



**HAL**  
open science

# Vers une théorie spatiale des réseaux d'interaction en écologie : méthodes, concepts et applications

Marc Ohlmann

► **To cite this version:**

Marc Ohlmann. Vers une théorie spatiale des réseaux d'interaction en écologie : méthodes, concepts et applications. Biodiversité. Université Grenoble Alpes, 2019. Français. NNT : 2019GREAV065 . tel-03051711

**HAL Id: tel-03051711**

**<https://theses.hal.science/tel-03051711v1>**

Submitted on 10 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biodiversité-Ecologie-Environnement

Arrêté ministériel : 25 mai 2016

Présentée par

**Marc OHLMANN**

Thèse dirigée par **Wilfried THULLER**, Université Grenoble Alpes

préparée au sein du **Laboratoire Laboratoire d'ECologie Alpine**  
dans **l'École Doctorale Chimie et Sciences du Vivant**

### **Vers une théorie spatiale des réseaux d'interaction en écologie : méthodes, concepts et applications**

### **Towards a spatial theory of interaction networks in ecology: methods, concepts and applications**

Thèse soutenue publiquement le **17 décembre 2019**,  
devant le jury composé de :

**Monsieur WILFRIED THULLER**

DIRECTEUR DE RECHERCHES CNRS, CNRS DELEGATION ALPES,  
Directeur de thèse

**Madame EMMANUELLE PORCHER**

PROFESSEUR, MUSEUM D'HISTOIRE NATURELLE - PARIS,  
Rapporteur

**Monsieur FRANÇOIS MASSOL**

DIRECTEUR DE RECHERCHES CNRS, CNRS DELEGATION HAUTS-  
DE-FRANCE, Rapporteur

**Monsieur MAHENDRA MARIADASSOU**

CHARGE DE RECHERCHE, INRA CENTRE JOUY-EN-JOSAS,  
Examineur

**Monsieur BJORN REINEKING**

DIRECTEUR DE RECHERCHE, IRSTEA CENTRE DE GRENOBLE,  
Président

**Madame CORINNE VACHER**

CHARGE DE RECHERCHE HDR, INRA CENTRE BORDEAUX  
AQUITAINE, Examineur





# Remerciements

---

A Wilfried THULLER pour avoir rendu possible cette excursion scientifique, pour sa motivation et son enthousiasme à toute épreuve et pour sa confiance permettant de mener, en relative indépendance, cet exercice périlleux qu'est la thèse.

A Sébastien LAVERGNE, pour ses conseils avisés de lecture, son initiation à l'Orphisme scientifique et tout le reste.

A François MUNOZ, pour les discussions à haute dérive intellectuelle nécessaire à l'exploration d'idées nouvelles.

A Stéphane DRAY, pour son regard perçant et son accueil au LBBE.

A Vincent MIELE, pour son intransigeance concernant le code et la transmission de l'amour des graphes.

A Stéphane ROBIN pour son accueil chaleureux à l'Agro et sa rigueur intellectuelle.

A Catherine MATIAS pour son dynamisme et sa sympathie au sein de l'ANR Econet.

A Fabien LAROCHE pour son œil avisé sur les modèles de méta-population.

A Mickaël HEDDE, pour son accueil à Montpellier et la transmission de sa passion concernant la faune du sol.

A Emmanuelle PORCHER, pour m'avoir initié à l'écologie scientifique et fasciné lors de ses cours à l'Ecole Polytechnique.

A Emmanuelle TOSEL, qui m'a éveillé à la beauté de mathématiques formelles en classe préparatoire.

A mes parents, pour avoir su m'emmener, enfant, dans la nature

A l'ensemble des rapporteurs et examinateurs qui ont bien voulu examiner cette thèse.

Aux étudiants, chercheurs et personnel technique, en visite ou enracinés au LECA : compagnons de routes scientifiques durant ces dernières années.

A la mémoire de Jean-Christophe WOLF, qui m'a initié au monde des livres, de l'écologie politique et de la poésie.



# Table des matières

---

<b>Introduction</b>	<b>v</b>
0.1 Introduction écologique . . . . .	v
0.1.1 Sciences et arts dans les représentations de la nature . . . . .	v
0.1.2 Collection, contemplation et désir de conservation . . . . .	viii
0.1.3 Des technologies nouvelles pour documenter et représenter la nature	ix
0.1.4 Des espèces et des interactions . . . . .	xii
0.1.5 Plan du manuscrit . . . . .	xxiii
0.2 Éléments mathématiques sur les graphes, la diversité et les modèles gra-	
phiques . . . . .	xxv
0.2.1 Éléments et vocabulaire de théorie des graphes . . . . .	xxv
0.2.2 Mesures de diversité et comparaison de graphes . . . . .	xxxii
0.2.3 Modèles graphiques . . . . .	xxxv
<b>I Corps principal de la thèse</b>	<b>1</b>
<b>Productions académiques</b>	<b>3</b>
<b>1 Indices de diversité pour les réseaux écologiques</b>	<b>5</b>
1.1 Synopsis . . . . .	5
1.2 Corps principal de l'article (en anglais) . . . . .	6
1.3 Annexes de l'article (en anglais) . . . . .	18
<b>2 Théorie des métacommunautés</b>	<b>33</b>
2.1 Synopsis . . . . .	33
2.2 Proposition d'article (en anglais) . . . . .	33
2.3 Annexes de l'article (en anglais) . . . . .	50
2.4 Équivalence des espèces dans le réseau et distribution stationnaire . . . . .	62
<b>3 L'empreinte des interactions biotiques sur la <math>\beta</math>-diversité</b>	<b>63</b>
3.1 Synopsis . . . . .	63
3.2 Corps principal de l'article (en anglais) . . . . .	64
3.3 Annexes de l'article (en anglais) . . . . .	75

<b>4</b>	<b>Mélange de marqueur pour le metabarcoding</b>	<b>107</b>
4.1	Synopsis . . . . .	107
4.2	Analyses préliminaires . . . . .	107
4.2.1	Introduction . . . . .	107
4.2.2	Notations . . . . .	108
4.2.3	Modèle statistique . . . . .	109
4.2.4	Le jeu de données de plantes . . . . .	109
4.2.5	Résultat sur les données de plantes . . . . .	110
4.2.6	Incertitude sur les abondances et estimation de la diversité . . . . .	110
<b>5</b>	<b>Discussion</b>	<b>115</b>
5.1	Conclusions des différents chapitres . . . . .	115
5.1.1	Chapitre 1 . . . . .	115
5.1.2	Chapitre 2 . . . . .	116
5.1.3	Chapitre 3 . . . . .	117
5.1.4	Chapitre 4 . . . . .	118
5.2	Vers de l'inférence et de la comparaison de réseaux pour l'observatoire ORCHAMP . . . . .	119
5.2.1	L'observatoire ORCHAMP . . . . .	119
5.2.2	Un chemin méthodologique escarpé . . . . .	121
5.3	Projets périphériques . . . . .	125
5.4	Représentations de la nature en écologie mathématique . . . . .	127
<b>II</b>	<b>Annexes</b>	<b>131</b>
5.5	Rapport de stage M2 'Maths en action' d'Alexis Teskrat . . . . .	133
5.6	Vignette du paquet R 'econetwork' . . . . .	176
	<b>Bibliographie</b>	<b>185</b>

## Résumé

Cette thèse s'intéresse aux liens entre réseaux d'interactions en écologie, espace et temps. On assiste à un changement croissant de représentation d'une communauté d'espèces, d'un ensemble d'espèces à un ensemble d'espèces et leurs interactions : un réseau d'interactions. On s'attachera alors à élaborer les prémisses d'une théorie spatiale des réseaux, en développant des méthodes, des modèles et en les appliquant sur des données écologiques. La thèse s'articule autour de quatre chapitres. Dans un premier chapitre, on se penchera sur le problème de comparaison de réseaux en différents points du temps et de l'espace. Nous étendrons les mesures de diversité, jusque-là développées pour des abondances uniquement, aux réseaux avec abondances des espèces et des interactions. Nous nous attacherons à définir des indices à plusieurs niveaux d'agrégation des nœuds dans le réseau et montrerons l'intérêt de la méthode sur des données de réseaux trophiques. Dans un deuxième chapitre, nous nous intéresserons au développement d'une théorie des méta-communautés qui modélise explicitement l'espace comme un réseau spatial et la communauté comme un réseau d'interaction. Nous définirons la notion de capacité de persistance de la méta-communauté. Dans un troisième chapitre, nous nous intéresserons aux problèmes d'inférence d'interactions sur des données de sol d'ADN environnemental le long d'un gradient d'altitude dans les Alpes. Nous montrerons que la méthode proposée permet d'estimer l'influence des variables environnementales et de reconstruire un réseau cohérent vis-à-vis de la littérature. Dans un quatrième chapitre, nous nous intéresserons à la combinaison d'abondances provenant de différents marqueurs d'ADN environnemental et montrerons l'efficacité de la méthode proposée pour reconstruire des abondances plus proche des abondances mesurées sur des données de plantes.



## Résumé (anglais)

This thesis focuses on the links between interaction networks, space and time. There is a paradigm shift in community ecology concerning the representation of a species community from a collection of species towards species and their interactions, represented by an interaction network. We aim to build the bricks for a spatial network theory, by developing new methods, new models and applying it on ecological data. This manuscript contains four chapters. In a first chapter, we extend the diversity indices, built on Hill numbers, to network diversity indices. We define diversity indices across species aggregation levels and show the interest of this method on a trophic network data set. In a second chapter, we develop a spatially explicit meta-community theory, with various kind of interactions. We then define the notion of meta-community persistence capacity. In a third chapter, we focus on network reconstruction from environmental DNA data along an environmental gradient. We show that the proposed method allows to evaluate the influence of environmental variables on community and infer a network in agreement with the literature on soil interactions. Finally, in a fourth chapter, we develop a method to combine environmental DNA data coming from different primers and show the efficiency of the method to reconstruct abundances on a plant data set.

# Introduction

---

La science elle aussi, arrivée au terme de ses paradoxes, cesse de proposer et s'arrête à contempler et dessiner le paysage toujours vierge des phénomènes.

---

Albert Camus, *le mythe de Sisyphe*

## 0.1 Introduction écologique

### 0.1.1 Sciences et arts dans les représentations de la nature

La description de la nature est une discipline où se mêlent arts et sciences. Le romantisme a nourri des idéaux de voyage vers des terres inconnues, où l'imagination pourrait s'exalter. Goethe met en vers l'impératif romantique de voyage allié à une expérience de la nature vue comme une initiation :

Et vous, allez ! Le monde est ouvert devant vous,  
Vaste est la terre, grand et sublime le ciel ;  
Observez, étudiez, rassemblez les détails ;  
Que la nature en vous balbutie son mystère.  
*Elégie de Marienbad*, Goethe, 1827

Les travaux d'Alexandre von Humboldt, père de la biogéographie, sont empreints de cette exaltation romantique qui invite au voyage. Dans *l'introduction à la géographie des plantes*<sup>1</sup>, il écrit :

Le simple aspect de la nature, la vue des champs et des bois, causent une jouissance qui diffère essentiellement de l'impression que fait l'étude particulière de la structure d'un être organisé. Ici, c'est le détail qui nous intéresse ; là, c'est l'ensemble, ce sont des masses qui agitent notre imagination.

Leur expédition avec Aimé Bonpland les emmène, de 1799 à 1804, sur les pentes du volcan Chimborazo, en Équateur, pour étudier la répartition des plantes le long d'un gradient d'altitude. Loin d'une description systématique, hors sol, de chaque organisme et de ses organes, ils soulignent le caractère idiosyncrasique des communautés d'êtres vivants. Les planches de la géographie des plantes (Fig. 1) nous rappellent une époque où le dessin et la peinture étaient les seuls moyens de représenter la nature. Ainsi, esthétique et documentation habitaient conjointement les représentations de la nature. Humboldt<sup>2</sup> nous rappelle la cohabitation de ces deux affects :

---

1. HUMBOLDT, 1805.  
2. HUMBOLDT, 1834.

Un livre de la nature doit produire la même impression que la nature elle-même. Mais ce à quoi j'ai donné une attention toute particulière dans mes vues de la nature, et par où ma manière diffère essentiellement de celle de Chateaubriand, c'est que j'ai constamment cherché en décrivant, *en peignant*, à être toujours vrai, même scientifiquement, sans tomber dans la sécheresse de la science pure.



FIGURE 1 – Une planche de la Géographie des Plantes, Alexandre von Humboldt & Aimé Bonpland, 1805

D'art et de nature, il en est justement question dans les mouvements artistiques du milieu du XIXème siècle. En France, l'école de Barbizon témoigne d'une puissante contemplation en forêt de Fontainebleau, alliée au caractère plus mobile des outils utilisés pour la peinture. Dans les Alpes, l'école dauphinoise laisse des représentations réalistes de paysages de montagne (Fig. 2). L'aquarelliste Karl Bodmer, ami de Humboldt, entreprend un voyage qui l'amènera à remonter les rives des fleuves Mississippi et Missouri pour décrire la diversité des coutumes et des costumes des autochtones d'Amérique du Nord ainsi que la faune et les paysages de ce continent. Il laisse alors des aquarelles<sup>3</sup> à la croisée du désir d'esthétique et de documentation du voyage (Fig. 3).

La biogéographie à ses débuts insiste sur les végétaux et leur singularité d'occupation de l'espace. Bien plus tard, le poète Francis Ponge aborde la question du lien entre le temps et l'espace chez les végétaux<sup>4</sup> :

Le temps des végétaux se résout à leur espace, à l'espace qu'ils occupent peu à peu, remplissant un canevas sans doute à jamais déterminé. Lorsque c'est fini, alors la lassitude les prend et c'est le drame d'une certaine saison.

3. GOETZMANN, 1996.

4. PONGE, 1942.



FIGURE 2 – La Bérarde en Oisans, Laurent Guétal, 1882, musée de Grenoble



FIGURE 3 – Chef de la société mandan du bison mâle, aquarelle et crayon sur papier.

La montagne remplit un double objectif chez les premiers biogéographes ; d'une part, elle permet d'observer des changements abrupts de composition de communautés végétales, d'autre part, parce qu'elle est parfois difficile d'accès, elle nourrit l'attrait pour les mystères de la nature chers aux romantiques. Dans *Du sentiment de la nature dans les sociétés modernes*<sup>5</sup>, le géographe Elisée Reclus affirme et s'interroge :

---

5. RECLUS, 1866.

La vue des hautes cimes exerce sur un grand nombre d'hommes une sorte de fascination ; c'est par un instinct physique, et souvent sans mélange de réflexion, qu'ils se sentent portés vers les monts pour en gravir les escarpements. Par la majesté de leur forme et la hardiesse de leur profil dessiné en plein ciel, par la ceinture de nuées qui s'enroule à leurs flancs, par les variations incessantes de l'ombre et de la lumière qui se produisent dans les ravins et sur les contreforts, les montagnes deviennent pour ainsi dire des êtres doués de vie, et c'est afin de surprendre le secret de leur existence qu'on cherche à les conquérir. En outre on se sent attiré vers elles par le contraste qu'offre la beauté virginale de leurs pentes incultes avec la monotonie des plaines cultivées et souvent enlaidies par le travail de l'homme. Et puis les monts ne comprennent-ils pas, dans un petit espace, un résumé de toutes les splendeurs de la terre ? Les climats et les zones de végétation s'étagent sur leur pourtour : on peut y embrasser d'un seul regard les cultures, les forêts, les prairies, les rochers, les glaces, les neiges, et chaque soir la lumière mourante du soleil donne aux sommets un merveilleux aspect de transparence, comme si l'énorme masse n'était qu'une légère draperie rose flottant dans les cieux.

### 0.1.2 Collection, contemplation et désir de conservation

Si l'observation et la description de la nature mènent parfois à une volonté de conservation, ce lien est loin d'être systématique. Aux Etats-Unis, John Muir documente les trésors de la vallée du Yosemite<sup>6</sup> et contribue à la création du parc national en 1890.

A l'inverse, l'intérêt porté aux lépidoptères et leur capture, en amateur, au milieu du XIX<sup>ème</sup> siècle en Angleterre connaît un tel engouement que l'état des populations de certaines espèces devient préoccupant. En témoigne ce texte des éditeurs de l'*Entomologist's Annual* rapporté dans *le Détail du Monde*<sup>7</sup> :

L'extinction des papillons n'est pas une contingence aussi improbable que certains se plaisent à l'imaginer : de nos jours, des espèces rares sont systématiquement capturées et épinglées avec une ardeur implacable dont n'ont jamais eu à pâtir les papillons de jadis.

Bates et Wallace s'adonnent à la capture de carabes et de papillons. Ils se rendent maîtres dans l'art de conserver la beauté des spécimens morts : la thanatopraxie. Wallace, nourri aux lectures d'Humboldt, entreprend des expéditions pour documenter la diversité de mammifères à Bornéo. La chasse devient systématique pour assouvir la soif de connaissance de la nature mais aussi pour enrichir les collections et financer les expéditions grâce au commerce lucratif des spécimens. Une telle entreprise se mue parfois en un rapport curieux aux éléments de la nature comme en témoigne ce passage de l'autobiographie de Wallace repris dans l'ouvrage *le Détail du Monde*<sup>8</sup> :

---

6. MUIR, 1890.

7. BERTRAND, 2019, p. 25.

8. BERTRAND, 2019, p. 76.

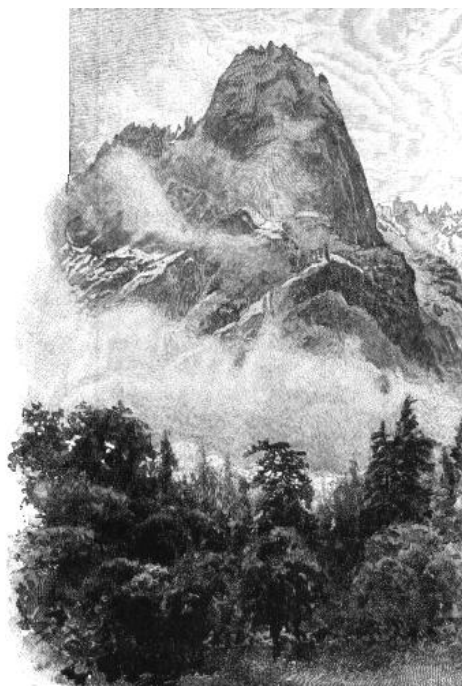


FIGURE 4 – Sentinel rock, vallée du Yosemite, Californie, Etats-Unis, John Muir, 1890, *The Century Magazine*, Vol. XL. August, 1890. No. 4

Je dois maintenant vous entretenir de l'ajout à la maisonnée d'un nourrisson orphelin, un curieux petit bébé à demi-nègre, que j'ai pris en nourrice voici plus d'un mois désormais. Je vous dirai bientôt de quelle manière je me le suis procuré, mais laissez-moi à présent vous faire récit de mes talents et de mon inventivité comme nourrice. Le petit innocent n'était pas encore sevré et je n'avais rien pour le nourrir de façon appropriée, aussi fus-je contraint de lui donner de l'eau de riz. [...] Mais je dois maintenant vous dire comment il m'a incombé d'avoir à m'en occuper : je fus la cause de la mort de sa mère. Voici comment les choses se sont passées : je chassais dans la jungle et vis soudain, au haut d'un arbre, quelque chose qui me parut être un grand singe ou un orang-outan. Je fis feu dessus, et aussitôt chuta ce petit bébé - dans les bras de sa mère. [...] De celle-ci j'ai conservé la peau et le squelette, et me voici éduquant sa seule et unique fille afin de pouvoir, un jour, la présenter à la bonne société des Jardins.

### 0.1.3 Des technologies nouvelles pour documenter et représenter la nature

L'utilisation du microscope (décrit par Athanasius Kircher en 1646) nourrit des représentations d'un monde jusqu'alors inaccessible : celui des micro-organismes. L'allemand Ernst Haeckel fournit des planches avec une volonté esthétique délibérée pour documenter la variété des formes des organismes son ouvrage *Kunstformen der Natur*<sup>9</sup> (Fig. 5). Francis

9. HAECKEL, 1904.

Ponge<sup>10</sup> propose une explication poétique à la variabilité des formes des végétaux :

La variété infinie des sentiments que fait naître le désir dans l'immobilité a donné lieu à l'infinie diversité de leur formes.

Certaines de ces planches, comme celle des diatomées (Fig. 5), ont pu être réalisées grâce à l'observation au microscope. La découverte des formes du monde microbien aurait même

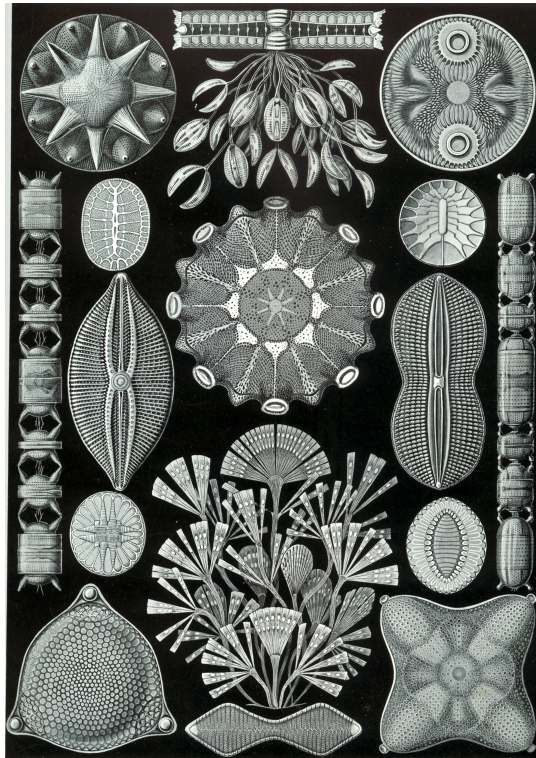


FIGURE 5 – Kunstformen der Natur, planche 84 : Diatomeae, Ernst Haeckel, 1904

été jusqu'à inspirer les œuvres de la fin de la vie du peintre russe Vassily Kandinsky, comme Complexité simple (Fig. 6).

La biogéographie des organismes visibles à l'œil nu a été documentée dès le XVIII<sup>ème</sup> siècle. LINNAEUS, 1781 s'intéresse à la relation entre la composition des communautés floristiques et l'environnement ; il émet l'hypothèse qu'à conditions environnementales identiques, les communautés floristiques ont même composition. BUFFON, 1791 met en avant l'effet de la géographie, à conditions environnementales identiques, en remarquant que l'Ancien monde et le Nouveau monde sont occupés par des mammifères distincts. WALLACE, 1876 découpe la carte du monde terrestre en six régions zoogéographiques en fonction de la similarité des vertébrés qu'elles abritent. Il insiste sur l'importance de la prise en compte des capacités de dispersion des organismes. Du fait des difficultés d'observation, la biogéographie des micro-organismes a mis plus de temps à être documentée. Une hypothèse formulée par le microbiologiste Martinus Wilhelm Beijerinck au début du

---

10. PONGE, 1942.



FIGURE 6 – Complexité simple, Vassily Kandinsky, 1939, Centre Pompidou Paris

XX<sup>ème</sup> suppose que, du fait des capacités de dispersion, supposées illimitées, des micro-organismes, seul l'environnement expliquerait leur répartition spatiale : 'everything is everywhere : but the environment selects'. Cette hypothèse d'ubiquité des micro-organismes a été soutenue jusqu'au début du XXI<sup>ème</sup> siècle (FINLAY, 2002). Toutefois, FONTANETO et HORTAL, 2012 rassemblent des preuves empiriques d'existence de biogéographie (liée à des limites dans les capacités de dispersion) pour les procaryotes, les eucaryotes unicellulaires ainsi que les micro-organismes eucaryotes multicellulaires.

Une telle direction a pu être suivie grâce au développement des technologies de barcoding, metabarcoding et de l'ADN environnemental (TABERLET et al., 2012). Ces techniques permettent d'amplifier de l'ADN extra-cellulaire dans le sol ou dans l'eau et de cibler des parties de gènes de certains groupes d'organismes. Les premières analyses d'ADN environnemental ont été effectuées dans des sédiments, pour étudier les populations de bactéries (OGRAM, SAYLER et BARKAY, 1987). Ces techniques ont ensuite été utilisées pour étudier la faune. Par exemple, AUSTEN et al., 2006 étudient les nématodes en milieu marin en ciblant le gène 18S rRNA. ARRIBAS et al., 2016 proposent d'étudier les arthropodes du sol en ciblant le gène Cox1. De telles technologies demandent des moyens en laboratoire mais aussi des outils bioinformatiques. En effet, il est nécessaire d'amplifier les fragments d'ADN grâce à une réaction en chaîne par polymérase et de lire les séquences. En ce qui concerne la bioinformatique, il est nécessaire de mesurer une distance d'édition entre plusieurs séquences (alignement de séquences). Cela permet alors de former des



groupes de séquences similaires, construisant ainsi des Unités Moléculaires Taxonomiques Opérationnelles (Molecular Operational Taxonomy Units, MOTUs). Une telle opération permet de réduire la dimension du problème mais aussi de former des entités que l'on imagine représentatives des espèces qui ont été au contact de l'échantillon. Des moyens de stockage et de calcul mutualisés sont alors nécessaires pour mener à bien le traitement bioinformatique des données d'ADN environnemental. Les techniques moléculaires ont permis l'approfondissement de la connaissance de certains organismes comme les lichens; PENNISI, 2016 nous apprend qu'ils sont constitués dans la plupart des cas d'une symbiose entre trois organismes (dont deux champignons), alors que précédemment on imaginait qu'ils étaient une symbiose entre une mousse et un champignon. De même les gradients latitudinaux et altitudinaux, connus pour les végétaux et les animaux, ont commencé à être documentés pour les micro-organismes grâce à l'ADN environnemental. Chez les champignons, les endomycorhizes et les ectomycorhizes suivent différents patrons de diversité le long de gradients altitudinaux (KIVLIN et al., 2017). De même la composition des communautés fongiques de la phyllosphère chez les hêtres varie le long d'un gradient altitudinal (CORDIER et al., 2012) et la variation des communautés fongiques associées aux racines ne suivent pas les mêmes patrons de diversité (COINCE et al., 2014).

#### 0.1.4 Des espèces et des interactions

##### Visée descriptive des réseaux d'interaction

Dans le terme écologie, forgé par Ernst Haeckel, il y a *Oikos*, l'environnement. L'environnement peut être constitué de facteurs abiotiques mais aussi d'autres organismes. On touche alors au phénomène d'interaction. A l'image de la diversité des organismes présents dans la nature, le phénomène d'interaction est extrêmement diversifié. Une approche systématique liée aux équations de Lotka-Volterra consiste à classer les interactions en fonction du signe de l'action d'une espèce A sur l'abondance future de l'espèce B et du signe de l'action de l'espèce B sur l'abondance de A. Le signe appartenant à  $\{+,-,0\}$ . On distingue ainsi la prédation  $\{+,-\}$ , la compétition  $\{-,-\}$ , le mutualisme  $\{+,+\}$ , l'amensalisme  $\{0,-\}$ , le commensalisme  $\{0,+ \}$  et le parasitisme, qui a la même combinaison de signe que la prédation. Néanmoins, dans un cas, la proie (qui a une action positive sur l'abondance du prédateur) supposée plus petite que son prédateur alors que le parasite (qui a une action négative sur l'abondance de son hôte) est supposé plus grand. Toutefois, même au sein d'un type d'interaction, il peut y avoir une diversité des phénomènes. Par exemple le mutualisme encapsule des phénomènes bien distincts : les interactions plantes-pollinisateurs ou encore les interactions entre plantes et champignons mycorhiziens. Mais même au sein des interactions mutualistes plantes-champignons, différents mécanismes sont présents. On distingue par exemple les champignons endo-mycorhiziens (*e.g.* glomeromycètes), qui pénètrent à l'intérieur des cellules des racines des plantes, des champignons

ecto-mycorhiziens (*e.g.* des basidiomycetes comme les russules ou les lactaires) qui restent à l'extérieur des cellules.

Dans l'Antiquité, le naturaliste romain Pline s'intéresse aux liens entre les arbres et les champignons, affirmant que les espèces fongiques sont « produites par l'humeur pituiteuse des arbres »<sup>11</sup>. Plus récemment, Bernardin de Saint-Pierre propose une classification des organismes par paire d'organismes présentant des affinités<sup>12</sup>. Mais plus que la focalisation sur une espèce ou une paire d'espèces en interaction, une partie de l'écologie scientifique s'intéresse à l'ensemble d'une communauté d'espèces ainsi qu'aux interactions qu'elles entretiennent. On s'intéresse alors au réseau d'interactions entre espèces, que l'on représente mathématiquement par un graphe (*cf.* introduction mathématique). Le réseau est la carte des relations entre les organismes. Pour appuyer la notion de carte et de son intérêt en modélisation, il est nécessaire d'introduire la sémantique générale du linguiste Alfred Korzybski. Dans *Une carte n'est pas le territoire*<sup>13</sup>, il propose une alternative à la logique aristotélicienne (vrai, faux) pour comprendre le langage, en creusant la métaphore de la carte. Il définit trois prémisses pour bâtir son système logique :

1. la carte n'est pas le territoire
2. la carte est toujours moins détaillée que le territoire qu'elle cherche à représenter
3. on peut faire la carte d'une carte (réflexivité)

Korzybski met en garde contre l'identification de la carte et du territoire. Ainsi, dans le cas de l'écologie des réseaux, le territoire consiste en l'ensemble des organismes vivants et les phénomènes d'interactions entre ces organismes. La carte est, dans ce cas, le réseau d'interactions, représenté par un graphe. Les nœuds de ce graphe représentent les espèces alors que les arêtes représentent les interactions. On distingue alors plusieurs types de réseaux en fonctions des interactions qui sont cartographiées. Les réseaux trophiques représentent les interactions trophiques entre les organismes. Le premier réseau échantillonné est probablement celui rapporté dans *Animal Ecology* (ELTON, 1927). Dans la mesure où l'interaction trophique est asymétrique, les réseaux trophiques sont des graphes dirigés. Les arêtes d'un réseau trophique peuvent également être pondérées pour décrire plus finement le système. Mais l'interprétation du poids des arêtes est délicate. Dans THOMPSON et al., 2012, les interprétations diverses des réseaux trophiques sont mises en avant. Les auteurs soulignent que les réseaux d'interactions sont utilisés d'une part en écologie des communautés, où les liens représentent des influences causales sur la présence ou l'abondance de différentes, et d'autre part en écologie des écosystèmes, où ils représentent des flux de matière ou d'énergie entre différents compartiments de l'écosystème. De nombreuses façons de mesurer/représenter les pondérations des arêtes d'un réseau trophique sont rapportées dans BERLOW et al., 2004. Parmi les façons de

11. PLINE, vers 77.

12. SAINT-PIERRE, 1804.

13. KORZYBSKI, 1998.

pondérer les arêtes, certaines appartiennent au monde de l'écologie théorique et ne sont pas directement quantifiables sur le terrain. Ce peut être par exemple en utilisant les paramètres de modèles dynamiques. D'autres peuvent être quantifiées par l'observation ou l'expérience. Ainsi, ces pondérations peuvent représenter les flux de biomasse de la proie vers le prédateur par unité de temps (*e.g.* COHEN, JONSSON et CARPENTER, 2003). Mais elles peuvent aussi représenter une fréquence d'interaction. Par exemple, dans le cas des réseaux hôtes-parasites, cette pondération peut représenter la fréquence des hôtes qui sont parasités (MULLER et al., 1999).

