Methods for Reconstructing Paleo Food Webs

Tanya Strydom

Baran Karapunar

Andrew P. Beckerman

Alexander Dunhill

2025-11-21

Abstract

Food webs represent the feeding relationships between species and can help infer ecosystem-level processes. Alongside the development of food web theory, methods for constructing food webs have been developed to infer species interactions when empirical data is lacking. Food web construction methods are diverse, each utilising different approaches to infer species interactions —such as the use of traits to infer mechanistic relationships vs using gut content as a proxy for species diets. These methods have distinct theories, mechanisms, and data requirements. In paleoecology, where direct evidence of feeding interactions are rare, food web construction methods are especially valuable and affords us the opportunity to make inferences about paleo communities beyond simply a record of species composition. However, the limitations of paleontological data (e.g., information of species traits is limited to that which can be preserved) restrict which methods can reliably be used. By considering both ecological theory and the constraints of what can be derived from the fossil record, we identify the methods best suited for the construction of paleo food webs. Specifically, we focus on how these methods differ in the networks they produce and what these networks can reveal about species interactions. In doing so we hope to clarify the ecological nuances of network prediction and help prevent the accidental misuse or misinterpretation of paleo food webs.

There has been a growing interest in understanding community responses to environmental changes across deep time events as a means to help understand current and future biodiversity changes [1,2]. Species interactions and the resulting networks have gained popularity in contemporary settings as a means to help us to understand aspects of community composition and biodiversity (eg [3] and ??) and so it is perhaps unsurprising that there has been a growing interest in using paleo food webs in a similar manner [*e.g.,* 4 looked at…,5 looked at…,6 looked at…]. However, one of the core challenges and limitations of being able to effectively *use* food webs is the challenge of *creating* them [7], although this is a challenge within contemporary settings it is compounded in paleo contexts where, in the absence of being able to observe interactions, we are dependent on the fossil record (and the inherent limitation it imposes) to infer interactions. As a way to address the challenges with recording species interactions there has been the development of a large number of models and tools that can be used to infer either species interactions [see *e.g.,* 8,9,10 for broader reviews] or networks [see *e.g.,* 11]. Although there has been the development of models and tools that are specific for inferring paleo food webs [*e.g.,* 12,13,14], it should be noted that these models only occupy a subset of the broader family of approaches that are used to predict networks, as they typically only focus on assessing the feasibility of interactions between species. Being able to construct only one ‘type’ of network means that we are limited in the scope of questions that we can appropriately answer with those networks [see Strydom in prep; [15]]. However, there is scope that models and tools that have been developed in contemporary settings have the potential to be used for paleo settings [*e.g.,* 6], which opens the door for researchers to ask a broader and more complete range of questions about community responses to environmental change.

Here we aim to provide an overview of the different models that can be used to construct food webs using paleo data. Specifically we focus on identifying a suite of models that are appropriate for use with paleo data that can feasibly be constructed within the limitations that are imposed by fossil data while still spanning the larger network space. Additionally we use the data from [4] as a case study to understand how different models recover different networks, both in terms of structure as well as pairwise interactions and establish if there are consequences for using networks that are constructed using different models in terms of making inferences about the behaviour of the system by looking at how the model type influences what we infer to be the dominant driver of extinctions across a mass extinction event.

# 1. Constructing paleo webs

# 2. Challenges specific to building paleo networks

Although there has been a push for the development of tools and methods that allow us to predict species interactions and networks they will not all be suitable for the prediction of paleo communities. This is primarily due to limitations that we are faced with in terms of the information that can be inferred from the fossil record (such as species traits, abundances, and assemblages), which is needed as input data for the different models. The limited information available from the fossil record is compounded by the incomplete and biased preservation of species [REF], which part of a species is preserved (part vs whole), the ambiguity of the ‘true’ community composition [were communities conserved *in situ* or were they there owing to geological processes?; REF], as well as the availability/accessibility of different rock layers (and thus the completeness of data we might have for a specific era in time). Additionally there is an increasing degree of ‘fuzziness’ around the ecology and life histories of species the further one moves back in geological time [REF]. This is not to say that because we have imperfect data we should not be attempting to construct paleo food webs but rather that we need to be aware of what the uncertainties are and how these might impact the assumptions that we need to make when constructing a network (as well as how this will intersect with the intended end use of the network). This will allow us to best identify an approach that minimises the assumption and potential uncertainties within the data while still constructing a suitable network. This includes thinking about both assumptions you are making about the actual data *e.g.,* trying to extrapolate body size from fossil data but also assumptions across time *e.g.,* assuming modern trait-feeding modes are the same or that assumptions about network structure will hold across deep time.

