of the different 'traits' of the model families refer to SuppMat 2.

Table 1: A summary of the different families of tools that can be used to generate food webs, this includes a brief description of the underlying philosophy of the family as well as how the different elements (nodes and edges) of the generated network represents.

Model family	Theory	Network predicted	Nodes represent	Links represent	Interaction	Key reference
null	Links are randomly distributed within a network	structural	agnostic	feeding links	binary	
neutral	Network structure is random, but species abundance determines links between nodes	structural	species	feeding links	binary	
resource	Networks are interval, species can be ordered on a 'niche axis'	structural	trophic species	subdivision of resource	binary	Williams & Martinez (2008)
generative	Networks are determined by their structural features	structural	agnostic	links	binary	

Model family	Theory	Network predicted	Nodes represent	Links represent	Interaction	Key reference
energetic	Interactions are determined by foraging theory (feeding links)	interaction	species	feeding links	quantitative	
graph embedding	Interactions can be predicted from the latent traits of networks	interaction	species	potential feeding links	probabilistic	Strydom et al. (2023)
trait matching	Interactions can be inferred by a mechanistic frame-work/relationships	interaction	species	feeding links	binary	Morales-Castilla et al. (2015)
binary classifiers	Interactions can be predicted by learning the relationship between interactions and ecologically relevant predictors	interaction	species	feeding links	binary	Pichler et al. (2020)

Model family	Theory	Network predicted	Nodes represent	Links represent	Interaction	Key reference
expert knowledge	‘Boots on the ground’ ecological knowledge and observations	interaction	species	feeding links	binary	
data scavenging	Webscraping to create networks from online databases	interaction	species	feeding links	binary	Poisot, Gravel, et al. (2016) (f you squint?)
co-occurrence	co-occurrence patterns arise from interactions so we can use these patterns to reverse engineer the interactions	co-occurrence patterns	species	association links	binary	

157 [Figure 2 about here.]

Box 2 - Assessing model outputs

Although understanding the underlying philosophy of the different model families is beneficial it is also important to understand in what situations the different families are likely to preform well or poorly. When we are assessing the performance of the different model families it is beneficial to think of benchmarking these assessments based on a broader basis than just its ability to correctly recover network structure or pairwise interactions. When thinking about how to benchmark models it is perhaps beneficial to take a step back and once again assess what are the needs of the researcher (Section 2) and linking this back to what aspects of the network (Section 1) are of importance and assess the performance of a model within those parameters.

Benchmarking

Benchmarking how well a model is doing to capture the desired elements of a network is also a task that required some thought and contemplation. Even if we think about the predicting the structure of a network it is possible that two networks may have the same number of nodes and links but that those links may be distributed in very different ways. Thus it is important to think critically about the suite of summary statistics that are used to assess a model, since there is no one 'silver bullet' summary statistic that will be able to assess if a model is able to fully replicate an empirical network (Allesina et al., 2008). One of the main challenges when assessing the ability to retrieve pairwise interactions is that food webs are sparse (that means that there are few links given the number of species) and it is important that we are able to discern between a model that is able to correctly predict interactions that do (true positives) and not (true negatives) occur and one that is simply predicting a lack of interactions (Poisot, 2023). For more detailed methods as to how benchmarking was done refer to SuppMat 3

[Figure 3 about here.]

Data cost

This includes thinking about the need for additional data sources (such as trait or phylogenetic data), the computational cost, as well as the time it might take to generate a network, *e.g.,* binary classifiers require an (often times) extensive list of additional trait data for the model training process, which limits predictions to communities for which you do have the relevant auxiliary data available.

Philosophical constraints

Probably mentioned elsewhere but basically are we constructing networks because we want to make real-world, case-specific predictions *e.g.,* for a conservation area or do we want to just have a set of ecologically plausible networks we can use for theoretical stuffs. Need to discuss the key differences and implications between predicting a **metaweb** (*sensu* Dunne (2006)) and a network realisation. (In a way the idea of predicting a metaweb vs realisation is what makes me hesitant to use the Mangal networks to test the structural models because do we even know what the Mangal networks represent and what the structural models are predicting…) Maybe also Poisot et al. (2015) that discuss how the local factors are going to play a role.