## Interactions et causalité

Si l'orientation des arêtes dans le cas où le réseau trophique représente des flux de matière est sans ambiguïté, l'orientation des arêtes dans le cas où les liens représentent des influences causales entre présences ou abondances d'espèces est plus délicate. En effet, la représentation utilisée implique une hypothèse causale sur le système. Dans le cas du système proie-prédateur, deux types de causalité (non exclusives) sont possibles. D'une part le contrôle *top-down* : c'est l'abondance du prédateur qui contrôle l'abondance de la proie, d'autre part le contrôle *bottom-up* : c'est l'abondance de la proie qui contrôle l'abondance du prédateur. Dans le modèle de Lotka-Volterra généralisé, les deux orientations sont présentes (avec un signe opposé) (ISTAS, 2006a). Dans BERLOW et al., 2004, les méthodes de pondération des arêtes reposent à la fois sur les perspectives 'top-down' et 'bottom-up'. SALA et GRAHAM, 2002 estiment les interactions trophiques entre ressources primaires et herbivore, dans un lac, en pondérant les liens du réseau par une mesure de la capacité d'ingestion de la proie par le prédateur, ignorant ainsi la réponse de la proie (c'est donc une approche 'bottom-up'). Une approche 'top-down' consiste, par exemple, à évaluer l'impact de l'exclusion d'une espèce de la communauté sur les autres espèces. Ces approches ont été utilisées pour quantifier les interactions au sein de plantes le long de gradient d'altitude (*e.g.* dans les Alpes CHOLER, MICHALET et CALLAWAY, 2001) mais aussi pour des interactions trophiques (*e.g.* exclusion de serpents de la communauté SALO et al., 2010). Les deux sens de causalité entre proie et prédateur sont parfois vus comme des alternative exclusives. C'est le fondement de la *green world hypothesis* (HAIRSTON, SMITH et SLOBODKIN, 1960), qui cherche à expliquer l'omniprésence des végétaux sur la terre par une pression des prédateurs sur les herbivores. Toutefois, elle propose une représentation simplifiée du réseau trophique, vue comme une chaîne (un graphe linéaire) en omettant le système du sol pourtant clé dans la survie des plantes. De plus, une telle représentation s'appuie sur le concept de niveau trophique et de hiérarchie des niveaux trophiques. Un concept qui a pénétré jusque dans la culture populaire ; en témoignent ces paroles du groupe de rap TSR Crew (Savanne sur Ciment) : « paraît qu'on est en haut de la chaîne, j'nous voie au plus bas niveau ». Néanmoins, ce concept a été critiqué (COUSINS, 1987) tout comme la terminologie 'bottom-up' et 'top down' qui sous-tend une représentation

hiérarchisée du réseau d'interaction. POLIS et STRONG, 1996 proposent une terminologie alternative plus proche du phénomène d'interaction en observant que lors d'une interaction trophique, il y a échange de matière organique entre la proie et le prédateur. L'interaction trophique est alors *donor control* ou *recipient control*. Sans faire référence aux terminologies 'bottom-up' ou 'top-down', TURCHIN et al., 2000 contrastent les dynamiques des lemmings et des campagnols en Europe du Nord. Les auteurs montrent alors, à partir de séries temporelles et de la forme des pics d'abondance de proies et de prédateurs dans le modèle de Lotka-Volterra, que les lemmings ont plutôt une dynamique de prédateur alors que les campagnols ont plutôt une dynamique de proie. Ainsi, ils mettent en évidence que les campagnols ont une abondance dictée par leurs prédateurs (contrôle 'top-down') alors que les lemmings ont une abondance dictée par leur ressources, les plantes (contrôle bottom-up).

### Réseau d'interactions et équivalence des espèces

Le concept de niveau trophique est construit à partir de la notion de niche d'Elton (ELTON, 1927), définie comme :

the animal's place in its community, its relations to food and enemies

Il pose une relation d'équivalence entre différentes espèces en fonctions des liens qu'elles entretiennent entre elles dans le réseau d'interaction. Différentes représentations mathématiques permettent d'approcher ce concept, de stricte équivalence mathématique entre les nœuds d'un graphe à des méthodes plus probabilistes (*cf.* introduction mathématique). Outre l'intérêt écologique, ces méthodes permettent de réduire la complexité de la carte des relations qu'est le réseau en conservant, en partie, sa structure. Ainsi, un réseau avec un grand nombre de nœud peut devenir lisible par un humain grâce à ces techniques d'agrégation des nœuds fondées sur la structure. GAUZENS et al., 2015 ont exploré dans quelle mesure ces méthodes statistiques appliquées à des données de réseaux trophiques reflétaient une image plus ou moins fidèle de la notion de groupe trophique. Sur le concept de niche d'Elton et d'équivalence entre différentes espèces s'est construit une partie de l'écologie fonctionnelle. Les travaux fondateurs de Lindemann (LINDEMAN, 1942), qui s'intéresse à l'écologie d'un lac, décrit l'écosystème comme composés de différents compartiments (eux mêmes constitués de différentes espèces) avec des fonctions précises. Ainsi, plus qu'un simple relation d'équivalence entre des espèces, il associe à chaque groupe une fonction bien précise. Le lien entre trait (grandeur mesurable chez un individu), fonction et interaction a été depuis largement étudié. En témoigne le modèle de niche (*cf.* introduction mathématique) qui utilise un trait continu (WILLIAMS et MARTINEZ, 2000). Ce modèle a été étendu en utilisant un deuxième trait, représentant la *foraging activity* de l'espèce (KLECKA, 2014). GRAVEL, ALBOUY et THUILLER, 2016 précisent les traits déterminants dans les interactions trophiques. Ils distinguent les traits

de vulnérabilités, qui conditionnent dans quelle mesure une proie va être plus ou moins vulnérable à un prédateur, des traits de *foraging activity*, qui conditionnent dans quelle mesure un prédateur va être plus ou moins capable d'attraper une proie. Ainsi, des proies ou des prédateurs ayant des traits trophiques (traits de vulnérabilité et de *foraging activity*) auront des interactions similaires et donc la même niche d'Elton. D'autres types de réseaux d'interactions sont échantillonnés, représentés et étudiés. Par exemple, BELLO et al., 2017 proposent un jeu de données spatialisé de réseaux bipartis de frugivorie dans la forêt atlantique brésilienne. Les réseaux bipartites plantes-champignons mycorrhiziens ont fait l'objet d'une attention particulière. PEAY et al., 2015 mettent en évidence une absence de spécificité des plantes envers leur cortège de champignon ecto-mycorrhiziens. TAUDIERE et al., 2015 étudient la structure d'un réseau bipartite plante-champignon ectomycorhizien et montre que le cortège de champignons change en fonction du stade de succession de la végétation.

### Variabilité de la structure des réseaux d'interaction

Nous avons vu que l'on pouvait représenter l'ensemble des interactions entre des organismes en utilisant des graphes, pour former des réseaux d'interactions, cartes de relations entre les organismes. Intéressons nous maintenant à la structure de ces réseaux et aux causes écologiques qui la façonnent. Diverses statistiques sur les graphes permettent d'appréhender numériquement la structure des réseaux d'interactions (*cf.* introduction mathématique). DORMANN, FRÜND et SCHAEFER, 2017 s'intéressent en particulier aux causes de la modularité (le fait que certains groupes de nœuds sont plus connectés entre eux qu'aux autres groupes). Ils distinguent plusieurs causes : l'aléa, la séparation d'habitat, la compétition entre espèces, des contraintes liées aux traits ou encore des mécanismes éco-évolutifs.

Comme la biogéographie souligne le caractère idiosyncratique des communautés d'organismes, on peut s'interroger sur le caractère variable dans l'espace des interactions biotiques et des ensembles d'interactions que sont les réseaux. Dans le monde fongique, nombre d'espèces peuvent passer de saprophyte à parasite, comme les *dark septate endophytes*, ou de mutualiste à parasite ; à tel point que le caractère mutualiste ou parasite d'une espèce de champignon ectomycorhizien est parfois vue comme un continuum (KARST et al., 2008). Des études se sont intéressées à la variabilité des interactions à échelle biogéographique. ROSLIN et al., 2017 mettent en évidence un gradient latitudinal dans la probabilité d'interaction trophique entre arthropodes et chenille ; la probabilité d'attaque augmente des pôles vers l'équateur. POISOT et al., 2017 s'intéressent à la biogéographie des réseaux de parasites de rongeurs à l'échelle de l'Eurasie. Les auteurs montrent que la variabilité des interactions biotiques s'expliquent par des variables climatiques différentes de celles qui expliquent la variabilité de composition de communautés d'espèces.

On peut alors interroger les mécanismes biologiques qui conduisent à la variabilité des

interactions biotiques. POISOT et al., 2017 donnent un exemple de rongeurs dont la phénologie change en fonction du climat. Ainsi, les rongeurs n’ont pas les mêmes parasites et les réseaux hôtes-parasites sont variables dans l’espace environnemental. Au sein d’un même groupe trophique (*e.g.* les plantes), l’hypothèse de gradient de stress (Stress-Gradient Hypothesis) stipule que les interactions (non-trophiques) basculent de compétition à facilitation en milieu stressant (BERTNESS et CALLAWAY, 1994). Cette hypothèse a pu être vérifiée expérimentalement dans les Alpes françaises, chez les plantes avec des expériences d’exclusion d’espèces de la communauté le long d’un gradient altitudinal (CHOLER, MICHALET et CALLAWAY, 2001). A l’inverse, DVORSKÝ et al., 2013 ne vérifient pas cette hypothèse dans des plantes en coussin situées au Ladakh, Himalaya.

L’échantillonnage de réseaux trophiques dans différents écosystèmes peut permettre d’expliquer des différences de fonctionnement entre ces écosystèmes. Ainsi, BARNES et al., 2014 modélisent des différences de fonctionnement (*e.g.* production de biomasse) en forêt tropicale en fonction du type d’utilisation des sols (forêt, culture de caoutchouc ou de palme). Pour ce faire, les auteurs ont échantillonné à la main la faune du sol dans les différents sites, assigné ces organismes à une guildes (omnivores, détritivores, prédateurs ou herbivores), mesuré leur longueur et construit un réseau trophique à partir des différentes guildes. Ainsi, en utilisant des modèles métaboliques, ils modélisent la dynamique des flux de matière dans le système.

Toutefois, si pour décrire des patrons de diversité de communautés d’espèces, les concepts et les outils que sont les mesures de diversité (*cf.* introduction mathématique) ont été largement développés et utilisés, les méthodes sont plus balbutiantes en ce qui concerne la comparaison de réseaux d’interactions. S’intéresser à la biogéographie des réseaux d’interactions plutôt que la biogéographie des assemblages d’espèces requiert d’abord de changer de représentation d’une communauté d’espèce. Assemblages d’espèces (mathématiquement représentés par un vecteur d’abondances) dans un cas, assemblages d’espèces et interactions entre ces espèces dans l’autre (représentés par un vecteur d’abondances et une matrice d’adjacence ou un graphe avec des pondérations sur les nœuds et sur les arêtes). Comme le concept de métacommunauté désigne l’union/le mélange des communautés locales en une communauté régionale, le concept de métaréseau (DUNNE, 2006) désigne l’union des réseaux locaux en un réseau régional. POISOT et al., 2012 proposent un indice de dissimilarité pour les réseaux binaires (espèces et liens présents ou absents) en décomposant de façon additive la dissimilarité totale ( $\beta_{WN}$ ) en une dissimilarité due au renouvellement d’espèces ( $\beta_{ST}$ ) et une dissimilarité due à la plasticité des interactions biotiques ( $\beta_{OS}$ ). POISOT et al., 2017 distinguent interactions potentielles et interactions réalisées. Une interaction est dite potentielle si elle est présente dans le métaréseau, c’est-à-dire qu’elle est réalisée dans au moins un réseau de la région. Une interaction est dite réalisée si elle est observée localement. Les auteurs soulignent alors le fait que les interactions potentielles sont conditionnées par les traits alors que les interac-

tions réalisées sont en plus modulées par l’environnement abiotique. La variabilité dans l’espace d’autres statistiques de réseaux ont été étudiées sans avoir recours à des mesures de dissimilarité comme la connectance ou la longueurs des chaînes trophiques (BRAGA et al., 2019). KÉFI et al., 2016 s’intéressent à la variabilité de la structure de réseaux multicouches, faisant intervenir plusieurs types d’interaction à différentes échelles d’agrégation des nœuds le long de la côte chilienne. Bien que les outils permettant de décrire les patrons de biogéographie aient été classifiés de façon systématique entre statistiques proches de l’ $\alpha$ -diversité ou de la  $\beta$ -diversité, un cadre formel unificateur fait défaut. La variabilité temporelle des réseaux d’interactions a été également étudiée. En particulier pour les réseaux plantes-pollinisateurs, où la saisonnalité engendre une variabilité de la diversité des plantes et des pollinisateurs et où les interactions peuvent être plastiques (CARADONNA et al., 2017). Le hasard peut également être une cause des patrons observés de diversité des réseaux. A l’image de la théorie neutre de la biodiversité (HUBBELL, 2001) qui peut générer des patrons de diversité sans processus déterministes, la théorie neutres des réseaux d’interaction (CANARD et al., 2012) met en lumière le fait que des structures non uniformes puissent être observées dans des réseaux sans que des processus déterministes les façonnent. Sans plaider ici pour le réalisme des théories déterministes ou stochastiques en écologie, rappelons qu’elles peuvent éclairer tour à tour la complexité écologique en changeant d’échelle d’agrégation ; des individus en entités plus larges comme les espèces ou les groupes fonctionnels. La frontière d’échelle d’agrégation entre processus déterministes et stochastiques pour l’assemblage des communautés est la notion de groupe émergent (HÉRAULT, 2007). L’importance du hasard dans l’explication des patrons écologiques est même présent chez Francis Ponge<sup>14</sup> :

Un ensemble de lois compliquées à l’extrême, c’est-à-dire le plus parfait des hasards, préside à la naissance, et au placement des végétaux sur la surface du globe. La loi des *indéterminés déterminants*.

## Échantillonnage et inférence d’interactions

Si l’échantillonnage d’organismes peut être une tâche difficile en particulier pour les micro-organismes mais possible toutefois grâce aux techniques de biologie moléculaire, l’échantillonnage des interactions est plus ardu. Pour l’interaction plante-pollinisateurs (*e.g.* plante-colibri), il peut être effectué grâce à des observations de visites à l’oeil nu (IBANEZ, 2012) ou à l’aide de caméras (*e.g.* WEINSTEIN et GRAHAM, 2017). En ce qui concerne la faune du sol, des techniques isotopiques permettent de déterminer le régime alimentaire des invertébrés (*e.g.* DIGEL et al., 2014). De même, les techniques moléculaires peuvent être appliquées à l’estomac des organismes échantillonnés pour reconstruire les réseaux d’interaction trophiques (PIÑOL et al., 2014). Dans ce cas, les arêtes (où la valuation des arêtes) représentent des grandeurs mesurables, un flux de matière caracté-

---

14. PONGE, 1942.

risé par les isotopes ou la présence d'ADN d'organismes dans l'estomac. Afin d'obtenir des réseaux en différents points de l'espace, une autre approche consiste à construire un métaréseau à dire d'expert et de littérature et en déduire les réseaux locaux comme les sous-graphes induits par les nœuds que représentent les espèces présentes localement (*e.g.* BRAGA et al., 2019). Les réseaux locaux représentent alors les interactions potentielles. Si cette approche est moins coûteuse en terme de temps passé sur le terrain (l'échantillonnage des interactions en différents points de l'espace étant besogneux), elle nécessite toutefois la mobilisation d'importantes connaissances naturalistes. De plus, dans la mesure où les réseaux locaux sont des sous-graphes induits (*cf.* introduction mathématique) du métaréseau, la seule cause de variabilité de la structure des réseaux dans ce cas est la variabilité des communautés d'espèces. En d'autres termes, à communautés locales d'espèces identiques, les réseaux locaux sont identiques.

Face aux efforts qui doivent être déployés pour échantillonner des interactions, des méthodes statistiques visent à reconstruire les réseaux d'interaction à partir de présences-absences ou d'abondances d'espèces. Ainsi, ces méthodes proposent des modèles probabilistes ou statistiques d'interaction qui sont fondées sur des données de co-occurrences (non-dynamiques). Un recours massif est alors fait aux modèles graphiques (*cf.* introduction mathématique). En ce qui concerne les modèles non-dynamiques où chaque communauté est un échantillon spatial tiré de façon indépendante dans un modèle probabiliste (une communauté d'espèces représentée par un vecteur binaire où pondéré par des abondances), on distinguera les approches conditionnelles des approches non-conditionnelles. Soulignons d'abord le caractère un peu paradoxal étymologiquement de l'objectif d'inférence d'interactions dans ce cas. « Interaction » est forgé à partir du préfixe « inter- » et du terme « action », qui vient du latin *ago* et qui signifie se mouvoir. On parlera alors plutôt de réseaux d'associations dans ce cas non-dynamique. Les approches non-conditionnelles se proposent d'utiliser la co-occurrence de paires d'espèces à travers l'espace. Ainsi, les corrélations de Pearson ont été utilisées pour construire des réseaux de co-occurrence (MANDAKOVIC et al., 2018). Dans le cas où les données de communautés sont binaires, VEECH, 2013 propose un modèle de co-occurrence dont on peut extraire certaines statistiques pour construire un réseau. L'un des problèmes majeurs de ces approches vient du fait que de trop nombreux phénomènes peuvent expliquer la co-occurrence de deux espèces, représentée mathématiquement par un scalaire. Ainsi, dans ce cas, les outils pour dessiner la carte des interactions ne parviennent pas à penser un territoire fait de phénomènes très diversifiés. Deux espèces peuvent certes être au même endroit, au même moment, si elles interagissent, bien que cette affirmation soit discutée en fonction de l'échelle spatiale (ARAÚJO et ROZENFELD, 2014) ; mais elles peuvent aussi être au même endroit, au même moment parce qu'elles ont la même niche abiotique ou à cause de la présence d'une troisième espèce. Afin d'écarter ce dernier phénomène du discours mathématique sur les interactions, les approches conditionnelles permettent de



caractériser les associations entre paires d'espèces conditionnellement à la connaissance de la distribution du reste des espèces et de l'environnement. Que ces modèles aient comme paramètre des réseaux dirigés ou non dirigés, ils ont été développés et utilisés en écologie au cours des deux dernières décennies. MILNS, BEALE et SMITH, 2010 proposent une des premières cartes d'associations entre espèces d'oiseaux qui utilisent les réseaux bayésiens. Toutefois, l'interprétation de tels objets est rendu complexe car plusieurs structures peuvent représenter les même indépendances conditionnelles (*cf.* introduction mathématique). STANICZENKO et al., 2017 proposent d'utiliser les réseaux bayésiens pour améliorer les prédictions des modèles de distributions d'espèces. HARRIS, 2016 utilise un champ de Markov pour inférer des réseaux d'association à partir de données binaires. Toutefois, dans la mesure où la fonction de partition est calculée dans l'implémentation (paquet R 'rosalia'), cette méthode est difficilement applicable à des communautés qui comportent beaucoup d'espèces. Dans ce cas, une méthode d'inférence exacte (SCHWALLER, ROBIN et STUMPF, 2015) permet d'inférer un arbre couvrant d'association entre les différentes espèces. Cette méthode a été appliquée pour inférer des réseaux d'associations faisant intervenir des parasites de plantes dans la phyllosphère (JAKUSCHKIN et al., 2016). Une des difficultés majeures consiste à valider le réseau prédit ; cette tâche est d'autant plus complexe que le nombre de nœuds est grand (ce qui proscrit la discussion de chaque arête du réseau) et que les interactions ne sont pas connues (ce qui proscrit toute comparaison avec un graphe de référence). Les études qui ont comparé des réseaux trophiques échantillonnés avec des réseaux reconstruits grâce à diverses techniques statistiques ont mis en lumière la difficulté d'un tel objectif (SANDER, WOOTTON et ALLESINA, 2017, FREILICH et al., 2018). Des développements récents qui utilisent un modèle Poisson log-Normal couplé avec le graphical lasso permettent à la fois d'utiliser une approche conditionnelle pour inférer des associations entre espèces tout en utilisant une régression sur les données environnementales (CHIQUET, MARIADASSOU et ROBIN, 2018). Ainsi, les associations inférées entre paires d'espèces sont faites conditionnellement à la distribution des autres espèces et aux variables environnementales. Deux espèces qui ont la même niche abiotique ne seront donc pas associées *a priori*. Une autre méthode propose, en utilisant des champs de Markov, d'inférer des associations dont le poids change en fonction des covariables (CHENG et al., 2014). Une telle méthode a été appliquée sur des données hôte-parasite (CLARK, WELLS et LINDBERG, 2018)

## Modèles dynamiques

Jusqu'à présent, nous nous sommes penchés sur le réseau comme description statique des espèces et de leurs interactions ou comme modèle spatial, où chaque observation (co-occurrence) est tirée de façon indépendante dans un modèle probabiliste. Nous allons nous intéresser maintenant aux réseaux d'interactions comme supports d'un modèle dynamique à la fois pour étudier des trajectoires temporelles d'abondances ou de probabilité

de présences d'espèces (écologie théorique) ou comme méthode de reconstruction de réseau à partir de données de co-occurrence temporelles (écologie statistique). Les équations de Lotka-Volterra et leurs extensions ont été utilisées pour décrire les trajectoires temporelles d'espèces en interaction. Les paramètres du modèle peuvent se représenter en utilisant un graphe pondéré. Certains résultats clés de l'écologie théorique, autour de la stabilité, utilisent ce genre de modèle. Par exemple, MAY, 1972 tire les poids des arêtes du réseau de façon indépendante dans une loi normale et montre que plus le nombre de nœuds du réseau est grand, plus le système est instable. Plus récemment, MAYNARD, SERVÁN et ALLESINA, 2018 testent l'impact de divers scénarios d'évolution de la structure du réseau sur la stabilité. Néanmoins, le modèle de Lotka-Volterra peut présenter des comportements chaotiques même pour un faible nombre d'espèces (par exemple dans le cas de quatre espèces en compétition VANO et al., 2006). Pour plus de détail sur les modèles de Lotka-Volterra, nous renvoyons aux ouvrages de Takeuchi<sup>15</sup> et Istas<sup>16</sup>. En ce qui concerne les modèles dynamiques en présence-absence, la théorie de la biogéographie insulaire (MACARTHUR et WILSON, 1965) comporte un modèle stochastique où chaque espèce a la même probabilité de coloniser l'île ou de s'éteindre sur l'île, sans interaction entre les espèces. La théorie trophique de la biogéographie insulaire se propose d'ajouter à la théorie de Mac-Arthur & Wilson un réseau d'interaction trophique, avec des hypothèses fortes sur la structure du réseau, la causalité du système et les règles d'interactions (GRAVEL et al., 2011, MASSOL et al., 2017). Le réseau trophique dans le modèle doit être un graphe acyclique orienté. La causalité supposée est donc de type 'donor control' (ou 'bottom-up'). De plus, les espèces basales (celles qui ont un degré entrant nul) suivent la dynamique du modèle de Mac-Arthur & Wilson alors que les espèces non-basales peuvent coloniser l'île seulement s'il existe au moins un chemin qui la relie à une espèce basale ; une espèce s'éteint s'il n'existe plus aucun chemin qui la relie à une espèce basale.

En ce qui concerne la reconstruction de réseaux d'interaction à partir de données temporelles, des méthodes utilisant des réseaux bayésiens dynamiques ont été mises à l'œuvre avec des résultats toutefois contrastés (SANDER, WOOTTON et ALLESINA, 2017). Des développements méthodologiques récents proposent d'inférer des réseaux bayésiens dynamiques à partir de données temporelles, en ajoutant une étiquette sur les arêtes de sorte que différents types d'interaction soient encodées dans les probabilités conditionnelles (AUCLAIR, PEYRARD et SABBADIN, 2017). La méthode a été testée afin d'inférer la structure d'un réseau trophique entre mammifères marins en Alaska. A l'image de la diversité des organismes sur la terre, nous avons évoqué la diversité des modèles permettant d'appréhender le phénomène d'interaction en terre mathématique. Ces modèles peuvent servir à modéliser la dynamique des communautés d'espèces en écologie théorique ou à reconstruire des réseaux d'interactions en écologie statistique. Si ces modèles appartiennent

---

15. TAKEUCHI, 1996.

16. ISTAS, 2006b.

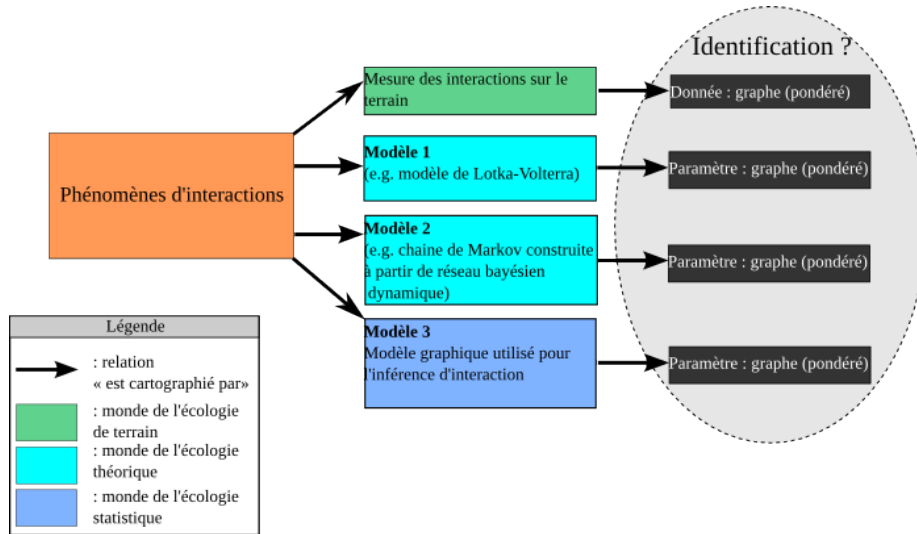


FIGURE 7 – Phénomènes, modèles, observations et paramètres. Graphe acyclique orienté représentant les différentes façon de cartographier les phénomènes d'interaction en écologie (modèles et observations) et les cartes de ces cartes (paramètres et données).

à des familles mathématiques distinctes (systèmes dynamiques ou modèle graphique), les paramètres de ces modèles (*i.e.* la carte de ces modèles qui sont eux mêmes la carte des phénomènes d'interaction) sont tous des graphes, éventuellement pondérés, visant à appréhender le phénomène d'interaction. Se pose alors la question de l'identification de tels objets (Fig. 7). Dans quelle mesure les réseaux construits à partir de l'observation des interactions sur le terrain et les réseaux qui représentent les modèles d'écologie statistique et théorique peuvent-ils être identifiés ?

## Réseaux spatiaux, méta-populations, méta-communautés

Nous nous sommes intéressés pour le moment aux graphes comme moyens de représenter les communautés d'espèces et d'interactions entre les espèces. Néanmoins, les graphes peuvent également être utilisés pour représenter l'espace. Introduits en écologie dans le contexte des modèles de métapopulation par Levins (LEVINS, 1969, une méta-population étant une population spatiale de populations à l'image d'une méta-communauté qui est une communauté de communautés), ils ont pénétré la littérature en écologie théorique (*e.g.* HANSKI et OVASKAINEN, 2003) ainsi qu'en écologie statistique (DRAY, 2011). Dans les réseaux spatiaux, les nœuds représentent des sites et les arêtes représentent les chemins de colonisation qu'une espèce peut emprunter (DALE et FORTIN, 2010). Représenter l'espace en écologie de façon discrète en utilisant des graphes permet de modéliser la fragmentation du paysage et donc de comprendre l'impact de la fragmentation du territoire (due à des perturbations d'origine anthropique ou à la géographie) sur la dynamique des populations ou communautés d'espèces. Des résultats concernant la persistance sont connus en ce qui concerne les modèles de colonisation-extinction d'une population d'une

espèce sur des réseaux spatiaux, tant pour les modèles déterministes que pour les modèles stochastiques (stochastique : OVASKAINEN, 2001, déterministe : OVASKAINEN et HANSKI, 2001, comparaison des deux : ETIENNE et NAGELKERKE, 2002). OVASKAINEN et HANSKI, 2001 définissent les notions de capacité d’invasion et de persistance d’une méta-population, qui sont liées à la structure des graphes spatiaux. Les modèles de méta-communauté (*i.e.* les modèles de méta-population à plusieurs espèces) permettent de modéliser la dynamique spatiale de plusieurs espèces, avec ou sans interaction. THULLER et al., 2007 explorent ainsi la coexistence locale et régionale de plantes en Afrique du Sud en utilisant un modèle stochastique de méta-communauté sans interaction. Les modèles théoriques de méta-communauté s’intéressent surtout à la compétition et au compromis colonisation-extinction. Cela peut être en environnement homogène (TILMAN, LEHMAN et YIN, 1997 ; CALCAGNO et al., 2006) ou en environnement hétérogène (MOUQUET et al., 2011, AMARASEKARE et al., 2004). Des modèles de méta-communauté sur grille ont été étudiés dans le cas d’interactions trophiques avec un faible nombre d’espèces (chaîne trophique à trois espèces ARAÚJO et ROZENFELD, 2014) ou pour des interactions mutualistes (FILOTAS et al., 2010). BRECHTEL et al., 2018 propose un modèle de méta-communauté trophique pour réseau spatial et réseau d’interaction avec structures arbitraires en utilisant un modèle de réaction-diffusion. Toutefois, une des hypothèses mathématique est que le modèle de réaction admet un équilibre non-nul, hypothèse difficile à vérifier. Ainsi la notion de capacité de méta-communauté n’est pas définie pour le moment. Si ces modèles ont été développés dans le cadre de l’écologie théorique, certains résultats peuvent avoir des implications pour les politiques de conservation. Par exemple, OVASKAINEN et HANSKI, 2001 proposent une méthode pour quantifier l’impact de la suppression d’un nœud du réseau spatial (un site) sur la capacité de persistance de la méta-population. Ils fournissent ainsi aux décideurs politiques un critère pour définir des zones à protéger en priorité.