## 2.1 Understanding the approaches to network construction

Broadly we can think about network construction as being nested within two different schools of thought (and thus methodological approaches, [Figure 1](#fig-concept)), models that focus on assessing the *mechanistic* feasibility of an interaction being able to occur between two species or models that are more closely married to specific bodies of ecological *theory* - such as niche theory or foraging ecology. The former of which will construct ‘metawebs’ and the latter ‘realised networks’ [Strydom et al in prep]. Models that have specifically been developed in the paleo space tend to be mechanistic in nature in that they focus on using a trait-based approach to formalise feeding interactions (*e.g.,* [12]; [13]), are assembled by expert opinion (*e.g.* [16]), or make assumptions based on the evolutionary signals of interactions (*e.g.,* [14]). Thus paleo models typically only construct metawebs, and there is the need for the intentional adoption of theoretical models if we want to realise the full potential of questions and information that we can glean from the fossil record. However, there is an argument that the fundamental ‘currencies of life’ to have remained constant - *e.g.,* the energetic constraints of foraging or foraging niches, meaning that theoretical models that have been developed and tested on contemporary food webs should still hold for paleo communities.

|  |
| --- |
| Figure 1: This obviously needs work but a variation on this to try and articulate the different approaches and broadly how they may differ. |

Here we present six different models ([Table 1](#tbl-models)) that can be used to construct food webs for both this specific community but are also broadly suited to paleo network prediction. These models span all facets of the network representation space (metaweb, realised, and structural network) and are suitable for an array of different paleo communities as the data requirements fall within the limitations set by the fossil record.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: A summary of the different families of tools that can be used to generate paleo food webs.   | Model family | Assumptions | Data needs | ‘Limitation’ | Network type | Key reference | | --- | --- | --- | --- | --- | --- | | random | Links are randomly distributed within a network | richness, number of links | parameter assumptions, species agnostic | structural network | [17] | | niche | Networks are interval, species can be ordered on a ‘niche axis’ | richness, connectance | parameter assumptions, species agnostic | structural network | [18] | | allometric diet breadth model (ADBM) | Interactions are determined by energetic costs (foraging ecology) | body mass, biomass (abundance) | does not account for forbidden links in terms of trait compatibility, assumptions on body size and biomass (abundance) from fossil data | theoretical network | [19] | | l-matrix | Interactions inferred using allometric rules (ratio of body sizes between predator and prey), with links being constrained by a Ricker function | body mass, number of producer species | does not account for forbidden links in terms of trait compatibility, assumptions on body size from fossil data, assumptions as to the number of producer species | theoretical network | [20] | | paleo food web inference model (PFIM) | Interactions can be inferred by a mechanistic framework/relationships | feeding traits for taxa, mechanistic feeding rules | Assumption made as to the feeding mechanisms, need to elucidate traits from models (although this is a way smaller issue) | mechanistic web | [12] | | body size ratio model | Interactions inferred using allometric rules (ratio of body sizes between predator and prey). Logit of the linking probability used to further constrain links to an ‘optimal size range’ for prey. | body mass | does not account for forbidden links in terms of evolutionary compatibility, assumptions on body size from fossil data | theoretical network | [21] | |

# 3. Case study: Toarcian mass extinction event

## 3.1 Dataset overview

### 3.1.1 Species occurrence

Here we use the fossil occurrence data over an interval extends from the upper Pliensbachian (~185 Ma) to the upper Toarcian (~175 Ma) of the Cleveland Basin [see 4 for a more comprehensive overview]. The data set consists of a subset of four broad time periods (pre-extinction, post-extinction, early recovery, and late recovery). The assemblages are treated as communities of interacting organisms. Something about the total number of taxa as well as numbers per a time period? Probbaly also make a comment that this is a ‘deep time’ community we are looking at.

