

**Université de Montréal**

**Vers une théorie de l'entropie maximale  
des réseaux trophiques**

par

**Francis Banville**

Département de sciences biologiques  
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de  
Philosophiæ Doctor (Ph.D.)  
en sciences biologiques

19 août 2024



**Université de Montréal**

Faculté des arts et des sciences

Cette thèse intitulée

**Vers une théorie de l'entropie maximale  
des réseaux trophiques**

présentée par

**Francis Banville**

a été évaluée par un jury composé des personnes suivantes :

*Frédérique Dubois*

\_\_\_\_\_  
(présidente-rapporteuse)

*Timothée Poisot*

\_\_\_\_\_  
(directeur de recherche)

*Dominique Gravel*

\_\_\_\_\_  
(codirecteur)

*Andrea Paz Velez*

\_\_\_\_\_  
(membre du jury)

*Jean-Gabriel Young*

\_\_\_\_\_  
(examinateur externe)

\_\_\_\_\_  
(représentant du doyen de la FESP)



## Résumé

---

Pour bien comprendre le fonctionnement des écosystèmes et mettre en place des mesures de conservation efficaces, nous devons développer une compréhension approfondie des interactions entre espèces. Parmi ces interactions, celles de nature trophique, comme la prédation et l'herbivorie, jouent un rôle écologique important en affectant la dynamique des populations. Cependant, de nombreuses incertitudes subsistent concernant l'occurrence des interactions trophiques. D'abord, ces incertitudes résultent de notre manque de connaissances empiriques sur ces interactions et sur les conditions favorisant leur réalisation. Ensuite, elles découlent de leur variabilité intrinsèque dans le temps et l'espace. La prise en compte de ces deux grandes sources d'incertitude (manque de connaissances et variabilité locale) a conduit plusieurs écologistes à représenter les interactions entre espèces de manière probabiliste au sein de réseaux écologiques complexes. Dans cette thèse, je soutiens qu'une telle vision probabiliste des systèmes écologiques complexes constitue le fondement d'une théorie de l'entropie maximale des réseaux trophiques.

Le Chapitre 1 examine en détail les sources d'incertitude des interactions entre espèces. J'y distingue deux types de représentation probabiliste des interactions entre espèces: les interactions locales décrivant la probabilité qu'elles soient réalisées à un endroit et moment particuliers, et les interactions régionales décrivant la probabilité qu'elles soit biologiquement faisables. À l'aide d'un jeu de données d'interactions hôtes-parasites en Europe, je montre que ces deux représentations diffèrent dans leurs propriétés statistiques, la probabilité d'interaction locale variant dans le temps et l'espace, mais pas la probabilité d'interaction

régionale. Bien comprendre les sources d'incertitude des interactions entre espèces permet d'identifier les réseaux susceptibles de respecter le principe d'entropie maximale.

Le Chapitre 2 présente pourquoi les réseaux locaux ont une entropie (ou complexité) maximale dans la limite d'un nombre restreint de contraintes écologiques. J'ai développé deux approches (analytique et heuristique) pour prédire la structure des réseaux trophiques locaux avec le principe d'entropie maximale. L'approche analytique m'a permis d'obtenir la distribution du nombre de proies et de prédateurs pour chaque espèce dans un réseau à partir du nombre total d'espèces et d'interactions. L'approche heuristique, quant à elle, m'a permis de trouver que cette même distribution est l'une des contraintes écologiques principales qui déterminent la structure des réseaux trophiques. Mes modèles prédisent mieux la structure des réseaux trophiques aquatiques et terrestres échantillonnés à l'échelle globale, que des modèles nuls basés sur les mêmes contraintes.

Cette thèse jette les bases de la théorie de l'entropie maximale des réseaux trophiques. Cette théorie permet de faciliter l'étude des réseaux trophiques locaux en réduisant la quantité d'informations nécessaires à leur prédiction, tout en mettant en lumière la simplicité des mécanismes biologiques qui régissent les systèmes écologiques complexes.

**Mots clés** : réseaux trophiques, interactions entre espèces, prédation, parasitisme, incertitude, variabilité locale, entropie, modélisation mathématique, écologie

# Abstract

---

To better understand ecosystem functioning and implement effective conservation measures, we need to develop a thorough understanding of species interactions. Among these interactions, trophic interactions such as predation and herbivory, play a significant ecological role by affecting population dynamics. However, uncertainties remain regarding the occurrence of trophic interactions. First, these uncertainties stem from our lack of empirical knowledge about these interactions and the conditions favoring their occurrence. Secondly, they arise from their intrinsic variability in time and space. Considering these two major sources of uncertainty (lack of knowledge and local variability) has led several ecologists to represent species interactions probabilistically within complex ecological networks. In this thesis, I argue that such a probabilistic view of complex ecological systems is the foundation of a maximum entropy theory of food webs.

Chapter 1 examines the sources of uncertainty in species interactions. I distinguish two types of probabilistic representation of species interactions: local interactions describing the probability that they occur at specific locations and times, and regional interactions describing the probability that they are biologically feasible. Using a dataset of host-parasite interactions in Europe, I show that these two representations differ in their statistical properties, with local interaction probabilities varying over time and space, but not regional interaction probabilities. Understanding the sources of uncertainty in species interactions helps identify networks likely to follow the principle of maximum entropy.

Chapter 2 presents why local networks are of maximum entropy (or internal complexity) within a limited number of ecological constraints. I developed two approaches (analytical

and heuristic) to predict local food-web structure using the principle of maximum entropy. The analytical approach allowed me to obtain the distribution of the number of prey and predators for each species in a network from the total number of species and interactions. The heuristic approach, on the other hand, revealed that this same distribution is one of the main ecological constraints determining food-web structure. My models better predict the structure of aquatic and terrestrial food webs sampled globally than null models based on the same constraints.

This thesis lays the foundations for the maximum entropy theory of food webs. This theory can facilitate the study of food webs by reducing the amount of information required for their prediction, while highlighting the simplicity of the biological mechanisms governing complex ecological systems.

**Keywords:** food webs, species interactions, predation, parasitism, uncertainty, local variability, entropy, mathematical modeling, ecology

# Table des matières

---

Résumé .....	5
Abstract .....	7
Liste des tableaux .....	15
Liste des figures .....	17
Liste des sigles et des abréviations .....	25
Remerciements .....	27
Avant-propos .....	31
Introduction .....	35
0.1. Les réseaux d'interactions entre espèces en tant que systèmes complexes .....	35
0.1.1. Quels mécanismes écologiques sous-tendent les interactions entre espèces? ..	36
0.1.1.1. Les interactions impliquent différents types d'échange .....	37
0.1.1.2. Les interactions sont réalisées par des individus .....	38
0.1.1.3. Les interactions locales résultent de plusieurs mécanismes écologiques ..	39
0.1.2. Comment mesurer une interaction entre espèces? .....	40
0.1.2.1. Les interactions mesurent différents aspects des échanges de matière et d'énergie .....	41
0.1.2.2. Les interactions sont intrinsèquement probabilistes .....	42
0.1.3. Quelle est la structure émergente d'un réseau d'interactions entre espèces? ..	43
0.1.3.1. Plusieurs mesures décrivent la complexité écologique .....	44

0.1.3.2. Les mesures de la structure sont corrélées entre elles .....	45
0.1.3.3. La structure est contrainte par des variables écologiques.....	46
0.2. Vers une théorie de l'entropie maximale des réseaux trophiques.....	47
0.2.1. Quels sont les fondements de la théorie?.....	47
0.2.1.1. Une théorie basée sur le principe d'entropie maximale .....	47
0.2.1.2. La théorie de l'entropie maximale de l'écologie appliquée aux réseaux..	48
0.2.2. Quels sont les champs d'application de la théorie?.....	49
0.2.2.1. Une théorie pouvant être appliquée à différents types de réseaux .....	49
0.2.2.2. Une théorie développée pour les réseaux locaux .....	49
0.2.3. Quelle est l'utilité de cette théorie?.....	50
0.2.3.1. Prédire la structure des réseaux trophiques.....	50
0.2.3.2. Identifier les mécanismes écologiques sous-jacents aux réseaux .....	52
0.3. Objectifs et organisation de la thèse.....	52
0.3.1. Chapitre 1: Décrypter les réseaux d'interactions probabilistes .....	53
0.3.2. Chapitre 2: Des modèles d'entropie maximale prédisant la structure des réseaux trophiques .....	54
<b>Chapter 1. First article. Deciphering probabilistic species interaction networks .....</b>	<b>55</b>
1.1. Introduction .....	59
1.1.1. Species interactions are variable and uncertain .....	59
1.1.2. Species interactions as probabilistic objects.....	61
1.1.3. We lack a clear understanding of probabilistic species interactions .....	63
1.2. Pairwise interactions: the building blocks of ecological networks .....	65
1.2.1. What are probabilistic interactions?.....	65

1.2.2.	What is the outcome of probabilistic interactions? .....	66
1.2.2.1.	The outcome of probabilistic interactions is usually binary .....	66
1.2.2.2.	The outcome of probabilistic interactions may also be quantitative .....	68
1.3.	Local networks: communities interacting in space and time .....	69
1.3.1.	What are local probabilistic interactions? .....	69
1.3.2.	What are local probabilistic interactions conditioned on? .....	70
1.3.2.1.	Local interactions may be conditioned on co-occurrence .....	70
1.3.2.2.	Local interactions may be conditioned on different environmental and biological factors .....	71
1.3.2.3.	Local interactions may be conditioned on biological feasibility .....	72
1.3.2.4.	Conditional variables must be explicitly stated .....	73
1.3.3.	How are local probabilistic interactions estimated? .....	76
1.4.	Metawebs: regional catalogs of interactions .....	83
1.4.1.	What are regional probabilistic interactions? .....	83
1.4.2.	What are regional probabilistic interactions conditioned on? .....	84
1.4.2.1.	Regional interactions describing biological feasibility are conditioned on traits .....	84
1.4.2.2.	Regional interactions describing ecological feasibility are conditioned on traits and environmental conditions .....	86
1.4.3.	How are regional probabilistic interactions estimated? .....	87
1.5.	Future perspectives .....	96
1.5.1.	Predicting local networks from metawebs .....	96
1.5.2.	Quantifying and reducing interaction uncertainty .....	98
1.5.3.	Relaxing the independence assumption .....	99
1.6.	Acknowledgment .....	100

<b>Chapter 2. Second article. What constrains food webs? A maximum entropy framework for predicting their structure with minimal biases .....</b>	101
2.1. Introduction .....	105
2.1.1. The constrained structure of ecological networks .....	105
2.1.2. The principle of maximum entropy: A primer for ecologists.....	108
2.1.3. Analytical and heuristic approaches .....	111
2.2. Methods .....	112
2.2.1. Analytical maximum entropy models.....	112
2.2.1.1. Joint degree distribution .....	113
2.2.1.2. Degree distribution .....	116
2.2.2. Heuristic maximum entropy models .....	117
2.2.2.1. From Shannon's to SVD entropy .....	117
2.2.2.2. Types I and II heuristic MaxEnt models .....	119
2.2.2.3. Structure of MaxEnt food webs .....	119
2.3. Results and Discussion.....	125
2.3.1. Analytical maximum entropy models.....	125
2.3.2. Heuristic maximum entropy models .....	129
2.4. Conclusion.....	137
2.4.1. Alternative MaxEnt models .....	138
2.4.2. Applications .....	139
2.4.3. Final remarks .....	141
2.5. Acknowledgments .....	141
<b>Conclusions générales .....</b>	143

2.6.	Développement de la théorie de l'entropie maximale des réseaux trophiques	143
2.6.1.	Où en sommes-nous?	143
2.6.2.	Comment améliorer et étendre la théorie?	143
2.7.	Validation de la théorie	143
2.7.1.	Quels réseaux sont d'entropie maximale?	143
2.7.2.	Un réseau d'entropie maximale est-il à l'équilibre?	143
2.8.	Applications potentielles de la théorie	143
2.8.1.	Compréhension des mécanismes sous-jacents aux réseaux d'interactions	143
2.8.2.	Prévision de la structure des réseaux trophiques	143
<b>Références</b>		145
<b>Appendix A.</b>	<b>Supplementary material for chapter 1</b>	173
<b>Appendix B.</b>	<b>Supplementary material for chapter 2</b>	181



## Liste des tableaux

---

- 1    **Mathematical expressions of probabilistic interactions.** The probability of interaction between two taxa  $i$  and  $j$  is interpreted differently in a local network  $L_k$  of realized interactions, a local network  $O_k$  of observed interactions, a metaweb  $M$  of potential interactions (representing the *biological* feasibility of interactions), and a metaweb  $M^*$  of potential interactions (representing the *ecological* feasibility of interactions). Each expression emphasizes a different conditional variable, the ellipsis serving as a placeholder for other variables not explicitly stated in the expression. The outcome of each of these probabilistic events, along with common models used for estimation, is presented alongside examples of studies that employed them (with specific variables indicated in parentheses, when applicable). The study marked with an asterisk has been conducted on binary interaction networks. The boxes in our study that discuss these expressions are also specified. .... 75
- 1    **Standardized mean difference between predicted network measures and empirical data for all food webs in our complete dataset ( $N = 257$ ).** Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence.  $\rho$ : nestedness measured by the spectral radius of the adjacency matrix.  $maxtl$ : maximum trophic level.

*diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self-loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy. .... 130

2 **Standardized mean difference between predicted network measures and empirical data for all food webs in our abundance dataset ( $N = 19$ )**. Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence.  $\rho$ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level.  
*diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self-loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy. .... 131

## Liste des figures

---

- 1    **Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (Eq 1.3.8) under different values of  $\lambda_k$  (interaction rate) and  $P(X_{i,j,k})$  (probability of co-occurrence), with  $t_0 = 1$  (duration). The probability of local interaction represents the probability that the two taxa will interact at least once within the given time interval. Parameters  $t_0$  and  $\lambda_k$  have complementary units (e.g.,  $t_0$  in months and  $\lambda_k$  in number of interactions per month). The parameter values used in the right panel are denoted by the white stars. (b) Scaling of the probability of interaction with the duration parameter  $t_0$ , for different values of  $\lambda_k$  and  $P(X_{i,j,k})$ ..... 79
- 2    **Network accumulation curves.** (a) Dissimilarity in species composition and (b) dissimilarity of interactions between common species between aggregated local networks and the metaweb of binary host-parasite interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of interactions and (d) scaling of connectance with the number of sampled (aggregated) binary and probabilistic local interaction networks. For a better comparison with binary interactions, local networks of probabilistic interactions were derived from a metaweb of probabilistic interactions with a false positive and false negative rate of zero. A specific value of  $P(L_{i,j,k}|M_{i,j})$  (the local probability of interaction among potentially interacting species) was used for all non-aggregated local networks within a particular curve. Aggregated local networks were obtained

by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions (with the value of $P(L_{i,j,k} M_{i,j})$ increasing in aggregated local networks of probabilistic interactions).....	82
<b>3 Spatial scaling of interactions.</b> Expected number of host-parasite interactions in a network aggregating all (a) local and (b) regional probabilistic interactions within a latitudinal window of a given width. Every dashed curve corresponds to a different window centered at a given latitude (color bar), with the pink solid line representing the median number of interactions across windows. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified width and position (central latitude). Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from regional probabilistic interactions by setting the value of $P(L_{i,j,k} M_{i,j})$ (the local probability of interaction among potentially interacting species) to 1. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of $P(L_{i,j,k} M_{i,j})$ remaining at their maximum value of 1.....	90
<b>4 Connectance of sampled binary interaction networks.</b> (a-c) Average connectance of binary interaction networks obtained from the two sampling techniques for 20 randomly selected host-parasite networks. Cross markers represent the connectance of a single sample for each network, diamond markers the average connectance across 10 samples, hexagon markers the average connectance across 50 samples, and the colored circles the average connectance across 100 samples (marker size proportional to the number of samples). (d-f) Reduction in the mean squared logarithmic error between the average connectance of binary interaction networks (all 233 host-parasite networks) obtained from	

these two sampling techniques as the number of samples increases. The local probability of interaction between potentially interacting species was set to three different values: (a,d)  $P(L_{i,j,k}|M_{i,j}) = 1.0$ , (b,e)  $P(L_{i,j,k}|M_{i,j}) = 0.75$ , and (c,f)  $P(L_{i,j,k}|M_{i,j}) = 0.50$ . Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Regional samples were obtained by randomly sampling binary interactions from the probabilistic interaction metaweb, and then propagating this result to all local networks that include the species potentially engaged in the interactions. Local samples were obtained by independently sampling binary interactions for each local network of probabilistic interactions. .... 95

- 1 **Maximum entropy degree distributions with predicted numbers of interactions.** (a) Probability density of the mean degree of a food web obtained using different values of species richness  $S$ . The number of interactions  $L$  was simulated 1000 times using the flexible links model fitted to all empirical networks. The mean degrees  $2L/S$  were then obtained from these simulated values. (b) Degree distributions of maximum entropy for a network of  $S = 27$  species and different numbers of interactions. The numbers of interactions correspond to the lower and upper bounds of the 67%, 89%, and 97% percentile intervals (PI), as well as the median (in blue), of the counterfactuals of the flexible links model. Each degree distribution of maximum entropy was obtained using Eq 2.2.13 after solving numerically Eq 2.2.14 using different values of the mean degree constraint. 124
- 2 **Prediction errors of the relative number of predators and prey.** The relative number of predators ( $k_{in}$ ) is plotted against the relative number of prey ( $k_{out}$ ) for each species in all (a) empirical and (b) predicted joint degree sequences. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping

the total number of interactions constant. (c) Difference between predicted and empirical values when species are ordered according to their total degree. Due to significant data overlap, all relationships are represented as 2D histograms. The color bar indicates the number of species that fall within each bin. ....	125
<b>3 Shape of empirical and predicted joint degree sequences.</b> (a) Probability density of KL divergence between in and out-degree sequences of empirical and predicted joint degree sequences. (b) Difference between the KL divergence of empirical and predicted joint degree sequences as a function of connectance. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant.....	128
<b>4 Relationship between the structure of empirical and maximum entropy food webs.</b> Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (a) Nestedness (estimated using the spectral radius of the adjacency matrix), (b) the maximum trophic level, (c) the network diameter, and (d) the SVD entropy were measured on these empirical and maximum entropy food webs. The identity line is plotted in each panel. ....	133
<b>5 Proportions of single-link three-species motifs in empirical and predicted food webs.</b> S1: Tri-trophic chain (a top predator feeds on a meso-predator which feeds on a basal prey). S2: Omnivory (a top predator feeds on a meso-predator and a basal prey). S3: Tri-trophic feeding loop (a cyclic three-species predator-prey system). S4: Apparent competition (a predator feeds on two prey). S5: Exploitative competition (two predators feed on the same prey). Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the	

joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. Boxplots display the median proportion of each motif in food webs (middle horizontal lines), as well as the first (bottom horizontal lines) and third (top horizontal lines) quartiles. Vertical lines encompass all data points that fall within 1.5 times the interquartile range from both quartiles, and dots are data points that fall outside this range. Only the single-link motifs S1-S5 are shown given the scarcity of double-link motifs in most empirical and predicted networks. .... 135

6 **Pairwise relationships between the proportions of single-link three-species motifs in empirical and predicted food webs.** S1: Tri-trophic chain. S2: Omnivory. S4: Apparent competition. S5: Exploitative competition. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. Regression lines are plotted in each panel. Motif S3 is not shown because of its low proportion in most empirical and predicted networks... 136

7 **Prediction errors of the absolute number of predators  $k_{in}$  and prey  $k_{out}$ .** Species were ordered according to their total degree in their network. Networks were sorted into different groups based on their total number of species. In each panel, each dot corresponds to a single species within one of the networks whose total species count is within the specified range. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant. .... 182

8 **Prediction errors of the relative number of predators  $k_{in}$  and prey  $k_{out}$ .** Species were ordered according to (a) their out-degree and (b) their in-degree.

The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant. Due to significant data overlap, all relationships are represented as 2D histograms. The color bar indicates the number of species that fall within each bin. .... 183

9 **Predicted probability that a species is isolated in its food web.** We derived many degree distributions of maximum entropy given a range of values of  $S$  and  $L$  and plotted the probability that a species has a degree  $k$  of 0 (log-scale color bar). Species richness varies between 5 and 100 species, by increment of 5 species. For each level of species richness, the numbers of interactions correspond to all 20-quantiles of the interval between 0 and  $S^2$ . The black line marks the  $S - 1$  minimum number of interactions required to have no isolated species. .... 184

10 **SVD entropy of empirical and predicted food webs.** (a) Distribution of the SVD entropy of empirical and maximum entropy food webs. Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (b) Distribution of z-scores of the SVD entropy of all empirical food webs. Z-scores were computed using the mean and standard deviation of the distribution of SVD entropy of MaxEnt food webs (type II heuristic MaxEnt model). The dashed line corresponds to the median z-score. .... 185

11 **Prediction errors of SVD entropy.** Difference in SVD entropy between maximum entropy and empirical food webs as a function of (a) the number of interactions, (b) connectance, and (c) species richness. (d) Standardization of the difference in SVD entropy with respect to species richness as a function of species richness. The exponential decrease in the difference of SVD entropy per species with species richness offers a complementary perspective supporting the lack of

relationship depicted in panel c. Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. .... 186

- 12 **Structure of empirical and maximum entropy food webs as a function of species richness.** Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (a) Nestedness (estimated using the spectral radius of the adjacency matrix), (b) the maximum trophic level, (c) the network diameter, and (d) the SVD entropy were measured on these empirical and maximum entropy food webs and plotted against species richness. Regression lines are plotted in each panel. .... 187



## Liste des sigles et des abréviations

---

cannib	Proportion d'espèces cannibales
diam	Diamètre du réseau
KL	Divergence de Kullback–Leibler
MaxEnt	Entropie maximale, de l'anglais <i>Maximum entropy</i>
METE	Théorie de l'entropie maximale de l'éologie, de l'anglais <i>Maximum entropy theory of ecology</i>
maxtl	Niveau trophique maximale, de l'anglais <i>Maximum trophic level</i>
MxSim	Similarité maximale moyenne entre paires d'espèces, de l'anglais <i>Average maximum similarity between species pairs</i>

Omniv	Proportion d'espèces omnivores
PI	Intervalle de percentile, de l'anglais <i>Percentile interval</i>
SVD	Décomposition en valeurs singulières, de l'anglais <i>Singular value decomposition</i>
Trophique-METE	Théorie de l'entropie maximale des réseaux trophiques, de l'anglais <i>Maximum entropy theory of ecology for food webs</i>

## Remerciements

---

Rédiger cette thèse de doctorat a été pour moi un pèlerinage intellectuel plus grand que toutes ces longues excursions pédestres que j'ai pu faire dans ma jeunesse. Cette promenade sur les sentiers parfois tumultueux de la connaissance scientifique n'aurait pas été possible sans le soutien indéfectible de nombreuses personnes qui m'ont accompagné tout au long de cette aventure.

Tout d'abord, merci à mes superviseurs, Timothée Poisot et Dominique Gravel, d'avoir été de si bons guides. Vous m'avez accueilli dans la cité des écologistes quantitatifs et computationnels et m'avez formé pour que je devienne l'un des vôtres. Vous m'avez aidé à préciser ma destination et m'avez montré la route pour l'atteindre. Merci pour votre support, vos conseils et votre bienveillance.

Merci aux autres professeurs qui m'ont supporté dans mes travaux et rassuré que j'allais dans la bonne direction. Merci à Allison Barner et Éric Harvey pour m'avoir conseillé et aidé à préciser mes besoins et objectifs de recherche lors des rencontres de mon comité-conseil. Merci à Éric Harvey et Erica Newman pour vos questions et propositions lors de mon examen de synthèse. Un grand merci également aux trois membres du jury qui ont accepté d'évaluer cette thèse, Andrea Paz Velez, Frédérique Dubois et Jean-Gabriel Young. Merci à Timothée Poisot et Dominique Gravel pour votre soutien et participation sur ces trois comités.

Merci à mes collègues de labo et ami-es, Alexandre Fuster, Alice Doucet Beaupré, Andrew MacDonald, Ariane Bussières-Fournel, Benjamin Mercier, Camille Rondeau Saint-Jean, Claire-Cécile Juhasz, Cole Brookson, Cristian Cruz Rodriguez, Daphnée Lecours-Tessier,

Dominique Caron, Emily Beasley, Eva Delmas, Gabriel Bergeron, Gabriel Dansereau, Gracielle Higino, Kari Norman, Katherine Hébert, Maria Isabel Arce Plata, Mathilde Besson, Mathis Gheno, Michael D. Catchen, Nerea Montes Perez, Norma Forero-Muñoz, Philippe Desjardins-Proulx, Salomé Bouskila, Samuel Provencher-Tardif, Steve Vissault, Tanya Strydom, Valentine Volz, Victor Cameron, Vincent Beauregard et Willian Vieira. Vous avez été mes compagnons de voyage, partageant avec moi différents tronçons de cette excursion qui, sans vous, aurait été bien plus solitaire.

Un merci spécial à Andrew MacDonald, qui m'a encouragé et accompagné dans mes premiers pas dans la recherche scientifique. Tu as été mon mentor et seras toujours pour moi un modèle d'écologiste attentionné, généreux et passionné. Un autre merci spécial à Gabriel Dansereau, qui a été à mes côtés tout au long de mon doctorat. Ton calme, ta sagesse et ton expertise m'ont permis d'avancer lorsqu'un obstacle se dressait sur mon chemin.

Merci à l'ensemble de mes coauteurs pour nos échanges d'idées et discussions stimulantes. Merci à Chris Brimacombe, Dominique Gravel, Gabriel Dansereau, Gracielle Higino, Hana Mayall, Kari Norman, Michael Catchen, Penelope S. A. Blyth, Tanya Strydom, Thomas Malpas et Timothée Poisot pour m'avoir aidé à améliorer la qualité de cette thèse. Grâce à vous, j'ai pu avancer beaucoup plus loin que ce que j'aurais imaginé. Merci également à mes autres coauteurs de m'avoir accueilli dans des quêtes secondaires. Ensemble, nous avons rédigé pas moins de 11 articles scientifiques au-delà de cette thèse, dont 8 ont déjà été publiés au moment où j'écris ces lignes. Loin d'être une distraction, ces détours m'ont permis de me retrouver et d'élargir mes horizons. Ce fut un réel plaisir d'explorer avec vous des sujets aussi variés que la prédiction des réseaux écologiques, la vulgarisation scientifique et la transmission des connaissances en programmation.

Merci à BIOS<sup>2</sup>, en particulier à Andrew MacDonald, Gracielle Higino et Kim Gauthier Schampaert, d'avoir développé une communauté de support pour les biologistes computationnels comme moi et d'avoir organisé des ateliers et cours d'été en statistique et programmation appliqués à l'écologie. Vous avez fait une énorme différence dans mon parcours.

Merci à Andrew Blakney, Emmanuelle Chrétien, Morgan Botrel et Stéphanie Shousha de m'avoir accueilli dans votre club de lecture. Nos discussions sur la philosophie des sciences, le féminisme et le colonialisme ont changé ma vision des sentiers académiques et ont permis de remettre le « Ph » dans mon PhD.

Merci à Émmélie Leroux, Francis Robitaille, Julie Brémaud, Mobina Gholamhosseini et Vincent Melançon pour avoir organisé avec moi le symposium 2023 du département de sciences biologiques de l'Université de Montréal. Vous avez fait d'une épreuve éprouvante l'une de mes meilleures expériences au doctorat.

Merci à Alexandre Collin de m'avoir donné l'opportunité d'enseigner une partie de ton cours de biostatistique pendant deux ans. Cette expérience m'a donné la piqûre pour l'enseignement des sciences et m'a donné le goût de montrer à mon tour le chemin à la nouvelle génération de scientifiques.

Merci à ma famille et à mes ami-es d'avoir été à mes côtés. Merci à Rebecca Burdayron d'avoir été là pour moi lorsque je débutais tout juste cette aventure. Un merci spécial à Léa Rancourt et Céleste Sorel de m'avoir aidé à me relever lorsque je traversais une période plus difficile en pleine pandémie. Merci également à mes parents et beaux-parents qui, sans connaître ma destination, m'ont encouragé et supporté tout au long de mon parcours universitaire, s'étant échelonné sur pas moins de 11 longues et belles années. Vous m'avez déposé devant des sentiers nouveaux avec tout ce dont j'avais besoin pour réussir la traversée, ayant confiance que j'allais trouver mon chemin.

Finalement, merci Sufia, ma jaan. Tu m'as accompagné dans la dernière ligne droite de notre traversée de l'île de Montréal à pied alors qu'on ne se connaissait à peine, et tu

es toujours à mes côtés maintenant que je complète les dernières étapes de cette odyssée doctorale. Merci pour ta positivité et tes encouragements constants. Avec toi mon avenir après le doctorat est radieux.

Ce travail n'aurait pas été possible sans le financement de l’Institut pour la valorisation des données (IVADO), du Programme de formation en sciences et services de la biodiversité computationnelle (BIOS<sup>2</sup>), financé par le Programme de formation orientée vers la nouveauté, la collaboration et l’expérience en recherche (FONCER) du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), des Études supérieures et postdoctorales (ESP) de l’Université de Montréal, du Département de sciences biologiques de l’Université de Montréal et des fonds du laboratoire Poisot.

## Avant-propos

---

La première moitié de mon doctorat fut marquée par la pandémie mondiale de COVID-19, qui eut des conséquences sociales et économiques sans précédent. Un virus respiratoire, le SRAS-CoV-2, a mis à rude épreuve les systèmes de santé partout dans le monde, tout en nécessitant la mise en œuvre de politiques de distanciation sociale aux conséquences psychologiques bien visibles. Maintenant derrière nous, cette épreuve nous a collectivement fait prendre conscience de la vulnérabilité de notre système de santé, en plus de mettre en lumière les conséquences possibles d'un contact trop étroit avec la faune. En dépit de cela, les activités humaines, comme la modification de l'utilisation du territoire et le rejet de gaz à effet de serre dans l'atmosphère, n'ont pas ralenti depuis la fin de la pandémie, même si nous savons qu'elles peuvent altérer la distribution et la biologie des espèces hôtes de pathogènes. Rien pour atténuer son éco-anxiété, le risque est réel de voir émerger une nouvelle crise sanitaire causée par des changements environnementaux d'origine anthropique.

C'est dans ce contexte que s'est développé mon intérêt pour les réseaux écologiques. Pour mieux anticiper et nous préparer adéquatement aux changements environnementaux à venir, que ce soit en lien avec le risque d'une future pandémie, la sécurité alimentaire mondiale ou la pérennité d'autres services écosystémiques, il est nécessaire de connaître la distribution des espèces et leurs interactions. Si les interactions hôtes-parasites, comme celles entre les chauves-souris et les bétacoronavirus, sont une préoccupation majeure en matière de santé publique, il est crucial de préserver les interactions plantes-polliniseurs afin de maintenir la culture de nombreuses plantes comestibles. De même, les interactions prédateurs-proies sont une composante clé de la réponse des écosystèmes aux changements environnementaux

puisqu'elles déterminent les flux de matière et d'énergie. Cependant, la difficulté d'observer les interactions entre espèces dans les milieux naturels complique considérablement notre capacité à développer des stratégies de gestion environnementale éclairées et à évaluer les risques qu'une nouvelle maladie infectieuse se transmette à l'humain. Je crois que l'écologie doit s'élever au rang d'une science prédictive pour répondre à ces enjeux.

Lors de mon doctorat, j'ai participé à plusieurs projets de recherche portant sur la pré-diction des réseaux d'interactions entre espèces à partir des données écologiques disponibles. Dans Strydom, Catchen, et al., 2021, nous avons proposé une feuille de route pour améliorer la pré-diction de différents types de réseaux écologiques dans le temps et l'espace, reconnaissant que les interactions entre espèces peuvent varier localement. Nous avons notamment avancé que la pré-diction des réseaux écologiques à l'échelle locale est facilitée lorsque nous disposons d'un réseau régional. Dans Strydom et al., 2022, nous avons montré qu'il est possible de reconstituer fidèlement un tel réseau régional en transférant l'information écolo-gique d'un réseau analogue. Nous avons ainsi pu prédire les interactions prédateurs-proies régionales entre les mammifères canadiens à partir d'un réseau trophique analogue européen, qui avait été auparavant échantillonné. Nous avons expliqué en détail la méthode utilisée d'intégration de graphes (une méthode d'apprentissage automatique) dans Strydom et al., 2023. Pour ce qui est des réseaux locaux, nous avons développé un modèle pour prédire le nombre total d'interactions prédateurs-proies à partir du nombre local d'espèces dans MacDonald et al., 2020. Finalement, nous avons proposé une méthode dans Higino et al., 2023 pour vérifier le réalisme écologique des réseaux trophiques locaux prédits. Bien que ces projets aient été principalement appliqués aux interactions prédateurs-proies à cause de leur importance historique en écologie, les méthodes présentées peuvent être appliquées à différents types d'interactions entre espèces.

Ma thèse s'inscrit donc dans un effort plus large visant à prédire les interactions entre espèces à différentes échelles spatiales. J'ai approfondi et développé dans ma thèse plusieurs

des idées soulevées lors de ces projets de recherche. Par exemple, lors d'un groupe de travail portant sur le développement d'une méthode d'identification des sites d'échantillonnage maximisant l'information recueillie sur les interactions entre espèces, nous avons constaté que nous n'avions pas la même interprétation des interactions probabilistes. Puisque nos modèles prédictifs estiment la probabilité qu'une interaction soit réalisée, il est crucial d'avoir une interprétation claire de leurs résultats pour pouvoir adéquatement les utiliser. La réalisation qu'il manquait une définition précise des interactions probabilistes fut la bougie d'allumage du Chapitre 1. De la même manière, c'est en travaillant sur ces modèles prédictifs que j'ai pris connaissance que la structure des réseaux d'interactions entre espèces était contrainte par un nombre limité de variables écologiques. Cette constatation m'a incité à développer les bases de la théorie de l'entropie maximale des réseaux trophiques, exposée au Chapitre 2. Cette méthode simplifie grandement la prédiction des réseaux d'interactions entre espèces en diminuant le nombre d'information requise.

Mon travail est ancré dans les principes de la reproductibilité et de la science ouverte. Je me suis engagé à partager mes données, scripts et résultats pour que ma recherche soit plus transparente. Dans Dansereau et al., 2020, nous avons reproduit l'article classique de Hastings and Powell, 1991 sur le chaos d'une chaîne alimentaire à trois espèces, montrant la faisabilité et l'importance d'un tel exercice de reproduction. J'ai confiance que le lecteur ou la lectrice désirant reproduire mes résultats de recherche trouvera cette démarche aisée. Pour aider, nous avons rédigé un article (Banville et al., 2021) expliquant les librairies principales utilisées dans cette thèse. En outre, dans Lawlor et al., 2022, nous avons rédigé un guide pour aider les biologistes à apprendre le langage de programmation R par eux et elles-mêmes. Les principes généraux de ce guide peuvent également être utiles pour débuter avec le langage de programmation Julia, qui a permis d'obtenir la majorité des résultats présentés dans cette thèse. Enfin, je suis convaincu que notre devoir en tant que scientifiques est de travailler ensemble et ouvertement pour relever les défis mondiaux tels que l'adaptation aux

changements climatiques, la préservation des habitats naturels et la prévention de futures pandémies. Je crois fermement que la compréhension des interactions entre espèces joue un rôle crucial dans cet objectif.

# **Introduction**

---

## **0.1. Les réseaux d’interactions entre espèces en tant que systèmes complexes**

Plusieurs phénomènes naturels sont régis par des systèmes complexes, qui forment un ensemble cohérent d’éléments interagissant entre eux (Rind, 1999). Ces interactions impliquent des échanges directs d’information, de matière ou d’énergie (Ladyman et al., 2013) dont l’effet conjoint génère des propriétés émergentes de niveau supérieur (Foote, 2007). Par exemple, le réseau de neurones dans un cerveau humain orchestre des milliers de signaux électriques à chaque seconde pour créer une expérience subjective du monde (Sporns, 2011). De même, une colonie de fourmis produit des comportements complexes tels que la construction de nids et la défense du territoire par le biais d’individus coopératifs qui s’échangent ressources et informations (Bonabeau et al., 1999). Caractéristiques fondamentales des systèmes complexes, ces propriétés émergentes ne sont pas déductibles des éléments pris individuellement (Nielsen and Muller, 2000). Du désordre apparent des éléments et de leurs interactions émerge donc une structure ordonnée qui est au cœur des recherches sur les systèmes complexes.

Pour faciliter l’étude des systèmes complexes, il est courant de les représenter sous forme de réseaux, où les éléments forment des nœuds et les interactions, des liens entre ces nœuds (M. E. J. Newman, 2003). Dans cette thèse, je me suis intéressé aux réseaux d’interactions entre espèces, qui servent d’outil pour mieux comprendre les systèmes écologiques complexes. Ces réseaux lient chaque paire d’espèces qui interagissent entre elles, comme un prédateur et sa proie, un parasite et son hôte ou une plante et son pollinisateur. Bien ancrée dans

la science des réseaux et la théorie des graphes, l'écologie des réseaux fournit un cadre d'analyse pour étudier les propriétés émergentes des écosystèmes, comme leur fonctionnement (p. ex. productivité primaire et cycle des nutriments) et résilience face aux changements environnementaux (Proulx et al., 2005, McCann, 2007, McCann, 2011, Rooney and McCann, 2012, Valiente-Banuet et al., 2019). Pour ce faire, elle a recours à des méthodes pour mesurer les interactions entre espèces (p. ex. Jordano, 2016) et analyser la structure émergente des réseaux d'interactions (p. ex. Delmas et al., 2019), nous permettant ainsi de mieux comprendre les caractéristiques de ces systèmes complexes.

Cette thèse jette les bases de la théorie de l'entropie maximale des réseaux trophiques, qui constitue une approche novatrice en écologie des réseaux permettant de prédire la structure émergente des réseaux complexes d'interactions à partir d'un nombre limité d'informations écologiques. Avant d'exposer cette théorie, il convient de passer en revue quelques caractéristiques importantes des systèmes écologiques complexes, notamment en ce qui a trait aux mécanismes régissant les interactions entre espèces et aux mesures des interactions et de la structure des réseaux. Plus spécifiquement, j'explique, dans les sous-sections suivantes, que les interactions entre espèces sont le résultat de nombreux mécanismes écologiques souvent difficiles à départager, que les interactions entre espèces sont intrinsèquement probabilistes et que la structure émergente des réseaux d'interactions est écologiquement et statistiquement contrainte. Ces trois observations sous-tendent la théorie proposée en décrivant plusieurs aspects fondamentaux de la complexité écologique.

### **0.1.1. Quels mécanismes écologiques sous-tendent les interactions entre espèces?**

Une communauté biologique est composée d'espèces différentes interagissant entre elles dans un espace donné (Stroud et al., 2015) lorsque les conditions environnementales le permettent. Par exemple, les milieux forestiers tempérés du Québec sont riches en espèces animales et végétales. À cause des changements climatiques et de la fragmentation des habitats naturels, on y retrouve maintenant des tiques *Ixodes scapularis* infectées par la bactérie

*Borrelia burgdorferi*, responsable de la maladie de Lyme (Ogden et al., 2009, Simon et al., 2014). Les tiques juvéniles se nourrissent sur des petits mammifères, dont la souris à pattes blanches (*Peromyscus leucopus*) qui est un réservoir compétent de la bactéries (Donahue et al., 1987). Les tiques adultes, quant à elles, se nourrissent et s'accouplent sur les cerfs de Virginie (*Odocoileus virginianus*, Lane et al., 1991). Ces deux espèces (*P. leucopus* et *O. virginianus*) s'alimentent à leur tour de glands de chêne (McShea and Schwede, 1993, Elkinton et al., 1996, Wolff, 1996, McShea, 2000), ce qui peut expliquer l'abondance élevée de tiques dans les chênaies (Ostfeld et al., 2006). Comprendre l'écologie et l'épidémiologie de la maladie de Lyme nécessite une compréhension approfondie des interactions entre espèces au sein des communautés où la tique *I. scapularis* est présente. Plus généralement, il est important d'identifier les mécanismes écologiques sous-tendant les interactions entre espèces pour mieux comprendre les conditions favorisant leur réalisation.