Also need to take into consideration inherent constraints that the model imposes on itself and how it will affect our ability to test hypotheses/ask questions using the *e.g.,* from Petchey et al. (2011) - models that are constrained by connectance means that we are unable to explain connectance itself and you would need a different approach if understanding connectance is your goal. Another way of phrasing this is thinking about what is needed (input data/parameters), produced (final network characteristics), and desired (end-use).

An interesting thing to also think about is data dependant and data independent 'parametrisation' of the models…

4 Concluding remarks

- As discussion about the different model families and in what areas they do/do not do well. This will depend probably a fair bit on how Figure 3 end up looking… But it will also be important to tie in some of the other considerations/constraints that are listed in what is currently Box 2
  - In certain situations structure is 'enough' but there may be use cases where we are really interested in the node-level interactions *i.e.,* species identity is a thing we care about and need to be able to retrieve specific interactions at specific nodes correctly.
- Why do interaction models do so badly at predicting structure? Nuance of metaweb vs realisation but also time? At the core of it interaction models are trained on existing interaction data; this is data that are most likely closer to a metaweb than a local realisation even if they are being inventoried at a small scale…
  - We can briefly shoehorn downsampling here maybe??
- It will be interesting to bring up the idea that if a model is missing a specific pairwise link but doing well overall then when does it matter?
  - The fact that *some* people are concerned about the taxonomic resolution and cascading effects those might have on our understanding of network structure (Pringle, 2020; Pringle & Hutchinson, 2020), but that puts us in a place where we are at risk of losing our ability to distinguish the wood from the tree - are we not (at least at times) concerned more with understanding ecosystem level processes than with needing to understand things *perfectly* at the species level.
  - I don't think these 'rare'/nuanced links (e.g. carnivorous hippos) are going to rock the boat when we think about networks at the structural level.

"The resolution of food-web data is demonic because it can radically change network topology and associated biological inferences in ways that are unknowable in the absence of better data." - Pringle & Hutchinson (2020) The counter to this is that structural models are often not working at the species level and thus the structure remains 'unchanged' when you increase the resolution - I don't think that people are that concerned with the structure of real world networks barring connectance and since that scales with species richness anyway your final proportion will probably still remain the same…

- 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. I feel like I need to look at Berlow et al. (2008) but maybe not exactly in this context but vaguely adjacent.
    - I think this is sort of the crux of the argument presented in Brimacombe et al. (2024) as well.

"we highlight an interesting paradox: the models with the best performance measures are not necessarily the models with the closest reconstructed network structure." - Poisot (2023)

- Do we need network models to predict interactions and interaction models to predict structure?
    - "Another argument for the joint prediction of networks and interactions is to reduce circularity and biases in the predictions. As an example, models like linear filtering generate probabilities of non-observed interactions existing, but do so based on measured network properties." - Strydom et al. (2021)
    - Aligning (dove-tailing) with this the idea of ensemble modelling as presented by Becker et al. (2022)
- Close out with a call to action that we have models that predict networks very well and models that predict interactions very well but nothing that is doing well at predicting both - this is where we should be focusing our attention when it comes to furthering model development…
- Do we expect there to be differences when thinking about unipartite vs bipartite networks? Is there underlying ecology/theory that would assume that different mechanisms (and thus models) are relevant in these two 'systems'.
    - The Terry & Lewis (2020) looks at some methods but is specifically looking at a bipartite world…

4.1 Downsampling

do we bring this up? this could be a box… if we have the 'finances' for it… otherwise it should go to the outstanding questions fur sure

- Dansereau et al. (2023)
- "That being said, there is a compelling argument for the need to 'combine' these smaller functional units with larger spatial networks (Fortin et al., 2021) and that we should also start thinking about the interplay of time and space (Estay et al., 2023). Although deciding exactly what measure might actually be driving differences between local networks and the regional metaweb might not be that simple (Saravia et al., 2022)."