### 3.1.2 Defining modes of life (traits)

We used the modes of life (traits) as identified in [4], who defined four traits: motility (fast, slow, facultative, non-motile), tiering (pelagic, erect, surficial, semi-infaunal, shallow infaunal, deep infaunal), feeding (predator, suspension feeder, deposit feeder, mining, grazer), and size: gigantic (>500 mm), very large (>300–500 mm), large (>100–300 mm), medium (>50–100 mm), small (>10–50 mm), tiny (≤10 mm), for each fossil species based on the ecological traits defined in the Bambach ecospace model [22].

### 3.1.3 Constructing networks

For each paleo community (time bin) we constructed **100** networks for each model (so 6 \* 100) networks. These networks were simplified so as to remove any disconnected species. In total 2 400 networks were constructed. When a quantitative measure of body size is needed (ADBM, body size ratio, and l-matrix) we drew a body mass for each species from a uniform distribution, with ranges being defined by the different size classes *e.g.,* a species classed as ‘very large’ would have a body mass drawn from . This was repeated for each run in order to add variation to the networks constructed, however the same body sizes were kept consistent for the relevant models *i.e.,* an ADBM and l-matrix network from the same replicate have the same bodysizes. For both the random and niche model the desired connectance was randomly selected between the range 0.07 - 0.15 for each replicate but kept consistent for both models. For each network we calculated the properties listed in [Table 2](#tbl-properties)

## 3.2 Models capture different network structure but in unexpected ways

Broadly when we talk about quantifying the structure of a network we are interesting in capturing some aspect of how the links are distributed between nodes, or alternatively about properties of the nodes. Structure is useful as it is gives information as to how the interactions between species are distributed within the community, informing us on *e.g.,* energy flows and fluxes [REF], propagation of stress [REF], and something about trophic levels [REF]. We are also able to glean information on interaction strategies between smaller interacting units in the bigger community in the form of motifs [23,24]. Motifs allow us to identify *e.g.,* the prevalence of competition, as well as smaller chains within the network. Node-level properties look at the the number of links coming in to (prey) or out of (predators) a node and are informative of diet specialisation.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Network properties used for analysis.   | Metric | Definition | Scale | Reference (for maths), can make footnotes probs | | --- | --- | --- | --- | | Richness | Number of nodes in the network | Macro |  | | Links | Normalized standard deviation of links (number of consumers plus resources per taxon) | Micro |  | | Connectance | , where is the number of species and the number of links | Macro |  | | Max trophic level | Prey-weighted trophic level averaged across taxa | Macro | [25] | | Diameter | Diameter can also be measured as the average of the distances between each pair of nodes in the network | Macro | [26] | | Complexity | SVD complexity of a network, defined as the Pielou entropy of its singular values | Macro | [10] | | Redundancy | , where is the number of species and the number of links. Indicates the number of edges beyond what is needed for a minimum-connected tree | Macro |  | | S1 | Number of linear chains, normalised | Meso | [23]; [24] | | S2 | Number of omnivory motifs, normalised | Meso | [23]; [24] | | S4 | Number of apparent competition motifs, normalised | Meso | [23]; [24] | | S5 | Number of direct competition motifs, normalised | Meso | [23]; [24] | | Generality | Normalized standard deviation of generality of a species standardized by | Micro | [27] | | Vulnerability | Normalized standard deviation of vulnerability of a species standardized by | Micro | [27] | |

In terms of wanting to asses and compare across the different models it is beneficial to approach this task by thinking about the different aspects of the network as well as interactions that are being predicted by the different models across different ‘scales’ of organisation within the network, namely macro (the entire network), meso (smaller interacting units within the network), and micro (species-level attributes). Although there are a myriad of possible ways to ‘measure’ and analyse ecological networks [26] we have selected those outlined in [Table 2](#tbl-properties) as they span different scales within thr network and have been shown to be informative of different ecological processes.

Here we used a Multivariate Analysis Of Variance (MANOVA) to assess the differences between networks generated by different models based on the combined information of the multiple structural network measures. Model defined as network structure values ~ model additionally we used a Linear Discriminant Analysis (LDA) to determine if different models produced networks with differing structure. In order to do the MANOVA and LDA we had to create within model variation for the different networks, with the exception of the PFIM model all models have some inherent variation. In order to generate variation within the PFIM metawebs we applied a *minimal* degree of downsampling following the protocol described in [28]. This downsampling approach uses a power law distribution to essentially ‘prune’ links from the most generalist species (See SUPP MATT for a more detailed overview).

