T is for Topology

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

Abstract: 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.

Keywords: food web, network construction

It can be argued that the interaction between species (or individuals) is one of the main determinants of the emergent properties that are studied in other fields of ecology, *e.g.,* the range of plant will be determined by the range of its pollinator, and although the importance of species interactions and the resulting networks that they form has been an acknowledged part of the ecological canon since the penning of the 'entangled bank' [1], the adoption of network ecology into other disciplines of ecology has been limited. This has primarily been driven by two limitations; firstly, it is extremely challenging to actually record species interactions in the field [2,3], which has resulted in a limited coverage of 'real world' interaction data [4], and secondly has been the need to develop terminology and tools that help us to construct, conceptualise, and analyse these networks. Although recording interactions in the field remains a challenge, the development of both practical tools [*i.e.,* tools that help us to record or measure interactions, 5], as well as discussions around the development of tools to predict or infer them [6,7], has allowed us to begin filling in these 'global gaps', albeit in a, potentially, more synthetic manner [8]. Additionally, there has been extensive development in in the ways in which we formalise networks [9,10], and the tools and language that we use to quantify the structure and properties of networks [11]. All together these tools mean that, as a field, network ecology can (and should) be integrated into the broader fields of ecology [*e.g.,* 12] and conservation biology [*e.g.,* 13]. However (as with any new tool or model), it is important that one has a firm grasp of how the underlying philosophy that underpins the construction of networks (particularly synthetic ones) can have an impact on the interpretation of the questions being asked. In this manuscript we will discuss three themes that should help provide clarity and understanding for those wishing to take a step into network (particularly food web) ecology this includes; thinking about and understanding the underlying assumptions that are made when we attempt to delimit and describe a food webs, a synthesis of the different families of tools that are commonly used to construct food webs, and a discussion linking network ecology to some of the outstanding questions in ecology.

[Figure 1 about here.]

1 The anatomy of a food web

Although we specifically focus on food webs (interactions representing feeding links) it is beneficial to take a step back and acknowledge the diversity of form that an interaction network can encapsulate. The idea of an interaction network 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 [14] networks can be constructed at the population (the links between individuals), community (the links between species), or metacommunity (fluxes between locations) level. Even if we are to limit our definition of a network to represent community-level processes there are still many ways to define what is captured by the edges and nodes [insert some e.g.]. It is thus clear that the way that a network is coded (constructed) can influence the resulting observations and conclusions that are made [15,16], and it is important to have a strong grasp of what information a network is attempting to convey.

Even if one were to limit their scope to thinking of interaction networks only in terms of food webs there are still many ways to define the various components of the network, 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?, what is the spatial extent?). All these decisions will have an impact on the resultant structure and potential use-cases of the network.

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 species, the reality is that nodes can often represent an aggregate of different (taxonomic) species - so called 'trophic species', and it is not uncommon that networks can have nodes that represent both taxonomic and trophic species (*e.g.,* there are many that do the basal 'plant/phytoplankton' node but include at least one REF). 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 [mutualism ref, at least there should be one of them], 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 [17] feeding links, potential [18] feeding links, or energy transfer and material flow [19]. How we quantify 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 (*i.e.,* a 'present' interaction is one implies that species a has the ability to consume species b but it does not mean that this interaction is realised in the field) will be meaningless if you are interested in understanding the flow of energy through the system as the links within the network are over connected. 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, 20] or continuous measurements [which quantifies the effect of one species on another, 21]. Although there is a clear argument for moving away from a purely binary way of representing interactions [probabilities preprint] this of course also means that there is an additional layer to the interpretation these links.

1.3 Aggregating networks

Here I think we need to talk about realised vs potential links (i.e. the concept of a metaweb) but also the idea that we are often aggregating over time and space which makes boundaries and whatnot all a bit fuzzy

> [22] states that *"[Their] approach is more like gross anatomy than like physiology… that is, the gross anatomy is frozen, rather than in motion."*.

