Title: A quantitative framework for network biogeography

Authors: Dominique Gravel[1,2,*], Timothée Poisot[1,2,3], Neo Martinez[4], Jennifer Dunne[5], Jason Tylianakis[3], Nicolas Mouquet[6], Daniel Stouffer[3]

1: Canada Research Chair in Biogeography and community ecology. Département de biologie, chimie et géographique, Université du Québec à Rimouski, 300 Allée des Ursulines, Québec, Canada. G5L 3A1.

2: Québec Centre for Biodiversity Sciences

3:

4:

5:

6:

Keywords: networks, spatial ecology, co-occurrence, probability of interaction

Words in the abstract:
Words in the main text:
Words in the legends:
Figures:
Tables:
References:

Abstract

Introduction

- The community is more than a species list. Interactions are central.

- Network structure do vary in space in time.

- We don't know yet to what extent interactions are varying with the environment.

- No theory to explain and interpret the meaning of network variation in space. Current interpretation fo species turnover involves the effect of the environment and stochasticity.

- Objective: Propose a theoretical framework to understand and predict the spatial and temporal variation in network structure.

A probabilistic representation of ecological networks

Interaction networks do vary in space and time because any given pairwise interaction could either occur or not at any particular location. We seek to represent the probability an interaction between species i and j occurs at location y. We consequently define L_{ijy} as a stochastic variable and we are thus looking at the probabilty this event occurs, $P(L_{ijy})$. There are several factors that could impact the occurrence of an interaction and we will describe them below. But ultimately, this probability depends on the spatial and temporal scale of observations. As long as the interaction probability is not null, the probability of observing an interaction will tend to 1 as the scale of observation increases. There is nonetheless a considerable amount of information in the probability matrix of interaction $mathbf L_y$ because of the uneven distribution of interaction probabilities between species, just as there is in quantiative description of networks (Banasek-Richter2002).

The occurrence of an interaction is dependent on the co-occurrence of species i and j. This argument might seem trivial at first, but the explicit consideration of this condition in the probabilistic representation of ecological networks will prove fundamental to understand their variation. We thus define X_{iy} as a stochastic variable representing the occurrence of a species i at location y, and similarly X_{ijy} the co-occurrence of species i and j. The quantity we seek to understand is the probability of a joint event:

$$P(X_{iy}, X_{jy}, L_{ijy},) \tag{1}$$

Which reads as the probability of observing species i, species j and an interaction between them.

The probability an interaction occurs could be decomposed in two parts using the product rule of probabilities:

$$P(X_{iy}, X_{jy}, L_{ijy}) = P(L_{ijy}|X_{iy}, X_{jy}, E_y)P(X_{iy}, X_{jy}|E_y) \tag{2}$$

We will refer to the left term as the metaweb. It is a conditional probability, representing the probability that an interaction occurs if species i and j are co-occurring. The right term is the probability of observing the two species co-occurring at location y.

The metaweb concept is making its way through the network litterature even though it has never been formally and technically defined. It is usually conceived as a network of interactions among species that could potentially co-occur. Here we define it as the matrix of interaction probabilities between co-occurring species. In other words, it represents the probability of interactions after factoring out the effect of distribution. It thus represents potential interactions and should therefore include interactions between species that never co-occurred but are susceptible to. The problem with most representations of metawebs to date is that the effect of co-occurrence is never factored out. The traditional approach to build a metaweb is to cumulate observations across replicated networks. The main problem with that approach is that the co-occurrence of rare species is extremely unlikely and thus most often appear as an absence of interactions in the metaweb. This approach is inappropriate because the observe co-occurrence will have a strong signature on the evaluation of interactions. If built with observations of interactions, then the only way to fill a metaweb is by running cafeteria experiments between all pairs of species. Otherwise, the metaweb should be inferred using traits and phylogenetic information. Most of the published metawebs are therefore incomplete because of their sensitivity to sampling heterogeneity. A rarefaction analysis previously shown that interactions accumulate with the addition of networks at a slower rate than species richness. It indicates that it is harder to have a direct evaluation of interactions from observeration than it is to evaluate species richness.