The multivariate effect of model was statistically significant, Pillai’s Trace = 3.89, F(45, 11 950) = 925.64, p < .001, indicating systematic differences across multiple ecological or network properties simultaneously. Follow-up univariate ANOVAs revealed that model type had significant effects on all nine dependent variables [Table 3](#tbl-manova), show that the network structure differed markedly across the model types on every measured dimension. Model type accounts for the vast majority of variance in most network metrics (66%–92%), indicating profound differences in structure between models. The only exception is trophic level (η² = .19), which still shows a large effect but is much smaller relative to the other metrics.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Manova/univariate ANOVA results.   | Metric | F(df = 5, 2394) | df | partial η² | | --- | --- | --- | --- | | Connectance | 2717.8 | 5 | 0.85 | | Complexity | 2356.6 | 5 | 0.83 | | Max trophic level | 108.86 | 5 | 0.19 | | Generality | 5646 | 5 | 0.92 | | Vulnerability | 3266.9 | 5 | 0.87 | | S1 | 1968.5 | 5 | 0.80 | | S2 | 1527.5 | 5 | 0.76 | | S4 | 940.79 | 5 | 0.66 | | S5 | 1919.4 | 5 | 0.80 | |

Post-hoc pairwise comparisons using Tukey-adjusted estimated marginal means further clarified the differences among models. The PFIM differed significantly from all other models (all p < 0.001). The niche and random models are similar to each other, and the adbm and lmatrix were also similar to each other [Figure 2](#fig-marginal). This although there are clear structural difference between the models as a whole we still broadly see the grouping between the theoretical models (ADBM, l-matrix), structural models (Niche and Random), and a metaweb (PFIM). Although the bodymass ratio model deviates from this neat grouping it is perhaps not as surprising since this simplified version of the bodymass ratio model is strongly rooted in the niche-based processes that are also assumed in the Niche model and so it having some overlap with the other structural models is not that surprising.

|  |
| --- |
| Figure 2: Estimated marginal means (EMMs) of ecological network metrics across six model types with 95% confidence intervals. Bars represent the predicted values for each model, and error bars indicate the 95% confidence limits. Letters above each bar denote Tukey-adjusted pairwise significance: models sharing the same letter are not significantly different, while models with different letters are significantly different (p < 0.05). The plot reveals three tiers of model performance, with pfim consistently higher, log ratio, niche, and random at intermediate levels, and adbm and lmatrix lower, consistent with the MANOVA and post-hoc analyses. |

From the LDA the first two discriminant functions explained 72% and 18% of the variance, respectively. Wilks’ λ indicated that the discriminant functions significantly differentiated among models (λ = 0.12, χ² = 1024, p < 0.001). The LDA plot [Figure 3](#fig-structure) shows clear separation of the pfim model from the others along LD1, with adbm and lmatrix clustering closely together, and niche and random occupying intermediate positions. Classification accuracy was 85%, confirming that the combination of dependent variables reliably distinguishes model types.

|  |
| --- |
| Figure 3: Linear discriminant analysis (LDA) of ecological network metrics for six model types. The first two discriminant functions (LD1 and LD2) explain 72% and 18% of the variance, respectively. Each point represents a replicate, and ellipses indicate 95% confidence regions for each model. The PFIM model is strongly separated along LD1, reflecting the highest values of network metrics, while adbm and lmatrix cluster closely together, indicating similar, lower metric values. The niche and random models occupy intermediate positions. Classification accuracy of the LDA was 85%, demonstrating that the combination of dependent variables effectively discriminates among model types. |

The implications of the above results is that it is clear that different models will recover different structures - across all structural measures and highlight how model selection has the potential to strongly shape ecological inferences. Using a model that overestimates connectivity could exaggerate our inferences about redundancy or disturbance risk, while overly sparse models could underestimate network complexity and functional links. Therefore, the choice of model should align with the specific ecological question *e.g.,* in interest in exploring *potential* redundancy, robustness, versus trying to understand *realistic* energy flow pathways. Ideally we should couple our analyses with sensitivity analyses to assess how conclusions depend on model assumptions. It also means that we cannot compare inferences made using different models but and any generalisations about observed patterns should be be standardised across network model *type* at minimum. That is two say it may not be completely illogical to make comparisons between two metawebs, however it would be unwise to compare a metaweb to a theoretical network.