Glossary

Term	Definition
food web	a representation of feeding links between species
topology generator	a model that predicts a network based on assumptions of structure, this network is species agnostic in the sense that it does not necessarily contain information at the node level
interaction predictor	a model that predicts species interactions, these interactions can be used to construct a network but there are no *a priori* assumptions as that will constrain the network structure
model	A tool that can be used to construct food webs, where the resulting network is a representation of a real world network. Models typically only capture specific elements of real world networks and are intended to be used in specific settings

Term	Definition
model family	A family of models that share an underlying philosophy when it comes to the mapping, pragmatism, and reduction of a network. Families have the same underlying philosophies and assumptions that determine the links between nodes as well as how these may be encoded
metaweb	A network that represents *all* the potential links between species. Importantly these links will not necessarily all be realised in a specific location for a specific time
realised network	A network that represents the links between species that are occurring. These networks represent a very localised network…
potential feeding link	links that indicate that an interaction is ecologically feasible but not realised *per se* (a metaweb would contain potential feeding links)
realised feeding link	links that indicate that the interaction is realised 'in the field'. (a realised network contains realised feeding links)
confusion matrix	captures the number of true positives (interaction predicted as present when it is present), false negatives (interaction predicted as absent when it is present), false positives (interaction predicted as present when it is absent), and true negatives (interaction predicted as absent when it is absent)

Outstanding questions

- non-consumptive effects
- can we develop a model that is both a topology generator as well as an interaction predictor?

- how do we define the spatial and temporal ‘boundaries’ of a network

References

Allesina, S., Alonso, D., & Pascual, M. (2008). A General Model for Food Web Structure. *Science*, *320*(5876), 658–661. https://doi.org/10.1126/science.1156269

Barberán, A., Bates, S. T., Casamayor, E. O., & Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal*, *6*(2), 343–351. https://doi.org/10.1038/ismej.2011.119

Becker, D. J., Albery, G. F., Sjodin, A. R., Poisot, T., Bergner, L. M., Chen, B., Cohen, L. E., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Guth, S., Han, B. A., Simmons, N. B., Stock, M., Teeling, E. C., & Carlson, C. J. (2022). Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *The Lancet Microbe*, *3*(8), e625–e637. https://doi.org/10.1016/S2666-5247(21)00245-7

Berlow, E. L., Brose, U., & Martinez, N. D. (2008). The “Goldilocks factor” in food webs. *Proceedings of the National Academy of Sciences*, *105*(11), 4079–4080. https://doi.org/10.1073/pnas.0800967105

Berlow, E. L., Neutel, A.-M., Cohen, J. E., de Ruiter, P. C., Ebenman, B., Emmerson, M., Fox, J. W., Jansen, V. A. A., Iwan Jones, J., Kokkoris, G. D., Logofet, D. O., McKane, A. J., Montoya, J. M., & Petchey, O. (2004). Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, *73*(3), 585–598. https://doi.org/10.1111/j.0021-8790.2004.00833.x

Bhatia, U., Dubey, S., Gouhier, T. C., & Ganguly, A. R. (2023). Network-based restoration strategies maximize ecosystem recovery. *Communications Biology*, *6*(1), 1–10. https://doi.org/10.1038/s42003-023-05622-3

Brimacombe, C., Bodner, K., & Fortin, M.-J. (2024). *Applying a method before its proof-of-concept: A cautionary tale using inferred food webs*. https://doi.org/10.13140/RG.2.2.22076.65927

Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T., & Fortin, M.-J. (2023). Shortcomings of reusing species interaction networks created by different sets of researchers. *PLOS Biology*, *21*(4), e3002068. https://doi.org/10.1371/journal.pbio.3002068

Caron, D., Brose, U., Lurgi, M., Blanchet, F. G., Gravel, D., & Pollock, L. J. (2024). Trait-matching models predict pairwise interactions across regions, not food web properties. *Global Ecology and Biogeography*, *33*(4), e13807. https://doi.org/10.1111/geb.13807

Cirtwill, A. R., Eklf, A., Roslin, T., Wootton, K., & Gravel, D. (2019). A quantitative framework for investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, *10*(6), 902–911. https://doi.org/10.1111/2041-210X.13180

Clegg, T., Ali, M., & Beckerman, A. P. (2018). The impact of intraspecific variation on food web structure. *Ecology*, *99*(12), 2712–2720. https://doi.org/10.1002/ecy.2523

Cohen, J. E., Briand, F., & Newman, C. (1990). *Community Food Webs: Data and Theory*. Springer-Verlag.

Dansereau, G., Barros, C., & Poisot, T. (2023). *Spatially explicit predictions of food web structure from regional level data*.

