T is for Topology

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

[1]School of Biosciences, University of Sheffield, Sheffield, UK,
[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA,
[3]Université de Montreal,
[4]Québec Centre for Biodiversity Sciences,

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

Abstract

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.

Plain Language Summary

Although it has been acknowledged that communities consist not only of co-occurring species but that they also interact being able to quantify those interactions and assemble them into interaction networks has been a limiting factor in the integration of network ecology into other fields of ecology. As the field of network ecology has matured there has been an accompanying expansion in the development of theory and tools that are centred around generating networks or predicting the interactions between species. Notably many of these tools have been developed with different underlying philosophies, ideas, and mechanisms as to what structures the interactions between species. It is thus critically important that those wanting to adopt these network generating tools be aware of how the the specific questions being asked maps to the underlying assumptions made when generating networks, as well as the limitations of how the networks/interactions are delimited. Here we provide an overview of the canonical network generating models, comparing and contrasting the underlying assumptions, data requirements, and resulting network predictions made by the different families in an attempt to provide guidance for those interested in adopting the generation of networks into their workflow. [R1. a discussion on the underlying assumptions we are making when we delimit a network]. [R2. an overview of how the different model families differ - ordination space/benchmarking]. [R3. identifying the relevant questions/bodies of theory that the networks generated by different families are suited to answer]. When choosing to construct an interaction network the researcher is faced with many assumptions and considerations that should be made and it is important to be aware of these limitations to avoid constructing (something poetic to capture the idea of falsity/false idols). Being aware of these choices is particularly important as the availability of these tools grows and network ecology starts to be adopted into other aspects of ecology and conservation biology.

0.1 Introduction

Some more general thotsTM

0.2 What does a 'network' mean?

0.3 What do we mean when we generate a network?

0.4 How does network ecology link to the outstanding questions in ecology

- In order to construct a 'perfect' network *i.e.,* one which *perfectly* captures the dynamics for a specific community at a specific time 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.

- Bring up the fact that delimiting a network is in and of itself fuzzy - we tend to think of them in terms of snapshots but in reality the final (empirical) network is often the result of aggregation over multiple timescales.
- Also 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), 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. To say this in a different way maybe it comes down to thinking about the scale of organisation within a network… The classical levels of organisation within ecology (population, community, …) are also relevant when we think about a networks.
- This is also going to be very important as network ecology as a field starts being incorporated into other fields *e.g.,* the biogeography one from Thuiller et al. (2024) since there is nuance in the method you use to generate the network but also in a more general way in which you actually think about/delimit/conceptualise your network.

- What do we mean when we talk about an 'interaction network'? Is it the flow of biomass/energy? Is it an effect on fitness (feeding)? (can we really represent indirect interactions as actual links?)

  - "For example, nodes can be individuals and the links contacts between those individuals (Melian et al. 2011), nodes can be species and the links the interactions (e.g. feeding, mutualistic, parasitic) between them (de Ruiter et al. 1995), or nodes can be ecological communities or ecosystems and the links fluxes between these localities (Proulx et al. 2005; Chadès et al. 2011)" - Poisot, Stouffer, et al. (2016)
  - "Further study is needed to determine whether the discrepancies in observed degree distributions are best explained by the differences in the coding of network structure or by biological differences between food webs and plant–animal interaction networks." Proulx et al. (2005)

- 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 ('real world', on the ground applications). Broadly this means that we either want to construct/predict a collection of interactions (generate networks) or a network of interactions (predict interactions).

  - Also need to discuss the nuance between predicting something vs generating something, which also then bleeds into need to be aware when we use the term 'model' as well…
  - "For example, dynamic models focus on how structure relates to population dynamics and community stability (2, 7–11). Evolutionary models incorporate the processes that control the formation and expansion of food webs (12, 13). Static models include rules that determine structural attributes of food webs (14–19)." - Petchey et al. (2008)

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 networks for the purpose of running further simulations (*e.g.,* extinction simulations) or understanding some higher-level process/concept

(*e.g.,* understanding energy flows), importantly these networks do not require any level of species specificity *per se* and it is more the arrangements of the nodes (species) within the context of network structure that is of value. This researcher is contrasted by one that is interested in constructing real-world, location specific, interaction data for a specific collection of species (community). This is driven by the need for researchers to find alternative ways to infer the interactions between species as a way to overcome the inherit challenges of inventorying interaction in the field (Morales-Castilla et al., 2015; Strydom, Catchen, et al., 2021). 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 (*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 (*i.e.,* local-level predictions between specific species).

- Brief history of the development of tools within the context of the two different fields? Sort of where the theory/body of work was based and how that has changed?

  - 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

  - 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.

  - 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).

- A breakdown of wanting to generate a network; statement of need and core philosophies

- A breakdown of wanting to predict an interaction; statement of need (Jordano, 2016b, 2016a; Poisot et al., 2021) and core philosophies (trait-matching, coexistence, evolutionary backbones)

- Stands to reason then that we have developed methods that 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.

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 functioning at a specific 'network 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…

"It makes no sense to describe the interaction structure of nodes which in themselves are poorly defined." — Roslin et al. (2013, p. 2)