We will come back to the issue of evaluating the metaweb in the section Applications

There are many variants of the metaweb representing different hypotheses about the origin of temporal and spatial variation in network structure (see the explicit formulations at Table 1). First, the interaction could be considered deterministic instead of probabilistic. In other words, $P(L_{ijy}) = 1$ if $X_{ijy} = 1$, and 0 otherwise. This representation of the metaweb is the one mostly used so far, as soon as the species are found together they are assumed to interact. It is also the only way to represent interactions when there is not enough information available to evaluate the probability of interaction. It is a reasonnable approximation when the sampling and inferrence scales are large enough and that the only variation of networks considered arises from species distribution. Ecological interactions could also depend on the environment. Although it is not common to see a conditional representation of ecological interactions, experimental studies of pairwise interactions revealing their sensitivity to the environment are common. For instance, it has been documented that the predation risks of shorebirds do vary at the continental scale, from the south to the north. Here the environment is considered in a very broad sense, as any factor potentially influencing the probability of a pairwise interaction, provided that the species co-occur. It thus includes both the biotic and the abiotic environment. We note however that here the biotic environment includes organisms that are not considered in the co-occurrence matrix. In such a case, any pairwise interaction could be conditional on higher order interactions. An interaction modifier occurs for instance when the predation risk by species j might be impacted by a parasite k changing the behaviour of the prey i. We note that a conditional probability approach could thus be used represent non-trophic interactions into ecological networks (REF). This topic is however beyond the scope of the current paper.

There are also variants to the co-occurrence matrix. Akin to the metaweb, co-occurrence could be conditional or not. A simple representation of it is simply to model co-occurrence as a function of the environment. In this situation there is nounderlying assumption about the ecological processes responsible for co-occurrence. Alternatively, the co-occurrence probability could be a function of the environment because of shared ecological requirements. We call the later neutral because species are specifically responding to the environment but are independently distributed. Co-occurrence is then simply obtained by multiplying the result of two independent and specific species distribution models (SDM).Finally, the co-occurrence probability itself could be dependent on ecological interactions. Direct pairwise interactions such as competition, facilitation and predation have long been studied for their impact on co-distribution. Second and higher order interactions (e.g. trophic cascade) could also impact co-occurrence. There is however currently no general theory on the expected co-occurrence in complex ecological networks. For instance, we do not know how far co-occurrence is not-random when going along the chain of indirect interactions. Berlow(2009) shown previously that almost only first and second order interactions do matter in food webs, but we don't know for co-distribution. We neither know what is the sensitiviy to species richness: do interactions tend to buffer each other? Generalizing knowledge aquired by the study of small community modiles will require future research.

Interpretation: the integrated niche

The niche concept is central in biogeography to understand and predict species distribution. Several attempts have been made to refresh the concept, but its main usage still follows Hutchinson's idea that species interactions restrict the fundamental niche to a realized one. Despite its intuitive interpretation and translation into species distribution models, the concept has been constantly criticized (Hardin, 1960; Peters, 1991; Chase2003; Silvertown, 2004; Soberon, 2007) and several attempts have been made to expand and reinforce it. Part of the problem surrounding the definition of the niche has been clarified with the distinction between Eltonian and Grinnellian definitions (ChaseLeibold 2003). The Grinnellian dimension of the niche is the effect of the environment on the demography of a species, while the Eltonian dimension is the effect of a species on its environment sensu lato. The Grinnellian niche is the most intuitive one to apply and is the conceptual backbone of species distribution models. The Eltonian niche is well known by food web ecologists, but it is much more difficult to formulate into distribution models. Nonetheless, the development of the niche model of food web structure (Williams and Martinez, 2000) and its parameterization (Williams et al. 2010; Gravel et al. 2013) made it more operational.

While it is easy to represent statistically the hyper volume where a species occurs, it is much more challenging to account for ecological interactions. Chase and Leibold (2003) attempted this representation in their definition: *[The niche is] the joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal or greater than its death rate along with the set of per capita effects of that species on these environmental conditions.* They represented the niche graphically with zero-net growth isoclines (the Grinnelian niche) and impact vectors (the Eltonian niche). While this representation has been very influential in community ecology at the local scale, it remains unpracticable at the biogeographical one. The absence of any mathematical representation of the niche that could easily be fit to ecological data perhaps explain why biogeographers are still struggling to develop species distribution models taking into account ecological interactions.

