Methods for Reconstructing Paleo Food Webs

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

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.

1   Why build paleo food webs?

- Because its interesting? Point to works that showcase how we can learn form the past and inform on the current but possibly also the future

- Value in using hindcasting to aid in forecasting. *e.g.,* the Toarcian ms (Dunhill et al., 2024) shows how we can use these paleo communities to understand trophic-level responses to extinctions.

2   How do we do it?

- There is an evolving body of work that focuses on developing tools specifically for the task of predicting food webs.

- There are a handful that have been developed specifically in the context of paleo settings *e.g.,* TODO but we can also talk about those that might have been developed/tested in contemporary settings but still have applicability in paleo ones.

- Different underlying theory though

    - Focus here on the idea of different 'currencies' but also aggregations - energy vs compatibility.

- Insert brief overview of the different methods as they pertain to approach (so the T4T triangle)

- Challenges we face (even in contemporary settings)?

    - keep high level - I think the argument here should fall more in the data trade offs...

3   Understanding how networks are different

It is important to be aware that networks can be configured in different ways depending on how the interactions are defined (Strydom, in prep). Basically we have metawebs (which represent *potential* interactions), realised networks (which represent the subset of potential that are realised as a result of community and environmental context), and structural networks (species agnostic networks that are structurally informative). Here also talk about the implications of these different networks types - different uses and capturing different processes. Speciifcally link this to models - *i.e.,* different models have been developed to construct a *specific* network representation.

4 Challenges specific to paleo communities/networks

Although there are a suite of tools and methods that have been developed to predict species interactions and networks they will not all be suitable for the prediction of paleo communities. Some of these include the fact that the fossil record is incomplete/preservation is biased [REF] which means that we have an incomplete picture of the entire community. Fossils are 2D and only represent specific 'parts' of an individual (hard and bone-y bits), this means we don't have a complete picture of the physical traits of species *e.g.,* no body mass (but yes size), behaviours, or ability to construct well resolved phylogenetic trees the deeper we go back in time. Also owing to the patchy nature of fossils one often has to aggregate over large spatial scales, and also fossils are preserved in 2D so no *real* idea of spatial arrangements, compounded that fossils aren't necessarily conserved/found 'in situ' but can be moved (*e.g.,* alluvial deposits). Methodologically speaking some tools that 'learn' from contemporary communities (*e.g.,* Strydom et al. (2023), Caron et al. (2022)) will become 'worse' the further one goes back in time since species then look very different from now but can still be useful for 'recent' communities (*e.g.,* Fricke et al. (2022)). Something about the intersectionality of the data we don't have for paleo communities and the data we need for some of the different modelling approaches.

5 Methods

5.1 Dataset Overview

5.1.1 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 Dunhill et al., 2024 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.

5.1.2 Defining modes of life (traits)

We used the modes of life (traits) as identified in Dunhill et al. (2024), 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 (Bambach et al., 2007).It

would be very sexy if we could get a figure that looks something like this together...

5.2 Models

Here we present six different models (@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 are 'paleo friendly'.

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	Erdős & Rényi (1959)
niche	Networks are interval, species can be ordered on a 'niche axis'	richness, connectance	parameter assumptions, species agnostic	structural network	Williams & Martinez (2008)
allometric diet breadth model (ADBM)	Interactions are determined by energetic costs	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	realised network	Petchey et al. (2008)

Model family	Assumptions	Data needs	'Limitation'	Network type	Key reference
l-matrix	Interactions inferred using allometric rules (ratio of body sizes between predator and prey)	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	realised network	Schneider et al. (2016)
paleo food web inference model (PFIM)	Interactions can be inferred by a mechanistic frame-work/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)	metaweb	Shaw et al. (2024)

Model family	Assumptions	Data needs	'Limitation'	Network type	Key reference
body size ratio model	Interactions inferred using allometric rules (ratio of body sizes between predator and prey)	body mass	does not account for forbidden links in terms of evolutionary compatibility, assumptions on body size from fossil data	metaweb??	Rohr et al. (2010)

5.2.1 Random model

The Erdős–Rényi random graph model (Erdős & Rényi, 1959) uniformly at random assigns an L number of links to an S number of nodes (species richness). From an ecological perspective this model assumes that 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. This creates a food web that is as free as possible from biological structuring while maintaining the expected richness (S) and connectance (L/S^2)

We could theoretically use the other 'null models' BUT I feel like in the context of constructing a network for a given community the Erdős–Rényi is the better choice than the other models that (IMO) are more suited to hypothesis testing e.g. do observed networks differ from the null network...