0.5 Model families

Given the large number of models that have been developed it is perhaps more meaningful to group models into families with the idea that models from the same family will yield similar results because they play by similar rules. These rules referring to the underlying philosophy as to what structures either networks or the interactions within them (see Figure 1 panel A). Although there have been efforts to compare and contrast different models (*e.g.,* Williams & Martinez, 2008 looked at 'structural' models; and Pichler et al., 2020 looked at 'machine learning algorithms') there still lacks an overall synthesis as to how the different model families differ from each other.
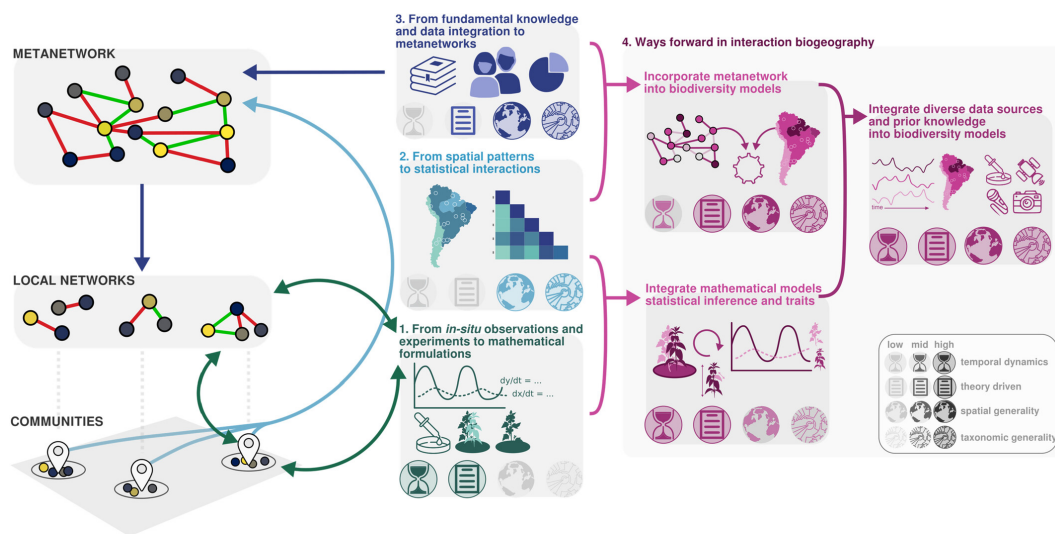

Figure 2: I like the use of the different source indicator items (not too dissimilar from Tall Tom's nature paper but also different). This is from Thuiller et al. (2024)

Null models: The interactions between species occurs regardless of the identity of the species (*i.e.,* species have no agency) and links are randomly distributed throughout the network. There is however the assumption that a network will be constrained by the number of links. Type I (Fortuna & Bascompte, 2006), where interactions happen proportionally to connectance and Type II (Bascompte et al., 2003), where interactions happen proportionally to the joint degree of the two species involved. These two models are equivalent to the Erdos-Renyi and Configuration models (Newman, 2010) respectively (check that though).

Neutral models: Based on the theory that interactions occur as the result of the abundance of species (*i.e.,* the species still has no agency but its abundance does?). See Pomeranz et al. (2019)

Resource models: Based on the idea that networks follow a trophic hierarchy and that species interactions can be determined using a single dimension [the "niche axis"; Allesina et al. (2008)]. Essentially these models can be viewed as being based
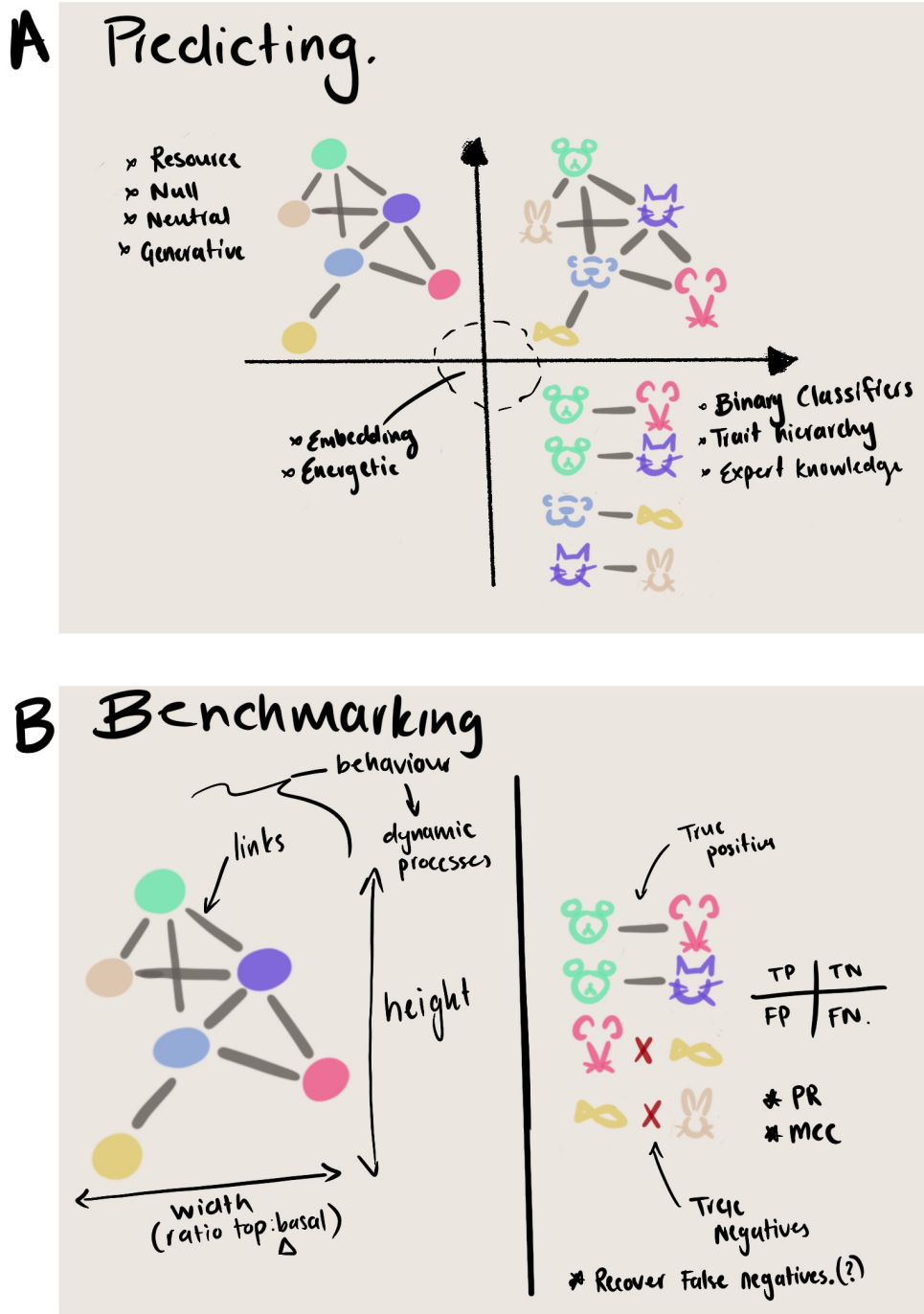
Figure 1: Conceptual figure of the 'network prediction'. Panel A shows where the model families fall in the the context of being models that predict networks or models that predict interactions space. Panel B serves to highlight the characteristics one might like to 'test'/benchmark for a model based on it being either a network or interaction predicting model