1.4 Putting the parts together; what does it mean?

It it clear that there are many ways to define, code, and construct food webs, however what may be less clear is understanding *why* there is such a diversity of thought. Here it may be meaningful to contextualise the different 'types' of food webs within the larger questions (or needs) that have been driving them. Some of the earliest work on food webs was linked to the idea of niche space, and more specifically, the idea of trophic niches and how this would influence the dimensionality of a networks [23]. This introduced the idea that a single dimension [the "niche axis," 24] constrains the interactions between species; in this instance it makes sense to think of species in terms of what they consume and what they are consumed by, as they are occupying the same space in the niche axis. Networks that are defined in this way may be useful for understanding how the flow of energy (resources) are constrained between 'species', particularly how it moves through the trophic levels. This 'niche-based' way of thinking might be beneficial when thinking about networks at the structural level, and when trying to map large-scale processes [ref?] however there was also a need to develop ways of thinking that were more geared to thinking about why does species *a* predate species *b*, broadly this is the result of two things; a predator needs to have the correct traits to be able to capture, kill, and consume, its prey (a mismatch between predator and prey is termed a forbidden link, [3]) and it needs to be energetically feasible [feeding ecology ref]. When we think of interactions in these terms it makes sense that nodes are defined at the species level (or at least as species that have the same traits and/or energy content), however the links between them can be quantified in different ways… [this is lazy writing]

something, something, introducing that the same problem (different philosophies) is also a thing that you need to think about when aggregating interactions/generating networks.

2 Why do we want to predict food webs?

Arguably the need for methods and tools that can be used to construct synthetic food webs 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 understanding some higher-level process/concept (*e.g.,* understanding energy flows) in a more synthetic setting, whereby these networks do not require any level of species specificity *per se* and it is more the arrangement of the nodes and links 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 interactions in the field (see [6] for a more mechanistic, and [7] for a more statistical overview of ways to approach this specific issue). 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). 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 they desire.

3 How do we predict food webs?

> maybe a more direct link here to the fact that when working with networks its often synthetic ones *i.e.,* the product of some sort of modelling exercise; alternatively there has also been a push to develop predictive tools to create hypothetical (but plausible) networks for real world situations. Also talk about even deciding to create a network from field observations is in and of itself still a 'model' that has assumptions… for example decisions are made about delimiting, aggregation, and observation, the idea of aggregating over time or aggregating over space. Same can e said for different food web generating tools , they have their own underlying rules and assumptions that are made when constructing a food web, which will determine and influence the resulting structure or inferred interactions [25]

Although there have been efforts to compare and contrast different models [27] there still lacks an overall synthesis as to how the different model families differ from each other - both in terms of what they are actually predicting as well as how well they are preforming in the different facets of constructing a food web.

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 highlights the differences and similarities of the philosophies and assumptions that determine a network. It should be noted that although we provide some examples of specific use cases within each model family this by no means an exhaustive list of of all the different approaches ever used but rather a representative collection of some of the more canonical approaches used within each model family.

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	Network structure is random	structure	agnostic	feeding links	binary	
neutral	Network structure is random, but species abundance plays a role	structure	species	feeding links	binary	
resource	Networks are interval, species can be ordered on a 'niche axis'	structure	trophic species	subdivision of resource	binary	[26]

Model family	Theory	Network predicted	Nodes represent	Links represent	Interaction	Key reference
generative	Networks are determined by their structural features	structure	agnostic	links	binary	
energetic	Interactions are determined by foraging theory (feeding links)	interactions	species	feeding links	quantitative	
graph embedding	Interactions can be predicted from the latent traits of networks	interactions	species	potential feeding links	probabilistic	[28]
trait matching	Interactions can be inferred by a mechanistic framework/relationships	interactions	species	feeding links	binary	[6]

Model family	Theory	Network predicted	Nodes represent	Links represent	Interaction	Key reference
binary classifiers	Interactions can be predicted by learning the relationship between interactions and ecologically relevant predictors	interactions	species	feeding links	binary	[27]
expert knowledge	'Boots on the ground' ecological knowledge and observations	interactions	species	feeding links	binary	
data scavenging	Webscraping to create networks from online databases	interactions	species	feeding links	binary	

Model family	Theory	Network predicted	Nodes represent	Links represent	Interaction	Key reference
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	

[Figure 2 about here.]