**0.1.1.1. Les interactions impliquent différents types d'échange.** Les interactions entre espèces peuvent avoir des effets positifs, négatifs ou neutres sur les organismes impliqués, en fonction du type d'interaction. Les interactions antagonistes, telles que les interactions hôtes-parasites, prédateurs-proies et plantes-herbivores, profitent à une espèce au détriment de l'autre. Les interactions hôtes-parasites procurent un habitat et une source d'alimentation au parasite au prix du *fitness* de l'hôte, alors que les interactions prédateurs-proies et plantes-herbivores transfèrent l'énergie de la ressource vers le consommateur, entraînant généralement la mort de la proie, mais pas nécessairement celle de la plante. Les interactions mutualistes, telles que les interactions plantes-polliniseurs, profitent au contraire aux deux espèces. Le pollinisateur se nourrit du nectar de la plante, qui en retour en profite pour disperser son pollen à des fins de reproduction. Selon le niveau de complexité représenté, les réseaux d'interactions entre espèces peuvent être composés d'un seul type d'interaction (p. ex. les réseaux trophiques composés exclusivement d'interactions trophiques prédateurs-proies et plantes-herbivores) ou de plusieurs (p. ex. les réseaux multicouches composés d'interactions antagonistes et mutualistes, Pilosof et al., 2017).

Les interactions prédateurs-proies et plantes-herbivores ont une importance particulière et historique en écologie. Les réseaux trophiques ont été introduits il y a près d'un siècle, lorsque Charles Sutherland Elton (1900-1991) a développé les concepts de chaînes et de niveaux trophiques (Elton, 1927, Elton, 1958). Plusieurs publications marquantes en écologie, comme celles de May, 1972 sur la stabilité des écosystèmes complexes et de Lotka, 1925 et Volterra, 1927 sur la dynamique des populations animales, portent sur les interactions trophiques. Les interactions prédateurs-proies et plantes-herbivores déterminent les flux de matière et d'énergie au sein des écosystèmes. Elles nous permettent de mieux comprendre des phénomènes tels que la régulation de la productivité primaire par cascade trophique (Carpenter et al., 1987), la stabilité des écosystèmes face à l'extirpation d'espèces (Dunne et al., 2002, Srinivasan et al., 2007, Staniczenko et al., 2010) et la capacité des espèces à migrer et s'adapter aux changements climatiques (Tylianakis et al., 2008, Gilman et al., 2010) et à la perte d'habitats naturels (D. M. Evans et al., 2013). De plus, les données disponibles sur les réseaux trophiques couvrent une diversité d'habitats à l'échelle mondiale (Poisot et al., 2021), ce qui nous permet de tester une variété de modèles statistiques et d'hypothèses écologiques sur ces interactions. C'est pourquoi les réseaux trophiques occupent une place centrale dans cette thèse.

Les mécanismes écologiques qui sous-tendent différents types d'interactions sont essentiellement les mêmes, bien que leurs spécificités puissent varier. Par exemple, toutes les interactions directes reposent sur des règles de correspondance des traits (Poisot et al., 2015). Cependant, les traits pertinents varient d'un type d'interaction à un autre (p. ex. dentition d'un prédateur pour une interaction trophique ou longueur de l'étamine pour une interaction plante-pollinisateur). L'ensemble des interactions écologiques directes est réalisé localement par des individus, permettant ainsi de discuter communément de leurs mécanismes écologiques.

**0.1.1.2. Les interactions sont réalisées par des individus.** Les individus sont à la base des interactions entre espèces puisque ce sont eux qui effectuent les échanges d'information, de matière ou d'énergie. Par exemple, c'est du sang d'un cerf de Virginie particulier que

s'alimentera une tique adulte. Les mécanismes sous-tendant les interactions entre espèces sont donc basés sur ceux agissant à l'échelle individuelle (p. ex. le comportement d'une proie qui détermine sa vulnérabilité à un prédateur, Choh et al., 2012). Il va sans dire que les réseaux d'interactions entre individus (réseaux dont les noeuds sont des individus) représentent plus exhaustivement les échanges ayant lieu au sein des écosystèmes. Ils sont cependant difficiles à échantillonner, le peu de données disponibles sur les interactions à cette échelle étant biaisées en faveur d'organismes plus simples à observer (Guimarães, 2020). Pour cette raison, les interactions entre individus sont le plus souvent groupées en des niveaux d'organisation supérieures, connectant différentes populations, espèces ou clades entre eux (Elton, 1927). Les réseaux d'interactions entre espèces, qui lient deux espèces lorsque certains de leurs individus interagissent, ne capturent donc qu'une partie de la complexité écologique. Ils constituent néanmoins une représentation adéquate du rôle écologique des espèces (Delmas et al., 2019) et des propriétés émergentes des écosystèmes (Loreau, 2010, McCann, 2011, Bascompte and Jordano, 2013, Gonzalez et al., 2020). Comprendre comment les processus écologiques agissant à l'échelle individuelle se combinent pour supporter les interactions entre espèces nous aide à mieux saisir les déterminants des réseaux d'interactions entre espèces.

**0.1.1.3. Les interactions locales résultent de plusieurs mécanismes écologiques.** Cette thèse distingue les interactions cataloguées à l'échelle régionale de celles réalisées à l'échelle locale. Les réseaux régionaux (ou *metaweb*, Pascual and Dunne, 2006) sont constitués d'interactions potentielles, c'est-à-dire d'interactions biologiquement ou écologiquement faisables sans nécessairement être réalisées localement (Tylianakis and Morris, 2017). Par exemple, les interactions trophiques régionales du cerf de Virginie incluent l'ensemble des plantes et champignons dont il peut se nourrir dans différents contextes. Une interaction régionale indique si deux espèces possèdent des traits leur permettant d'interagir dans des conditions idéales. Les réseaux régionaux sont donc un catalogue de l'ensemble des interactions observées (p. ex. Maiorano et al., 2020) ou prédites (p. ex. Strydom et al., 2022) entre un groupe d'espèces données. Contrairement aux réseaux locaux, les réseaux régionaux ne varient pas spatialement ni temporellement puisqu'ils incluent l'ensemble des interactions

locales entre un groupe d'espèces. Les mécanismes biologiques sous-tendant les interactions régionales sont relativement peu nombreux et plus facile à identifier, reposant presqu'exclusivement sur les traits des espèces.

Les réseaux locaux (Poisot et al., 2012), quant à eux, sont constitués d'interactions observées ou réalisées à un endroit et moment particuliers. Ce sont eux qui décrivent les échanges directs d'information, de matière ou d'énergie réalisés par les individus à l'échelle locale. Étant donné qu'une interaction locale ne peut être réalisée que si elle est biologiquement faisable, les réseaux locaux sont des sous-réseaux du *metaweb* (Saravia et al., 2022). Deux espèces interagissent localement si leurs abondances (nombres d'individus) sont suffisamment élevées pour qu'elles puissent se rencontrer (Canard et al., 2012, Canard et al., 2014) et que leurs traits (caractéristiques biologiques telles que la taille corporelle) supportent une interaction lorsqu'elles se rencontrent (Bolnick et al., 2011, Gravel et al., 2013, Stouffer, Rezende, and Amaral, 2011). Par exemple, les habitudes alimentaires du cerf de Virginie varient localement selon la disponibilité saisonnière en nourriture (Short, 1975) et sa masse corporelle (Luna et al., 2013). Les conditions environnementales, comme la température (Angilletta et al., 2004), les précipitations (Woodward et al., 2012) et l'utilisation du territoire (Tylianakis et al., 2007), influent également sur la réalisation des interactions à l'échelle locale. Ces trois facteurs (abondances, traits, conditions environnementales) contribuent à la variation spatiale et temporelle des interactions locales (Poisot et al., 2015). Les mécanismes sous-tendant les interactions locales sont nombreux, et il peut être difficile d'identifier lesquels contribuent le plus à la réalisation des interactions.

### 0.1.2. Comment mesurer une interaction entre espèces?

Outre le type d'interaction, la manière dont nous mesurons les interactions entre espèces affecte également notre représentation des réseaux et notre interprétation de leurs propriétés. Les mesures des interactions varient selon leur niveau d'information écologique. Alors que les interactions binaires et quantitatives évaluent respectivement la présence et le poids d'une interaction, les interactions probabilistes mesurent son incertitude. Étant donné qu'il y aura

toujours une incertitude dans notre évaluation des interactions, celles-ci sont intrinsèquement probabilistes. Cependant, le choix d'une mesure d'interaction dépend des questions posées, de la qualité des données disponibles et de notre capacité de quantifier cette incertitude. Les propriétés des réseaux sont généralement calculées à partir d'une matrice d'adjacence (représentation matricielle d'un réseau) contenant la valeur (binaire, quantitative ou probabiliste) des interactions entre chaque paire d'espèces (Delmas et al., 2019), d'où l'importance du choix de mesure.

**0.1.2.1. Les interactions mesurent différents aspects des échanges de matière et d'énergie.** Les interactions binaires indiquent si un lien est présent ou non dans un réseau. La présence d'une interaction locale peut signifier que deux espèces ont été observées en train d'interagir à l'intérieur des frontières du réseau. Par exemple, observer une tique *I. scapularis* sur le pelage d'un cerf de Virginie *O. virginianus* dans une forêt donnée témoigne d'une interaction hôte-parasite entre ces deux espèces. Ne pas détecter cette interaction suggère qu'elle ne se produit pas à cet emplacement particulier. Cependant, une absence d'observation peut également être un faux négatif (Blüthgen, 2010, Chacoff et al., 2012, Stock et al., 2017), c'est-à-dire une interaction non observée qui est en réalité réalisée. Étant donné la difficulté d'échantillonner l'ensemble des interactions entre espèces dans un lieu donné (Jordano, 2016), ces faux négatifs sont fréquents dans les jeux de données d'interactions entre espèces. Les méthodes prédictives des interactions entre espèces (Strydom, Catchen, et al., 2021) pallient cette difficulté d'observation en prédisant lesquelles sont réalisées à l'échelle locale. Il est cependant nécessaire de tester ces méthodes sur des données empiriques fiables de présence-absence pour pouvoir utiliser les réseaux prédits en toute confiance (Brimacombe et al., 2024). Bien qu'elles contiennent peu d'information écologique comparativement aux autres mesures d'interaction, les interactions binaires ont été plus fréquemment étudiées en écologie (Pascual and Dunne, 2006, Delmas et al., 2019), notamment en raison de leur plus grande facilité d'échantillonnage (Jordano, 2016) et de prédiction (Strydom, Catchen, et al., 2021).

Les interactions quantitatives mesurent quant à elles le poids d'une interaction (Berlow et al., 2004). Ce poids peut représenter le flux d'énergie ou de biomasse (p. ex. Benke et al., 2001, Post, 2002, Bersier et al., 2002, Borrett and Scharler, 2019), l'impact démographique (p. ex. Paine, 1992, Kokkoris et al., 2002, Emmerson and Raffaelli, 2004) ou la fréquence d'interaction (p. ex. Herrera, 1989, Montoya et al., 2003) entre deux espèces. Par exemple, le taux de cerfs de Virginie infectés par *I. scapularis* est une mesure du poids d'une interaction locale. Quel que soit l'effet mesuré, le poids d'une interaction est souvent difficile à évaluer, surtout lorsque cet effet est variable dans le temps ou qu'il suit une fonction non linéaire ou densité-dépendante (Wootton and Emmerson, 2005). De plus, puisque les interactions binaires peuvent être déduites des interactions quantitatives, il est nettement plus ardu d'échantillonner et de prédire ces dernières. Toutefois, lorsqu'ils sont bien mesurés, les réseaux d'interactions quantitatives décrivant les flux d'énergie ou de biomasse entre espèces semblent être une bonne représentation des échanges directs se produisant au sein des systèmes écologiques complexes.

**0.1.2.2. Les interactions sont intrinsèquement probabilistes.** Que nous ayons évalué la présence ou le poids d'une interaction, il est important de quantifier et de bien comprendre l'incertitude de nos estimations (c.-à-d. notre niveau de croyance quant à la présence ou valeur d'une interaction). En raison des défis que posent l'échantillonnage des interactions entre espèces (Jordano, 2016), celles-ci possèdent presque toujours une incertitude, surtout quand elles n'ont pas été directement observées ou qu'elles ont été prédites à l'aide d'une méthode numérique. Cette incertitude peut avoir plusieurs sources, incluant les erreurs de mesure des données servant à prédire une interaction (p. ex. données de taille corporelle, Gravel et al., 2013), l'intervalle de confiance des paramètres du modèle et l'existence de modèles prédictifs alternatifs (Simmonds et al., 2022, Simmonds et al., 2024). La variabilité des interactions locales dans le temps et l'espace (Poisot et al., 2015) introduit une source additionnelle d'incertitude en limitant le transfert d'information d'un réseau à un autre, par exemple lors de l'inférence d'interactions locales à partir d'un réseau régional (Dansereau

et al., 2024). Documenter l'incertitude des interactions entre espèces offre une vision plus juste de la fiabilité des données d'interactions.

Les interactions probabilistes mesurent la probabilité qu'une interaction soit présente (Poisot, Cirtwill, et al., 2016). À l'échelle locale, elles représentent la probabilité qu'une interaction soit réalisée à un endroit et moment particuliers. Par exemple, si nous n'avons pas observé d'interaction entre un cerf de Virginie et *I. scapularis* dans une forêt donnée, nous pouvons évaluer la probabilité que ces deux espèces interagissent localement à partir de leurs abondances relatives (Canard et al., 2014) ou de nos connaissances sur le cycle de vie du parasite. À l'échelle régionale, les interactions probabilistes représentent plutôt la probabilité qu'une interaction soit biologiquement ou écologiquement faisable (Strydom et al., 2023). Cette probabilité d'interaction régionale est particulièrement utile lorsque deux espèces n'ont jamais été observées en train d'interagir, mais qu'elles pourraient possiblement le faire dans les bonnes circonstances. Qu'elles soient locales ou régionales, les interactions probabilistes sont généralement évaluées de manière indépendante. Cela peut introduire des biais lorsque nous propagons l'incertitude des interactions locales vers la structure du réseau (Poisot, Cirtwill, et al., 2016), à cause de l'impact que les interactions peuvent exercer les unes sur les autres au sein d'un réseau local (Golubski and Abrams, 2011, Ims et al., 2013). Néanmoins, les interactions probabilistes demeurent le moyen le plus courant de représenter l'incertitude inhérente aux interactions entre espèces.

### **0.1.3. Quelle est la structure émergente d'un réseau d'interactions entre espèces?**

Les propriétés émergentes d'un système écologique complexe peuvent être évaluées en mesurant la structure de son réseau d'interactions, c'est-à-dire en calculant les propriétés de la matrice d'adjacence. Plusieurs mesures ont été proposées pour analyser la structure des réseaux d'interactions binaires (Delmas et al., 2019), quantitatives (Bersier et al., 2002) et probabilistes (Poisot, Cirtwill, et al., 2016). Plusieurs de ces mesures sont de bons descripteurs de la complexité des écosystèmes (Landi et al., 2018). Elles peuvent décrire des valeurs

uniques (p. ex. nombre d'espèces, nombre d'interactions, connectance) ou des distributions relatives aux espèces (p. ex. distribution de degrés) ou aux interactions (p. ex. distribution des poids d'interaction). La structure des réseaux est associée à la dynamique (Pascual and Dunne, 2006) et au fonctionnement des écosystèmes, qui détermine les flux d'énergie, le cycle des nutriments et la régulation des populations (McCann, 2011, Thompson et al., 2012). Malgré le nombre élevé de mesures développées, plusieurs covarient entre elles ou sont contraintes par des variables écologiques. Cela suggère que la structure des réseaux puisse être déterminée par un nombre limité de variables. Ici, je propose un aperçu des mesures de la structure d'un réseau, en mettant l'accent sur celles décrivant la complexité des réseaux d'interactions entre espèces.

**0.1.3.1. Plusieurs mesures décrivent la complexité écologique.** Le niveau de complexité d'un écosystème peut être évalué à l'aide de différents descripteurs (Landi et al., 2018). Les plus simples d'entre eux sont le nombre d'espèces (p. ex. May, 1972) et le nombre d'interactions (p. ex. Okuyama and Holland, 2008). Ces deux mesures déterminent la connectance d'un réseau, soit la proportion des interactions possibles qui sont réalisées. La connectance est un descripteur clé de la structure des réseaux d'interactions entre espèces (Martinez, 1992) et de sa complexité (p. ex. Rozdilsky and Stone, 2001), une connectance élevée indiquant que les espèces sont davantage connectées. Elle a été associée à plusieurs propriétés émergentes des écosystèmes, notamment à l'invasibilité des habitats par des espèces envahissantes, qui ont plus de difficulté à intégrer un réseau fortement connecté (Smith-Ramesh et al., 2017). La stabilité des réseaux trophiques face à l'extirpation d'espèces augmente également avec la connectance puisqu'un prédateur peut plus facilement s'adapter en cherchant d'autres sources de nourriture après la disparition d'une de ses proies (Dunne et al., 2002). Toutefois, on observe la relation inverse dans les réseaux mutualistes lorsque l'on tient compte de la dépendance mutuelle des espèces et de l'importance relative de chaque interaction (Vieira and Almeida-Neto, 2015). Ces descripteurs capturent tous un aspect important de la complexité écologique, soit sa faible prédictibilité (Strydom, Dalla Riva, and Poisot, 2021).

Un autre descripteur de la complexité d'un écosystème est la distribution de degrés (Landi et al., 2018). Le degré d'une espèce est le nombre d'interactions qu'elle réalise avec les autres espèces dans son réseau, et la distribution de degrés est la distribution de probabilité de ces degrés. Dans les réseaux dirigés, où une interaction se fait dans un sens particulier (p. ex. prédateurs vers proies), la distribution conjointe des degrés est la distribution de probabilité conjointe décrivant les degrés entrants (nombre d'interactions entrantes) et sortants (nombre d'interactions sortantes). La distribution de degrés est une mesure écologiquement informative, nous éclairant sur le rôle écologique des espèces (Solé and Montoya, 2001, Dunne et al., 2002, Memmott et al., 2004), les mécanismes sous-jacents aux interactions entre espèces (Williams, 2011) et l'assemblage des réseaux (Vázquez, 2005).

D'autres mesures ont été proposées pour décrire la complexité d'un réseau écologique. Par exemple, Strydom, Dalla Riva, and Poisot, 2021 suggèrent l'utilisation de l'entropie SVD (de l'anglais *Singular Value Decomposition*) comme descripteur de la complexité interne d'un réseau. Cette mesure, basée sur l'entropie de Shannon (Shannon, 1948), quantifie l'information contenue dans les rangs de la matrice d'adjacence. Elle représente une façon plus robuste de quantifier la complexité d'un réseau que les autres mesures mentionnées. L'entropie SVD a été utilisée comme mesure de complexité dans l'application de la théorie de l'entropie maximale des réseaux trophiques présentée dans cette thèse.

**0.1.3.2. Les mesures de la structure sont corrélées entre elles.** La structure des réseaux est caractérisée par des mesures qui covarient fortement les unes avec les autres, limitant ainsi l'information écologique unique qu'elles procurent. Par exemple, un réseau contenant plus d'espèces réalise un plus grand nombre d'interactions, ces deux mesures ayant été liées par une loi de puissance (Brose et al., 2004, Riede et al., 2010). Également, la connectance détermine plusieurs autres mesures de la structure des réseaux, comme la distribution de degrés (Poisot and Gravel, 2014). D'autres mesures de la structure des réseaux covarient entre elles, comme l'emboîtement et la modularité dans les réseaux antagonistes et mutualistes (Fortuna et al., 2010). L'emboîtement décrit à quel point les diètes des espèces spécialistes sont un sous-ensemble de celles des espèces généralistes (Staniczenko

et al., 2013), alors que la modularité représente le niveau de division des espèces en modules étroitement connectées. Un réseau fortement emboîté, et donc peu modulaire, a tendance à être moins stable écologiquement (Okuyama and Holland, 2008, Bastolla et al., 2009, Thébault and Fontaine, 2010). Plusieurs propriétés des réseaux d’interactions sont corrélées les unes avec les autres, ce qui suggère que la structure des réseaux puisse être déterminée par quelques mesures seulement. Une théorie permettant d’unifier et de réconcilier ces différentes mesures dans un cadre commun fait actuellement défaut.

**0.1.3.3. La structure est contrainte par des variables écologiques.** Au-delà des corrélations observées entre les propriétés des réseaux, des principes écologiques contraignent la structure des réseaux. Ces valeurs imposent une limite biologique à la configuration que peut prendre un réseau d’interactions. Par exemple, l’obligation pour les espèces de se nourrir impose une limite inférieure au nombre d’interactions réalisées dans un réseau (MacDonald et al., 2020). Le modèle des liens flexibles développé par MacDonald et al., 2020 prédit de manière réaliste le nombre d’interactions dans les réseaux trophiques en utilisant le nombre d’espèces comme contrainte écologique. Prédire le nombre d’interactions constitue une bonne première étape vers la reconstitution des réseaux trophiques à partir de la richesse en espèces lorsque les données d’interactions sont rares (Strydom, Catchen, et al., 2021). En effet, le nombre d’espèces et le nombre d’interactions peuvent à leur tour être utilisés comme contraintes pour dériver la distribution de degrés avec le principe d’entropie maximale (Williams, 2011). Également, le niveau trophique maximal est le nombre maximal d’espèces séparant un producteur primaire d’un prédateur apical le long de chaînes trophiques (Cohen, 1978). Il représente une limite biologique au transfert d’énergie dans les réseaux trophiques (Williams and Martinez, 2004). Tenir compte de ces contraintes peut faciliter la prédiction de la structure des réseaux d’interactions en réduisant le nombre de configurations écologiquement possibles.

## 0.2. Vers une théorie de l'entropie maximale des réseaux trophiques

Cette thèse pose les fondements d'une théorie des réseaux trophiques prédisant leur structure émergente à partir d'un nombre limité de variables d'état (variables écologiques caractérisant la structure des réseaux trophiques). Cette théorie utilise ces variables écologiques pour contraindre des distributions de probabilité représentant différents aspects de la structure des réseaux trophiques (p. ex. leur distribution de degrés). Pour ce faire, elle identifie la distribution la moins biaisée parmi celles qui satisfont les contraintes utilisées, c.-à-d. celle qui ne fait aucune supposition sur la forme de la distribution au-delà des informations fournies par les contraintes écologiques. Cette théorie est enracinée dans la vision probabiliste des réseaux d'interactions entre espèces puisqu'elle permet d'inférer des distributions de probabilité à différents niveaux (c.-à-d. à l'échelle de l'espèce, de l'interaction ou du réseau). Selon les variables utilisées et les mesures prédites, différentes versions de la théorie peuvent être développées, celles présentant peu de support empirique étant écartées en faveur des versions mieux soutenues.

### 0.2.1. Quels sont les fondements de la théorie?

**0.2.1.1. Une théorie basée sur le principe d'entropie maximale.** Le principe d'entropie maximale (MaxEnt) est une méthode d'inférence permettant d'identifier la distribution de probabilité la moins biaisée compte tenu d'un ensemble de contraintes sur la forme de la distribution (Jaynes, 1957a, Jaynes, 1957b). MaxEnt stipule que notre meilleure estimation d'une distribution de probabilité est celle qui maximise son entropie d'information (une mesure de l'incertitude d'un évènement aléatoire telle que l'entropie de Shannon, Shannon, 1948), dans la limite des contraintes utilisées. Par exemple, si nous connaissons la moyenne d'une distribution, nous pouvons montrer que la distribution exponentielle est celle ayant la plus grande entropie parmi l'ensemble des distributions possédant la même moyenne (Frank and Smith, 2011). Cette distribution unique est dérivée uniquement à partir de nos connaissances préalables sur le système étudié, sans faire appel à des suppositions additionnelles sur

la forme de la distribution. MaxEnt ne repose sur aucun mécanisme explicite et ne nécessite pas d'ajustement aux données. La distribution de probabilité d'entropie maximale est plutôt dérivée directement à l'aide de la méthode des multiplicateurs de Lagrange (une méthode d'optimisation mathématique), dont le fonctionnement est décrit plus loin dans cette thèse. Bien qu'il n'y ait aucune garantie que cette distribution soit la mieux soutenue empiriquement, elle est celle qui représente le plus fidèlement nos connaissances préalables (c.-à-d. celle qui minimise les biais causés par des suppositions infondées).

#### **0.2.1.2. La théorie de l'entropie maximale de l'écologie appliquée aux réseaux.**

La théorie de l'entropie maximale de l'écologie (METE) prédit de nombreuses distributions d'intérêt en macroécologie à l'aide de MaxEnt (Harte, 2011). Différentes versions de cette théorie peuvent être développées en fonction des variables d'état choisies. La version ANSE repose sur quatre variables d'état : la superficie  $A_0$ , le nombre total d'individus  $N_0$ , le nombre d'espèces  $S_0$  et l'énergie métabolique totale de la communauté  $E_0$  (Harte et al., 2008, Harte and Newman, 2014). Les ratios entre ces quatre variables (p. ex. le nombre moyen d'individus par espèce) contraignent les distributions prédites par la théorie (p. ex. la distribution du nombre d'individus par espèce, Brummer and Newman, 2019). Les prédictions de la version ANSE de METE sont en grande partie conformes aux données empiriques (Harte, 2011, McGlenn et al., 2015), surtout pour les écosystèmes moins perturbés (E. A. Newman et al., 2020).

La théorie de l'entropie maximale des réseaux trophiques (Trophique-METE), mise de l'avant dans cette thèse, est une application de METE aux réseaux trophiques. Tout comme METE, elle permet de prédire plusieurs distributions écologiques à partir d'un nombre limité de variables d'état les contraignant. MaxEnt a été utilisé pour prédire la distribution de degrés de réseaux mutualistes et antagonistes à partir du nombre total d'espèces et d'interactions (Williams, 2011). Cependant, cette thèse présente la première utilisation de MaxEnt permettant de prédire simultanément plusieurs mesures de la structure des réseaux trophiques. Trophique-METE est le cadre théorique sur lequel reposent les modèles MaxEnt développés dans cette thèse.

## 0.2.2. Quels sont les champs d'application de la théorie?

**0.2.2.1. Une théorie pouvant être appliquée à différents types de réseaux.** Cette thèse développe Trophique-METE autour des réseaux trophiques locaux d'interactions entre espèces. Cependant, cette théorie peut être adaptée aux réseaux représentés à d'autres niveaux d'organisation taxonomique, en sélectionnant les contraintes écologiques conformément au niveau d'organisation choisi. Par exemple, une version de la théorie développée pour les réseaux d'interactions entre espèces peut être basée sur le nombre d'espèces, alors qu'une autre version développée pour les réseaux d'interactions entre individus pourrait être basée sur le nombre total d'individus. La théorie peut également être étendue pour prédire *simultanément* les propriétés des réseaux d'interactions entre individus et d'interactions entre espèces. À défaut d'avoir développé une telle théorie élargie, cette thèse explore comment les interactions trophiques varient en fonction de leur niveau d'organisation taxonomique, alimentant une réflexion quant à la manière de généraliser la théorie au-delà des interactions entre espèces. Par ailleurs, Trophique-METE peut être adaptée et testée sur d'autres types d'interactions impliquant un échange direct d'information, de matière ou d'énergie (comme les interactions hôtes-parasites ou plantes-polliniseurs). Une telle généralisation de la théorie au-delà des interactions prédateurs-proies et plantes-herbivores nécessite de tester plusieurs variables d'état susceptibles de contraindre la structure de différents types de réseaux écologiques. Malgré le fait que les versions de la théorie présentées dans cette thèse ne s'appliquent qu'à un nombre restreint de réseaux écologiques, le champ d'application de la théorie est en réalité plus étendu.

**0.2.2.2. Une théorie développée pour les réseaux locaux.** La théorie de l'entropie maximale des réseaux trophiques est mieux adaptée aux réseaux locaux qu'aux réseaux régionaux. En effet, Trophique-METE trouve son fondement dans la complexité des écosystèmes locaux. Les écosystèmes complexes sont constitués, par définition, d'interactions impliquant un échange direct d'information, de matière ou d'énergie (Ladyman et al., 2013), ce qui se produit à l'échelle locale. Les réseaux régionaux décrivent plutôt le potentiel que de tels échanges aient lieu, sans nécessairement que ces derniers soient réalisés. De plus,

cette théorie repose sur la prémissse selon laquelle la structure macroscopique des réseaux locaux résulte d'une agrégation des processus écologiques microscopiques se produisant à l'échelle individuelle (Frank, 2009). Le principe d'entropie maximale permet d'identifier la structure pouvant être réalisée par le plus grand nombre de configurations microscopiques différentes respectant les contraintes utilisées (Dewar and Porté, 2008). Il n'est pas clair que ces principes puissent être respectés par les réseaux régionaux contenant des interactions non réalisées. Finalement, je montre dans cette thèse que la variabilité des interactions locales introduit une incertitude qui est non réductible (c.-à-d. qui ne diminue pas avec davantage de données). Cela a pour conséquence de diminuer la prédictibilité des interactions locales, rendant possible les réseaux locaux plus complexes. La théorie de l'entropie maximale des réseaux trophiques permet d'identifier la structure la plus complexe possible des réseaux locaux, dans la limite des contraintes écologiques. Pour toutes ces raisons, les versions de Trophique-METE développées dans cette thèse ont été conçues uniquement pour les réseaux locaux.

### 0.2.3. Quelle est l'utilité de cette théorie?

La théorie de l'entropie maximale des réseaux trophiques explique la structure des réseaux d'interactions entre espèces en montrant qu'elle est écologiquement et statistiquement contrainte. À l'aide de principes statistiques simples, elle permet de prédire quantitativement et de manière cohérente différents aspects de la structure des réseaux, sans avoir recours à des méthodes d'ajustement de modèles aux données. Ces prédictions dépendent des variables écologiques utilisées pour contraindre la structure des réseaux. Identifier les variables qui génèrent les meilleures prédictions nous permet de mieux comprendre les facteurs qui déterminent la structure des réseaux trophiques. Ainsi, en tant que théorie scientifique, elle permet à la fois de générer des prédictions vérifiables sur la structure des réseaux trophiques et d'identifier les mécanismes écologiques les plus influents.

**0.2.3.1. Prédire la structure des réseaux trophiques.** L'écologie doit s'élever au rang d'une science prédictive pour que nous puissions mieux anticiper les effets des changements

environnementaux à venir (M. R. Evans et al., 2012) et informer adéquatement les processus de prises de décision environnementale (Clark et al., 2001). Étant donné la difficulté d'échantillonner localement les interactions entre espèces (Jordano, 2016) et les bias d'échantillonnage présents dans les jeux de données existants (de Aguiar et al., 2019, Poisot et al., 2021), nous sommes encore loin d'une compréhension exhaustive des interactions entre espèces (Hortal et al., 2015). Les modèles prédictifs peuvent être utilisés pour combler partiellement ce manque de données et faciliter l'étude des réseaux d'interactions à différentes échelles spatiales et temporelles. Prédire en premier lieu la structure des réseaux permet également de générer de meilleures prédictions sur les interactions entre paires d'espèces (Strydom, Catchen, et al., 2021). La plupart des modèles développés prédisant la structure des réseaux se basent sur des mécanismes particuliers, comme ceux liés à la correspondance des traits (p. ex. correspondance entre la niche écologique des espèces, Williams and Martinez, 2000) ou la dispersion et démographie des espèces (Canard et al., 2012). Cependant, puisque les mécanismes sous-jacents aux réseaux d'interactions entre espèces sont nombreux, il peut être hasardeux de construire une théorie autour d'un mécanisme particulier. De plus, la disponibilité limitée de plusieurs types de données écologiques peut rendre difficile l'utilisation des modèles prédictifs actuels (Strydom, Catchen, et al., 2021).

En étant basée sur un petit nombre de variables écologiques, Trophique-METE facilite la prédiction de la structure des réseaux trophiques en réduisant la quantité d'informations requises. Les seules données d'entrée du modèle sont les valeurs des variables d'état. Par exemple, une version de la théorie prédisant la structure des réseaux à partir du nombre d'espèces et du nombre d'interactions n'a besoin que de ces deux variables comme entrées du modèle. Elle peut prédire simultanément et de façon cohérente différents aspects de la structure des réseaux, comme la distribution de degrés ou la matrice d'adjacence. Elle le fait en maximisant l'entropie des distributions de probabilité prédictes, c'est-à-dire en trouvant la distribution qui respecte le plus fidèlement possible les connaissances préalables représentées par les contraintes écologiques.

**0.2.3.2. Identifier les mécanismes écologiques sous-jacents aux réseaux.** Bien que cette théorie ne soit pas basée sur des mécanismes écologiques explicites, elle permet d’inférer les mécanismes sous-jacents aux réseaux trophiques. Premièrement, les variables d’état capturent les mécanismes écologiques sous-jacents aux réseaux (White et al., 2012, McGill and Nekola, 2010). Par exemple, le nombre total d’espèces, qui peut être utilisé comme variable d’état dans certaines versions de la théorie, résulte de l’effet conjoint de plusieurs mécanismes écologiques (p. ex. mécanismes de dispersion, d’immigration, de spéciation et d’extinction, Whittaker et al., 2001). Les mécanismes écologiques sont donc implicites dans la théorie, étant exprimés à travers les valeurs des variables d’état. Deuxièmement, à l’instar de METE (Harte and Newman, 2014), la théorie de l’entropie maximale des réseaux trophiques peut être utilisée comme modèle nul (Fortuna and Bascompte, 2006) générant des distributions de référence (c.-à-d. qui conservent certaines propriétés désirées, comme la forme de la distribution imposée par les contraintes écologiques). En calculant l’écart entre les données empiriques et ces distributions de référence, nous pouvons évaluer dans quelle mesure les mécanismes implicites au modèle (c.-à-d. ceux capturés par les variables d’état) peuvent générer les valeurs observées. Des erreurs systématiques de prédiction suggèrent que des mécanismes importants n’ont pas été capturés par les variables d’état utilisées, ce qui peut guider le développement de versions améliorées de la théorie qui tiennent compte de ces mécanismes.

## 0.3. Objectifs et organisation de la thèse

Cette thèse formule les principes de base de la théorie de l’entropie maximale des réseaux trophiques. Pour ce faire, elle décrit d’abord les mécanismes écologiques et sources d’incertitude des réseaux d’interactions entre espèces (Chapitre 1). Puis, elle propose et teste deux premières versions de la théorie permettant de prédire une variété de mesures des réseaux trophiques (Chapitre 2). Les résultats et observations de ces deux chapitres pourraient inspirer le développement de versions améliorées et élargies de la théorie, en tenant compte des caractéristiques des systèmes écologiques complexes ainsi que des réussites et limites des

deux versions présentées. Globalement, mon travail permet de mieux comprendre les déterminants et mesures des réseaux d'interactions entre espèces, tout en fournissant un cadre d'analyse permettant de prédire différents aspects de la structure des réseaux trophiques de manière cohérente, rigoureuse et facilitée.

### **0.3.1. Chapitre 1: Décrypter les réseaux d'interactions probabilistes**

Le Chapitre 1 présente les mécanismes écologiques et sources d'incertitude de différents types de réseaux écologiques. Il décrit comment les interactions entre espèces peuvent être représentées de manière probabiliste, que ce soit au sein de réseaux locaux ou régionaux. Ce chapitre montre que ces deux types de réseaux diffèrent quant à leurs sources d'incertitude: alors que l'incertitude des réseaux locaux provient à la fois de la variation spatiale et temporelle des interactions entre espèces et d'un manque de connaissances empiriques sur leur réalisation, celle des réseaux régionaux provient exclusivement d'un manque de connaissances sur la faisabilité des interactions. Il est important d'identifier ces différentes sources d'incertitude afin de les quantifier et, le cas échéant, de les réduire. De manière connexe, ce chapitre explique également que ces deux types de réseaux diffèrent quant à leurs mécanismes sous-jacents, la réalisation des interactions locales étant conditionnelle à un plus grand nombre de facteurs, incluant leur faisabilité biologique (ou régional). Ces constats permettent de mieux comprendre les champs d'application de la théorie de l'entropie maximale des réseaux trophiques, notamment en expliquant que ce sont les réseaux locaux qui représentent le plus fidèlement les échanges directs d'information, de matière ou d'énergie se produisant au sein des systèmes écologiques complexes. La vision probabiliste des réseaux d'interactions entre espèces, approfondie dans ce chapitre, nous permet de mieux interpréter les prédictions probabilistes de la théorie.

### **0.3.2. Chapitre 2: Des modèles d'entropie maximale prédisant la structure des réseaux trophiques**

Le Chapitre 2 développe la théorie de l'entropie maximale des réseaux trophiques. Il met de l'avant deux versions de la théorie: la première utilise le nombre d'espèces et le nombre d'interactions comme variables d'état pour prédire la distribution jointe de degrés et la matrice d'adjacence, alors que la seconde utilise la distribution jointe de degrés comme variable d'état pour prédire la matrice d'adjacence. Dans ces deux versions, la distribution jointe de degrés a été prédite (le cas échéant) à l'aide d'une approche dite analytique, alors que la matrice d'adjacence a été prédite à l'aide d'une approche dite heuristique basée sur l'entropie SVD. Les deux versions ont été testées à l'aide de données empiriques de réseaux trophiques échantillonnés à l'échelle mondiale (milieux aquatiques et terrestres). La version utilisant la distribution jointe de degrés comme variable d'état prédit avec précision plusieurs mesures de la structure des réseaux calculées à partir de la matrice d'adjacence d'entropie maximale, ce qui suggère que la structure des réseaux trophiques est principalement déterminée par la distribution jointe de degrés. Cependant, cette version n'a pas réussi à prédire de manière fiable le niveau trophique maximal. Pour mieux prédire cette mesure, une version améliorée de la théorie pourrait intégrer des contraintes liées à l'énergie totale de la communauté. Pour ce faire, ce chapitre présente, de manière précise et détaillée, le principe d'entropie maximale et la méthode des multiplicateurs de Lagrange, permettant à d'autres écologistes de développer leur propre version de la théorie.

# Chapter 1 First article

---

## Deciphering probabilistic species interaction networks

by

Francis Banville<sup>1</sup>, Tanya Strydom<sup>2</sup>, Penelope S.A. Blyth<sup>3</sup>, Chris Brimacombe<sup>4</sup>, Michael D. Catchen<sup>5</sup>, Gabriel Dansereau<sup>6</sup>, Gracielle Higino<sup>7</sup>, Thomas Malpas<sup>8</sup>, Hana Mayall<sup>9</sup>, Kari Norman<sup>10</sup>, Dominique Gravel<sup>11</sup>, and Timothée Poisot<sup>12</sup>

- (<sup>1</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3
- (<sup>2</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3
- (<sup>3</sup>) School of Biosciences, The University of Sheffield  
Western Bank, Sheffield, United Kingdom S10 2TN
- (<sup>4</sup>) Department of Ecology and Evolutionary Biology, University of Toronto  
25 Willcocks St, Toronto, ON, Canada M5S 3B2
- (<sup>5</sup>) Department of Biology, McGill University  
1205 avenue du Docteur-Penfield, Montréal, QC, Canada H3A 1B1
- (<sup>6</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3
- (<sup>7</sup>) Département de biologie, Université de Sherbrooke  
2500 boulevard de l'Université, Sherbrooke, Qc, Canada J1K 2R1
- (<sup>8</sup>) School of Biosciences, The University of Sheffield  
Western Bank, Sheffield, United Kingdom S10 2TN
- (<sup>9</sup>) School of Biosciences, The University of Sheffield  
Western Bank, Sheffield, United Kingdom S10 2TN
- (<sup>10</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3
- (<sup>11</sup>) Département de biologie, Université de Sherbrooke  
2500 boulevard de l'Université, Sherbrooke, Qc, Canada J1K 2R1
- (<sup>12</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3

This article was submitted in EcoEvoRxiv and can be found at <https://doi.org/10.32942/X28G8Z>.