5.2.2 Niche model

The niche model (Williams & Martinez, 2000) introduces the idea that species interactions are based on the 'feeding niche' of a species. Broadly, all species are randomly assigned a 'feeding niche' range and all species that fall in this range can be consumed by that species (thereby allowing for cannibalism). The niche of each species is randomly assigned and the range of each species' niche is (in part) constrained by the specified connectance (Co) of the network. The niche model has also been modified, although it appears that adding to the 'complexity' of the niche model does not improve on its ability to generate a more ecologically 'correct' network (Williams & Martinez, 2008).

Each of S species assigned a 'niche value' parameter n_i drawn uniformly from the interval [0,1]. Species i

consumes all species falling in a range (r_i) that is placed by uniformly drawing the center of the range (c_i)

from $[\frac{r_i}{2}, n_i]$ The size of r_i is assigned by using a beta function to randomly draw values from [0,1] whose

expected value is $2 \times Co$ and then multiplying that value by n_i to obtain the desired Co.

5.2.3 Allometric diet breadth model

The Allometric diet breadth model (ADBM; Petchey et al. (2008)) 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, where the diet (K) is defined as follows:

$$K = \frac{\sum_{i=1}^{k} \lambda_{ij} E_i}{1 + \sum_{i=1}^{k} \lambda_{ij} H_{ij}}$$

where λ_{ij} is the handling time, which is the product of the attack rate A_i and resource density N_i, E_i is the

energy content of the resource and H_{ij} is the ratio handling time, with the relationship being dependent on

the ratio of predator and prey bodymass as follows:

$$H_{ij} = \frac{h}{b - \frac{M_i}{M_j}} if \frac{M_i}{M_j} < b$$

or

$$H_{ij} = \infty \geq b$$

Refer to Petchey et al. (2008) for more details as to how these different terms are parametrised.

5.2.4 L matrix

For now we can link to the ATNr package (Gauzens et al., 2023) until I can find a more suitable manuscript

that breaks down this construction method. Schneider et al. (2016) Interactions are determined by allometric

rules (ratio of consumer (M_i) and resource (M_j) body sizes) and a Ricker function as defined by R_{opt} and

γ and returns The probability of a link (P_{ij}) existing between a consumer and resource, and is defined as

follows:

$$P_{ij} = (L \times \exp(1 - L))^{\gamma}$$

where

$$L = \frac{M_i}{M_j \times R_{opt}}$$

It is also possible to apply a threshold value to P_{ij}, whereby any probabilities below that threshold are set to zero.

5.2.5 Paleo food web inference model

The Paleo food web 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 feasibly occur between a species pair. If all conditions are met for the different rule classes then an interaction is deemed to be feasible. The original work put forward in Shaw et al. (2024) also includes a 'downsampling' step developed by Roopnarine (2006) that uses a power law, defined by the link distribution, to 'prune' down some of the links. It is worth mentioning that this approach is similar to that developed by Roopnarine (2017) with the exception that Shaw does not specifically bin species into guilds, and so we choose to use the method developed by Shaw since both methods should produce extremely similar networks as they are built on the same underlying philosophy.

5.2.6 Body size ratio model

The body size ratio model (Rohr et al., 2010) determines feeding interactions using the ratio between consumer (M_i) and resource (M_j) body sizes - which supposedly stems from niche theory (still trying to reconcile that). The probability of a link existing between a consumer and resource (in its most basic form) is defined as follows:

$$P_{ij} = \frac{p}{1 + p}$$

where

$$p = exp[\alpha + \beta log(\frac{M_i}{M_j}) + \gamma log^2(\frac{M_i}{M_j})] \tag{1}$$

The original latent-trait model developed by Rohr et al. (2010) also included an additional latent trait term $v_i \delta f_j$ however for simplicity we will use Equation 1 as per Yeakel et al. (2014) . Based on Rohr et al. (2010)

7

₁₁₉ it is possible to estimate the parameters α, δ, and γ using a GLM but we will use the parameters from Yeakel
₁₂₀ et al. (2014), which was 'trained' on the Serengeti food web data and are as follows: $\alpha = 1.41$, $\delta = 3.75$, and
₁₂₁ $\gamma = 1.87$.