Source: Model family traits

3.2 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 two broader criteria, namely the ability of the model to correctly capture different elements of the structure of the network and the ability of the model to correctly retrieve 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. For example if we are concerned with being able to successfully predict pairwise interactions we want to ensure that we are able to retrieve interactions that really exist but also those that cannot exist (*sensu* forbidden links [3])

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 [24]. 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 [29].

benchmarking requires the use of empirical networks and comparing that to the predicted one

3.2.0.1 Benchmarking for structure

Despite structural models being some of the older model families there is a distinctive lack of clear guidelines as to how we assess the ability of these models to replicate the *entire* structure of a network. In part this may perhaps be driven by the underlying research agenda and interest in different aspects of capturing the structure of networks *e.g.,* the obsession with intervality [ref] or link distributions [ref]. However, it is still a good idea to think about the network in its entirety and to benchmark structural models in a more holistic manner. Some useful ways to assess how well the model predicts the shape (*e.g.,* the height (chain length) and…), links (*e.g.,* connectance), internal structure (*e.g.,* SVD entropy, [30]), and meso-level features (*e.g.,* motifs, [31]) of a network. This is shown in Figure 3…

- Maybe look at some of the historic papers that compare some of the 'resource models'
- See also [24] and the likelihood function that they use for model selection
- Look at [32]

[Figure 3 about here.]

Source: Quantitative approach to topology generators

3.2.0.2 Benchmarking for interactions

Broadly speaking the task of assessing the ability of a model to predict interactions as being an assessment of the model's classification ability (does it correctly predict the presence and absence of interactions?) and so we want to benchmark the model on how well it is able to correctly predict these presences and absences. This can be done in a myriad of ways [7,29] but is always based off of the confusion matrix [ref maybe?]. Essentially the 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).

Using the confusion matrix it is then possible to assess the 'quality' of the model predictions such as their accuracy or informedness.

As mentioned above one of the main challenges we are faced with when trying to benchmark interaction predictions is the high class imbalance (inherit sparsity) of networks, and as highlighted by [29] we can very easy to lull ourselves into a false sense of predictive accuracy if we use the wrong benchmarking tools — even a low skill (fails to predict interactions that are present) model can appear to do well if we assess it on its ability to correctly predict interactions, this is because most interactions are absent and so a model that predicts interactions as being absent will still predict most interactions correctly (*i.e.,* getting the 'right' answers but for the wrong reasons). Another aspect of assessing these types of predictions is quantifying the bias of the model, this will give an indication if the model tends to systematically over predict one of the classes. As per [29] the best ways to assess the classification performance of the different models is to use the Precision-Recall (PR-AUC) to assess precision [ref?], and the Matthews correlation coefficient (MCC) to assess accuracy [33].

- 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
- "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.,* introducing false negatives into the training data *sensu* what we did in [34]

[Figure 4 about here.]

3.3 The bigger picture

In addition to thinking about the 'performance' if a model it is also important to be aware of the 'unseen' costs and limitations of the different modelling families. What data do I need? Can a make *de novo* predictions? What are the related 'sinks' *e.g.,* computational or time? What does the network I am constructing actually represent?

3.3.1 Data need vs availability

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.

3.3.2 Theory vs 'real world'

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* [18]) 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 [35] that discuss how the local factors are going to play a role.

3.3.3 The target system?

3.3.4 Philosophy limits theory

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 [36] - 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).

4 Discussion

- 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 [5,17], 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.

- 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).

Interestingly [26] 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." - [5]
> 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)

- 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 [36] might have (and share) some thoughts on this (thanks Andrew). I feel like I need to look at [37] but maybe not exactly in this context but vaguely adjacent.

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

- 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.