### 0.1.5 Plan du manuscrit

Dans le cadre de cette thèse, nous allons d’abord nous focaliser, dans un premier chapitre, sur le réseau d’interaction comme description de la communauté d’espèces et de leurs interactions en un point de l’espace et du temps, dont la structure varie dans l’espace et dans le temps. Nous nous intéresserons au développement et à l’implémentation d’indices de diversité utilisant les nombres de Hill pour appréhender le concept de diversité de réseau d’interaction. Nous nous pencherons sur la construction de ces indices à travers plusieurs niveau d’agrégations des nœuds (c’est-à-dire des espèces) et à l’application sur des données simulées et des données de réseaux trophiques.

Nous nous intéresserons ensuite à une théorie de dynamique de méta-communauté qui modélise explicitement l’espace et les interactions entre les espèces par un réseau spatial

et un réseau d'interaction. Le modèle est un modèle stochastique construit à partir de modèle graphique (réseau bayésien dynamique) dont les paramètres sont le réseau spatial et le réseau d'interaction (qui peuvent avoir différentes structures). Nous dériverons un modèle déterministe de type champ moyen du modèle stochastique pour les interactions mutualistes, définirons la notion de capacité de persistance de la méta-communauté et étudierons dans quelle mesure cette grandeur dépend à la fois de la structure du réseau spatial et du réseau d'interaction. En ce qui concerne le modèle stochastique, nous étudierons dans quelle mesure les équivalences entre espèces influencent la dynamique et permettent de réduire la dimension du problème.

Nous nous pencherons sur le problème de la reconstruction de réseau d'association à partir de données de metabarcoding. Le site d'étude est un site pilote, dans les Alpes Françaises, de l'observatoire ORCHAMP. Nous montrerons, en utilisant des réseaux de corrélation partielles sur de la  $\beta$ -diversité (en présence-absence), que la méthode permet de retrouver des associations entre groupes clés dans le fonctionnement du sol et entre groupes trophiques dans une simulation.

Enfin, dans le but de reconstruire des réseaux d'interactions à partir de données d'abondances issues de metabarcoding, données par différentes technologies d'observation (les marqueurs), nous proposerons une méthode statistique permettant de combiner ces différentes données.

Nous discuterons enfin les différents chapitres, la possibilité d'inférer et de comparer des réseaux à partir de données de metabarcoding sur les Alpes françaises et les travaux en cours auxquels sont reliés cette thèse.

## 0.2 Éléments mathématiques sur les graphes, la diversité et les modèles graphiques

### 0.2.1 Éléments et vocabulaire de théorie des graphes

On introduira ici quelques notions de théorie des graphes nécessaires à la compréhension de cette thèse. Pour un panorama plus exhaustif, on se réfère aux ouvrages de Claude Berge<sup>17</sup> et de Barabasi<sup>18</sup>.

#### Définitions et vocabulaire de base

**Définition 0.2.1** *Un graphe non-dirigé  $G$  est un couple d'ensembles finis  $(V, E)$  tel que :*

- $V$ , l'ensemble des nœuds (ou sommets) de  $G$ , est non vide et fini
- $E$ , l'ensemble des arêtes du graphe, est un sous ensemble de  $V \times V \setminus \{(v, v), v \in V\}$ .

Le graphe est dirigé si les arêtes sont constituées de couples de nœuds  $(u, v)$  et non dirigé si les arêtes sont constitués de sous-ensembles de taille 2 des nœuds  $\{u, v\}$ .

**Définition 0.2.2** *Soit  $G = (V, E)$  un graphe avec  $|V| = n \in \mathbb{N}^*$ . On note  $\{v_1, v_2, \dots, v_n\}$  les nœuds du graphe  $G$ .*

*On appelle matrice d'adjacence, la matrice  $\mathbf{A}$  de taille  $n \times n$  dont les coefficients sont :*

$$a_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon} \end{cases}$$

Dans le cas des graphes non-dirigés,  $\mathbf{A}$  est symétrique.

**Définition 0.2.3** *Un graphe valué sur ses arêtes est un graphe auquel on adjoint une fonction de valuation.*

$$\Phi : \begin{array}{ll} E & \longrightarrow \mathbb{I} \\ (u, v) & \longmapsto \Phi(u, v) \end{array}$$

Si  $\mathbb{I}$  est de taille fini, on parle d'étiquette, si  $\mathbb{I}$  est infini on parle de pondération. Dans le cas où  $\mathbb{I} = \mathbb{R}^{+*}$ , la matrice d'adjacence du graphe est alors dans  $M_n(\mathbb{R}^{+*})$ .

---

17. BERGE, 1973.

18. BARABÁSI, 2016.

## Quelques mesures sur les graphe

**Définition 0.2.4** Soit  $G = (V, E)$  un graphe non-dirigé. Si  $(u, v) \in E$ , alors on dit que  $u$  et  $v$  sont voisins. On appelle voisinage de  $u \in V$  l'ensemble des nœuds voisins de  $u$  et on le note  $N_u$  :

$$N_u = \{v \in V, \{u, v\} \in E\}$$

Le degré de  $u$ , noté  $deg(u)$ , est alors le cardinal de  $N_u$ .

On a également  $deg(u) = \sum_k A_{k,u} = \sum_k A_{u,k}$ . On dit qu'un graphe est  $k$ -régulier si tous les nœuds sont de degré  $k$

Dans le cas dirigé, on distingue les parents de  $u$  :

$$Par_G(u) = \{v \in V, (v, u) \in E\} \quad (1)$$

Et les enfants de  $u$  :

$$Enf_G(u) = \{v \in V, (u, v) \in E\} \quad (2)$$

On a alors  $N_u = Par_u \cup Enf_u$ .

Le degré entrant de  $u$ , noté  $deg_{in}(u)$  est le cardinal de  $Par_G(u)$  et le degré sortant, noté  $deg_{out}(u)$  est le cardinal de  $Enf_G(u)$ . On a :

$$deg(u) = deg_{in}(u) + deg_{out}(u) \quad (3)$$

Dans le cas d'un graphe pondéré, le degré pondéré d'un nœud  $u$  est :

$$deg_u^{pond} = \sum_{(u,v) \in E} \phi((u, v)) \quad (4)$$

On peut également définir la notion de chemin partant du nœud  $u$  et arrivant en  $v$ , que l'on note  $\mu[u, v]$ . Un chemin est une suite d'arêtes consécutives (le nœud terminal d'une arête doit être le nœud de départ de l'arête suivante). S'il existe, on note  $|\mu[u, v]|$  sa longueur : c'est le nombre d'arêtes comprises dans le chemin. Dans le cas où le graphe est pondéré, on pondère la longueur de chemin par une fonction du poids des arêtes.

On dit qu'un graphe non dirigé est connecté s'il existe un chemin entre tous les paires de nœuds du graphe.

On dit qu'un graphe dirigé est fortement connecté si quelques soient les nœuds  $u$  et  $v$ , il existe un chemin de  $u$  vers  $v$  et de  $v$  vers  $u$ .

On dit que  $G_{sub} = (V_{sub}, E_{sub})$  est un sous-graphe de  $G = (V, E)$  si  $V_{sub} \subset V$  et  $E_{sub} \subset E \cap (V_{sub})^2$ . On dira que  $G_{subind} = (V_{subind}, E_{subind})$  est une sous graphe induit par  $V_{subind}$  de  $G$  si  $G_{subind}$  est un sous-graphe de  $G$  et que :

$$\forall (u, v) \in V_{subind}^2, (u, v) \in E \Rightarrow (u, v) \in E_{subind} \quad (5)$$

## Opérations sur les graphes, équivalence des nœuds

On peut munir l'ensemble des graphes de différentes opérations, nous en définirons ici quelques-unes. Pour un panorama sur les produits de graphe, nous renvoyons à IMRICH et KLAVZAR, 2000.

On considère ici deux graphes  $G_1 = (V_1, E_1)$  et  $G_2 = (V_2, E_2)$  (dont les matrices d'adjacence sont  $\mathbf{A}_1$  et  $\mathbf{A}_2$ ). On note  $n_1 = |V_1|$  et  $n_2 = |V_2|$ .

**Définition 0.2.5** *Le graphe union de  $G_1$  et  $G_2$ ,  $G_{\cup} = (V_{\cup}, E_{\cup})$ , est tel que  $V_{\cup} = V_1 \cup V_2$  et  $E_{\cup} = E_1 \cup E_2$ .*

On peut également munir les graphes de divers produits.

**Définition 0.2.6** *Le produit tensoriel de  $G_1$  et  $G_2$ , noté  $G_{\text{tens}} = G_1 \otimes G_2$ , est le graphe dont la matrice d'adjacence est  $\mathbf{A}_{\text{tens}} = \mathbf{A}_1 \otimes \mathbf{A}_2$  (où  $\otimes$  est le produit de Kronecker).*

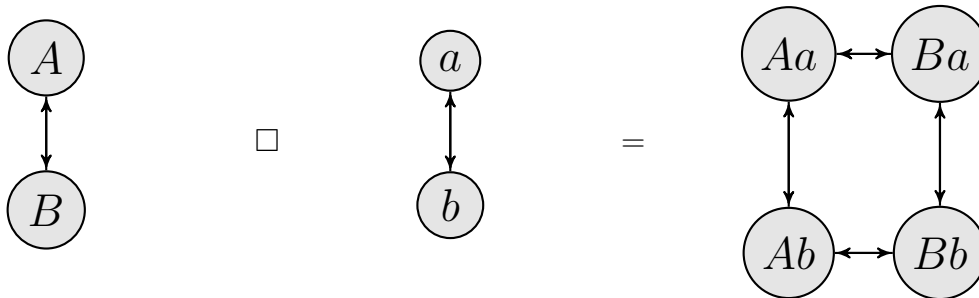
**Définition 0.2.7** *On définit le produit cartésien des graphes  $G_1$  et  $G_2$ , noté  $G_1 \square G_2$ , le graphe dont l'ensemble des nœuds est  $V_1 \times V_2$  et dont l'ensemble des arêtes est tel que deux nœuds  $u_1u_2$  et  $v_1v_2$  de ce graphe sont voisins si  $(u_1 = v_1 \text{ et } (u_2, v_2) \in E(G_2))$  ou  $(u_2 = v_2 \text{ et } (u_1, v_1) \in E(G_1))$ .*

Ainsi,

$$V(G_1 \square G_2) = \{(u_1, u_2) \mid u_1 \in G_1, u_2 \in G_2\}$$

$$E(G_1 \square G_2) = \{((u_1, u_2), (v_1, v_2)) \mid u_1 = v_1, (u_2, v_2) \in E(G_2) \text{ ou } u_2 = v_2, (u_1, v_1) \in E(G_1)\}$$

On a par exemple la multiplication suivante :



**Proposition 0.2.8** *La matrice d'adjacence du graphe produit  $G_1 \square G_2$ , notée  $\mathbf{A}_{1 \square 2}$ , est donnée par  $\mathbf{A}_{1 \square 2} = \mathbf{A}_1 \otimes \mathbf{I}_{n_2} + \mathbf{I}_{n_1} \otimes \mathbf{A}_2$*

Afin d'approcher ce qui pourrait être qualifié de niche d'Elton dans un réseau écologique, on introduit une relation d'équivalence entre les nœuds du graphe qui représente le réseau écologique. Pour cela il est nécessaire d'introduire la notion d'automorphisme de graphe.



**Définition 0.2.9**  $\phi : V \rightarrow V$  est un automorphisme de  $G = (V, E)$  si  $(u, v) \in E \iff (\phi(u), \phi(v)) \in E$ .

On note  $Aut(G)$  l'ensemble des automorphismes de  $G$ .

**Proposition 0.2.10**  $Aut(G)$  est un groupe. On l'appelle groupe des automorphismes de  $G$ .

On va montrer que  $Aut(G)$  est une sous-groupe du groupe des permutations de  $V$ .

On a  $Id \in Aut(G)$ .

De plus, soient  $\phi_1$  et  $\phi_2$  deux éléments de  $Aut(G)$ . Alors  $(u, v) \in E \iff (\phi_1(u), \phi_1(v)) \in E \iff (\phi_2 \circ \phi_1(u), \phi_2 \circ \phi_1(v)) \in E$ . Donc  $\phi_2 \circ \phi_1 \in Aut(G)$

Enfin, si  $\phi \in Aut(G)$ , on considère alors la permutation  $s$  de  $V$  telle que  $\phi \circ s = Id$  (elle existe car l'ensemble des permutations de  $V$  est un groupe). On a maintenant  $(u, v) \in E \iff (\phi(u), \phi(v)) \in E$ . D'où  $(s(u), s(v)) \in E \iff (u, v) \in E$ . Donc  $s \in Aut(G)$  et  $s = \phi^{-1}$ .

On peut maintenant définir la notion d'orbite d'un nœuds.

**Définition 0.2.11** Soit  $G = (V, E)$  un graphe. On note  $\mathcal{R}$  la relation d'équivalence sur les nœuds de  $G$  définie de la manière suivante :

$$\forall (u, v) \in V^2 \quad u\mathcal{R}v \iff \exists \phi \in Aut(G) / \phi(u) = v$$

Les classes d'équivalence associées à la relation  $\mathcal{R}$  sont appelées orbites du graphe  $G$ , notées  $\mathcal{O}(u)$ , avec  $u \in V$  un représentant de la classe.

La relation  $u\mathcal{R}v$  est bien une relation d'équivalence. En effet, elle est réflexive ( $u$  et  $u$  appartiennent à la même orbite car  $Id \in Aut(G)$ ), symétrique (si  $\phi \in Aut(G)$  envoie  $u$  sur  $v$ , alors  $\phi^{-1}$  envoie  $v$  sur  $u$  et  $\phi^{-1} \in Aut(G)$  car  $Aut(G)$  est un groupe) et transitive par stabilité de  $Aut(G)$  par composition.

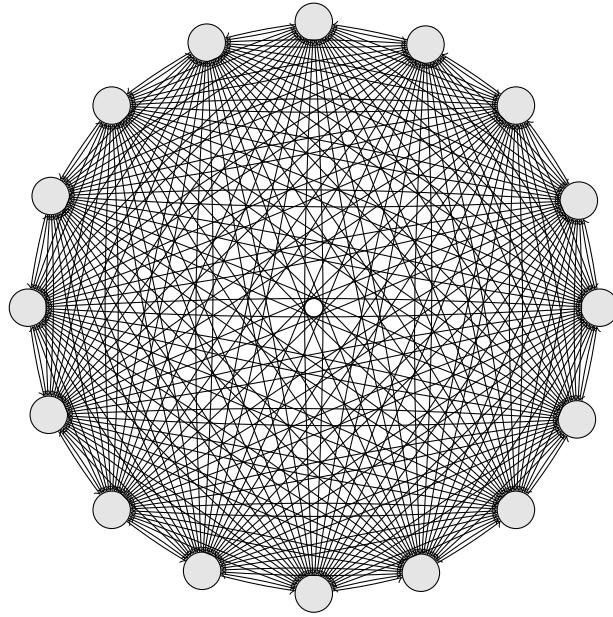
Appartenir à la même orbite est donc une relation d'équivalence mathématique des nœuds d'un graphe. En ce qui concerne les réseaux trophiques, on peut également interpréter cette équivalence comme une équivalence écologique des espèces à travers le concept de niche d'Elton. D'autres approches du concept de niche d'Elton ont été proposées dans la littérature comme l'équivalence structurale ou l'équivalence régulière (LUCZKOVICH et al., 2003).

On dit qu'un graphe est asymétrique si son groupe d'automorphisme est réduit à l'identité.

On dit qu'un graphe  $G$  est sommet-transitif s'il n'a qu'une seule orbite :

$$\forall (u, v) \in V \times V, \exists \Phi \in Aut(G) / \Phi(u) = v \tag{6}$$

Ainsi, dans un graphe somme-transitif, tous les nœuds sont équivalents.

FIGURE 8 –  $K_{16}$ , graphe complet à 16 nœuds

## Bêtes communes et curieuses, graphes aléatoires

### Bêtes communes et curieuses

**Définition 0.2.12** *Un graphe est dit complet, dans le cas non-orienté, lorsque chaque nœud est voisin de tous les autres nœuds. Dans le cas orienté, le graphe est dit complet lorsque chaque nœud est parent de tous les autres nœuds. Un graphe complet dont l'ensemble des nœuds est de cardinal  $n$  se note  $K_n$ .*

Le groupe d'automorphisme de  $K_n$  est l'ensemble des permutations des nœuds.  $K_n$  est donc sommet-transitif.

**Définition 0.2.13** *Un graphe est dit biparti si son ensemble de sommets peut être divisé en deux sous-ensembles disjoints  $V_1$  et  $V_2$  tels que chaque arête ait une extrémité dans  $V_1$  l'autre dans  $V_2$ . L'ensemble des graphes bipartis et complets (toutes les arêtes entre  $V_1$  et  $V_2$  sont présentes) dont l'ensemble des sommets peut être divisé en deux sous-ensemble de taille  $m$  et  $n$ , se note  $K_{mn}$ .*

Plus généralement, un graphe est dit multiparti si son ensemble de sommets peut être divisé en plusieurs sous-ensembles disjoints tels que chaque arête ait une extrémité dans un de ces ensembles et l'autre dans un autre de ces ensembles. On note  $K_{\underbrace{n_1 \dots n_1}_{r_1 \text{ fois}} \underbrace{n_2 \dots n_2}_{r_2 \text{ fois}} \dots \underbrace{n_p \dots n_p}_{r_p \text{ fois}}}$ ,

avec  $p \in \mathbb{N}^*$  et  $\forall i \in \{1, \dots, p\} (n_i, r_i) \in (\mathbb{N}^*)^2$ , l'ensemble des graphes multipartis complets, dont la tailles des ensembles partitionnant le l'ensemble des nœuds est décrite par les indices.

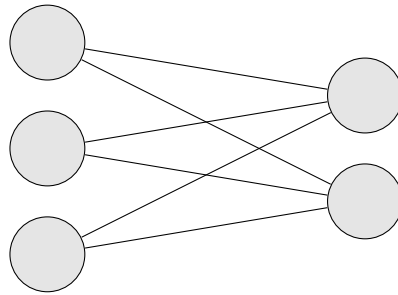


FIGURE 9 –  $K_{3,2}$

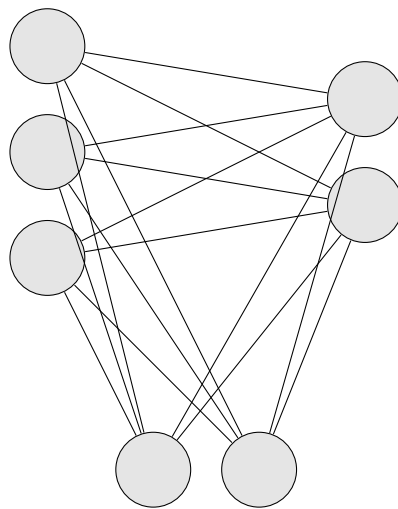


FIGURE 10 –  $K_{3,2,2}$

Les  $K_{n,\dots,n}$  sont sommet-transitifs.

Un graphe est dit acyclique orienté (Directed Acyclic Graph, DAG) si les arêtes sont dirigées et qu'il ne contient pas de cycle. On dit qu'un graphe non-orienté est un arbre s'il ne contient pas de cycle (un cycle est un chemin de longueur non-nulle qui a même noeud de départ et d'arrivée). Certains graphes réguliers comme le graphe de Frucht (12 noeuds, 3-régulier) sont asymétriques. Si l'on imagine que le graphe de Frucht représente un réseau d'interaction, alors cela signifie que dans ce réseau, toutes les espèces interagissent avec le même nombre d'espèce mais qu'aucune d'entre elles ne partage la même niche d'Elton.

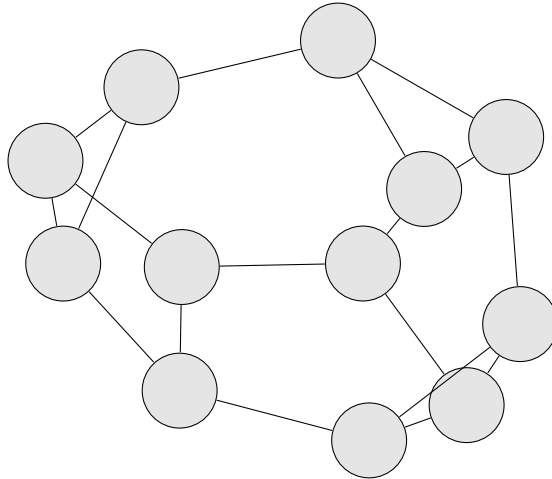


FIGURE 11 – Le graphe de Frucht

### Graphes aléatoires et écologie

Un graphe d’Erdős-Rényi à  $n$  nœuds et avec une probabilité d’arête  $p$ , noté  $G(n, p)$  est un graphe aléatoire à  $n$  nœuds dont les arêtes sont tirées de façon indépendante et identiquement distribuée dans des lois de Bernoulli de paramètre  $p$ .

Une statistique commune en écologie des réseaux est la connectance (généralement notée  $C$ ), qui est l’estimation, pour un réseau donnée, de la probabilité d’arête entre deux nœuds en supposant que le réseau est un graphe d’Erdős-Rényi. Un modèle stochastique à blocs (NEWMAN et LEICHT, 2007, DAUDIN, PICARD et ROBIN, 2008, ALLESINA et PASCUAL, 2009) est un graphe aléatoire qui est un mélange de graphes d’Erdős-Rényi. Si  $V = \{1, \dots, n\}$  est son ensemble de nœuds et étant donné une partition des nœuds en  $r$  blocs,  $\{B_1, \dots, B_r\}$ , et une matrice de probabilité d’arêtes entre les groupes  $\Pi$  (de dimension  $r * r$ ), chaque arête du graphe est tirée indépendamment dans une loi de Bernoulli dont le paramètre est donné par le groupe d’appartenance des deux sommets que cette arête lie. On a :

$$Pr(u \rightarrow v | u \in B_i, v \in B_j) = \pi_{i,j} \quad (7)$$

En écologie, si les nœuds représentent des espèces et les liens des interactions, un modèle stochastique à blocs peut représenter la façon dont plusieurs groupes équivalents d’espèces en termes de traits (l’ensemble des traits est ici fini) interagissent entre elles. Les interactions sont alors aléatoires mais conditionnées par les groupes auxquels appartiennent les espèces impliquées dans l’interaction. Une telle interprétation est suggérée par GRAVEL, ALBOUY et THUILLER, 2016.

Des modèles de graphe aléatoire propres à l’écologie théorique ont également été développés. Le modèle en cascade propose de représenter un réseau trophique par un graphe acyclique orienté dont les arêtes sont tirées dans des lois de Bernoulli (COHEN, BRIAND et NEWMAN, 1990). La hiérarchie des nœuds induites par le réseau (*i.e.* l’ordre topologique

des nœuds dans le graphe acyclique orienté) s'interprète comme la hiérarchie d'un trait dans la communauté (par exemple la taille moyenne des espèces modélisées).

Le modèle de niche propose un modèle de graphe aléatoire plus complexe mais toutefois fondé sur la notion de trait (WILLIAMS et MARTINEZ, 2000). Chaque espèce  $i$  se voit assignée une valeur de trait  $n_i$  (le paramètre de niche) tirée dans une loi uniforme sur  $[0; 1]$ . Chaque espèce peut alors consommer les espèces dans un intervalle  $r_i$  tiré dans une loi beta d'espérance  $2^*C$  et de paramètre de forme  $\alpha = 1$ , où  $C$  est la connectance désirée du réseau. Le centre  $c_i$  de l'intervalle dans lequel les espèces peuvent consommer d'autres espèces  $c_i$  est placé uniformément dans  $[\frac{r_i}{2}; n_i]$ . Le modèle de niche peut produire des cycles.

Une extension du modèle de niche utilise un trait continu et un deuxième trait fini (trois modalités) censé représenter l'activité de recherche de proies (KLECKA, 2014).

## 0.2.2 Mesures de diversité et comparaison de graphes

On introduira ici quelques éléments sur les mesures de diversité des communautés d'espèces représentées par des abondances.

### Mesures de diversité et entropie statistique

On représente une communauté de  $n$  espèces par un vecteur d'abondances relatives  $\mathbf{p}$ . On associe une variable aléatoire catégorielle  $Y \sim \text{Cat}(\mathbf{p})$  à l'expérience : « On tire uniformément un individu de la communauté, de quelle espèce est cet individu ? ». On mesure alors la diversité de la communauté en utilisant les nombres de Hill, qui sont construits à partir de l'entropie de Rényi, elle même une extension de l'entropie de Shannon (RÉNYI, 1961, HILL, 1973)

L'entropie de Shannon de  $Y$  est :

$$H(Y) = H(\mathbf{p}) = -\mathbb{E}(\ln(Y)) = -\sum_{i=1}^n p_i \ln(p_i) \quad (8)$$

En définissant la norme  $L_q$  (pour  $q \in \mathbb{R}^+ \setminus \{1\}$ ) d'un vecteur de  $\mathbb{R}^n$  comme :

$$\|\cdot\|_q : \begin{array}{l} \mathbb{R}^n \longrightarrow \mathbb{R}^+ \\ \mathbf{p} \longmapsto (\sum_{i=1}^n p_i^q)^{\frac{1}{q}} \end{array}$$

On définit l'entropie de Rényi d'ordre  $q$  ( $q \in \mathbb{R}^+ \setminus \{1\}$ ) de  $Y$  comme :

$$H_q(\mathbf{p}) = \frac{q}{1-q} (\ln(\|\mathbf{p}\|_q)) = \frac{1}{1-q} \ln \left( \sum_{i=1}^n p_i^q \right) \quad (9)$$

Les nombres de Hill, utilisés en écologie afin de mesurer la diversité d'une communauté d'espèces (représentée par une loi catégorielle) sont des transformations exponentielles des

entropies de Rényi. Ainsi, le nombre de Hill d'ordre  $q$  ( $q \in \mathbb{R}^+ \setminus \{1\}$ ) est défini comme :

$$D_q(\mathbf{p}) = \left( \sum_{i=1}^n p_i^q \right)^{\frac{1}{1-q}} \quad (10)$$

On prolonge par continuité  $D^q$  pour  $q = 1$  pour obtenir l'indice de Shannon.

$$D_1(\mathbf{p}) = \lim_{q \rightarrow 1} D_q(\mathbf{p}) = \exp \left( \sum_{i=1}^n -p_i \log p_i \right) \quad (11)$$

Si, en théorie de l'information, l'entropie se mesure en bits (le logarithme est en base 2); en écologie, la diversité se mesure en nombre effectif d'espèces (le logarithme est en base naturelle). Ainsi, si une communauté a une diversité  $K$ , cela revient à dire que cette communauté a une diversité identique à une communauté composée de  $K$  espèces (*i.e.* égale à l'exponentielle de l'entropie différentielle d'une loi uniforme sur  $[0; K]$ ). Les nombres de Hill permettent ainsi de mesurer la diversité d'une communauté d'espèces. On a, pour  $q \in \mathbb{R}^+$ ,  $1 \leq D_q(\mathbf{p}) \leq n$ , avec  $D_q(\mathbf{p}) = n$  si  $Y$  est une loi uniforme.

## Indices de diversité

Dans le cas d'une métacommunauté, qui est une communauté de communautés, mesurer la diversité nécessite l'emploi d'indices qui permettent d'appréhender la diversité globale ( $\gamma$ -diversité), la moyenne des diversités locales ( $\alpha$ -diversité) et la similarité des communautés locales ( $\beta$ -diversité). Une telle partition a été proposée par WHITTAKER, 1960. Différents indices de diversité utilisant les nombres de Hill ont été développés et étudiés (*e.g.* REEVE et al., 2014, CHAO et CHIU, 2016). Nous présenterons ici le cadre de CHAO et CHIU, 2016. Une métacommunauté n'est plus représentée par un vecteur d'abondances relatives mais par une matrice de communauté. Dans le cas où une métacommunauté comprend  $n$  espèces et  $m$  sites, on la représente par une matrice  $\mathbf{P}$ .  $P_{ij}$  est l'abondance relative de l'espèce  $j$  dans la communauté  $i$ .

### $\gamma$ -diversité

La  $\gamma$ -diversité est la diversité de la communauté régionale (*i.e.* la communauté dont les abondances sont les abondances des communautés locales moyennées à travers les sites). On a, pour  $q \in \mathbb{R}^+ \setminus \{1\}$  :

$$G_q(\mathbf{P}) = \left( \sum_{j=1}^m \left( \frac{P_{+j}}{P_{++}} \right)^q \right)^{\frac{1}{1-q}} \quad (12)$$

où  $P_{+j} = \sum_i P_{ij}$  et  $P_{++} = \sum_{i,j} P_{ij}$ .

Pour  $q = 1$ , on a :

$$G_q(\mathbf{P}) = \exp\left(\sum_{j=1}^m -\frac{P_{+j}}{P_{++}} \log\left(\frac{P_{+j}}{P_{++}}\right)\right) \quad (13)$$

L'unité de mesure de la  $\gamma$ -diversité est un nombre effectif d'espèces dans la métacommunauté. Ainsi :  $1 \leq G_q(\mathbf{P}) \leq n$

### $\alpha$ -diversité

L' $\alpha$ -diversité est la moyenne généralisée des diversité des communautés locales. On a, pour  $q \in \mathbb{R}^+ \setminus \{1\}$  :

$$A_q(\mathbf{P}) = \frac{1}{m} \left( \sum_{i=1}^n \sum_{j=1}^m \left( \frac{P_{ij}}{P_{++}} \right)^q \right)^{\frac{1}{1-q}} \quad (14)$$

Pour  $q = 1$ , on a :

$$A_q(\mathbf{P}) = \exp\left(\sum_{i=1}^n \sum_{j=1}^m -\frac{P_{ij}}{P_{++}} \log\left(\frac{P_{ij}}{P_{++}}\right) - \log(m)\right) \quad (15)$$

L'unité de mesure de l' $\alpha$ -diversité est un nombre effectif d'espèces à travers les communautés locales. Ainsi :  $1 \leq A_q(\mathbf{P}) \leq n$ . Dans le cadre de CHAO et CHIU, 2016, on a toujours :  $A_q(\mathbf{P}) \leq G_q(\mathbf{P})$ .