on the idea of resource partitioning (niches) along a one-dimensional resource and that the number of links scale with species richness (linear link scaling). That is, there is some sort of hierarchical feeding based on how a 'resource' is partitioned. Broadly this family consists of three core models; the cascade model (Joel E. Cohen et al., 1990), which rests on the idea that species feed on one another in a hierarchical manner; the niche model (Williams & Martinez, 2000), broadly all species are randomly assigned a 'feeding niche' and all species that fall in this niche can be consumed by that species; and the nested hierarchy model (Cattin et al., 2004), which adds some component of phylogenetic clustering/signal... so not a single dimension? **TODO**. Williams & Martinez (2008) provides a broader overview of some of the variations in these models as well as comparison between them regarding their ability to retrieve elements of networks structure (see also Allesina et al. (2008)).

Generative models: (this is maybe a bit of a bold term to use). MaxEnt (Banville et al., 2023), (maybe) stochastic block (Xie et al., 2017).

Feeding models: Broadly this family of models is rooted in feeding theory and allocates the links between species based on energetics, which predicts the diet of a consumer based on energy intake. This means that the model is focused on predicting not only the number of links in a network but also the arrangement of these links based on the diet breadth of a species. The diet breadth model (Beckerman et al., 2006) as well as its allometrically scaled cousin the allometric diet breadth model (ADBM) (Petchey et al., 2008) determine links between species based on the energetic content, handling time, and density of species. See also DeAngelis et al. (1975)

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

Binary classifiers: The task of predicting if an interaction will occur between a species pair is treated as a statistical binary classification task, where the task is to correlate 'real world' interaction data with a suitable ecological proxy for which data is more widely available (*e.g.,* traits). Model families often used include generalised linear models (*e.g.,* Caron et al., 2022), random forest (*e.g.,* Llewelyn et al., 2023), trait-based k-NN (*e.g.,* Desjardins-Proulx et al., 2017), and Bayesian models (Cirtwill et al., 2019; *e.g.,* Eklöf et al., 2013). See Pichler et al. (2020) for a more detailed overview on the performance of machine learning and statistical approaches for inferring trait-trait relationships.

Graph embedding: This family of approaches has been extensively discussed in Strydom et al. (2023) but can be broadly explained as an approach that estimates latent features from observed networks that can be used to predict interactions. Strydom et al. (2022) uses a transfer learning framework (specifically using a random dot product graph for embedding) based around the idea that interactions are evolutionarily conserved and that we can use known networks, and phylogenetic relationships, to predict interactions for a given species pool. **TODO** Log-ratio (Rohr et al., 2010)

Trait matching: Interactions are determined by a series of 'feeding rules', whereby the interaction between a species pair will only occur if all feeding rules are met. These rules are determined on an *a priori* basis using expert/ecological knowledge to determine the underlying feeding hierarchy using ecological proxies (Morales-Castilla et al., 2015). For example the Paleo Foodweb Inference Model (PFIM, Shaw et al., 2024) uses a series of rules for a set of trait categories (such as habitat and body size) to determine if an interaction can occur between a species pair. What sets this family of models apart from **expert knowledge** ones is that there is a formalisation of the feeding rules and thus there is some ability to transfer these rules to different communities.

Expert knowledge: Not so much about empirical observations but more the value of 'local' knowledge and having specific individuals sitting around a table and assigning a value of how confident they are that a specific species pair are likely to interact (*e.g.,* Jennifer A. Dunne et al., 2008), this has the added advantage that interactions can be scored in a more categorical as opposed to binary fashion, *e.g.,* Maiorano et al. (2020) score interactions as either obligate (typical food resources) or occasional (opportunistic feeding) interactions. I feel like its worth also mentioning downfalls *a la* Brimacombe et al. (2023)…

Data scavenging: There are also a lot of published *interaction* data that are publicly available *e.g.,* the Global Biotic Interactions (GloBI) database (Poelen et al., 2014) and these can also be used to construct an interaction network by mining these sources to look for interactions for specific species pairs. This is done by matching species pairs against those within a dataset of trophic interactions to determine if an interaction is present or absent between the two species (*e.g.,* the WebBuilder tool developed by Gray et al., 2015). It is important to note that this methodology is only going to be able to infer observations that have been recorded in the field, and given the relative scarcity [*I say Poisot et al. (2021) but that's more an overview of complete networks but one can also get pairwise interactions from these types of data so I feel like its okay?*] and localised sampling of these types of datasets it is very likely that there will be many false negatives (missing pairwise interactions) using this approach.

Co-occurrence: Trying to infer interactions from the co-occurrence patterns of species pairs within the community *e.g.,* the geographical lasso (Ohlmann et al., 2018). This (for me) seems fundamentally flawed and Blanchet et al. (2020) seems to agree with me at least a little bit.

0.6 Model benchmarking

- 'Testing' the performance of a model is going to depend on some of the core limitations of the model itself thus it makes sense to think of two sets benchmarking rules for network and interaction prediction models respectively (see Figure 1 panel B).

- When it comes to network models we are concerned with the 'preservation' of structure and distribution of links across the network. For interaction models we want to ensure that we are able to retrieve interactions that really exist but also those that cannot exist (*sensu* forbidden links Jordano (2016b))

  "As long as these predictions are not perfect, some interactions will be predicted at the 'wrong' position in the network; these measures cannot describe the structural effect of these mistakes. On the other hand, measures of network structure can have the same value with interactions that fall at drastically different positions; this is in part because a lot of these measures covary with connectance, and in part because as long as these values are not 0 or their respective maximum, there is a large number of network configurations that can have the same value." - Poisot (2023)

0.6.1 Benchmarking network models

- Maybe look at some of the historic papers that compare some of the 'resource models'

- See also Allesina et al. (2008) and the likelihood function that they use for model selection

- Look at Vermaat et al. (2009)

"Possibly, the most striking caveat of the use of summary statistics is that it cannot tell us whether or not a model is able to fully replicate empirical networks." - Allesina et al. (2008)

0.6.2 Benchmarking interaction models

- Main concern with predicting interactions is that we want to test the 'quality' of the links we are predicting (both true positives and true negatives), but the inherit sparsity (meaning high class imbalance) means that we also need to look at the balance of these predictions.