  - I think this is sort of the crux of the argument presented in [38]

"we highlight an interesting paradox: the models with the best performance measures are not necessarily the models with the closest reconstructed network structure." - [29]

- *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." - [7]
  - Aligning (dove-tailing) with this the idea of ensemble modelling as presented by [39]
- 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)

4.1 Downsampling

- [40]
- "That being said, there is a compelling argument for the need to 'combine' these smaller functional units with larger spatial networks [41] and that we should also start thinking about the interplay of time and space [42]. Although deciding exactly what measure might actually be driving differences between local networks and the regional metaweb might not be that simple [43]."

Glossary

Term	Definition
food web	a representation of feeding links between species

Term	Definition
model	A tool that can be used to construct networks, 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
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 betweens 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…

Box 1 - A text box

Nonsense outro.

References

1. Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, J. Murray
2. Jordano, P. (2016) Chasing Ecological Interactions. *PLOS Biology* 14, e1002559
3. Jordano, P. (2016) Sampling networks of ecological interactions. *Functional Ecology* DOI: 10.1111/1365-2435.12763
4. Poisot, T. *et al.* (2021) Global knowledge gaps in species interaction networks data. *Journal of Biogeography* n/a

1 5. Pringle, R.M. and Hutchinson, M.C. (2020) Resolving Food-Web Structure. *Annual Review of Ecology, Evolution and Systematics* 51, 55–80

2 6. Morales-Castilla, I. *et al.* (2015) Inferring biotic interactions from proxies. *Trends in Ecology & Evolution* 30, 347–356

3 7. Strydom, T. *et al.* (2021) A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 20210063

4 8. Poisot, T. *et al.* (2016) Synthetic datasets and community tools for the rapid testing of ecological hypotheses. *Ecography* 39, 402–408

5 9. Dale, M.R.T. and Fortin, M.-J. (2010) From Graphs to Spatial Graphs. *Annual Review of Ecology, Evolution, and Systematics* 41, 21–38

6 10. Fortin, M.-J. *et al.* (2012) Spatial statistics, spatial regression, and graph theory in ecology. *Spatial Statistics* 1, 100–109

7 11. Delmas, E. *et al.* (2019) Analysing ecological networks of species interactions. *Biological Reviews* 94, 16–36

8 12. Thuiller, W. *et al.* (2024) Navigating the integration of biotic interactions in biogeography. *Journal of Biogeography* 51, 550–559

9 13. Bhatia, U. *et al.* (2023) Network-based restoration strategies maximize ecosystem recovery. *Communications Biology* 6, 1–10

10 14. Poisot, T. *et al.* (2016) Describe, understand and predict: Why do we need networks in ecology? *Functional Ecology* 30, 1878–1882

11 15. Proulx, S.R. *et al.* (2005) Network thinking in ecology and evolution. *Trends in Ecology & Evolution* 20, 345–353

12 16. Brimacombe, C. *et al.* (2023) Shortcomings of reusing species interaction networks created by different sets of researchers. *PLOS Biology* 21, e3002068

13 17. Pringle, R.M. (2020) Untangling Food Webs. In *Untangling Food Webs*, pp. 225–238, Princeton University Press

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

15 19. Lindeman, R.L. (1942) The Trophic-Dynamic Aspect of Ecology. *Ecology* 23, 399–417

16 20. Poisot, T. *et al.* (2016) The structure of probabilistic networks. *Methods in Ecology and Evolution* 7, 303–312

1 21. Berlow, E.L. *et al.* (2004) Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology* 73, 585–598

2 22. Cohen, J.E. *et al.* (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, 421–448

3 23. Cohen, J.E. (1977) Food webs and the dimensionality of trophic niche space. *Proceedings of the National Academy of Sciences* 74, 4533–4536

4 24. Allesina, S. *et al.* (2008) A General Model for Food Web Structure. *Science* 320, 658–661

5 25. Petchey, O.L. *et al.* (2008) Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences* 105, 4191–4196

6 26. Williams, R.J. and Martinez, N.D. (2008) Success and its limits among structural models of complex food webs. *Journal of Animal Ecology* 77, 512–519

7 27. Pichler, M. *et al.* (2020) Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution* 11, 281–293

8 28. Strydom, T. *et al.* (2023) Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations. *Methods in Ecology and Evolution* 14, 2917–2930