These structural differences have consequences for predicting species persistence, stability, and ecosystem functioning. For example, metrics like generality and vulnerability influence top-down and bottom-up control, affecting how energy and biomass flow through trophic levels (Dunne et al., 2002). Similarly, connectance and trophic coherence influence stability and resilience; more coherent networks tend to resist perturbations, whereas very dense, highly connected webs may either buffer or amplify disturbances depending on interaction strengths (Johnson et al., 2014).

## 3.3 Some networks don’t share any interactions and some share a lot

In addition to wanting to measure network structure researchers may also be interested in understanding aspects about the diets and predators of *specific* species in a community. In this instance the interest should be in understanding how the pairwise links predicted between species pairs differ between models. Here we look at the interaction turnover both within and between the different models [29]. This can be thought of as the equivalent of species turnover and tells us which interactions are ‘conserved’ (shared) across the networks but only between species pairs that are shared - *i.e.,* this turnover is only driven by interaction and not species turnover. Here we only compared networks that we constructed for the same period (as our interest is only in between model differences) and excluded the random and niche networks from consideration as these two models are essentially species agnostic.

Across the four network models, turnover in species interactions varied substantially, revealing clear differences in how each model appraoches determining the presence of links between species pairs [Figure 4](#fig-beta_div). The log-ratio model consistently showed high turnover relative to all other approaches, indicating that it produces interaction pairs that are the most distinct from other models. In contrast the ADBM and l-matrix exhibited the lowest turnover and suggests strong agreement between how pairwise interactions are determined. This is unsurprising given the underlying inference mechanisms of the models. The PFIM displays an intermediate turnover, aligning most closely with log-ratio and least with ADBM and l-matrix. Although this result is unsurprising as the mechanisms that determine interactions in ADBM and l-matrix (a single trait (bodysize) + paramterisiation of links by ecological theory) is very different from the PFIM model that makes assumptions on a trait-based, mechanistic hierarchy. Taken together, these results demonstrate that model choice strongly influences inferred pairwise interactions.

|  |
| --- |
| Figure 4: Pairwise beta turnover in species interactions among four ecological network models (adbm, lmatrix, log-ratio, and pfim). Each cell represents the mean turnover value between a pair of models, with warmer colors indicating greater dissimilarity in inferred interactions. The diagonal is omitted. High turnover values (yellow) indicate strong disagreement in network structure between models, whereas lower values (blue–purple) indicate greater similarity. |

In terms of how model choice will influence our inference - this will have the biggest consequence when thinking about diet related questions. In [Figure 4](#fig-beta_div) we can see that the ADBM and PFIM are recovering (almost) totally different pairwise links and so will have very different answers when we want to start interrogating the specific interactions that may established by a specific species within the network. Pragmatically when if comes to deciding which model is the most appropriate choice when ir comes to diet-based questions it makes sense to contextualise this in the feasible vs realised interaction spectrum. Specifically that from a ‘philosophical’ basis if you are asking questions about possible diets of species then it makes sense to use models that fall firmly in the ‘feasible’ space *e.g.,* PFIM model or even something like the [14] model.

## 3.4 Model choice changes the narrative

As shown in both [Figure 3](#fig-structure) and [Figure 4](#fig-beta_div) different models do indeed create networks that look different despite being created using the same species pool (or at least richness in the case of the Niche and Random models). As we have alluded to multiple times these underlying difference could have consequences on inferences that we make about the behaviour of the system. Here we will look at how model choice alters the inference we make about the underlying mechanisms that were driving extinctions in the Toarcian extinction event. Here we again we use we followed the approach outlined in [4] when simulating extinctions as well as assessing which extinction mechanism results in a simulated network most closely match the real post extinction network. Extinction simulations were only run on the pre extinction networks whereby species were removed until they reached the ‘target richness’, which is the richness of the post extinction community. **TODO** The whole shebang about the different extinction simulation scenarios?

Extinctions were simulated using different plausible mechanisms based on species traits (size, motility), their position within the network (generality, vulnerability), as well as randomly. Each network was subjected to **50** extinction runs for each extinction mechanism. The extinctions themselves were cascading in nature meaning that after the target species was removed all species that no longer had any prey were also deemed as extinct (secondary extinction), checking for secondary extinctions was then repeated until there were no longer any species without prey. This represents one extinction event and only then would the proceeding target species be removed from the network and cascading extinctions assessed again. Note that for extinction simulations which use the network position of a species to determine extinction order we follow the protocol from [30] and reassess the vulnerability/generality of each species after each extinction event to ‘redetermine’ the extinction order based on their position in the newly generated network.