Dormann, C. F. (2023). The rise, and possible fall, of network ecology. In *Defining Agroecology – A Festschrift for Teja Tscharntke* (pp. 143–159.). Tredition.

Dunn, R. R., Harris, N. C., Colwell, R. K., Koh, L. P., & Sodhi, N. S. (2009). The sixth mass coextinction: Are most endangered species parasites and mutualists? *Proceedings. Biological Sciences*, *276*(1670), 3037–3045. https://doi.org/10.1098/rspb.2009.0413

Dunne, J. A. (2006). The Network Structure of Food Webs. In J. A. Dunne & M. Pascual (Eds.), *Ecological networks: Linking structure and dynamics* (pp. 27–86). Oxford University Press.

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

Estay, S. A., Fortin, M.-J., & López, D. N. (2023). Editorial: Patterns and processes in ecological networks over space. *Frontiers in Ecology and Evolution*, *11*.

Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1949), rspb.2020.1889, 20201889. https://doi.org/10.1098/rspb.2020.1889

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton University Press. https://www.jstor.org/stable/j.ctt7rj8w

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

Jordano, P. (2016b). Sampling networks of ecological interactions. *Functional Ecology*. https://doi.org/10.1111/1365-2435.12763

Lindeman, R. L. (1942). The Trophic-Dynamic Aspect of Ecology. *Ecology*, *23*(4), 399–417. https://doi.org/10.2307/1930126

Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution*, *30*(6), 347–356. https://doi.org/10.1016/j.tree.2015.03.014

Petchey, O. L., Beckerman, A. P., Riede, J. O., & Warren, P. H. (2008). Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences*, *105*(11), 4191–4196. https://doi.org/10.1073/pnas.0710672105

Petchey, O. L., Beckerman, A. P., Riede, J. O., & Warren, P. H. (2011). Fit, efficiency, and biology: Some thoughts on judging food web models. *Journal of Theoretical Biology*, *279*(1), 169–171. https://doi.org/10.1016/j.jtbi.2011.03.019

Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M., & Hartig, F. (2020). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution*, *11*(2), 281–293. https://doi.org/10.1111/2041-210X.13329

Poisot, T. (2023). Guidelines for the prediction of species interactions through binary classification. *Methods in Ecology and Evolution*, *14*(5), 1333–1345. https://doi.org/10.1111/2041-210X.14071

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., Mercier, B., Violet, C., & Vissault, S. (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography*, *48*(7), 1552–1563. https://doi.org/10.1111/jbi.14127

Poisot, T., Cirtwill, A., Cazelles, K., Gravel, D., Fortin, M.-J., & Stouffer, D. (2016). The structure of probabilistic networks. *Methods in Ecology and Evolution*, *7*(3), 303–312. https://doi.org/10

Poisot, T., Gravel, D., Leroux, S., Wood, S. A., Fortin, M.-J., Baiser, B., Cirtwill, A. R., Araújo, M. B., & Stouffer, D. B. (2016). Synthetic datasets and community tools for the rapid testing of ecological hypotheses. *Ecography*, *39*(4), 402–408. https://doi.org/10.1111/ecog.01941

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., Stouffer, D. B., & Kéfi, S. (2016). Describe, understand and predict: Why do we need networks in ecology? *Functional Ecology*, *30*(12), 1878–1882. https://www.jstor.org/stable/48582345

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*(5), 397–406. https://doi.org/10.1111/2041-210X.12180

Pringle, R. M. (2020). Untangling Food Webs. In *Unsolved Problems in Ecology* (pp. 225–238). Princeton University Press. https://doi.org/10.1515/9780691195322-020

Pringle, R. M., & Hutchinson, M. C. (2020). Resolving Food-Web Structure. *Annual Review of Ecology, Evolution and Systematics*, *51*(Volume 51, 2020), 55–80. https://doi.org/10.1146/annurev-ecolsys-110218-024908

Proulx, S. R., Promislow, D. E. L., & Phillips, P. C. (2005). Network thinking in ecology and evolution.