### $\beta$ -diversité

La  $\beta$ -diversité mesure l'hétérogénéité de composition des communautés locales à travers les sites. On a, pour  $q \in \mathbb{R}^+$  :

$$B_q(\mathbf{P}) = \frac{G_q(\mathbf{P})}{A_q(\mathbf{P})} \quad (16)$$

L'unité de mesure de la  $\beta$ -diversité est un nombre effectif de communautés. On a :  $1 \leq B_q(\mathbf{P}) \leq m$ . On a  $B_q(\mathbf{P}) = 1$  si toutes les communautés locales ont les mêmes abondances et  $B_q(\mathbf{P}) = m$  si aucune paire de communautés ne partage d'espèces en commun.

### Dissimilarité construite à partir de la $\beta$ -diversité

On peut construire des mesures de dissimilarités à partir de la  $\beta$ -diversité. On a :

$$\delta_q(\mathbf{P}) = \frac{1 - (B_q(\mathbf{P}))^{q-1}}{1 - (1/m)^{q-1}} \quad (17)$$

On a :  $0 \leq \delta_q(\mathbf{P}) \leq 1$ . Avec  $\delta_q(\mathbf{P}) = 0$  si toutes les communautés locales ont les mêmes abondances et  $\delta_q(\mathbf{P}) = 1$  si aucune paire de communauté ne partage d'espèce en commun.

Si  $q = 0$ , alors  $\delta_0$  est l'indice de Sorensen, qui quantifie la dissimilarité de composition entre deux communautés représentées par des vecteurs binaires (les abondances sont omises). Pour  $q = 1$ ,  $\delta_1$  est l'indice de dissimilarité de Horn.

## Comparaison de graphes

En informatique, les méthodes de comparaison de graphes sont connues sous le nom de 'network alignment problem', 'graph matching' et 'network comparison' (EMMERT-STREIB, DEHMER et SHI, 2016).

En écologie, POISOT et al., 2012 proposent des indices de dissimilarité pour les réseaux binaires (*i.e.* sans pondération des arêtes). Soient  $G_1 = (V_1, E_1)$  et  $G_2 = (V_2, E_2)$  deux graphes qui représentent des réseaux d'interaction.

La dissimilarité en terme de composition d'espèces,  $\beta_S$ , est définie en utilisant l'indice de Sorensen :

$$\beta_S = \frac{|V_1| + |V_2| - 2|V_1 \cap V_2|}{|V_1| + |V_2|} \quad (18)$$

La dissimilarité totale en terme de composition d'arêtes,  $\beta_{WN}$ , est définie comme l'indice de Sorensen sur les arêtes :

$$\beta_{WN} = \frac{|E_1| + |E_2| - 2|E_1 \cap E_2|}{|E_1| + |E_2|} \quad (19)$$

Néanmoins, une partie de la dissimilarité en terme de composition d'arêtes est due à la variation de composition en terme de nœuds (c.a.d d'espèces). On définit également la dissimilarité due à la plasticité des interactions biotiques,  $\beta_{OS}$  (*i.e.* en regardant uniquement les arêtes faisant intervenir des nœuds communs aux deux graphes). On note  $V_\cap = V_1 \cap V_2$  et  $G_1^\cap = (V_\cap, E_1^\cap)$ ,  $G_2^\cap = (V_\cap, E_2^\cap)$  les sous-graphes induits de  $G_1$  et  $G_2$  par  $V_\cap$

$$\beta_{OS} = \frac{|E_1^\cap| + |E_2^\cap| - 2|E_1^\cap \cap E_2^\cap|}{|E_1^\cap| + |E_2^\cap|} \quad (20)$$

De façon additive, on en déduit la dissimilarité due à la variation de composition d'espèces,  $\beta_{ST}$  :

$$\beta_{ST} = \beta_{WN} - \beta_{OS} \quad (21)$$

### 0.2.3 Modèles graphiques

Les modèles graphiques permettent de représenter, en utilisant un graphe, des relations d'indépendances conditionnelles entre des variables aléatoires. Il existe deux grandes familles de modèles suivant que le graphe soit dirigé ou non : les réseaux bayésiens et les champs de Markov. Tous deux ont été utilisés en écologie pour tenter d'inférer des associations statistiques, souvent interprétées comme des interactions entre espèces. Nous



renvoyons aux ouvrages de Murphy<sup>19</sup> et Koller & Friedman<sup>20</sup>.

## Réseaux bayésiens

Soit  $n$  variables aléatoires  $(X_1, \dots, X_n)$  (on note  $I = \{1, \dots, n\}$ )

Deux variables  $X_i$  and  $X_j$  sont indépendantes conditionnellement à  $X_{I \setminus \{i,j\}}$  si :

$$Pr(X_i, X_j | X_{I \setminus \{i,j\}}) = Pr(X_i | X_{I \setminus \{i,j\}}) Pr(X_j | X_{I \setminus \{i,j\}}) \quad (22)$$

On note alors :  $X_i \perp\!\!\!\perp X_j | X_{I \setminus \{i,j\}}$

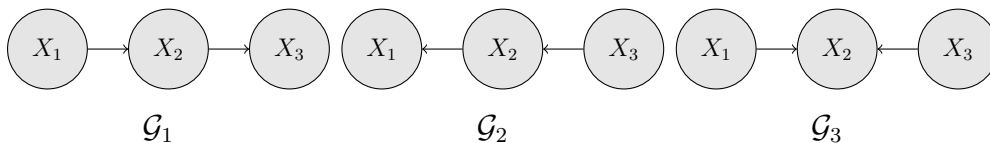
Les réseaux Bayésiens visent à représenter les relations de dépendances et d'indépendances conditionnelles entre les variables en utilisant un graphe acyclique orienté,  $\mathcal{G}$ . On note  $V = \{1, \dots, n\}$  l'ensemble des nœuds de  $\mathcal{G}$  et  $Par_{\mathcal{G}}(u)$  les parents de  $u$ . La probabilité jointe se factorise alors selon  $\mathcal{G}$  de la façon suivante :

$$Pr(X_1, \dots, X_n) = \prod_i Pr(X_i | Pa_{\mathcal{G}}(i)) \quad (23)$$

Les paramètres du modèle sont alors le graphe et les tables de probabilités conditionnelles. On peut alors inférer la structure de  $\mathcal{G}$  avec divers algorithmes. Par exemple en effectuant des test d'indépendances conditionnelles ou utiliser une heuristique pour parcourir l'espace des graphes en utilisant un score. Ces méthodes sont implémentées dans le paquet R 'bnlearn'.

Une difficulté d'interprétation de la structure des réseaux bayésiens provient du fait que plusieurs DAGs peuvent représenter les mêmes indépendances conditionnelles.

Considérons d'abord un exemple comprenant 3 DAGs à 3 nœuds :



$\mathcal{G}_1$  et  $\mathcal{G}_2$  peuvent représenter, par exemple, une chaîne trophique à trois espèces. Toutefois, si on interprète les arêtes du réseaux comme causales, on a une causalité 'top-down' et une causalité 'bottom-up'. Néanmoins, les indépendances conditionnelles formulées par ces deux DAGs sont les mêmes. En effet, dans les deux cas, on a :

$$X_1 \perp\!\!\!\perp X_3 | X_2$$

$\mathcal{G}_1$  et  $\mathcal{G}_2$  sont donc équivalents au sens où ils représentent les mêmes indépendances conditionnelles. Ce n'est pas le cas de  $\mathcal{G}_3$ , qui peut représenter deux espèces en compétition

19. MURPHY, 2012.

20. KOLLER et FRIEDMAN, 2009.

pour une ressource. Dans ce cas,  $X_1$  et  $X_3$  ne sont pas indépendantes conditionnellement à  $X_2$ . Pour s'en convaincre, imaginons que  $X_1$ ,  $X_2$  et  $X_3$  sont des variables binaires.  $X_1$  et  $X_3$  représentent la présence de consommateurs et  $X_2$  représente la ressource. On suppose que si la ressource est présente, alors les espèces associées à  $X_1$  et  $X_3$  sont présentes. Alors, dans les sites où la ressource est présente, on observera les deux espèces de consommateurs. Ainsi, la présence des deux espèces de consommateur n'est pas indépendante conditionnellement à la présence de la ressource. La structure de DAG de  $\mathcal{G}_3$  s'appelle un  $v$ -structure. VERMA et PEARL, 1990 fournissent une caractérisation des structures de réseaux bayésiens qui sont équivalentes.

**Théorème 0.2.14 (Pearl & Verma, 1990)** *Deux structures de réseau bayésiens sont équivalentes si et seulement si elles ont le même squelette (i.e. les mêmes arêtes sans tenir compte de l'orientation) et les mêmes  $v$ -structures.*

## Champs de Markov

Les champs de Markov sont les cousins non-orientés des réseaux bayésiens. Dans ces modèles graphiques, deux variables aléatoires sont indépendantes conditionnellement si les nœuds qui les représentent n'ont pas d'arête qui les relie (propriété de Markov locale). Pour  $n$  variables aléatoires  $(X_1, \dots, X_n)$ , la distribution jointe  $Pr(\mathbf{X})$  ne se représente pas avec des conditionnements comme dans le cas des réseaux bayésiens. Soit  $G$  un graphe non-dirigé à  $n$  nœuds. Alors la distribution  $Pr(\mathbf{X})$  satisfait les indépendances conditionnelles formulées par  $G$  si elle s'écrit (théorème de Hammersley-Clifford) :

$$Pr(\mathbf{X}|\Theta) = \frac{1}{Z(\Theta)} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{X}_c|\Theta_c) \quad (24)$$

où  $\phi_c$  est la fonction de potentiel,  $\Theta$  est le paramètre du modèle,  $\mathcal{C}$  est l'ensemble des cliques de  $G$  (i.e. l'ensemble des sous-graphes induits complets) et  $Z(\Theta)$  est la fonction de partition :

$$Z(\Theta) = \sum_{\mathbf{X}} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{X}_c|\Theta_c) \quad (25)$$

On peut alors construire différents champs de Markov. Par exemple, pour un champ de Markov gaussien dont le support est  $G = (V, E)$  on a :

$$Pr(\mathbf{X}|\Theta) = \prod_{\{s,t\} \in E} \phi_{st}(x_s, x_t) \prod_t \phi_t(x_t) \quad (26)$$

où

$$\phi_{st}(x_s, x_t) = \exp\left(-\frac{1}{2}x_s \Lambda_{st} x_t\right) \quad (27)$$

et

$$\phi_t(x_t) = \exp\left(-\frac{1}{2}x_t^2 \Lambda_{tt} + \eta_t y_t\right) \quad (28)$$

On a alors :

$$Pr(\mathbf{X}|\Theta) = \exp(\eta^T \mathbf{X} - \frac{1}{2} \mathbf{X}^T \mathbf{\Lambda} \mathbf{X}) \quad (29)$$

On reconnaît alors la densité de probabilité d'une loi normale  $\mathcal{N}(\mu, \Sigma)$  avec  $\mathbf{\Lambda} = \Sigma^{-1}$ .

On peut inférer des réseau de corrélations partielles (dont la matrice de précision, l'inverse de la matrice de variance-covariance, est creuse) à partir de données en utilisant le graphical lasso (FRIEDMAN, HASTIE et TIBSHIRANI, 2008). Cette méthode est implémentée dans le paquet R 'glasso'.

PREMIÈRE PARTIE

# Corps principal de la thèse

---



# Productions académiques

---

Cette thèse s'appuie sur les articles suivants, publiés ou en préparation :

**En lien avec le chapitre 1** (méthodes et applications pour la biogéographie des réseaux d'interaction) :

- **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.
- O'Connor, L., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., Montemaggiore, A., **Ohlmann, M.** Thuiller, W. Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *Journal of Biogeography* (*in press*)

**En lien avec le chapitre 2** (théorie des métacommunautés) :

- **Ohlmann, M.**, Munoz, F., Laroche, F., Thuiller, W. A spatial theory of mutualistic metacommunities : assesment of metacommunity capacity (*in prep*)

**En lien avec le chapitre 3** (Dépendances des patrons de diversité de groupes trophiques) :

- **Ohlmann, M.**, Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L.,... Thuiller, W. (2018). Mapping the imprint of biotic interactions on  $\beta$ -diversity. *Ecology letters*, 21(11), 1660-1669.
- Martinez-Almoyna, C., Thuiller, W., Chalmandrier, L., Clément, J.C., Foulquier, A., **Ohlmann, M.**, Zinger, L. Münkemüller, T. Multi-trophic -diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions. *Funct Ecol.* 2019; 00 : 1– 12.

**En lien avec le chapitre 4** (mélange de marqueurs pour le metabarcoding) :

- **Ohlmann, M.**, Robin, S., Boyer, F., Thuiller, W. Combining abundances from different markers for environmental DNA (*in prep*)



# Indices de diversité pour les réseaux écologiques, un cadre unifié avec les nombres de Hill

---

Les symboles n'ont pas une signification intrinsèque et invariable ; ils ne sont pas autonomes vis-à-vis du contexte. Leur signification est d'abord de position.

---

Claude Levi-Strauss, *Mythologiques 1 : Le cru et le cuit*

## 1.1 Synopsis

Les indices de diversité sont des outils qui permettent d'appréhender quantitativement le concept de diversité en écologie. Ils sont des transformations des entropies de Shannon et de Rényi. Dans le cas des méta-communautés, les indices de diversité permettent de mesurer la diversité globale ( $\gamma$ -diversité), la moyenne des diversités locales ( $\alpha$ -diversité) et la dissimilarité entre les communautés ( $\beta$ -diversité). Ils satisfont des propriétés mathématiques qui permettent aux mesures de diversité d'être cohérentes. Toutefois, ils représentent une méta-communauté comme une matrice ; c'est à dire que chaque communauté est représentée par une collection d'espèces avec des abondances. Une représentation plus complète des communautés d'espèces consiste en une collection d'espèces et les interactions entre ces espèces ; c'est-à-dire un réseau d'interaction. Les interactions (et donc le réseau) peuvent varier dans l'espace et dans le temps. Ainsi, des méthodes de comparaison de réseau permettent de quantifier ces variations de structure. Toutefois, elles ne sont pas directement reliées à la théorie des nombres de Hill. Cette partie propose de lier indices de diversité et méthodes de comparaison de réseaux. On s'intéressera au cas général, dans lequel les espèces ont une abondance et les interactions une certaine probabilité et l'on proposera de nouveaux indices de diversité pour les réseaux. On montrera que les indices de diversité qui ont été préalablement développés sont des cas particuliers du cadre développé ici. On s'intéressera également à l'agrégation des nœuds (les espèces) en groupes de nœuds (qui peuvent par exemple représenter des niches d'Elton) et au développement d'indices de diversité permettant d'effectuer des analyses à plusieurs niveaux



d'agrégation dans le réseau. A travers une simulation et une application sur des données libres de réseaux trophiques, nous montrerons comment ces indices peuvent être utilisés et comment les facteurs qui influencent la diversité de réseau dépendent de l'échelle d'agrégation considérée. Ce chapitre a fait l'objet d'une publication dans *Ecology Letters* (section 'methods') :

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.

Un paquet R, 'econetwork', a également été développé en lien avec l'ANR 'econet'.

## 1.2 Corps principal de l'article (en anglais)

## METHOD

# Diversity indices for ecological networks: a unifying framework using Hill numbers

Marc Ohlmann,<sup>1\*</sup>   
 Vincent Miele,<sup>2,3,4</sup>  
 Stéphane Dray,<sup>2,3,4</sup>  
 Loïc Chalmardrier<sup>5</sup>  
 Louise O'Connor,<sup>6</sup> and  
 Wilfried Thuiller<sup>1</sup> 

### Abstract

Describing how ecological interactions change over space and time and how they are shaped by environmental conditions is crucial to understand and predict ecosystem trajectories. However, it requires having an appropriate framework to measure network diversity locally, regionally and between samples ( $\alpha$ -,  $\gamma$ - and  $\beta$ -diversity). Here, we propose a unifying framework that builds on Hill numbers and accounts both for the probabilistic nature of biotic interactions and the abundances of species or groups. We emphasise the importance of analysing network diversity across different species aggregation levels (e.g. from species to trophic groups) to get a better understanding of network structure. We illustrate our framework with a simulation experiment and an empirical analysis using a global food-web database. We discuss further usages of the framework and show how it responds to recent calls on comparing ecological networks and analysing their variation across environmental gradients and time.

### Keywords

Hill numbers, interaction network diversity, Metanetwork, species aggregation level.

Ecology Letters (2019)

## INTRODUCTION

Since the pioneering work of Humboldt (von Humboldt 1805), understanding the patterns of biodiversity across space and time has been a question central to both biogeography and community ecology (Gaston 2003). The recent upsurge of large-scale databases has made possible to produce comprehensive syntheses of biodiversity patterns (Belmaker *et al.* 2012; Mazel *et al.* 2017) by analysing local assemblages on the one hand ( $\alpha$ -diversity, Hawkins *et al.* 2003). and composition turnover between such assemblages on the other ( $\beta$ -diversity, Mazel *et al.* 2017). A plethora of diversity indices and unifying frameworks have thus been proposed to partition biodiversity into  $\alpha$ - and  $\beta$ -diversity components (Whittaker 1960; Routledge 1979; Ellison 2010; Chao *et al.* 2014b; Chao & Chiu 2016). However, not only does biodiversity reflect species coexistence but also the trophic and non-trophic interactions that link them to one another (Kéfi *et al.* 2016). The development of the trophic theory of island biogeography (Gravel *et al.* 2011; Massol *et al.* 2017) has recently paved the way for a new biogeography synthesis by accounting for trophic interactions in theoretical predictions of biodiversity patterns. Similarly, empiricists do not only investigate species distribution patterns but also analyse how ecological interactions (i.e. ecological networks) vary over space and time (Pellissier *et al.* 2017; Tylaniakis & Morris 2017). To this aim, the *metanetwork* concept generalises the regional species-pool of classic community ecology by adding to this representation of biodiversity the potential trophic and non-trophic interactions

between species (Dunne 2006; Kéfi *et al.* 2016) at a regional scale. Thus, in the same way local assemblages are conceptualised as subsets of a regional species pool, local ecological networks are realisations of a subset of the regional metanetwork. This opens new perspectives in understanding the processes that shape the distribution of biodiversity in space and time. For instance, mapping, describing and comparing ecological networks along environmental or disturbance gradients are the first steps of a fascinating era to understand the organisation of life on Earth (Pellissier *et al.* 2017) and its effects on ecosystem functioning and associated services (Brose & Hillebrand 2016). The realisation, the frequency and the intensity of interactions within networks across space and time are driven by the compositional turnover of species or groups of species, changes in their abundances, their plasticity or behavioural variations, and finally by the environmental constraints on biotic interactions. Any of these variations may have direct or indirect consequences on ecosystem functioning (Barnes *et al.* 2014). Such knowledge would thus help not only to improve our understanding of multi-trophic assemblages and their influence on ecosystem functioning but also to help build a more robust predictive ecology at the interface between trophic ecology, community ecology and ecosystem ecology (Thompson *et al.* 2012).

There is thus a strong need to develop a framework to understand the structure and composition of ecological networks across spatial and temporal scales and along environmental gradients (Pellissier *et al.* 2017). To date, such a framework remains hampered by several issues.

<sup>1</sup>Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Écologie Alpine, F-38000 Grenoble, France

<sup>2</sup>Université de Lyon, F-69000 Lyon, France

<sup>3</sup>Université Lyon 1, F-69007 Lyon, France

<sup>4</sup>CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France

<sup>5</sup>Department of Botany, University of Wyoming, Laramie, WY 82071, USA

<sup>6</sup>École Normale Supérieure, F-75005 Paris, France

\*Correspondence: E-mail: marc.ohlmann@univ-grenoble-alpes.fr

First, no appropriate diversity measure is available to describe the diversity of ecological networks, and partition it into  $\alpha$ ,  $\beta$  and  $\gamma$  components, that would account for both species abundances and the probabilistic nature of interactions, and that would relate to existing frameworks in biogeography or community ecology (Pellissier *et al.* 2017). Recent years have seen a prolific development of frameworks to measure diversity at both the taxonomic (Jost 2007; Ellison 2010; Chao & Chiu 2016) and phylogenetic or trait levels (Chao *et al.* 2014a; Tucker *et al.* 2016). These indices need to satisfy four mathematical properties (see Jost 2010): (1)  $\alpha$  and  $\beta$  should be mathematically unrelated; (2)  $\alpha$ ,  $\beta$  and  $\gamma$  should be effective numbers (this enables to interpret a given measure of diversity in terms of the diversity of an evenly distributed community and therefore guarantees the comparability of diversity measures); (3)  $\gamma$  should be completely determined by  $\alpha$  and  $\beta$ ; (4)  $\alpha$  cannot be larger than  $\gamma$ . A fifth additional practical property of  $\beta$ -diversity, *invariance under shattering* (Reeve *et al.* 2014) assumes that each community represents a portion of the geographical space. This assumption implies that if a community is split into two and the abundances of the two resulting communities are equal, then the  $\beta$ -diversity of the overall metacommunity should not change. The framework recently proposed by Reeve *et al.* (2014) satisfies each of these fundamental properties (only when similarities between species are not considered) while Jost's framework satisfies the first four properties, and the fifth only for some particular cases. These indices are based on Hill numbers (Hill 1973), which are derived from Rényi's entropy (Rényi 1961) and have enabled a generalisation of the well-established diversity measures such as the Shannon entropy or the Simpson diversity index. An additional and interesting feature of Hill numbers is the introduction of a *viewpoint parameter* linked to the weight given to dominant vs. rare species onto assembly rules (Chalmandrier *et al.* 2015). While this framework could potentially be very useful for ecological networks, it is not yet applicable. So far, the few network-specific metrics are built on graph theory, with the aim to summarise the structure of a network through a single quantity (Poisot *et al.* 2012, 2016; Pellissier *et al.* 2017) – but none of them satisfy the five requirements listed above, nor are they able to manipulate species abundances or the probability of a given interaction occurring.

Second, diversity metrics depend on the way individuals are aggregated into larger groups (e.g. species, guilds, functional groups). In trophic networks, species can be aggregated based on their equivalent roles. Indeed, species richness or taxonomic turnover do not reveal much on how assemblages are truly structured in terms of resource exploitation, niche partitioning and co-existence mechanisms (Thompson *et al.* 2012), whereas functional or trophic groups enable to encapsulate more of the underlying ecological processes. In this respect, insights from graph theory (Luczkovich *et al.* 2003) and random models of networks (Group model, Allesina & Pascual 2009 or stochastic block model, Newman & Leicht 2007) have helped to mathematically formalise equivalence relationships between species using the topology of the ecological network alone. However, aggregating species in trophic or functional groups is often challenging (e.g. choosing an optimal number of groups) and there is no way of knowing whether a given grouping will provide a better

understanding of the diversity pattern than others. To address this issue, we believe it is necessary to describe and analyze patterns at different levels of species aggregation, as is now done in community phylogenetic analyses (Chalmandrier *et al.* 2015; Graham *et al.* 2018). In fact, analyses of ecological networks diversity ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) should be carried out along a profile of species aggregation levels, ranging from characterising all species (their abundances, the abundance of their links and their probability of interactions, hereafter named the *microscopic scale*), to various species aggregation levels (hereafter coined the *mesoscopic scale*) until the *macroscopic scale*, represented by the single value of connectance (the probability of interaction between any two species). This multi-scale approach should provide novel insights to understand the processes that shape ecological networks. To overcome these limitations, we introduce a novel framework that allows measuring  $\alpha$ ,  $\beta$ , and  $\gamma$  diversities of ecological networks and combines Hill numbers (Hill 1973; Jost 2006; Chao & Chiu 2016) with different species aggregation levels. First, we build on the existing mathematical frameworks to derive new indices for ecological network diversity, which we partition into  $\alpha$ ,  $\beta$ , and  $\gamma$  components. We then demonstrate that existing network diversity indices (Bersier *et al.* 2002; Poisot *et al.* 2012, 2016) are particular cases of the proposed unified framework. We further extend our framework so it can be used across multiple levels of species aggregation. We then apply this framework to an intercontinental data set of stream water trophic networks (Thompson & Townsend 2003) and show that the drivers of the dissimilarity of ecological networks vary with the level of species aggregation. We finally provide an implementation of the framework in the R package *econetwork* available on CRAN (<https://cran.r-project.org>), guidelines for the interpretation of the results, and recommendations for the analyses of networks across space and time.

## DIVERSITY FOR A SINGLE NETWORK

For the sake of simplicity, we first introduce the formalism behind our new indices with a single trophic network. We then generalise the framework to the case of a metanetwork, and provide the details and mathematical proofs in the Supporting Information.

### A probabilistic model of interaction networks

We propose a generic model of ecological networks that considers both species abundances and the probabilities of interaction between species. It is an extension of the probabilistic network model (Poisot *et al.* 2016) with the additional property that it accounts for species abundances.

We consider a given region that contains individuals belonging to  $n$  different species with relative abundances  $p = (p_1, \dots, p_n)$ .  $p_q$  represents the probability of picking an individual of species  $q$ . We also assume that the probability of interaction between two individuals of species  $q$  and  $l$  follows a Bernoulli law of parameter  $\pi_{ql}$ . This allows to account for the potential variability on the realisation of an interaction event at the individual level (Albert *et al.* 2010; Gonzalez-Varo & Traveset 2016). We also assume that all interactions occur independently. We represent this regional model using a

weighted network  $G$ , with  $p_q$  the relative abundance of the node  $V_q$ , and  $\pi_{ql}$  the weight of the link  $(V_q, V_l)$  ( $\mathbf{\Pi}$  is the adjacency matrix of  $G$ , see Table 1). The probability of picking a link that connects two individuals of species  $q$  and  $l$  is thus:

$$L_{ql} = \Pr(i \rightarrow j, i \in q, j \in l) \quad (1)$$

$$L_{ql} = \pi_{ql} p_q p_l \quad (2)$$

where

$$\pi_{ql} = \Pr(i \rightarrow j | i \in q, j \in l) \quad (3)$$

If this model represents the most complete case of a single network (abundances on nodes and weights on links), simpler cases can easily be derived by omitting the weights on links (i.e. for binary networks,  $\pi_{ql}$  is either 0 or 1) or the abundances of nodes (i.e. assuming evenly distributed species abundances).

### Navigating across species aggregation levels

We initially described the probabilistic network model at a species level. However, species can have similar positions, roles (e.g. Eltonian niche, Elton 1927) or functions (Lindeman 1942; Lavorel & Garnier 2002; Luck *et al.* 2012), leading to inflated or deflated estimates with respect to functional diversity. It is thus crucial to represent and analyze the diversity of ecological networks at different aggregation levels, by grouping species into larger and more relevant entities. Here, we propose to use mathematical methods that group nodes using the topology of the ecological network without any prior knowledge (see Supporting Information for a brief review of these methods).

Assuming that we have established  $Q$  groups ( $C_1, \dots, C_Q$ ) from the previous network ( $Q \leq n$ ), we can represent the network at a coarser resolution (Fig. 1, mesoscopic scale), called the image network (Luczkovich *et al.* 2003; Allesina & Pascual 2009). The new set of nodes is  $\tilde{V} = (\tilde{V}_1, \dots, \tilde{V}_Q)$  and each node is assigned a weight  $\tilde{p}_q$  that corresponds to the abundance of the group  $q$ .

$$\tilde{p}_q = \sum_{k \in C_q} p_k \quad (4)$$

Similarly, each link of the image network is assigned a weight  $\tilde{\pi}_{ql}$  that corresponds to the probability of interaction between individuals from classes  $C_q$  and  $C_l$ .

$$\tilde{\pi}_{ql} = \frac{\sum_{k \in C_q, k' \in C_l} \pi_{kk'} p_k p_{k'}}{\sum_{k \in C_q} p_k \sum_{k' \in C_l} p_{k'}} \quad (5)$$

The link abundances between individuals of classes  $q$  and  $l$ ,  $\tilde{L}_{ql}$  thus equates to:

$$\tilde{L}_{ql} = \sum_{k \in C_q, k' \in C_l} \pi_{kk'} p_k p_{k'} \quad (6)$$

We thus define the scale of the image network considered as:

$$s = \frac{Q}{n} \quad (7)$$

If  $s = 1$ , the network is considered at a microscopic scale (the image network corresponds to the original one). If  $s = \frac{1}{n}$ , the network is considered at a macroscopic scale. In this case, the image network is then made of a single node (with

**Table 1** Notations, name of the different indices and ranges of values

Object	Name	Total margin
$\mathbf{p}$	Vector of relative group abundances	$\sum_{1 \leq q \leq Q} p_q = 1$
$\mathbf{L}$	Matrix of link abundances	$\sum_{1 \leq q, l \leq Q} L_{ql} = C$
$\mathbf{\Pi}$	Matrix of link probabilities (adjacency matrix of the weighted network)	$\sum_{1 \leq q, l \leq Q} \pi_{ql}$
$\mathbf{P}$	Matrix of group abundances of groups (metanetwork case)	$\sum_{q=1}^Q \sum_{k=1}^K P_{qk} = K$
$\mathbf{L}$	Tensor of links abundances (metanetwork case)	$\sum_{q,l=1}^Q \sum_{k=1}^K L_{qlk} = \sum_{k=1}^K C_k$
$\mathbf{\Pi}$	Tensor of link probabilities (metanetwork case)	$\sum_{q,l=1}^Q \sum_{k=1}^K \pi_{qlk}$

	Diversity index	Name & interpretation	Range	
$\alpha$ -diversity	$A_p^\alpha$	Overall $\alpha$ -diversity in group abundances		
		Average diversity in group abundances across local networks	$1 \leq A_p^\alpha \leq Q$	
	$A_L^\alpha$	Overall $\alpha$ -diversity in link abundances		
		Average diversity in link abundances across local networks	$1 \leq A_L^\alpha \leq N_L$	
$\gamma$ -diversity	$A_\Pi^\alpha$	Overall $\alpha$ -diversity in link probabilities		
		Average diversity in link probabilities across local network	$1 \leq A_\Pi^\alpha \leq N_L$	
	$G_p^\gamma$	$\gamma$ -diversity in group abundances		
		Diversity in group abundances of the metanetwork	$1 \leq G_p^\gamma \leq Q$	
$\beta$ -diversity	$G_L^\gamma$	$\gamma$ -diversity in link abundances		
		Diversity in link abundances of the metanetwork	$1 \leq G_L^\gamma \leq N_L$	
	$G_\Pi^\gamma$	$\gamma$ -diversity in link probabilities		
		Diversity in link probabilities of the metanetwork	$1 \leq G_\Pi^\gamma \leq N_L$	
$\beta$ -diversity	$B_p^\beta$	$\beta$ -diversity of group abundances		
		Effective numbers of distinct communities of groups	$1 \leq B_p^\beta \leq K$	
	$B_L^\beta$	$\beta$ -diversity of link abundances		
		Effective numbers of distinct networks	$1 \leq B_L^\beta \leq K$	
$\beta$ -diversity	$B_\Pi^\beta$	$\beta$ -diversity of link probabilities		
		Effective numbers of distinct networks (with abundances rescaled at evenly distributed values)	$1 \leq B_\Pi^\beta \leq K$	
	dissimilarity	$\delta_p^\beta$	Dissimilarity of group abundances	
			Effective average proportion of shared groups	$0 \leq \delta_p^\beta \leq 1$
$\delta_L^\beta$		Dissimilarity of link abundances		
		Effective average proportion of shared links	$0 \leq \delta_L^\beta \leq 1$	
dissimilarity	$\delta_\Pi^\beta$	Dissimilarity of link probabilities		
		Effective average proportion of shared links (with abundances rescaled at evenly distributed values)	$0 \leq \delta_\Pi^\beta \leq 1$	

Notes.  $C$  is the connectance of the considered network,  $Q$  is the number of groups of the considered metanetwork and  $N_L$  its number of different links and  $C_k$  is the connectance of the local network  $k$ .

abundance 1, the sum of species relative abundances) and a single link. The weight of this link represents the probability that any two nodes of the original networks are connected and is, consequently, the connectance of the original network.

$$C = \sum_{1 \leq q, l \leq n} \pi_{ql} p_q p_l \quad (8)$$

If  $\frac{1}{n} < s < 1$ , the network is considered at a mesoscopic scale.