The main contributions of Francis Banville for this articles are presented.

FB developed and performed the analyses and wrote the first and final versions of the manuscript. All authors participated in the conceptualisation and revision of the manuscript. DG and TP provided guidance on the analyses and interpretation of the results. All authors read and approved the manuscript.

**RÉSUMÉ.** Représenter les interactions entre espèces de manière probabiliste (quelle est leur probabilité d'occurrence?) plutôt que de manière déterministe (se produisent-elles?) nous informe de leur variabilité et des incertitudes dans nos connaissances des interactions. Les sources d'incertitude capturées par les probabilités d'interaction dépendent de la méthode utilisée pour les évaluer : incertitude des modèles prédictifs, évaluation subjective des experts ou mesure empirique de la variabilité spatio-temporelle des interactions. Cependant, des directives pour l'estimation et la documentation des données d'interactions probabilistes sont manquantes. Cela est préoccupant car notre compréhension et l'analyse des probabilités d'interaction dépendent de leur définition parfois ambiguë et de leurs sources d'incertitude. Nous passons en revue comment les interactions probabilistes sont définies à différentes échelles spatiales, des interactions locales aux réseaux régionaux (*metawebs*), en mettant l'accent sur les interactions hôte-parasites et trophiques (préation et herbivorie). Ces définitions reposent sur la distinction entre la réalisation d'une interaction à un moment et un lieu spécifiques (local) et sa faisabilité biologique ou écologique (régional). En utilisant des données d'interactions hôtes-parasites en Europe, nous illustrons comment ces deux représentations de réseau diffèrent dans leurs propriétés statistiques, en particulier : comment les réseaux locaux et les *metawebs* diffèrent dans leurs mises à l'échelle spatiale et temporelle des probabilités d'interaction, mais pas dans leur mise à l'échelle taxonomique. Nous présentons deux approches pour inférer des interactions binaires à partir d'interactions probabilistes qui tiennent compte de ces différences et montrons que des biais systématiques apparaissent lorsque l'on infère directement des réseaux locaux à partir d'un *metaweb*. Nos résultats soulignent l'importance de descriptions plus rigoureuses des réseaux d'interactions probabilistes entre espèces qui spécifient leur type d'interaction (locale ou régionale), leurs variables conditionnelles et leurs sources d'incertitude.

**Mots clés :** réseaux écologiques, réseaux trophiques, metaweb, réseaux probabilistes, échelle spatiale, interactions entre espèces, échelle temporelle

**ABSTRACT.** Representing species interactions probabilistically (how likely are they to occur?) as opposed to deterministically (are they occurring?) conveys uncertainties in our knowledge of interactions and information on their variability. The sources of uncertainty captured by interaction probabilities depend on the method used to evaluate them: uncertainty of predictive models, subjective assessment of experts, or empirical measurement of interaction spatiotemporal variability. However, guidelines for the estimation and documentation of probabilistic interaction data are lacking. This is concerning because our understanding and analysis of interaction probabilities depend on their sometimes elusive definition and uncertainty sources. We review how probabilistic interactions are defined at different spatial scales, from local interactions to regional networks (metawebs), with a strong emphasis on host-parasite and trophic (predatory and herbivory) interactions. These definitions are based on the distinction between the realization of an interaction at a specific time and space (local) and its biological or ecological feasibility (regional). Using host-parasite interactions in Europe, we illustrate how these two network representations differ in their statistical properties, specifically: how local networks and metawebs differ in their spatial and temporal scaling of probabilistic interactions, but not in their taxonomic scaling. We present two approaches to inferring binary interactions from probabilistic ones that account for these differences and show that systematic biases arise when directly inferring local networks from metawebs. Our results underscore the importance of more rigorous descriptions of probabilistic species interaction networks that specify their type of interaction (local or regional), conditional variables and uncertainty sources.

**Keywords:** ecological networks, food webs, metaweb, probabilistic networks, spatial scale, species interactions, temporal scale

## 1.1. Introduction

### 1.1.1. Species interactions are variable and uncertain

As we try to navigate global biodiversity change, filling in knowledge gaps about biodiversity becomes instrumental to monitoring and mitigating those changes (Hortal et al., 2015, Gonzalez and Londoño, 2022, Abrego et al., 2021). However, cataloging species, populations and, in particular, ecological interactions (e.g., predation, parasitism, and pollination) is a substantial challenge (Polis, 1991, Pascual and Dunne, 2006). There are methodological and biological constraints that hinder our ability to document species interactions, inevitably leading to uncertainty in our knowledge of interactions. For example, the spatial and temporal uncoupling of species (e.g., nocturnal and diurnal species coexisting in the same space with different daily activity timings, Jordano, 1987) and the large number of rare and cryptic interactions in a community, contribute to these knowledge gaps by making it more difficult to observe interactions (Jordano, 2016).

Several conditions must be satisfied for an interaction to be observed locally. First, both species must have overlapping geographic ranges, i.e. they must co-occur within the region of interest (Morales-Castilla et al., 2015, Cazelles et al., 2016). Second, they must have some probability of meeting within a defined time frame (Poisot et al., 2015). Probabilities of interspecific encounters are typically low, especially for rare species with low abundances (Vázquez et al., 2007, Canard et al., 2012, Canard et al., 2014). The probability that species meet also depends on their biology, such as their phenology (Olesen et al., 2010, Singer and McBride, 2012) and discoverability (Broom and Ruxton, 2005). Finally, when species do come into contact, an interaction occurs only if their traits, such as their phenotypes (Bolnick et al., 2011, Stouffer, Rezende, and Amaral, 2011, Gravel et al., 2013) and behavior (Pulliam, 1974, Choh et al., 2012), are locally compatible in that specific environment (Poisot et al., 2015). Because these conditions are not consistently met locally, there will inevitably be instances where interactions will be observed and others where they will not.

Documenting the location and timing of interactions is difficult when accounting for the spatiotemporal variability of ecological interactions (Poisot et al., 2012, Poisot et al., 2015). Knowing the biological capacity of two species to interact directly (via e.g., trophic interactions) is necessary but not sufficient for inferring their interaction at a specific time and space. Environmental factors, such as temperature (Angilletta et al., 2004), drought (Woodward et al., 2012), climate change (Gilman et al., 2010, Woodward et al., 2010, Araujo et al., 2011), habitat characteristics (e.g., presence of refuges where prey can hide from predators, Grabowski, 2004), and land use change (Tylianakis et al., 2007), contribute to this spatiotemporal variability by impacting species abundance and traits. Interactions may also be influenced by a third species (e.g., a more profitable prey species, Golubski and Abrams, 2011, Sanders and van Veen, 2012). Even under favorable circumstances, there remains a possibility that the interaction does not occur locally, either due to the intricate nature of the system or simply by chance. If it does occur, it might go undetected, particularly if it happens infrequently. In this context, it is unsurprising that our knowledge of ecological interactions remains limited (Hortal et al., 2015) despite extensive biodiversity data collection (Schmeller et al., 2015).

We distinguish the variability of interactions from their uncertainty. Interaction variability is defined as the changes in the occurrence or strength of interactions along spatial, temporal, or environmental axes (Poisot et al., 2015). It is a property of interactions that should be quantified if we aim for a comprehensive understanding of ecological networks. Stochasticity is the inherent randomness or unpredictability of interactions that lead to this variability. Conversely, uncertainty is defined as a lack of knowledge about the occurrence of interactions. When using statistical models to infer interactions, uncertainty sources include input data, parameter, and model structure uncertainties (Simmonds et al., 2024). Input data uncertainty arises from our inability to empirically observe all interactions and from measurement errors in environmental and biological variables used for inference. Parameter uncertainty represents a plausible range of values for a parameter whose exact value is unknown. For example, we may calculate a range of plausible values for interaction variability

(e.g., there could be a 50% certainty that an interaction occurs 50% of the time). Model structure uncertainty recognizes that different statistical models may adequately predict interactions. In contrast to variability, uncertainty can be reduced by sampling additional data (except for model uncertainty, which will persist regardless of sampling effort). Simmonds et al., 2024 underscores the importance of quantifying and reporting these diverse sources of uncertainty, alongside ensuring their appropriate propagation to model output (such as predicted interactions) and higher-level measures (such as network structure). While recognizing that these definitions may not be universally accepted, clarifying the distinction between variability and uncertainty enables us to better comprehend the sources of our knowledge gaps about ecological interactions.

### **1.1.2. Species interactions as probabilistic objects**

The recognition of the intrinsic variability and uncertainty of species interactions has led ecologists to expand their representation of ecological networks to include a probabilistic view of interactions (Poisot, Cirtwill, et al., 2016, Dallas et al., 2017, Fu et al., 2021). This allows filling in the Eltonian shortfall (i.e., the gap between our current knowledge and a comprehensive understanding of interactions, Hortal et al., 2015) by modeling the probability of occurrence of interactions (e.g., Gravel et al., 2019), which can be an important tool for directing efforts and taking action (Carlson et al., 2021), especially in places where access and resources for research are scarce. A probability is a measure of how likely a specific outcome is, based on both the uncertainty and variability of interactions. Interaction probabilities may be uncertain when there is a distribution of plausible probability values. The probabilistic representation of interactions has been applied to direct interactions, which are conceptually and mathematically analogous regardless of their biological type (e.g., predation and pollination). This is in contrast with indirect interactions (e.g., interspecific competition), which arise from distinct ecological processes and are often not directly observable (Kéfi et al., 2015, Kéfi et al., 2016). By accounting for the uncertainty and variability of direct interactions, networks of probabilistic interactions (which differ from

*probabilistic networks* describing the uncertainty and variability of the *whole* network) may provide a more realistic portrait of species interactions.

Probabilistic interactions differ from binary interactions. Networks of probabilistic interactions, within a Bayesian perspective, express our degree of belief (or confidence) regarding the occurrence or observation of interactions. In a frequentist approach, they represent the expected relative frequencies of interactions over many repeated trials or sampling events. In contrast, interactions are simply regarded as either occurring or not in networks of deterministic binary interactions. Based on the scale at which they are estimated, interaction probabilities may reflect our level of confidence in whether interactions will be observed, realized locally, or biologically feasible. Our level of confidence should be more definitive (approaching either 0 or 1) as we extend our sampling to a broader area and over a longer duration, thereby diminishing the uncertainty of our knowledge of interactions (but not necessarily the estimation of their variability). In the broadest sense, binary interactions are also a type of probabilistic interaction, in which the numerical value of an interaction is restrained to 0 (non-occurring) or 1 (occurring). In networks of probabilistic interactions, only forbidden interactions (i.e., interactions prohibited by biological traits or species absence, Jordano et al., 2003, Olesen et al., 2010) have a probability value of zero, provided that intraspecific trait variability is considered (Gonzalez-Varo and Traveset, 2016). Understanding the nuances between probabilistic and binary interactions is essential for accurately modeling and interpreting ecological networks.

The application and development of computational methods in network ecology, often based on a probabilistic representation of interactions, can alleviate (and guide) the sampling efforts required to document species interactions (Strydom, Catchen, et al., 2021). For example, statistical models can be used to estimate the uncertainty of pairwise interactions (Cirtwill et al., 2019) and the probability of missing (false negatives) and spurious (false positives) interactions (Guimerà and Sales-Pardo, 2009), helping us identify places where sampling is most needed to reduce this uncertainty. Statistical models can also predict networks without prior knowledge of pairwise interactions. They may do so using body size

(Petchey et al., 2008, Gravel et al., 2013, Caron et al., 2024), phylogeny (Elmasri et al., 2020, Strydom et al., 2022), or a combination of niche and neutral processes (Bartomeus et al., 2016, Pomeranz et al., 2019) for inference. Before being used to test ecological hypotheses, predicted networks must be validated against empirical data (Brimacombe et al., 2024), which could be sampled strategically to optimize the validation process. Topological null models, which generate networks of probabilistic interactions by preserving chosen characteristics of the adjacency matrix of binary interactions while intentionally omitting others (Bascompte et al., 2003, Fortuna and Bascompte, 2006), are examples of common probabilistic interaction models. Null models can produce underlying distributions of network measures for null hypothesis significance testing. However, how the uncertainty of pairwise interactions propagates to network structure (i.e., community-level properties driving the functioning, dynamics, and resilience of ecosystems, McCann, 2007, McCann, 2011, Proulx et al., 2005, Rooney and McCann, 2012) remains to be elucidated. Many measures have been developed to describe the structure (Poisot, Cirtwill, et al., 2016) and diversity (Ohlmann et al., 2019, Godsoe et al., 2022) of probabilistic interaction networks. These models and measures support the use of probabilistic interactions for the study of a wide range of ecological questions, from making better predictions of species distribution (Cazelles et al., 2016) to forecasting the impact of climate change on ecological networks (Gilman et al., 2010).

### **1.1.3. We lack a clear understanding of probabilistic species interactions**

We still lack a precise definition of probabilistic interactions, which makes the estimation and use of these data more difficult. In this manuscript, we aim to take a step back by outlining different ways in which probabilistic interactions are defined and used in network ecology. We distinguish two broad categories of probabilistic interaction networks that necessitate distinct approaches: local networks describing probabilities of realized interactions, and regional networks (metawebs) describing probabilities of potential interactions. We highlight the distinctions in the ecological meaning of these two representations of interactions

and examine their properties and relationships (particularly with space, time, and between each other).

The lack of clear guidelines on the use of probabilistic interaction data is worrisome, as it affects both data producers and re-users who generate and manipulate these numbers. This is concerning because sampling strategies and decisions regarding network construction can affect our understanding of network properties (Brimacombe et al., 2023). There is currently no reporting standard that could guide the documentation of all types of probabilistic interactions (Salim et al., 2022 discuss data standards for deterministic mutualistic networks). Clear reporting standards for probabilistic interactions would support more adequate manipulation and integration of interaction data from different sources and guard against possible misinterpretations arising from ambiguous definitions of probabilistic interaction networks. Data documentation should outline the nature (i.e., local or regional) and type (e.g., predatory or pollination) of interactions, provide information regarding the taxonomic level, identities, and characteristics (e.g., life stages) of the individuals involved in an interaction, present the mathematical formulation of probabilities, including clearly identified conditional variables (e.g., spatial and temporal scales), and describe the methods and contexts (e.g., location, time, environmental conditions) in which interactions were estimated. Inadequately documented probabilistic interaction data should be used with caution when analyzing ecological networks. These broad principles remain relevant and applicable across different types of direct interactions. In the following sections, we discuss the definitions, conditions, and estimation of probabilistic interactions as we scale up from pairwise interactions to interactions within local and regional networks.

## 1.2. Pairwise interactions: the building blocks of ecological networks

### 1.2.1. What are probabilistic interactions?

Consider a scenario where an avian predator has just established itself in a northern habitat home to a small rodent. Suppose their interaction has not been previously observed, either because these species have never co-occurred before or because previous sampling failed to detect an interaction despite their co-occurrence. What is the probability that the rodent is part of the diet of the avian predator, or put differently, what is the probability that they interact? Answering this question requires some clarification, as there are multiple ways to interpret and calculate interaction probabilities. We could calculate the probability that the traits of these species match, i.e. that the avian predator possesses the biological attributes to capture and consume the rodent. We could also calculate the probability that their traits support an interaction under the typical environmental conditions of the new habitat. For example, because avian predators hunt by sight, predation could be possible in the absence of snow but highly improbable when snow is present, as rodents may use it as a shelter to hide from predators. Finally, we could calculate the probability that the avian predator will consume the rodent at *that* particular location, for which the spatial and temporal boundaries need to be specified. The estimation of the probability of interaction between these two species, whether through predictive models or informative prior probabilities, hinges on our understanding of these probabilities and the specific ecological processes we aim to capture.

An important aspect to consider when estimating or using interaction probabilities is knowing if they describe the probability of potential or realized interactions, as these two types of interactions have distinct meanings and sources of uncertainty and variability. A potential (regional) interaction is defined as the biological or ecological capacity of two taxa to interact (i.e., the probability that they interact if they were to encounter each other, given sufficient time and appropriate environmental conditions) whereas a realized (local) interaction is the occurrence or observation of this interaction in a well-defined space and time

(i.e., the probability that they interact locally). For two co-occurring taxa and over enough time, the probability of local interaction tends toward the probability of regional (potential) interaction. A longer duration increases the probability that species will eventually encounter each other and that local environmental conditions supporting an interaction will occur, provided that species have the biological capacity to interact. Recognizing the distinction between probabilistic regional and local interactions is crucial for accurately interpreting interaction probabilities in ecological networks.

We use the terms *metaweb* (Dunne, 2006) to designate regional networks of potential interactions and *local networks* (Poisot et al., 2012) for those of realized interactions. Metawebs are the network analogs of the species pool, where local networks originate from a subset of both species (nodes) and interactions (edges) of the regional metaweb (Saravia et al., 2022). Without clear documentation, it can be challenging to know if published probabilistic interaction networks describe local or regional interactions. When probabilistic local interactions are used and interpreted incorrectly as regional interactions (and conversely), this may generate misleading findings during data analysis. A better understanding of probabilistic local and regional interaction networks would facilitate a more adequate use of interaction data (e.g., when studying network-area relationships in local networks and metawebs) and prevent misinterpretations of the biological meaning of probabilistic interactions.

### **1.2.2. What is the outcome of probabilistic interactions?**

**1.2.2.1. The outcome of probabilistic interactions is usually binary.** Local networks and metawebs, like any type of network, are made of nodes and edges that may be represented at different levels of organization. The basic units of ecological networks are individuals that interact with each other (e.g., by predation in food webs, Elton, 2001), forming individual-based networks (Melián et al., 2011). The aggregation of these individuals into more or less homogeneous groups (e.g., populations, species, families, feeding guilds) allows us to represent nodes at broader taxonomic scales, which affects our interpretation of the properties of these systems (Guimarães, 2020, Hemprich-Bennett et al., 2021).

Ecologists have traditionally represented interactions (edges) as binary objects that were considered realized after observing at least one individual from group  $i$  interact with at least another individual from group  $j$ . In an adjacency matrix  $B$  of binary interactions, the presence or absence of an interaction  $B_{i,j}$  between two taxa can be viewed as the result of a Bernoulli trial  $B_{i,j} \sim \text{Bernoulli}(P(B_{i,j}))$ , with  $P(B_{i,j})$  being the probability of interaction. This interaction probability characterizes our limited ecological knowledge and/or the intrinsic spatiotemporal variability of the interaction. It may be estimated through predictive models (e.g., those based on biological traits and species abundances) or expert (prior) knowledge about the interaction. In networks of probabilistic interactions, the edge values  $P(B_{i,j})$  are probabilistic events whose only two possible outcomes are the presence ( $B_{i,j} = 1$ ) or absence ( $B_{i,j} = 0$ ) of an interaction between each pair of nodes. Depending on the type of probabilistic interaction network (local network or metaweb), the mathematical formulation and interpretation of stochastic parameters like  $P(B_{i,j})$  can be linked to environmental and biological factors such as species abundances, species traits, area, and time, for example using logistic regression with continuous explanatory variables. This allows us to model the probability that at least two individuals interact under these conditions.

The variability of an interaction determines the fraction of networks in which it occurs. This fraction can be predicted by using a Binomial distribution, assuming a constant interaction probability and independence between interactions in different networks (trials). When considering uncertainties around the estimation of  $P(B_{i,j})$ , a Beta distribution may be used to represent the relative likelihood of different probability values. For example, when calculating the probability of interaction between two taxa based on their local abundances, any uncertainty in their abundances would introduce uncertainty in the interaction probability at the local scale. If we take into account the uncertainty of the interaction probability, a Beta-Binomial distribution can be used to predict the number of networks in which the interaction occurs. Empirically observing an interaction between two taxa at a given location and time provides important information that can be used to update previous estimates of  $P(B_{i,j})$ , informing us of the conditions that enabled them to interact locally. By sampling

binary interactions in different contexts, we can thus estimate their local variability more precisely.

**1.2.2.2. The outcome of probabilistic interactions may also be quantitative.** Even though binary interaction networks constitute a highly valuable source of ecological information (Pascual and Dunne, 2006), they overlook interaction strengths. Represented in a quantitative adjacency matrix  $W$ , interaction strengths better describe the energy flows, demographic impacts or frequencies of interactions between nodes (Berlow et al., 2004, Borrett and Scharler, 2019), with  $W_{i,j}$  being a natural  $\mathbb{N}$  or real  $\mathbb{R}$  number depending on the measure. For example, they may represent local interaction rates (e.g., the flower-visiting rates of pollinators in a mutualistic network, Herrera, 1989). Relative frequencies of interactions may be used as a measure of both the strength and probability of local interactions that are biologically feasible. When interaction strengths characterize predation pressure on prey, they can serve as parameters in a Lotka-Volterra model (e.g., Emmerson and Raffaelli, 2004). The extra amount of ecological information in quantitative networks typically comes at a cost of greater sampling effort and data volume (Strydom, Catchen, et al., 2021), especially when using predictive models that quantify the uncertainty and variability of quantitative interactions (Berlow et al., 2004). However, if two taxa are repeatedly found together without interacting, there may be more uncertainty about their capacity to interact than their interaction strength (which would assuredly be close to 0).

Like binary interaction networks, the uncertainty and variability of interaction strengths can be represented probabilistically. Interaction strengths can follow many probability distributions depending on the measure. For instance, they can follow a Poisson distribution  $W_{i,j} \sim \text{Poisson}(\lambda_{i,j}t_0)$  when predicting the number of interactions between individuals during a time interval  $t_0$ , with  $\lambda_{i,j}$  being the expected rate at which individuals of taxa  $i$  and  $j$  interact (e.g., the expected number of prey  $j$  consumed by all predators  $i$ ). The Poisson distribution can also be 0-inflated when taking into account non-interacting taxa (e.g., Boulangeat et al., 2012 employ a 0-inflated model to analyze species abundance following the modeling of species presence and absence), which constitute the majority of taxa pairs in

most local networks (Jordano, 2016). Regardless of the measure, estimating the uncertainty of quantitative interactions enables us to consider a range of possible values of interaction strength.

Because of the methodological difficulties typically encountered when building deterministic quantitative networks, binary interaction networks, which are usually easier to sample (Jordano, 2016) and predict (Strydom, Catchen, et al., 2021), have been more frequently studied and used. Mathematical models such as Ecopath (Plagányi and Butterworth, 2004) partially mitigate these difficulties, but the number of biological parameters required to make predictions hinders their application in many systems. Moreover, most published probabilistic interaction networks (e.g., Strydom et al., 2022) and methods (e.g., Poisot, Cirtwill, et al., 2016) involve probabilistic interactions whose outcome is binary. This underlines the need for better guidelines on the interpretation and manipulation of probabilistic interactions with binary outcomes first, to ensure the appropriate use of these networks and methods. For these reasons, the primary focus of the remainder of this manuscript is on the interpretation of interaction probabilities that determine the presence or absence of interactions, in both local networks and metawebs.

## 1.3. Local networks: communities interacting in space and time

### 1.3.1. What are local probabilistic interactions?

Local networks of probabilistic interactions describe how likely taxa are to interact in a local context. Local interactions are contingent upon the environmental conditions experienced by the community and the matching of taxa's local biological traits. In local networks, edges commonly represent our degree of belief that two taxa interact in nature, but can also represent the probability of empirically *observing* this interaction (Catchen et al., 2023). Local interactions may thus arise from both the ecological (realized interactions) and sampling (observed interactions) processes taking place locally.

Local networks are delineated within a particular location and time. We define space as the collection of geographic coordinates  $(x, y, z)$ , with  $(x, y)$  representing longitude and latitude coordinates, and  $z$  denoting either altitudes or depths. These point coordinates delineate the spatial boundaries of the system, which may be portrayed as a polyhedron. Ecological interactions may vary along latitudinal and altitudinal gradients, as evidenced by changes in hummingbird-plant interactions (Weinstein and Graham, 2017a, Weinstein and Graham, 2017b) and mosquito biting rates (e.g., Kulkarni et al., 2006) at different elevations. On the other hand, time is defined as the specific time period within which interactions were either observed or predicted. Even though space and time are continuous variables that should yield probability *densities* of interactions (i.e., relative likelihoods of interactions occurring at infinitesimal locations and instants in time), these definitions enable them to be conceptualized as distinct patches and time segments. Treating space and time as discrete dimensions aligns with the common sampling methods of ecological networks and provides probabilities of interactions, which can be obtained by integrating probability densities over space and time. We can quantify both an area  $A_0$  and a duration  $t_0$  with these definitions. By studying probabilistic local interaction networks, we may thus conduct spatiotemporal analyses of local interactions (Box 1), enhancing our understanding of interactions occurring in distinct environmental contexts.

### 1.3.2. What are local probabilistic interactions conditioned on?

**1.3.2.1. Local interactions may be conditioned on co-occurrence.** The probability that two taxa  $i$  and  $j$  interact in a local network  $L_{x,y,z,t}$  (spatial and temporal subscripts hereafter replaced by the shorter subscript  $k$  for clarity) can be conditioned on many environmental and biological factors. In addition to network area (or volume)  $A_0$  and duration  $t_0$ , they may be conditioned on taxa co-occurrence  $X_{i,j,k}$ , which is usually Boolean, describing if the geographic distributions of both taxa overlap within the study area. As illustrated in Box 1, co-occurrence may be modeled probabilistically, in which case it may conform to a Bernoulli distribution  $X_{i,j,k} \sim \text{Bernoulli}(P(X_{i,j,k}))$ . The probability of co-occurrence can be calculated

using the individual (marginal) occurrence probabilities  $P(X_{i,k})$  and  $P(X_{j,k})$ . Given that taxa occurrences are not independent of each other, the probability of co-occurrence can be calculated by multiplying the probability of occurrence of one taxon by the probability of occurrence of the other given that the first one is present:

$$P(X_{i,j,k}) = P(X_{i,k}, X_{j,k}) = P(X_{i,k}|X_{j,k}) \times P(X_{j,k}) \quad (1.3.1)$$

The probability of co-occurrence  $P(X_{i,j,k})$  can be estimated through the application of joint species distribution models (e.g., Pollock et al., 2014), potentially taking into account biotic interactions (Staniczenko et al., 2017). Given that the probability that two non-co-occurring taxa interact locally is zero (i.e.,  $P(L_{i,j,k}|X_{i,j,k} = 0) = 0$ ), the probability of local interaction can be obtained by multiplying the probability of interaction given co-occurrence with the probability of co-occurrence:

$$P(L_{i,j,k}) = P(L_{i,j,k}|X_{i,j,k}) \times P(X_{i,j,k}). \quad (1.3.2)$$

Knowing that two taxa co-occur improves our estimation of the probability that they interact locally by mitigating a potential source of uncertainty.

**1.3.2.2. Local interactions may be conditioned on different environmental and biological factors.** Local interactions may also be conditioned on local environmental factors such as temperature (Angilletta et al., 2004), precipitation (Woodward et al., 2012), habitat structure (Klecka and Boukal, 2014), and presence of other taxa in the network (Pilosof et al., 2017, Kéfi et al., 2012). We use the variable  $E_k$  to describe the local environmental context in which interaction probabilities were estimated. For example, in a mesocosm experiment estimating interaction probabilities between predators and prey with and without refuges,  $E_k$  would represent the presence or absence of these refuges. Like co-occurrence,  $E_k$  can also be modeled probabilistically when the variability or uncertainty of environmental factors is considered.  $E_k$  represents all environmental variables that were

taken into consideration when measuring interaction probabilities; it is therefore a subset of all environmental factors acting on ecological interactions.

Other important factors that can impact interaction probabilities at the local scale are taxa local abundances  $N_{i,k}$  and  $N_{j,k}$ , which affect encounter probabilities (Canard et al., 2012), and local traits  $T_{i,k}$  and  $T_{j,k}$  (e.g., movement rates, Beardsell et al., 2021, Cherif et al., 2024), which may also impact encounter probabilities as well as the ability of individuals to interact after encountering each other (Caron et al., 2024, Poisot et al., 2015). Local interaction probabilities may also be conditioned on higher-level properties of the community (i.e., the emerging structure of ecological networks), which we denote by  $f(L_k)$ . Many topological null models (i.e., statistical models that randomize interactions by retaining certain properties of the network while excluding others) provide interaction probabilities from selected measures of network structure, such as connectance (Fortuna and Bascompte, 2006) and the degree distribution (Bascompte et al., 2003). Biological factors, whether at the scale of individual taxa pairs or the community, may thus impact how we estimate and define interaction probabilities.

**1.3.2.3. Local interactions may be conditioned on biological feasibility.** Local interactions must be biologically feasible before occurring at a specific time and space. A local probability of interaction  $P(L_{i,j,k})$  can be expressed as the product of the probability of local interaction given that the two taxa can potentially interact  $P(L_{i,j,k}|M_{i,j} = 1)$ , which we sometimes denote as  $P(L_{i,j,k}|M_{i,j})$  for the sake of simplicity, with their probability of regional interaction  $P(M_{i,j})$ :

$$P(L_{i,j,k}) = P(L_{i,j,k}|M_{i,j} = 1) \times P(M_{i,j} = 1). \quad (1.3.3)$$

Low values of  $P(L_{i,j,k}|M_{i,j})$  indicate that feasible interactions rarely occur locally, intermediate values around 50% suggest considerable spatiotemporal variability, while high values indicate that regional interactions are nearly always realized locally. The local probability of interaction between a given pair of taxa is thus always equal to or below their probability

of regional interaction. Taking into account biological feasibility in our estimation of local interaction probabilities leverages information from the metaweb to better predict the local occurrence of interactions (Strydom, Catchen, et al., 2021, Dansereau et al., 2024).

**1.3.2.4. Conditional variables must be explicitly stated.** The probability that two taxa  $i$  and  $j$  interact in a local network  $L_k$  can thus be conditioned on their co-occurrence  $X_{i,j,k}$  (or more explicitly on their occurrences  $X_{i,k}$  and  $X_{j,k}$ ), local abundances  $N_{i,k}$  and  $N_{j,k}$ , local traits  $T_{i,k}$  and  $T_{j,k}$ , local environmental conditions  $E_k$ , network area (or volume)  $A_0$ , time interval  $t_0$ , network properties  $f(L_k)$ , and biological feasibility  $M_{i,j}$ . When these conditions are not explicit, it may be because they have been marginalized over, which would be reflected in the overall uncertainty of the interaction. The local probability of interaction is described by the following expression when all of these conditional variables are included:

$$P(L_{i,j,k}|X_{i,k}, X_{j,k}, N_{i,k}, N_{j,k}, T_{i,k}, T_{j,k}, E_k, A_0, t_0, f(L_k), M_{i,j}). \quad (1.3.4)$$

These conditional variables do not all need to be considered at all times. The representation of the local context in which probabilities are estimated and the variables that should be taken into consideration depend on the study system, the objectives of the study, and the resources available to the researchers. For example, Gravel et al., 2019 analyzed local European host-parasite networks of willow-galling sawflies and their natural enemies, all referenced in space and time, to infer probabilities of local interactions between co-occurring species. This was achieved by including temperature and precipitation as conditional variables in their models. In Box 2, we reuse these data to show the extent of variation among these local networks. We do so by measuring their dissimilarity with the regional network (metaweb aggregating all local interactions), both in terms of species composition and interactions. We built local probabilistic networks following Eq 1.3.3, showing that insufficient local variation (high probability of local interaction among potentially interacting species) results in an overestimation in both the number of interactions and connectance (i.e., the proportion of all of the non-forbidden links that are realized). This analysis was conducted

for illustrative purposes, and other conditional variables could have been used to make these comparisons.

When accounted for, conditional variables should be clearly described in the documentation of the data (Brimacombe et al., 2023), preferentially in mathematical terms to avoid any confusion in their interpretation and to limit manipulation errors during their re-use. For instance, ecologists should be explicit about their consideration ( $P(L_{i,j,k}|X_{i,j,k})$ ) or not ( $P(L_{i,j,k})$ ) of co-occurrence in their estimation of local interaction probabilities, as this can change our interpretation of the data and understanding of potential uncertainty sources. Reporting the scale and level of aggregation of the data enables us to more accurately study the underlying ecological processes (Clark et al., 2011) and manipulate or propagate uncertainty to different aggregation levels (Simmonds et al., 2024). In Table 1, we present examples of studies that used different expressions of probabilistic interactions with different conditional variables. We included in this table the probability of empirically observing an interaction that is realized locally  $P(O_{i,j,k}|L_{i,j,k})$  to underscore the distinction between local observations and actual realizations of interactions.

Expression	Type	Outcome	Common models	Reference
$P(L_{i,j,k} X_{i,k}, X_{j,k}, \dots)$	local	realization of the interaction given taxa co-occurrence	species distribution models	Gravel et al., 2019, Dansereau et al., 2024, Boxes 1 and 5
$P(L_{i,j,k} N_{i,k}, N_{j,k}, \dots)$	local	realization of the interaction given taxa abundances	neutral models	Canard et al., 2014
$P(L_{i,j,k} T_{i,k}, T_{j,k}, \dots)$	local	realization of the interaction given local traits	trait matching models	Caron et al., 2024, Box 4
$P(L_{i,j,k} E_k, \dots)$	local	realization of the interaction given local environmental conditions	environmental-based models	Gravel et al., 2019 (temperature and precipitation)
$P(L_{i,j,k} A_0, \dots)$	local	realization of the interaction in a given area or volume	spatial models	Galiana et al., 2018 *, Box 3
$P(L_{i,j,k} t_0, \dots)$	local	realization of the interaction during a given time period	temporal models	Weinstein and Graham, 2017a, Boxes 1 and 3
$P(L_{i,j,k} f(L_k), \dots)$	local	realization of the interaction given network structure	topological models	Fortuna and Bascompte, 2006 (connectance), Stock et al., 2017
$P(L_{i,j,k} M_{i,j}, \dots)$	local	realization of the interaction given that it is biologically feasible	spatiotemporal models	Dansereau et al., 2024, Boxes 2, 3, and 5
$P(O_{i,j,k} L_{i,j,k}, \dots)$	local	observation of the interaction given that it is realized locally	sampling models	Catchen et al., 2023
$P(M_{i,j} T_i, T_j)$	regional	biological feasibility of the interaction given regional traits (non-forbiddenness)	trait matching models	Strydom et al., 2022, Box 4
$P(M_{i,j}^* T_i, T_j, E)$	regional	ecological feasibility of the interaction given regional traits and environmental conditions	trait matching and environmental-based models	this study

**Table 1. Mathematical expressions of probabilistic interactions.** The probability of interaction between two taxa  $i$  and  $j$  is interpreted differently in a local network  $L_k$  of realized interactions, a local network  $O_k$  of observed interactions, a metaweb  $M$  of potential interactions (representing the *biological* feasibility of interactions), and a metaweb  $M^*$  of potential interactions (representing the *ecological* feasibility of interactions). Each expression emphasizes a different conditional variable, the ellipsis serving as a placeholder for other variables not explicitly stated in the expression. The outcome of each of these probabilistic events, along with common models used for estimation, is presented alongside examples of studies that employed them (with specific variables indicated in parentheses, when applicable). The study marked with an asterisk has been conducted on binary interaction networks. The boxes in our study that discuss these expressions are also specified.

### 1.3.3. How are local probabilistic interactions estimated?

Various statistical models can be used to estimate local interaction probabilities, some of which are presented in Table 1. These models can be based on multiple conditional variables. Although these variables correspond to distinct ecological inquiries or mechanisms related to ecological interactions, they may covary with each other, such as the possible dependence of  $X_{i,j,k}$  and  $E_k$  on spatial and temporal scales. When estimating interaction probabilities using e.g. a generalized linear model with multiple explanatory variables that might not all be independent, it may become important to address collinearity. In such cases, it may be necessary to use variable selection techniques before fitting the model to data to mitigate this issue. Other challenges and opportunities associated with predictive models of species interactions are reviewed in Strydom, Catchen, et al., 2021.

When using multiple competing models to estimate local interaction probabilities, rather than selecting a single model that best fits the data, model averaging may enhance our estimations. Model weights represent the probability that each model is the most suitable for explaining the data, and may be measured using Akaike weights (Burnham and Anderson, 2004, Wagenmakers and Farrell, 2004). For instance, given two competing models  $mod_1$  and  $mod_2$  with respective probabilities (or weights)  $P(mod_1)$  and  $P(mod_2)$ , the average probability of interaction  $P(L_{i,j,k})$  can be calculated as follows:

$$P(L_{i,j,k}) = P(L_{i,j,k}|mod_1) \times P(mod_1) + P(L_{i,j,k}|mod_2) \times P(mod_2). \quad (1.3.5)$$

Model averaging takes into account the uncertainty of model structure in our estimation of local interaction probabilities. Regardless of the model used for prediction, it is crucial to quantify and disclose all sources of uncertainty to understand better the validity and limitations of our predictions (Simmonds et al., 2024).

### Box 1: A spatiotemporally explicit model of interactions



Ecologists may resort to predictive models to reconstruct local networks across time and space. We introduce and develop a simple generative Bayesian model for probabilistic local interactions, which explicitly accounts for their spatiotemporal variability. Our model is not designed for regional interactions, which do not vary spatially nor temporally. Rather, it could prove valuable for generating new data on local interactions across time and space, following parameter inference.

As indicated by Eq 1.3.2, the probability that two taxa  $i$  and  $j$  interact locally can be obtained by multiplying their probability of interaction given co-occurrence with their probability of co-occurrence. The probability of interaction given co-occurrence can be made temporally explicit by modeling it as a Poisson process with rate parameter  $\lambda_k$ . This parameter represents the local expected frequency of interaction between co-occurring taxa. The probability that two co-occurring taxa interact at least once during a time interval  $t_0$  can be given by:

$$P(L_{i,j,k}|X_{i,j,k}) = 1 - e^{-\lambda_k t_0}, \quad (1.3.6)$$

which tends toward 1 as  $t_0 \rightarrow \infty$  if  $\lambda_k > 0$ . In other words, two co-occurring taxa with a nonzero rate of interaction will inevitably interact at least once in a sufficiently long time interval.