- "Both precision and recall may be useful in cases where there is imbalanced data. However, it may be valuable to prioritize one over the other in cases where the outcome of a false positive or false negative is costly."

- Caveat regarding the use of real world interaction data both for training and validating predictions? *e.g.,* Poisot, Ouellet, et al. et al 2021 and Catchen et al 2023

- See Poisot (2023)

  - skill (ability to make the right prediction; evaluate whether low prevalence can lull us into a false sense of predictive accuracy)

  - bias (trends towards systematically over-predicting one class)

  - class imbalance (the relative number of cases representing interactions)

- "These results suggest that learning from a dataset with very low connectance can be a different task than for more connected networks: it becomes increasingly important to capture the mechanisms that make an interaction exist, and therefore having a slightly more biased training dataset might be beneficial. As connectance increases, the need for biased training sets is less prominent, as learning the rules for which interactions do not exist starts gaining importance"

- Maybe also looking at how well a model can recover 'missing links' *i.e.,* false negatives *sensu* what we did in Strydom et al. (2022)

- Need to discuss the key differences and implications between predicting a metaweb (*sensu* Jennifer A. Dunne (2006)) and a network realisation. Maybe also Poisot et al. (2015) that discuss how the local factors are going to play a role.

0.7 Data & Methods

0.7.1 Selecting models

This section depends on if we go the family route and where we introduce them. But a more extended description of each model can be found in the `Extended Model Description` notebook (I'm trying to work out how to embed this…)

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 model families and some of their features. *A column that captures naïve vs a priori knowledge of interactions/structure i.e., a 'parameter' of sorts?*

Model family	Theory	Network predicted	Links predict	Make 'de novo' predictions (node/species identity)	Needs (minimum)	Assembly mechanism	Constraints	Interaction
null	Network structure is random	structure		no	network (species agnostic)	random	link	binary
neutral	Network structure is random, but species abundance plays a role	structure		yes	abundance, number of links	mass effect	link	binary
resource	Networks are interval, species can be ordered on a 'niche axis'	structure	flow of biomass (resource?)	no	richness, connectance	'random'	link	binary
generative	Networks are determined by their structural features	structure		no	network (species agnostic)	'random'		binary

Model family	Theory	Network predicted	Links predict	Make 'de novo' predictions (node/species identity)	Needs (minimum)	Assembly mechanism	Constraints	Interaction
energetic	Interactions are determined by foraging theory (feeding links)	interactions	flow of energy	yes	body size	deterministic	energy	
graph embedding	Interactions can be predicted from the latent traits of networks	interactions	potential feeding links	yes	interactions, phylogenetic tree, list of target species (species pool)			probabilistic
trait matching	Interactions can be inferred by a mechanistic framework/relationships	interactions	potential feeding links	yes	prior (expert) knowledge of trait hierarchy/relationships, traits, list of target species (species pool)	mechanistic	trait matching (*sensu* forbidden links in a way)	

Model family	Theory	Network predicted	Links predict	Make 'de novo' predictions (node/species identity)	Needs (minimum)	Assembly mechanism	Constraint	Interaction
binary classifiers	Interactions can be predicted by learning the relationship between interactions and ecologically relevant predictors	interactions	potential feeding links	yes	interactions traits, list of target species (species pool)	statistical		
expert knowledge	'Boots on the ground' ecological knowledge and observations	interactions	potential feeding links	yes	list of target species (species pool)	mechanistic	forbidden links	
data scavenging	Webscraping to create networks from online databases	interactions	potential feeding links	no	list of target species (species pool)			binary

Model family	Theory	Network predicted	Links predict	Make 'de novo' predictions (node/species identity)	Needs (minimum)	Assembly mechanism	Constraints	Interaction
co-occurrence	co-occurrence patterns arise from interactions so we can use these patterns to reverse engineer the interactions	co-occurrence links? (or am I being a bit too mean here)	association links		co-occurrence (so a species list?)			

0.7.2 Datasets used

0.7.2.1 Mangal networks

We queried the Mangal (Poisot, Baiser, et al., 2016) database and extracted a total of **TODO** networks. [*Some sort of summary as to the geographic/taxonomic range??*] Although these networks represent a high volume of interaction data they do not have accompanying 'metadata' that we would need for some of the more data-hungry model families (*e.g.,* local abundance), the Mangal networks were used to provide the 'starting values' for the random, resource, and generative families. This allows us to generate a large number of different networks that we can use to compare and contrast the performance of the various model families (see Section 0.7.3 for a more detailed breakdown). For each network from Mangal we generated **TODO** versions of that network using each model family.

> "These complex food webs differ in their level of resolution and sampling effort, which may introduce noise in the estimation of their properties, especially given their large number of interacting elements. However, because our MaxEnt models are applied on imperfect data, they aim at reproducing the sampled structure of food webs, not their actual structure." - Banville et al. (2023) (something to think about…)

0.7.2.2 Empirical networks

'Elite' number of datasets for interaction models

Although the availability of empirical interaction data is growing as techniques begin to improve and grow (Pringle & Hutchinson, 2020), we still lack a way to define what is the 'ideal' interaction dataset.

New Zealand dataset(s): Pomeranz et al. (2019)

> 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 re-
> gional 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 poten-
> tially interact and obviously not constrained by real-world scenarios) *sensu*
> Caron et al. (2024). 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 Jennifer A. Dunne et al. (2008) work. 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?

0.7.3 Model benchmarking

For now the (still essentially pending) workflow/associated code can be found at the following repository BecksLab/topology_generators. This will reflect that which is shown in panel *B* of Figure 1.

- 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

0.7.3.1 Network models

Want to compare real vs predicted and then get something that looks like Figure 7

- connectance, nestedness (Bastolla et al., 2009), modularity (Barber, 2007), asymmetry (Delmas et al., 2018), and Jaccard network dissimilarity (Canard et al., 2014)

- *Shape:* do the models construct tall 'pencil' vs flat 'pancake' networks (Beckerman 2024, pers comms), generality/vulnerability, chain length (?)

- *Structure:* Predicting 'structure' - SVD (Strydom, Dalla Riva, et al., 2021) but maybe something like nestedness as well (?)