5.3 Assessing model performance

₁₂₃ In terms of wanting to asses and compare across the different models it is beneficial to approach this task
₁₂₄ by thinking about the different aspects of the network as well as interactions that are being predicted by the
₁₂₅ different models. It is perhaps beneficial to think of these across different 'scales' of organisation within the
₁₂₆ network, namely macro (the entire network), meso (smaller interacting units within the network), and micro
₁₂₇ (species-level attributes). Although there are a myriad of possible ways to 'measure' and analyse ecological
₁₂₈ networks (Delmas et al., 2018) we do still lack a clear set of guidelines for assessing how well models recover
₁₂₉ network structure (Allesina et al., 2008) and it is beneficial to use a small subset of metrics that can clearly
₁₃₀ be tied to broader aspects of network function or capturing a ecological process.

5.3.1 Macro network properties

₁₃₂ **Connectance** (Martinez, 1992) has been shown to be the feature of networks that underpin a series of other
₁₃₃ properties and function (Strydom, Catchen, et al., 2021) and so it is perhaps the most important structural
₁₃₄ attribute for a model to be able to retrieve correctly. Additionally we consider the **complexity** of networks by
₁₃₅ calculating their SVD entropy (this gives us an estimate of the physical as opposed to behavioural complexity
₁₃₆ of networks; Strydom, Dalla Riva, et al. (2021)), we could also look at the rank/rank deficiency of networks
₁₃₇ which (theoretically) represents the number fo unique interaction strategies in the network (Strydom, Dalla
₁₃₈ Riva, et al., 2021), which may be specifically interesting in terms of looking at pre and post extinction but
₁₃₉ also as a way to unpack 'functional redundancy' that some models may introduce.

5.3.2 Meso network properties

₁₄₁ Motifs represent smaller subset of interactions between three species, and are argued to capture dynamics
₁₄₂ that are likely to be ecologically relevant (Milo et al., 2002; Stouffer et al., 2007). Here we specifically look
₁₄₃ at the number of **linear chains**, **omnivory**, **apparent competition**, and **direct competition** motifs. In
₁₄₄ the broader context the ability of a model in being able to capture these smaller motifs will inform as to its
₁₄₅ suitability of use understanding the more dynamic component of network ecology.

5.3.3 Micro network properties

The number of interactions established (**generality**) or received (**vulnerability**) by each species (Schoener, 1989), are (broadly) indicative of consumer-resource relationships and diet breadth of species [ref]. Although this is usually determined at the species level the standard deviation of the generality and vulnerability of species is often used when benchmarking predicted networks (Petchey et al., 2008; *e.g.,* Williams & Martinez, 2008).

The **specificity** of species in a network is measured as a function of the proportion of resources they effectively use (Poisot et al., 2012)

> **Shape:** to determine if the 'shape' of the network is correct we are looking at the ratio of 'top':'basal' species (where 'top' species are those that have a vulnerability of 0 and 'basal' species have a generality of 0) as well as the distance to base from one of the top species (this will represent the shortest path but a large discrepancy between the real and predicted network would be indicative that the model is not predicting a similar 'shape'). This will allow is to see if the models construct tall 'pencil' vs flat 'pancake' networks (Beckerman 2024, pers comms). A small (<1) number will thus be indicative of a 'bottom-heavy' network and the opposite for larger numbers

5.3.4 Interactions

Interaction turnover (Poisot et al., 2012) tells us which interactions are 'conserved' (shared) across the networks from the same period but constructed using different models.

5.3.5 Robustness

R_{50} see Jonsson et al. (2015)

6 Results

6.1 Comparing predicted networks

[Figure 1 about here.]

9

6.2 Comparing inference

6.3 Extinctions

[Figure 2 about here.]