9 29. Poisot, T. (2023) Guidelines for the prediction of species interactions through binary classification. *Methods in Ecology and Evolution* 14, 1333–1345

10 30. Strydom, T. *et al.* (2021) SVD Entropy Reveals the High Complexity of Ecological Networks. *Frontiers in Ecology and Evolution* 9

11 31. Stouffer, D.B. *et al.* (2007) Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences* 274, 1931–1940

12 32. Vermaat, J.E. *et al.* (2009) Major dimensions in food-web structure properties. *Ecology* 90, 278–282

13 33. Matthews, B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405, 442–451

14 34. Strydom, T. *et al.* (2022) Food web reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution* 13, 2838–2849

15 35. Poisot, T. *et al.* (2015) Beyond species: Why ecological interaction networks vary through space and time. *Oikos* 124, 243–251

16 36. Petchey, O.L. *et al.* (2011) Fit, efficiency, and biology: Some thoughts on judging food web models. *Journal of Theoretical Biology* 279, 169–171

17 37. Berlow, E.L. *et al.* (2008) The "Goldilocks factor" in food webs. *Proceedings of the National Academy of Sciences* 105, 4079–4080

38. Brimacombe, C. *et al.* (2024) Applying a method before its proof-of-concept: A cautionary tale using inferred food webs

39. Becker, D.J. *et al.* (2022) Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *The Lancet Microbe* 3, e625–e637

40. Dansereau, G. *et al.* (2023) Spatially explicit predictions of food web structure from regional level data

41. Fortin, M.-J. *et al.* (2021) Network ecology in dynamic landscapes. *Proceedings of the Royal Society B: Biological Sciences* 288, rspb.2020.1889, 20201889

42. Estay, S.A. *et al.* (2023) Editorial: Patterns and processes in ecological networks over space. *Frontiers in Ecology and Evolution* 11

43. Saravia, L.A. *et al.* (2022) Ecological network assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology* 91, 630–642