The occurrence of an interaction between  $i$  and  $j$  may be the result of a Bernoulli trial with probability  $P(L_{i,j,k})$ . A Bayesian model can be built using the preceding equations to generate new interaction data, following the inference of the  $\lambda_k$  and  $P(X_{i,j,k})$  parameters.

$$L_{i,j,k} \sim \text{Bernoulli}(P(L_{i,j,k})) \quad (1.3.7)$$

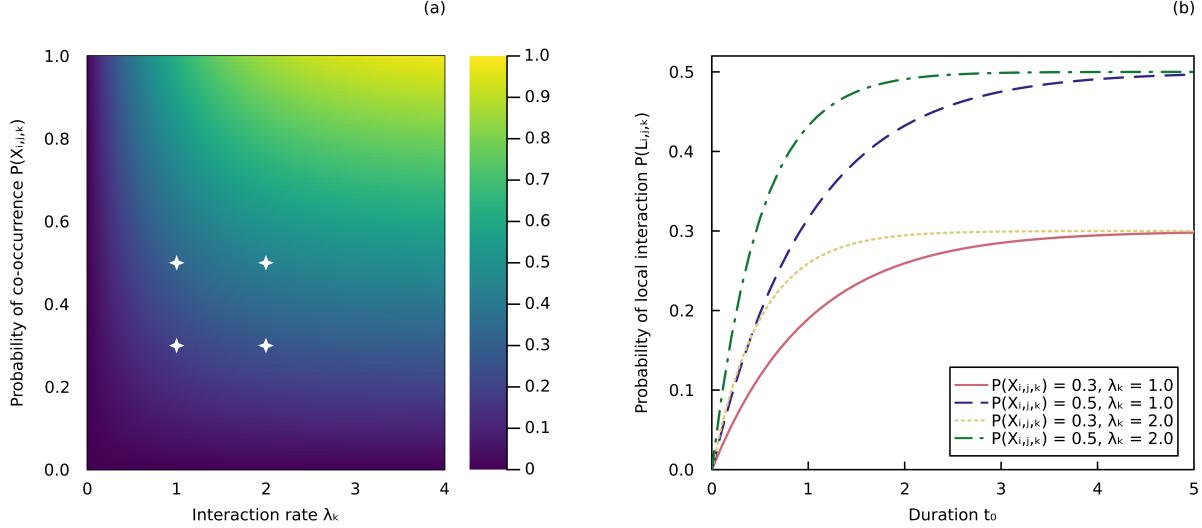
$$P(L_{i,j,k}) = P(X_{i,j,k})(1 - e^{-\lambda_k t_0}) \quad (1.3.8)$$

$$P(X_{i,j,k}) \sim \text{Beta}(2,2) \quad (1.3.9)$$

$$\lambda_k \sim \text{Exponential}(2) \quad (1.3.10)$$

In Fig 1, we show the variation in the probability of interaction under different parameter values. In the right panel, we notice that the probability of interaction always converges toward the probability of co-occurrence  $P(X_{i,j,k})$ , for all positive values of the interaction rate.

This model can be customized in different ways, such as linking both parameters to specific environmental or biological variables. For instance, the probability of co-occurrence could be modeled as a function of climatic variables, while the interaction rate parameter could be modeled based on taxa abundances.



**Fig. 1. Parameters of the spatiotemporally explicit model of interactions.** (a) Probability of local interaction given by the process model (Eq 1.3.8) under different values of  $\lambda_k$  (interaction rate) and  $P(X_{i,j,k})$  (probability of co-occurrence), with  $t_0 = 1$  (duration). The probability of local interaction represents the probability that the two taxa will interact at least once within the given time interval. Parameters  $t_0$  and  $\lambda_k$  have complementary units (e.g.,  $t_0$  in months and  $\lambda_k$  in number of interactions per month). The parameter values used in the right panel are denoted by the white stars. (b) Scaling of the probability of interaction with the duration parameter  $t_0$ , for different values of  $\lambda_k$  and  $P(X_{i,j,k})$ .

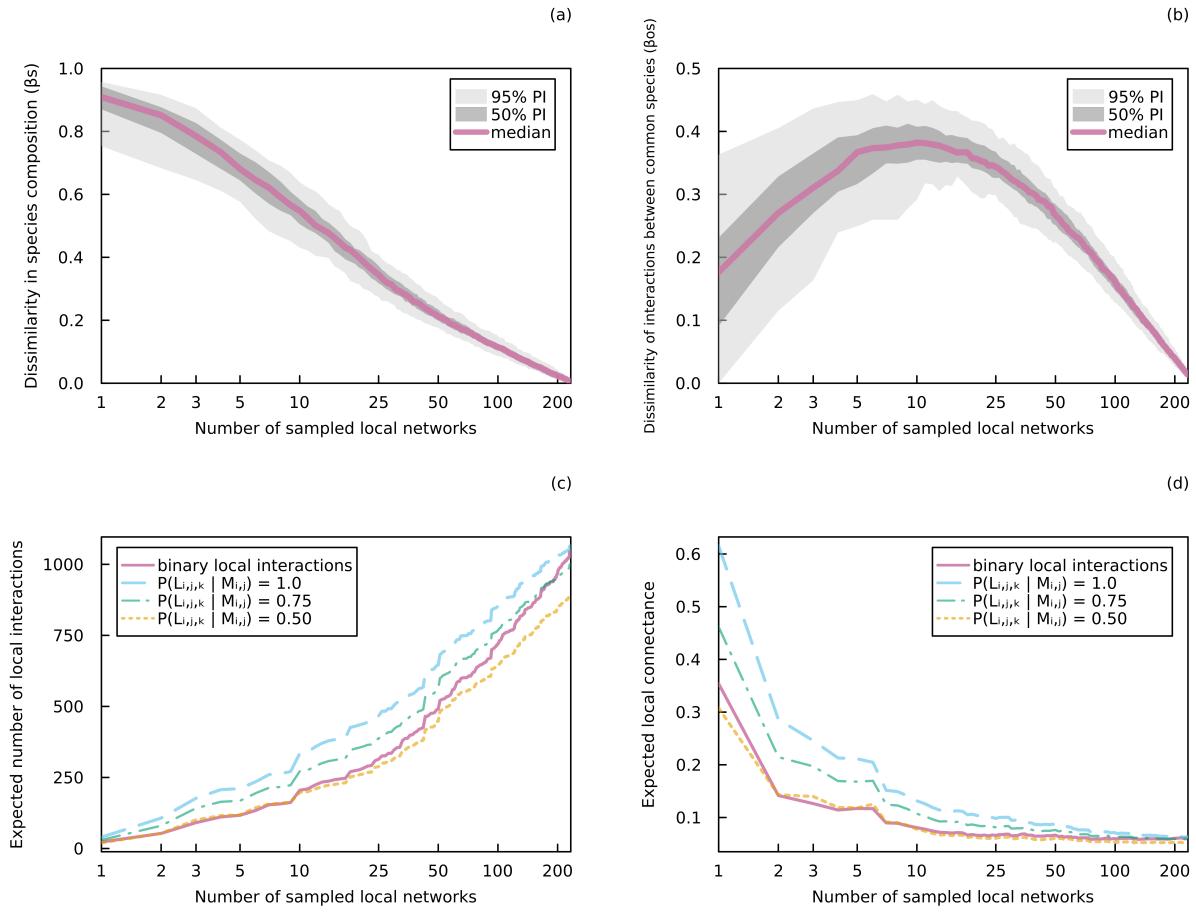
## Box 2: Dissimilarity of local host-parasite networks ➤➤

We present a way to assess local network variability and dissimilarity regarding species composition and interactions. We do so by comparing local tripartite host-parasite networks to the metaweb using data from Kopelke et al., 2017. This collection of networks consists of interactions between willows, willow-galling sawflies, and their natural enemies sampled across Europe. All data manipulation and methods are described in Appendix 1. All code and data to reproduce these analyses are available on Zenodo (<https://doi.org/10.5281/zenodo.12802326>).

In Fig 2a-b, we show how the dissimilarity between the metaweb of binary interactions and aggregated local networks changes with the number of sampled local networks. We compared the metaweb and the aggregated local networks using the dissimilarity in species composition ( $\beta_S$ , Fig 2a) and the dissimilarity of interactions between common species ( $\beta_{OS}$ , Fig 2b) indices (Poisot et al., 2012). Expectedly, local networks are highly dissimilar from the metaweb in terms of species composition, especially when only a limited number of sites have been sampled. This is because few species from the metaweb (species pool) occur locally. Moreover, we observe a peak in the dissimilarity of interactions between common species at intermediate sampling levels. This suggests that species are collected faster than their interactions. With a limited number of sampled local networks, few regional interactions are observed locally. Adding more sites brings new species, but not always their interactions. Quadratic relationships of network properties with sampling effort were also observed by McLeod et al., 2021.

Next, we investigate how the number of local interactions and connectance scale with the number of sampled (aggregated) local networks of probabilistic or binary interactions (Fig 2c-d). By comparing the scaling relationships observed in local networks of binary and probabilistic interactions, we observe that high values of  $P(L_{i,j,k}|M_{i,j})$  lead to systematic overestimations in the number of interactions and connectance, especially when  $P(L_{i,j,k}|M_{i,j}) = 1$  (i.e., when local and regional probabilities of interactions are equivalent). This suggests that high values of  $P(L_{i,j,k}|M_{i,j})$  do not

adequately capture the variability of local interactions. However, these biases tend to diminish as the number of sampled networks increases, indicating that most interactions are eventually captured when  $P(L_{i,j,k}|M_{i,j})$  is high. In contrast, low values of  $P(L_{i,j,k}|M_{i,j})$  lead to missing interactions, resulting in an underestimation of the number of interactions and connectance. These results underscore the importance of using the appropriate level of variability when estimating local interaction probabilities.



**Fig. 2. Network accumulation curves.** (a) Dissimilarity in species composition and (b) dissimilarity of interactions between common species between aggregated local networks and the metaweb of binary host-parasite interactions. In both panels, the colored line represents the median dissimilarity across simulations and the grey areas cover the 50% and 95% percentile intervals. (c) Scaling of the number of interactions and (d) scaling of connectance with the number of sampled (aggregated) binary and probabilistic local interaction networks. For a better comparison with binary interactions, local networks of probabilistic interactions were derived from a metaweb of probabilistic interactions with a false positive and false negative rate of zero. A specific value of  $P(L_{i,j,k} | M_{i,j})$  (the local probability of interaction among potentially interacting species) was used for all non-aggregated local networks within a particular curve. Aggregated local networks were obtained by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions (with the value of  $P(L_{i,j,k} | M_{i,j})$  increasing in aggregated local networks of probabilistic interactions).

## 1.4. Metawebs: regional catalogs of interactions

### 1.4.1. What are regional probabilistic interactions?

Metawebs (Dunne, 2006) are networks of potential interactions over broad spatial, temporal, and taxonomic scales (e.g., food webs at the continental scale). They correspond to the temporal and spatial asymptotes of local interactions (Box 1). Potential interactions describe the biological capacity of taxa to interact under optimal or feasible environmental conditions given enough time, which is typically assessed at the regional scale. Metawebs of probabilistic interactions are particularly useful in situations where there is uncertainty in the ability of taxa to interact (Strydom et al., 2023). They may also be used as informative priors of local interactions. Therefore, building a metaweb of probabilistic interactions may be an important first step before predicting networks at finer scales.

In contrast to local networks, where interaction probabilities arise from the variability of interactions and the lack of information on the conditions, interaction probabilities in metawebs solely result from a lack of knowledge. This uncertainty arises due to insufficient interaction data, especially for taxa that have not yet been observed to co-occur, and uncertainties in trait-matching models. As data accumulates, interactions in metawebs should tend towards binarity, either taking a value of 1 (observing an interaction at least once) or approaching 0 (repeatedly failing to observe an interaction between co-occurring taxa). Confidently observing an interaction once confirms its biological feasibility, but failing to observe it (even on multiple occasions) does not ensure that it is non-feasible (e.g., due to false negatives, Catchen et al., 2023). While local interaction probabilities are irreducible because of local variability, the uncertainty of regional interactions reduces to 0 with the addition of information. Moreover, although *neutrally* forbidden interactions (i.e., forbidden interactions between rare species, Canard et al., 2012) have low probability values in local networks, they would have a probability of 1 in the metaweb (this is because the species' traits could support an interaction if they were to encounter each other at high enough abundances). Likewise, non-co-occurring taxa may have a non-zero probability of interaction in

the metaweb. Regional interaction probabilities are thus fundamentally different from local interaction probabilities, both in terms of uncertainty sources and probability values.

The extent of sampling effort influences our evaluation of probabilities of regional interactions, as sampling over a larger area or for a longer duration enables us to capture a greater number of interactions (Box 1, McLeod et al., 2021). However, in contrast with local networks of probabilistic interactions, regional interactions are not evaluated for any particular local context (they are rather a collection of local contexts), which impacts how they scale with space and time (notably through the extent of the region covered and sampling duration). In Box 3, we discuss the differences in spatial and temporal scaling of regional interactions compared to local interactions. We do so using the host-parasite networks of Kopelke et al., 2017 as an illustration of spatial scaling (Box 3). Understanding the effect of spatial and temporal scales (including sampling effort) on local and regional interaction probabilities is important for effectively propagating uncertainty across scales and highlighting the fundamental differences between these two types of networks.

### 1.4.2. What are regional probabilistic interactions conditioned on?

**1.4.2.1. Regional interactions describing biological feasibility are conditioned on traits.** Potential interactions describe what we refer to as the *biological* feasibility of interactions, which is based solely on the regional traits distributions  $T_i$  and  $T_j$  of taxa  $i$  and  $j$ , respectively. We define regional traits distributions as the range of phenotypes that a taxon can express across various environments. Local traits  $T_{i,k}$  and  $T_{j,k}$ , which vary spatially and temporally because of phenotypic plasticity and local environmental variability (Berg and Ellers, 2010), are a subset of regional traits. A probability of potential interaction in a metaweb  $M$  describing the biological feasibility of interactions may be expressed as:

$$P(M_{i,j}|T_i, T_j), \quad (1.4.1)$$

which, in contrast with local networks, is not conditioned on any spatial, temporal, co-occurrence or environmental variables (Table 1). Because phylogenetically close species often share similar traits, we should expect that closely related species will have similar interacting partners. We can thus use phylogeny to predict species traits and infer regional interactions (Strydom et al., 2022, Eklöf and Stouffer, 2016, Stouffer et al., 2012). The taxonomic level at which interactions are evaluated also influences the distribution of regional traits. However, as explained in Box 4, there is no fundamental difference in the taxonomic scaling of regional and local interactions (i.e., how interaction probabilities change with taxonomic level) because they both depend on trait aggregation.

The biological feasibility of interactions expresses our degree of belief that there exists at least one combination of phenotypes that could support an interaction if they were to encounter each other, assuming they had enough time to interact. Evaluating this probability is conducted without incorporating the environmental conditions under which they encounter each other into the model. It is the complement of the probability  $P(F_{i,j}|T_i, T_j)$  of forbidden interactions (i.e., the probability that their traits do not support an interaction), which is based uniquely on biological traits:

$$P(M_{i,j}|T_i, T_j) = 1 - P(F_{i,j}|T_i, T_j). \quad (1.4.2)$$

For example, let  $i$  be a western diamondback rattlesnake (*Crotalus atrox*) and  $j$ , a wood lemming (*Myopus schisticolor*). These two taxa never co-occur, the rattlesnake being adapted to warm regions of North America (Castoe et al., 2007) and the lemming, to northern habitats of Eurasia (Fedorov et al., 2008). As we lack direct observations of an interaction between these two species, we have to rely on expert knowledge or trait-matching models to estimate their probability of potential interaction. To accurately estimate this probability using trait-matching models, it is crucial to ensure that the set of traits considered reflects the overall traits distributions of both taxa. We could for instance consider their average body mass and the average phylogenetic distance of lemmings to rattlesnakes' prey. Doing

so, we might find a high probability of potential interaction based on these traits. This example illustrates how regional interactions describing biological feasibility may be estimated solely based on traits, without taking into account environmental conditions (which could be important to consider when e.g. an interaction is forbidden at all temperature values).

**1.4.2.2. Regional interactions describing ecological feasibility are conditioned on traits and environmental conditions.** The biological feasibility of interactions should not be confused with what we refer to as the *ecological* feasibility of interactions. A probability of potential interaction in a metaweb  $M^*$  describing the ecological feasibility of interactions may be expressed as:

$$P(M_{i,j}^*|T_i, T_j, E), \quad (1.4.3)$$

where  $E$  is the environmental conditions under which potential interactions are evaluated (Table 1). Unlike  $E_k$ , these environmental conditions do not represent conditions occurring at specific locations. Ecological feasibility represents the probability that two taxa interact if they were to encounter each other under given environmental conditions, assuming they had enough time to interact. Incorporating environmental conditions into a trait-matching model may be important when there is high covariance between the environment and traits. For instance, in our example involving rattlesnakes and lemmings, the probability of potential interaction between these two species may be low in most environmental conditions. Western diamondback rattlesnakes may be unactive under low temperatures (Kissner et al., 1997), whereas wood lemmings may have low tolerance to high temperatures (Kausrud et al., 2008). The probability that an interaction is ecologically feasible is always lower than the probability that it is biologically feasible, even across all environmental conditions:

$$\int_E P(M_{i,j}^*|T_i, T_j, E) dE \leq P(M_{i,j}|T_i, T_j). \quad (1.4.4)$$

This is because the biological feasibility of an interaction is a prerequisite for its ecological feasibility. In other words, biological feasibility is necessary but not sufficient for an interaction to be ecologically feasible. Our discussion of metawebs focuses on the biological feasibility of interactions since most methods developed for inferring probabilities of regional interactions do not explicitly take into account environmental conditions (e.g., Strydom et al., 2022).

### 1.4.3. How are regional probabilistic interactions estimated?

Starting from a selected set of taxa, which are usually distributed within a broad region of interest, metawebs can be built using different data sources, including literature review (e.g., Maiorano et al., 2020), aggregated interaction data (e.g., Gravel et al., 2019, Saravia et al., 2022), trait-matching models (e.g., Shaw et al., 2024, Strydom et al., 2022), and expert knowledge, which is not a trivial challenge. Every pair of taxa that have confidently been observed to interact at least once can be given a probability of 1 (i.e.,  $P(M_{i,j}) = 1$ ) since we know that they *can* interact. This differs from local networks of probabilistic interactions, where interaction events may remain stochastic (i.e.,  $P(L_{i,j,k}) < 1$ ) even after empirically observing interactions due to their spatiotemporal variability. Interactions that were never observed typically have low probability values in local networks and vary from low to high values in metawebs, contingent upon taxa traits distributions (reaching 0 for forbidden links). The aggregation of model predictions and data from different sources thus tends to raise the number of potential interactions in metawebs.

When using local interaction data to estimate probabilities of regional interactions, repeatedly failing to observe an interaction between two co-occurring taxa should decrease the probability that the interaction is biologically feasible. Using Bayes' theorem, the probability that the interaction is biologically feasible given that it was never observed locally,  $P(M_{i,j} = 1 | O_{i,j,k} = 0)$ , may be calculated as follows:

$$P(M_{i,j} = 1 | O_{i,j,k} = 0) = \frac{P(O_{i,j,k} = 0 | M_{i,j} = 1) \times P(M_{i,j} = 1)}{P(O_{i,j,k} = 0)}. \quad (1.4.5)$$

The reduction in the probability of regional interaction after considering that it was never observed locally (i.e.,  $P(M_{i,j} = 1 | O_{i,j,k} = 0) < P(M_{i,j} = 1)$ ) occurs because  $P(O_{i,j,k} = 0 | M_{i,j} = 1)$  must be lower than  $P(O_{i,j,k} = 0)$ , i.e. there is a higher chance of observing an interaction when it is biologically feasible.

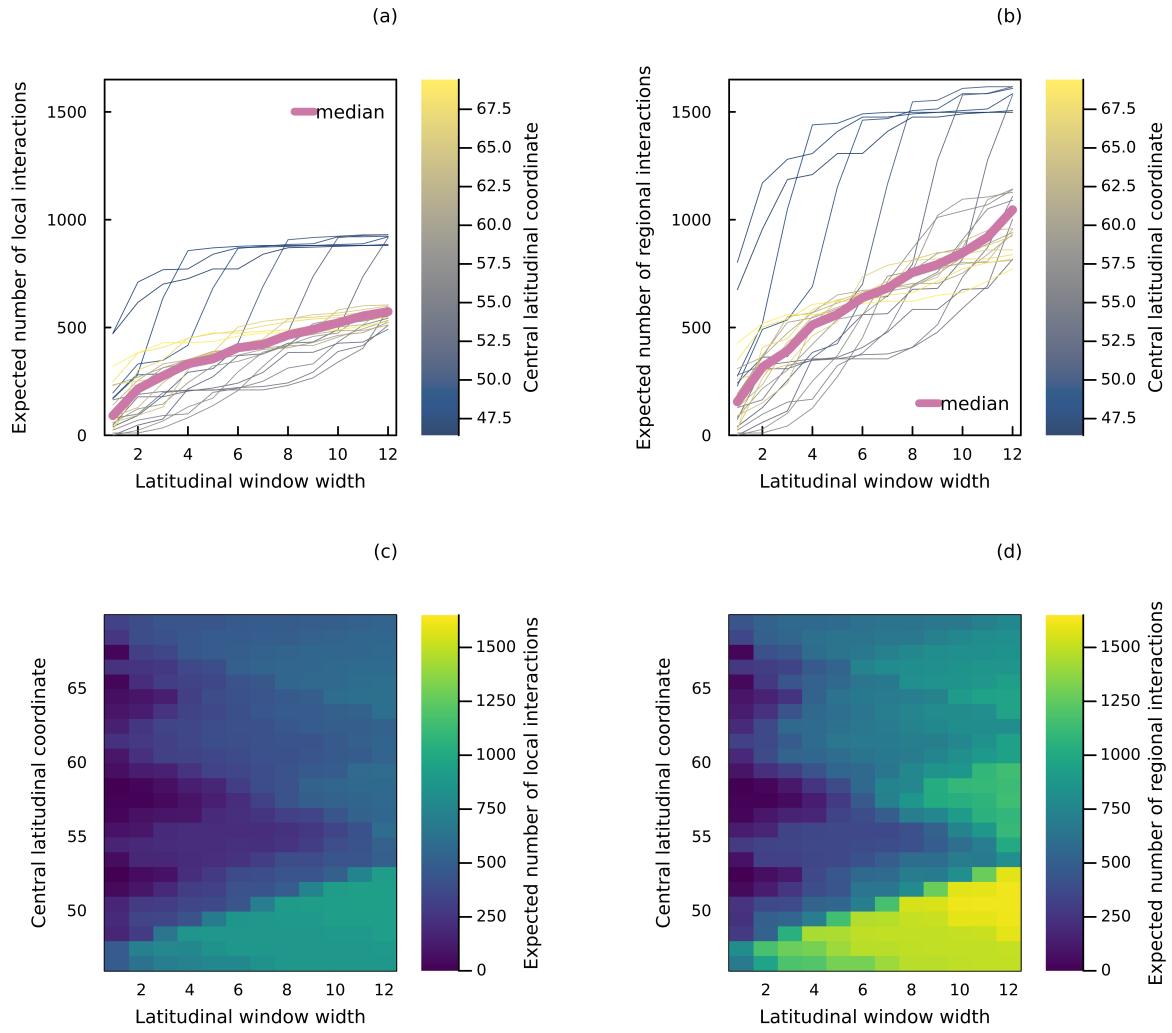
Observations of interactions may be false positives because of observation errors due to taxonomic misidentifications and ecological misinterpretations, such as those involving phylogenetically close species or cryptic species and interactions (Pringle and Hutchinson, 2020). Likewise, forbidden interactions may be false negatives, e.g. if they have been evaluated based on unrepresentative or incomplete traits distributions. Employing Bayesian models proves valuable when estimating interaction probabilities in metawebs (e.g., Bartomeus et al., 2016, Cirtwill et al., 2019). This improvement is achieved by updating prior information regarding the feasibility of interactions (e.g., experts' prior assessments of interaction probabilities) with empirical data on interactions and traits. By improving our estimation of potential interaction probabilities, we may build more reliable metawebs that adequately reflect our uncertainty on the biological feasibility of interactions.

### Box 3: Spatial and temporal scaling of interactions

Local networks and metawebs have distinct relationships with space (area or volume) and time (sampling effort or duration). Local probabilities of interaction scale both spatially and temporally, because local interactions have more opportunities to be realized in larger areas and longer durations. In a larger sampling area and duration, we increase the likelihood of sampling favorable conditions for interactions to occur. If a local network of probabilistic interactions  $L_1$  with an area  $A_1$  is compared to a larger network  $L_0$  with an area  $A_0$ , and  $A_1$  is entirely nested within  $A_0$ , interaction probabilities should be lower in the smaller network, i.e.  $P(L_{i,j,1} | A_1 < A_0) \leq P(L_{i,j,0} | A_0)$ .

However, if  $A_1$  and  $A_0$  are disjoint, interaction probabilities could be higher in the smaller area, contingent upon local environmental and biological conditions. In contrast, regional probabilities of interaction do not scale with space and time. The probability of two taxa potentially interacting should be the same in all metawebs in which they are present regardless of scale, provided that the data and methods used for estimation are consistent. This is because they depend solely on the biological capacity of two taxa to interact, regardless of co-occurrence and local environmental conditions. However, probabilities of regional interactions may change, tending to become more definitive, with increased sampling effort.

In Fig 3, we show how the expected *number* of local host-parasite interactions scales with the spatial boundary of the network (represented by an expanding latitudinal window) in comparison with regional interactions. We do so using the host-parasite networks of Kopelke et al., 2017. The increase in the number of regional interactions is due to the inclusion of more species in a larger area. To ensure a conservative comparison between aggregated local and regional networks, we employed equal interaction probabilities (i.e., using  $P(L_{i,j,k}|M_{i,j}) = 1$ ) in both types of network. This means that local interaction probabilities could not increase further when aggregating them. Despite this, we notice that the total number of regional interactions scales more rapidly than local interactions. This is because numerous regional interactions involve species that never co-occur, and as a result, these interactions are not captured in local networks. All data manipulation and methods are described in Appendix 1.



**Fig. 3. Spatial scaling of interactions.** Expected number of host-parasite interactions in a network aggregating all (a) local and (b) regional probabilistic interactions within a latitudinal window of a given width. Every dashed curve corresponds to a different window centered at a given latitude (color bar), with the pink solid line representing the median number of interactions across windows. Heatmaps of the expected number of (c) local and (d) regional interactions found in windows of specified width and position (central latitude). Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Local probabilistic interactions were derived from regional probabilistic interactions by setting the value of  $P(L_{i,j,k}|M_{i,j})$  (the local probability of interaction among potentially interacting species) to 1. Aggregated local networks were obtained by aggregating both the species and interactions found within a particular latitudinal window, with the values of  $P(L_{i,j,k}|M_{i,j})$  remaining at their maximum value of 1.

#### Box 4: Taxonomic scaling of interactions ➤

Given that our interpretation of the properties of ecological networks depends on their taxonomic level (Melián et al., 2011), investigating the taxonomic scaling of interactions (i.e., how interaction probabilities change with taxonomic level) is important. There are no inherent differences between the taxonomic scaling of local and regional interactions. The taxonomic level of interactions impacts the definition of nodes. Local and regional interaction probabilities are not directly conditioned on taxonomic scale. However, some conditional variables (e.g., trait distribution) may covary with taxonomic scale. In such cases, local and regional interaction probabilities would change taxonomically following the scaling of these variables.

In both types of interactions, transitioning to a broader level of organization (e.g., from a species-level network  $S$  to a genus-level network  $G$ ) can be done using interaction probabilities from finer scales. For example, in a network with  $n_1$  species of genus  $g_1$  and  $n_2$  species of genus  $g_2$ , one can calculate the probability that at least one species from genus  $g_1$  interacts with at least one species from genus  $g_2$  (i.e., the probability that the genus-level interaction occurs) as follows:

$$P(G_{g_1,g_2}) = 1 - \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (1 - P(S_{g_1,i,g_2,j})), \quad (1.4.6)$$

where  $g_{1,i}$  and  $g_{2,j}$  are the species of the corresponding genus and assuming independence between species-level interactions. In contrast, a different approach is necessary when transitioning from a broader to a finer level of organization. This is because the knowledge of an interaction between two genera does not guarantee that all possible pairwise species combinations will also interact. One possible method is to build a finer-scale network by generating probabilities of interaction through random sampling from a beta distribution, parameterized by the broader-scale network.

Fundamentally, the taxonomic scaling of interactions involves aggregating interactions between individuals into larger groups. Interaction probabilities at broader taxonomic scales should thus conform to probabilities of interactions between individuals. For example, Canard et al., 2012 built a species-based network using simulated individual-based networks. In local individual-based food webs, the probability that two individuals interact reflects our degree of belief that one individual will consume the other. Likewise, in local species-based food webs, the probability that two species interact represents our degree of belief that *at least* one individual from the predator species will consume at least another individual from the prey species. In that regard, taxonomic scaling is analogous to the spatial and temporal scaling of interactions, as they all represent different ways to aggregate individuals into broader groups (either spatially, temporally, or taxonomically).

#### Box 5: Sampling for binary interaction networks ➤

Local networks of binary interactions may be predicted by performing independent Bernoulli trials for each probabilistic interaction. This is particularly useful when analyzing the structure of probabilistic interaction networks in the absence of specific analytical formulas (Poisot, Cirtwill, et al., 2016), even though it may introduce biases in our estimations when connectance is low (Poisot and Gravel, 2014, Chagnon, 2015). There are at least two techniques to sampling binary interaction networks across space, each predicting a binary interaction network for each location  $k$  within a given region. The first technique involves performing a single Bernoulli trial for each pair of taxa based on their regional probability of interaction:

$$M_{i,j} \sim \text{Bernoulli}(P(M_{i,j})). \quad (1.4.7)$$

In employing this technique, we predict a single metaweb of binary interactions for each simulation. Every pair of taxa predicted to interact in this metaweb will be treated as interacting in all localized networks where they co-occur, i.e.  $L_{i,j,k} = M_{i,j}$  when  $X_{i,j,k} = 1$ . This will result in local pairwise interactions without spatial variation. The second technique is to independently sample each local network of probabilistic interactions:

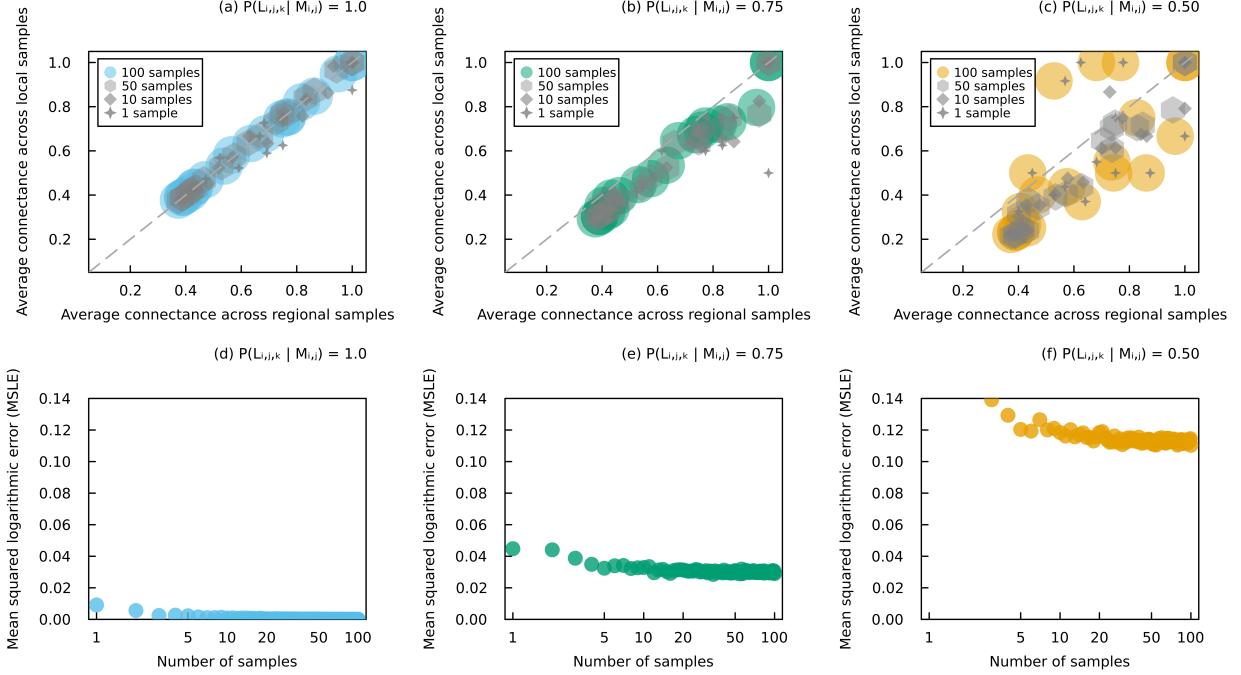
$$L_{i,j,k} \sim \text{Bernoulli}(P(L_{i,j,k})). \quad (1.4.8)$$

This can be achieved by first generating distinct probabilistic interaction networks for each location. Because binary interactions are sampled independently for each location, this second technique captures network structure across space and time more effectively. When sampling binary interactions from local interaction probabilities, it is crucial to sample at the same spatial scale for which probabilities were estimated to prevent systematic biases in predictions.

In Fig 4, we compare the average connectance of binary interaction networks resulting from these two sampling techniques. We sampled regional and local interactions from our host-parasite networks of probabilistic interactions (Kopelke et al., 2017), generating a number of binary interaction network realizations for each site in the dataset. These two sampling techniques yield different outcomes, particularly for intermediate values of  $P(L_{i,j,k}|M_{i,j})$  of 0.50, which represent instances where regional interactions do not consistently manifest locally (i.e., with the largest local variability). As anticipated, we observe that sampling binary interactions from the metaweb tends to overestimate connectance on average compared to sampling them from local networks (Fig 4). We also observe an increase in the variability of connectance when employing

a single simulation (Fig 4a-c, cross markers), which is a more tangible representation of the process leading to the realization of local interactions in nature. All data manipulation and methods are described in Appendix 1.

Both sampling techniques assume independence between interactions, which might not hold true in reality. Covariation among interactions could exist even if we do not explicitly condition interactions on others. For example, an interaction between two taxa could be more probable when another interaction occurs. The consequences of this assumption of independence on the prediction of network structure have yet to be empirically examined. Sampling whole networks (or graphs) instead of pairwise interactions may eliminate the need for this assumption of independence (Battiston et al., 2020).



**Fig. 4. Connectance of sampled binary interaction networks.** (a-c) Average connectance of binary interaction networks obtained from the two sampling techniques for 20 randomly selected host-parasite networks. Cross markers represent the connectance of a single sample for each network, diamond markers the average connectance across 10 samples, hexagon markers the average connectance across 50 samples, and the colored circles the average connectance across 100 samples (marker size proportional to the number of samples). (d-f) Reduction in the mean squared logarithmic error between the average connectance of binary interaction networks (all 233 host-parasite networks) obtained from these two sampling techniques as the number of samples increases. The local probability of interaction between potentially interacting species was set to three different values: (a,d)  $P(L_{i,j,k}|M_{i,j}) = 1.0$ , (b,e)  $P(L_{i,j,k}|M_{i,j}) = 0.75$ , and (c,f)  $P(L_{i,j,k}|M_{i,j}) = 0.50$ . Probabilities of regional interactions were obtained with a false positive rate of 5% and a false negative rate of 10%. Regional samples were obtained by randomly sampling binary interactions from the probabilistic interaction metaweb, and then propagating this result to all local networks that include the species potentially engaged in the interactions. Local samples were obtained by independently sampling binary interactions for each local network of probabilistic interactions.

## 1.5. Future perspectives

In this contribution, we underline the importance of network documentation for adequately interpreting and manipulating probabilistic interaction data. The mathematical representation of probabilities and their statistical properties depend on the type of interactions (local or regional) and the conditions under which these interactions were evaluated. We show that local networks and metawebs of probabilistic interactions differ in their relationship to spatial and temporal scales (Box 3), with regional interactions remaining consistent across scales. In contrast with metawebs, local interactions are measured in a specific context (e.g., in a given area, time, and biological and environmental conditions) and depend on taxa co-occurrence. These differences bring to light the need to use probabilistic data with caution, for instance when generating network realizations of binary interactions across space (Box 5). Clear documentation describing the type of interaction and the variables used in their estimation are required to ensure adequate data manipulation. Sound data practices and foundations for probabilistic thinking in network ecology facilitate reliable assessments of the spatiotemporal variability and uncertainty of biotic interactions. Here we identify key research priorities for improving our understanding of probabilistic local and regional interactions.

### 1.5.1. Predicting local networks from metawebs

Metawebs are a valuable source of ecological information for predicting local networks across time and space. Local networks of binary interactions can be reconstructed by selecting a subset of taxa and interactions from the metaweb (Dunne, 2006). Determining the list of taxa to select can be achieved empirically (e.g., observed occurrence data for a site) or numerically (e.g., species distribution models). As species composition is arguably easier to sample and predict than pairwise interactions, the primary challenge lies in deciding which interactions to select from the metaweb. Inferring the structure of local networks from the metaweb before predicting local pairwise interactions could hold promise (Strydom, Catchen,

et al., 2021), considering that the structure of local networks is constrained by the metaweb (Saravia et al., 2022).

While predicting local binary interactions from a metaweb is not be a simple task, inferring local networks of probabilistic interactions from a metaweb comes with its own set of challenges. For example, Dansereau et al., 2024 inferred spatially-explicit food webs from a metaweb of probabilistic trophic interactions between Canadian mammals. Their predicted localized food webs are downscaled versions of the metaweb (i.e., localized metawebs with the same interaction probabilities as those in the regional metaweb). To infer local networks as defined in this manuscript (i.e., describing local realizations of interactions), local interaction probabilities must be smaller than regional interaction probabilities. Inferring local networks from a metaweb by maintaining identical interaction probability values introduces systematic biases into the predictions, as discussed in Box 2 (unless networks are seen as downscaled metawebs).

As suggested by McLeod et al., 2021, metawebs establish an upper limit for local interactions (similarly for metawebs of probabilistic interactions, Strydom et al., 2023). In other words, the probability that two taxa interact at a specific location and time is consistently lower or equal to the probability of their regional interaction, regardless of the conditional variables considered:

$$P(L_{i,j,k} | \dots) \leq P(M_{i,j} | T_i, T_j). \quad (1.5.1)$$

Moreover, the probability that two taxa possess the biological capacity to interact must be higher than the probability of them interacting at any location and time because they may never co-occur or encounter locally. Specifically, the cumulative probability of local interaction across all spatial, temporal, and environmental conditions must be less than the probability of regional interaction, i.e.

$$\int_{E_k} \int_{A_0} \int_{t_0} P(L_{i,j,k}|E_k, A_0, t_0) dt_0 dA_0 dE_k \leq P(M_{i,j}|T_i, T_j). \quad (1.5.2)$$

Estimating more precisely the probability  $P(L_{i,j,k}|M_{i,j})$  that two taxa interact locally if they can potentially interact allows for improved predictions of local networks from the metaweb of probabilistic interactions. This task is challenging due to the variability of this probability across space and time, as well as its variability across pairwise interactions within a network. Using simple models of  $P(L_{i,j,k}|M_{i,j})$ , as shown in Appendix 1, represents an initial step toward the overarching objective of reconstructing local networks from metawebs.

### 1.5.2. Quantifying and reducing interaction uncertainty

While sampling biological communities decreases the uncertainty of interactions by accumulating evidence for their feasibility and local realization, there is a limit to how much we can reduce uncertainty. In metawebs, probabilities reflect our limited knowledge of interactions, which is expected to improve with a larger volume of data. Regional interactions should become more definitive (with probabilities approaching 0 or 1) as we investigate various conditions, including different combinations of species traits.

In comparison, local interaction probabilities represent both our knowledge uncertainty and their spatiotemporal variability. Owing to environmental heterogeneity, there will invariably be instances in which an interaction occurs and others in which it does not, across different times and locations, irrespective of the extent to which we can improve our knowledge of its biological feasibility and the local conditions that facilitate its occurrence. When local networks describe probabilities of observing interactions rather than their realization, we must also consider observation uncertainty (sampling error) as an additional source of uncertainty. Quantifying and partitioning this uncertainty will enable us to make more accurate predictions about ecological interactions at various spatial and temporal scales, and to identify priority sampling locations to reduce this uncertainty. This will prove to be of

vital importance as our time to understand nature runs out, especially at locations where the impacts of climate change and habitat loss hit harder.