The key point to integrate dimensions of the niche is to represent the Eltonian niche into a Grinnelian space. - We do so by considering that the Eltonian niche is the hypoervolume in the trait-space allowing an interaction. - Doing so, we could project both niches in a plane and find the hypervolume where an interaction should occur (Fig. 2). - This visual representation is parallel to the probabilistic definition of interaction probability. - We propose that the metaweb is the Eltonian dimension of the niche, while the matrix of co-occurrence is the Grinnellian dimension. - Feedbacks between dimensions occur through the inclusion of co-occurrence in the metaweb, and interactions in the co-occurrence matrix. - This approach radically change the representation of the niche, putting species distribution and ecological interactions at the same level. - Fitting the probabilistic model allows the evaluation of link distribution and species distribution models. - Moreover, the integrated niche concept facilitates the formulation of species distribution models taking into account biotic interactions (see the section Applications)

Example: network structure in different habitats

In this section we provide an analysis illustrating the framework with an empirical dataset of host-parasitoid networks. Data come from the study of Tylianakis(2007) on the impacts of habitat modifications to the network structure. The data consists of 48 networks with 4090 recorded interactions. The advantage of replicated host-parasitoid networks is that usually every interaction is observed, not inferred from a stationary metaweb. It thus allows to evaluate interaction probability and to factor out the effect of co-occurrence. Five habitats were sampled along a gradient of habitat modification: forest, abandonned coffee agroforest, coffee agroforest, pasture and rice culture. The metaweb consists of 9 parasitoids and kleptoparasites (Hymenoptera: Eulophidae, Ichneumonidae, Leucospidae, Megachilidae and Chrysididae; Dyptera: Bombyliidae) of 33 species of bees and wasps (Hymenoptera: Apidae, Megachilidae, Mutilidae, Pompilidae, Sphecidae, Vespidae). The metaweb is illustrated at Fig. 2, along with an example of one iteration of the metaweb.

Tylianakis (2007) investigated if habitat modification affects the structure of these networks They found a significant impact of the habitat on their structure, despite little variation in species richness. Increasing habitat modification led to a higher parasitoid to host species ratio and a paraistoid were also more specialized, thus impacting considerably vulnerability. A closer inspection of the networks revealed that intensive agricultural systems were dominated by a strong interaction and a specialization of the most abundanc parasitoid. Although the discussion made clear that both the turnover in species composition and the interaction probability changed with habitat modification, it was not possible to partition these components.

We developped a R package (REF) to fit alternative formulations of the metaweb and the co-occurrence matrix along an environmental gradient and run it to re-interpret the data of Tylianaks (2007). The package provides a general interface facilitating the development of different species and link distribution models. It is also built to facilitate the interaction iwth the Mangal database of ecological interactions (REF). The first step consists of fitting a probablistic model from the observation of a pairwise interaction (binary) and the environment (could be categorical or continuous) from the subset of the data where the two species are co-occurring. In other words, it fits the equation $P(L_{ijy}|X_{iy}, Xjy, E_y)$ to the data where $X_{iy} = 1$ and $X_{iy} = 1$. Logistic regression was used and is currently programmed, but alternative models could be used as well. The second steps consists of fitting a a probabilistic model for co-occurrence over the whole dataset, $P(X_{iy}, Xjy|E_y)$, independently of the observation of an interaction. The two probabilities are then multiplied to obtain the probability of observing an interaction (Eq. 2). We used this probability to compute the likelihood of each observation ($\zeta(\theta|D) = P(L_{ijy}, X_{iy}, X_{jy})$ if $L_{ijy} = 1$ and $\zeta(\theta|D) = 1 - P(L_{ijy}, X_{iy}, X_{jy})$ otherwise). We then after compare the models by their AIC.

We considered the gradient of habitat modification as a ordered categorical variable and compared XX models (results are summarized at Table 2). Not surprisingly the best model takes into account the effect of the environment on both the metaweb and co-occurrence. What is most interesting are the comparisons to the best model. First, we find that using a constant metaweb has a dramatic impact on the fit of the model to the data (the AIC drops from X for model 1 to X for model 2), indicating a strong effect of the environment on pairwise interactions. Secondly, we find that the deterministic metaweb is the worst model (model 3, AIC =). This result indicate that the traditional appraoch to consider that species interact as soon as they co-occur is definitely wrong. Thirdly, we also find that using a constant co-occurrence does have a significant impact on the model (the AIC drops to X, model 4), indicating there is a non-random change in community composition with habitat modification. Taken together, these two results better explain why network structured changed with habitat modification, even though here we only used binary information about the network structure. Another interesting result is that considering a neutral co-occurrence did not impact much the fit of the modeel. The AIC drops to XX with model 6, indicating that considering indepdenent SDMs yields similar networks over this environmental gradient. This means that for this particular dataset, ecological interactions does not have a strong impact on species distribution since; a strong dependence of parasitoids to the host for instance would have a occurrence probability higher than expected by chance, while a repulsion would have had the opposite.