Trends in Ecology & Evolution, *20*(6), 345–353. https://doi.org/10.1016/j.tree.2005.04.004

Saravia, L. A., Marina, T. I., Kristensen, N. P., De Troch, M., & Momo, F. R. (2022). Ecological network assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, *91*(3), 630–642. https://doi.org/10.1111/1365-2656.13652

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

Staniczenko, P. P. A., Lewis, O. T., Jones, N. S., & Reed-Tsochas, F. (2010). Structural dynamics and robustness of food webs. *Ecology Letters*, *13*(7), 891–899. https://doi.org/10.1111/j.1461-0248.2010.01485.x

Stouffer, D. B., Camacho, J., Jiang, W., & Nunes Amaral, L. A. (2007). Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1621), 1931–1940. https://doi.org/10.1098/rspb.2007.0571

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V., & Poisot, T. (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., Fortin, M.-J., Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V., & Poisot, T. (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

Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz, N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (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

Terry, J. C. D., & Lewis, O. T. (2020). Finding missing links in interaction networks. *Ecology*, *101*(7), e03047. https://doi.org/10.1002/ecy.3047

Thuiller, W., Calderón-Sanou, I., Chalmandrier, L., Gaüzère, P., O'Connor, L. M. J., Ohlmann, M., Poggiato, G., & Münkemüller, T. (2024). Navigating the integration of biotic interactions in biogeography. *Journal of Biogeography*, *51*(4), 550–559. https://doi.org/10.1111/jbi.14734