### 1.5.3. Relaxing the independence assumption

Estimating local interaction probabilities independently for each taxa pair and assembling them into a network of probabilistic interactions comes with limitations. Predicting local networks of binary interactions based on these interaction probabilities assumes independence among interactions, a condition seldom respected in practice (Golubski and Abrams, 2011). Relaxing this assumption is the next logical step in the stochastic representation of interactions.

A more accurate representation of the uncertainty and variability of ecological networks involves creating *probabilistic networks* ( $P(L_k)$  and  $P(M)$ ), rather than networks of *probabilistic interactions* ( $P(L_{i,j,k})$  and  $P(M_{i,j})$ ). Probabilistic networks describe the probability that a particular network of binary (or quantitative) interactions (its whole adjacency matrix) is realized. For example, Young et al., 2021 used a Bayesian approach to estimate the probability of different plant-pollinator network structures derived from imperfect observational data. A probability distribution of ecological networks may also be derived using the principle of maximum entropy given structural constrained (e.g., Cimini et al., 2019, Park and Newman, 2004).

Regardless of the method used, generating probabilistic local networks could lead to more accurate predictions of local networks of binary interactions by bypassing the independence assumption. Probabilistic networks could serve as an alternative to null hypothesis significance testing when comparing the structure of a local network to some random expectations or, as done in Pellissier et al., 2018 and Box 2, to the metaweb. These random expectations are typically derived by performing a series of Bernoulli trials on probabilistic interactions, assuming independence, to generate a distribution of networks of binary interactions to calculate their structure (Poisot, Cirtwill, et al., 2016). One could instead compare the likelihood of an observed network to the one of the most likely network structure (according to the

probabilistic network distribution), thereby directly obtaining a measure of discrepancy of the empirical network. Generating probabilistic ecological networks represents a tangible challenge, one that, in the coming years, promises to unlock doors to more advanced and adequate analyses of ecological networks.

## 1.6. Acknowledgment

We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was supported by the Institute for Data Valorisation (IVADO) and the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the Computational Biodiversity Science and Services (BIOS<sup>2</sup>) program. A special thanks to all members of the Black Holes and Revelations working group (organized by BIOS<sup>2</sup>) and the Poisot Lab for their insightful discussions and valuable feedback on this manuscript.

## Chapter 2 Second article

---

# What constrains food webs? A maximum entropy framework for predicting their structure with minimal biases

by

Francis Banville<sup>1</sup>, Dominique Gravel<sup>2</sup>, and Timothée Poisot<sup>3</sup>

- (<sup>1</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3
- (<sup>2</sup>) Département de biologie, Université de Sherbrooke  
2500 boulevard de l'Université, Sherbrooke, QC, Canada J1K 2R1
- (<sup>3</sup>) Département de sciences biologiques, Université de Montréal  
1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3

This article was submitted in PLOS Computational Biology and can be found at <https://doi.org/10.1371/journal.pcbi.1011458>.

The main contributions of Francis Banville for this articles are presented.

FB developed and performed the analyses and wrote the first and final versions of the manuscript. DG and TP provided guidance on the analyses and interpretation of the results and revised the manuscript. All authors read and approved the manuscript.

**RÉSUMÉ.** Les réseaux trophiques sont des réseaux complexes dont la structure est éco- logiquement et statistiquement contrainte, plusieurs de leurs propriétés émergentes étant fortement corrélées entre elles. Malgré la reconnaissance de ces relations invariantes au sein des réseaux trophiques, l'utilisation du principe d'entropie maximale (MaxEnt) en écologie des réseaux est rare. Cela est surprenant étant donné que MaxEnt est un outil statistique conçu spécifiquement pour comprendre et prédire de nombreux types de systèmes contraints. Ce principe stipule que la distribution de probabilité la moins biaisée d'une propriété d'un système, contrainte par nos connaissances préalables sur ce système, est celle avec l'entropie d'information maximale. MaxEnt s'est révélé utile dans de nombreux problèmes de modéli- sation écologique, mais son application aux réseaux trophiques et écologiques est limitée. Ici, nous montrons comment MaxEnt peut être utilisé pour dériver de nombreuses propriétés des réseaux trophiques, à la fois analytiquement et heuristiquement. Premièrement, nous mon- trons comment la distribution conjointe des degrés (la distribution de probabilité conjointe des nombres de proies et de prédateurs pour chaque espèce dans le réseau) peut être dé- rivée analytiquement en utilisant le nombre total d'espèces et d'interactions des réseaux trophiques. Deuxièmement, nous présentons une approche heuristique et flexible pour trou- ver la matrice d'adjacence d'un réseau (la représentation du réseau sous forme de matrice) en utilisant la méthode de recuit simulé et l'entropie SVD. Nous avons construit deux mo- dèles heuristiques utilisant respectivement la connectance et la séquence de degrés conjointe comme contraintes statistiques. Nous avons comparé les prédictions des deux modèles à celles de modèles nuls et neutres couramment utilisés en écologie des réseaux en utilisant des données en accès libre de réseaux trophiques terrestres et aquatiques échantillonnés à l'échelle globale ( $N = 257$ ). Nous avons constaté que le modèle heuristique contraint par la séquence de degrés conjointe était un bon prédicteur de nombreuses mesures de la struc- ture des réseaux trophiques, en particulier l'emboîtement et la distribution des motifs. Plus précisément, nos résultats suggèrent que la structure des réseaux trophiques terrestres et aquatiques est principalement déterminée par leur distribution conjointe de degrés.

**Mots clés :** modélisation écologique, réseaux écologiques, réseaux trophiques, entropie maximale, modèles nuls

**ABSTRACT.** Food webs are complex ecological networks whose structure is both ecologically and statistically constrained, with many network properties being correlated with each other. Despite the recognition of these invariable relationships in food webs, the use of the principle of maximum entropy (MaxEnt) in network ecology is still rare. This is surprising considering that MaxEnt is a statistical tool precisely designed for understanding and predicting many types of constrained systems. This principle asserts that the least-biased probability distribution of a system's property, constrained by prior knowledge about that system, is the one with maximum information entropy. MaxEnt has been proven useful in many ecological modeling problems, but its application in food webs and other ecological networks is limited. Here we show how MaxEnt can be used to derive many food-web properties both analytically and heuristically. First, we show how the joint degree distribution (the joint probability distribution of the numbers of prey and predators for each species in the network) can be derived analytically using the number of species and the number of interactions in food webs. Second, we present a heuristic and flexible approach of finding a network's adjacency matrix (the network's representation in matrix format) based on simulated annealing and SVD entropy. We built two heuristic models using the connectance and the joint degree sequence as statistical constraints, respectively. We compared both models' predictions against corresponding null and neutral models commonly used in network ecology using open access data of terrestrial and aquatic food webs sampled globally ( $N = 257$ ). We found that the heuristic model constrained by the joint degree sequence was a good predictor of many measures of food-web structure, especially the nestedness and motifs distribution. Specifically, our results suggest that the structure of terrestrial and aquatic food webs is mainly driven by their joint degree distribution.

**Keywords:** ecological modelling, ecological networks, food webs, maximum entropy, null models

## 2.1. Introduction

### 2.1.1. The constrained structure of ecological networks

A variety of measures of the structure of ecological networks have been used to describe the organization of species interactions in a biological community (Delmas et al., 2019). These measures provide valuable information on the functioning of ecosystems and their responses to environmental change (e.g. Pascual and Dunne, 2006, Gómez et al., 2011). For instance, two main properties of a network are its order (the number of nodes or species  $S$ ) and its size (the number of edges or species interactions  $L$ ), whose ratio  $L/S^2$  (the proportion of possible interactions that are realized) is called connectance. In food webs, Dunne et al., 2002 showed that a high connectance can promote the robustness of the community to species lost. Moreover, a different investigation of network structure revealed that plant–pollinator and seed-disperser networks have a highly nested structure that can promote species persistence (Bascompte et al., 2003). Despite the recognition of the ecological significance of network structure, our knowledge of the properties of ecological networks remains limited and is unevenly distributed across space (Poisot et al., 2021). This can largely be attributed to the difficulties encountered when sampling interactions (Jordano, 2016), which impede our ability to describe network structure on a global scale. Moreover, these complex systems (Williams and Martinez, 2000) have an emerging structure with intricate dynamics (i.e. non-linear behaviors that cannot be predicted from individual interactions). Their multi-faceted nature and high level of complexity represent a challenge when studying the ecological processes driving their structure, especially given the strong relationship between many of their properties. For example, nestedness and modularity are highly correlated in ecological networks (Fortuna et al., 2010) and many emerging properties are linked to connectance (Poisot and Gravel, 2014). In light of these observations, it is difficult to assess whether the presumed effects of a particular measure on network structure and behavior are the artifacts of other, perhaps simpler, measures.

One way to address this issue is to recognize that food webs and other ecological networks are constrained systems. In other words, the space of possible network configurations is shaped by our partial knowledge of their structure. For example, there is a finite set of networks with a given number of species and interactions (i.e. a given connectance). As shown by Poisot and Gravel, 2014, connectance can constrain many aspects of network structure such as the degree distribution (i.e. the probability distribution of the number of interactions realized by a species). Because of the scaling relationship between the total number of species and interactions, network structure is also largely constrained by its order (food webs with high species richness typically have a low connectance compared to smaller networks, MacDonald et al., 2020). Other measures, such as the maximum trophic level (i.e. the maximum number of times energy is transformed along food chains through biomass consumption), may also constrain the space of feasible networks, e.g. by imposing a limit on energy transfer (Williams and Martinez, 2004).

Prior knowledge of the structure of ecological networks, even partial, can be useful in the current context of data scarcity about species interactions. The Eltonian shortfall, which describes the gap between our present-day knowledge of ecological networks and a comprehensive understanding of interactions (Hortal et al., 2015), can be partially alleviated using known information about a network's attributes. First, knowing a given set of network properties (e.g. the number of species and interactions), it is possible to identify various structures that meet these constraints and estimate their respective probabilities of occurrence. For instance, a set of adjacency matrices (i.e. a representation of interactions in matrix format) that satisfy the known properties can be found. Selecting the right adjacency matrix among all of these suitable configurations or evaluating their relative probabilities may be done using different modeling tools (e.g. using the principle of maximum entropy or Bayesian inference). Second, as suggested by Strydom, Catchen, et al., 2021, network structure can also be used to facilitate the estimation of pairwise species interactions. For example, by knowing a network's adjacency matrix (or the constrained space of feasible networks), we can decrease the volume of information we need to infer interactions (e.g. by knowing how

many prey each species should have). Moreover, network structure can be used as validation to ensure that inferred interactions collectively satisfy known structural properties. Overall, partial knowledge of a network's properties can help us reconstruct its emerging structure and individual interactions when data is lacking.

Predicting ecological networks and understanding the biological mechanisms that shape species interactions are two complementary aims of network ecology. On one hand, prediction is the process of estimating the value of an unknown variable using available data and appropriate statistical models and assumptions. The goal of predictive models is to minimize predictive errors and provide reliable estimates of the unknown variable. These models have the potential to fill many data gaps about species interactions and network structure by making sound predictions and generating new data. A variety of such models have recently been developed using machine learning and other statistical tools, most of which are presented in Strydom, Catchen, et al., 2021. On the other hand, statistical models can also be built to describe or understand a parameter or mechanism of interest. For example, null models are statistical models that generate baseline distributions to evaluate the significance of an observed measure. These distributions are obtained from different randomizations of a network that maintain some of its properties while deliberately omitting others. By assessing the deviation of empirical data from these reference distributions, we can evaluate the degree to which the processes underlying the null model can accurately capture the observed measures, providing a benchmark to assess the importance of additional factors such as the omitted properties (Fortuna and Bascompte, 2006, Delmas et al., 2019). This makes null models a valuable tool to identify the ecological mechanisms that drive species interactions and constrain the structure of ecological networks. The distinction between understanding and predicting is crucial when using statistical and mathematical models as this can impact how we use and interpret them. Numerous predictive and null models have been used to analyze ecological networks (Delmas et al., 2019, Strydom, Catchen, et al., 2021). Nevertheless, given the constrained nature of ecological networks, it is surprising that

the principle of maximum entropy, a mathematical method designed for both the description and prediction of constrained systems, has been barely used in network ecology.

### 2.1.2. The principle of maximum entropy: A primer for ecologists

The principle of maximum entropy (MaxEnt) is a mathematical method of finding probability distributions, strongly rooted in statistical mechanics and information theory (Jaynes, 1957b, Jaynes, 1957a, Harremoës and Topsøe, 2001). Starting from a set of constraints given by prior knowledge of a system (i.e. what we call state variables), this method helps us find least-biased probability distributions subject to the constraints. These probability distributions are guaranteed to be unique given our prior knowledge and represent the most we can say about a system without making more assumptions. For example, if we know the number of species and total number of individuals in a biological community (our state variables), we can calculate the average number of individuals per species (our constraint). Using the principle of maximum entropy, one could show that the least-biased species abundance distribution (i.e. the distribution of the number of individuals of each species), constrained by the average species abundance, follows an exponential distribution whose parameter is determined by the values of the state variables (Frank and Smith, 2011, Harte and Newman, 2014). However, this does not imply that this distribution will be the best fit to empirical data. The challenge is to find the right set of constraints that would best reproduce distributions found in nature.

MaxEnt states that the least-biased probability distribution given the constraints used is the one with the highest entropy (a measure of the uncertainty of a random process) among all probability distributions that satisfy these constraints. Many measures of entropy have been developed in physics (Beck, 2009), but only a fraction of them could be used as an optimization measure with the principle of maximum entropy. According to Beck, 2009 and Khinchin, 2013, a measure of entropy  $H$  should satisfy four properties: (1) it should be a function of a probability distribution  $p(n)$  only; (2) it should be maximized when  $p(n)$  is uniform; (3) it should not be influenced by outcomes with a null probability; and (4) it should

be independent of the order of information acquisition. The Shannon's entropy (Shannon, 1948)

$$H = - \sum_n p(n) \log p(n) \quad (2.1.1)$$

satisfies all of these properties. Finding the probability distribution  $p(n)$  that maximizes  $H$  under a set of  $m$  constraints  $g$  can be done using the method of Lagrange multipliers. These constraints can include one or many properties of the probability distribution (e.g. its mean, variance, and range). However, the normalization constraint always needs to be included in  $g$  to make sure that  $p(n)$  sums to 1. The objective is then to find the values of the Lagrange multipliers  $\lambda_i$  that optimize a function  $F$ :

$$F = H - \sum_{i=1}^m \lambda_i (g_i - c_i), \quad (2.1.2)$$

where  $g_i$  is the mathematical formulation of the constraint  $i$  and  $c_i$ , its value. Note that  $F$  is just Shannon's entropy to which we added terms that each sums to zero ( $g_i = c_i$ ).  $F$  is maximized by setting to 0 its partial derivative with respect to  $p(n)$ .

The principle of maximum entropy has been used in a wide range of disciplines, from thermodynamics, chemistry and biology (Martyushev and Seleznev, 2006) to graph and network theory (e.g., Park and Newman, 2004, van der Hoorn et al., 2018). MaxEnt has also been proven useful in ecology, e.g. in species distribution (Phillips et al., 2006) and macroecological (Harte et al., 2008, Harte and Newman, 2014) models. In network ecology, it can be used to generate null models of network structure with either fixed or fluctuating constraints (Caruso et al., 2022) and predict important network properties using limited data. For example, it has been used to predict the degree distribution of bipartite networks from the number of species and the number of interactions (Williams, 2011) and the strength of biotic interactions from species relative abundances (Stock et al., 2021). However, to the best of our knowledge, MaxEnt has never been used to predict food-web structure directly,

even though food webs are among the most documented and widespread ecological networks (Ings et al., 2009).

Food-web properties that can be derived using MaxEnt are varied and pertain to different elements of the network (i.e. at the species, interaction and community levels). Because MaxEnt is a method of finding least-biased probability distributions given partial knowledge about a system, these properties need to be represented probabilistically. For example, at the species level, MaxEnt can be used to predict the distribution of trophic levels among species, as well as the distribution of species' vulnerability (number of predators) and generality (number of prey). By contrast, at the interaction level, predictions can be made on the distribution of interaction strengths in weighted food webs. At the community level, it can generate probability distributions of many measures of a network's emerging structure and of networks themselves (i.e. a probability distribution governing the occurrence of different network configurations). Because the decomposition of a network's adjacency matrix into a product of matrices can yield a vector of relative values (e.g. a vector of relative singular values), MaxEnt can also be used to find the configuration whose distribution of relative values is of maximum entropy. This configuration would be the one with the greatest entropy (or complexity, Strydom, Catchen, et al., 2021) among all configurations that satisfy the constraints. Overall, the potential of this method in the study of food webs is broad. The applicability and performance of MaxEnt mostly depend on the ecological information available and on our capacity to find the right set of state variables that best represent natural systems and to translate them into appropriate statistical constraints. Having a validated maximum entropy model for the system at hand allows us to make least-biased predictions using a minimal amount of data, as well as identify the most important ecological processes shaping that system. In other words, MaxEnt can help us better understand and predict the structure of ecological networks worldwide.

### 2.1.3. Analytical and heuristic approaches

In this contribution, we used two complementary approaches to predict the structure of food webs using the principle of maximum entropy. The first approach consists in deriving constrained probability distributions of given network properties analytically, whereas the second approach consists in finding the adjacency matrix of maximum entropy heuristically, from which network properties can be measured. We compared our predictions against empirical data and null and neutral models commonly used in network ecology. We focus on deterministic (non-probabilistic) and unweighted (Boolean) food webs in both approaches for two reasons. (1) Most sampled food webs lack estimates of interaction probabilities (e.g. spatial variability of interactions) and strengths (e.g. energy fluxes between species) because of inherent difficulties in measurement. (2) The methods and measures to analyze the structure of binary food webs are more developed and tested (Delmas et al., 2019), which increases the range of analysis we can perform and simplifies their interpretation. However, our framework can be applied to all types of ecological networks and a wider variety of measures.

For the first approach (analytic), we focus on species-level properties. Specifically, we derived the joint degree distribution (i.e. the joint probability distribution that a species has a given number of prey and predators in its network) of maximum entropy using only the number of species  $S$  and the number of interactions  $L$  as state variables. Then, we calculated the degree distribution of maximum entropy directly from the joint degree distribution by summing probabilities over the joint degree distribution (since the degree of a species is its total number of prey and predators combined). Because of the scarcity of empirical data on the number of interactions in food webs, we also present a method to predict  $L$  from  $S$  (Box 1), thus allowing the prediction of the joint degree distribution from  $S$  solely.

For the second approach (heuristic), we focus on community-level properties. We used a flexible and heuristic model based on simulated annealing (an optimization algorithm, Kirkpatrick et al., 1983) to find the network configuration *close* to maximum entropy and measured its structure. We developed this heuristic model because the analytical derivation

of a maximum entropy graph model of food webs is difficult, and because this model is readily applicable to other types of ecological networks and measures. Indeed, the mathematical representation of food webs (i.e. directed simple graphs frequently having self-loops) makes the optimization of maximum entropy graph models more complicated than with many other types of (non-ecological) networks. In other words, deriving a probability distribution on the graphs themselves is difficult when working with food webs. We built two types of heuristic MaxEnt models depending on the constraint used. Our type I MaxEnt model uses the connectance of the network (i.e. the ratio  $L/S^2$ ) as a constraint, whereas our type II MaxEnt model uses the whole joint degree sequence as a constraint.

## 2.2. Methods

### 2.2.1. Analytical maximum entropy models

The analytical approach is the most common way to use and develop maximum entropy models. As shown above, starting from a defined set of constraints, a mathematical expression for the target distribution is derived using the method of Lagrange multipliers. The derived distribution of maximum entropy is unique and least biased given the constraints used. Although we refer to this approach as analytic, finding the values of the Lagrange multipliers usually requires the use of numerical methods. Here we use MaxEnt to derive two species-level properties in food webs: the joint degree distribution and the degree distribution. The degree distribution has driven the attention of ecologists because of its role in determining the assembly of ecological networks (Vázquez, 2005), shaping their emerging structure (Fortuna et al., 2010), and understanding interaction mechanisms (Williams, 2011). As noted above, although the degree distribution of maximum entropy has already been derived in bipartite networks (Williams, 2011), we show in much greater detail its mathematical derivation in food webs. But first, we derived the joint degree distribution, a related property that holds significantly more ecological information than the degree distribution.

We tested our analytical MaxEnt models against open food-web data queried from three different sources and integrated into what we call our *complete dataset*. These sources include

(1) terrestrial and aquatic food webs sampled globally and archived on the ecological interactions database Mangal; (2) freshwater stream food webs from New Zealand; and (3) aquatic food webs from Tuesday Lake (Michigan, United States). First, all food webs archived on mangal.io (Poisot, Baiser, et al., 2016, Banville et al., 2021) were directly queried from the database ( $N = 235$ ). Most ecological networks archived on Mangal are multilayer networks, i.e. networks that describe different types of interactions. We kept all networks whose interactions were mainly of predation and herbivory types and removed the largest network ( $S = 714$ ) for computational efficiency reasons. Then, to this set, we added food webs from two different sources: the New Zealand dataset ( $N = 21$ , Pomeranz et al., 2018) and the Tuesday Lake dataset ( $N = 2$ , Cohen et al., 2003). Our complete dataset thus contained a total of 257 food webs. These complex food webs differ in their level of resolution and sampling effort, which may introduce noise in the estimation of their properties, especially given their large number of interacting elements. However, because our MaxEnt models are applied on imperfect data, they aim at reproducing the *sampled* structure of food webs, not their actual structure. All code and data to reproduce this article are available at the Open Science Framework (<https://osf.io/kt4gs/files/github>). Data cleaning, simulations and analyses were conducted in Julia v1.8.0.

**2.2.1.1. Joint degree distribution.** The joint degree distribution  $p(k_{in}, k_{out})$  of a food web with  $S$  species is a joint discrete probability distribution describing the probability that a species has  $k_{in}$  predators and  $k_{out}$  prey, with  $k_{in}$  and  $k_{out} \in [0, S]$ . Basal species (e.g. plants) have a  $k_{out}$  of 0, whereas top predators have a  $k_{in}$  of 0. In contrast, the maximum number of prey and predators a species can have is set by the total number of species in the food web. Here we show how the joint degree distribution of maximum entropy can be obtained given knowledge of the number of species  $S$  and the number of interactions  $L$ .

We want to maximize Shannon's entropy

$$H = - \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S p(k_{in}, k_{out}) \log p(k_{in}, k_{out}) \quad (2.2.1)$$

subject to the following constraints:

$$g_1 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S p(k_{in}, k_{out}) = 1; \quad (2.2.2)$$

$$g_2 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{in} p(k_{in}, k_{out}) = \langle k_{in} \rangle = \frac{L}{S}; \quad (2.2.3)$$

$$g_3 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{out} p(k_{in}, k_{out}) = \langle k_{out} \rangle = \frac{L}{S}. \quad (2.2.4)$$

The first constraint  $g_1$  is our normalizing constraint, whereas the other two ( $g_2$  and  $g_3$ ) fix the average of the marginal distributions of  $k_{in}$  and  $k_{out}$  to the linkage density  $L/S$ . It is important to notice that  $\langle k_{in} \rangle = \langle k_{out} \rangle$  because every edge is associated with a predator and a prey. Therefore, without using any further constraints, we would expect the joint degree distribution of maximum entropy to be a symmetric probability distribution with regards to  $k_{in}$  and  $k_{out}$ . However, this does not mean that the joint degree *sequence* will be symmetric since the joint degree sequence is a random realization of its probabilistic counterpart.

The joint probability distribution of maximum entropy given these constraints is found using the method of Lagrange multipliers. To do so, we seek to maximize the following expression:

$$F = H - \lambda_1(g_1 - 1) - \lambda_2 \left( g_2 - \frac{L}{S} \right) - \lambda_3 \left( g_3 - \frac{L}{S} \right), \quad (2.2.5)$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the Lagrange multipliers. The probability distribution that maximizes entropy is obtained by finding these values. As pointed out above,  $F$  is just Shannon's entropy to which we added terms that each sums to zero (our constraints).  $F$  is maximized by setting to 0 its partial derivative with respect to  $p(k_{in}, k_{out})$ . Because the derivative of a constant is zero, this gives us:

$$\frac{\partial H}{\partial p(k_{in}, k_{out})} = \lambda_1 \frac{\partial g_1}{\partial p(k_{in}, k_{out})} + \lambda_2 \frac{\partial g_2}{\partial p(k_{in}, k_{out})} + \lambda_3 \frac{\partial g_3}{\partial p(k_{in}, k_{out})}. \quad (2.2.6)$$

Evaluating the partial derivatives with respect to  $p(k_{in}, k_{out})$ , we obtain:

$$-\log p(k_{in}, k_{out}) - 1 = \lambda_1 + \lambda_2 k_{in} + \lambda_3 k_{out}. \quad (2.2.7)$$

Then, solving Eq 2.2.7 for  $p(k_{in}, k_{out})$ , we obtain:

$$p(k_{in}, k_{out}) = \frac{e^{-\lambda_2 k_{in} - \lambda_3 k_{out}}}{Z}, \quad (2.2.8)$$

where  $Z = e^{1+\lambda_1}$  is called the partition function. The partition function ensures that probabilities sum to 1 (our normalization constraint). It can be expressed in terms of  $\lambda_2$  and  $\lambda_3$  as follows:

$$Z = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S e^{-\lambda_2 k_{in} - \lambda_3 k_{out}}. \quad (2.2.9)$$

After substituting  $p(k_{in}, k_{out})$  in Eqs 2.2.3 and 2.2.4, we get a nonlinear system of two equations and two unknowns:

$$\frac{1}{Z} \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{in} e^{-\lambda_2 k_{in} - \lambda_3 k_{out}} = \frac{L}{S}; \quad (2.2.10)$$

$$\frac{1}{Z} \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{out} e^{-\lambda_2 k_{in} - \lambda_3 k_{out}} = \frac{L}{S}. \quad (2.2.11)$$

We solved Eqs 2.2.10 and 2.2.11 numerically using the Julia library JuMP.jl v0.21.8 (Dunning et al., 2017). JuMP.jl supports nonlinear optimization problems by providing exact

second derivatives that increase the accuracy and performance of its solvers. The estimated values of  $\lambda_2$  and  $\lambda_3$  can be substituted in Eq 2.2.8 to have a more workable expression for the joint degree distribution.

We assessed the empirical support of this expression using all food webs in our complete dataset. First, we predicted the joint degree distribution of maximum entropy for each of these food webs, i.e. using their number of species and number of interactions as state variables. Then, we sampled one realization of the joint degree sequence for each network using the probabilities given by the joint degree distribution, while fixing the total number of interactions. This gave us a random realization of the number of prey and predators for each node in each network. We standardized the predicted  $k_{out}$  and  $k_{in}$  by the total number of species in their network to generate relative values, which can be compared across networks.

**2.2.1.2. Degree distribution.** The joint degree distribution derived using MaxEnt (Eq 2.2.8) can be used to obtain the degree distribution of maximum entropy. The degree distribution  $p(k)$  represents the probability that a species has  $k$  interactions in its food web, with  $k = k_{in} + k_{out}$ . It can thus be obtained from the joint degree distribution as follows:

$$p(k) = \sum_{i=0}^k p(k_{in} = k - i, k_{out} = i). \quad (2.2.12)$$

The degree distribution can also be obtained directly using the principle of maximum entropy, as discussed in Williams, 2011. This gives the following distribution:

$$p(k) = \frac{e^{-\lambda_2 k}}{Z}, \quad (2.2.13)$$

with  $Z = \sum_{k=0}^S e^{-\lambda_2 k}$ .

This can be solved numerically using the constraint of the average degree  $\langle k \rangle = \frac{2L}{S}$  of a species, yielding an identical solution to the one obtained using the joint degree distribution as an intermediate.

$$\frac{1}{Z} \sum_{k=0}^S k e^{-\lambda_2 k} = \frac{2L}{S} \quad (2.2.14)$$

Note that the mean degree is twice the value of the linkage density because every interaction must be counted twice when we add in and out-degrees together.

### 2.2.2. Heuristic maximum entropy models

With the analytical approach, we showed how important measures of food-web structure (e.g. the degree distribution and the joint degree distribution) can be derived with the principle of maximum entropy using minimal knowledge about a biological community. This type of model, although useful to make least-biased predictions of many network properties, can be hard to apply for other measures. There are dozens of measures of network structure (Delmas et al., 2019) and many are not directly calculated with mathematical equations but are instead estimated algorithmically. Moreover, the applicability of this method to empirical systems is limited by the state variables we can measure and use. Here, we suggest a more flexible method to predict many measures of network structure simultaneously, i.e. by finding heuristically the network configuration of maximum SVD (singular value decomposition) entropy given partial knowledge of its emerging structure.

**2.2.2.1. From Shannon's to SVD entropy.** The principle of maximum entropy can be applied on the network itself if we decompose its adjacency matrix into a non-zero vector of relative values. This is a necessary step when working with food webs, which are frequently expressed as a matrix  $A = [a_{ij}]$  of Boolean values representing the presence ( $a_{ij} = 1$ ) or absence ( $a_{ij} = 0$ ) of an interaction between two species  $i$  and  $j$ . Knowing one or many properties of a food web of interest (e.g. its number of species and number of interactions), we can simulate its adjacency matrix randomly by using this known ecological information to constrain the space of potential networks. The entropy of this hypothetical matrix can then be measured after decomposing it into appropriate values. Simulating a series of networks until we find the one having the highest entropy allows us to search for the most complex

food-web configuration given the ecological constraints used. This configuration is the least-biased one considering the information available. In other words, the most we can say about a network's adjacency matrix, without making more assumptions than the ones given by our incomplete knowledge of its structure, is the one of maximum entropy. Generating the most complex network that corresponds to this structure allows us to explore more easily other properties of food webs under MaxEnt.

Shannon's entropy can only be calculated on conventional probability distributions such as the joint degree distribution. This is an issue when working with the adjacency matrix of ecological networks. For this reason, we need to use another measure of entropy if we want to predict a network's configuration directly using MaxEnt. We used the SVD entropy as our measure of entropy, which is an application of Shannon's entropy to the relative non-zero singular values of a food web's adjacency matrix (Strydom, Dalla Riva, and Poisot, 2021). These values are obtained by performing singular value decomposition (SVD) on the matrix. SVD is a mathematical method that decomposes a matrix into three separate matrices, one of which contains singular values on the diagonal. We selected the non-zero singular values (i.e. the informative singular values that reflect a non-negligible contribution to the structure of the matrix) by keeping the largest  $R$  values, where  $R$  is the rank of the matrix (Golub et al., 1987). We measured SVD entropy as follows:

$$J = - \sum_{i=1}^R s_i \log s_i, \quad (2.2.15)$$

where  $s_i$  are the relative singular values of the adjacency matrix ( $s_i = \sigma_i / \sum_{i=1}^R \sigma_i$ , where  $\sigma_i$  are the singular values). Note that the distribution of relative singular values is analogous to a probability distribution, with  $0 < s_i < 1$  and  $\sum s_i = 1$ . This measure also satisfies all four properties of an appropriate entropy measure above-mentioned, while being a proper measure of the internal complexity of food webs (Strydom, Dalla Riva, and Poisot, 2021). Following Strydom, Dalla Riva, and Poisot, 2021, we standardized it with the rank of the

matrix (i.e.  $J/\ln(R)$ ) to account for the difference in dimensions between networks (*sensu* Pielou's evenness, Pielou, 1975).

**2.2.2.2. Types I and II heuristic MaxEnt models.** We used SVD entropy to predict the network configuration of maximum entropy (i.e. of maximum complexity) heuristically given different constraints for all food webs in our complete dataset. We built two types of heuristic MaxEnt models that differ on the constraint used. The type I heuristic MaxEnt model is based on connectance, whereas the type II heuristic MaxEnt model is based on the joint degree sequence. These models are thus based on the same constraints as the types I (Fortuna and Bascompte, 2006) and II (Bascompte et al., 2003) null models (Box 2) frequently used to generate random networks topologically. This allows a direct comparison of the performance of null and heuristic MaxEnt models in reproducing the emerging structure of empirical food webs.

For each network, we estimated their configuration of maximum entropy given each of these constraints. For both types of heuristic MaxEnt models, we used a simulated annealing algorithm with 4 chains, 2000 steps and an initial temperature of 0.2. For each food web, we first generated one random Boolean matrix per chain while fixing the number of species. We also maintained the total number of interactions (i.e. the sum of all elements in the matrix) in the type I MaxEnt model and the joint degree sequence (i.e. the rows and columns sums) in the type II MaxEnt model. These were our initial configurations. Then, we swapped interactions sequentially while maintaining the original connectance or joint degree sequence. Configurations with a higher SVD entropy than the previous one in the chain were always accepted, whereas they were accepted with a probability conditional to a decreasing temperature and the difference in SVD entropy when lower. The final configuration with the highest SVD entropy among the four chains constitutes the estimated maximum entropy configuration of a food web given the constraint used.

**2.2.2.3. Structure of MaxEnt food webs.** We measured various properties of these configurations of maximum entropy to evaluate how well they reproduce the structure of sampled food webs. Specifically, we evaluated their nestedness  $\rho$ , their maximum trophic level *maxtl*,

their network diameter  $diam$ , their average maximum similarity between species pairs  $MxSim$  (Williams and Martinez, 2000), their proportion of cannibal species  $Cannib$ , their proportion of omnivorous species  $Omniv$ , their SVD entropy, and their motif profile. Nestedness indicates how much the diet of specialist species is a subset of the one of generalists (Delmas et al., 2019) and was measured using the spectral radius of the adjacency matrix (Staniczenko et al., 2013). Next, the network diameter represents the longest of the shortest paths between all species pair (Albert and Barabasi, 2002). Further, cannibal species are species that can eat individuals of their own species (i.e. species having self-loops), whereas omnivorous species can prey on different trophic levels (Williams and Martinez, 2000). Finally, motifs are unique n-species connected subgraphs that can be considered simple building blocks of ecological networks (Milo et al., 2002, Stouffer et al., 2007). There are 13 possible three-species motifs in food webs, including 5 with only single links ( $i \rightarrow j$ ) and 8 with double links ( $i \leftrightarrow j$ ), labeled S1-S5 and D1-D8 by Stouffer et al., 2007, respectively. A motif profile represents the proportion of each of these motifs in a network (i.e. their relative frequencies). All of these properties are relatively easy to measure and were chosen based on their ecological importance and prevalent use in network ecology (McCann, 2011, Delmas et al., 2019).

We compared the performance of both heuristic MaxEnt models in predicting these measures to the one of the null and neutral models (Box 2). We conducted these comparisons using two different datasets: (1) our complete dataset including most food webs archived on Mangal, as well as all food webs in the New Zealand and Tuesday Lake datasets, and (2) our *abundance dataset*, a subset of the complete dataset comprising all food webs having data on their species' relative abundances ( $N = 19$ ). Indeed, of the New Zealand and Tuesday Lake datasets, 19 networks had data on species' relative abundances that were used in the neutral model to better assess the performance of our heuristic models. We compared our models' predictions using these two datasets separately to minimize biases and to better represent food webs with abundance data.

### Box 1: Working with predicted numbers of interactions

Our analytical MaxEnt models require information on the number of species and the number of interactions. However, since the latter is rarely measured empirically, ecologists might need to use another predictive model to estimate the total number of interactions in a food web before using MaxEnt. Here we show how this can be done by combining both models sequentially.

We used the flexible links model of MacDonald et al., 2020 to predict the number of interactions from the number of species in food webs. The flexible links model has been shown to make reliable predictions of  $L$  while taking into account meaningful ecological constraints on its lower and upper bounds. Specifically, recognizing that there is a minimum of  $S - 1$  (if no isolated species) and a maximum of  $S^2$  (if all species interact) interactions in food webs, it estimates the number of the  $S^2 - (S - 1)$  *flexible links* that are realized. This represents the number of realized interactions above the minimum (i.e.  $L_{FL} = L - (S - 1)$ ). The probability that a flexible link is realized is treated as constant within a particular food web but differs across food webs to capture variations in connectance. Therefore, assuming that flexible links are independent Bernoulli events, the total number of flexible links that are realized in a food web can be obtained from a beta-binomial distribution BB with  $S^2 - (S - 1)$  trials and parameters  $\alpha = \mu e^\phi$  and  $\beta = (1 - \mu)e^\phi$ :

$$L_{FL} \sim \text{BB}(S^2 - (S - 1), \mu e^\phi, (1 - \mu)e^\phi), \quad (2.2.16)$$

where  $\mu$  is the average probability across food webs that a flexible link is realized and  $\phi$  is the concentration parameter around  $\mu$ .

We fitted the flexible links model on all food webs in our complete dataset and estimated the parameters of Eq 2.2.16 using a Hamiltonian Monte Carlo sampler with static trajectory (4 chains and 3000 iterations):

$$[\mu, \phi | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^m \text{BB}(L_i - (S_i - 1) | S_i^2 - (S_i - 1), \mu e^\phi, (1 - \mu) e^\phi) \\ \times \text{B}(\mu | 3, 7) \times \mathcal{N}(\phi | 3, 0.5),$$

where  $m$  is the number of food webs ( $m = 257$ ) and  $\mathbf{L}$  and  $\mathbf{S}$  are the vectors of their numbers of interactions and numbers of species, respectively. Our weakly-informative prior distributions were chosen following MacDonald et al., 2020, i.e. a beta distribution for  $\mu$  and a normal distribution for  $\phi$ . The Monte Carlo sampling of the posterior distribution was conducted using the Julia library Turing v0.15.12.

The flexible links model is a generative model, i.e. it can generate plausible values of the predicted variable. We thus simulated 1000 values of  $L$  for different values of  $S$  using the joint posterior distribution of our model parameters and calculated the mean degree for each simulated value. The resulting distributions are shown in the left panel of Fig 1 for three different values of species richness. In the right panel of Fig 1, we show how the probability distribution for the mean degree constraints can be used to generate a distribution of maximum entropy degree distributions since each simulated value of mean degree generates a different maximum entropy degree distribution (Eqs 2.2.13 and 2.2.14).

## Box 2: Corresponding null and neutral models

### Null models (types I and II)

The predictions of our heuristic maximum entropy models were compared against two topological null models. These null models use the same ecological information as our heuristic models and thus constitute an adequate baseline for comparison. The first is the type I null model of Fortuna and Bascompte, 2006, in which the probability that a species  $i$  predares on another species  $j$  is given by

$$p(i \rightarrow j) = \frac{L}{S^2}. \quad (2.2.17)$$

The second is the type II null model of Bascompte et al., 2003, in which the probability of interaction is instead given by

$$p(i \rightarrow j) = \frac{1}{2} \left( \frac{k_{in}(j)}{S} + \frac{k_{out}(i)}{S} \right), \quad (2.2.18)$$

where  $k_{in}(j)$  and  $k_{out}(i)$  are the in and out-degrees of species  $j$  and  $i$ , respectively. The type I null model is based on connectance, whereas the type II null model is based on the joint degree sequence. Therefore, the type I and II topological null models correspond to our type I and II heuristic MaxEnt models, respectively, since they use similar constraints.