An important output of this analysis is a more explicit representation of the uncertainty in the evaluation of the metaweb. We find that among the XX pairs of host and parasitoids, XX did not co-occur. There were therefore many forbidden links based on co-occurrence. These might never occur in reality, but we do not know without doing extra experiments. Therefore, any analysis of the structure of the metaweb would be inappropriate without filling those gaps. In addition to specific experiments, the gaps could be filled with a trait-based approach, using phylogenies or with a null hypothesis (e.g. the interaction probability is equal to connectance computed on the observed interactions).

It is also possible to obtain for each pairwise interaction an estimate of the uncertainty. Not surprisingly, the confidence interval is usually very high for the estimation of a probability with a very small sample size. The standard error on the evaluation of the interaction probability is provided along with the metaweb at Fig. 3. It reveals that the uncertainty is very high for most interactions, even if 48 networks were sampled. Such an approach could be used to detect wich pairwise interaction requires additional sampling in order to reduce the uncertainty to a manageable level.

Applications

Network descriptors

Partitionning beta diversity

Null model testing

Species & link distribution models

Conclusion

- New research agenda

- List of new questions

Acknowledgements

This is tis a contribution to the Santa Fe Institute working group *Networks over ecological gradients*. DG received financial support from NSERC and Canada Research Chair program. TP is supported by a MELS-FQRNT post-doctoral fellowship.

Figure legends

Figure 1

Non-random sampling of the metaweb. The sampling of the metaweb is illustrated with a local interaction network from the Tylianakis et al (2007) dataset. Here the metaweb is simply the number of observed interactions over the 48 networks. Arrows points to species that are present in the local network #34 (represented on the right).

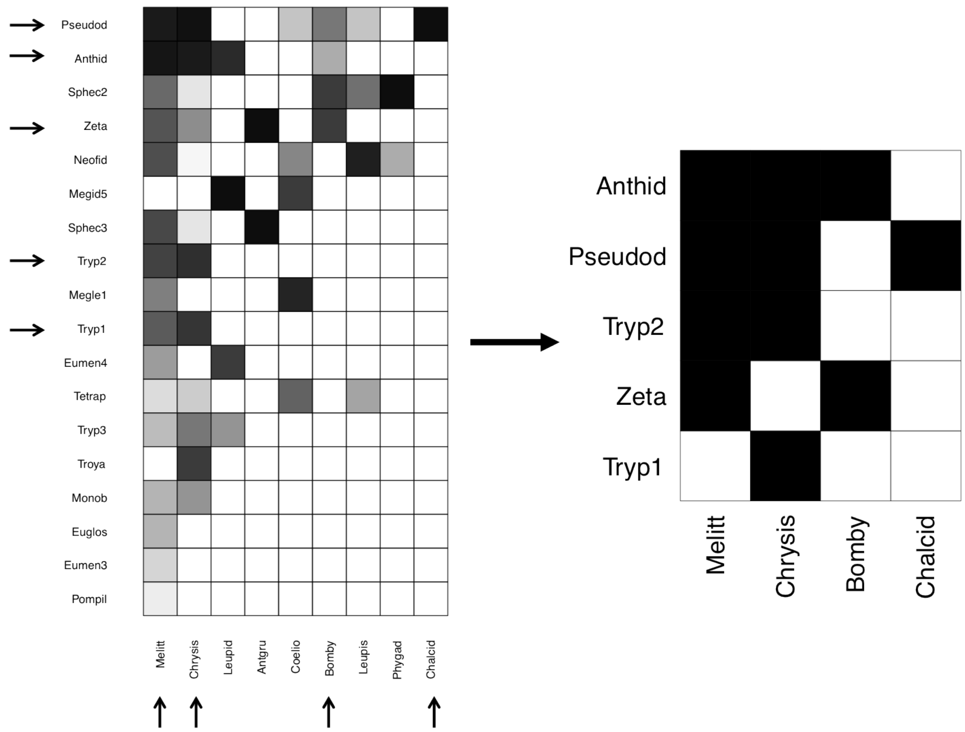
Figure 2

The integrated niche. The Grinnellian niche is the set of environmental conditions where the intrinsic growth rate r is positive (axioms i, ii & iii). Contingencies (axiom iv) such as disturbances and stochastic extinctions, in conjunction with limited dispersal (axiom v), restrict species distribution to the conditions where the colonization rate c is larger than extinction rate e. The Eltonian niche on the other dimension is represented by the set of traits allowing species to interact (axiom viii). The red species is a predator with a trophic position n, feeding on species whose niche is within a certain range around the niche optimum c. The integrated niche combines the effects of the environment and ecological interactions. The central square represents the area where the joint probability of observing interactions and co-occurrence is positive.