[Figure 3 about here.]

References

Allesina, S., Alonso, D., & Pascual, M. (2008). A general model for food web structure. *Science*, *320*(5876), 658–661. https://doi.org/10.1126/science.1156269

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

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

Delmas, E., Besson, M., Brice, M.-H., Burkle, L. A., Dalla Riva, G. V., Fortin, M.-J., Gravel, D., Guimarães, P. R., Hembry, D. H., Newman, E. A., Olesen, J. M., Pires, M. M., Yeakel, J. D., & Poisot, T. (2018). Analysing ecological networks of species interactions. *Biological Reviews*, 112540. https://doi.org/10.1111/brv.12433

Dunhill, A. M., Zarzyczny, K., Shaw, J. O., Atkinson, J. W., Little, C. T. S., & Beckerman, A. P. (2024). Extinction cascades, community collapse, and recovery across a Mesozoic hyperthermal event. *Nature Communications*, *15*(1), 8599. https://doi.org/10.1038/s41467-024-53000-2

Erdős, P., & Rényi, A. (1959). On random graphs. i. *Publicationes Mathematicae Debrecen*, *6*(3-4), 290–297. https://doi.org/10.5486/pmd.1959.6.3-4.12

Fricke, E. C., Hsieh, C., Middleton, O., Gorczynski, D., Cappello, C. D., Sanisidro, O., Rowan, J., Svenning, J.-C., & Beaudrot, L. (2022). Collapse of terrestrial mammal food webs since the Late Pleistocene. *Science*, *377*(6609), 1008–1011. https://doi.org/10.1126/science.abn4012

Gauzens, B., Brose, U., Delmas, E., & Berti, E. (2023). ATNr: Allometric Trophic Network models in R. *Methods in Ecology and Evolution*, *14*(11), 2766–2773. https://doi.org/10.1111/2041-210X.14212

Jonsson, T., Berg, S., Pimenov, A., Palmer, C., & Emmerson, M. (2015). The reliability of R50 as a measure of vulnerability of food webs to sequential species deletions. *Oikos*, *124*(4), 446–457. https://doi.org/10.1111/oik.01588

Martinez, N. D. (1992). Constant connectance in community food webs. *The American Naturalist*, *139*(6), 1208–1218. http://www.jstor.org/stable/2462337

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

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

Poisot, T., Canard, E., Mouquet, N., & Hochberg, M. E. (2012). A comparative study of ecological specialization estimators. *Methods in Ecology and Evolution*, *3*(3), 537–544. https://doi.org/10.1111/j.2041-210x.2011.00174.x

Rohr, R., 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

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

Roopnarine, P. D. (2017). *Ecological Modelling of Paleocommunity Food Webs* (pp. 201–226). University of Chicago Press.

Schneider, F. D., Brose, U., Rall, B. C., & Guill, C. (2016). Animal diversity and ecosystem functioning in dynamic food webs. *Nature Communications*, *7*(1), 12718. https://doi.org/10.1038/ncomms12718

Schoener, T. W. (1989). Food Webs From the Small to the Large: The Robert H. MacArthur Award Lecture. *Ecology*, *70*(6), 1559–1589. https://doi.org/10.2307/1938088

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

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

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V., & Poisot, T. (2023). Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations. *Methods in Ecology and Evolution*, *14*(12), 2917–2930. https://doi.org/10.1111/2041-210X.14228

Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz, N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of*

233   *the Royal Society B: Biological Sciences, 376*(1837), 20210063. https://doi.org/10.1098/rstb.2021.0063

234   Strydom, T., Dalla Riva, G. V., & Poisot, T. (2021). SVD entropy reveals the high complexity of ecological

235   networks. *Frontiers in Ecology and Evolution, 9.* https://doi.org/10.3389/fevo.2021.623141

236   Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature, 404*(6774), 180–183.