**TODO** not sure if we also want to unpack/showcase robustness [31]

|  |
| --- |
| Figure 5: stuff… Recreation of the figure from Dunhill 2024. I think it makes sense to split and divide into the below sections (depending on potential figure limitations of target journal) |

### 3.4.1 Trends over time

Generalized additive mixed modelling showed strong differences in both the magnitude and temporal dynamics of network structure across models. Model results indicated that both parametric and smooth terms significantly influenced network metrics across models, with some exceptions. For connectance, all parametric coefficients were negative except the intercept, and most smooth terms showed highly significant effects (p < 0.05), except for the Niche (p = 0.48) and Random (p = 0.016) models. Complexity and generality similarly exhibited significant parametric and smooth effects, whereas for vulnerability, the smooth term the Random model was non-significant (p = 0.73). Trophic level effects varied among models: modellmatrix had a non-significant parametric effect (p = 0.89), while others were strongly significant. Across the motifs (S1, S2, S4, S5), most parametric and smooth effects were significant (p < 0.05), though a few, such as the intercepts of S1 and S5were non-significant. Overall, these results indicate model- and metric-specific variation in both baseline (parametric) and temporal (smooth) effects, highlighting the importance of model structure in influences the observed temporal trends.

### 3.4.2 Inferred extinction drivers

To assess how consistently different modelling approaches evaluate extinction scenarios, we quantified the agreement in scenario rankings produced by multiple models across several network metrics. For each model, each extinction scenario, and each network metric we calculated the mean absolute difference (MAD) between the observed metric value and the value predicted following the simulated extinction sequence. Lower MAD values indicate a closer match to the empirical network structure and therefore a better-performing extinction scenario for that model and metric. Additionally, we used a derivative of [32] true skill statistic (TSS, see [Equation 1](#eq-1)), where a score below zero indicates that the simulated extinction performs no better than random, and a score of one indicates a perfect match between real and simulated. Here we calculated both a node-level TSS as well as link-level TSS, by parsing out the TSS into two components we are able to assess if differences between real and simulated networks are due to node-level (the wrong species being removed) or link-level (the wrong links be recovered) mismatches. Because the extinction simulations do not allow for the origination of species, when calculating the TSS we only retained species that were present in both the pre and post extinction community and so any node-level mismatches between real and simulated networks was due to the wrong species being removed and not because new species were not.

For each network metric, we treated each model as an independent evaluator of scenario performance. MAD and TSS values were converted to within-model rankings, with rank 1 assigned to the scenario with the smallest MAD (i.e., the closest match to the empirical value) or highest TSS score. Ranking was performed independently for each combination of model and network metric to avoid assumptions about comparability across metrics. To evaluate whether different models produced consistent rankings of extinction scenarios, we quantified rank correlation among models separately for each network metric. Agreement among model rankings was assessed using Kendall’s rank correlation coefficient (τ), which measures the degree of concordance between two ordinal rankings. Kendall’s τ was selected because it is robust for small sample sizes, handles tied ranks appropriately, and provides a direct measure of the probability that model pairs agree or disagree on the relative ordering of scenarios. Kendall’s τ ranges from –1 to +1, where +1 indicates perfect agreement between rankings, 0 reflects no relationship, and –1 represents complete disagreement such that one ranking is the exact reverse of the other.

When we look at Kendall’s τ for the MAD across network structure and models [Figure 6](#fig-mad) we see that generally there is a positive correlation between the different different models. This implies that different models are often recovering a similar ranking of extinction mechanisms (as in the ‘signal’ as to which extinction mechanisms may be the most plausible are the same). Although there is not a strong agreement between models as values tend to be low it is positive to observe that it is not often that we have a completely different ranking of extinction mechanisms, with the exception of complexity and to a lesser degree connectance. Interestingly we once again the the strong similarity between the l-matrix and the ADBM. Broadly this is positive, as despite the fact that we show that these models construct very different networks structurally [Figure 3](#fig-structure) they are still (broadly) capturing a similar extinction mechanism signal.