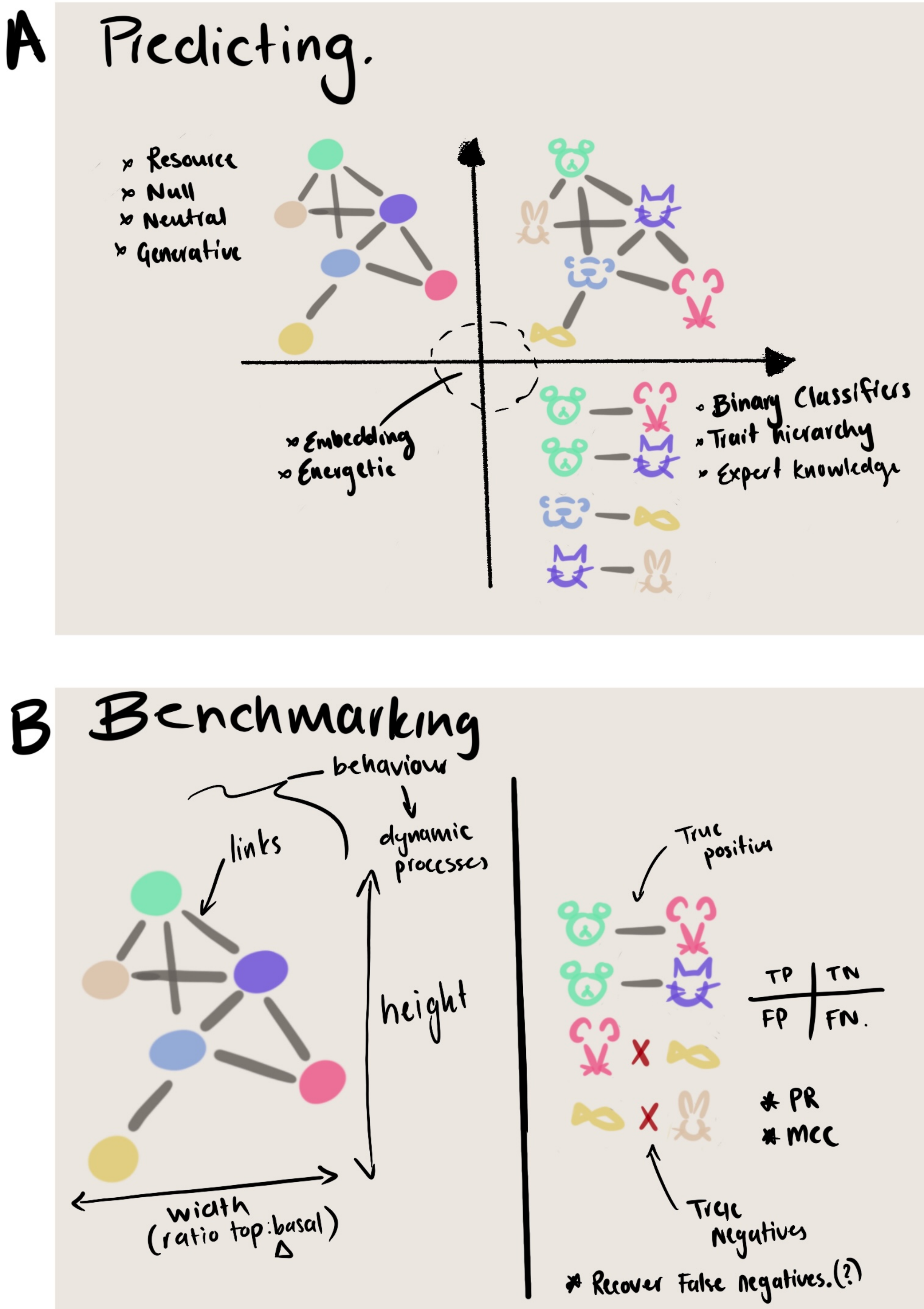

Figure 1: Conceptual figure of the ‘network prediction’. Panel A shows where the model families fall in the the context of being models that predict networks or models that predict interactions space. Panel B serves to highlight the characteristics one might like to ‘test’/benchmark for a model based on it being either a network or interaction predicting model