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

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

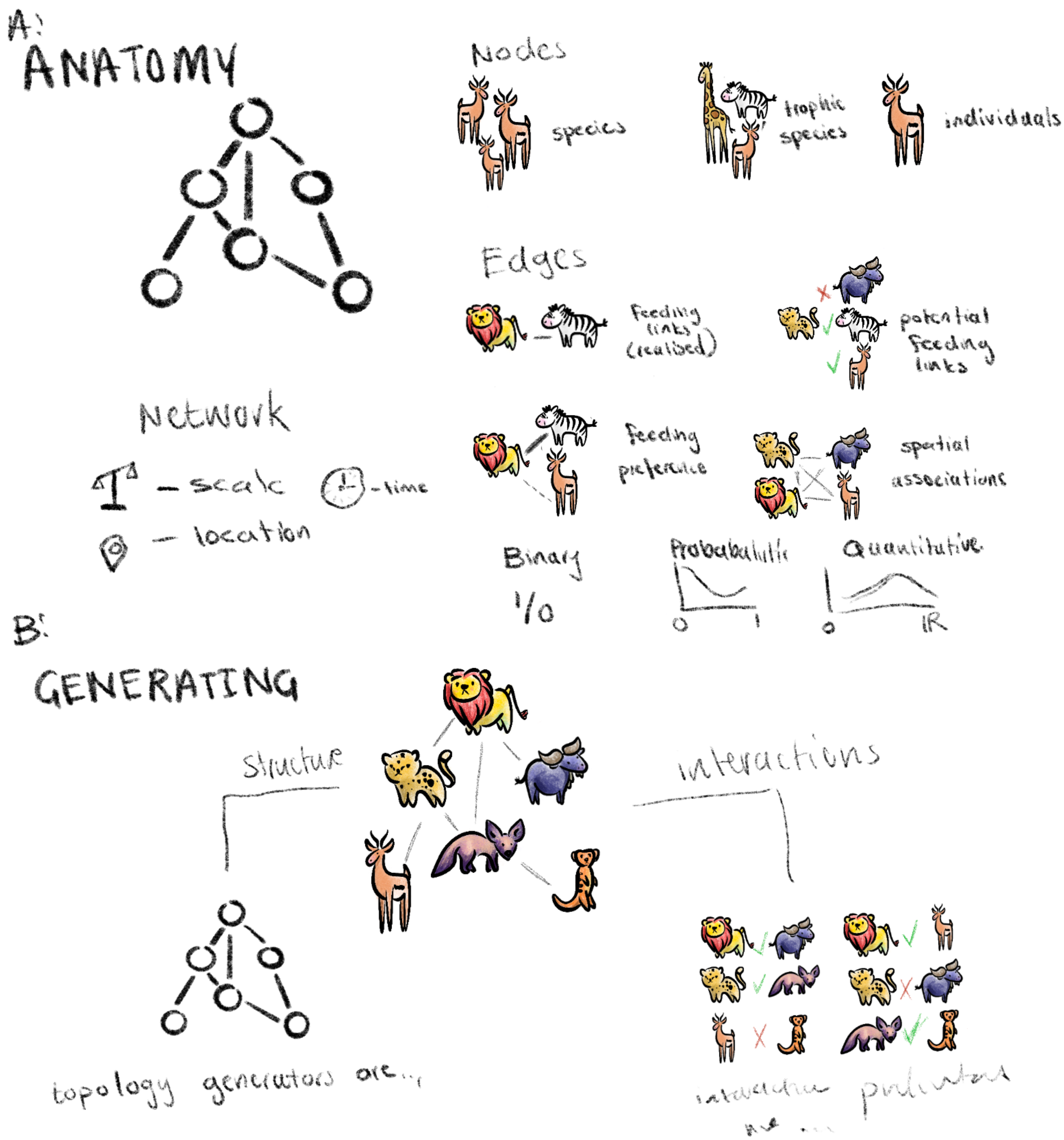

Figure 1: Panel **A** shows the many ways in which a food web can be defined and described at the node, edge, and even network level. Panel **B** (will) shows how the way in which we predict networks also limited and often focuses only only predicting the structure of a network (the final networks is parametrised by the expected structure of the network) or the interactions between species (the final network is determined by the behaviour of the nodes). These different models also encode different philosophies/hypotheses not only as to what determines how a network will look like but also how the final network itself is encoded *i.e.,* its anatomy. (*aside:* there is the potential to either try and visually summarise how the different model families define a network (so repeating the motifs used in the ANATOMY panel) alternatively it would be cool to try and have a panel C that tries to quantify the different 'data ingredients' you would need to try and construct a network, this would probably be very visually overwhelming though...)

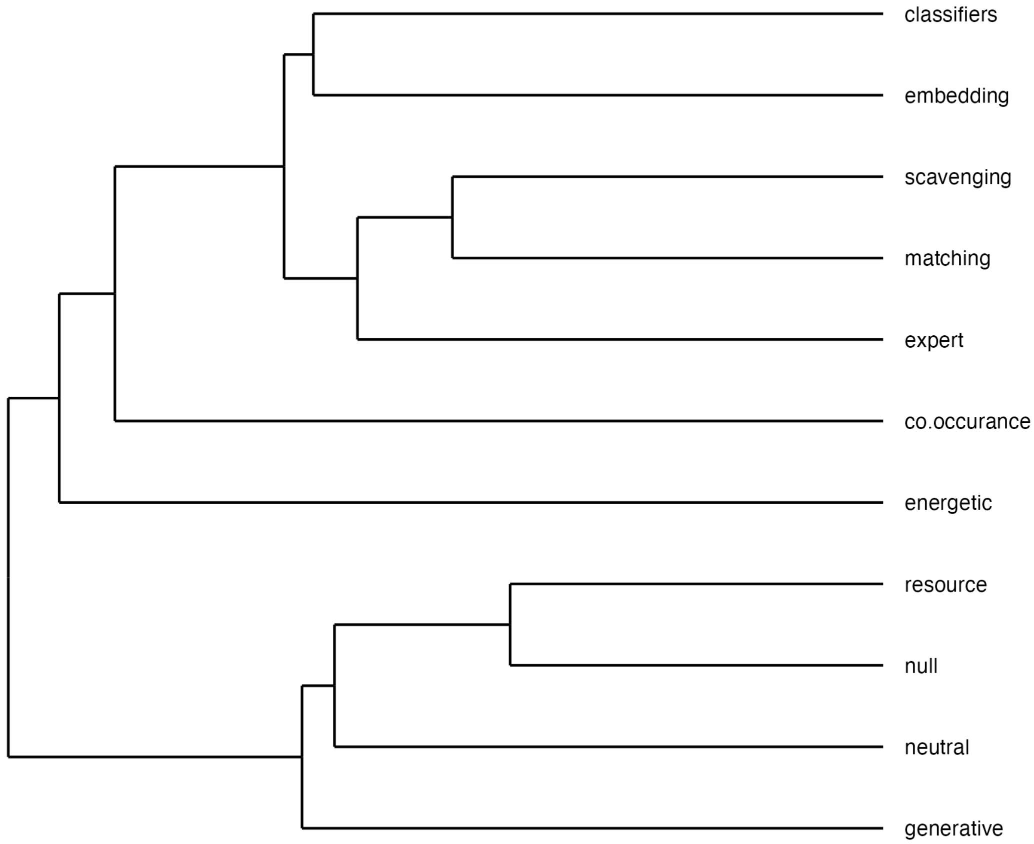

Figure 2: Dendrogram of the trait table using a hierarchical clustering model, This is based off of the traits table in SuppMat 2)