- *Links:* are the number of links preserved (most network predicting models are to some extend link constrained but useful to see)

- *Motifs:* Staniczenko et al. (2010) uses S1, S2, S4, S5 from Stouffer et al. (2007)

  - S1: Number of linear chains

  - S2: Number of omnivory motifs

394   − S4: Number of apparent competition motifs

395   − S5: Number of direct competition motifs

396   *0.7.3.2 Interaction models*

397   • Based on Poisot (2023):
398   − Precision-Recall (PR-AUC) - performance
399   − Matthews correlation coefficient (MCC) - accuracy
400   • Maybe same measures we use for the network models

401   *0.7.3.3 Action plan*

402   1. Shortlist/finalise the different topo generators
403   2. collate/translate into `Julia`
404   • *e.g.,* some models wil be in SpeciesInteractionNetworks.jl (new EcoNet); I
405   know (parts of) the transfer learning stuff is and the niche model
406   • others will need to be coded out (the more simpler models should be easier)
407   3. Curate networks for the different datasets/scenarios we select - I feel like
408   there might be some scenarios that we can't do all models for all datasets but
409   maybe I'm being a pessimist.
410   • Need to also think about where one might find the additional data for some
411   of the models...
412   − Body size: Herberstein et al. (2022) - Although maybe Andrew has
413   strong thotsTM RE the one true body size database to rule them all...
414   − Other trait sources: Wilman et al. (2014) and Jones et al. (2009)
415   − This is where we'll get the paleo traits from if I'm correct Bambach et
416   al. (2007)
417   − Phylogeny stuff: Upham et al. (2019) (what we used for TL but its
418   only mammals...) but I'm sure there will be others
419   • Also limitation of scope... *e.g.,* do we even dare to think about including
420   plants/basal producers (see *e.g.,* Valdovinos et al. (2023))
421   • Taxonomic harmonisation - something to think about and check

0.8 Results

422

423 Joel E. Cohen et al. (1985) actually tells us that the cascade model only really
424 works for communities that range from 3-33 species... and Williams & Martinez
425 (2008) also highlights how structural models really only work for small communities

0.8.1 Qualitative stuff

426

427 Maybe not the best term to use but thinking here about practical limitations of the
428 different families. This can include thinking about:

429   • scale limitations (time or space); *e.g.,* a metaweb is going to encapsulate but
430   not distinguish between different seasons or locations
431   • data needed. I think this can be in the form of real world datasets (*e.g.,*
432   traits) but also *a priori* knowledge (*e.g.,* having to define the constraints of a
433   niche model)
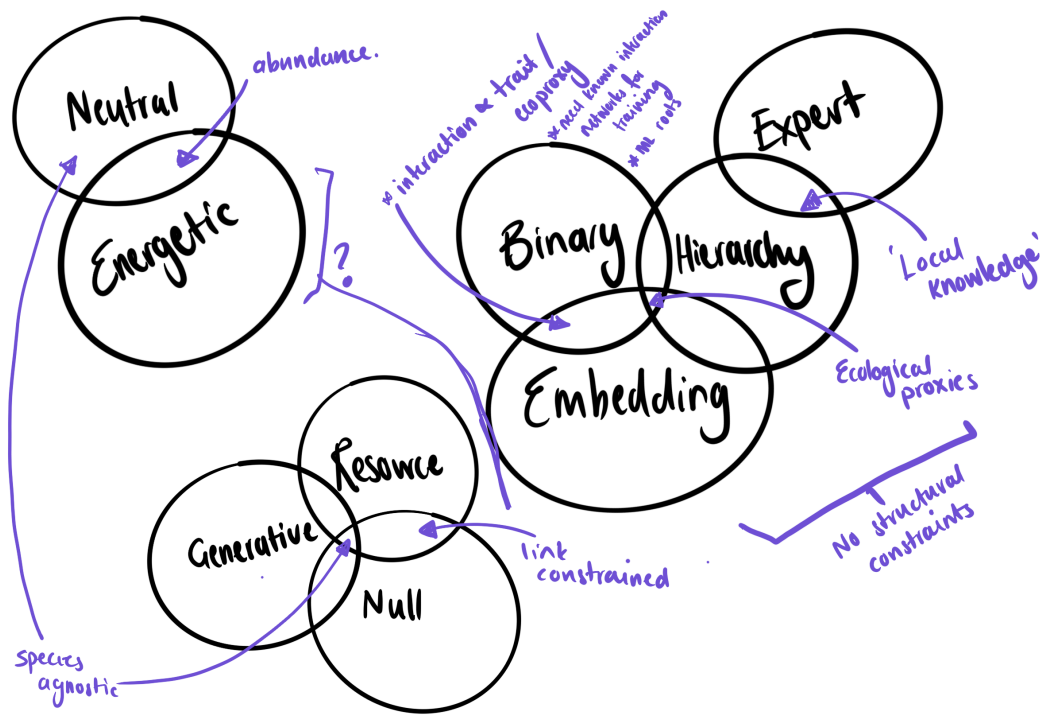434   • computational costs