237   https://doi.org/10.1038/35004572

238   Williams, R. J., & Martinez, N. D. (2008). Success and its limits among structural models of complex food

239   webs. *The Journal of Animal Ecology, 77*(3), 512–519. https://doi.org/10.1111/j.1365-2656.2008.01362.x

240   Yeakel, J. D., Pires, M. M., Rudolf, L., Dominy, N. J., Koch, P. L., Guimarães, P. R., & Gross, T. (2014).

241   Collapse of an ecological network in ancient egypt. *PNAS, 111*(40), 14472–14477. https://doi.org/10.
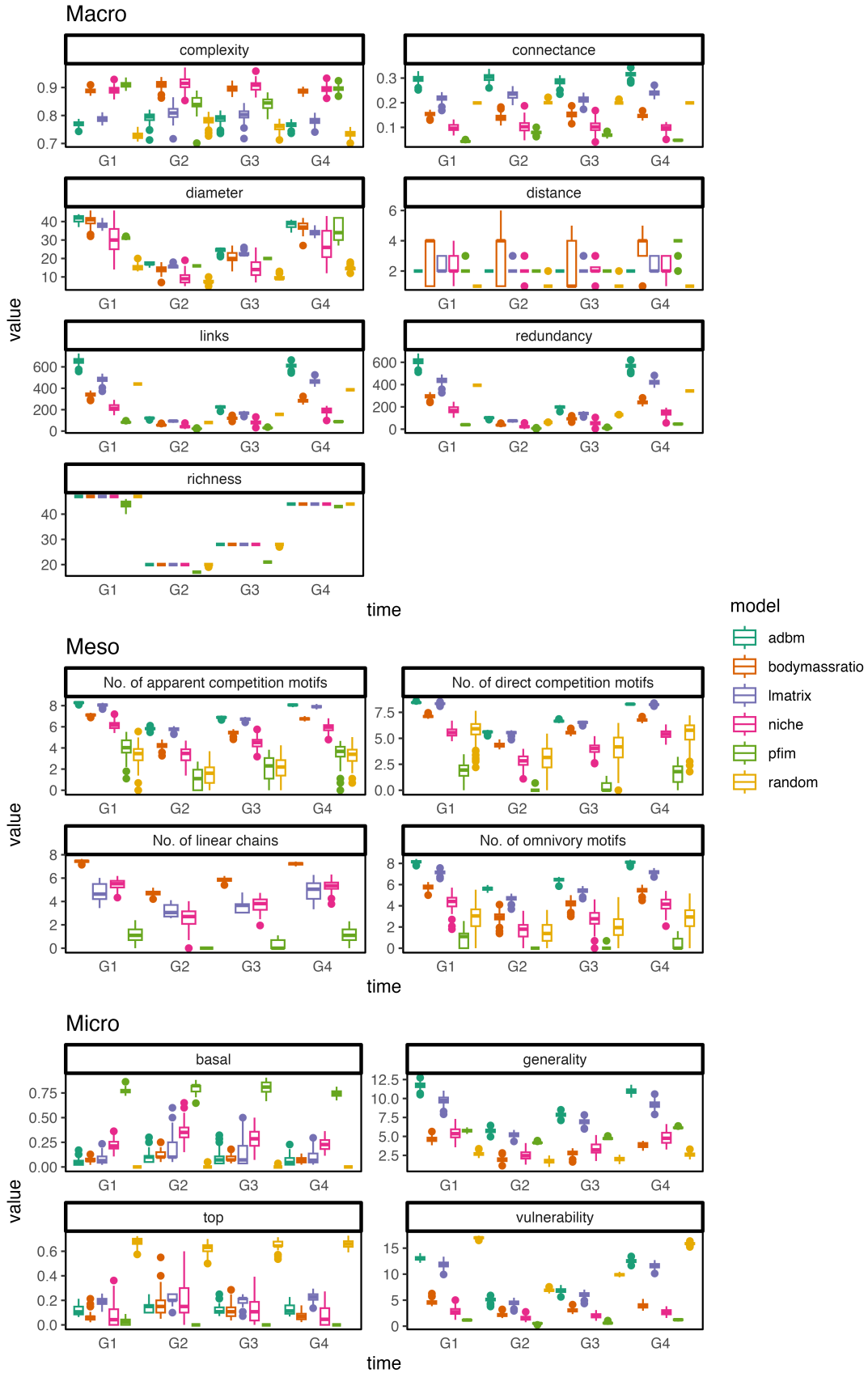
242   1073/pnas.1408471111