|  |
| --- |
| Figure 6: Heatmaps showing pairwise Kendall rank correlation coefficients (τ) between models for each network metric. Each panel corresponds to a different metric and displays the degree of agreement in extinction-scenario rankings across models based on mean absolute differences (MAD) between observed and predicted network values. Positive τ values (blue) indicate concordant rankings between models, whereas negative τ values (red) indicate opposing rankings. Warmer colours approaching zero represent little or no agreement. Panels illustrate how consistently different modelling approaches evaluate the relative realism of extinction scenarios across multiple network properties. |

|  |
| --- |
| Figure 7: Heatmaps showing pairwise Kendall rank correlation coefficients (τ) between models for link and node level TSS. Each panel corresponds to a different metric and displays the degree of agreement in extinction-scenario rankings across models based on mean TSS scores between observed and predicted network values. Positive τ values (blue) indicate concordant rankings between models, whereas negative τ values (red) indicate opposing rankings. Warmer colours approaching zero represent little or no agreement. Panels illustrate how consistently different modelling approaches evaluate the relative realism of extinction scenarios across multiple network properties. |

When looking at the node-level TSS scores ([Figure 7](#fig-tss) panel 2) we see that in general the signal of the extinction mechanism is maintained across the different models. However as many of the extinction mechanisms are determine by the *traits* of the node it is not surprising that we see a similar signal as the species are being removed in the exact same order. The only two extinction mechanisms that are driven by The link-level TSS scores ([Figure 7](#fig-tss) panel 1) do not show the same signal within extinction mechanism. We see that the random and PFIM models have high TSS scores (*i.e.,* have a ‘good fit’), however in the case of the PFIM this is to be expected as the links are deterministic and so if you have the same two species pools you will recover the same links. The ‘stochastic’ element of the theoretical models (ADBM, l-matrix, and log-ratio) means that they create a degree of ‘noise’ at the link-level and thus they are probably inappropriate to use for the type of extinction mechanism question we are asking here - specifically does the real and the simulated network look the same. Link-level TSS is perhaps also not an appropriate approach to determine the ‘best fit’ extinction mechanism if used in isolation and we advocate that the node-level TSS score (or alternatively some measure of diversity is used). FInally, if we were to focus only on node level TSS we do not observe any strong differences between the models and it suggests that node-level driven (topological) extinction processes are insensitive to model type.

# 4. Discussion (need a catchier heading)

I want this section to be more about contextualising model choice within the bigger research question discussion - i.e. mapping question and model choice more tightly…

Points to discuss:

* Guidlines - as a box? Can we give something concrete?? Something like [Figure 8](#fig-guidelines)
* How to we synthesise these results? As in should we give clear directives ot is it enough to do a bit more handwaving and have the bigger message be that model choice matters?

|  |
| --- |
| Figure 8: TODO. |

# References

1. Kiessling, W. *et al.* (2019) [Addressing priority questions of conservation science with palaeontological data](https://doi.org/10.1098/rstb.2019.0222). *Philosophical Transactions of the Royal Society B: Biological Sciences* 374, 20190222

2. Dillon, E.M. *et al.* (2022) [What is conservation paleobiology? Tracking 20 years of research and development](https://doi.org/10.3389/fevo.2022.1031483). *Frontiers in Ecology and Evolution* 10

3. Thuiller, W. *et al.* (2024) [Navigating the integration of biotic interactions in biogeography](https://doi.org/10.1111/jbi.14734). *Journal of Biogeography* 51, 550–559

4. Dunhill, A.M. *et al.* (2024) [Extinction cascades, community collapse, and recovery across a Mesozoic hyperthermal event](https://doi.org/10.1038/s41467-024-53000-2). *Nature Communications* 15, 8599

5. Hao, X. *et al.* (2025) [Global Projection of Terrestrial Vertebrate Food Webs Under Future Climate and Land-Use Changes](https://doi.org/10.1111/gcb.70061). *Global Change Biology* 31, e70061

6. Yeakel, J.D. *et al.* (2014) [Collapse of an ecological network in ancient egypt](https://doi.org/10.1073/pnas.1408471111). *PNAS* 111, 14472–14477

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

8. Morales-Castilla, I. *et al.* (2015) [Inferring biotic interactions from proxies](https://doi.org/10.1016/j.tree.2015.03.014). *Trends in Ecology & Evolution* 30, 347–356

9. Pichler, M. and Hartig, F. (2023) [Machine learning and deep learningA review for ecologists](https://doi.org/10.1111/2041-210X.14061). *Methods in Ecology and Evolution* 14, 994–1016