#### Measuring diversity at different species aggregation levels

For a community vector  $\mathbf{p} = (p_1, \dots, p_Q)$ , the Hill number of order  $\eta$  is defined as:

$$D^\eta(\mathbf{p}) = \left( \sum_{i=1}^Q p_i^\eta \right)^{\frac{1}{1-\eta}}, \quad \eta \geq 0, \eta \neq 1 \quad (9)$$

This number ranges between 1 and  $Q$  (Table 1), and translates into an effective number of groups (which can be species or group of species i.e., we define diversity indices on the image network while keeping the notations of the original one for the sake of simplicity). A Hill measure of  $\Delta$  hence means that the system holds a diversity equivalent to a system made of  $\Delta$  equally distributed groups.  $\eta$  is considered as a viewpoint parameter that modulates the weight given to group abundances. When  $\eta = 0$ , all groups equally contribute to the

index and  $D^0$  is the richness of groups. For  $\eta = 1$ , eqn 8 is not defined but it converges towards the exponential of the Shannon entropy :

$$D^1(\mathbf{p}) = \lim_{\eta \rightarrow 1} D^\eta(\mathbf{p}) = \exp \left( \sum_{i=1}^Q -p_i \log p_i \right) \quad (10)$$

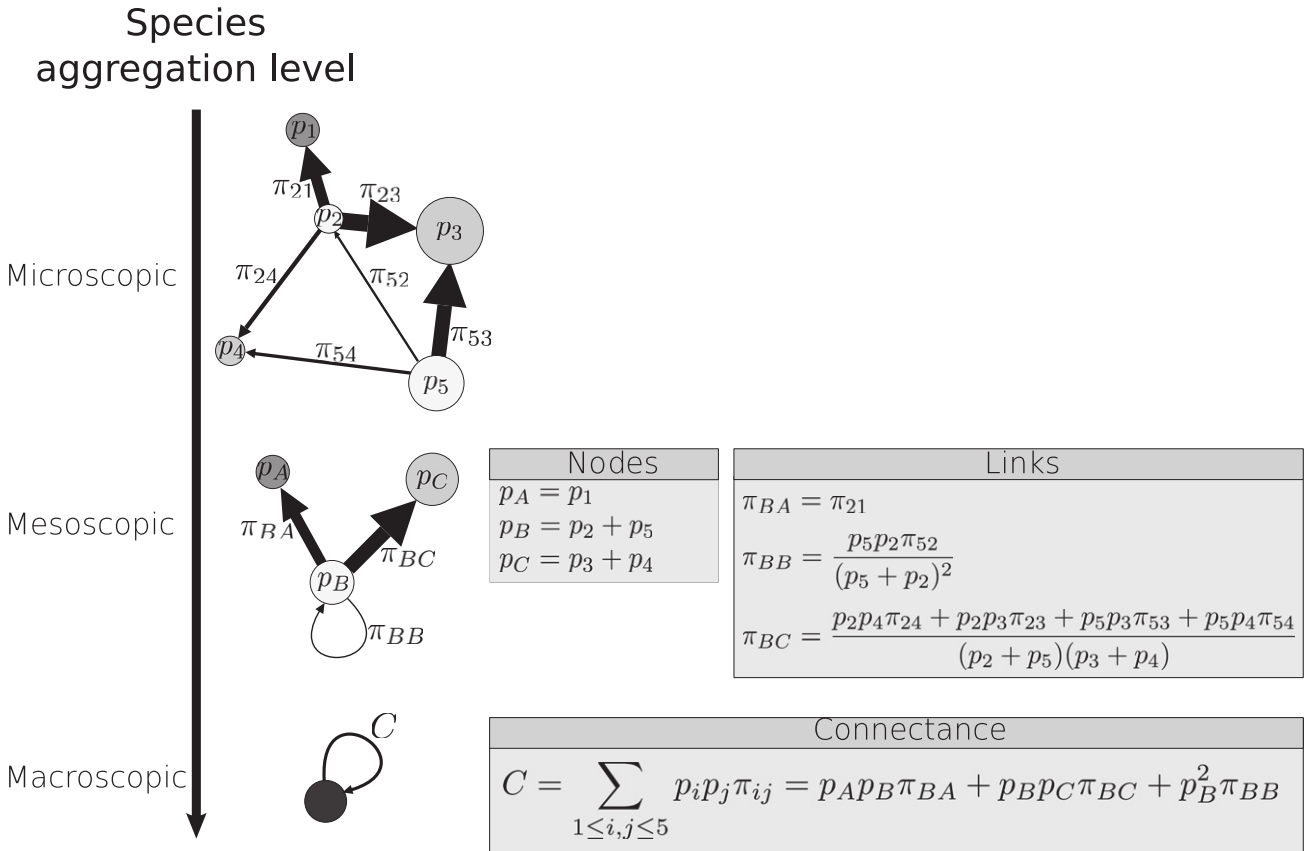
We propose to extend the use of Hill numbers to compute the diversity in link abundances and the diversity in link probabilities between groups. More precisely, we measure the entropy of the random variable associated to the experience: ‘A link is drawn uniformly in the network, what is the label of this link (the label is defined by the identity of the two groups that are connected by the link)’. Assuming that  $\mathbf{L}$  is the matrix of link abundances, the diversity in link abundances is:

$$D^\eta(\mathbf{L}) = \left( \sum_{1 \leq q, l \leq Q} \left( \frac{L_{ql}}{C} \right)^\eta \right)^{\frac{1}{1-\eta}} \quad (11)$$

where  $C = \sum_{1 \leq q, l \leq Q} L_{ql}$ . Similarly, assuming that  $\mathbf{\Pi}$  is the adjacency matrix of the image graph, the diversity in link probabilities is defined as:

$$D^\eta(\mathbf{\Pi}) = \left( \sum_{1 \leq q, l \leq Q} \left( \frac{\pi_{ql}}{\pi_{++}} \right)^\eta \right)^{\frac{1}{1-\eta}} \quad (12)$$

where  $\pi_{++} = \sum_{1 \leq q, l \leq Q} \pi_{ql}$ .



**Figure 1** Navigating through species aggregation levels. From the original weighted network to image networks at mesoscopic and macroscopic scales, with the formulas giving the group abundances and link probabilities of the image networks.

$D^\eta(\mathbf{\Pi})$  is unrelated to  $D^\eta(\mathbf{p})$  (*sensu* Chao & Chiu 2016), since the value of one of this measure does not constrain the value of the other. When  $\eta = 0$ , it measures the number of links of the image network. When  $\eta \rightarrow 1$ , it converges towards the exponential of the Shannon entropy of the probability of links (Bersier *et al.* 2002)

Both group abundances and the interaction probabilities determine the range of values of  $D^\eta(\mathbf{L})$  which are therefore related. These last two indices translate into an effective numbers of links, either weighted ( $D^\eta(\mathbf{L})$ ) or not ( $D^\eta(\mathbf{\Pi})$ ) by the group abundances. Note that ( $D^\eta(\mathbf{\Pi})$ ) could be used in studies where groups have different relative abundances, but these are not important in the analysis.

## DIVERSITY FOR A METANETWORK

### Measuring $\alpha$ -, $\beta$ - and $\gamma$ -diversity at different species aggregation levels

Mirroring the single network case, we propose to analyze the diversity of the metanetwork and its local realisations through different species aggregation levels. Importantly, we assume that any grouping is defined on the metanetwork. We thus define  $\alpha$ -,  $\beta$ - and  $\gamma$ -diversity measures on the set of local networks and on the metanetwork at different species aggregation levels. We measure the diversity of group abundances, link abundances and link probabilities using Hill numbers. We extend the framework presented in Chao & Chiu 2016 since it satisfies the first four properties listed in the introduction and elegantly link the variance and decomposition perspective on  $\beta$ -diversity (see Chao & Chiu 2016 for details). For the sake of simplicity, we present the case  $\eta \rightarrow 1$  (and therefore omit the exponent in the indices). The general case is presented in Supporting Information, together with our framework as an extension of Jost's and Reeve's framework and the mathematical links between the existing network diversity indices and the proposed unified framework. The proposed indices can be applied in several subcases (Fig. 2). We use the same probabilistic network model as presented before. The metanetwork is thus a weighted network, divided in  $K$  local networks (see Table 1 for notations and total margins).

### $\alpha$ -diversity

For each local network, the  $\alpha$  diversity is computed using Hill numbers (for  $\eta \rightarrow 1$ , it converges towards Shannon entropy). The overall  $\alpha$ -diversity of groups across local networks is:

$$A_P = \exp\left(\sum_{q=1}^Q \sum_{k=1}^K -\frac{P_{qk}}{P_{++}} \log\left(\frac{P_{qk}}{P_{++}}\right) - \log(K)\right) \quad (13)$$

where  $P_{++} = \sum_{1 \leq k \leq K} \sum_{1 \leq q \leq Q} P_{qk}$ .

This is the mean equivalent number of groups across local networks. Similarly, the overall  $\alpha$ -diversities in link abundances and link probabilities are equal to:

$$A_L = \exp\left(\sum_{q,l=1}^Q \sum_{k=1}^K -\frac{L_{qlk}}{L_{++++}} \log\left(\frac{L_{qlk}}{L_{++++}}\right) - \log(K)\right) \quad (14)$$

where  $L_{++++} = \sum_{1 \leq k \leq K} \sum_{1 \leq q,l \leq Q} L_{qlk}$ .

$$A_\Pi = \exp\left(\sum_{q,l=1}^Q \sum_{k=1}^K -\frac{\pi_{qlk}}{\pi_{++++}} \log\left(\frac{\pi_{qlk}}{\pi_{++++}}\right) - \log(K)\right) \quad (15)$$

where  $\pi_{++++} = \sum_{1 \leq k \leq K} \sum_{1 \leq q,l \leq Q} \pi_{qlk}$ .

### $\gamma$ -diversity

The  $\gamma$ -diversity of group abundances is defined as:

$$G_P = \exp\left(\sum_{q=1}^Q -\frac{P_{q+}}{P_{++}} \log\left(\frac{P_{q+}}{P_{++}}\right)\right) \quad (16)$$

where  $P_{q+} = \sum_{1 \leq k \leq K} P_{qk}$ . This corresponds to the equivalent number of groups in the metanetwork. The  $\gamma$ -diversity of the link abundances is defined as:

$$G_L = \exp\left(\sum_{q,l=1}^Q -\frac{L_{ql+}}{L_{++++}} \log\left(\frac{L_{ql+}}{L_{++++}}\right)\right) \quad (17)$$

where  $L_{ql+} = \sum_{1 \leq k \leq K} L_{qlk}$ .

This corresponds to the equivalent number of links in the metanetwork. The  $\gamma$ -diversity in link probabilities is defined as:

$$G_\Pi = \exp\left(\sum_{q,l=1}^Q -\frac{\pi_{ql+}}{\pi_{++++}} \log\left(\frac{\pi_{ql+}}{\pi_{++++}}\right)\right) \quad (18)$$

where  $\pi_{ql+} = \sum_{1 \leq k \leq K} \pi_{qlk}$ . This corresponds to the equivalent number of links in a network that contains the same probabilities of links as in the metanetwork, but where the relative abundances of groups are arbitrarily considered evenly distributed.

### $\beta$ -diversity and dissimilarity measures

The overall  $\beta$ -diversity can be calculated in group abundances, link abundances and link probabilities. The  $\beta$ -diversity in groups abundance is equal to:

$$B_P = \frac{G_P}{A_P} \quad (19)$$

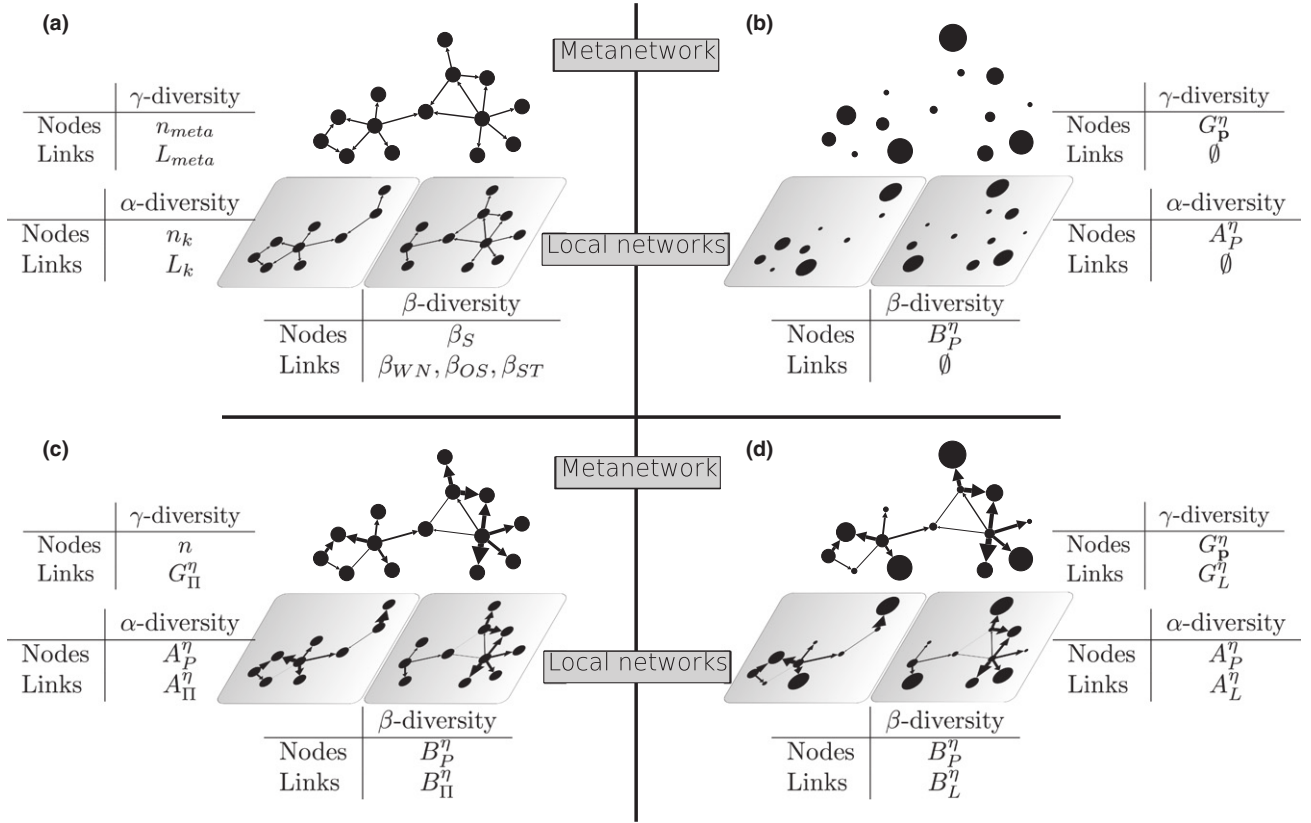
This is the effective number of equally large and completely distinct communities of groups. It represents how many completely distinct communities of groups are present in the set of networks. The  $\beta$ -diversity in link abundances is equal to:

$$B_L = \frac{G_L}{A_L} \quad (20)$$

This is the effective number of equally large and completely distinct networks i.e., the number of networks made of distinct links across the considered region. The  $\beta$ -diversity in link probabilities is equal to:

$$B_\Pi = \frac{G_\Pi}{A_\Pi} \quad (21)$$

This translates to an effective number of equally large and completely distinct networks where group abundances would have arbitrarily been considered equal.



**Figure 2** The metanetwork and the local realised networks in different cases: (a) binary network, unweighted links and without node abundances, (b) node abundances but absence of links, (c) weights on links but no node abundances and (d) weights and links and node abundances. The different indices to measure  $\alpha$ -,  $\beta$ - and  $\gamma$ -diversity are associated to each particular case and presented more generally in Table 1.

Overlap measures can be built from  $\beta$ -diversity to obtain dissimilarity measures (Jost 2007, 2010; Chao & Chiu 2016). A class of parameterised Sorensen's based dissimilarity measures can be defined as non-linear transformation of  $\beta$ -diversity. When  $\eta = 1$ , it equals to the Horn dissimilarity index (the general case is presented in Supporting Information).

$$\delta_P = \frac{\log(G_P) - \log(A_P)}{\log K} \quad (22)$$

$$\delta_L = \frac{\log(G_L) - \log(A_L)}{\log K} \quad (23)$$

$$\delta_\Pi = \frac{\log(G_\Pi) - \log(A_\Pi)}{\log K} \quad (24)$$

These measures quantify the effective average proportion of shared groups/links/probability of links across networks and range between 0 and 1.

The framework is implemented in the R package *econetwork* available on CRAN (<https://www.cran.r-project.org>).

#### APPLICATION TO A CASE STUDY: WHEN THE AGGREGATION LEVEL REVERSES THE ASSESSMENT OF THE DRIVERS OF NETWORK DISSIMILARITY

Here, we re-analysed a data set used in Thompson & Townsend (2003). Using groups built *a priori* with three trophic levels, the authors concluded that stream water networks surrounded by pine or tussock grassland in New Zealand differ

in their structure at a mesoscopic level. They attributed this change of structure to differences of energy supply in the two systems. We proposed to extend this analysis to the entire dataset [ten stream water trophic networks sampled in the United States of America (USA hereafter) and New Zealand surrounded either by pines or not (Table 1)] using our novel framework together with trophic groups built using the topology of the metanetwork (Allesina & Pascual 2009; Gauzens *et al.* 2015). We hypothesised that at a species level, geographic location should have a major impact on network dissimilarity due to the different biogeographical histories of the two continents (e.g. dispersal limitation that leads to small species overlap, different life history traits due to different environmental constraints), whereas at a trophic group level, vegetation should have much more impact due to energy supply provided by the riparian vegetation (e.g. vegetation types select for certain groups of species and network structure, which is not discernible at a species level).

The data set consists of ten stream water trophic networks sampled in the USA and New Zealand (Thompson & Townsend 2003, [https://www.nceas.ucsb.edu/interactionweb/resource.html#predator\\_prey](https://www.nceas.ucsb.edu/interactionweb/resource.html#predator_prey), Fig. S1). The riparian vegetation of the American networks is a native species of pine, *Pinus strobus*. Two of the New Zealand networks are surrounded by planted pines, *Pinus radiata* (Table 2). All other networks in New Zealand are surrounded by bush and tussock. The networks contain species of algae, invertebrates and fishes. We

**Table 2** The set of trophic networks and the covariates (adapted from Thompson & Townsend 2003)

Location	Site	Vegetation
Maine (USA)	Troy	<i>Pinus strobus</i>
Maine (USA)	Martins	<i>Pinus strobus</i>
North Carolina (USA)	Herlzler	<i>Pinus strobus</i>
North Carolina (USA)	Cooper	<i>Pinus strobus</i>
New Zealand	Venlaw	<i>Pinus radiata</i>
New Zealand	Berwick	<i>Pinus radiata</i>
New Zealand	North col	Native bush
New Zealand	Powder	Native bush
New Zealand	Trib C	Tussock
New Zealand	Sutton	Tussock

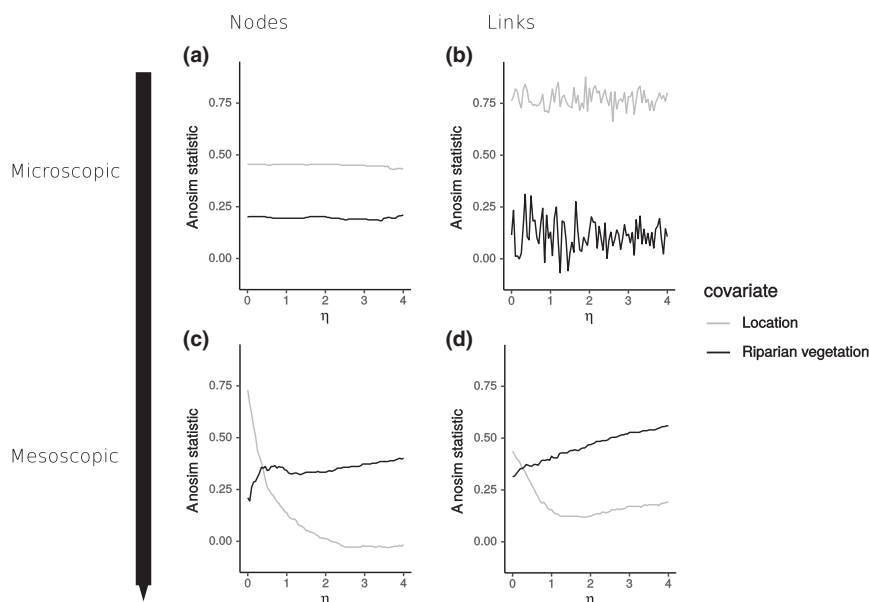
kept only the largest connected component of the metanetwork (Fig. S2). It contains 532 species and has a connectance value of 0.01.

To work at the mesoscale, we first determined the most relevant trophic groups using the stochastic block model implemented in the R package ‘mixer’ (Daudin *et al.* 2008). The optimal number of groups, 14, was identified using an information criterion (for simplicity we only used the optimal number of groups, but could have navigated through a wider range of aggregation levels). Therefore, the scale used to analyse the mesoscopic network dissimilarity is 14/532. We thereafter computed the dissimilarity matrices of link and group abundances at the microscopic and mesoscopic scales (using pairwise  $\delta_L^\eta$  at two different aggregation levels) and the dissimilarity matrix of groups (using pairwise  $\delta_p^\eta$  at two different aggregation levels) along a profile of weights attributed to

abundant groups or links by varying the values of  $\eta$ . We then assessed the influence of the riparian vegetation (presence/absence of pine trees) and the location (USA or New Zealand) on the four dissimilarity matrices per value of  $\eta$  using ANOSIM (Clarke 1993) for both covariates (location and riparian vegetation) along the range of  $\eta$  values. (Fig. 3).



These analyses revealed that at a microscopic scale, the pairwise dissimilarities of both group and link abundances ( $\delta_p^\eta$  and  $\delta_L^\eta$ ) are best explained by the geographic location. At the mesoscopic scale, however, the riparian vegetation was the variable that best explains both the dissimilarity of group and link abundances for medium to high values of  $\eta$  ( $\eta > 0.35$  for the groups dissimilarity and  $\eta > 0.15$  for the links dissimilarity).

Since New Zealand and the USA have drastically different biogeographical histories, they have very few species in common (New Zealand and USA streams share, for example, almost no invertebrate species, Thompson & Townsend 2003). Consequently, the location is indeed expected to be a more powerful explanatory variable of the species dissimilarity (i.e.  $\delta_p^\eta$  at a microscopic scale). Moreover, given that species turnover is partially responsible for the links turnover (i.e.  $\delta_L^\eta$  at a microscopic scale), the latter is also expected to be predominantly explained by the location. Studying  $\delta_p^\eta$  and  $\delta_L^\eta$  at a mesoscopic scale allows to look beyond species turnover, and accounts for the role of the riparian vegetation in diversity, both for the group abundances and the link abundances. Importantly, riparian vegetation best explains group and link dissimilarities for medium to high values of  $\eta$ . So, the abundances of the largest trophic groups and the links between these groups are shaped by the riparian vegetation whereas



**Figure 3** Assessing the drivers of dissimilarity in group abundances and link abundances at different species aggregation levels. Relative importance (ANOSIM statistic) of the location vs. the riparian vegetation regarding the (a) microscopic pairwise dissimilarity in groups abundances ( $\delta_p^\eta$  at a microscopic scale) (b) microscopic pairwise  $\beta$ -diversity in link abundances ( $\delta_L^\eta$  at a microscopic scale) (c) mesoscopic pairwise dissimilarity in group abundances ( $\delta_p^\eta$  at a mesoscopic scale) (d) mesoscopic pairwise dissimilarity in link abundances ( $\delta_L^\eta$  at a mesoscopic scale) across a range of  $\eta$  values (i.e. the viewpoint parameter controlling the weight given to entities (group abundances or link abundances) in the measure of the dissimilarity).



		Local network	Networks in space and time								
Network aggregation	(a)	 <p>To what extent do ecological processes shape network structure?</p>	(c)								
		<p><i>Studies:</i> Cohen <i>et al.</i> 2003 Dormann <i>et al.</i> 2017 (review)</p> <p><i>Hypothesis:</i> Traits and eco-evolutionary processes shape network structure</p>	<p><i>Studies:</i> Caradonna <i>et al.</i> 2017 Tylianakis &amp; Morris 2017 (review)</p> <p><i>Hypothesis:</i> Species composition and interactions change across space and time</p>								
		<p><i>Diversity index</i> <math>C, D^n(\mathbf{p}), D^n(\mathbf{L}), D^n(\mathbf{\Pi})</math></p>	<p style="text-align: center;"><i>Mechanisms</i></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%;">Change in species composition</th> <th style="width: 33%;">Change that alters probability of interactions</th> <th style="width: 33%;">Change in energy fluxes</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;"><math>A_P^\eta, G_P^\eta</math> <math>B_P^\eta, \delta_P^\eta</math></td> <td style="text-align: center;"><math>A_\Pi^\eta, G_\Pi^\eta</math> <math>B_\Pi^\eta, \delta_\Pi^\eta</math></td> <td style="text-align: center;"><math>A_L^\eta, G_L^\eta</math> <math>B_L^\eta, \delta_L^\eta</math></td> </tr> </tbody> </table>			Change in species composition	Change that alters probability of interactions	Change in energy fluxes	$A_P^\eta, G_P^\eta$ $B_P^\eta, \delta_P^\eta$	$A_\Pi^\eta, G_\Pi^\eta$ $B_\Pi^\eta, \delta_\Pi^\eta$	$A_L^\eta, G_L^\eta$ $B_L^\eta, \delta_L^\eta$
	Change in species composition	Change that alters probability of interactions	Change in energy fluxes								
$A_P^\eta, G_P^\eta$ $B_P^\eta, \delta_P^\eta$	$A_\Pi^\eta, G_\Pi^\eta$ $B_\Pi^\eta, \delta_\Pi^\eta$	$A_L^\eta, G_L^\eta$ $B_L^\eta, \delta_L^\eta$									
	<p><i>Diversity index</i> <math>D^n(\mathbf{p}), D^n(\mathbf{L}), D^n(\mathbf{\Pi})</math> (meso)</p>	<p><i>Diversity index</i> <math>A_P^\eta, G_P^\eta</math> (meso) <math>B_P^\eta, \delta_P^\eta</math> (meso)</p> <p><math>A_\Pi^\eta, G_\Pi^\eta</math> (meso) <math>B_\Pi^\eta, \delta_\Pi^\eta</math> (meso)</p> <p><math>A_L^\eta, G_L^\eta</math> (meso) <math>B_L^\eta, \delta_L^\eta</math> (meso)</p>									
(b)	 <p>To what extent can species in a network be aggregated into groups while preserving the overall structure?</p>	(d)									
	<p><i>Studies:</i> Allesina &amp; Pascual 2009 Gauzens <i>et al.</i> 2015 Cirtwill <i>et al.</i> 2018 (review)</p> <p><i>Hypothesis:</i> Species that have similar traits are equivalent in terms of position in the network. Interpretation as trophic group (or trophospecies) in the case of trophic networks</p>	<p><i>Studies:</i> Thompson &amp; Townsend 2003 (groups formed a priori) This study</p> <p><i>Hypothesis:</i> (Trophic) groups composition and interactions change across space and time</p>									
	<p><i>Diversity index</i> <math>D^n(\mathbf{p}), D^n(\mathbf{L}), D^n(\mathbf{\Pi})</math> (meso)</p>	<p style="text-align: center;"><i>Mechanisms</i></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%;">Change in (trophic) groups composition</th> <th style="width: 33%;">Change that alters probability of interactions between groups</th> <th style="width: 33%;">Change in energy fluxes</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;"><math>A_P^\eta, G_P^\eta</math> (meso) <math>B_P^\eta, \delta_P^\eta</math> (meso)</td> <td style="text-align: center;"><math>A_\Pi^\eta, G_\Pi^\eta</math> (meso) <math>B_\Pi^\eta, \delta_\Pi^\eta</math> (meso)</td> <td style="text-align: center;"><math>A_L^\eta, G_L^\eta</math> (meso) <math>B_L^\eta, \delta_L^\eta</math> (meso)</td> </tr> </tbody> </table>			Change in (trophic) groups composition	Change that alters probability of interactions between groups	Change in energy fluxes	$A_P^\eta, G_P^\eta$ (meso) $B_P^\eta, \delta_P^\eta$ (meso)	$A_\Pi^\eta, G_\Pi^\eta$ (meso) $B_\Pi^\eta, \delta_\Pi^\eta$ (meso)	$A_L^\eta, G_L^\eta$ (meso) $B_L^\eta, \delta_L^\eta$ (meso)	
Change in (trophic) groups composition	Change that alters probability of interactions between groups	Change in energy fluxes									
$A_P^\eta, G_P^\eta$ (meso) $B_P^\eta, \delta_P^\eta$ (meso)	$A_\Pi^\eta, G_\Pi^\eta$ (meso) $B_\Pi^\eta, \delta_\Pi^\eta$ (meso)	$A_L^\eta, G_L^\eta$ (meso) $B_L^\eta, \delta_L^\eta$ (meso)									

**Figure 4** Reconciling two perspectives in ecological network analyses. Here, we represent the key questions, seminal studies and underlying hypotheses usually considered in studies of ecological networks, and our specific indices for investigating them. (a) Studying the structure of a local network. (b) Studying the structure of a network where species have been aggregated in meaningful groups (trophic groups in the case of trophic networks). (c) Studying how networks vary in space and time. (d) Studying how networks change in space and time at various aggregation levels (trophic groups in the case of trophic networks). (meso) means the diversity indices are computed at a mesoscopic scale.

their presence (i.e. while omitting their abundances) is explained by the location.