Figure 3

Uncertainty in the evaluation of the metaweb. The inferred metaweb for the XX environment is represented (left), along with the uncertainty in the evaluation of the interaction probability (right). Note that the standard error for probabilities is not symetric and thus only the upper bound is represented.

Figure 1


Pseudod
Anthid
Sphec2
Zeta
Neofid
Megid5
Sphec3
Tryp2
Megle1
Tryp1
Eumen4
Tetrap
Tryp3
Troya
Monob
Euglos
Eumen3
Pompil
Melitt
Chrysis
Leupid
Antgru
Coelio
Bomby
Leupis
Phygad
Chalcid
Anthid
Pseudod
Tryp2
Zeta
Tryp1
Melitt
Chrysis
Bomby
Chalcid

292 **Figure 2**

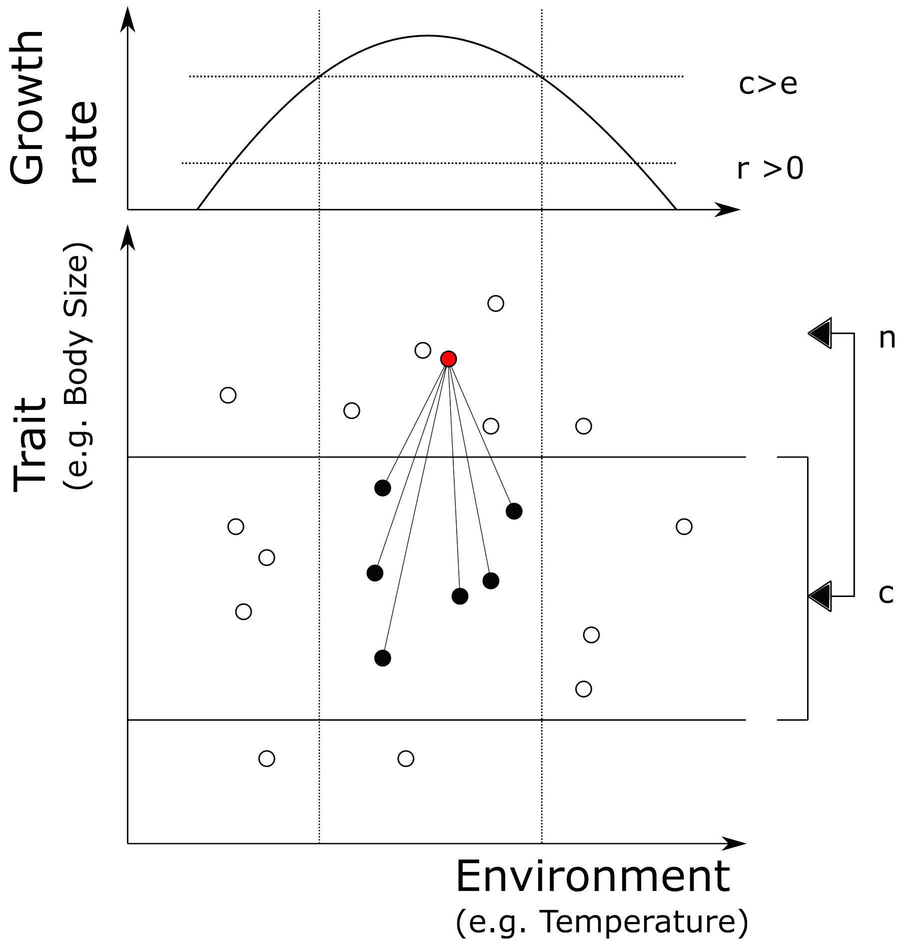
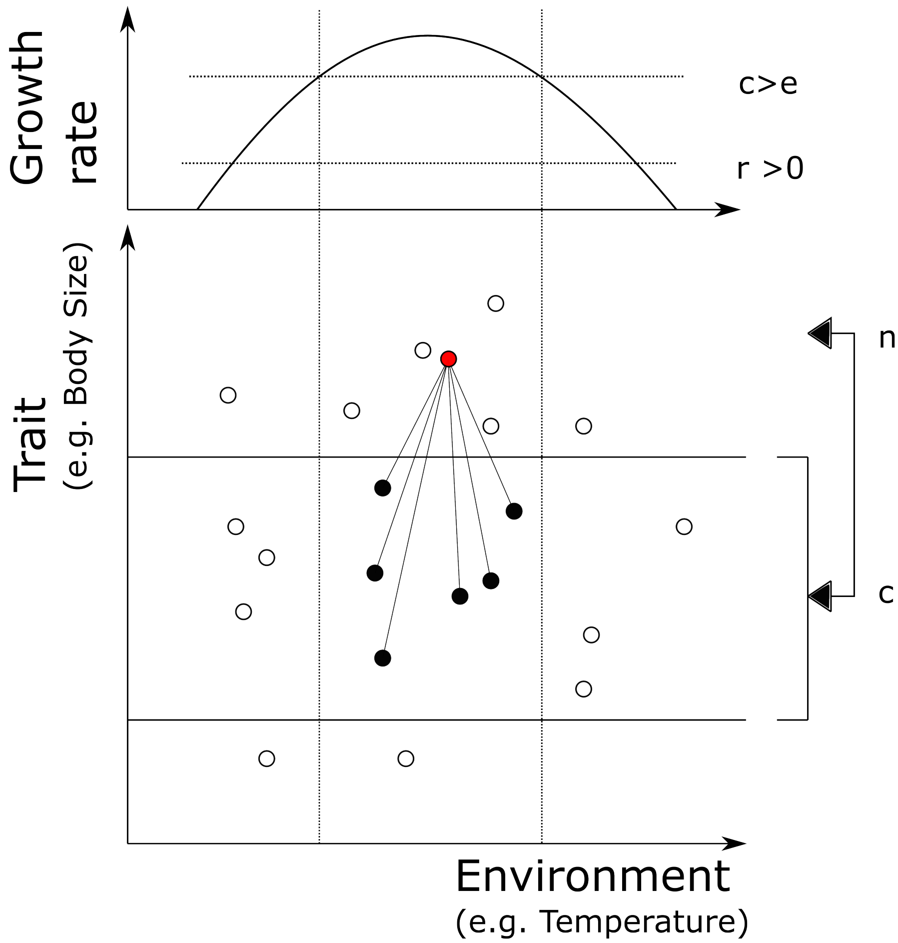