10. Strydom, T. *et al.* (2021) [A roadmap towards predicting species interaction networks (across space and time)](https://doi.org/10.1098/rstb.2021.0063). *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 20210063

11. Allesina, S. *et al.* (2008) [A general model for food web structure](https://doi.org/10.1126/science.1156269). *Science* 320, 658–661

12. Shaw, J.O. *et al.* (2024) [A framework for reconstructing ancient food webs using functional trait data](https://doi.org/10.1101/2024.01.30.578036)bioRxiv, 2024.01.30.578036

13. Roopnarine, P.D. (2006) [Extinction cascades and catastrophe in ancient food webs](https://www.jstor.org/stable/4096814). *Paleobiology* 32, 1–19

14. Fricke, E.C. *et al.* (2022) [Collapse of terrestrial mammal food webs since the Late Pleistocene](https://doi.org/10.1126/science.abn4012). *Science* 377, 1008–1011

15. Gauzens, B. *et al.* (2025) Tailoring interaction network types to answer different ecological questions. *Nature Reviews Biodiversity* DOI: [10.1038/s44358-025-00056-7](https://doi.org/10.1038/s44358-025-00056-7)

16. Dunne, J.A. *et al.* (2014) [Highly resolved early eocene food webs show development of modern trophic structure after the end-cretaceous extinction](https://doi.org/10.1098/rspb.2013.3280). *Proceedings of the Royal Society B: Biological Sciences* 281, 20133280

17. Erdős, P. and Rényi, A. (1959) [On random graphs. i.](https://doi.org/10.5486/pmd.1959.6.3-4.12) *Publicationes Mathematicae Debrecen* 6, 290–297

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

19. Petchey, O.L. *et al.* (2008) [Size, foraging, and food web structure](https://doi.org/10.1073/pnas.0710672105). *Proceedings of the National Academy of Sciences* 105, 4191–4196

20. Schneider, F.D. *et al.* (2016) [Animal diversity and ecosystem functioning in dynamic food webs](https://doi.org/10.1038/ncomms12718). *Nature Communications* 7, 12718

21. Rohr, R. *et al.* (2010) [Modeling food webs: Exploring unexplained structure using latent traits.](https://doi.org/10.1086/653667) *The American Naturalist* 176, 170–177

22. Bambach, R.K. *et al.* (2007) [Autecology and the Filling of Ecospace: Key Metazoan Radiations](https://doi.org/10.1111/j.1475-4983.2006.00611.x). *Palaeontology* 50, 1–22

23. Milo, R. *et al.* (2002) [Network motifs: Simple building blocks of complex networks](https://doi.org/10.1126/science.298.5594.824). *Science* 298, 824–827

24. Stouffer, D.B. *et al.* (2007) [Evidence for the existence of a robust pattern of prey selection in food webs](https://doi.org/10.1098/rspb.2007.0571). *Proceedings of the Royal Society B: Biological Sciences* 274, 1931–1940

25. Williams, R.J. and Martinez, N.D. (2004) [Stabilization of chaotic and non-permanent food-web dynamics](https://doi.org/10.1140/epjb/e2004-00122-1). *The European Physical Journal B - Condensed Matter* 38, 297–303

26. Delmas, E. *et al.* (2018) Analysing ecological networks of species interactions. *Biological Reviews* DOI: [10.1111/brv.12433](https://doi.org/10.1111/brv.12433)

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

28. Roopnarine, P.D. (2017) Ecological Modelling of Paleocommunity Food Webspp. 201–226, University of Chicago Press

29. Poisot, T. *et al.* (2012) [A comparative study of ecological specialization estimators](https://doi.org/10.1111/j.2041-210x.2011.00174.x). *Methods in Ecology and Evolution* 3, 537–544

30. Curtsdotter, A. *et al.* (2011) [Robustness to secondary extinctions: Comparing trait-based sequential deletions in static and dynamic food webs](https://doi.org/10.1016/j.baae.2011.09.008). *Basic and Applied Ecology* 12, 571–580

31. Jonsson, T. *et al.* (2015) [The reliability of R50 as a measure of vulnerability of food webs to sequential species deletions](https://doi.org/10.1111/oik.01588). *Oikos* 124, 446–457

32. Gupta, A. *et al.* (2022) [Simultaneously estimating food web connectance and structure with uncertainty](https://doi.org/10.1002/ece3.8643). *Ecology and Evolution* 12, e8643