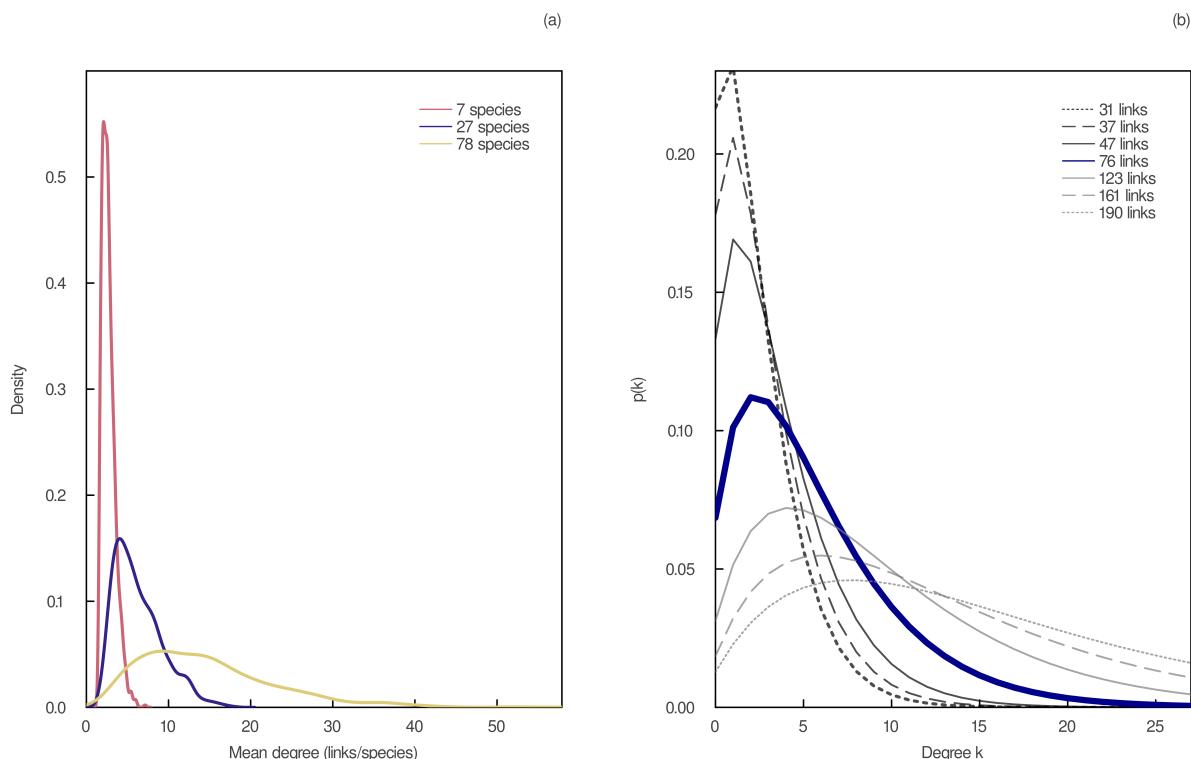
We generated probabilistic networks using both types of null models for all empirical food webs in our complete dataset. Then, we converted these networks to adjacency matrices of Boolean values by generating 100 random networks for each of these probabilistic webs and kept the  $L$  entries that were sampled the most amount of times, with  $L$  given by the number of interactions in each food web. This ensured that the resulting null networks had the same number of interactions as their empirical counterparts. Thus, for each null model, we ended up with one null adjacency matrix for each empirical network.

### Neutral model

We also compared our heuristic MaxEnt models to a neutral model of relative abundances, in which the probabilities of interaction are given by

$$p(i \rightarrow j) \propto \frac{n_i}{N} \times \frac{n_j}{N}, \quad (2.2.19)$$

where  $n_i$  and  $n_j$  are the abundances (or biomass) of both species and  $N$  is the total abundance (or biomass) of all species in the network. We generated neutral abundance matrices for all empirical food webs in our abundance dataset and converted these weighted networks to adjacency matrices of Boolean values using the same method as the one we used for our null models.

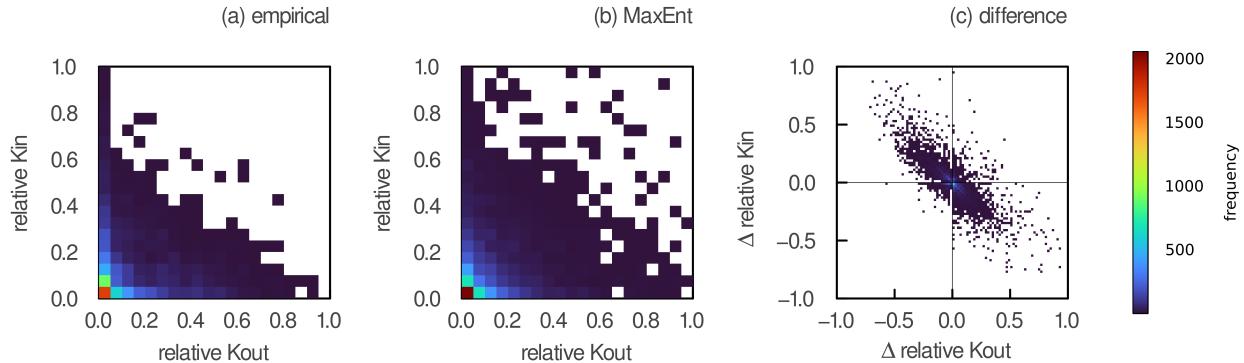


**Fig. 1. Maximum entropy degree distributions with predicted numbers of interactions.** (a) Probability density of the mean degree of a food web obtained using different values of species richness  $S$ . The number of interactions  $L$  was simulated 1000 times using the flexible links model fitted to all empirical networks. The mean degrees  $2L/S$  were then obtained from these simulated values. (b) Degree distributions of maximum entropy for a network of  $S = 27$  species and different numbers of interactions. The numbers of interactions correspond to the lower and upper bounds of the 67%, 89%, and 97% percentile intervals (PI), as well as the median (in blue), of the counterfactuals of the flexible links model. Each degree distribution of maximum entropy was obtained using Eq 2.2.13 after solving numerically Eq 2.2.14 using different values of the mean degree constraint.

## 2.3. Results and Discussion

### 2.3.1. Analytical maximum entropy models

We first discuss the predictive capacity of our analytical models. The relationship between the relative numbers of prey  $k_{out}$  and predators  $k_{in}$  in empirical networks and obtained from the joint degree distributions of maximum entropy is depicted in the left and central panels of Fig 2, respectively. We observe that our analytical model predicts higher values of generality and vulnerability compared to empirical food webs (i.e. relative values of  $k_{out}$  and  $k_{in}$  both closer to 1) for many species. In other words, our model predicts that species that have many predators also have more prey than what is observed empirically (and conversely). This is not surprising, given that our model did not include biological factors preventing generalist predators from having many prey. Nevertheless, with the exception of these generalist species, MaxEnt adequately predicts that most species have low generality and vulnerability values.

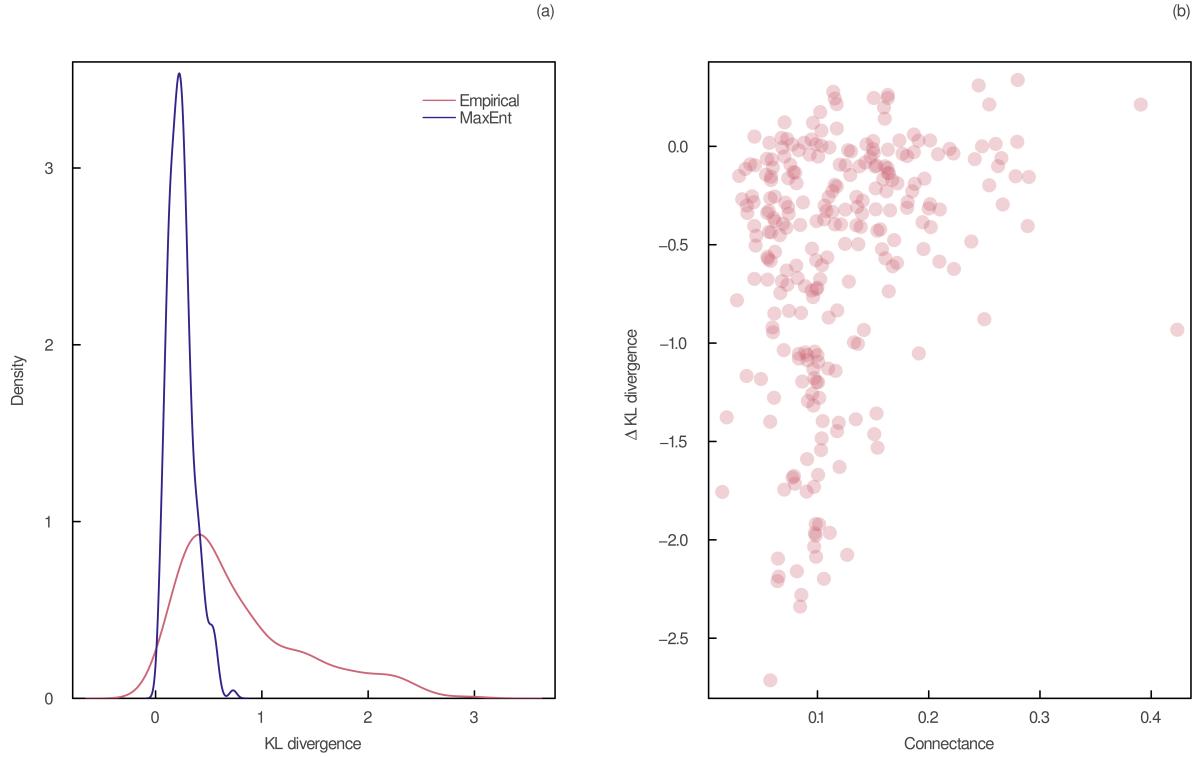


**Fig. 2. Prediction errors of the relative number of predators and prey.** The relative number of predators ( $k_{in}$ ) is plotted against the relative number of prey ( $k_{out}$ ) for each species in all (a) empirical and (b) predicted joint degree sequences. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant. (c) Difference between predicted and empirical values when species are ordered according to their total degree. Due to significant data overlap, all relationships are represented as 2D histograms. The color bar indicates the number of species that fall within each bin.

Examining the difference between predicted and empirical values for each species gives a slightly different perspective (right panel of Fig 2). To make that comparison, we must first associate each of our predictions with a specific species in a network. Indeed, our predicted joint degree sequences have the same number of species (elements) as their empirical counterparts, but they are species agnostic. In other words, instead of predicting a pair of values for each species directly (i.e. the number of prey and predators of a given species  $i$ ), we predicted the entire joint degree sequence without taking into account species' identity (i.e. the distribution of the number of prey and predators for the entire set of species, without knowing which values belong to which species). The challenge is thus to adequately associate predictions with empirical data. In Fig 2, we present these differences when species are ordered by their total degree in their respective networks (i.e. by the sum of their in and out-degrees). This means that the species with the highest total degree in its network will be associated with the highest prediction, and so forth. Doing so, we see that species predicted to have a higher number of predators than what is observed generally have a lower number of prey than what is observed (and conversely). This is also shown in Fig 7 (Appendix B), which represents the relationship between prediction errors in the *absolute* (non-relative) values of  $k_{out}$  and  $k_{in}$  across networks of varying levels of species richness. This is because the difference in total degree ( $k_{out} + k_{in}$ ) between predictions and empirical data is minimized when species are ranked by their total degree (i.e. the average deviation of the sum of relative  $k_{out}$  and  $k_{in}$  is close to 0 across all species). This result thus shows that the difference between predicted and empirical total degrees is low for most species when ordered by their total degrees. There are no apparent biases towards in or out degrees. In Fig 8 (Appendix B), we show how these differences change when species are instead ordered by their out-degrees (left panel) and in-degrees (right panel), i.e. when minimizing the error in the estimation of the out and in-degrees, respectively.

Another way to evaluate the empirical support of the sampled joint degree sequences is to compare their shape with the ones of empirical food webs. We described the shape of a joint degree sequence by measuring the distance between its in and out-degree sequences (i.e. the

distance between its marginal distributions). To do so, we calculated the Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951) between the in and out-degree sequences of each predicted and empirical distribution. The KL divergence is a measure of relative entropy describing the difference between two distributions. Low values indicate high similarity between the in and out-degree sequences and suggest that the joint degree sequence has a high level of symmetry. We compared the shape of the empirical and predicted joint degree sequences in the left panel of Fig 3. As expected, our model predicts more similar in-degree and out-degree sequences than empirical data (shown by lower KL divergence values). However, the difference between the KL divergence of predicted and empirical joint degree sequences decreases with connectance (right panel of Fig 3). This might be because food webs with a low connectance are harder to predict than food webs with a high connectance. Indeed, in low connectance systems, what makes two species interact may be more important for prediction than in high connectance systems, in which what prevents species from interacting may be more meaningful. This implies that more ecological information may be needed in food webs with a low connectance because more ecological processes determine interactions compared to non-interactions. Therefore, other ecological constraints might be needed to account for the asymmetry of the joint degree distribution, especially for networks with a lower connectance. Nevertheless, our MaxEnt model seems to capture quite well the shape of the joint degree sequence for networks having a high connectance.



**Fig. 3. Shape of empirical and predicted joint degree sequences.** (a) Probability density of KL divergence between in and out-degree sequences of empirical and predicted joint degree sequences. (b) Difference between the KL divergence of empirical and predicted joint degree sequences as a function of connectance. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant.

Regarding the degree distribution of maximum entropy, one aspect that informs us of its ecological realism is the number of isolated species it predicts. As MacDonald et al., 2020 pointed out, the size of food webs should at least be of  $S - 1$  interactions since a lower number would yield isolated species, i.e. species without any predator or prey. Because non-basal species must eat to survive, isolated species could indicate that other species are missing; otherwise, isolated species should be removed from the network. In Fig 9 (Appendix B), we show that the degree distribution of maximum entropy, given  $S$  and  $L$ , gives a very low probability that a species will be isolated in its food web (i.e. having  $k = 0$ ) when  $L > S - 1$ . However, under our purely information-theoretic model, the probability that a species is isolated is quite high when the total number of interactions is below  $S - 1$ . Moreover, the expected proportion of isolated species rapidly declines by orders of magnitude with increasing numbers of species and interactions. This supports the ecological realism of the degree distribution of maximum entropy derived above. Nevertheless, ecologists wanting to model a system without allowing isolated species could simply change the lower limit of  $k$  to 1 in Eq 2.2.14 and solve the resulting equation numerically.

### 2.3.2. Heuristic maximum entropy models

In this section, we explore the predictions of our heuristic models. Overall, we found that the models based on the joint degree sequence (i.e. the type II null and heuristic MaxEnt models) reproduced the structure of empirical food webs much better than the ones based on connectance (i.e. the type I null and heuristic MaxEnt models, Table 1). This suggests that the predictive capacity of connectance might be more limited than what was previously suggested (Poisot and Gravel, 2014). On the other hand, the neutral model of relative abundances was surprisingly good at predicting the maximum trophic level and the network diameter (Table 2). However, with the exception of the network diameter, the type II heuristic MaxEnt model was better at predicting network structure than the neutral model for most measures considered. This might be because, although neutral processes are important, they act in concert with niche processes in determining species interactions

(Bartomeus et al., 2016, Canard et al., 2014, Poisot et al., 2015, Pomeranz et al., 2019). The joint degree sequence captures information on both neutral and niche processes because the number of prey and predators a species has is determined by its relative abundance and biological traits. These results thus show that having information on the number of prey and predators for each species substantially improves the prediction of food-web structure, both compared to models solely based on connectance and to the ones solely based on species relative abundances.

<b>model</b>	$\rho$	<b>maxtl</b>	<b>diam</b>	<b>MxSim</b>	<b>Cannib</b>	<b>Omniv</b>	<b>entropy</b>
null 1	-0.167	0.980	1.428	-0.502	2.007	1.493	0.056
MaxEnt 1	-0.226	0.831	1.274	-0.524	1.982	1.863	0.106
null 2	0.160	-0.125	0.016	0.007	1.078	0.559	-0.023
MaxEnt 2	-0.015	0.178	0.565	-0.282	0.698	0.589	0.058

**Table 1. Standardized mean difference between predicted network measures and empirical data for all food webs in our complete dataset ( $N = 257$ ).** Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence.  $\rho$ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level. *diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self-loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy.

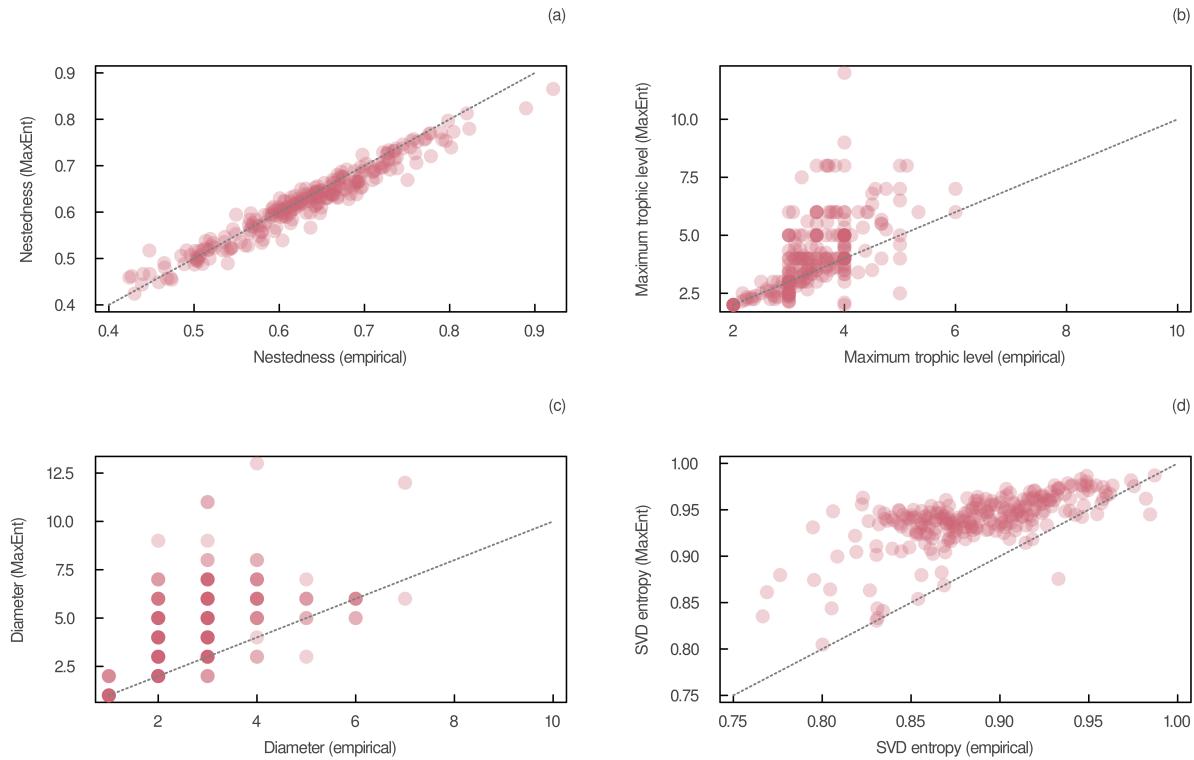
<b>model</b>	$\rho$	<b>maxtl</b>	<b>diam</b>	<b>MxSim</b>	<b>Cannib</b>	<b>Omniv</b>	<b>entropy</b>
neutral	0.367	-0.090	0.027	0.266	6.870	0.576	-0.083
null 1	-0.134	0.950	1.919	-0.369	2.077	0.614	0.068
MaxEnt 1	-0.229	1.020	1.946	-0.355	2.215	0.801	0.121
null 2	0.128	-0.115	-0.135	0.157	1.444	0.029	-0.021
MaxEnt 2	-0.010	0.054	0.243	-0.062	-0.038	0.083	0.038

**Table 2.** Standardized mean difference between predicted network measures and empirical data for all food webs in our abundance dataset ( $N = 19$ ). Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence.  $\rho$ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level. *diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self-loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy.

Next, the predictions of the type II heuristic MaxEnt model can be compared to its null model counterpart. On average, the type II heuristic MaxEnt model was better at predicting nestedness ( $0.62 \pm 0.08$ ) than its corresponding null model ( $0.73 \pm 0.05$ ; empirical networks:  $0.63 \pm 0.09$ ) for networks in our complete dataset (Table 1). This might in part be due to the fact that nestedness was calculated using the spectral radius of the adjacency matrix, which directly leverages information on the network itself just like the heuristic MaxEnt model. The proportion of self-loops (cannibal species) was also better predicted by the type II heuristic MaxEnt model in comparison to the type II null model. However, the type II null model was better at predicting network diameter and average maximum similarity between species pairs, and predictions of the maximum trophic level and the proportion of omnivorous species were similar between both types of models. We believe that this is because increasing the complexity of a food web might increase its average and maximum food-chain lengths. In comparison, the null model was more stochastic and does not necessarily produce more complex food webs with longer food-chain lengths.

Moreover, we found that the entropy of empirical food webs was slightly lower than their maximum entropy when constrained by their joint degree sequence Fig 10 (Appendix B). Empirical food webs had an SVD entropy of  $0.89 \pm 0.04$ , compared to an SVD entropy of  $0.94 \pm 0.03$  for networks generated using the type II heuristic MaxEnt model. The relationship between the SVD entropy of empirical food webs and their maximum entropy is plotted in the last panel of Fig 4. The slight increase in entropy confirms that our method generated more complex networks. Even though we found that many measures of empirical networks are close to the ones of their maximum entropy configuration, the relatively low predictability of entropy itself may be indicative of additional constraints shaping food-web structure, especially for networks with low SVD entropy. Incorporating more constraints into the model could increase its capacity to generate networks with an adequate level of complexity, as shown by the decrease in predictive errors of entropy of the type II heuristic MaxEnt model compared to the one based on connectance (Table 1). Additionally, we found no clear relationship between the increase in SVD entropy and the number of species, the

number of interactions, and connectance Fig 11 (Appendix B). This suggests that our model captured the complexity of small and large networks on a similar level and that its capacity to reproduce food-web structure was unrelated to the order and size of the network. In other words, the gap in entropy between empirical food webs and their maximum entropy configuration may be the result of additional constraints that were not taken into account in the model, regardless of the number of species and the number of interactions.

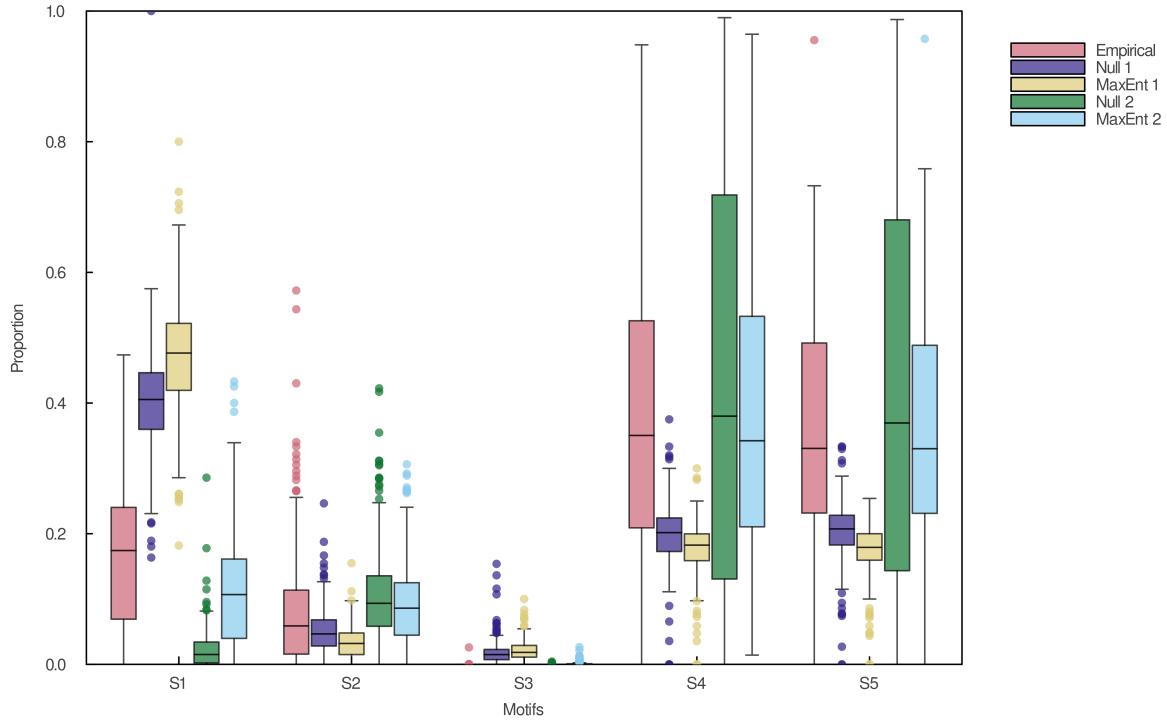


**Fig. 4. Relationship between the structure of empirical and maximum entropy food webs.** Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (a) Nestedness (estimated using the spectral radius of the adjacency matrix), (b) the maximum trophic level, (c) the network diameter, and (d) the SVD entropy were measured on these empirical and maximum entropy food webs. The identity line is plotted in each panel.

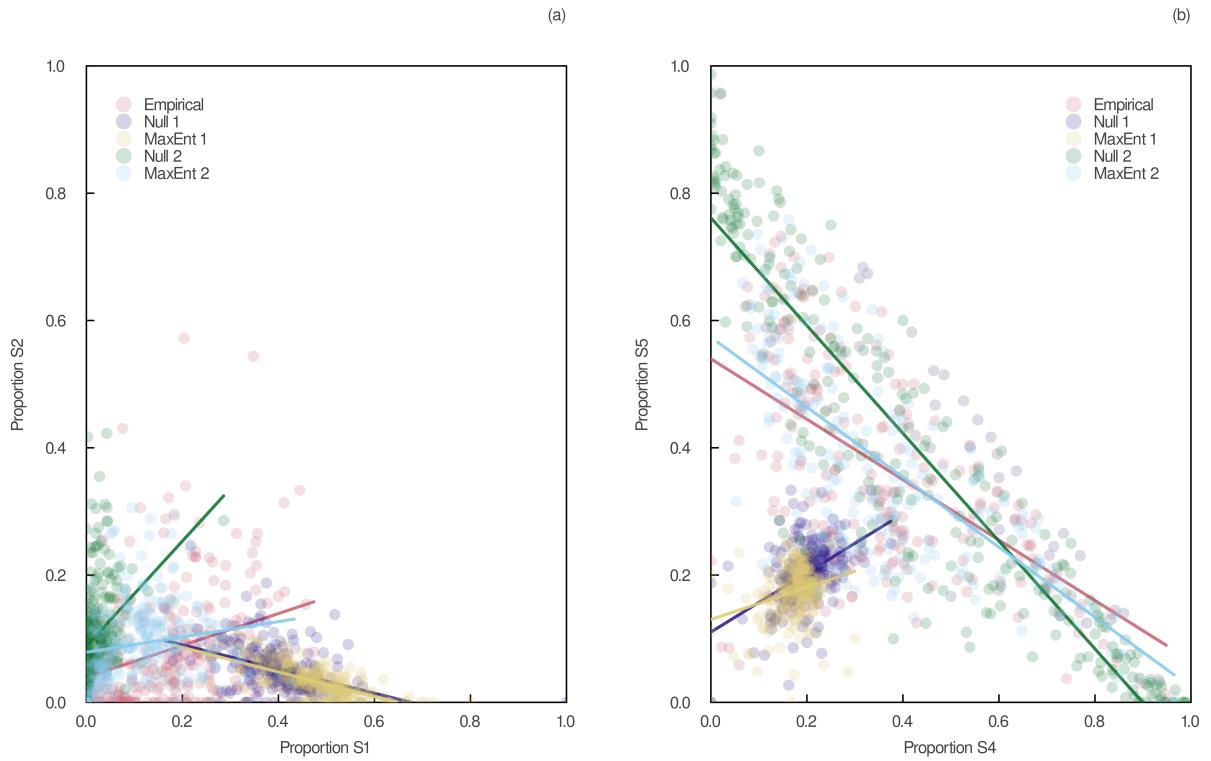
A direct comparison of the structure of maximum entropy food webs constrained by the joint degree sequence with empirical data also supports the results depicted in Table 1. In Fig 4, we show how well empirical measures are predicted by the type II heuristic MaxEnt model. Following our previous results, we found that nestedness was very well predicted by our model. However, the model overestimated the maximum trophic level and network diameter, especially when the sampled food web had intermediate values of these measures. In Fig 12 (Appendix B), we show that the pairwise relationships between the four measures in Fig 4 and species richness in empirical food webs are similar (in magnitude and sign) to the ones found in food webs generated using the type II heuristic MaxEnt model. This indicates that the number of species in the network does not seem to impact the ability of the model to reproduce food-web structure.

Notwithstanding its difficulties in reproducing adequate measures of food-chain lengths, the type II heuristic MaxEnt model can predict surprisingly well the proportions of three-species motifs in empirical food webs. Motifs have been shown to be the backbone of complex ecological networks on which network structure is built and play a crucial role in community dynamics and assembly (Stouffer and Bascompte, 2011). Differences in motif profiles between an observed food web and null model-generated ones can unveil important ecological mechanisms that contribute to network structure (Stouffer et al., 2007). In Fig 5, we show that the motif profile of networks generated using the type II heuristic MaxEnt model accurately reproduced the one of empirical data. This model made significantly better predictions than the ones based on connectance and the type II null model based on the joint degree sequence. This is also shown in Fig 6, which reveals that the relationships between the proportions of single-link motifs in empirical food webs are similar to the ones in networks generated using the type II heuristic MaxEnt model. This is in contrast with the type I null and MaxEnt models based on connectance, which produced opposite relationships than what was observed empirically. Our findings show that generating the most complex food web constrained by the joint degree sequence using maximum entropy does not alter the proportions of three-species motifs on the whole. This suggests that motif profiles may simply be

a statistical attribute of food webs driven by the joint degree sequence. However, given the incapacity of our MaxEnt models to accurately predict food-chain lengths, the way motifs interconnect with each other may hold greater biological significance than the proportion of motifs itself.



**Fig. 5. Proportions of single-link three-species motifs in empirical and predicted food webs.** S1: Tri-trophic chain (a top predator feeds on a meso-predator which feeds on a basal prey). S2: Omnivory (a top predator feeds on a meso-predator and a basal prey). S3: Tri-trophic feeding loop (a cyclic three-species predator-prey system). S4: Apparent competition (a predator feeds on two prey). S5: Exploitative competition (two predators feed on the same prey). Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. Boxplots display the median proportion of each motif in food webs (middle horizontal lines), as well as the first (bottom horizontal lines) and third (top horizontal lines) quartiles. Vertical lines encompass all data points that fall within 1.5 times the interquartile range from both quartiles, and dots are data points that fall outside this range. Only the single-link motifs S1-S5 are shown given the scarcity of double-link motifs in most empirical and predicted networks.



**Fig. 6. Pairwise relationships between the proportions of single-link three-species motifs in empirical and predicted food webs.** S1: Tri-trophic chain. S2: Omnivory. S4: Apparent competition. S5: Exploitative competition. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. Regression lines are plotted in each panel. Motif S3 is not shown because of its low proportion in most empirical and predicted networks.

One of the challenges in implementing and validating a maximum entropy model is to discover where its predictions break down. The results depicted in Table 1 and Fig 4 show that our type II heuristic MaxEnt model can capture many high-level properties of food webs, but does a poor job of capturing others. This suggests that, although the joint degree sequence is an important driver of food-web structure, other ecological constraints might be needed to account for some emerging food-web properties, especially the ones regarding food-chain lengths. Nevertheless, Figs 5 and 6 show that this model can reproduce surprisingly well motif profiles, one of the most ecologically informative properties of food webs. This suggests that the emerging structure of food webs is mainly driven by their joint degree sequence, although higher-level properties might need to be included in the model to ensure that food-chain lengths fall within realistic values.

## 2.4. Conclusion

The principle of maximum entropy is a mathematical method of finding least-biased probability distributions that have some specified properties given by prior knowledge about a system. We first applied this conventional MaxEnt approach on food webs to predict species-level properties, namely the joint degree distribution and the degree distribution of maximum entropy given known numbers of species and interactions. We found that the joint degree distributions of maximum entropy had a similar shape to the ones of empirical food webs in high-connectance systems. However, these MaxEnt distributions were more symmetric than the ones of empirical food webs when connectance was low, which suggests that other constraints might be needed to improve these predictions in low-connectance systems. Then, we used a slightly different approach that aimed at finding heuristically the network configuration with the highest SVD entropy, i.e. whose vector of relative singular values has maximum entropy. This network of maximum entropy is the most complex, or random, given the specified structure. We found that the heuristic maximum entropy model based on connectance did not predict the structure of sampled food webs very well. However, the heuristic maximum entropy model based on the entire joint degree sequence, i.e. on the

number of prey and predators for each species, gave more convincing results. Indeed, this model reproduced food-web structure surprisingly well, including the highly informative motif profile. Nevertheless, it was not able to predict realistic food-chain lengths.

Our results bring to the forefront the role of the joint degree distribution in shaping food-web structure. This echoes the work of Fortuna et al., 2010, who found that the degree distribution of ecological networks drives their emerging structure such as their nestedness and modularity. Several measures of food webs have been analyzed when studying the ecological consequences of network structure (McCann, 2011, Delmas et al., 2019). In fact, following Williams, 2011, we believe that there is a lot more ecological information in the deviation between these properties in empirical systems and their maximum entropy configuration given a fixed joint degree sequence.

#### 2.4.1. Alternative MaxEnt models

In this contribution, we used a method based on simulated annealing to find the network configuration with the highest SVD entropy while fixing some aspects of its structure. However, there are different ways to generate adjacency matrices using MaxEnt. Another technique, also based on simulated annealing, could begin by generating a food web randomly with fixed numbers of species and interactions and calculating its joint degree distribution. Pairs of interactions could then be swapped sequentially until we minimize the divergence between the calculated joint degree distribution and the one of maximum entropy obtained analytically. In that case, this is the entropy of the joint degree distribution that would be maximized, not the one of the network's topology. To a certain extent, this method would bridge the gap between the analytical and heuristic approaches presented in this article. More research is needed to compare the quality of different methods in generating adjacency matrices of food webs using MaxEnt.

Maximum entropy graph models are another type of method that predicts a distribution of adjacency matrices under soft or hard constraints (e.g. Park and Newman, 2004, Cimini et al., 2019). Under hard constraints, every network with a non-zero probability exactly

satisfies the constraints on its structure. This is in contrast with soft constraints, which require that networks satisfy them on average (i.e. many networks with a non-zero probability do not have the exact structure set by the constraints). Maximum entropy graph models are helpful because they can provide probability distributions for many network properties by measuring the structure of all adjacency matrices with a non-zero probability. However, we consider that our approach based on simulated annealing is more flexible and more computationally efficient. Indeed, many measures of food-web structure are hard to translate into mathematical constraints. Moreover, because food webs are directed networks that can have self-loops, it makes the mathematical derivation of maximum entropy graph models difficult. We believe that identifying heuristically what constrains the topology of food webs is a useful first step before attempting to derive the mathematical formulation of a maximum entropy graph model for food webs.

#### 2.4.2. Applications

Our analytical and heuristic models can be applied for different purposes. First, they could be used to generate first-order approximations of a network's properties when state variables are known empirically. For example, knowing the number of species in an ecological community, we can predict its number of interactions using the flexible links model and then predict its joint degree distribution with minimal biases using the principle of maximum entropy. This could prove particularly useful when predicting network structure at large spatial scales, subdividing the study area into smaller communities (e.g. grid cells). Indeed, because species richness and other ecological data are increasingly abundant (e.g. Dickinson et al., 2010), validated MaxEnt models can be used to respond to a wider range of macroecological questions regarding food webs.

Second, our analytical model can be used to generate informative priors in Bayesian analyses of the structure of ecological networks (e.g. Cirtwill et al., 2019). Indeed, the probability distribution of maximum entropy derived using MaxEnt can be used as a prior that can be updated with novel data. For instance, if we know the number of species and

the number of interactions, we can derive the degree distribution of maximum entropy, as shown in this contribution. The degree distribution represents the probability that a species can interact (as a predator or a prey) with a given number of other species. Data on species interactions can be used to update the prior degree distribution to generate a more accurate posterior distribution, thus improving our description and understanding of the system.

Third, our analytical and heuristic models can be used to make better predictions of pairwise species interactions by constraining the space of feasible networks, as discussed in Strydom, Catchen, et al., 2021. In other words, we can use the network configuration or specific measures of food-web structure derived using MaxEnt to ensure that our predictions of interspecific interactions form feasible networks. This means that the probability that two species interact may be conditional on the network structure and the probability of interactions of all other species pairs. When data are limited, MaxEnt can be used to predict network structure on which pairwise probabilities of interactions are conditional.

Finally, our analytical and heuristic models can be used as alternative null models of ecological networks to better understand and identify the ecological processes driving food-web structure. Indeed, these mechanisms can be better described when analyzing the deviation of empirical data from MaxEnt predictions (Caruso et al., 2022). A strong deviation would indicate that ecological mechanisms not captured by the statistical constraints are at play for the system at hand. For instance, the incapacity of a MaxEnt model to reliably predict entropy may be a compelling indication of additional constraints shaping food-web structure. If deviations are systematic, the maximum entropy model might need to be revised to include appropriate ecological constraints. This revision process helps us reflect on and identify what constrains food-web structure. However, it is important to note that tangible ecological mechanisms cannot be directly inferred from statistical distributions (Warren II et al., 2022). Instead, by identifying the constraints of a system and by analyzing empirical deviations from maximum entropy predictions, MaxEnt can only help us redirect research efforts toward understanding the biological mechanisms behind these constraints.

The principle of maximum entropy can thus be applied to both the prediction and understanding of natural systems. The model’s interpretation depends on how we use it. It can be used as a baseline distribution to identify the ecological processes organizing natural systems. It can also be used to generate predictions of ecological networks. This distinction between understanding and predicting is important when using and interpreting MaxEnt models.

### 2.4.3. Final remarks

One of the biggest challenges in using the principle of maximum entropy is to identify the set of state variables that best reproduce empirical data. We found that the number of species and the number of interactions are important state variables for the prediction of the joint degree distribution. Similarly, we found that the numbers of prey and predators for each species in a food web are important state variables for the prediction of the network configuration. However, our predictions overestimated the symmetry of the joint degree distribution for our analytical model and the maximum trophic level and network diameter for our heuristic model. We should thus continue to play the ecological detective to find these other topological constraints that would improve the predictions of MaxEnt models and help us understand better what drives food-web structure.

## 2.5. Acknowledgments

We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was supported by the Institute for Data Valorization (IVADO, PhD-2019a-5993304626), which supported FB, and the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the Computational Biodiversity Science and Services (BIOS2) program, which supported FB, DG, and TP. TP received funding from the

Canadian Institute for Ecology and Evolution and a donation from the Courtois Foundation. TP also acknowledges financial support from NSERC through the Discovery Grants and Discovery Accelerator Supplement programs.