Figure 3: I still haven't given up on a sort of venn diagram idea but maybe it going to be more of a venn-flow chart hybrid...
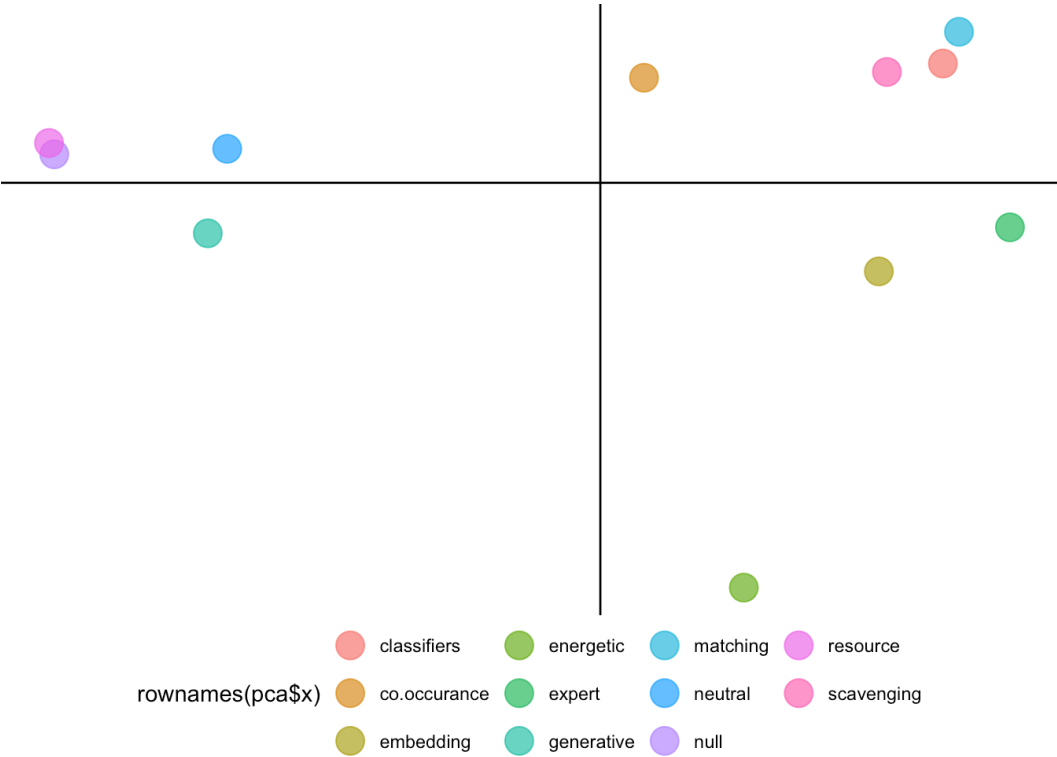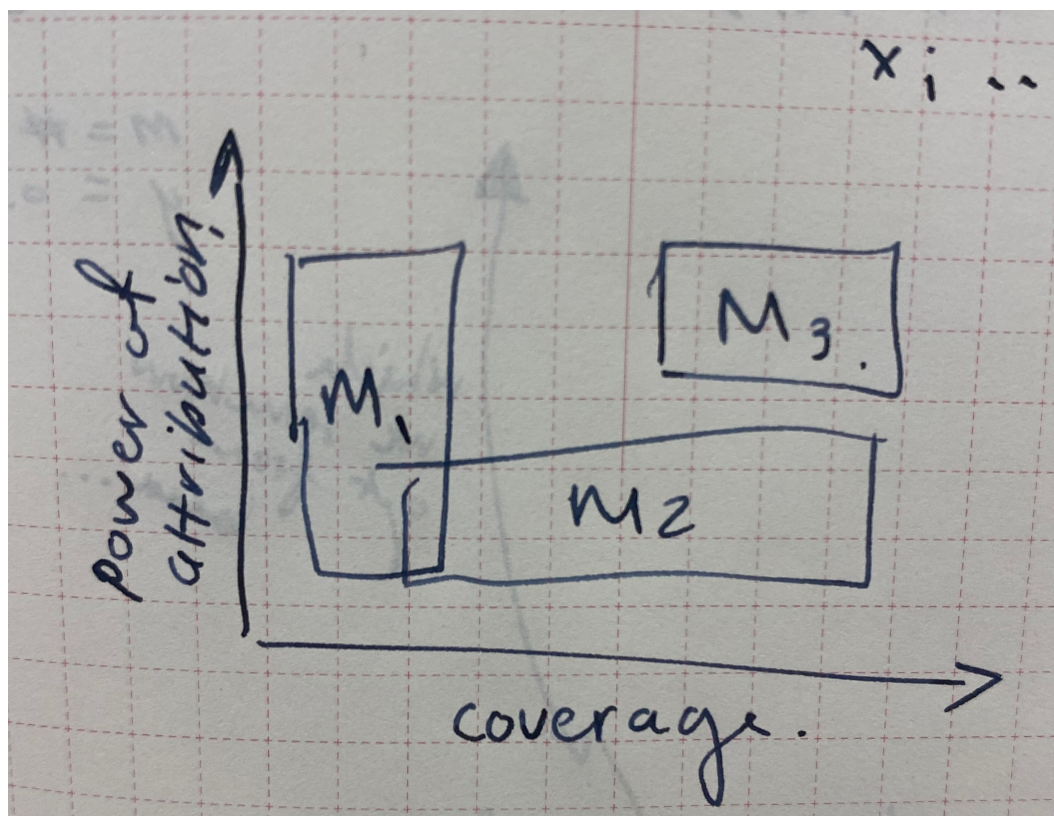