## DISCUSSION

Diversity indices aim to describe and quantify the structure of ecological communities across space and time. There is currently a paradigm shift in the representation of a community, from a species assemblage to an interaction network (Thompson *et al.* 2012; Pellissier *et al.* 2017; Tylianakis & Morris 2017). While deciding which species belong to a community is made easier using a network representation of biodiversity (since a community is no more than a connected network), measuring and partitioning the diversity of these interaction networks is much more complex (Poisot *et al.* 2016; Pellissier *et al.* 2017). Diversity indices using Hill numbers provide a robust framework when ignoring interactions, as it gradually takes into account species abundances and satisfies theoretical properties. To be generic enough and to embrace the complexity of natural systems, these indices should take into account species abundances and the probabilistic nature of biotic interactions, while unifying the existing diversity frameworks. Moreover, they should be able to measure diversity at different species aggregation levels, so as to not inflate diversity indices or overestimate link turnover. In this paper, we defined a set of diversity indices that address each of these

requirements. The proposed framework is a generalisation of the Hill numbers to measure  $\alpha$ -,  $\beta$ - and  $\gamma$ -diversity in link abundances and link probabilities. By doing so, we have extended the existing indices of network diversity (e.g. Poisot *et al.* 2012), while benefiting from key properties of Hill numbers. In other words, using this single framework on a single data set would enable one to not only investigate traditional relationships between species richness and energy as well as understand the compositional turnover across space, but also explore further by deciphering how variations in species abundance, probability of interactions and environmental gradients influence ecological networks.

The proposed framework is based on a probabilistic model of networks where parameters are species abundances and probabilities of interaction between species or groups of species. Consequently, it represents interactions as a random event rather than a deterministic event, thus assuming a plasticity of interactions at an individual level. While this constitutes an appealing representation from a theoretical standpoint, empirical datasets of interaction networks are often binary and lack abundance estimates. Binary networks constitute particular cases of our framework, that then connect with existing frameworks (Poisot *et al.* 2012). Our framework can also be applied to any weighted network (i.e. network containing interaction strength) even if the weights do not strictly represent a probability of interaction. The

viewpoint parameter  $\eta$  can then be used to modulate the weight given to interaction strength when assessing network diversity. The proposed diversity indices are based on Hill numbers that satisfy properties regarding group abundances but also link abundances. This is a fundamental condition to describe adequately network diversity over space and time and to build a robust spatial network ecology.

Additionally, our framework allows to compute diversity indices at different species aggregation levels (Fig. 4). In this paper, we have focused on methods that aggregate species based on the topology of the metanetwork (regular equivalence and stochastic block modelling). These methods aim to form trophic groups (Gauzens *et al.* 2015) and, in the general case, reduce the complexity of the network (i.e. the number of nodes) while preserving the overall structure. Grouping species using ecological and expert knowledge and computing diversity indices is possible using the developed framework. In this latter case, however, there is no guarantee that the structure of the image network will reflect the structure of the original network (Allsina & Pascual 2009; Gauzens *et al.* 2015; Leger *et al.* 2015).

Whatever the clustering method used, the image network can be viewed as a map at a coarser resolution than the original species network. A map which, depending on the method used, summarises faithfully the structure of the original network. Importantly, it changes the assessment of link turnover. Indeed, what appears as link turnover at a species level could disappear at a group level, provided that the species considered belong to the same group. In other words, network diversity patterns depend on the aggregation level we choose to study the network. This introduces a new notion of scale in the analysis of ecological networks and adds to the spatial and temporal scale used to describe network biogeographic patterns (Fig. 4). Studies aiming to describe network biogeography have so far mostly described macroscopic (i.e. connectance, Thompson & Townsend 2003) or microscopic (link turnover at a species level, Poisot *et al.* 2012, 2016; Carstensen *et al.* 2014; CaraDonna *et al.* 2017) scale patterns and occasionally mesoscopic scale using a priori groups based on the trophic level concept (Thompson & Townsend 2003). Statistical methods, such as the stochastic block model and regular equivalence, allow to select an optimal number of groups to cluster the nodes of a network, thus defining an appropriate scale to study network diversity when no ecological knowledge is available for the species described in the network. As shown by the case study, network diversity can be shaped by different ecological processes depending on the aggregation level considered, in the same way that species diversity is shaped by different processes depending on the spatial and aggregation level considered (Münkemüller *et al.* 2014). This encourages to study network diversity at micro-, macroscopic scale and along a profile of mesoscopic scales (i.e. by changing gradually the number of groups of the image network) to study the processes that govern network structure across space. Indeed, given that some empirical evidence suggests that network structure might be random at a species level (CaraDonna *et al.* 2017), one purpose of aggregating species into equivalent groups is to investigate beyond the stochastic plasticity of biotic interactions. For example, the simulation (Supporting Information) suggests a stochastic plasticity at a

species level but not at a group level since the image network is fixed at a given point of the ecological gradient. We hypothesise that, in the real world, there is an aggregation level below which stochastic processes drive the patterns of network diversity, and above which deterministic processes (i.e. ecological processes) are the main drivers. This hypothesis mirrors the use of the concept of emergent groups of organisms to assess the contribution of niche and neutral theory to community assembly (Héroult 2007).

Moreover, since ecological networks are now built using a wide spectrum of organisms, especially microorganisms with the advent of Next Generation Sequencing (Bohan *et al.* 2017) where the notion of species is hard to handle, using indices that allow to understand network diversity through different species aggregation levels will allow overcoming issues in the definition of the biological entity.

We thus believe that this unified framework should now pave the way for a better understanding of the spatial and temporal structure of biodiversity while considering biotic interactions. Indeed, it reconciles two perspectives on ecological networks analysis while building the associated indices. On the one hand, the study of how ecological processes shape an isolated network (Fig. 4a) and how meaningful groups can be derived from the topology of an ecological network (Fig. 4b), and on the other hand how networks vary across space and time (Fig. 4c). Importantly, it allows testing key ecological hypotheses on the processes shaping the spatial and temporal variation of ecological networks (case study, Fig. 3), by varying different aspects of the networks (Fig. 4d). Consequently, it should foster the emergence of spatial network ecology and allow the comparison, analysis and partitioning of multiple ecological networks, from the local community to the global metacommunity they are a part of, while considering various definitions of the organisms involved.

#### ACKNOWLEDGEMENTS

The research received funding from the French Agence Nationale de la Recherche (ANR) through the GlobNets (ANR-16-CE02-0009) and from 'Investissement d'Avenir' grants managed by the ANR (Trajectories: ANR-15-IDEX-02; Montane: OSUG@2020: ANR-10-LAB-56).

#### AUTHORSHIP

MO developed a draft version of the framework that was further elaborated with WT. VM, SD, WT and MO conceptualised the final version of the framework. LO and LC gave additional perspectives to the paper. MO and WT wrote the first version of the paper and all authors contributed substantially to the revisions.

#### DATA ACCESSIBILITY STATEMENT

We used an open data set : [https://www.nceas.ucsb.edu/interactionweb/resources.html#predator\\_prey](https://www.nceas.ucsb.edu/interactionweb/resources.html#predator_prey). The code used to generate the simulated data is provided in section 5 of the Supporting Information. We implemented the method in the

R package *econetwork* available on CRAN (<https://cran.r-project.org>).

## REFERENCES

- Albert, C.H., Thuiller, W., Yoccoz, N.G., Douzet, R., Aubert, S. & Lavorel, S. (2010). A multi-trait approach reveals the structure and the relative importance of intra- vs. interspecific variability in plant traits. *Funct. Ecol.*, **24**, 1192–1201.
- Allesina, S. & Pascual, M. (2009). Food web models: a plea for groups. *Ecol. Lett.*, **12**, 652–662.
- Barnes, A.D., Jochum, M., Mumme, S., Haneda, N.F., Farajallah, A., Widarto, T.H. & Brose, U. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nat. Commun.*, **5**, 5351.
- Belmaker, J., Sekercioglu, C.H. & Jetz, W. (2012). Global patterns of specialization and coexistence in bird assemblages. *J. Biogeogr.*, **39**, 193–203.
- Bersier, L.F., Banasek-Richter, C. & Cattin, M.F. (2002). Quantitative descriptors of food-web matrices. *Ecology*, **83**, 2394–2407.
- Bohan, D.A., Vacher, C., Tamaddon-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017). Next-generation global biomonitoring: large-scale, automated reconstruction of ecological networks. *Trends Ecol. Evol.*, **32**, 477–487.
- Brose, U. & Hillebrand, H. (2016). Biodiversity and ecosystem functioning in dynamic landscapes. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.*, **371**.
- CaraDonna, P.J., Petry, W.K., Ross, M.B., Cunningham, J.L., Bronstein, J.L., Waser, N.M. *et al.* (2017). Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecol. Lett.*, **20**, 385–394.
- Carstensen, D.W., Sabatino, M., Trøjelsgaard, K. & Morellato, L.P.C. (2014). Beta diversity of plant–pollinator networks and the spatial turnover of pairwise interactions. *PLoS ONE*, **9**, 1–7.
- Chalmandrier, L., Münkemüller, T., Lavergne, S. & Thuiller, W. (2015). Effects of species' similarity and dominance on the functional and phylogenetic structure of a plant meta-community. *Ecology*, **96**, 143–153.
- Chao, A. & Chiu, C.H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods Ecol. Evol.*, **7**, 919–928.
- Chao, A., Chiu, C.H. & Jost, L. (2014a). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annu. Rev. Ecol. Syst.*, **45**, 297–324.
- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. *et al.* (2014b). Rarefaction and extrapolation with hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.*, **84**, 45–67.
- Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.*, **18**, 117–143.
- Daudin, J.J., Picard, F. & Robin, S. (2008). A mixture model for random graphs. *Stat. Comput.*, **18**, 173–183.
- Dunne, J. (2006). The network structure of food webs. In: *Ecological Networks: Linking Structure to Dynamics in Food Webs* (eds Pascual, M. & Dunne, J.A.). Oxford University Press, Oxford, pp. 27–86.
- Ellison, A.M. (2010). Partitioning diversity1. *Ecology*, **91**, 1962–1963.
- Elton, C.S. (1927). *Animal Ecology*. Macmillan Co., New York. <https://www.biodiversitylibrary.org/bibliography/7435>
- Gaston, K.J. (2003). *The Structure and Dynamics of Geographic Ranges*. Oxford University Press on Demand, Oxford.
- Gauzens, B., Thébault, E., Lacroix, G. & Legendre, S. (2015). Trophic groups and modules: two levels of group detection in food webs. *J. R. Soc. Interface*, **12**, 20141176.
- Gonzalez-Varo, J.P. & Traveset, A. (2016). The labile limits of forbidden interactions. *Trends Ecol. Evol.*, **31**, 700–710.
- Graham, C.H., Storch, D. & Machac, A. (2018). Phylogenetic scale in ecology and evolution. *Glob. Ecol. Biogeogr.*, **27**, 175–187.
- Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecol. Lett.*, **14**, 1010–1016.
- Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.F., Kaufman, D.M. *et al.* (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.
- Héroult, B. (2007). Reconciling niche and neutrality through emergent group approach. *Perspect. Plant. Ecol.*, **9**, 71–78.
- Hill, M.O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427–432.
- von Humboldt, A. (1805). *Essai sur la géographie des plantes; accompagné d'un tableau physique des régions équinoxiales*. Levrault, Paris.
- Jost, L. (2006). Entropy and diversity. *Oikos*, **113**, 363–375.
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427–2439.
- Jost, L. (2010). Independence of alpha and beta diversities. *Ecology*, **91**, 1969–1974.
- 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 Biol.*, **14**, 1–21.
- Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the holy grail. *Funct. Ecol.*, **16**, 545–556.
- Leger, J.B., Daudin, J.J. & Vacher, C. (2015). Clustering methods differ in their ability to detect patterns in ecological networks. *Methods Ecol. Evol.*, **6**, 474–481.
- Lindeman, R.L. (1942). The trophic-dynamic aspect of ecology. *Ecology*, **23**, 399–417.
- Luck, G.W., Lavorel, S., McIntyre, S. & Lumb, K. (2012). Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *J. Anim. Ecol.*, **81**, 1065–1076.
- Luczkovich, J.J., Borgatti, S.P., Johnson, J.C. & Everett, M.G. (2003). Defining and measuring trophic role similarity in food webs using regular equivalence. *J. Theor. Biol.*, **220**, 303–321.
- Massol, F., Dubart, M., Calcagno, V., Cazelles, K., Jacquet, C., Kéfi, S. & Gravel, D. (2017). Island biogeography of food webs. *Adv. Ecol. Res.*, **56**, 183–262.
- Mazel, F., Wüest, R.O., Lessard, J.P., Renaud, J., Ficetola, F.G., Lavergne, S. *et al.* (2017). Global patterns of beta-diversity along the phylogenetic time-scale: the role of climate and plate tectonics. *Glob. Ecol. Biogeogr.*, **26**, 1211–1221.
- Münkemüller, T., Gallien, L., Lavergne, S., Renaud, J., Roquet, C., Abdulhak, S. *et al.* (2014). Scale decisions can reverse conclusions on community assembly processes. *Glob. Ecol. Biogeogr.*, **23**, 620–632.
- Newman, M. & Leicht, E. (2007). Mixture models and exploratory analysis in networks. *Proc. Natl Acad. Sci.*, **104**, 9564–9569.
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M. *et al.* (2017). Comparing species interaction networks along environmental gradients. *Biol. Rev.*, **93**, 785–800.
- Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecol. Lett.*, **15**, 1353–1361.
- Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.J. & Stouffer, D.B. (2016). The structure of probabilistic networks. *Methods Ecol. Evol.*, **7**, 303–312.
- Reeve, R., Leinster, T., Cobbold, C., Thompson, J.N., Brummitt, M. & Matthews, L. (2014). How to partition diversity. *arXiv:1404.6520*.
- Rényi, A. (1961). On measures of entropy and information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (ed Neyman, J.). University of California Press, Berkeley, CA. pp. 547–561.

- Routledge, R. (1979). Diversity indices: which ones are admissible? *J. Theor. Biol.*, 76, 503–515.
- Thompson, R.M. & Townsend, C.R. (2003). Impacts on stream food webs of native and exotic forest: an intercontinental comparison. *Ecology*, 84, 145–161.
- Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L. *et al.* (2012). Food webs: reconciling the structure and function of biodiversity. *Trends Ecol. Evol.*, 27, 689–697.
- Tucker, C.M., Cadotte, M.W., Carvalho, S.B., Davies, T.J., Ferrier, S., Fritz, S.A. *et al.* (2016). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev.*, 92, 698–715.
- Tylianakis, J.M. & Morris, R.J. (2017). Ecological networks across environmental gradients. *Annu. Rev. Ecol. Evol. Syst.*, 48, 25–48.
- Whittaker, R.H. (1960). Vegetation of the siskiyou mountains, oregon and california. *Ecol. Monogr.*, 30, 279–338.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, John Drake

Manuscript received 23 July 2018

First decision made 28 August 2018

Second decision made 19 November 2018

Manuscript accepted 12 December 2018

### **1.3 Annexes de l'article (en anglais)**

# Supporting Information

## 1 Proofs

### 1.1 Proof of Eq. 2

The abundance of the group  $q$  is:

$$p_q = Pr(i \in q)$$

The abundance of the link between group  $q$  and  $l$  is:

$$L_{ql} = Pr(i \rightarrow j, i \in q, j \in l)$$

Using Bayes formula, we have:

$$L_{ql} = Pr(i \rightarrow j | i \in q, j \in l) Pr(i \in q) Pr(j \in l)$$

$$L_{ql} = \pi_{ql} p_q p_l$$

### 1.2 Proof of Eq. 3

By definition,

$$\tilde{p}_q = Pr(i \in C_q)$$

Moreover,

$$C_q = \dot{\bigcup}_{k \in C_q} V_k$$

where  $\dot{\bigcup}$  denotes the disjoint union. So:

$$\tilde{p}_q = \sum_{k \in C_q} p_k$$

### 1.3 Proof of Eq. 4 and 5

By definition,

$$\tilde{\pi}_{ql} = Pr(i \rightarrow j | i \in C_q, j \in C_l)$$

Using Bayes' formula,

$$\tilde{\pi}_{ql} = \frac{Pr(i \rightarrow j, i \in C_q, j \in C_l)}{Pr(i \in C_q) Pr(j \in C_l)}$$

By definition,

$$\tilde{L}_{ql} = Pr(i \rightarrow j, i \in C_q, j \in C_l)$$

Using the law of total probabilities:

$$\tilde{L}_{ql} = \sum_{k \in C_q, k' \in C_l} Pr(i \rightarrow j | i \in V_k, j \in V_{k'}) Pr(i \in V_k) Pr(j \in V_{k'})$$

$$\tilde{L}_{ql} = \sum_{k \in C_q, k' \in C_l} \pi_{kk'} p_k p'_{k'}$$

Since

$$Pr(i \in C_q, j \in C_l) = \left( \sum_{k \in C_q} p_k \right) \left( \sum_{k' \in C'_l} p'_{k'} \right)$$

It follows

$$\tilde{\pi}_{ql} = \frac{\sum_{k \in C_q, k' \in C'_l} \pi_{kk'} p_k p'_{k'}}{\left( \sum_{k \in C_q} p_k \right) \left( \sum_{k' \in C'_l} p'_{k'} \right)}$$

## 1.4 Proof of Eq. 7

By definition,

$$C = Pr(i \rightarrow j)$$

Moreover,

$$L_{ql} = Pr(i \rightarrow j | i \in q, j \in l)$$

By summing over  $q$  and  $l$ , it follows that

$$C = \sum_{q,l} L_{ql}$$

$$C = \sum_{q,l} \pi_{ql} p_q p_l$$

This last result is a generalization of the weighted connectance presented in [Poisot \*et al.\* \(2016\)](#).

## 2 Clustering methods

We focus here on two network clustering methods that are particularly appropriate to aggregate nodes without any *a priori*: regular equivalence ([Luczkovich \*et al.\* 2003](#)) and the stochastic block model ([Daudin \*et al.\* 2008](#); [Allesina & Pascual 2009](#); [Newman & Leicht 2007](#)). These methods have already been used to study interaction networks (regular equivalence [Gauzens \*et al.\* 2015](#); the stochastic block model [Kéfi \*et al.\* 2016](#)). Importantly, in the case of trophic networks, the groups formed are interpreted in terms of trophic groups ([Allesina & Pascual 2009](#); [Gauzens \*et al.\* 2015](#)). We acknowledge that other methods exist (see e.g., [Henderson \*et al.\* 2012](#)) and could be applied to trophic networks. In regular equivalence, two nodes belong to the same equivalence class (i.e, same groups) if they have the same links (with the same weight in the case of weighted networks) to all other equivalence classes. Regular equivalence can thus be applied to binary or weighted networks. In contrast, the stochastic block model is a probabilistic perspective. Two nodes belong to the same group if they share the same probability of being linked to all the other groups. Similar to regular equivalence, the method is applicable to both binary networks ([Daudin \*et al.\* 2008](#)) and weighted networks ([Mariadassou \*et al.\* 2010](#)). A broader review of the links between clustering methods based on the topology of a network and the concepts of species' role in trophic networks can be found in [Cirtwill \*et al.\* \(2018\)](#).

## 3 General framework

In this section, we present the general case of our framework (i.e, for any  $\eta \geq 0$ ) to partition network diversity using the framework presented in [Chao & Chiu \(2016\)](#), but also the frameworks presented in [Jost \(2007\)](#) and [Reeve \*et al.\* \(2014\)](#).

### 3.1 Chao's framework

In the main text, we introduced our diversity indices for  $\eta = 1$ . Here, we present the generalisation for any  $\eta \geq 0$ . The interpretations of the indices are the same than those presented in the main document, as well as the ranges presented in table 1.

For  $\eta \in \mathbb{R}^+ \setminus \{1\}$ , the  $\alpha$ -diversities are:

$$A_P^\eta = \frac{1}{K} \left( \sum_{q=1}^Q \sum_{k=1}^K \left( \frac{P_{qk}}{P_{+++}} \right)^\eta \right)^{\frac{1}{1-\eta}}$$

$$A_L^\eta = \frac{1}{K} \left( \sum_{q,l=1}^Q \sum_{k=1}^K \left( \frac{L_{qlk}}{L_{++++}} \right)^\eta \right)^{\frac{1}{1-\eta}}$$

$$A_\Pi^\eta = \frac{1}{K} \left( \sum_{q,l=1}^Q \sum_{k=1}^K \left( \frac{\pi_{qlk}}{\pi_{++++}} \right)^\eta \right)^{\frac{1}{1-\eta}}$$

For  $\eta \in \mathbb{R}^+ \setminus \{1\}$ , the  $\gamma$ -diversities are:

$$G_P^\eta = \left( \sum_{q=1}^Q \left( \frac{P_{q+}}{P_{+++}} \right)^\eta \right)^{\frac{1}{1-\eta}}$$

$$G_L^\eta = \left( \sum_{q,l=1}^Q \left( \frac{L_{ql+}}{L_{++++}} \right)^\eta \right)^{\frac{1}{1-\eta}}$$

$$G_\Pi^\eta = \left( \sum_{q,l=1}^Q \left( \frac{\pi_{ql+}}{\pi_{++++}} \right)^\eta \right)^{\frac{1}{1-\eta}}$$

And the  $\beta$ -diversities are defined as:

$$B_P^\eta = \frac{G_P^\eta}{A_P^\eta}$$

$$B_L^\eta = \frac{G_L^\eta}{A_L^\eta}$$

$$B_\Pi^\eta = \frac{G_\Pi^\eta}{A_\Pi^\eta}$$

Overlap measures between the K local networks can be built using  $\beta$ -diversities giving the following dissimilarity measures (for  $\eta \neq 1$ ):

$$\delta_P^\eta = \frac{1 - (1/B_P^\eta)^{\eta-1}}{1 - (1/K)^{\eta-1}}$$

$$\delta_L^\eta = \frac{1 - (1/B_L^\eta)^{\eta-1}}{1 - (1/K)^{\eta-1}}$$

$$\delta_\Pi^\eta = \frac{1 - (1/B_\Pi^\eta)^{\eta-1}}{1 - (1/K)^{\eta-1}}$$

### 3.2 Chao's framework and the invariance under shattering

The uniform weighted framework presented in (Chao & Chiu, 2016) do not follow the invariance under shattering property. We propose here a counter-example to demonstrate it. Let's assume the following community matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

where  $P_{qk}$  is the relative abundance of species (or group)  $q$  in the community  $k$ . We have then:

$$\forall \eta \geq 0, \quad B_P^\eta = 2$$

Imagine now that the second community is shattered in two pieces with the same relative abundance, with  $\mathbf{P}'$  being the corresponding community matrix.

$$\mathbf{P}' = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Then, we have  $B_{P'}^2 = 1.8 < 2$ . This violates the invariance under shattering property. It is a consequence of the uniform weights used in the Chao's framework. More generally, we now show that whatever the community matrix  $P$  considered, if one shatters a community in an infinite number of pieces with the same relative abundances, the  $\beta$ -diversity converges towards 1 for all  $\eta$  values.

Let  $(P_{qk})_{1 \leq q \leq Q, 1 \leq k \leq K}$  be a community matrix.  $P_{qk}$  is the relative abundance of species (or group)  $q$  in the community  $k$ . We assume, without loss of generality, that  $P_{+k} = 1$ . So,  $P_{+++} = K$ . We now assume that we shatter the community  $K$  in  $S$  pieces with the same relative abundances (assuming



that this community is not empty) and denote  ${}^S P$  the corresponding community matrix. Written as a vector of columns:

$${}^S \mathbf{P} = \left( \mathbf{P}_{\cdot,1}, \dots, \underbrace{\mathbf{P}_{\cdot,K}, \dots, \mathbf{P}_{\cdot,K}}_{S \text{ times}} \right)$$

where  $\mathbf{P}_{\cdot,k} = {}^t (P_{1k}, \dots, P_{Qk})$ . We have also  ${}^S P_{++} = K + S - 1$ . Let  $\eta \in \mathbb{R}^+ \setminus \{1\}$  and:

$$\begin{aligned} B^\eta: \mathbb{N}^* &\rightarrow [1; +\infty[ \\ S &\mapsto B_{S\mathbf{P}}^\eta \end{aligned}$$

For  $S \in \mathbb{N}^*$ ,

$$B^\eta(S) = \left( \frac{\sum_{q=1}^Q \left( \frac{{}^S P_{q+}}{K+S-1} \right)^\eta}{(K+S-1)^{\eta-1} \sum_{q=1}^Q \sum_{k=1}^{K+S-1} \left( \frac{{}^S P_{qk}}{K+S-1} \right)^\eta} \right)^{\frac{1}{1-\eta}}$$

Defining

$$\begin{aligned} P_{q,max} &= \max({}^S P_{q1}, \dots, {}^S P_{q,K-1}) \\ P_{q,min} &= \min({}^S P_{q1}, \dots, {}^S P_{q,K-1}) \end{aligned}$$

We get

$$\sum_{q=1}^Q \left( \frac{{}^S P_{q+}}{K+S-1} \right)^\eta \leq \sum_{q=1}^Q \left( \frac{(K-1)P_{q,max} + {}^S P_{qK}}{K+S-1} \right)^\eta$$

and

$$\sum_{q=1}^Q \sum_{k=1}^{K+S-1} \left( \frac{{}^S P_{qk}}{K+S-1} \right)^\eta \geq \sum_{q=1}^Q (K-1) \left( \frac{P_{q,min}}{K+S-1} \right)^\eta + \sum_{q=1}^Q S \left( \frac{P_{qK}}{K+S-1} \right)^\eta$$

Let

$$N(S) = \sum_{q=1}^Q \left( \frac{(K-1)P_{q,max} + {}^S P_{qK}}{K+S-1} \right)^\eta$$

We have

$$\lim_{S \rightarrow +\infty} N(S) = \sum_{q=1}^Q (P_{qK})^\eta$$

Moreover, let

$$D(S) = (K+S-1)^{\eta-1} \left( \sum_{q=1}^Q (K-1) \left( \frac{P_{q,min}}{K+S-1} \right)^\eta + \sum_{q=1}^Q S \left( \frac{P_{qK}}{K+S-1} \right)^\eta \right)$$

$$D(S) = \frac{1}{K+S-1} \left( \sum_{q=1}^Q (K-1) (P_{q,min})^\eta + \sum_{q=1}^Q S (P_{qK})^\eta \right)$$

We have

$$\lim_{D \rightarrow +\infty} D(S) = \sum_{q=1}^Q (P_{qK})^\eta$$

Since

$$1 \leq B^\eta(S) \leq \left( \frac{N(S)}{D(S)} \right)^{\frac{1}{1-\eta}}$$

and

$$\lim_{S \rightarrow +\infty} \left( \frac{N(S)}{D(S)} \right)^{\frac{1}{1-\eta}} = 1$$

It follows

$$\lim_{S \rightarrow +\infty} B^\eta(S) = 1 \quad \eta \in \mathbb{R}^+ \setminus \{1\}$$

We extend the result for  $\eta = 1$  by continuity of the function  $\eta \mapsto B_{\mathbf{SP}}^\eta$ . The result holds now for  $\eta \in \mathbb{R}^+$ , as was to be shown. Similar results can be shown for the  $\beta$ -diversity in link abundances and link probabilities.

### 3.3 Jost's and Reeve's frameworks

In this section, we also present the partition of our network diversity indices using the frameworks presented in [Jost \(2007\)](#) and in [Reeve et al. \(2014\)](#). These two frameworks allow to measure diversity while putting weights (that are not necessarily evenly distributed contrary to the Chao and Chiu's framework) on local networks. Jost's framework satisfies the invariance under shattering property for  $\eta \in 0,1$  (and to all order for equal weights) while Reeve's framework satisfies this property whatever  $\eta$ . However, this last framework is not correct for similarity sensitive case i.e., when phylogenetic or functional information is incorporated to the indices (see Appendix 1 of [Chao et al. \(2014\)](#) for a counter-example). However, the framework is correct when no phylogenetic or functional information is used (as in our network diversity indices) as shown in S2.1.2 of [Reeve et al. \(2014\)](#). We thus encourage the reader to use Reeve's framework with special care.

We recommend to use Jost's framework for any potential extension with non-uniform weights that would incorporate phylogenetic or functional information. We recommend to use Reeve's in biodiversity surveys where sites might be re-sampled at a higher spatial resolution after the analysis. Indeed, in that case, it would be desirable to find the same  $\beta$ -diversity values if an area is shattered in two with the same group abundances, link abundances or link probabilities.