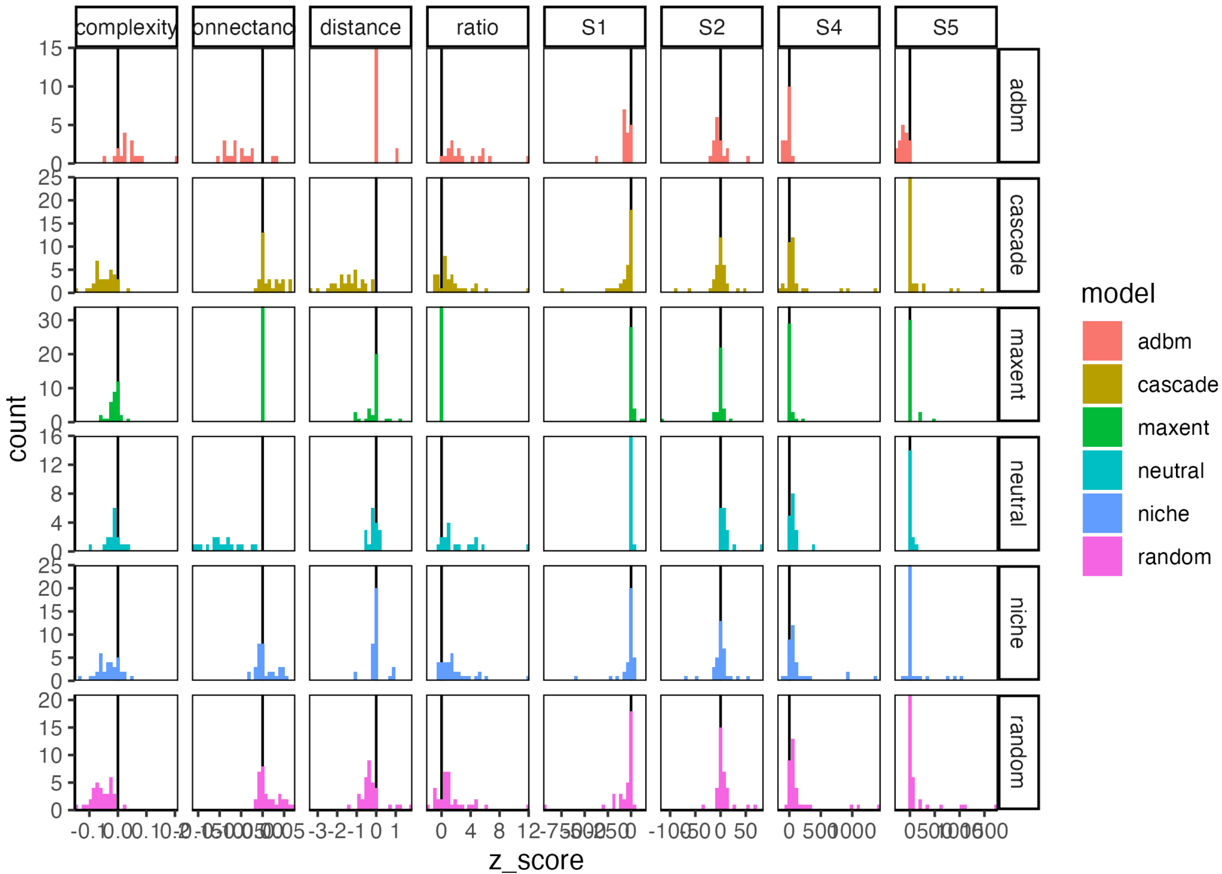

Figure 3: Difference between real and model network property. S1 - S5 represent the different motif structures identified in Stouffer et al. (2007) which are S1: Number of linear chains, S2: Number of omnivory motifs, S4: Number of apparent competition motifs, and S5: Number of direct competition motifs