Figure 5: PCA of the trait table

Figure 4: I like these schematics that Charlie Outhwaite presented at the EEB seminar (there was a series of them).
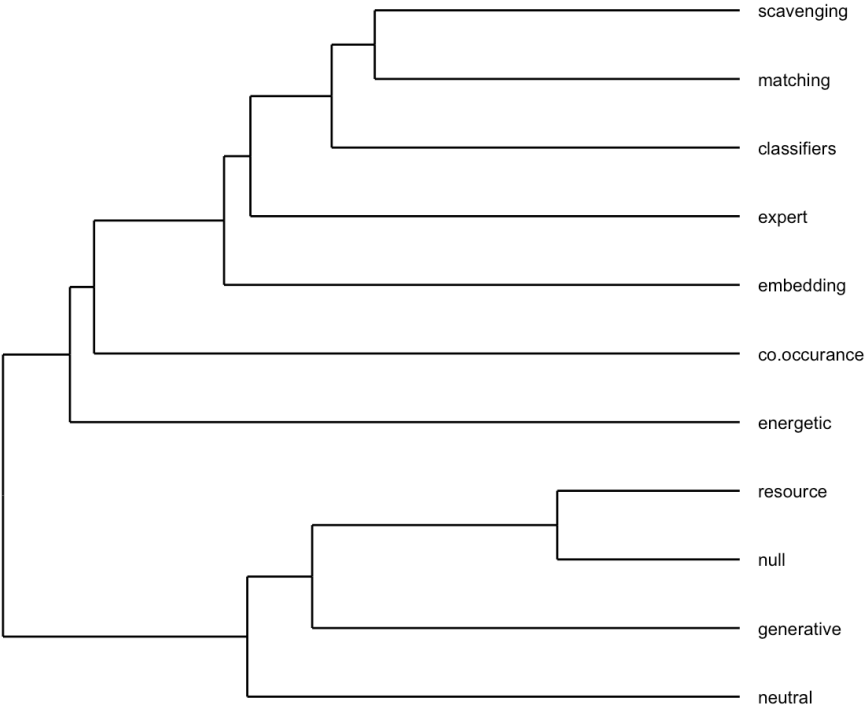
Figure 6: Dendrogram of the trait table
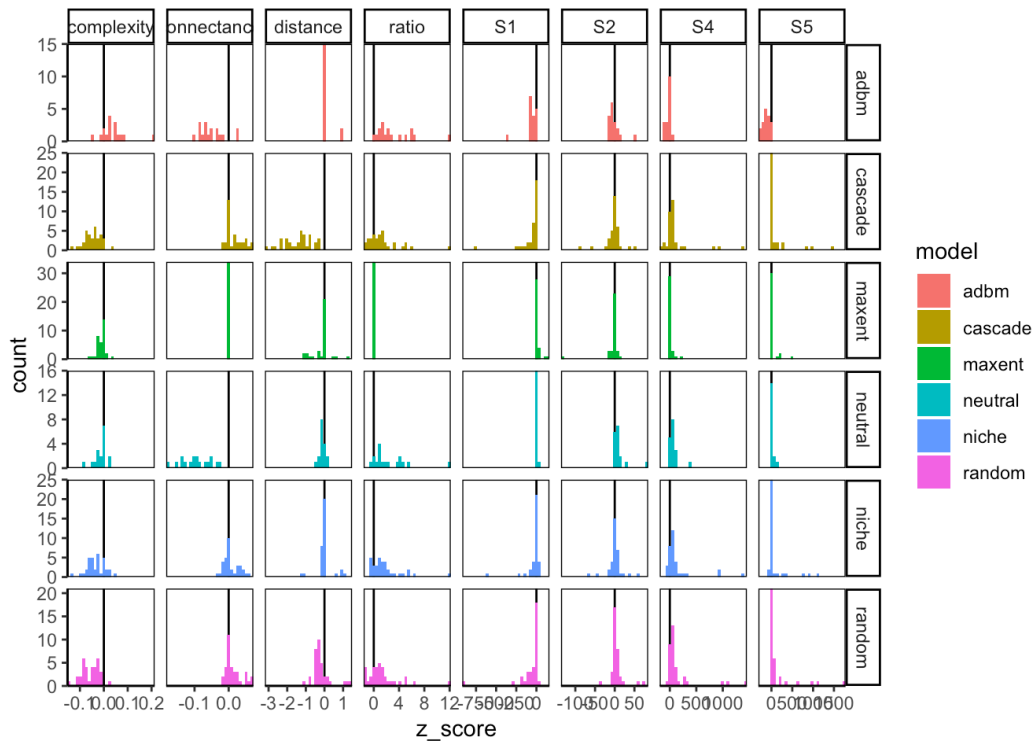
437 *0.8.2 Quantitative stuff*

Figure 7: Difference between real and model network property. S1 - S5 represent the different motif structures identified in Stouffer et al. (2007).

438 Source: Article Notebook

439 I really like this way of plotting results from Pichler et al. (2020)

0.9 Discussion

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

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

450 • Why do interaction models do so badly at predicting structure? Nuance of
451 metaweb vs realisation but also time? At the core of it interaction models are
452 trained on existing interaction data; this is data that are most likely closer
453 to a metaweb than a local realisation even if they are being inventoried at a
454 small scale.

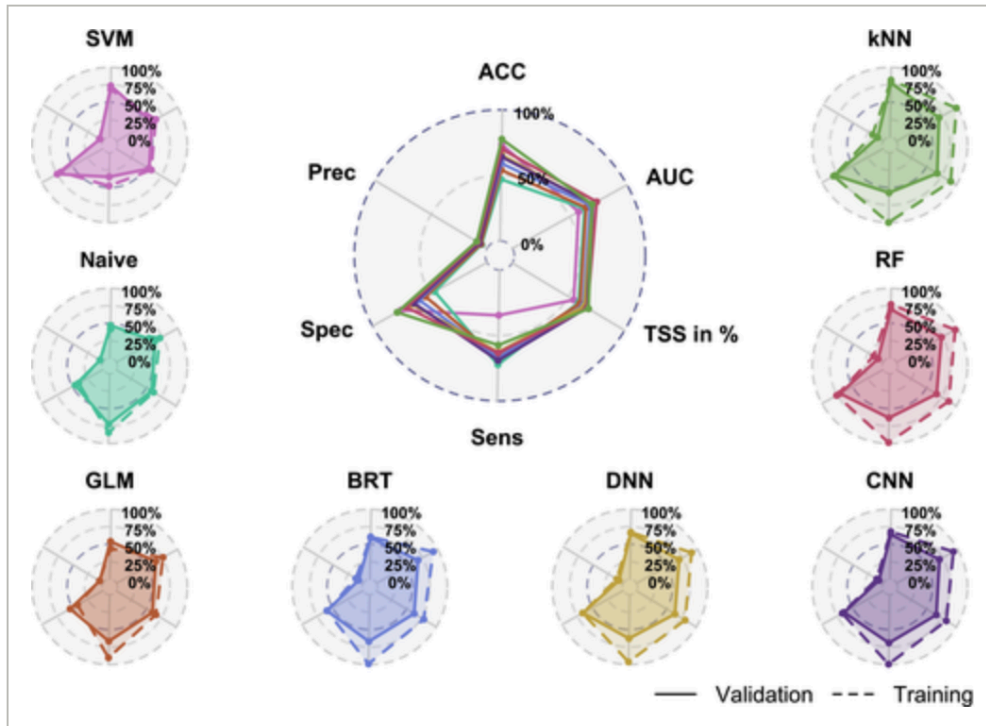455 – I think this is sort of the crux of the argument presented in Brimacombe et
456 al. (2024)

Figure 8: Cool way to conceptualise results from Pichler et al. (2020)

"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?* (lets not think about that too hard or I might just have to sit in silence for a while...)

  – "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, Catchen, et al. (2021)

  – Aligning (dove-tailing) with this the idea of ensemble modelling as presented by Becker et al. (2022)

- It 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?