Growth rate
c>e
r >0
Trait
(e.g. Body Size)
n
c
Environment
(e.g. Temperature)

Figure 3


Growth
rate
c>e
r >0
Trait
(e.g. Body Size)
n
c
Environment
(e.g. Temperature)

Name	Equation	Details
Metaweb		
Constant	$P(L_{ijy}\|X_{iy}, X_{jy})$	Interaction probability is invariant to the environment
Conditional	$P(L_{ijy}\|X_{iy}, X_{jy}, E_y)$	Interaction probability is a function of the local environment
Deterministic	$P(L^{*}_{ijy}\|X_{iy}, X_{jy})$	Interaction occurs whenever both species are present
Co-occurrence		
Constant	$P(X_{iy}, X_{jy})$	Species distribution independent of E
Conditional on E	$P(X_{iy}, X_{jy}\|E_y)$	Similar to a SDM applied to co-occurrence
Neutral	$P(X_{ix}\|E_y)P(X_{jy}\|E_y)$	Independent SDMs fit to both species; could be independent of E
Conditional on L_y	$P(X_{iy}, X_{jy}\|L_y)$	Could account for first and higher order interactions

Table 1: List of different models

Model	**Metaweb**			**Co-occurrence**			$L(H\|D)$	AIC
	Constant	Cond. on E	Deterministic	Constant	Cond. on E	Neutral		
1.	X				X		0	0
2.		X			X		0	0
3.			X		X		0	0
4.		X		X			0	0
5.		X			X		0	0
6.		X				X	0	0

Table 2: Model comparison with the host-parasitoid networks. The 48 networks were fitted to different models of interaction networks. Note that for the computation of the likelihood all null interaction probabilities, co-occurrences and the pairwise interactions without observed co-occurrences were removed.