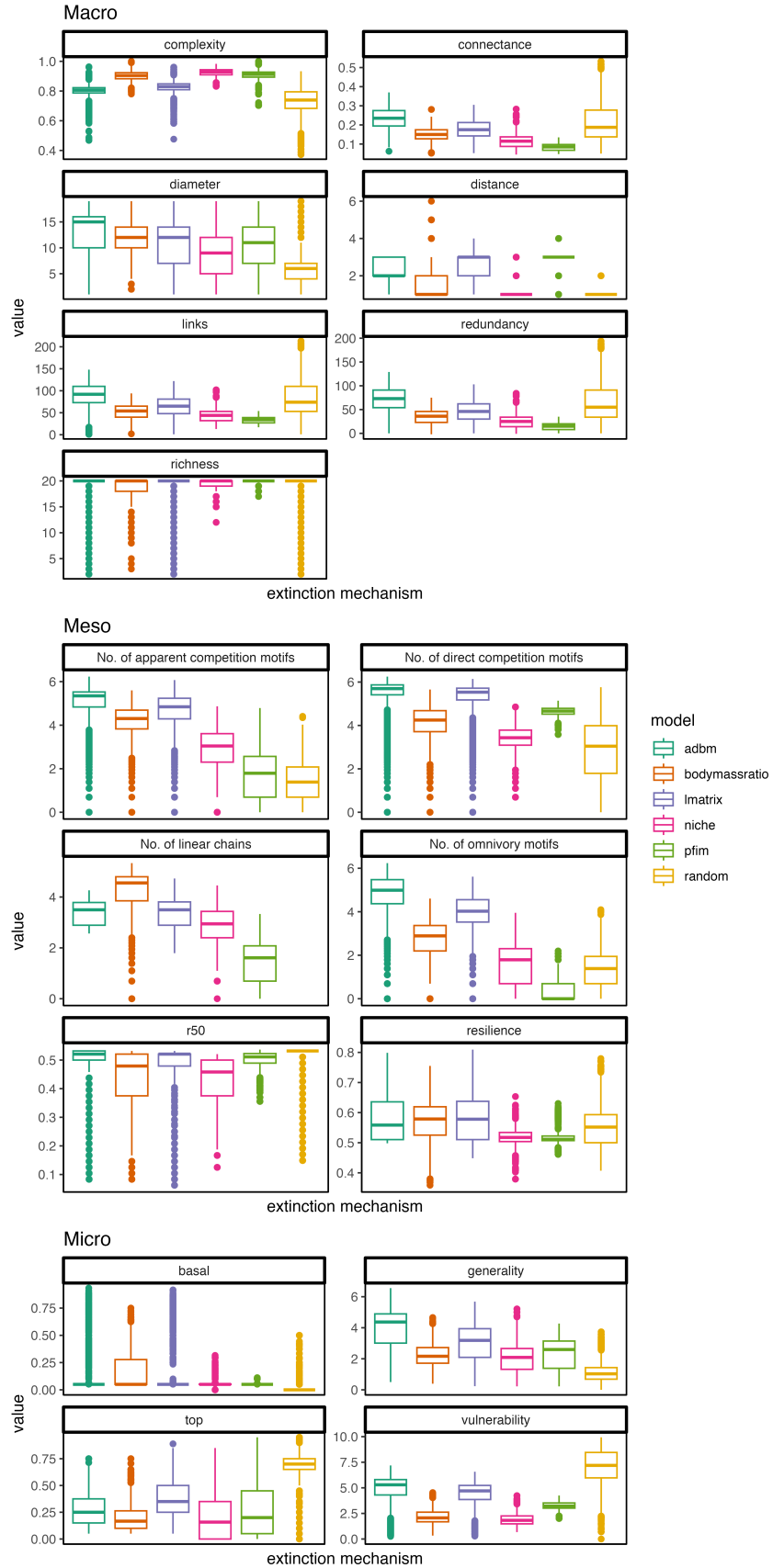Figure 1: stuff... For display purposes the counts for the different motifs are log transformed