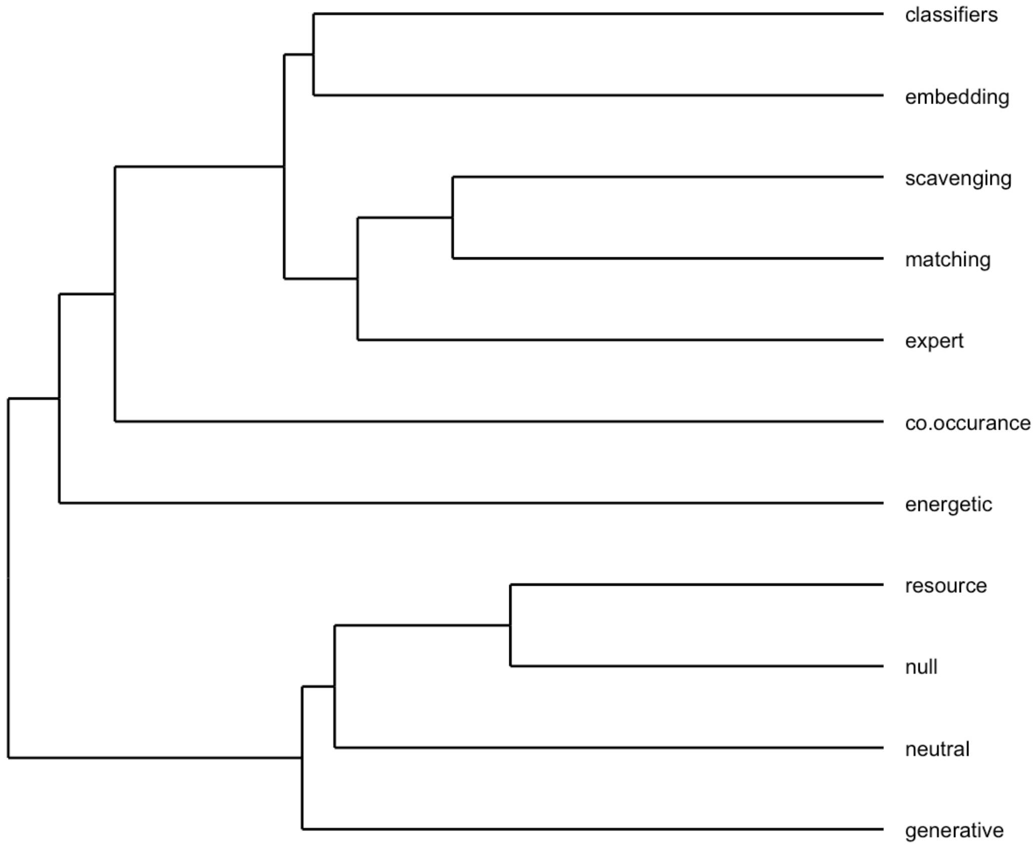

Figure 2: Dendrogram of the trait table

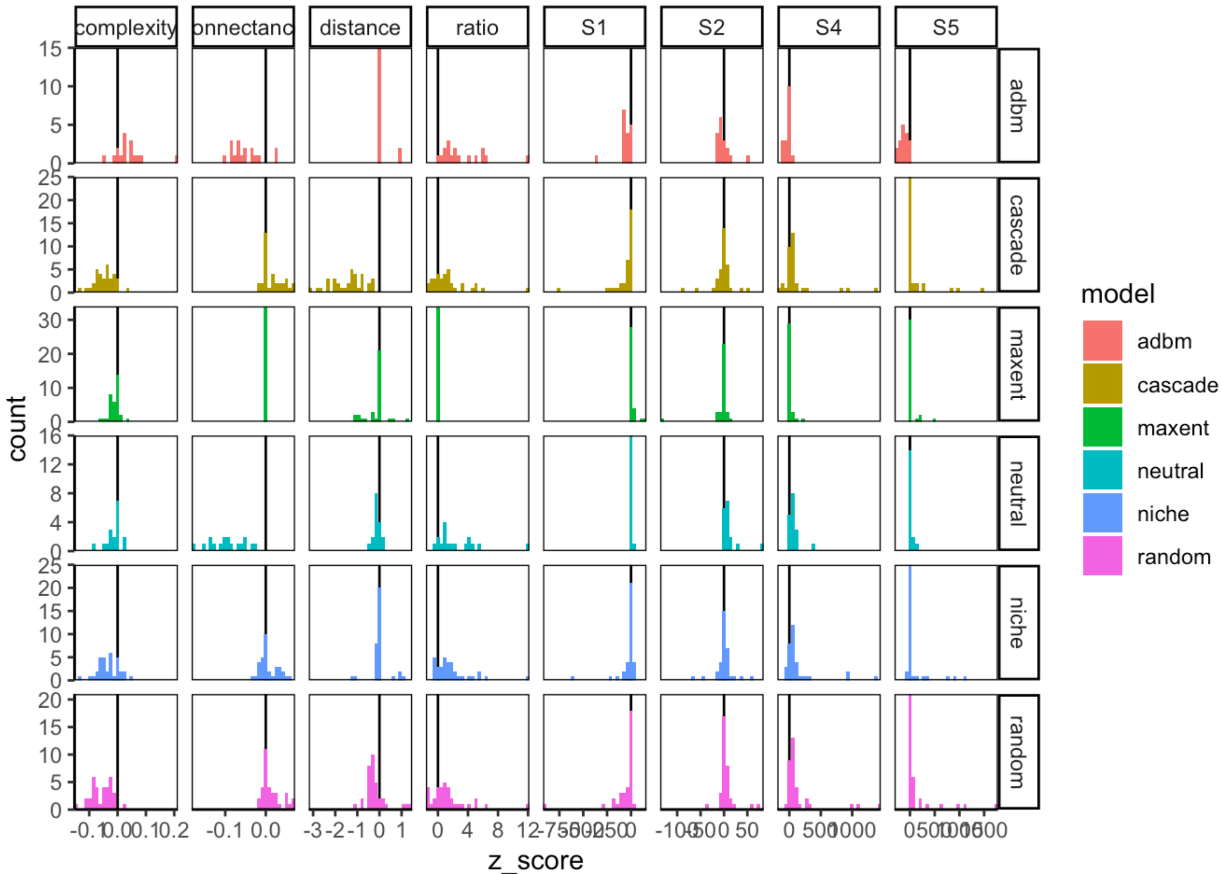

Figure 3: Difference between real and model network property. S1 - S5 represent the different motif structures identified in [31].