- 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. (we need models that will fill the space in the top right quadrant of panel A in Figure 1)

References

Source: Article Notebook

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

Bambach, R. K., Bush, A. M., & Erwin, D. H. (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

Banville, F., Gravel, D., & Poisot, T. (2023). What constrains food webs? A maximum entropy framework for predicting their structure with minimal biases. *PLOS Computational Biology*, *19*(9), e1011458. https://doi.org/10.1371/journal.pcbi.1011458

Bascompte, J., Jordano, P., Melian, C. J., & Olesen, J. M. (2003). The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387. https://doi.org/10.1073/pnas.1633576100

Becker, D. J., Albery, G. F., Sjodin, A. R., Poisot, T., Bergner, L. M., Chen, B., et al. (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

Beckerman, A. P., Petchey, O. L., & Warren, P. H. (2006). Foraging biology predicts food web complexity. *Proceedings of the National Academy of Sciences*, *103*(37), 13745–13749. https://doi.org/10.1073/pnas.0603039103

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

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, *23*(7), 1050–1063. https://doi.org/10.1111/ele.13525

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

Brimacombe, C., Bodner, K., & Fortin, M.-J. (2024, April). 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

Caron, D., Maiorano, L., Thuiller, W., & Pollock, L. J. (2022). Addressing the Eltonian shortfall with trait-based interaction models. *Ecology Letters*, *25*(4), 889–899. https://doi.org/10.1111/ele.13966

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

Cattin, M.-F., Bersier, L.-F., Banašek-Richter, C., Baltensperger, R., & Gabriel, J.-P. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature*, *427*(6977), 835–839. https://doi.org/10.1038/nature02327

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*, *0*(ja). https://doi.org/10.1111/2041-210X.13180

Cohen, Joel E., Newman, C. M., & Steele, J. H. (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), 421–448. https://doi.org/10.1098/rspb.1985.0042

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

DeAngelis, D. L., Goldstein, R. A., & O'Neill, R. V. (1975). A Model for Tropic Interaction. *Ecology*, *56*(4), 881–892. https://doi.org/10.2307/1936298

Desjardins-Proulx, P., Laigle, I., Poisot, T., & Gravel, D. (2017). Ecological interactions and the Netflix problem. *PeerJ*, *5*, e3644. https://doi.org/10.7717/peerj.3644

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

Dunne, Jennifer 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

Eklöf, A., Tang, S., & Allesina, S. (2013). Secondary extinctions in food webs: A Bayesian network approach. *Methods in Ecology and Evolution*, *4*(8), 760–770. https://doi.org/10.1111/2041-210X.12062

Fortuna, M. A., & Bascompte, J. (2006). Habitat loss and the structure of plant-animal mutualistic networks: Mutualistic networks and habitat loss. *Ecology Letters*, *9*(3), 281–286. https://doi.org/10.1111/j.1461-0248.2005.00868.x

Gravel, D., Poisot, T., Albouy, C., Velez, L., & Mouillot, D. (2013). Inferring food web structure from predator–prey body size relationships. *Methods in Ecology and Evolution*, *4*(11), 1083–1090. https://doi.org/10.1111/2041-210X.12103

Gray, C., Figueroa, D. H., Hudson, L. N., Ma, A., Perkins, D., & Woodward, G. (2015). Joining the dots: An automated method for constructing food webs from compendia of published interactions. *Food Webs*, *5*, 11–20. https://doi.org/10.1016/j.fooweb.2015.09.001

Herberstein, M. E., McLean, D. J., Lowe, E., Wolff, J. O., Khan, M. K., Smith, K., 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, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., et al. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, *90*(9), 2648–2648. https://doi.org/10.1890/08-1494.1

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

Llewelyn, J., Strona, G., Dickman, C. R., Greenville, A. C., Wardle, G. M., Lee, M. S. Y., et al. (2023). Predicting predator–prey interactions in terrestrial endotherms using random forest. *Ecography*, *2023*(9), e06619. https://doi.org/10.1111/ecog.06619

Maiorano, L., Montemaggiori, A., Ficetola, G. F., O'Connor, L., & Thuiller, W. (2020). TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, *29*(9), 1452–1457. https://doi.org/10.1111/geb.13138

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

Newman, M. E. J. (2010). *Networks. An introduction.* New York, NY: Oxford University Press.

Ohlmann, M., Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L., et al. (2018). Mapping the imprint of biotic interactions on β-diversity. *Ecology Letters*, *21*(11), 1660–1669. https://doi.org/10.1111/ele.13143

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

Poelen, J. H., Simons, J. D., & Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, *24*, 148–159. https://doi.org/10.1016/j.ecoinf.2014.08.005

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., 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. Retrieved from https://www.jstor.org/stable/48582345

Poisot, T., Baiser, B., Dunne, J., Kéfi, S., Massol, F., Mouquet, N., et al. (2016). Mangal – making ecological network analysis simple. *Ecography*, *39*(4), 384–390. https://doi.org/10.1111/ecog.00976

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

Pomeranz, J. P. F., Thompson, R. M., Poisot, T., & Harding, J. S. (2019). Inferring predator–prey interactions in food webs. *Methods in Ecology and Evolution*, *10*(3), 356–367. https://doi.org/10.1111/2041-210X.13125

Pringle, R. M. (2020). Untangling Food Webs. In *Untangling Food Webs* (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

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

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

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., 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., Dalla Riva, G. V., & Poisot, T. (2021). SVD Entropy Reveals the High Complexity of Ecological Networks. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.623141

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

Thuiller, W., Calderón-Sanou, I., Chalmandrier, L., Gaüzère, P., O'Connor, L. M. J., Ohlmann, M., et al. (2024). Navigating the integration of biotic interactions in biogeography. *Journal of Biogeography*, *51*(4), 550–559. https://doi.org/10.1111/jbi.14734

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

Vermaat, J. E., Dunne, J. A., & Gilbert, A. J. (2009). Major dimensions in food-web structure properties. *Ecology*, *90*(1), 278–282. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19294932

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

Xie, J.-R., Zhang, P., Zhang, H.-F., & Wang, B.-H. (2017). Completeness of Community Structure in Networks. *Scientific Reports*, *7*(1), 5269. https://doi.org/10.1038/s41598-017-05585-6