Figure 2: Dashed line indicates the (mean) extinction simulation results (post value, start values are those estimated by the relevant model). For display purposes the counts for the different motifs are log transformed
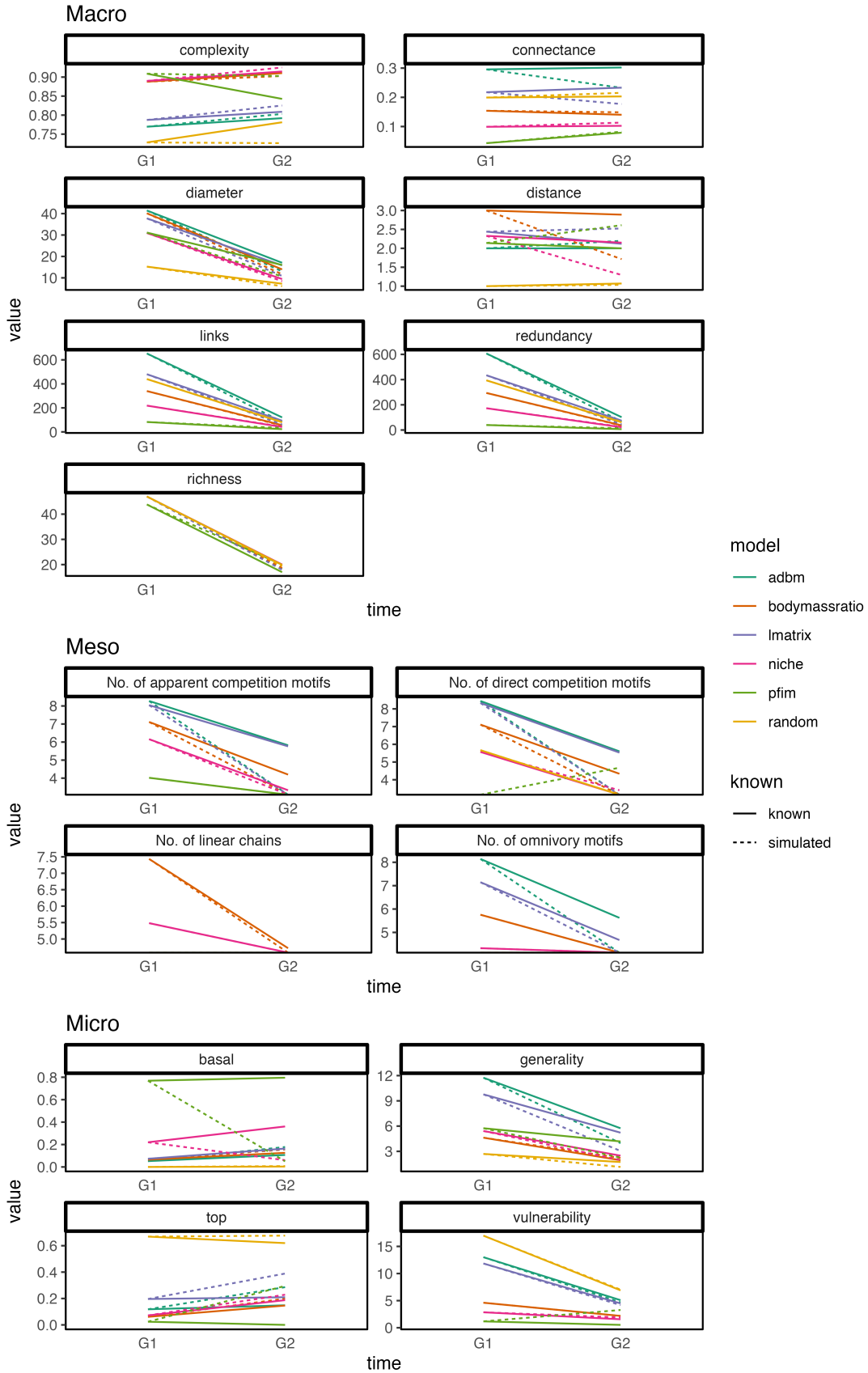
Figure 3: Dark line indicates 'real' extinction simulation results the lighter lines show each model individually, which is also denoted by the linetype. For display purposes the counts for the different motifs are log transformed