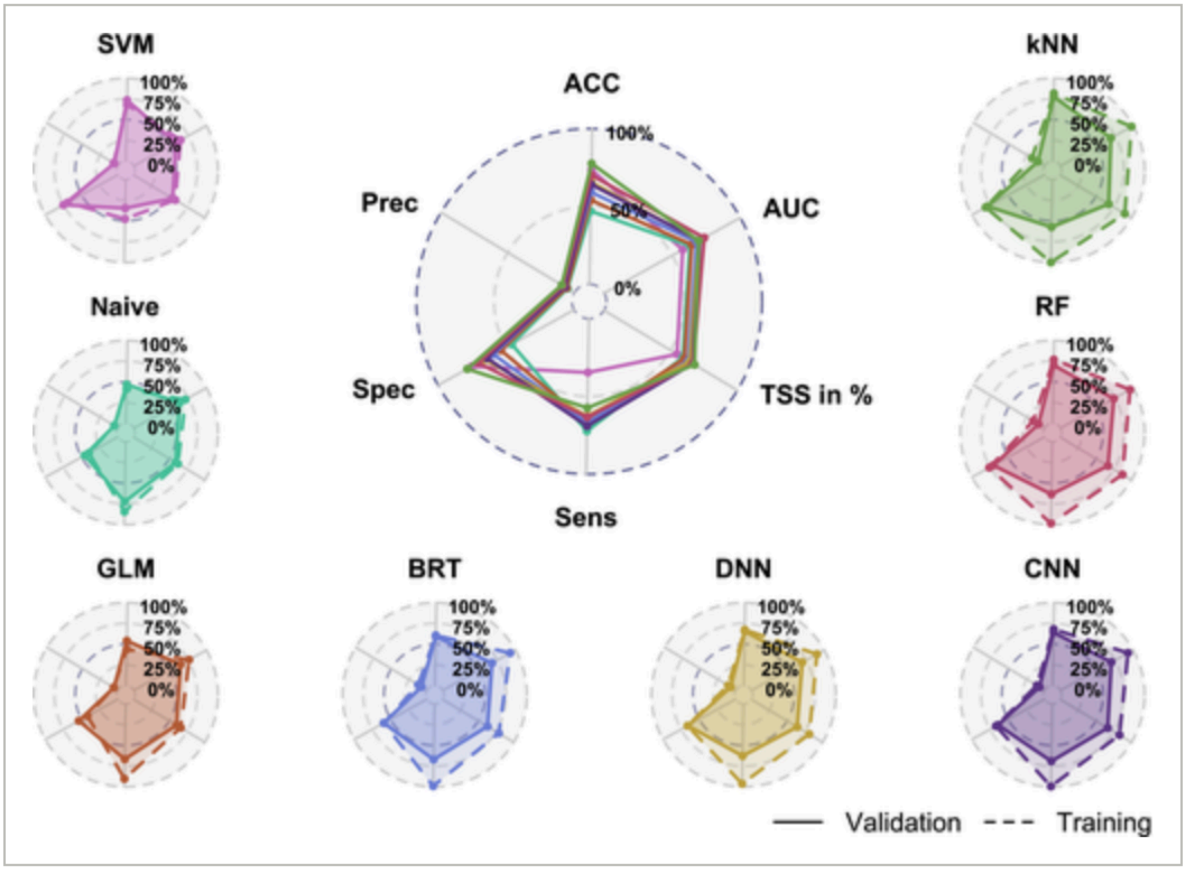

Figure 4: Moc result from [27]