#### 3.3.1 Notations

The relative abundance of the group  $q$  in the local network  $k$  is  $p_{qk}$ , giving the matrix  $\mathbf{P}$  of group abundances (so that  $\sum_{1 \leq q \leq Q, 1 \leq k \leq K} P_{qk} = 1$ ). The relative group abundances in the local network  $k$  is  $\mathbf{P}_{\cdot \mathbf{k}}$  and it constitutes a fraction  $\omega_k^p = \sum_{1 \leq q \leq Q} P_{qk}$  of total group abundances. The normalized relative group abundances in the local network  $k$  is  $\overline{\mathbf{P}}_{\cdot \mathbf{k}} = \frac{\mathbf{P}_{\cdot \mathbf{k}}}{\omega_k^p}$  (so that  $\sum_{1 \leq q \leq Q} \overline{P}_{qk} = 1$ ). It follows that  $p^{meta} = \sum_{1 \leq k \leq K} \omega_k^p \overline{\mathbf{P}}_{\cdot \mathbf{k}}$

Similarly, the relative link abundance between groups  $q$  and  $l$  in the local network  $k$  is  $L_{qlk}^*$ , giving the matrix of link abundances in the local network  $k$ ,  $\mathbf{L}_{\cdot \mathbf{k}}^*$  and the tensor of order three  $\mathbf{L}^* = (L_{qlk}^*)$  of link abundances (so that  $\sum_{1 \leq q, l \leq Q, 1 \leq k \leq K} L_{qlk}^* = 1$ ). If we note  $C_k$  the connectance of the local network  $k$  and  $\overline{\mathbf{L}}_{\cdot \mathbf{k}}$  the normalized link abundances of the network  $k$  (so that  $\sum_{1 \leq q, l \leq n} \overline{L}_{qlk} = 1$ ), we have  $\overline{\mathbf{L}}_{\cdot \mathbf{k}} = \frac{\mathbf{L}_{\cdot \mathbf{k}}^*}{\omega_k^L}$ , with  $\omega_k^L = \frac{C_k}{\sum C_p}$  (see next subsection for proof). We have then  $\mathbf{L}_{\mathbf{meta}}^* = \sum_{1 \leq k \leq K} \omega_k^L \overline{\mathbf{L}}_{\cdot \mathbf{k}}$  (see next subsection for a proof).

Moreover, the relative link probability between groups  $q$  and  $l$  in the local network  $k$  is  $\pi_{qlk}^*$ , giving the matrix of link probabilities in the local network  $k$ ,  $\mathbf{\Pi}_{\cdot \mathbf{k}}^*$ , and the tensor of order three  $\mathbf{\Pi}^* = (\pi_{qlk}^*)$  of link probabilities (so that  $\sum_{1 \leq q, l \leq Q, 1 \leq k \leq K} \pi_{qlk}^* = 1$ ). The matrix of link probabilities in the local network  $k$ ,  $\mathbf{\Pi}_{\cdot \mathbf{k}}^*$ , constitutes a fraction  $\omega_k^\pi = \sum_{1 \leq q, l \leq Q} \pi_{qlk}^*$  of total link probabilities. The normalized matrix of link probabilities in the local network  $k$  is  $\overline{\mathbf{\Pi}}_{\cdot \mathbf{k}} = \frac{\mathbf{\Pi}_{\cdot \mathbf{k}}^*}{\omega_k^\pi}$  (so that  $\sum_{1 \leq q, l \leq Q} \overline{\Pi}_{\cdot \mathbf{k}} = 1$ ). We have then  $\mathbf{\Pi}_{\mathbf{meta}}^* = \sum_{1 \leq k \leq K} \omega_k^\pi \overline{\mathbf{\Pi}}_{\cdot \mathbf{k}}$

#### 3.3.2 Weights on the networks when assessing link diversity

By definition :

$$\omega_k^L = \sum_{1 \leq q, l \leq Q} L_{qlk}^*$$

where  $L_{qlk}^*$  is an element of the tensor of relative link abundances  $L$ . This tensor is normalized so that  $\sum_{1 \leq q, l \leq Q, 1 \leq k \leq K} L_{qlk}^* = 1$ . By denoting  $L_{qlk}$  the absolute link abundance between group  $q$  and  $l$  in the network  $k$  and  $C_k$  the connectance of the network  $k$ , we have (from Eq. 7) :

$$C_k = \sum_{q, l} L_{qlk}$$

and

$$\sum_k C_k = \sum_{q,l,k} L_{qlk}$$

Since  $\frac{\sum_{q,l} L_{qlk}^*}{\sum_{q,l,k} L_{qlk}^*} = \frac{\sum_{q,l} L_{qlk}}{\sum_{q,l,k} L_{qlk}}$ , it follows :

$$\omega_k^L = \frac{C_k}{\sum_k C_k}$$

### 3.3.3 Reeve's framework

We first define the power mean function of order  $\eta$  of  $\mathbf{X}$  weighted by  $\mathbf{u}$ , where  $\mathbf{X}$  and  $\mathbf{u}$  are either a vectors or matrices.

If  $\mathbf{X}$  and  $\mathbf{u}$  are real vectors and  $\mu$  a positive real number:

$$M_\mu(\mathbf{u}, \mathbf{X}) = \begin{cases} \left( \sum_q u_q X_q^\mu \right)^{\frac{1}{\mu}} & \text{if } \mu \neq 0 \\ \exp \left( \sum_q u_q \log(X_q) \right) & \text{if } \mu = 0 \end{cases}$$

If  $\mathbf{X}$  and  $\mathbf{u}$  are real matrices and  $\mu$  a positive real number:

$$M_\mu(\mathbf{u}, \mathbf{X}) = \begin{cases} \left( \sum_{q,l} u_{ql} X_{ql}^\mu \right)^{\frac{1}{\mu}} & \text{if } \mu \neq 0 \\ \exp \left( \sum_{ql} u_{ql} \log(X_{ql}) \right) & \text{if } \mu = 0 \end{cases}$$

The local  $\alpha$ -diversities (groups, link abundances and link probabilities) of the network  $k$  are defined as :

$$\alpha_{P,k}^\eta = M_{1-\eta}(\overline{\mathbf{P}}_{\cdot k}, \frac{1}{\overline{\mathbf{P}}_{\cdot k}})$$

$$\alpha_{L,k}^\eta = M_{1-\eta}(\overline{\mathbf{L}}_{\cdot k}, \frac{1}{\overline{\mathbf{L}}_{\cdot k}})$$

$$\alpha_{\Pi,k}^\eta = M_{1-\eta}(\overline{\mathbf{\Pi}}_{\cdot k}, \frac{1}{\overline{\mathbf{\Pi}}_{\cdot k}})$$

The overall  $\alpha$ -diversities are:

$$A_P^\eta = M_{1-\eta}(\omega^{\mathbf{P}}, \alpha_{\mathbf{P}}^\eta)$$

$$A_L^\eta = M_{1-\eta}(\omega^{\mathbf{L}}, \alpha_{\mathbf{L}}^\eta)$$

$$A_{\Pi}^\eta = M_{1-\eta}(\omega^\pi, \alpha_{\mathbf{\Pi}}^\eta)$$

The local  $\beta$ -diversities of the network  $k$  are defined as :

$$\beta_{P,k}^\eta = \frac{1}{M_{1-\eta}(\overline{\mathbf{P}}_{\cdot k}, \mathbf{p}^{\text{meta}}/\overline{\mathbf{P}}_{\cdot k})}$$

$$\beta_{L,k}^\eta = \frac{1}{M_{1-\eta}(\overline{\mathbf{L}}_{\cdot k}, \mathbf{L}^{\text{meta}}/\overline{\mathbf{L}}_{\cdot k})}$$

$$\beta_{\Pi,k}^\eta = \frac{1}{M_{1-\eta}(\overline{\mathbf{\Pi}}_{\cdot k}, \mathbf{\Pi}^{\text{meta}}/\overline{\mathbf{\Pi}}_{\cdot k})}$$

The overall  $\beta$ -diversities are defined as:

$$B_P^\eta = M_{1-\eta}(\omega^{\mathbf{P}}, \beta_{\mathbf{P}}^\eta)$$

$$B_L^\eta = M_{1-\eta}(\omega^{\mathbf{L}}, \beta_{\mathbf{L}}^\eta)$$

$$B_{\Pi}^\eta = M_{1-\eta}(\omega^\pi, \beta_{\mathbf{\Pi}}^\eta)$$

The local  $\gamma$ -diversities of the network  $k$  are defined as:

$$\gamma_{P,k}^\eta = M_{1-\eta}(\overline{\mathbf{P}}_{\cdot k}, 1/\mathbf{p}^{\text{meta}})$$

$$\begin{aligned}\gamma_{L,k}^\eta &= M_{1-\eta}(\overline{\mathbf{L}\cdot\mathbf{k}}, 1/\mathbf{L}^{\mathbf{meta}}) \\ \gamma_{\mathbf{II},k}^\eta &= M_{1-\eta}(\overline{\mathbf{II}\cdot\mathbf{k}}, 1/\mathbf{II}^{\mathbf{meta}})\end{aligned}$$

The overall  $\gamma$ -diversities are defined as:

$$\begin{aligned}G_P^\eta &= M_{1-\eta}(\omega^{\mathbf{P}}, \gamma_{\mathbf{P}}) \\ G_L^\eta &= M_{1-\eta}(\omega^{\mathbf{L}}, \gamma_{\mathbf{L}}) \\ G_{\mathbf{II}}^\eta &= M_{1-\eta}(\omega^\pi, \gamma_{\mathbf{II}})\end{aligned}$$

The range of variation of the overall diversity indices are the same as these presented in table 1 for  $\eta = 1$  using the Chao's framework. The proofs can be found in [Reeve et al. \(2014\)](#) (Appendix S2.1.2).

### 3.3.4 Jost's Framework

The  $\alpha$ -diversities for  $\eta \neq 1$  are defined as:

$$A_P^\eta = \frac{\left(\sum_k (\omega_k^P)^\eta \left(\sum_q \overline{P_{qk}^\eta}\right)\right)^{\frac{1}{1-\eta}}}{\left(\sum_k (\omega_k^P)^\eta\right)}$$

For  $\eta = 1$ , it converges towards the same index than the one presented in the previous section (i.e.  $A_P^1$ ). It ranges between 1 and the number of groups in the metanetwork.

$$A_L^\eta = \frac{\left(\sum_k (\omega_k^L)^\eta \left(\sum_{q,l} \overline{L_{qlk}^\eta}\right)\right)^{\frac{1}{1-\eta}}}{\left(\sum_k (\omega_k^L)^\eta\right)}$$

For  $\eta = 1$ , it converges towards the same index than the one presented in the previous section (i.e.  $A_L^1$ ). It ranges between 1 and the number of links in the metanetwork.

$$A_{\mathbf{II}}^\eta = \frac{\left(\sum_k (\omega_k^{\mathbf{II}})^\eta \left(\sum_{q,l} \overline{\mathbf{II}_{qlk}^\eta}\right)\right)^{\frac{1}{1-\eta}}}{\left(\sum_k (\omega_k^{\mathbf{II}})^\eta\right)}$$

For  $\eta = 1$ , it converges towards the same index than the one presented in the previous section (i.e.  $A_{\mathbf{II}}^1$ ). It ranges between 1 and the number of links in the metanetwork.

The  $\gamma$ -diversities are defined as :

$$G_P^\eta = M_{1-\eta}(\mathbf{p}^{\mathbf{meta}}, 1/\mathbf{p}^{\mathbf{meta}})$$

It ranges between 1 and the number of groups in the metanetwork.

$$G_L^\eta = M_{1-\eta}(\mathbf{L}^{\mathbf{meta}}, 1/\mathbf{L}^{\mathbf{meta}})$$

$$G_{\mathbf{II}}^\eta = M_{1-\eta}(\mathbf{II}^{\mathbf{meta}}, 1/\mathbf{II}^{\mathbf{meta}})$$

It ranges between 1 and the number of links in the metanetwork. The  $\beta$ -diversities are defined as

$$B_P^\eta = \frac{G_P^\eta}{A_P^\eta}$$

$$B_L^\eta = \frac{G_L^\eta}{A_L^\eta}$$

$$B_{\mathbf{II}}^\eta = \frac{G_{\mathbf{II}}^\eta}{A_{\mathbf{II}}^\eta}$$

However, for  $\eta \neq 1$ ,  $\alpha$ -diversity can exceed  $\gamma$ -diversity ([Jost 2007](#)). So, these indices must be used for  $\eta = 1$  (or with equal weights for any  $\eta$  value). In this case ( $\eta = 1$ ), the  $\beta$  diversities have the following ranges ([Jost 2007](#)):

$$1 \leq B_P^1 \leq M_1(\omega^{\mathbf{P}}, \frac{1}{\omega^{\mathbf{P}}})$$

$$1 \leq B_L^1 \leq M_1(\omega^{\mathbf{L}}, \frac{1}{\omega^{\mathbf{L}}})$$

$$1 \leq B_{\mathbf{II}}^1 \leq M_1(\omega^{\mathbf{II}}, \frac{1}{\omega^{\mathbf{II}}})$$

Dissimilarity measures can be built from  $\beta$ -diversity as presented in section 2.1 (using only  $\eta = 1$  for the case of unequal weights).

### 3.4 Link with existing frameworks

While we originally defined our framework and associated indices for the general case of weighted networks (weights on links and abundances of nodes), they are obviously applicable to un-weighted networks. In this particular instance, our framework generalizes the metrics developed by [Poisot et al. \(2012\)](#) to assess the diversity of non-weighted networks. In addition to species turnover ( $\beta_S$ , Sorensen index on species composition), the authors proposed to measure the dissimilarity of links ( $\beta_{WN}$ , Sorensen index on links) and to partition it in an additive way between a component that measures the plasticity of interactions ( $\beta_{OS}$ ) and a component that measures the links turnover due to species turnover ( $\beta_{ST}$ ). We propose to show that, at a microscopic scale for a binary network,  $\delta_P^0 = \beta_S$ ,  $\delta_L^0 = \beta_{WN}$  and  $\delta_{\Pi}^0 = \beta_{OS}$  (using Chao's framework).

Let consider two networks  $G_1$  and  $G_2$ , we note  $V_1, V_2$  the corresponding sets of nodes and  $E_1, E_2$  the corresponding set of edges. Let  $G_{meta}$  be the metanetwork, defined (in the case of binary networks) as the union of  $G_1$  and  $G_2$  (i.e.  $V_{meta} = V_1 \cup V_2$  and  $E_{meta} = E_1 \cup E_2$ ).

Poisot's diversity indices ( $\beta_S$  and  $\beta_{WN}$ ) consist in Sorensen dissimilarity index on sets of nodes and links of the two networks. We have :

$$\beta_S = \frac{|V_1| + |V_2| - 2|V_1 \cap V_2|}{|V_1| + |V_2|}$$

where  $|\cdot|$  denotes the cardinal function. Moreover :

$$\beta_{WN} = \frac{|E_1| + |E_2| - 2|E_1 \cap E_2|}{|E_1| + |E_2|}$$

$\delta_P^0$  consists in Sorensen dissimilarity index on nodes of  $G_1$  and  $G_2$  (i.e. species composition, [Chao & Chiu 2016](#)), so  $\delta_P^0 = \beta_S$ . Similarly,  $\delta_L^0$  consists in Sorensen dissimilarity index on edges of  $G_1$  and  $G_2$ , so  $\delta_L^0 = \beta_{WN}$ .

We now show that  $\delta_{\Pi}^0 = \beta_{OS}$ . To do so, we introduce the induced subgraph function *ind.sub*. Let  $G$  be a network and  $V$  a subset of  $V(G)$ . *ind.sub*( $G, V$ ) is the network whose nodes are  $V$  and edges are  $E = E(G) \setminus (q, l) / (q \notin V \vee l \notin V)$ . We note  $V^{\cap} = V_1 \cap V_2$ ,  $G_1^{\cap} = \text{ind.sub}(G_1, V^{\cap})$  and  $G_2^{\cap} = \text{ind.sub}(G_2, V^{\cap})$ . The set of nodes of these graphs are  $V^{\cap}$  and we note  $E_1^{\cap}$  and  $E_2^{\cap}$  their sets of edges.. In order to estimate the probability of interaction between species  $q$  and  $l$  in the network  $k$ ,  $\pi_{qlk}^*$ , we need to observe at least one individual of species  $q$  and  $l$  in the network  $k$ . Consequently, in order to compute probabilities of interaction, we only consider species that belong to  $V^{\cap}$ . In the case of the two considered networks, the dimension of  $\Pi$  is then  $|V^{\cap}| * |V^{\cap}| * 2$ . Poisot's measure of plasticity of interactions is :

$$\beta_{OS} = \frac{|E_1^{\cap}| + |E_2^{\cap}| - 2|E_1^{\cap} \cap E_2^{\cap}|}{|E_1^{\cap}| + |E_2^{\cap}|}$$

At a microscopic level,  $\delta_{\Pi}^0$  consists in the Sorensen dissimilarity index on the edges of  $G_1^{\cap}$  and  $G_2^{\cap}$ . Consequently,  $\delta_{\Pi}^0 = \beta_{OS}$ .

## 4 Simulation study

Here, we simulate the change in the structure of an ecological network along a gradient (Fig. S3a) and show that the observed network diversity strongly depends on the aggregation level at which the network is considered. For simplicity, we hypothesized that we have a prior knowledge of the groups.

We assume that we observe three groups ( $C_1, C_2, C_3$ ), each containing 100 species, along an environmental gradient. We assume no species turnover along the gradient. In other words, at each location of the gradient, the local species pool is the regional species pool. At a given location  $t$  within the gradient, the matrix of link probabilities  $\Pi_t^*$  between the groups is thus given by:

$$\begin{pmatrix} C_1 & C_2 & C_3 \\ 0 & 0 & 0 \\ c-t & 0 & 0 \\ c+t & 0 & 0 \end{pmatrix} \begin{matrix} C_1 \\ C_2 \\ C_3 \end{matrix} \quad (1)$$

At a species level, the network is binary and its edges are drawn in independent Bernoulli laws of parameters depending on the group of species (the interaction between two species from group  $q$  and  $l$  is a random event that occurs with a probability of  $\pi_{qlt}^*$ ). Two species belonging to the same group are equivalent in the sense that they have the same link probabilities with all the other species. However, globally, the species are not equivalent since they do not all have the same probability of interacting. Moreover, the link probabilities between groups change along the gradient. This simulation setting may reflect the interactions between a group of rodents (C1) and two groups of parasites (C2, C3), where we observed a variation in parasitic interactions following a shift in the rodents' phenology induced by climate (see [Poisot \*et al.\* 2017](#)).

We now propose to analyze network diversity at micro-, meso-, and macroscopic scales along the gradient. The groups are evenly distributed along the gradient since they contain 100 species each (i.e.  $p_{C_1} = p_{C_2} = p_{C_3} = \frac{1}{3}$ ). The expected value of the connectance (i.e. macroscopic diversity) at a location  $t$  in the gradient,  $E(C_t)$ , is then (Eq. 8):

$$E(C_t) = \frac{c-t}{3} + \frac{c+t}{3} = \frac{2c}{3} \quad (2)$$

Therefore, the expected value of connectance is constant along the gradient and consequently, no pattern can appear at a macroscopic scale along the gradient.

We now propose to evaluate the dissimilarity of probabilities of interactions between each pair of networks along the gradient at both the microscopic (species-level) and mesoscopic (group level) scales using  $\delta_{\Pi}^1$ .

We used 21 values of  $t$ , regularly spaced across a range of -0.05 and 0.05 to represent the environmental gradient and set  $c = 0.1$ . At each point  $t$  of the gradient, we randomly sampled 20 networks. We then computed, on any two locations  $t_1$  and  $t_2$  in the gradient, the value of  $\delta_{\Pi}^1$  at a microscopic and mesoscopic scales. We then visualize  $\delta_{\Pi}^1$  at both aggregation level for each pair of values of  $t$ , revealing two different patterns (Fig. S3b and S3c). The simulation highlights that patterns of network diversity strongly depend on the aggregation level at which the network is considered. At the macroscopic scale, no spatial structure emerges, since the connectance is constant along the gradient. At the microscopic scale, the interaction rewiring is too high to distinguish any structure. The dissimilarity values shifted from 0 along the diagonal to high values on the off-diagonal terms. However, at the mesoscopic scale, we can observe the dissimilarity pattern since the dissimilarity values slowly increase around the diagonal. As a result, this encourages one to look beyond species to investigate the spatial and temporal diversity of networks.

## 5 R code to generate the simulated data and the R plots of Fig. S3

```
#This script generates the simulated data and
#the R plots in Fig. S3 of Ohlmann, M., Miele, V., Dray, S., Chalmardrier, L., O Connor, L. & Thuiller,
W. (2018) Diversity indices for ecological networks: a unifying framework using Hill numbers.
Ecology Letters. doi:10.1111/ele.13221

#Copyright (C) 2018 Marc Ohlmann et al.

#This program is free software: you can redistribute it and/or modify it under the terms of the GNU
General Public License as published by the Free Software Foundation, either version 3 of the
License, or (at your option) any later version. See the GNU General Public License for more details
(http://www.gnu.org/licenses/)

#required packages
library(econetwork)
library(parallel)
library(ggplot2)
library(igraph)

J<-seq(-0.05,0.05,by=0.005) #range of variation of connectivities along the gradient
S<-100 #number of species per group
C<-0.1 #expected connectance of each network
R<-20 #number of replicates

#microscopic link dissimilarity
#####

#Computation of the microscopic dissimilarities in parallel

#matrix of index to parallelize
seq.mat<-matrix(0,ncol=2,nrow=length(J)*length(J))
comp<-0
for(i in 1:length(J)){
  for(j in 1:length(J)){
    comp<-comp+1
    seq.mat[comp,]<-c(i,j)
  }
}
#function that gives the an element of the pairwise dissimilarity matrix
dis.compute=function(k){
```

```

    ind.i<-seq.mat[k,1]
    ind.j<-seq.mat[k,2]
    dis.micro.loc<-dis.pairwise(list(g.list[[ind.i]],g.list[[ind.j]]),type='L')
  return(dis.micro.loc)
}

#microscopic link dissimilarity along the gradient and the replicates
dis.mat.list<-list()
for(l in 1:R){
  plot(l)
  g.list=c()
  for(lambda in J){
    pm<-cbind(c(0,C+lambda,0),c(0,0,C-lambda),c(0,0,0)) #matrix of connectivity for the SBM
    g <- sbm.game(3*S, pref.matrix=pm, block.sizes=c(S,S,S),directed = T) # this samples a stochastic
      block model
    V(g)$name<-as.character(1:(3*S))
    g.list<-c(g.list,list(g)) # these are the simulated data strictly speaking
  }

  dis.micro<-matrix(0,length(I),length(I))
  res<-mclapply(1:nrow(seq.mat),dis.compute,mc.cores = 16)#the machine used has 16 cores
  res<-do.call(rbind, res)
  dis.micro<-res
  dis.mat<-matrix(dis.micro,ncol=length(J),nrow=length(J))
  dis.mat.list<-c(dis.mat.list,list(dis.mat))
}

dis.mat.mean<-Reduce("+", dis.mat.list) / length(dis.mat.list) #averaging the dissimilarities of the
  replicates

#plot the results
x <- J
y <- J
data <- expand.grid(t1=x, t2=y)
data$dissimilarity <- as.vector(dist.mat.mean)

plot1<-ggplot(data, aes(t1, t2, z= dissimilarity)) + geom_tile(aes(fill = dissimilarity)) + theme_
  classic()+scale_fill_gradient(low="white", high="black")+labs(fill=expression(atop(delta[Pi]^1)))
plot1

#mesoscopic link dissimilarity
#####
#we compute the mesoscopic dissimilarity
#directly from the group probability matrix (and not from the networks)
#since, in this case, it is known (i.e. we do not infer the groups)

dis.horn<-function(t1,t2,C=0.1,eta){ #function to compute the Horn dissimilarity index
  L<-matrix(NA,ncol=2,nrow=2)
  L[1,1]<-C-t1
  L[1,2]<-C+t1
  L[2,1]<-C-t2
  L[2,2]<-C+t2
  div<-econetwork:::abgDecompQ(spxp = L,q = eta)
  res<-(log(div$Gamma)-log(div$mAlpha))/(log(2))
  return(res)
}

dis.horn.mat<-matrix(NA,ncol=length(J),nrow=length(J)) #pairwise Horn dissimilarity index matrix
  for(i in 1:length(J)){
    for(j in 1:length(J)){
      dis.horn.mat[i,j]<-dis.horn(J[i],J[j],eta=1)
    }
  }

#plot the results
x <- J
y <- J
data <- expand.grid(t1=x, t2=y)
data$dissimilarity <- as.vector(dis.horn.mat)
plot2<-ggplot(data, aes(t1, t2, z= dissimilarity)) + geom_tile(aes(fill = dissimilarity)) + theme_
  classic()+scale_fill_gradient(low="white", high="black")+labs(fill=expression(atop(delta[Pi]^1)))
plot2

```

Points locations  
Berghaus star projection

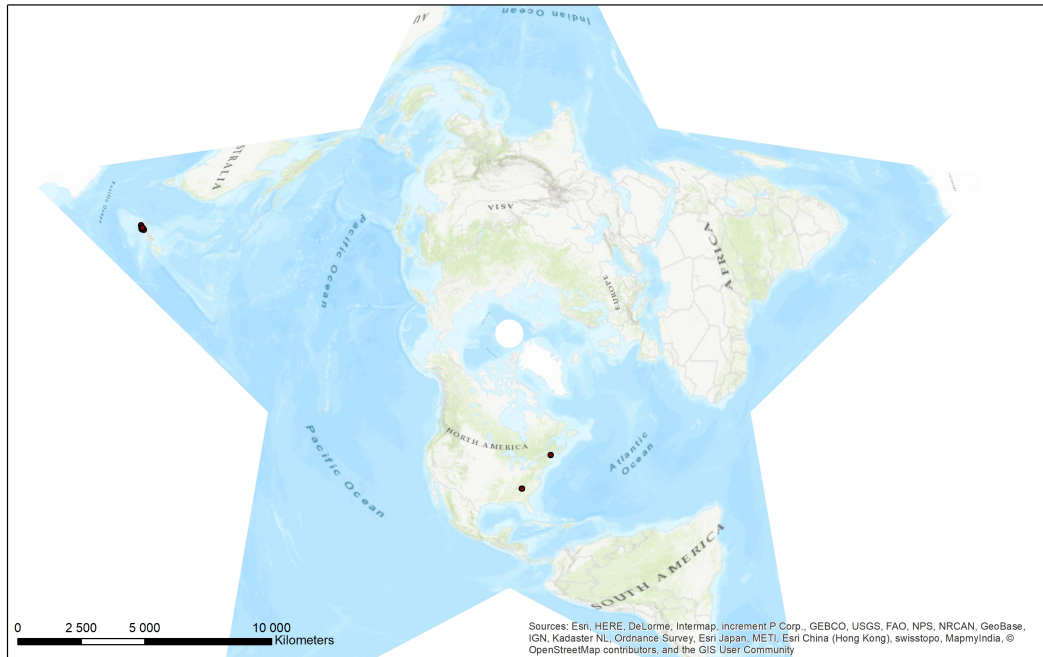


Figure S1: Locations of the ten trophic networks described in [Thompson & Townsend \(2003\)](#). Six networks were sampled in the USA and four networks in New Zealand





Figure S2: The metanetwork built from the 10 stream trophic networks. We only kept the largest connected component that contains 532 species and has a connectance value of 0.01. The network was visualized using the Gephi software (force Atlas 2 algorithm).

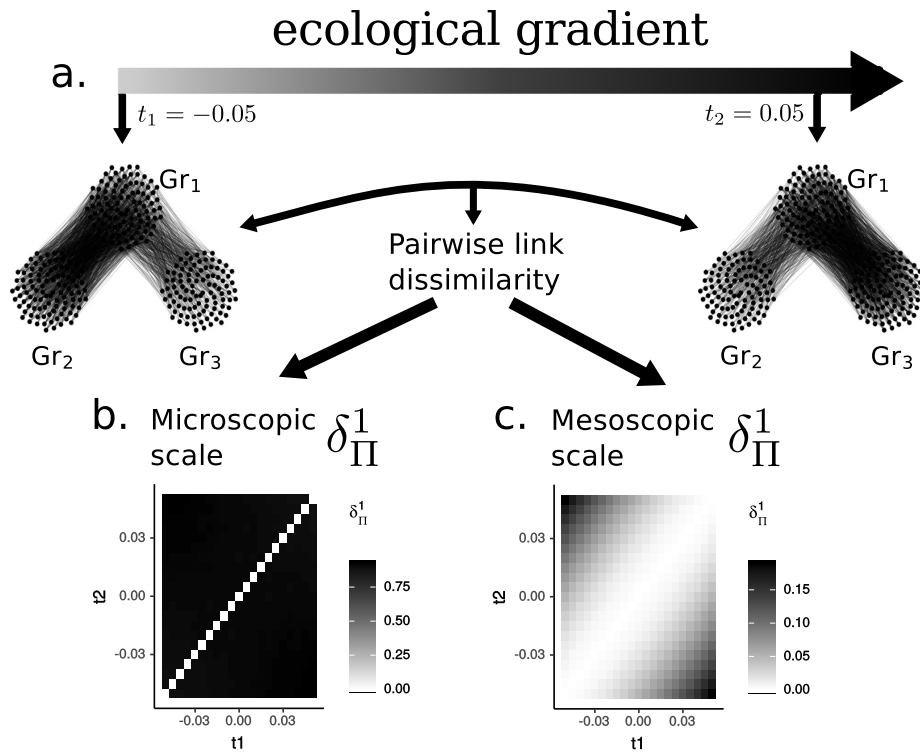


Figure S3: Change of network structure along an ecological gradient (a.) in function of two species aggregation levels (species and three equivalent groups of species in terms of probability of interactions with the others groups). We assess and represent the pairwise dissimilarity in link probabilities ( $\delta_{\Pi}^1$ ) at both a microscopic (b.) and mesoscopic scales (c.).