## **Conclusions générales**

---

### **2.6. Développement de la théorie de l'entropie maximale des réseaux trophiques**

#### **2.6.1. Où en sommes-nous?**

#### **2.6.2. Comment améliorer et étendre la théorie?**

### **2.7. Validation de la théorie**

#### **2.7.1. Quels réseaux sont d'entropie maximale?**

#### **2.7.2. Un réseau d'entropie maximale est-il à l'équilibre?**

### **2.8. Applications potentielles de la théorie**

#### **2.8.1. Compréhension des mécanismes sous-jacents aux réseaux d'interactions**

#### **2.8.2. Prévision de la structure des réseaux trophiques**



## Références

---

- Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N. M., Wang, J., Yu, D. W., & Ovaskainen, O. (2021). Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change. *Ecography*, 44(6), 885–896. <https://doi.org/10.1111/ecog.05547>
- Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47–97. <https://doi.org/10.1103/RevModPhys.74.47>
- Angilletta, M. J., Jr., Steury, T. D., & Sears, M. W. (2004). Temperature, Growth Rate, and Body Size in Ectotherms: Fitting Pieces of a Life-History Puzzle1. *Integrative and Comparative Biology*, 44(6), 498–509. <https://doi.org/10.1093/icb/44.6.498>
- Araujo, M. B., Rozenfeld, A., Rahbek, C., & Marquet, P. A. (2011). Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, 34(6), 897–908. <https://doi.org/10.1111/j.1600-0587.2011.06919.x>
- Banville, F., Vissault, S., & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6(61), 2721. <https://doi.org/10.21105/joss.02721>
- Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, 30(12), 1894–1903. <https://doi.org/10.1111/1365-2435.12666>
- Bascompte, J., Jordano, P., Melian, C. J., & Olesen, J. M. (2003). The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences*

*of the United States of America*, 100(16), 9383–9387. <https://doi.org/10.1073/pnas.1633576100>

- Bascompte, J., & Jordano, P. (2013). *Mutualistic Networks*. Princeton University Press.
- Bastolla, U., Fortuna, M. A., Pascual-Garcia, A., Ferrera, A., Luque, B., & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241), 1018–U91. <https://doi.org/10.1038/nature07950>
- Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., & Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874, 1–92. <https://doi.org/10.1016/j.physrep.2020.05.004>
- Beardsell, A., Gravel, D., Berteaux, D., Gauthier, G., Clermont, J., Careau, V., Lecomte, N., Juhasz, C.-C., Royer-Boutin, P., & Béty, J. (2021). Derivation of Predator Functional Responses Using a Mechanistic Approach in a Natural System. *Frontiers in Ecology and Evolution*, 9. <https://doi.org/10.3389/fevo.2021.630944>
- Beck, C. (2009). Generalised information and entropy measures in physics. *Contemporary Physics*, 50(4), 495–510. <https://doi.org/10.1080/00107510902823517>
- Benke, A. C., Wallace, J. B., Harrison, J. W., & Koebel, J. W. (2001). Food web quantification using secondary production analysis: Predaceous invertebrates of the snag habitat in a subtropical river. *Freshwater Biology*, 46(3), 329–346. <https://doi.org/10.1046/j.1365-2427.2001.00680.x>
- Berg, M. P., & Ellers, J. (2010). Trait plasticity in species interactions: A driving force of community dynamics. *Evolutionary Ecology*, 24(3), 617–629. <https://doi.org/10.1007/s10682-009-9347-8>
- Berlow, E. L., Neutel, A.-M., Cohen, J. E., De Ruiter, P. C., Ebenman, B., Emmerson, M., Fox, J. W., Jansen, V. A. A., Iwan Jones, J., Kokkoris, G. D., Logofet, D. O., McKane, A. J., Montoya, J. M., & Petchey, O. (2004). Interaction strengths in food webs: Issues and opportunities. *Journal of Animal Ecology*, 73(3), 585–598. <https://doi.org/10.1111/j.0021-8790.2004.00833.x>

- Bersier, L.-F., Banašek-Richter, C., & Cattin, M.-F. (2002). Quantitative Descriptors of Food-Web Matrices. *Ecology*, 83(9), 2394–2407. [https://doi.org/10.1890/0012-9658\(2002\)083\[2394:QDOFWM\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2394:QDOFWM]2.0.CO;2)
- Blüthgen, N. (2010). Why network analysis is often disconnected from community ecology: A critique and an ecologist's guide. *Basic and Applied Ecology*, 11(3), 185–195. <https://doi.org/10.1016/j.baae.2010.01.001>
- Bolnick, D. I., Amarasekare, P., Araújo, M. S., Bürger, R., Levine, J. M., Novak, M., Rudolf, V. H. W., Schreiber, S. J., Urban, M. C., & Vasseur, D. A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26(4), 183–192. <https://doi.org/10.1016/j.tree.2011.01.009>
- Bonabeau, E., Dorigo, M., & Theraulaz, a. G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.
- Borrett, S. R., & Scharler, U. M. (2019). Walk partitions of flow in Ecological Network Analysis: Review and synthesis of methods and indicators. *Ecological Indicators*, 106, 105451. <https://doi.org/10.1016/j.ecolind.2019.105451>
- Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, 15(6), 584–593. <https://doi.org/10.1111/j.1461-0248.2012.01772.x>
- Brimacombe, C., Bodner, K., & Fortin, M.-J. (2024). Applying a method before its proof of concept: A cautionary tale using inferred food webs. *Global Change Biology*, 30(6), e17360. <https://doi.org/10.1111/gcb.17360>
- Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T., & Fortin, M.-J. (2023). Shortcomings of reusing species interaction networks created by different sets of researchers. *PLOS Biology*, 21(4), e3002068. <https://doi.org/10.1371/journal.pbio.3002068>
- Broom, M., & Ruxton, G. D. (2005). You can run—or you can hide: Optimal strategies for cryptic prey against pursuit predators. *Behavioral Ecology*, 16(3), 534–540. <https://doi.org/10.1093/beheco/ari024>

- Brose, U., Ostling, A., Harrison, K., & Martinez, N. D. (2004). Unified spatial scaling of species and their trophic interactions. *Nature*, 428(6979), 167–171. <https://doi.org/10.1038/nature02297>
- Brummer, A. B., & Newman, E. A. (2019). Derivations of the Core Functions of the Maximum Entropy Theory of Ecology. *Entropy*, 21(7), 712. <https://doi.org/10.3390/e21070712>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Canard, E., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D., & Gravel, D. (2014). Empirical Evaluation of Neutral Interactions in Host-Parasite Networks. *The American Naturalist*, 183(4), 468–479. <https://doi.org/10.1086/675363>
- Canard, E., Mouquet, N., Marescot, L., Gaston, K. J., Gravel, D., & Mouillot, D. (2012). Emergence of Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7(8), e38295. <https://doi.org/10.1371/journal.pone.0038295>
- Carlson, C. J., Farrell, M. J., Grange, Z., Han, B. A., Mollentze, N., Phelan, A. L., Rasmussen, A. L., Albery, G. F., Bett, B., Brett-Major, D. M., Cohen, L. E., Dallas, T., Eskew, E. A., Fagre, A. C., Forbes, K. M., Gibb, R., Halabi, S., Hammer, C. C., Katz, R., ... Webala, P. W. (2021). The future of zoonotic risk prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1837), 20200358. <https://doi.org/10.1098/rstb.2020.0358>
- Caron, D., Brose, U., Lurgi, M., Blanchet, F. G., Gravel, D., & Pollock, L. J. (2024). Trait-matching models predict pairwise interactions across regions, not food web properties. *Global Ecology and Biogeography*, 33(4), e13807. <https://doi.org/10.1111/geb.13807>
- Carpenter, S. R., Kitchell, J. F., Hodgson, J. R., Cochran, P. A., Elser, J. J., Elser, M. M., Lodge, D. M., Kretchmer, D., He, X., & von Ende, C. N. (1987). Regulation of Lake Primary Productivity by Food Web Structure. *Ecology*, 68(6), 1863–1876. <https://doi.org/10.2307/1939878>

- Caruso, T., Clemente, G. V., Rillig, M. C., & Garlaschelli, D. (2022). Fluctuating ecological networks: A synthesis of maximum-entropy approaches for pattern detection and process inference. *Methods in Ecology and Evolution*, 13(11), 2306–2317. <https://doi.org/10.1111/2041-210X.13985>
- Castoe, T. A., Spencer, C. L., & Parkinson, C. L. (2007). Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): A perspective on North American desert biogeography. *Molecular Phylogenetics and Evolution*, 42(1), 193–212. <https://doi.org/10.1016/j.ympev.2006.07.002>
- Catchen, M. D., Poisot, T., Pollock, L. J., & Gonzalez, A. (2023). The missing link: Discerning true from false negatives when sampling species interaction networks.
- Cazelles, K., Araujo, M. B., Mouquet, N., & Gravel, D. (2016). A theory for species co-occurrence in interaction networks. *Theoretical Ecology*, 9(1), 39–48. <https://doi.org/10.1007/s12080-015-0281-9>
- Chacoff, N. P., Vázquez, D. P., Lomáscolo, S. B., Stevani, E. L., Dorado, J., & Padrón, B. (2012). Evaluating sampling completeness in a desert plant–pollinator network. *Journal of Animal Ecology*, 81(1), 190–200. <https://doi.org/10.1111/j.1365-2656.2011.01883.x>
- Chagnon, P.-L. (2015). Characterizing topology of ecological networks along gradients: The limits of metrics' standardization. *Ecological Complexity*, 22, 36–39. <https://doi.org/10.1016/j.ecocom.2015.01.004>
- Cherif, M., Brose, U., Hirt, M. R., Ryser, R., Silve, V., Albert, G., Arnott, R., Berti, E., Cirtwill, A., Dyer, A., Gauzens, B., Gupta, A., Ho, H.-C., Portalier, S. M. J., Wain, D., & Woottton, K. (2024). The environment to the rescue: Can physics help predict predator-prey interactions? *BIOLOGICAL REVIEWS*. <https://doi.org/10.1111/brv.13105>
- Choh, Y., Ignacio, M., Sabelis, M. W., & Janssen, A. (2012). Predator-prey role reversals, juvenile experience and adult antipredator behaviour. *Scientific Reports*, 2(1), 728. <https://doi.org/10.1038/srep00728>

- Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., & Caldarelli, G. (2019). The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1), 58–71. <https://doi.org/10.1038/s42254-018-0002-6>
- Cirtwill, A. R., Eklof, A., Roslin, T., Wootton, K., & Gravel, D. (2019). A quantitative framework for investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, 10(6), 902–911. <https://doi.org/10.1111/2041-210X.13180>
- Clark, J. S., Bell, D. M., Hersh, M. H., Kwit, M. C., Moran, E., Salk, C., Stine, A., Valle, D., & Zhu, K. (2011). Individual-scale variation, species-scale differences: Inference needed to understand diversity. *Ecology Letters*, 14(12), 1273–1287. <https://doi.org/10.1111/j.1461-0248.2011.01685.x>
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological Forecasts: An Emerging Imperative. *Science*, 293(5530), 657–660. <https://doi.org/10.1126/science.293.5530.657>
- Cohen, J. E. (1978). *Food Webs and Niche Space. (MPB-11), Volume 11*. Princeton University Press. <https://doi.org/10.2307/j.ctvx5wc04>
- Cohen, J. E., Jonsson, T., & Carpenter, S. R. (2003). Ecological community description using the food web, species abundance, and body size. *Proceedings of the National Academy of Sciences*, 100(4), 1781–1786. <https://doi.org/10.1073/pnas.232715699>
- Dallas, T., Park, A. W., & Drake, J. M. (2017). Predicting cryptic links in host-parasite networks. *PLOS Computational Biology*, 13(5), e1005557. <https://doi.org/10.1371/journal.pcbi.1005557>
- Dansereau, G., Banville, F., Basque, E., MacDonald, A., & Poisot, T. (2020). [Re] Chaos in a Three-Species Food Chain. *ReScience C*, 6(3), #5. <https://doi.org/10.5281/zenodo.4022518>

- Dansereau, G., Barros, C., & Poisot, T. (2024). Spatially explicit predictions of food web structure from regional-level data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379(1909), 20230166. <https://doi.org/10.1098/rstb.2023.0166>
- de Aguiar, M. A. M., Newman, E. A., Pires, M. M., Yeakel, J. D., Boettiger, C., Burkle, L. A., Gravel, D., Jr, P. R. G., O'Donnell, J. L., Poisot, T., Fortin, M.-J., & Hembry, D. H. (2019). Revealing biases in the sampling of ecological interaction networks. *PeerJ*, 7, e7566. <https://doi.org/10.7717/peerj.7566>
- Delmas, E., Besson, M., Brice, M.-H., Burkle, L. A., Dalla Riva, G. V., Fortin, M.-J., Gravel, D., Guimarães Jr., P. R., Hembry, D. H., Newman, E. A., Olesen, J. M., Pires, M. M., Yeakel, J. D., & Poisot, T. (2019). Analysing ecological networks of species interactions. *Biological Reviews*, 94(1), 16–36. <https://doi.org/10.1111/brv.12433>
- Dewar, R. C., & Porté, A. (2008). Statistical mechanics unifies different ecological patterns. *Journal of Theoretical Biology*, 251(3), 389–403. <https://doi.org/10.1016/j.jtbi.2007.12.007>
- Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen Science as an Ecological Research Tool: Challenges and Benefits. In D. J. Futuyma, H. B. Shafer, & D. Simberloff (Eds.), *Annual Review of Ecology, Evolution, and Systematics*, Vol 41 (pp. 149–172, Vol. 41). Annual Reviews. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
- Donahue, J. G., Piesman, J., & Spielman, A. (1987). Reservoir competence of white-footed mice for Lyme disease spirochetes. *The American journal of tropical medicine and hygiene*, 36(1), 92–96. <https://doi.org/10.4269/ajtmh.1987.36.92>
- Dunne, J. A. (2006). The Network Structure of Food Webs. In J. A. Dunne & M. Pasqual (Eds.), *Ecological networks: Linking structure and dynamics* (pp. 27–86). Oxford University Press.
- Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, 5(4), 558–567. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>

- Dunning, I., Huchette, J., & Lubin, M. (2017). JuMP: A Modeling Language for Mathematical Optimization. *SIAM Review*, 59(2), 295–320. <https://doi.org/10.1137/15M1020575>
- Eklöf, A., & Stouffer, D. B. (2016). The phylogenetic component of food web structure and intervality. *Theoretical Ecology*, 9(1), 107–115. <https://doi.org/10.1007/s12080-015-0273-9>
- Elkinton, J. S., Healy, W. M., Buonaccorsi, J. P., Boettner, G. H., Hazzard, A. M., & Smith, H. R. (1996). Interactions Among Gypsy Moths, White-footed Mice, and Acorns. *Ecology*, 77(8), 2332–2342. <https://doi.org/10.2307/2265735>
- Elmasri, M., Farrell, M. J., Davies, T. J., & Stephens, D. A. (2020). A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships. *The Annals of Applied Statistics*, 14(1), 221–240. <https://doi.org/10.1214/19-AOAS1296>
- Elton, C. S. (1927). *Animal Ecology*. University of Chicago Press.
- Elton, C. S. (1958). *The Ecology of Invasions by Animals and Plants*. Springer Nature.
- Elton, C. S. (2001). *Animal Ecology* (M. A. Leibold & T. J. Wootton, Eds.). University of Chicago Press.
- Emmerson, M. C., & Raffaelli, D. (2004). Predator-Prey Body Size, Interaction Strength and the Stability of a Real Food Web. *Journal of Animal Ecology*, 73(3), 399–409.
- Evans, D. M., Pocock, M. J. O., & Memmott, J. (2013). The robustness of a network of ecological networks to habitat loss. *Ecology Letters*, 16(7), 844–852. <https://doi.org/10.1111/ele.12117>
- Evans, M. R., Norris, K. J., & Benton, T. G. (2012). Predictive ecology: Systems approaches *Introduction. PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 367(1586), 163–169. <https://doi.org/10.1098/rstb.2011.0191>
- Fedorov, V. B., Goropashnaya, A. V., Boeskorov, G. G., & Cook, J. A. (2008). Comparative phylogeography and demographic history of the wood lemming (*Myopus schisticolor*):

- Implications for late Quaternary history of the taiga species in Eurasia. *Molecular Ecology*, 17(2), 598–610. <https://doi.org/10.1111/j.1365-294X.2007.03595.x>
- Foote, R. (2007). Mathematics and Complex Systems. *Science*, 318(5849), 410–412. <https://doi.org/10.1126/science.1141754>
- Fortuna, M. A., & Bascompte, J. (2006). Habitat loss and the structure of plant-animal mutualistic networks. *Ecology Letters*, 9(3), 278–283. <https://doi.org/10.1111/j.1461-0248.2005.00868.x>
- Fortuna, M. A., Stouffer, D. B., Olesen, J. M., Jordano, P., Mouillot, D., Krasnov, B. R., Poulin, R., & Bascompte, J. (2010). Nestedness versus modularity in ecological networks: Two sides of the same coin? *Journal of Animal Ecology*, 79(4), 811–817. <https://doi.org/10.1111/j.1365-2656.2010.01688.x>
- Frank, S. A. (2009). The common patterns of nature. *Journal of Evolutionary Biology*, 22(8), 1563–1585. <https://doi.org/10.1111/j.1420-9101.2009.01775.x>
- Frank, S. A., & Smith, E. (2011). A simple derivation and classification of common probability distributions based on information symmetry and measurement scale. *Journal of Evolutionary Biology*, 24(3), 469–484. <https://doi.org/10.1111/j.1420-9101.2010.02204.x>
- Fu, X., Seo, E., Clarke, J., & Hutchinson, R. A. (2021). Link Prediction Under Imperfect Detection: Collaborative Filtering for Ecological Networks. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3117–3128. <https://doi.org/10.1109/TKDE.2019.2962031>
- Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya, J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5), 782–790. <https://doi.org/10.1038/s41559-018-0517-3>
- Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W., & Holt, R. D. (2010). A framework for community interactions under climate change. *Trends in Ecology & Evolution*, 25(6), 325–331. <https://doi.org/10.1016/j.tree.2010.03.002>

- Godsoe, W., Murray, R., & Iritani, R. (2022). Species interactions and diversity: A unified framework using Hill numbers. *Oikos*, n/a(n/a), e09282. <https://doi.org/10.1111/oik.09282>
- Golub, G. H., Hoffman, A., & Stewart, G. W. (1987). A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra and its Applications*, 88–89, 317–327. [https://doi.org/10.1016/0024-3795\(87\)90114-5](https://doi.org/10.1016/0024-3795(87)90114-5)
- Golubski, A. J., & Abrams, P. A. (2011). Modifying modifiers: What happens when inter-specific interactions interact? *Journal of Animal Ecology*, 80(5), 1097–1108. <https://doi.org/10.1111/j.1365-2656.2011.01852.x>
- Gómez, J. M., Perfectti, F., & Jordano, P. (2011). The Functional Consequences of Mutualistic Network Architecture. *PLOS ONE*, 6(1), e16143. <https://doi.org/10.1371/journal.pone.0016143>
- Gonzalez, A., Germain, R. M., Srivastava, D. S., Filotas, E., Dee, L. E., Gravel, D., Thompson, P. L., Isbell, F., Wang, S., Kéfi, S., Montoya, J., Zelnik, Y. R., & Loreau, M. (2020). Scaling-up biodiversity-ecosystem functioning research. *Ecology Letters*, 23(4), 757–776. <https://doi.org/10.1111/ele.13456>
- Gonzalez, A., & Londoño, M. C. (2022). Monitor biodiversity for action. *Science*, 378(6625), 1147–1147. <https://doi.org/10.1126/science.adg1506>
- Gonzalez-Varo, J. P., & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends in Ecology & Evolution*, 31(9), 700–710. <https://doi.org/10.1016/j.tree.2016.06.009>
- Grabowski, J. H. (2004). Habitat Complexity Disrupts Predator–Prey Interactions but Not the Trophic Cascade on Oyster Reefs. *Ecology*, 85(4), 995–1004. <https://doi.org/10.1890/03-0067>
- Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B., Tylianakis, J. M., Wood, S. A., & Roslin, T. (2019). Bringing Elton and Grinnell together: A quantitative framework to represent the biogeography of

- ecological interaction networks. *Ecography*, 42(3), 401–415. <https://doi.org/10.1111/ecog.04006>
- Gravel, D., Poisot, T., Albouy, C., Velez, L., & Mouillot, D. (2013). Inferring food web structure from predator-prey body size relationships. *Methods in Ecology and Evolution*, 4(11), 1083–1090. <https://doi.org/10.1111/2041-210X.12103>
- Guimarães, P. R. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review of Ecology, Evolution, and Systematics*, 51(Volume 51, 2020), 433–460. <https://doi.org/10.1146/annurev-ecolsys-012220-120819>
- Guimerà, R., & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106(52), 22073–22078. <https://doi.org/10.1073/pnas.0908366106>
- Harremoës, P., & Topsøe, F. (2001). Maximum Entropy Fundamentals. *Entropy*, 3(3), 191–226. <https://doi.org/10.3390/e3030191>
- Harte, J., & Newman, E. A. (2014). Maximum information entropy: A foundation for ecological theory. *Trends in Ecology & Evolution*, 29(7), 384–389. <https://doi.org/10.1016/j.tree.2014.04.009>
- Harte, J., Zillio, T., Conlisk, E., & Smith, A. B. (2008). Maximum Entropy and the State-Variable Approach to Macroecology. *Ecology*, 89(10), 2700–2711. <https://doi.org/10.1890/07-1369.1>
- Harte, J. (2011). *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. OUP Oxford.
- Hastings, A., & Powell, T. (1991). Chaos in a Three-Species Food Chain. *Ecology*, 72(3), 896–903. <https://doi.org/10.2307/1940591>
- Hemprich-Bennett, D. R., Oliveira, H. F. M., Le Comber, S. C., Rossiter, S. J., & Clare, E. L. (2021). Assessing the impact of taxon resolution on network structure. *Ecology*, 102(3), e03256. <https://doi.org/10.1002/ecy.3256>

- Herrera, C. M. (1989). Pollinator abundance, morphology, and flower visitation rate: Analysis of the “quantity” component in a plant-pollinator system. *Oecologia*, 80(2), 241–248. <https://doi.org/10.1007/BF00380158>
- Higino, G. T., Banville, F., Dansereau, G., Muñoz, N. R. F., Windsor, F., & Poisot, T. (2023). Mismatch between IUCN range maps and species interactions data illustrated using the Serengeti food web. *PeerJ*, 11, e14620. <https://doi.org/10.7717/peerj.14620>
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Ims, R. A., Henden, J.-A., Thingnes, A. V., & Killengreen, S. T. (2013). Indirect food web interactions mediated by predator–rodent dynamics: Relative roles of lemmings and voles. *Biology Letters*, 9(6), 20130802. <https://doi.org/10.1098/rsbl.2013.0802>
- Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., Lauridsen, R. B., Ledger, M. E., Lewis, H. M., Olesen, J. M., Van Veen, F. F., Warren, P. H., & Woodward, G. (2009). Review: Ecological networks – beyond food webs. *Journal of Animal Ecology*, 78(1), 253–269. <https://doi.org/10.1111/j.1365-2656.2008.01460.x>
- Jaynes, E. T. (1957a). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620–630. <https://doi.org/10.1103/PhysRev.106.620>
- Jaynes, E. T. (1957b). Information Theory and Statistical Mechanics. II. *Physical Review*, 108(2), 171–190. <https://doi.org/10.1103/PhysRev.108.171>
- Jordano, P. (1987). Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance, Dependence Asymmetries, and Coevolution. *The American Naturalist*, 129(5), 657–677. <https://doi.org/10.1086/284665>
- Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30(12), 1883–1893. <https://doi.org/10.1111/1365-2435.12763>

- Jordano, P., Bascompte, J., & Olesen, J. M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters*, 6(1), 69–81. <https://doi.org/10.1046/j.1461-0248.2003.00403.x>
- Kausrud, K. L., Mysterud, A., Steen, H., Vik, J. O., Østbye, E., Cazelles, B., Framstad, E., Eikeset, A. M., Mysterud, I., Solhøy, T., & Stenseth, N. C. (2008). Linking climate change to lemming cycles. *Nature*, 456(7218), 93–97. <https://doi.org/10.1038/nature07442>
- Kéfi, S., Berlow, E. L., Wieters, E. A., Joppa, L. N., Wood, S. A., Brose, U., & Navarrete, S. A. (2015). Network structure beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, 96(1), 291–303. <https://doi.org/10.1890/13-1424.1>
- Kéfi, S., Berlow, E. L., Wieters, E. A., Navarrete, S. A., Petchey, O. L., Wood, S. A., Boit, A., Joppa, L. N., Lafferty, K. D., Williams, R. J., Martinez, N. D., Menge, B. A., Blanchette, C. A., Iles, A. C., & Brose, U. (2012). More than a meal... integrating non-feeding interactions into food webs. *Ecology Letters*, 15(4), 291–300. <https://doi.org/10.1111/j.1461-0248.2011.01732.x>
- Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., & Berlow, E. L. (2016). How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience. *PLOS Biology*, 14(8), e1002527. <https://doi.org/10.1371/journal.pbio.1002527>
- Khinchin, A. Y. (2013). *Mathematical Foundations of Information Theory*. Courier Corporation.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671–680. <https://doi.org/10.1126/science.220.4598.671>
- Kissner, K. J., Forbes, M. R., & Secoy, D. M. (1997). Rattling Behavior of Prairie Rattlesnakes (*Crotalus viridis viridis*, Viperidae) in Relation to Sex, Reproductive Status, Body Size, and Body Temperature. *Ethology*, 103(12), 1042–1050. <https://doi.org/10.1111/j.1439-0310.1997.tb00146.x>

- Klecka, J., & Boukal, D. S. (2014). The effect of habitat structure on prey mortality depends on predator and prey microhabitat use. *Oecologia*, 176(1), 183–191. <https://doi.org/10.1007/s00442-014-3007-6>
- Kokkoris, G. D., Jansen, V. A. A., Loreau, M., & Troumbis, A. Y. (2002). Variability in Interaction Strength and Implications for Biodiversity. *Journal of Animal Ecology*, 71(2), 362–371.
- Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S., & Roslin, T. (2017). Food-web structure of willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98(6), 1730–1730. <https://doi.org/10.1002/ecy.1832>
- Kulkarni, M. A., Kweka, E., Nyale, E., Lyatuu, E., Mosha, F. W., Chandramohan, D., Rau, M. E., & Drakeley, C. (2006). Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District, Northeastern Tanzania. *Journal of Medical Entomology*, 43(3), 580–588. <https://doi.org/10.1093/jmedent/43.3.580>
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <https://doi.org/10.1214/aoms/1177729694>
- Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, 3(1), 33–67. <https://doi.org/10.1007/s13194-012-0056-8>
- Landi, P., Minoarivelo, H. O., Brännström, Å., Hui, C., & Dieckmann, U. (2018). Complexity and stability of ecological networks: A review of the theory. *Population Ecology*, 60(4), 319–345. <https://doi.org/10.1007/s10144-018-0628-3>
- Lane, R. S., Piesman, J., & Burgdorfer, W. (1991). Lyme Borreliosis: Relation of Its Causative Agent to Its Vectors and Hosts in North America and Europe. *Annual Review of Entomology*, 36(Volume 36, 1991), 587–609. <https://doi.org/10.1146/annurev.en.36.010191.003103>
- Lawlor, J., Banville, F., Forero-Muñoz, N.-R., Hébert, K., Martínez-Lanfranco, J. A., Rogy, P., & MacDonald, A. A. M. (2022). Ten simple rules for teaching yourself R. *PLOS*

*Computational Biology*, 18(9), e1010372. <https://doi.org/10.1371/journal.pcbi.1010372>

Loreau, M. (2010). *From Populations to Ecosystems: Theoretical Foundations for a New Ecological Synthesis (MPB-46)*. Princeton University Press. <https://doi.org/10.1515/9781400834167>

Lotka, A. J. (1925). *Elements of Physical Biology*. Williams & Wilkins.

Luna, R. S., Duarte, A., & Weckerly, F. W. (2013). Influence of Body Size on Dietary Nutrition of White-Tailed Deer *Odocoileus virginianus*. *JOURNAL OF FISH AND WILDLIFE MANAGEMENT*, 4(1), 53–62. <https://doi.org/10.3996/092012-JFWM-085>

MacDonald, A. A. M., Banville, F., & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in Food Webs. *Patterns*, 0(0). <https://doi.org/10.1016/j.patter.2020.100079>

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

Martinez, N. D. (1992). Constant Connectance in Community Food Webs. *The American Naturalist*, 139(6), 1208–1218. <https://doi.org/10.1086/285382>

Martyushev, L. M., & Seleznev, V. D. (2006). Maximum entropy production principle in physics, chemistry and biology. *Physics Reports-Review Section of Physics Letters*, 426(1), 1–45. <https://doi.org/10.1016/j.physrep.2005.12.001>

May, R. M. (1972). Will a Large Complex System be Stable? *Nature*, 238(5364), 413–414. <https://doi.org/10.1038/238413a0>

McCann, K. S. (2007). Protecting biostructure. *Nature*, 446(7131), 29–29. <https://doi.org/10.1038/446029a>

McCann, K. S. (2011). *Food Webs (MPB-50)*. Princeton University Press. <https://doi.org/10.1515/9781400840687>

- McGill, B. J., & Nekola, J. C. (2010). Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. *Oikos*, 119(4), 591–603. <https://doi.org/10.1111/j.1600-0706.2009.17771.x>
- McGlinn, D. J., Xiao, X., Kitzes, J., & White, E. P. (2015). Exploring the spatially explicit predictions of the Maximum Entropy Theory of Ecology. *Global Ecology and Biogeography*, 24(6), 675–684. <https://doi.org/10.1111/geb.12295>
- McLeod, A., Leroux, S. J., Gravel, D., Chu, C., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Poisot, T., & Wood, S. A. (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*, n/a(n/a). <https://doi.org/10.1111/oik.08650>
- McShea, W. J. (2000). The Influence of Acorn Crops on Annual Variation in Rodent and Bird Populations. *Ecology*, 81(1), 228–238. [https://doi.org/10.1890/0012-9658\(2000\)081\[0228:TIOACO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[0228:TIOACO]2.0.CO;2)
- McShea, W. J., & Schwede, G. (1993). Variable Acorn Crops: Responses of White-Tailed Deer and other Mast Consumers. *Journal of Mammalogy*, 74(4), 999–1006. <https://doi.org/10.2307/1382439>
- Melián, C. J., Vilas, C., Baldó, F., González-Ortegón, E., Drake, P., & Williams, R. J. (2011). Eco-evolutionary Dynamics of Individual-Based Food Webs. In A. Belgrano (Ed.), *Advances in Ecological Research* (pp. 225–268, Vol. 45). Academic Press. <https://doi.org/10.1016/B978-0-12-386475-8.00006-X>
- Memmott, J., Waser, N. M., & Price, M. V. (2004). Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1557), 2605–2611. <https://doi.org/10.1098/rspb.2004.2909>
- 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>
- Montoya, J. M., Rodríguez, M. A., & Hawkins, B. A. (2003). Food web complexity and higher-level ecosystem services. *Ecology Letters*, 6(7), 587–593. <https://doi.org/10.1046/j.1461-0248.2003.00469.x>

- Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution*, 30(6), 347–356. <https://doi.org/10.1016/j.tree.2015.03.014>
- Newman, E. A., Wilber, M. Q., Kopper, K. E., Moritz, M. A., Falk, D. A., McKenzie, D., & Harte, J. (2020). Disturbance macroecology: A comparative study of community structure metrics in a high-severity disturbance regime. *Ecosphere*, 11(1), e03022. <https://doi.org/10.1002/ecs2.3022>
- Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2), 167–256. <https://doi.org/10.1137/S003614450342480>
- Nielsen, S. N., & Muller, F. (2000). Emergent properties of ecosystems. *Handbook of ecosystem theories and management*, 195–216.
- Ogden, N. H., Lindsay, L. R., Morshed, M., Sockett, P. N., & Artsob, H. (2009). The emergence of Lyme disease in Canada. *CANADIAN MEDICAL ASSOCIATION JOURNAL*, 180(12), 1221–1224. <https://doi.org/10.1503/cmaj.080148>
- Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L., & Thuiller, W. (2019). Diversity indices for ecological networks: A unifying framework using Hill numbers. *Ecology Letters*, 22(4), 737–747. <https://doi.org/10.1111/ele.13221>
- Okuyama, T., & Holland, J. N. (2008). Network structural properties mediate the stability of mutualistic communities. *Ecology Letters*, 11(3), 208–216. <https://doi.org/10.1111/j.1461-0248.2007.01137.x>
- Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C., & Jordano, P. (2010). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 278(1706), 725–732. <https://doi.org/10.1098/rspb.2010.1371>
- Ostfeld, R. S., Canham, C. D., Oggeneffuss, K., Winchcombe, R. J., & Keesing, F. (2006). Climate, deer, rodents, and acorns as determinants of variation in Lyme-disease risk. *PLoS BIOLOGY*, 4(6), 1058–1068. <https://doi.org/10.1371/journal.pbio.0040145>

- Paine, R. T. (1992). Food-web analysis through field measurement of per capita interaction strength. *Nature*, 355(6355), 73–75. <https://doi.org/10.1038/355073a0>
- Park, J., & Newman, M. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6), 066117. <https://doi.org/10.1103/PhysRevE.70.066117>
- Pascual, M., & Dunne, J. A. (2006). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford University Press, USA.
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., Magliaenesi, M. A., Melián, C. J., Pitteloud, C., Roslin, T., Rohr, R., Saavedra, S., Thuiller, W., Woodward, G., Zimmermann, N. E., & Gravel, D. (2018). Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93(2), 785–800. <https://doi.org/10.1111/brv.12366>
- 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>
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Pielou, E. C. (1975). Ecological diversity.
- Pilosof, S., Porter, M. A., Pascual, M., & Kéfi, S. (2017). The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1(4), 1–9. <https://doi.org/10.1038/s41559-017-0101>
- Plagányi, É. E., & Butterworth, D. S. (2004). A critical look at the potential of Ecopath with ecosim to assist in practical fisheries management. *African Journal of Marine Science*, 26(1), 261–287. <https://doi.org/10.2989/18142320409504061>
- Poisot, T., Baiser, B., Dunne, J. A., Kéfi, S., Massol, F., Mouquet, N., Romanuk, T. N., Stouffer, D. B., Wood, S. A., & Gravel, D. (2016). Mangal - making ecological network analysis simple. *Ecography*, 39(4), 384–390. <https://doi.org/10.1111/ecog.00976>

- Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., Mercier, B., Violet, C., & Vissault, S. (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography*, 48(7), 1552–1563. <https://doi.org/10.1111/jbi.14127>
- Poisot, T., Canard, E., Mouillot, D., Mouquet, N., & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecology Letters*, 15(12), 1353–1361. <https://doi.org/10.1111/ele.12002>
- Poisot, T., Cirtwill, A. R., Cazelles, K., Gravel, D., Fortin, M.-J., & Stouffer, D. B. (2016). The structure of probabilistic networks. *Methods in Ecology and Evolution*, 7(3), 303–312. <https://doi.org/10.1111/2041-210X.12468>
- Poisot, T., & Gravel, D. (2014). When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ*, 2, e251. <https://doi.org/10.7717/peerj.251>
- Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, 124(3), 243–251. <https://doi.org/10.1111/oik.01719>
- Polis, G. A. (1991). Complex Trophic Interactions in Deserts: An Empirical Critique of Food-Web Theory. *The American Naturalist*, 138(1), 123–155.
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5(5), 397–406. <https://doi.org/10.1111/2041-210X.12180>
- Pomeranz, J. P., Thompson, R. M., Poisot, T., & Harding, J. S. (2019). Inferring predator-prey interactions in food webs. *Methods in Ecology and Evolution*, 10(3), 356–367. <https://doi.org/10.1111/2041-210X.13125>
- Pomeranz, J. P., Thompson, R. M., Poisot, T., Harding, J. S., & Pomeranz, J. P. F. (2018). Data from: Inferring predator-prey interactions in food webs. <https://doi.org/10.5061/DRYAD.K59M37F>

- Post, D. M. (2002). The long and short of food-chain length. *Trends in Ecology & Evolution*, 17(6), 269–277. [https://doi.org/10.1016/S0169-5347\(02\)02455-2](https://doi.org/10.1016/S0169-5347(02)02455-2)
- Pringle, R. M., & Hutchinson, M. C. (2020). Resolving Food-Web Structure. *Annual Review of Ecology, Evolution, and Systematics*, 51(1), 55–80. <https://doi.org/10.1146/annurev-ecolsys-110218-024908>
- Proulx, S. R., Promislow, D. E. L., & Phillips, P. C. (2005). Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, 20(6), 345–353. <https://doi.org/10.1016/j.tree.2005.04.004>
- Pulliam, H. R. (1974). On the Theory of Optimal Diets. *The American Naturalist*, 108(959), 59–74. <https://doi.org/10.1086/282885>
- Riede, J. O., Rall, B. C., Banasek-Richter, C., Navarrete, S. A., Wieters, E. A., Emmerson, M. C., Jacob, U., & Brose, U. (2010). Chapter 3 - Scaling of Food-Web Properties with Diversity and Complexity Across Ecosystems. In G. Woodward (Ed.), *Advances in Ecological Research* (pp. 139–170, Vol. 42). Academic Press. <https://doi.org/10.1016/B978-0-12-381363-3.00003-4>
- Rind, D. (1999). Complexity and Climate. *Science*, 284(5411), 105–107. <https://doi.org/10.1126/science.284.5411.105>
- Rooney, N., & McCann, K. S. (2012). Integrating food web diversity, structure and stability. *Trends in Ecology & Evolution*, 27(1), 40–46. <https://doi.org/10.1016/j.tree.2011.09.001>
- Rozdilsky, I. D., & Stone, L. (2001). Complexity can enhance stability in competitive systems. *Ecology Letters*, 4(5), 397–400. <https://doi.org/10.1046/j.1461-0248.2001.00249.x>
- Salim, J. A., Saraiva, A. M., Zermoglio, P. F., Agostini, K., Wolowski, M., Drucker, D. P., Soares, F. M., Bergamo, P. J., Varassin, I. G., Freitas, L., Maués, M. M., Rech, A. R., Veiga, A. K., Acosta, A. L., Araujo, A. C., Nogueira, A., Blochtein, B., Freitas, B. M., Albertini, B. C., ... Brito, V. L. G. (2022). Data standardization of plant–pollinator interactions. *GigaScience*, 11, giac043. <https://doi.org/10.1093/gigascience/giac043>

- Sanders, D., & van Veen, F. J. F. (2012). Indirect commensalism promotes persistence of secondary consumer species. *Biology Letters*, 8(6), 960–963. <https://doi.org/10.1098/rsbl.2012.0572>
- Saravia, L. A., Marina, T. I., Kristensen, N. P., De Troch, M., & Momo, F. R. (2022). Ecological network assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 91(3), 630–642. <https://doi.org/10.1111/1365-2656.13652>
- Schmeller, D. S., Julliard, R., Bellingham, P. J., Böhm, M., Brummitt, N., Chiarucci, A., Couvet, D., Elmendorf, S., Forsyth, D. M., Moreno, J. G., Gregory, R. D., Magnusson, W. E., Martin, L. J., McGeoch, M. A., Mihoub, J.-B., Pereira, H. M., Proença, V., van Swaay, C. A. M., Yahara, T., & Belnap, J. (2015). Towards a global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51–57. <https://doi.org/10.1016/j.jnc.2015.03.003>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- 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. <https://doi.org/10.1101/2024.01.30.578036>
- Short, H. L. (1975). Nutrition of Southern Deer in Different Seasons. *The Journal of Wildlife Management*, 39(2), 321–329. <https://doi.org/10.2307/3799909>
- Simmonds, E. G., Adjei, K. P., Cretois, B., Dickel, L., González-Gil, R., Laverick, J. H., Mandeville, C. P., Mandeville, E. G., Ovaskainen, O., Sicacha-Parada, J., Skarstein, E. S., & O'Hara, B. (2024). Recommendations for quantitative uncertainty consideration in ecology and evolution. *Trends in Ecology & Evolution*, 39(4), 328–337. <https://doi.org/10.1016/j.tree.2023.10.012>
- Simmonds, E. G., Adjei, K. P., Andersen, C. W., Hettle Aspheim, J. C., Battistin, C., Bulso, N., Christensen, H. M., Cretois, B., Cubero, R., Davidovich, I. A., Dickel, L., Dunn, B., Dunn-Sigouin, E., Dyrstad, K., Einum, S., Giglio, D., Gjerløw, H., Godefroidt, A., González-Gil, R., ... O'Hara, R. B. (2022). Insights into the quantification and

- reporting of model-related uncertainty across different disciplines. *iScience*, 25(12), 105512. <https://doi.org/10.1016/j.isci.2022.105512>
- Simon, J. A., Marrotte, R. R., Desrosiers, N., Fiset, J., Gaitan, J., Gonzalez, A., Koffi, J. K., Lapointe, F.-J., Leighton, P. A., Lindsay, L. R., Logan, T., Milord, F., Ogden, N. H., Rogic, A., Roy-Dufresne, E., Suter, D., Tessier, N., & Millien, V. (2014). Climate change and habitat fragmentation drive the occurrence of *Borrelia burgdorferi*, the agent of Lyme disease, at the northeastern limit of its distribution. *EVOLUTIONARY APPLICATIONS*, 7(7), 750–764. <https://doi.org/10.1111/eva.12165>
- Singer, M. C., & McBride, C. S. (2012). Geographic mosaics of species' association: A definition and an example driven by plant–insect phenological synchrony. *Ecology*, 93(12), 2658–2673. <https://doi.org/10.1890/11-2078.1>
- Smith-Ramesh, L. M., Moore, A. C., & Schmitz, O. J. (2017). Global synthesis suggests that food web connectance correlates to invasion resistance. *Global Change Biology*, 23(2), 465–473. <https://doi.org/10.1111/gcb.13460>
- Solé, R. V., & Montoya, M. (2001). Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1480), 2039–2045. <https://doi.org/10.1098/rspb.2001.1767>
- Sporns, O. (2011). The human connectome: A complex network. *Annals of the New York Academy of Sciences*, 1224(1), 109–125. <https://doi.org/10.1111/j.1749-6632.2010.05888.x>
- Srinivasan, U. T., Dunne, J. A., Harte, J., & Martinez, N. D. (2007). Response of Complex Food Webs to Realistic Extinction Sequences. *Ecology*, 88(3), 671–682. <https://doi.org/10.1890/06-0971>
- Staniczenko, P. P. A., Kopp, J. C., & Allesina, S. (2013). The ghost of nestedness in ecological networks. *Nature Communications*, 4(1), 1391. <https://doi.org/10.1038/ncomms2422>
- Staniczenko, P. P. A., Lewis, O. T., Jones, N. S., & Reed-Tsochas, F. (2010). Structural dynamics and robustness of food webs. *Ecology Letters*, 13(7), 891–899. <https://doi.org/10.1111/j.1461-0248.2010.01485.x>

- Staniczenko, P. P., Sivasubramaniam, P., Suttle, K. B., & Pearson, R. G. (2017). Linking macroecology and community ecology: Refining predictions of species distributions using biotic interaction networks. *Ecology Letters*, 20(6), 693–707. <https://doi.org/10.1111/ele.12770>
- Stock, M., Poisot, T., & De Baets, B. (2021). Optimal transportation theory for species interaction networks. *Ecology and Evolution*, 11(9), 3841–3855. <https://doi.org/10.1002/ece3.7254>
- Stock, M., Poisot, T., Waegeman, W., & De Baets, B. (2017). Linear filtering reveals false negatives in species interaction data. *Scientific Reports*, 7(1), 45908. <https://doi.org/10.1038/srep45908>
- Stouffer, D. B., & Bascompte, J. (2011). Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9), 3648–3652. <https://doi.org/10.1073/pnas.1014353108>
- 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>
- Stouffer, D. B., Rezende, E. L., & Amaral, L. A. N. (2011). The role of body mass in diet contiguity and food-web structure. *Journal of Animal Ecology*, 80(3), 632–639. <https://doi.org/10.1111/j.1365-2656.2011.01812.x>
- Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335(6075), 1489–1492. <https://doi.org/10.1126/science.1216556>
- Stroud, J. T., Bush, M. R., Ladd, M. C., Nowicki, R. J., Shantz, A. A., & Sweatman, J. (2015). Is a community still a community? Reviewing definitions of key terms in community ecology. *Ecology and Evolution*, 5(21), 4757–4765. <https://doi.org/10.1002/ece3.1651>

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V., & Poisot, T. (2022). Food web reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution*, 13(12), 2838–2849. <https://doi.org/10.1111/2041-210X.13835>

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., 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 the Royal Society B: Biological Sciences*, 376(1837), 20210063. <https://doi.org/10.1098/rstb.2021.0063>

Strydom, T., Dalla Riva, G. V., & Poisot, T. (2021). SVD Entropy Reveals the High Complexity of Ecological Networks. *Frontiers in Ecology and Evolution*, 9, 623141. <https://doi.org/10.3389/fevo.2021.623141>

Thébault, E., & Fontaine, C. (2010). Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. *Science*, 329(5993), 853–856. <https://doi.org/10.1126/science.1188321>

Thompson, R. M., Brose, U., Dunne, J. A., Hall, R. O., Hladyz, S., Kitching, R. L., Martinez, N. D., Rantala, H., Romanuk, T. N., Stouffer, D. B., & Tylianakis, J. M. (2012). Food webs: Reconciling the structure and function of biodiversity. *Trends in Ecology & Evolution*, 27(12), 689–697. <https://doi.org/10.1016/j.tree.2012.08.005>

Tylianakis, J. M., Didham, R. K., Bascompte, J., & Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11(12), 1351–1363. <https://doi.org/10.1111/j.1461-0248.2008.01250.x>

- Tylianakis, J. M., & Morris, R. J. (2017). Ecological Networks Across Environmental Gradients. *Annual Review of Ecology, Evolution, and Systematics*, 48(Volume 48, 2017), 25–48. <https://doi.org/10.1146/annurev-ecolsys-110316-022821>
- Tylianakis, J. M., Tscharntke, T., & Lewis, O. T. (2007). Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature*, 445(7124), 202–205. <https://doi.org/10.1038/nature05429>
- Valiente-Banuet, A., Aizen, M. A., Alcántara, J. M., Arroyo, J., Cocucci, A., Galetti, M., García, M. B., García, D., Gómez, J. M., Jordano, P., Medel, R., Navarro, L., Obeso, J. R., Oviedo, R., Ramírez, N., Rey, P. J., Traveset, A., Verdú, M., & Zamora, R. (2019). Beyond species loss: Extinction of interactions in a changing world.
- van der Hoorn, P., Lippner, G., & Krioukov, D. (2018). Sparse Maximum-Entropy Random Graphs with a Given Power-Law Degree Distribution. *Journal of Statistical Physics*, 173(3), 806–844. <https://doi.org/10.1007/s10955-017-1887-7>
- Vázquez, D. P. (2005). Degree distribution in plant-animal mutualistic networks: Forbidden links or random interactions? *Oikos*, 108(2), 421–426. <https://doi.org/10.1111/j.0030-1299.2005.13619.x>
- Vázquez, D. P., Melián, C. J., Williams, N. M., Blüthgen, N., Krasnov, B. R., & Poulin, R. (2007). Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116(7), 1120–1127. <https://doi.org/10.1111/j.0030-1299.2007.15828.x>
- Vieira, M. C., & Almeida-Neto, M. (2015). A simple stochastic model for complex coextinctions in mutualistic networks: Robustness decreases with connectance. *Ecology Letters*, 18(2), 144–152. <https://doi.org/10.1111/ele.12394>
- Volterra, V. (1927). Fluctuations in the Abundance of a Species considered Mathematically. *Nature*, 119(2983), 12–13. <https://doi.org/10.1038/119012b0>
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196. <https://doi.org/10.3758/BF03206482>

- Warren II, R. J., Costa, J. T., & Bradford, M. A. (2022). Seeing shapes in clouds: The fallacy of deriving ecological hypotheses from statistical distributions. *Oikos*, 2022(11), e09315. <https://doi.org/10.1111/oik.09315>
- Weinstein, B. G., & Graham, C. H. (2017a). On comparing traits and abundance for predicting species interactions with imperfect detection. *Food Webs*, 11, 17–25. <https://doi.org/10.1016/j.fooweb.2017.05.002>
- Weinstein, B. G., & Graham, C. H. (2017b). Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions. *Ecology Letters*, 20(3), 326–335. <https://doi.org/10.1111/ele.12730>
- White, E. P., Thibault, K. M., & Xiao, X. (2012). Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*, 93(8), 1772–1778. <https://doi.org/10.1890/11-2177.1>
- Whittaker, R. J., Willis, K. J., & Field, R. (2001). Scale and species richness: Towards a general, hierarchical theory of species diversity. *JOURNAL OF BIOGEOGRAPHY*, 28(4), 453–470. <https://doi.org/10.1046/j.1365-2699.2001.00563.x>
- Williams, R. J. (2011). Biology, Methodology or Chance? The Degree Distributions of Bipartite Ecological Networks. *PLOS ONE*, 6(3), e17645. <https://doi.org/10.1371/journal.pone.0017645>
- Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, 404(6774), 180–183. <https://doi.org/10.1038/35004572>
- Williams, R. J., & Martinez, N. D. (2004). Limits to Trophic Levels and Omnivory in Complex Food Webs: Theory and Data. *The American Naturalist*, 163(3), 458–468. <https://doi.org/10.1086/381964>
- Wolff, J. O. (1996). Population Fluctuations of Mast-Eating Rodents Are Correlated with Production of Acorns. *Journal of Mammalogy*, 77(3), 850–856. <https://doi.org/10.2307/1382690>
- Woodward, G., Benstead, J. P., Beveridge, O. S., Blanchard, J., Brey, T., Brown, L. E., Cross, W. F., Friberg, N., Ings, T. C., Jacob, U., Jennings, S., Ledger, M. E., Milner,

- A. M., Montoya, J. M., O'Gorman, E., Olesen, J. M., Petchey, O. L., Pichler, D. E., Reuman, D. C., ... Yvon-Durocher, G. (2010). Chapter 2 - Ecological Networks in a Changing Climate. In G. Woodward (Ed.), *Advances in Ecological Research* (pp. 71–138, Vol. 42). Academic Press. <https://doi.org/10.1016/B978-0-12-381363-3.00002-2>
- Woodward, G., Brown, L. E., Edwards, F. K., Hudson, L. N., Milner, A. M., Reuman, D. C., & Ledger, M. E. (2012). Climate change impacts in multispecies systems: Drought alters food web size structure in a field experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1605), 2990–2997. <https://doi.org/10.1098/rstb.2012.0245>
- Wootton, J. T., & Emmerson, M. (2005). Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics*, 36(Volume 36, 2005), 419–444. <https://doi.org/10.1146/annurev.ecolsys.36.091704.175535>
- Young, J.-G., Valdovinos, F. S., & Newman, M. E. J. (2021). Reconstruction of plant-pollinator networks from observational data. *Nature Communications*, 12(1), 3911. <https://doi.org/10.1038/s41467-021-24149-x>



# **Appendix A. Supplementary material for chapter 1**

---

## **.1. Host-parasite network data**

### **Data description**

We use the collection of tripartite host-parasite networks sampled across Europe of Kopelke et al., 2017. This dataset contains well-resolved binary local interactions between willows (52 species), willow-galling sawflies (96 species), and their parasitoids (126 species). Out of a total of 374 local networks, we retained those containing at least 5 species, resulting in a set of 233 georeferenced local networks (networks sampled within areas of 0.1 to 0.3 km<sup>2</sup> during June and/or July spanning 29 years). Given its replicated networks spanning large spatiotemporal scales, this dataset is well-suited for analyzing network variability.

We built a metaweb of binary interactions by aggregating all local interactions, which gave us a regional network composed of 274 species and 1080 interactions.

### **Metawebs of probabilistic interactions**

We converted these binary regional interactions into probabilistic ones using simple assumptions. Our aim is not to estimate precise probability values, but to create plausible metawebs of probabilistic interactions for our illustrative examples.

We created two metawebs of probabilistic interactions by employing constant false positive and false negative rates for all regional interactions. In the first metaweb, we set both false positive and false negative rates to zero to prevent artificially inflating the total number of interactions, enabling a more accurate comparison with binary interaction networks. This

gave us a probability of regional interaction of 1 when at least one interaction has been observed locally and of 0 in the absence of any observed interaction between a given pair of species. This metaweb was used in Box 2.

In the second metaweb, we introduced a 5% false positive rate to account for spurious interactions and a 10% false negative rate to address the elevated occurrence of missing interactions in ecological networks (Catchen et al., 2023). We believe these rates represent reasonable estimates of missing and spurious potential interactions, but confirming their accuracy is challenging due to the unavailability of data on the actual feasibility of interactions. Observed interactions were thus given a probability of regional interaction of 95%, whereas unobserved ones were assigned a probability of 10%. This metaweb was used in Boxes 3 and 5.

## Local networks of probabilistic interactions

We built local networks of probabilistic interactions using the taxa found in the empirical local networks and attributing pairwise interaction probabilities based on the metawebs of probabilistic interactions  $P(M_{i,j})$  and a constant value of  $P(L_{i,j,k}|M_{i,j})$  across interactions:

$$P(L_{i,j,k}) = P(L_{i,j,k}|M_{i,j}) \times P(M_{i,j}). \quad (.1.1)$$

We set all values of  $P(L_{i,j,k}|M_{i,j})$  to 0.5, 0.75, or 1.0 depending on the simulation. Intermediate values of  $P(L_{i,j,k}|M_{i,j})$  around 50% indicate considerable spatiotemporal variability, while higher values close to 100% indicate that regional interactions are nearly always realized locally.

## .2. Additional methods for Box 2: Dissimilarity of local host-parasite networks

### Dissimilarity between local networks and the metaweb

We aggregated local networks of binary interactions by sequentially and randomly selecting a number of local networks and aggregating both their species and interactions.

We compared the metaweb of binary interactions and the aggregated local networks of binary interactions using the dissimilarity in species composition  $\beta_S$ , and the dissimilarity of interactions between common species  $\beta_{OS}$  indices. Both dissimilarity indices were calculated based on the number of items shared by the two networks ( $c_{LM}$ ) and the number of items unique to the metaweb ( $u_M$ ) and the aggregated local network ( $u_L$ ). The  $\beta_S$  dissimilarity index uses species (nodes) as items being compared, while the  $\beta_{OS}$  index assesses dissimilarity based on interactions between shared species. Both indices were calculated following the  $\beta_W$  index of Whittaker, 1960:

$$\beta_W = \frac{c_{LM} + u_L + u_M}{(2c_{LM} + u_L + u_M)/2} - 1. \quad (.2.1)$$

We repeated the aggregation process one hundred times and highlighted the median dissimilarity values across simulations, as well as the 50% and 95% percentile intervals.

### Aggregation of local networks of probabilistic interactions

We aggregated local networks of probabilistic interactions similarly to the networks of binary interactions, with the distinction that we also adjusted the value of  $P(L_{i,j,k})$  when sampling networks. The constancy of the probability of regional interaction across the entire study area means that any rise in the probability of local interaction is solely attributable to an increase in  $P(L_{i,j,k}|M_{i,j})$ . We adjusted the value of  $P(L_{i,j,k}|M_{i,j})$  as follows. Let  $L_1$  and  $L_2$  be two local networks and  $L_{1,2}$  the aggregated network. If  $P(L_{i,j,1}|M_{i,j})$  and  $P(L_{i,j,2}|M_{i,j})$  are the probabilities that two potentially interacting taxa interact respectively in  $L_1$  and

$L_2$ , the probability  $P(L_{i,j,1,2}|M_{i,j})$  that these taxa interact in the aggregated network  $L_{1,2}$  is obtained by:

$$P(L_{i,j,1,2}|M_{i,j}) = 1 - [1 - P(L_{i,j,1}|M_{i,j})] \times [1 - P(L_{i,j,2}|M_{i,j})], \quad (.2.2)$$

assuming independence between the interaction of the two taxa in different networks. This equation represents the probability that the interaction is realized in either (1) exclusively the local network  $L_1$ , (2) exclusively the local network  $L_2$  or (3) both, given that the two taxa have the biological capacity to interact.

We then calculated the probabilities of local interaction of the aggregated networks using Eq .1.1. The value of  $P(L_{i,j,k}|M_{i,j})$  for each curve in Figure 2 (panels c-d) is the probability before aggregating networks.

## Calculation of the expected number of local interactions and connectance

We investigated how the number of local interactions and connectance scale with the number of sampled (aggregated) local networks. We calculated the expected numbers of interactions by taking the sum of all binary or probabilistic interaction values. Connectance was calculated as the ratio of the expected number of interactions to the number of possible (non-forbidden) interactions. Because our networks are tripartite, connectance was calculated as follows:

$$Co = \frac{I}{S_S S_G + S_G S_P}, \quad (.2.3)$$

where  $I$  is the expected number of interactions,  $S_S$  the number of Salix species,  $S_G$  the number of galler species, and  $S_P$  the number of parasitoid species in the network.

### .3. Additional methods for Box 3: Spatial and temporal scaling of interactions

#### Aggregation of local and regional networks of probabilistic interactions

Local probabilistic interactions were derived from probabilistic regional interactions by setting the value of  $P(L_{i,j,k}|M_{i,j})$  (the local probability of interaction among potentially interacting species) to 1, ensuring a conservative comparison between aggregated local networks and metawebs. Aggregated local and regional networks were obtained by aggregating both the species and interactions found within a particular latitudinal window. The values of  $P(L_{i,j,k}|M_{i,j})$  in local networks remained at their maximum value of 1 following Eq .2.2. Latitudinal windows had different positions (central latitudes) and widths (latitudinal widths).

#### Calculation of the expected number of interactions

We calculated the expected number of local and regional interactions by taking the sum of all probabilistic interaction values of the aggregated networks.

### .4. Additional methods for Box 5: Sampling for binary interaction networks

#### Sampling using regional interaction probabilities

We sampled for binary interaction networks across space, predicting a binary interaction network for each location in our dataset. We performed a single Bernoulli trial for each pair of taxa based on their regional probability of interaction:

$$M_{i,j} \sim \text{Bernoulli}(P(M_{i,j})). \quad (.4.1)$$

Every pair of taxa predicted to interact in this metaweb was treated as interacting in all localized networks where they co-occurred, i.e.  $L_{i,j,k} = M_{i,j}$  when  $X_{i,j,k} = 1$ .

We performed between 1 and 100 simulations for each location to get a distribution of networks of binary interactions sampled using regional interaction probabilities.

## Sampling using local interaction probabilities

We sampled binary interaction networks across space, predicting a binary interaction network for each location in our dataset. We first generated distinct probabilistic interaction networks for each location. The local probability of interaction between potentially interacting species was set to three different values:  $P(L_{i,j,k}|M_{i,j}) = 1.0$ ,  $P(L_{i,j,k}|M_{i,j}) = 0.75$ , and  $P(L_{i,j,k}|M_{i,j}) = 0.50$ . We then sampled each local network of probabilistic interactions independently:

$$L_{i,j,k} \sim \text{Bernoulli}(P(L_{i,j,k})). \quad (.4.2)$$

We performed between 1 and 100 simulations for each location to get a distribution of networks of binary interactions sampled using local interaction probabilities.

## Calculation of connectance

We calculated the connectance of our predicted tripartite networks of binary interactions following Eq .2.3. We calculated the average connectance across simulations for each location.

## Calculation of the mean squared logarithmic error (MSLE)

The mean squared logarithmic error was calculated as follows:

$$MSLE = \frac{\sum(\log(\overline{Co_L}) - \log(\overline{Co_M}))^2}{n}, \quad (.4.3)$$

where  $\overline{Co_L}$  and  $\overline{Co_M}$  are the average connectance across simulations for each location, respectively for local and regional samples, and  $n$  is the number of locations.

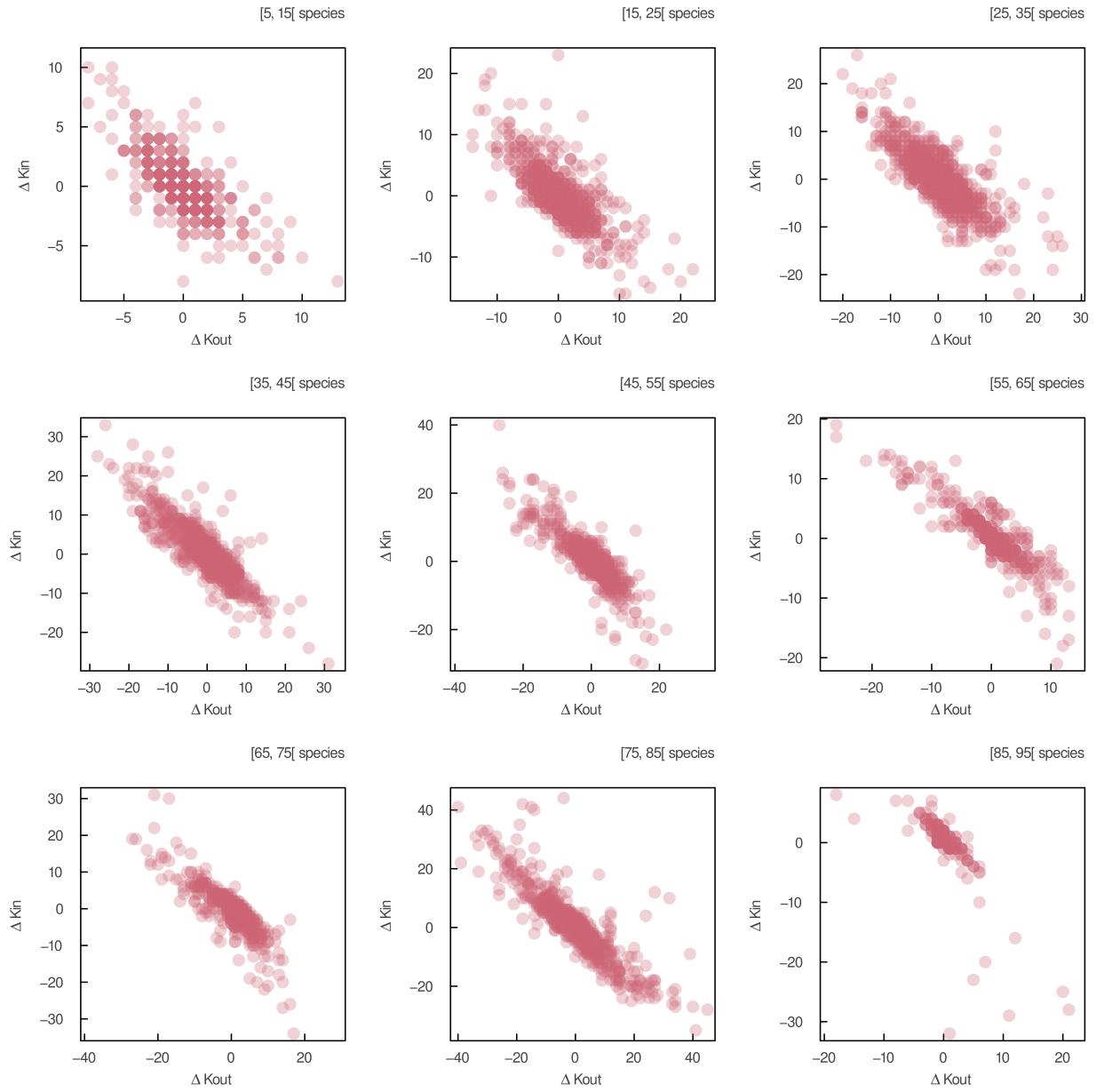
## References

---

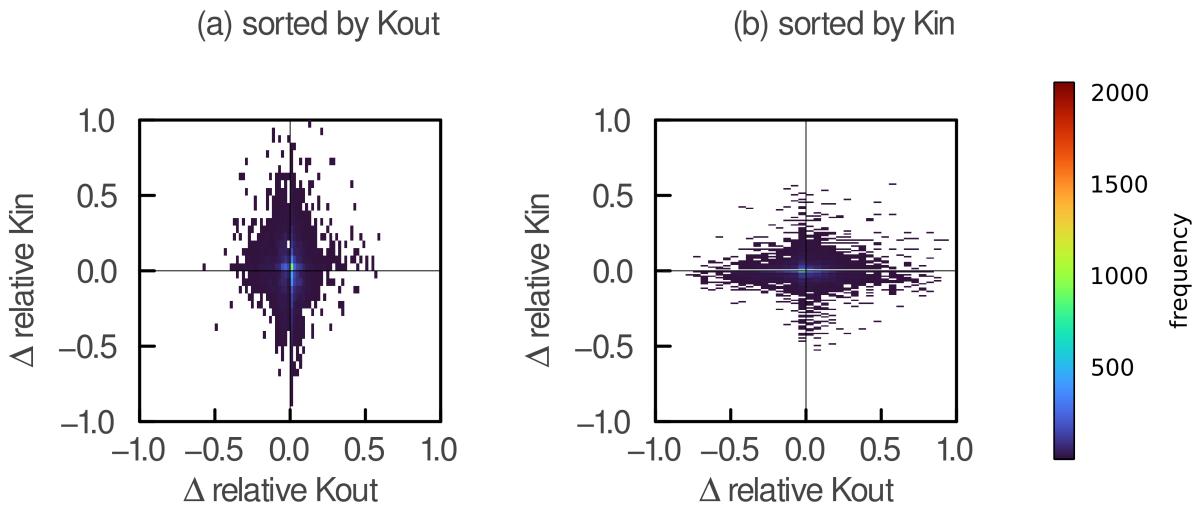
- Catchen, M. D., Poisot, T., Pollock, L. J., & Gonzalez, A. (2023). The missing link: Discerning true from false negatives when sampling species interaction networks.
- Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S., & Roslin, T. (2017). Food-web structure of willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98(6), 1730–1730. <https://doi.org/10.1002/ecy.1832>
- Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3), 279–338. <https://doi.org/10.2307/1943563>

## **Appendix B. Supplementary material for chapter 2**

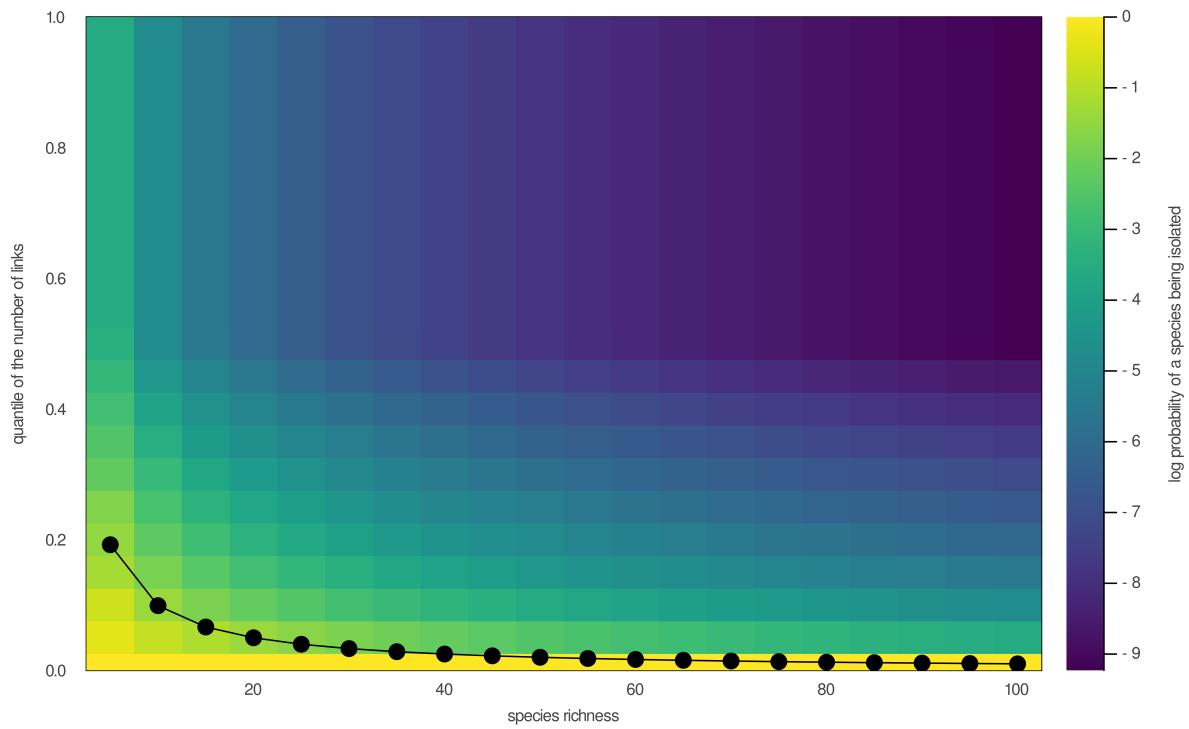
---



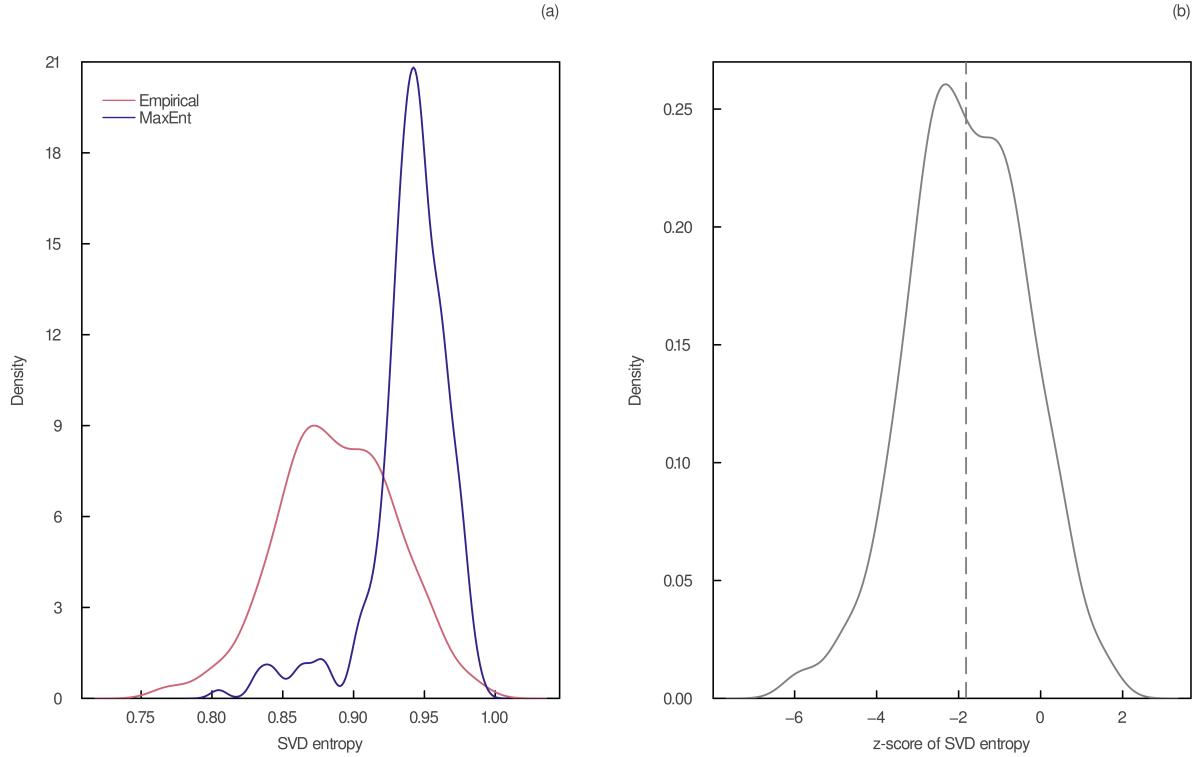
**Fig. 7. Prediction errors of the absolute number of predators  $k_{in}$  and prey  $k_{out}$ .** Species were ordered according to their total degree in their network. Networks were sorted into different groups based on their total number of species. In each panel, each dot corresponds to a single species within one of the networks whose total species count is within the specified range. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant.



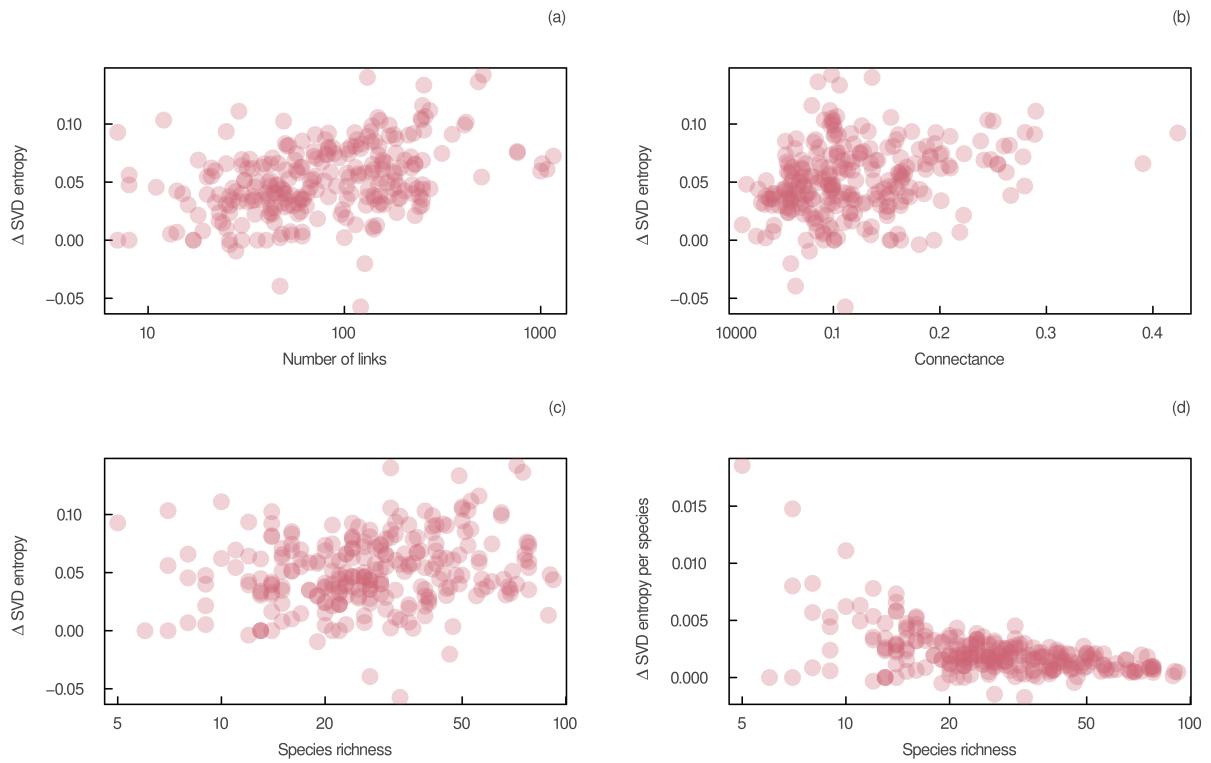
**Fig. 8. Prediction errors of the relative number of predators  $k_{in}$  and prey  $k_{out}$ .** Species were ordered according to (a) their out-degree and (b) their in-degree. The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network while keeping the total number of interactions constant. Due to significant data overlap, all relationships are represented as 2D histograms. The color bar indicates the number of species that fall within each bin.



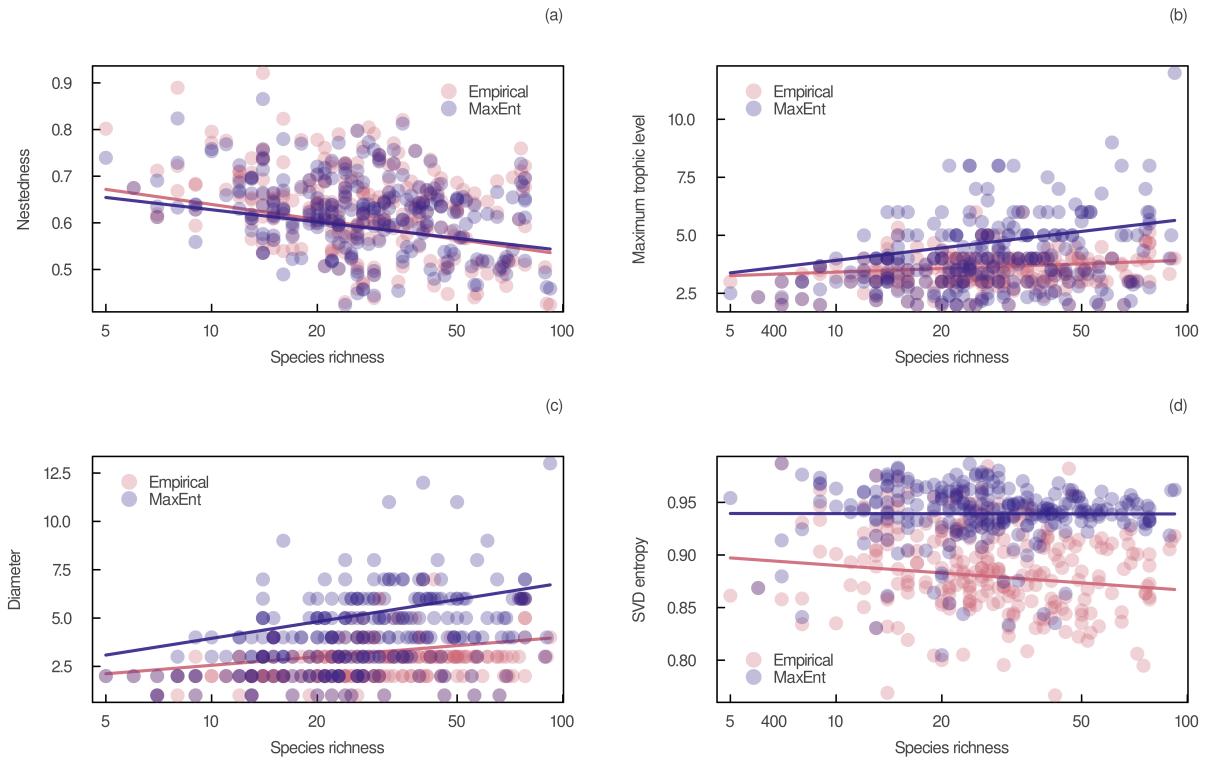
**Fig. 9. Predicted probability that a species is isolated in its food web.** We derived many degree distributions of maximum entropy given a range of values of  $S$  and  $L$  and plotted the probability that a species has a degree  $k$  of 0 (log-scale color bar). Species richness varies between 5 and 100 species, by increment of 5 species. For each level of species richness, the numbers of interactions correspond to all 20-quantiles of the interval between 0 and  $S^2$ . The black line marks the  $S - 1$  minimum number of interactions required to have no isolated species.



**Fig. 10. SVD entropy of empirical and predicted food webs.** (a) Distribution of the SVD entropy of empirical and maximum entropy food webs. Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (b) Distribution of z-scores of the SVD entropy of all empirical food webs. Z-scores were computed using the mean and standard deviation of the distribution of SVD entropy of MaxEnt food webs (type II heuristic MaxEnt model). The dashed line corresponds to the median z-score.



**Fig. 11. Prediction errors of SVD entropy.** Difference in SVD entropy between maximum entropy and empirical food webs as a function of (a) the number of interactions, (b) connectance, and (c) species richness. (d) Standardization of the difference in SVD entropy with respect to species richness as a function of species richness. The exponential decrease in the difference of SVD entropy per species with species richness offers a complementary perspective supporting the lack of relationship depicted in panel c. Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence.



**Fig. 12. Structure of empirical and maximum entropy food webs as a function of species richness.** Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (a) Nestedness (estimated using the spectral radius of the adjacency matrix), (b) the maximum trophic level, (c) the network diameter, and (d) the SVD entropy were measured on these empirical and maximum entropy food webs and plotted against species richness. Regression lines are plotted in each panel.