## References

- Allesina, S. & Pascual, M. (2009). Food web models: a plea for groups. *Ecology Letters* 12, 652–662.
- Chao, A. & Chiu, C.H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution* 7.
- Chao, A., Chiu, C.H. & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annual Review of Ecology, Evolution, and Systematics* 45, 297–324.
- Cirtwill, A.R., Riva, G.V.D., Gaiarsa, M.P., Bimler, M.D., Cagua, E.F., Coux, C. & Dehling, D.M. (2018). A review of species role concepts in food webs. *Food Webs* p. e00093.
- Daudin, J.J., Picard, F. & Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing* 18, 173–183.
- Gauzens, B., Thébault, E., Lacroix, G. & Legendre, S. (2015). Trophic groups and modules: Two levels of group detection in food webs. *J R Soc Interface* 12.
- Henderson, K., Gallagher, B., Eliassi-Rad, T., Tong, H., Basu, S., Akoglu, L., Koutra, D., Faloutsos, C. & Li, L. (2012). Rolx: Structural role extraction & mining in large graphs.
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439.
- 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, 1–21.
- Luczkovich, J.J., Borgatti, S.P., Johnson, J.C. & Everett, M.G. (2003). Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of Theoretical Biology* 220, 303 – 321.
- Mariadassou, M., Robin, S. & Vacher, C. (2010). Uncovering latent structure in valued graphs: A variational approach. *Ann. Appl. Stat.* 4, 715–742.
- Newman, M. & Leicht, E. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Science* 104, 9564–9569.
- Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecology Letters* 15, 1353–1361.
- 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, 303–312.
- Poisot, T., Guéveneux-Julien, C., Fortin, M.J., Gravel, D. & Legendre, P. (2017). Hosts, parasites and their interactions respond to different climatic variables. *Global Ecology and Biogeography* 26, 942–951.
- Reeve, R., Leinster, T., Cobbold, C., Thompson, J., N., Brummitt, M. & Matthews, L. (2014). How to partition diversity. *arXiv:1404.6520*.
- Thompson, R.M. & Townsend, C.R. (2003). Impacts on stream food webs of native and exotic forest: An intercontinental comparison. *Ecology* 84, 145–161.

# Théorie des métacommunautés d'espèces en interactions sur un réseau spatial

---

Ne cessons jamais de puiser à l'interminable source de la nature.

---

František Kupka

## 2.1 Synopsis

Les modèles de meta-population ont été introduits en écologie pour modéliser la dynamique de colonisation-extinction d'une espèce dans un environnement. Cet environnement peut être modélisé explicitement par un réseau spatial et le modèle stochastique de colonisation-extinction peut être formulé en utilisant des modèles graphiques. Nous montrerons que des modèles îles-continentes d'espèces en interaction peuvent être formulés en utilisant le même formalisme. Cela nous permettra de construire un modèle stochastique qui représente la dynamique de colonisation-extinction d'espèces qui interagissent à travers un réseau d'interaction sur un réseau spatial. Nous proposerons plusieurs modèles qui encodent différents mécanismes biologiques à partir de produits de réseaux d'interaction et spatial. Nous dériverons alors un modèle déterministe du type champ moyen à partir duquel nous définirons la notion de capacité de persistance de la méta-communauté. Nous nous intéresserons alors à la façon dont la structure du réseau spatial (ici sa modularité) et la structure du réseau d'interaction (ici l'hétérogénéité de la distribution des degrés) modulent conjointement la capacité de persistance de la méta-communauté pour les différents modèles proposés. Cette étude pourrait éclairer les questions concernant l'impact de la fragmentation du paysage sur les communautés d'espèces, en y intégrant l'impact de la structure du réseau d'interaction.

## 2.2 Proposition d'article (en anglais)

# A spatial theory of mutualistic metacommunity

Marc Ohlmann<sup>1</sup>, François Munoz<sup>1</sup>, Fabien Laroche<sup>2</sup>, Wilfried Thuiller<sup>1</sup>,

<sup>1</sup> Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Écologie Alpine, F-38000 Grenoble, France

<sup>2</sup> Irstea, Domaine des Barres, 45290, Nogent-sur-Vernisson, France

## Abstract

Predicting the long-term dynamics of a metacommunity in a complex spatial network and in which occupants interact in a heterogeneous way is a fundamental goal of Ecology. However, reaching such a goal requires to merge spatial network theory with interaction network theory. Here, we develop spatially realistic dynamic models of mutualistic metacommunities whose parameters are the spatial network and the biotic interaction network. To quantify the viability of metacommunity, we further extend the metapopulation capacity concept, that measures the viability of a metapopulation, to the metacommunity capacity. We then explore how the joint structure of the spatial and the biotic interaction networks shape the metacommunity capacity. We show that low-modular spatial network and power-law degree distribution of the interaction network is the most favourable configuration for the long-term persistence of a mutualistic metacommunity. Our models unify existing metapopulation models and island-mainland models of biotic interactions, allow to encode several explicit ecological hypothesis and has a close relationship with ecological network reconstruction.

## Introduction

One of the fundamental goals of community ecology is to predict the dynamics of many interacting species in a given region. Reaching such goal has direct implications for biodiversity management and conservation purposes to anticipate or mitigate the effects of global change on biodiversity.

Metapopulation models aim to predict the fate of a population that can colonise, persist or go extinct in a homogeneous or heterogeneous space. The heterogeneous space is often represented by a spatial network connecting suitable patches (Dale & Fortin 2010, Hagen *et al.* 2012). A network representation is suitable to represent the striking patchiness of the world, being natural or disturbed (Haddad *et al.* 2015). The simplest model of species occupancy was originally a deterministic model in a homogeneous landscape, whose long term-dynamics is determined various extinction and colonisation parameters (Levins 1969). This model is made of differential equations that predict the occupancy of a metapopulation, i.e. the probability of presence of a population in each patch (i.e. nodes of the spatial network) at equilibrium. Mirroring the emergence of stochastic theories in ecology and evolution (Hubbell 2001, Houchmandzadeh, 2014), a stochastic analogue of Levins model that encodes the mechanisms that act on species colonisation or extinction has been proposed and links between properties of the stochastic and deterministic model have been studied (Etienne & Nagelkerke 2002). To embed spatial heterogeneity in metapopulation models, spatially realistic metapopulation models have been developed. Two sources of heterogeneity can be embedded in metapopulation models, the heterogeneity of the parameters of colonisation/extinction (functional connectivity) and the heterogeneity on the spatial structure (structural connectivity). In the latter case, the impact of the topology of the spatial network on the stationary occupancy have been quantified (Gilarranz & Bascompte 2012) and give insights on longstanding debate on the impact of landscape fragmentation on metapopulations' viability (Fahrig 2003, Fletcher Jr *et al.* 2018). Importantly, the theory around the deterministic spatially realistic model allows to quantify the viability of a metapopulation from the mere structure of the spatial network (Ovaskainen & Hanski 2001, Hanski & Ovaskainen 2003), without simulating the whole metapopulation dynamics. To do so, Ovaskainen & Hanski introduced metapopulation capacities, quantities that can be seen as thresholds above which the metapopulation can survive.

However, metapopulation dynamics might be impacted by the dynamic of other species through biotic interactions, tangling so the fate of different species, and ultimately affecting the spatial coexistence of multiple metapopulations. In this respect, the use of metacommunity models to investigate the joint metapopulation dynamics is required. Interestingly, while biotic interactions have been formalized as interaction networks since they early days of community ecology (Elton 1927, Layman *et al.* 2015), most deterministic metacommunity did not have used so far the network representation. Moreover, these models have mostly focused on competition solely and the competition-colonisation trade-off in a homogeneous landscape (Tilman *et al.* 1997, Calcagno *et al.* 2006) and in some cases, on simplistic heterogeneous space (Amarasekare *et al.* 2004, Mouquet *et al.* 2011). Models focusing on other types of interactions (e.g. mutualism and trophic) have also been proposed but only for unrealistic species-poor communities (for two species, Nee *et al.* 1997), preventing the study of complex networks. Stochastic models of interactions where species are either present or absence have recently been developed. These models can encode mechanisms through specific rules such as having at least one prey present to survive in the Trophic Theory of Island Biogeography (Gravel *et al.* 2011, Massol *et al.* 2017) or through an increase of probability of presence with number of prey present in a model originally designed for inference (Auclair *et al.* 2017). Interestingly, the bestiary of probabilistic species interactions models have grown thanks to the statistical ecology literature. Indeed, aiming to reconstruct networks of interaction from co-occurrence data and/or time series, statistical ecology has defined *de facto* several phenomenological models of species interaction. Most of these models are graphical models, a class of statistical models of interaction that links probabilities to graph theory. Importantly, using the parametrisation of the model, several mechanisms can be encoded in the conditional probability tables as suggested in Staniczenko *et al.* (2017).

So far, only a few theoretical studies have explored the dynamic of a metacommunity with a spatially explicit environment and various interactions (e.g. for a three species food chain and with a lattice as spatial network Araújo & de Aguiar 2007), competition without using the graph formalism (Amarasekare *et al.* 2004) or mutualistic interactions on a lattice (Filotas *et al.* 2010). Despite a growing interest around mutualistic interactions in natural systems (Valdovinos 2019), no network based model of spatially realistic mutualistic metacommunities have yet been proposed.

Such model would allow to test jointly the impact of the structure of the spatial network and the biotic interaction network on the viability of a metacommunity. It would allow to reconcile the on-going debate on the impact of the structure of the spatial network (Fletcher Jr *et al.* 2018) and the impact of the power-law distribution of degrees in the interaction network (Dunne *et al.* 2002, Landi *et al.* 2018) on the dynamic and persistence of metacommunities. .

In this paper, we propose to explicitly model mutualistic biotic interactions and heterogeneous space using dynamic Bayesian networks. Importantly, various parameterisations of the model allow to encode different ecological hypotheses on extinction and colonisation. We derive the mean-field model and extend the metapopulation capacity *sensu* Ovaskainen & Hanski to mutualistic metacommunities, allowing so to assess the viability of a mutualistic metacommunity in a given spatial network. Using numerical methods, we show how metacommunity capacity relies on the structure of both the biotic interaction network and the spatial network. We contrast metacommunity capacities for different ecological hypothesis on the way spatial proximity of sites and biotic interactions modulate either colonisation or extinction probability. Moreover, we also explore the relationship between the degree of the nodes of both the spatial and the biotic interaction networks and occupancies at equilibrium. Such analysis allow to compute quantities on species across the sites (e.g. mean occupancy) or on sites across species (e.g. species diversity, interaction network diversity). We show that metacommunity viability might be seen in the light of both the structure of the spatial network and the biotic interaction network.

Table 1: Notations, name of the objects

Object	Name
$G_s$	Spatial network ( $n$ nodes)
$G_b$	Interaction network ( $m$ nodes)
$G_{s,b} = G_s \square G_b$	Cartesian product of the spatial and biotic interaction networks ( $n * m$ nodes)
$A_s$	Adjacency matrix of the spatial network
$A_b$	Adjacency matrix of the biotic interaction network
$A_{s,b} = A_s \otimes I_m + I_n \otimes A_b$	Adjacency matrix of the Cartesian product network)
$G_c$	colonisation network ( $n * m$ nodes)
$G_e$	Extinction network ( $n * m$ nodes)
$A_c$	Adjacency matrix of the colonisation network
$A_e$	Adjacency matrix of the extinction network
$\lambda_M$	Metacommunity persistence capacity
$\lambda_I$	Metacommunity invasion capacity
$\Lambda_s$	Dominant eigenvalue of the adjacency matrix of the spatial network
$\Lambda_b$	Dominant eigenvalue of the adjacency matrix of the biotic interaction network
$\Lambda_{s,b} = \Lambda_s + \Lambda_b$	Dominant eigenvalue of the adjacency matrix of the Cartesian product network

## Building stochastic models of metacommunity using dynamic Bayesian networks

We first present a formalism that unifies spatially realistic metapopulation models and mainland islands models of biotic interactions in discrete time using dynamic Bayesian networks, that describe dependencies between random variables at different time steps through a bipartite directed graph. While they are mostly used for inference (Auclair *et al.* 2017), they are also theoretical models whose parameters are networks (Lähdesmäki & Shmulevich 2008). More details on the theory of Dynamic Bayesian Networks are presented in Supporting Information. Since a dynamic Bayesian network can be interpreted as a causal Bayesian network (Pearl *et al.* 2009), the structure of the network represents the causal influences between species distribution between  $t$  and  $t + 1$ . Once the structure of causal influences is fixed, several distributions can be associated to a given network structure through different parametrisations. Those parameterisations allow to embed several mechanisms of interaction, by describing the effect of the presence of neighbour species or sites on the probability of presence of a given species at time  $t + 1$  chains that have desirable convergence properties (see Supporting Information).

The heterogeneous space is represented by a spatial network  $G_s = (V_s, E_s)$ . We assume that this network is undirected and connected, meaning that considering two nodes  $u$  and  $v$  of  $G_s$ , there

always exists a path from  $u$  to  $v$ . Moreover, interactions occurring in the metacommunity are represented by a biotic interactions network  $G_b = (V_b, E_b)$  that we also assume undirected and connected. We note  $n = |V_s|$  and  $m = |V_b|$  (see Table 1 for notations)

## Spatially realistic metapopulation model

Let  $X_i^t$  be the random variable associated to the presence of a population in the site  $i$  (i.e. the node  $v_i$  of  $G_s$ ) at time  $t$  ( $1 \leq i \leq n$ ,  $t \in \mathbb{N}^*$ ). The dependency structure is depicted using a dynamic Bayesian network, that is built from  $G_s$  (Fig. 1). This network is causal, meaning that the bipartite dependency network describes the causal influences between  $t$  and  $t + 1$ . Defining the neighbours of  $v_i$  in  $G_s$  as  $N_s(i)$ , the parents of  $X_i^{t+1}$  in the dynamic Bayesian network are  $\{X_i^t, \mathbf{X}_{N_s(i)}^t\}$ , meaning that the presence of a population at time  $t + 1$  is causally influenced by the presence of a population at time  $t$  in both site  $i$  and adjacent sites to  $i$ . Importantly, no other variables influence the presence of a population in site  $i$  at time  $t + 1$ . Through conditional probabilities, the parameterisation allows to encode the way the presence or absence of a population in adjacent sites will impact the presence of a population in the focal site. Here, we chose the same parametrisation as in [Gilarranz & Bascompte 2012](#).

$$Pr(X_i^{t+1} = 1 | X_i^t, \mathbf{X}_{N_s(i)}^t) = (1 - (1 - c)^{\sum_{k \in N_s(i)} X_k^t})(1 - X_i^t) + (1 - e)X_i^t \quad (1)$$

where  $c$  and  $e$  are the colonisation ( $0 < c < 1$ ) and extinction ( $0 < e < 1$ ) parameters, respectively. In this parametrisation, the probability of presence grows with the number of occupied adjacent sites. This model is a homogeneous Markov chain on  $\chi = \{0, 1\}^n$ . A state of the metapopulation is a binary vector of length  $n$  indicating whether each site is occupied or not. The dimension of the transition matrix is  $2^n * 2^n$  and the probability of transition between a state  $s_k = (x_1, \dots, x_n)$  and  $s_l = (\tilde{x}_1, \dots, \tilde{x}_n)$  is

$$P_{k,l} = Pr(X_1^{t+1} = x_1, \dots, X_n^{t+1} = x_n | X_1^t = \tilde{x}_1, \dots, X_n^t = \tilde{x}_n) \quad (2)$$

$$P_{k,l} = \prod_i Pr(X_i^{t+1} = x_i | X_i^t = \tilde{x}_i, \mathbf{X}_{N_s(i)}^t = (x_{N_s(i)})) \quad (3)$$

0 is an absorbing state in which the chain will be absorbed with probability equals to 1 in a finite time. However, the model will reach a quasi-stationary distribution before extinction giving, which gives a distribution of all possible states of the metapopulation across sites. Getting extinction time and quasi-stationary distribution require to compute eigenvectors and eigenvalues of  $P$  that are intractable in the general case. Nevertheless, the spatially implicit stochastic Levins model, where  $G_s$  is a complete graph have been studied ([Etienne & Nagelkerke 2002](#)).

## A mainland-island model with biotic interactions

Dynamic Bayesian networks can also be used to build mainland-island stochastic models of species interactions.

Let  $X_j^t$  be the random variable associated to the presence of population of species  $j$  on the island. The dependency structure is depicted using a dynamic Bayesian network built from  $G_b$  (Fig. 1). We also interpret this network as causal. Defining as  $N_b(j)$  the neighbours of  $v_j$  in  $G_b$ , the parents of  $X_j^{(t+1)}$  in the dynamic Bayesian network are  $\{X_j^t, \mathbf{X}_{N_b(j)}^t\}$ , meaning that the presence of a species  $v_j$  and adjacent species to  $v_j$  at time  $t$  on the island, causally influences the presence of a species at time  $t + 1$ . Importantly, other variables do not influence the presence of a species  $v_j$  at time  $t + 1$ . We chose a parameterisation similar to [Auclair \*et al.\* 2017](#):

$$Pr(X_j^{t+1} = 1 | X_j^t = 0) = c \quad (4)$$

$$Pr(X_j^{t+1} = 1 | X_j^t = 1, \mathbf{X}_{N_b(j)}^t) = 1 - e + (1 - (1 - e)^{\sum_{k \in N_b(j)} X_k^t})/M \quad (5)$$

where  $M$  is in normalising constant so that the probability of extinction belongs to  $]0, 1[$  (Supporting Information). The model is a homogeneous Markov chain on  $\chi = \{0, 1\}^m$  with no absorbing state. A state of the mainland-island model of species interaction is a binary vector of length  $m$ ,



representing the composition of the community. The dimension of the transition matrix is  $2^m * 2^m$  and the probability of transition between a state  $s_k = (x_1, \dots, x_m)$  and  $s_l = (\tilde{x}_1, \dots, \tilde{x}_m)$  is

$$P_{k,l} = \prod_j Pr(X_j^{t+1} = x_j | X_j^t = \tilde{x}_j, X_{N_b(j)}^t = (x_{N_b(j)})) \quad (6)$$

The chain converges towards a stationary distribution, a distribution of probability over all possible species communities. However, as in the metapopulation case, computing the stationary distribution is intractable in the general case. Importantly, in the metapopulation model, the spatial network is acting on the probability of colonisation whereas in the interaction model, the biotic network is acting on the probability of extinction.

## Spatially realistic models of mutualistic metacommunities

From the two previous models, we could then build a spatially explicit metacommunity model using  $G_s$  and  $G_b$ . To do so, we used the Cartesian product of graphs that allows to build a network from  $G_b$  and  $G_s$  (Imrich & Klavzar 2000), which is the network  $G_{s,d} = G_s \square G_b$  whose the set of nodes is  $V_d \times V_b$ . A node of this network is identified by a pair of species and site identities. Moreover, there is an edge between  $(u_s, u_b)$  and  $(v_s, v_b)$  if  $(u_s = v_s \text{ and } (u_b, v_b) \in E_s)$  or  $(u_b = v_b \text{ and } (u_s, v_s) \in E_b)$ .

Let  $X_{ij}^t$  be the random variable associated to the presence of a population of species  $i$  in node  $j$  at time  $t$ . The dependency structure is depicted using a dynamic Bayesian network that is built from  $G$  (Fig. 1). Defining as  $N(i, j)$  the neighbours of  $(v_i, v_j)$  in  $G_{s,b}$ , the parents of  $X_{ij}^{t+1}$  in the dynamic Bayesian network are  $\{X_{ij}^t, \mathbf{X}_{N(i,j)}^t\}$ . This means that the presence of a population of species  $j$  in site  $i$  at time  $t + 1$  is causally influenced by the presence of population of the same species in adjacent site to  $i$  at time  $t$  and by the presence of populations of species that interact with  $j$  in the same site. In the parameterisation of the previous mainland-island models of interaction, we assumed that colonisation probability is constant and that interactions act on the extinction probability contrary, to the metapopulation model.

Here, several parameterisations can be built from a given product graph, using either the spatial network or the colonisation network to modulate colonisation and/or extinction probability.

Let  $G_s^0$  be the network that has the same set of nodes as  $G_s$  but an empty set of edges, and let  $G_b^0$  be the network that has the same set of nodes as  $G_b$  but an empty set of edges. We introduce then the colonisation network  $G_c$  ( $A_c$  is its adjacency matrix) and the extinction network  $G_e$  ( $A_e$  is its adjacency matrix). Then, depending on the hypothesis to test, four sets of parameterisations can be built from a given product graph (Table 2) :

- a Levins type model where both the spatial and biotic interaction networks modulate the colonisation probability ( $G_c = G_s \square G_b$ ) and the extinction probability is constant ( $G_e = G_s^0 \square G_b^0$ )
- a separated-effect model where the spatial network modulates the colonisation probability ( $G_c = G_s \square G_b^0$ ) and where the biotic interaction network modulates the extinction probability ( $G_e = G_s^0 \square G_b$ )
- a combined model where both the spatial and the biotic interaction networks modulate the colonisation probability ( $G_c = G_s \square G_b$ ) and the biotic interaction network modulates the extinction probability ( $G_e = G_s^0 \square G_b^0$ )
- a rescue effect model where the spatial network modulates the colonisation probability ( $G_c = G_s \square G_b^0$ ) and both the spatial and the biotic interaction networks modulate the extinction probability ( $G_e = G_s \square G_b$ ).

For the four models, the conditional probabilities of colonisation and non-extinction are expressed as:

$$Pr(X_{i,j}^{t+1} = 1 | X_{i,j}^t = 0, \sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t) = \epsilon + (1 - \epsilon) * (1 - (1 - c)^{\sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t}) \quad (7)$$

$$Pr(X_{i,j}^{t+1} = 0 | X_{i,j}^t = 1, \sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t) = e - (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t})/M \quad (8)$$

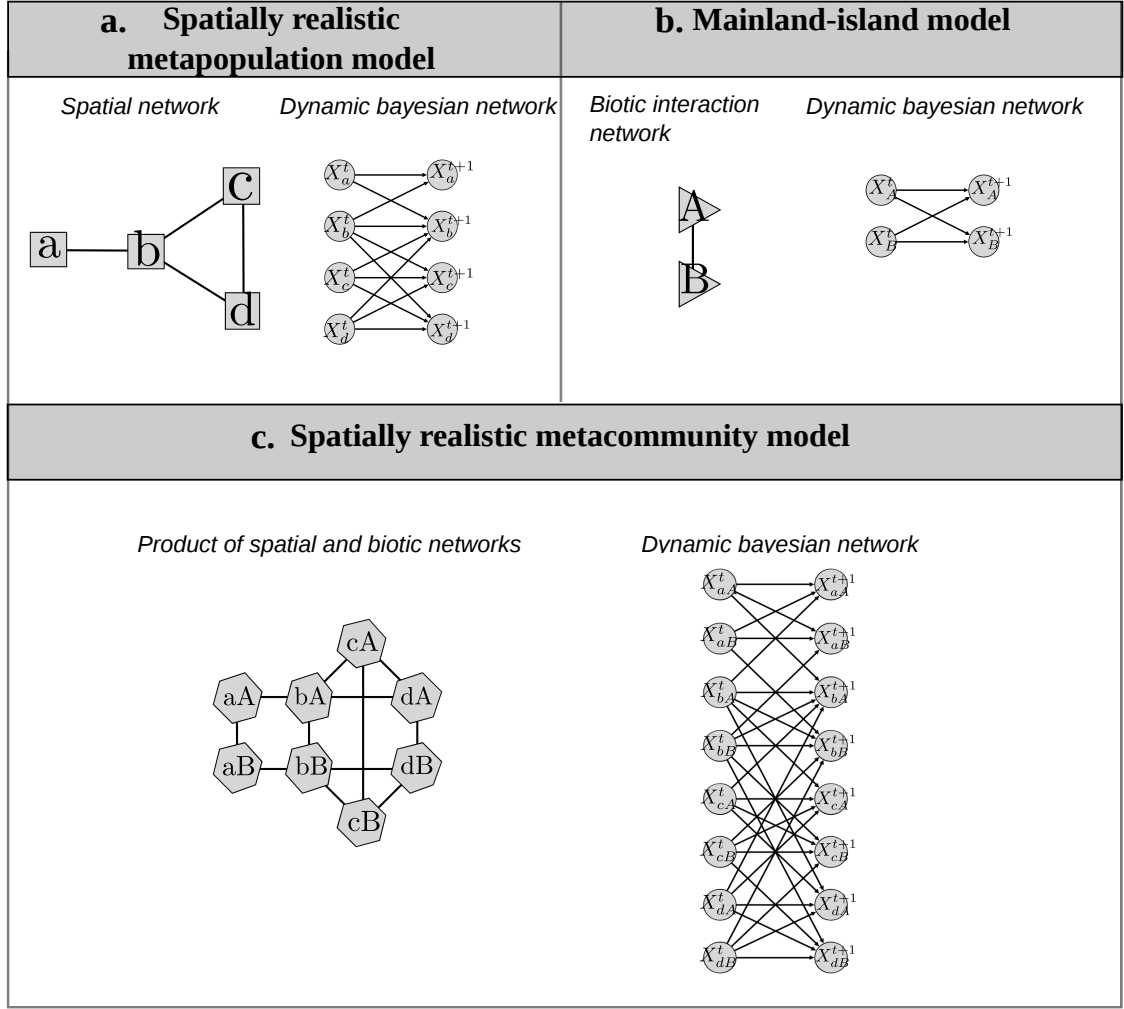


Figure 1: Metapopulation models, island-mainland interaction models and metacommunity models as dynamic Bayesian networks: a. the metapopulation model. The spatial network is represented on the left and the dynamic Bayesian network, represented on the right, described the map of causal influences between the variables that represent the presence of a population of a given species between  $t$  and  $t + 1$ . b. Mainland island model of biotic interaction. The biotic interaction network is represented on the left whereas the associated dynamic Bayesian network is represented on the right. c. Spatially realistic metacommunity model. On the left, the product network of the spatial and the biotic interaction network. On the right, the dynamic Bayesian network that maps causal influences between the variables that represent the presence of a species in a site.

where  $\epsilon \in ]0; 1[$  is a constant that guarantess the convergence of the model,  $M$  is a normalising constant so that the probability of extinction is in  $]0, 1[$  (Supporting Information), and  $N_{G_c}(i, j)$  (resp.  $N_{G_e}(i, j)$ ) denotes the neighbours of  $(i, j)$  in  $G_c$  (resp.  $G_e$ ). The model is a homogeneous Markov chain on  $0, 1^{mn}$ . A state of the metacommunity is a binary matrix of size  $m * n$  that is a metacommunity binary matrix. The dimension of the transition matrix is  $2^{mn} * 2^{mn}$  and the probability of transition between a state  $s_k = (x_{1,1}, \dots, x_{m,n})$  and  $s_l = (x_{1,1}, \dots, x_{m,n})$  is :

$$P_{k,l} = \prod_{i,j} Pr(X_{i,j}^{t+1} = x_{i,j} | X_{1,1}^t = x_{1,1}, \dots, X_{m,n}^t = x_{m,n}) \quad (9)$$

The chain converges towards a stationnary distribution. Computing the stationary distribution is also intractable in the general case, but, it is however possible to simulate the dynamic of the metacommunity as [Gillarranz & Bascompte \(2012\)](#) did for metapopulation model.

## The mean-field deterministic model

Since studying the stochastic model is intractable, we propose to study the model under the mean field approximation, extending so the spatially realistic Levins model and benefiting from the mathematical results of it (Ovaskainen & Hanski 2001). The approximation is derived from Bianconi (2018). The aim is to study the dynamic of occupancy of each species  $j$  in each site  $i$ , i.e.  $p_{ij}(t) = \mathbb{E}(X_{ij}^t)$ . For all  $i$  and  $j$ , we have

$$X_{i,j}^{t+1} = (1 - X_{i,j}^t)(\epsilon + (1 - \epsilon)(1 - (1 - c)^{\sum_{(k,l) \in N_{G_c(i,j)}} X_{k,l}^t})) + (1 - e + (1 - (1 - e)^{\sum_{(u,v) \in N_{G_e(i,j)}} X_{k,l}^t})/M)X_{i,j}^t \quad (10)$$

It follows :

$$p_{i,j}(t+1) = \mathbb{E}((1 - X_{i,j}^t)(\epsilon + (1 - \epsilon)(1 - (1 - c)^{\sum_{(k,l) \in N_{G_c(i,j)}} X_{k,l}^t}))) + \mathbb{E}((1 - e + (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e(i,j)}} X_{k,l}^t})/M)X_{i,j}^t) \quad (11)$$

This approach leads to a hierarchy of equations that cannot be solved (i.e. we need to consider  $\mathbb{E}(X_{1,1}^t, \dots, X_{m,n}^t)$  to find a solution to the system). A drastic approximation consists in the mean field approximation, for any sequence of indices  $n(1), n'(1); \dots, n(r), n'(r)$ , we assume :

$$\mathbb{E}(X_{n(1),n'(1)}^t, \dots, X_{n(r),n'(r)}^t) \simeq \mathbb{E}(X_{n(1),n'(1)}^t) \dots \mathbb{E}(X_{n(r),n'(r)}^t) \quad (12)$$

After some calculus, introducing a new single index  $v$  for the nodes of the product network and assuming that  $c \ll 1, e \ll 1$  and  $\epsilon \ll c$  (see Supp. Inf.), it follows :

$$p_v(t+1) - p_v(t) = C_v(p(t))(1 - p_v(t)) - E_v(p(t))(p_v(t)) \quad (13)$$

where  $C_v(p(t)) = c \sum_u [A_c]_{vu} p_u(t)$  and  $E_v(p(t)) = e(1 - \sum_u [A_e]_{vu} p_u(t)/M)$ . This is the equation (1) of Ovaskainen & Hanski (2001). We are interested in determining the equilibrium states and assess their stability to assess the viability of a given mutualistic metacommunity.

## Metacommunity capacities

Originally defined for metapopulation (Hanski & Ovaskainen 2000, Ovaskainen & Hanski 2001) metapopulation capacities allow to predict the fate of a population in function of extinction and colonisation parameters and the spatial network, without simulating the dynamic. The metapopulation invasion capacity  $\lambda_I$  assesses whether a species can invade a spatial network, and the metapopulation persistence capacity  $\lambda_M$ , measures under what conditions on colonisation and extinction parameters a metapopulation will persist at equilibrium. We propose to extend the metapopulation capacities to metacommunity capacities. In our mutualistic metacommunity setting, starting from a non-null occupancy, the metacommunity will persist at equilibrium if  $\lambda_M \geq \frac{\epsilon}{c}$ , leading to a non-null occupancy in every site for every species.

The Levins type model (Table 2) is actually a spatially realistic Levins model where the product network is analogue to the spatial network in the classic spatially realistic metapopulation model. In this case,  $\lambda_M = \lambda_I = \Lambda_b + \Lambda_c$  (see Supporting Information for proof). Notice that in this case, the biotic interaction and the spatial networks play interchangeable roles. For the combined effect model, the separated effect model and the rescue effect model, we computed the metacommunity capacity  $\lambda_M$  using Appendix D of Ovaskainen & Hanski 2001 and simulating annealing. See Supporting Information for more details on metacommunity capacities and implementation in R and Python.

## Degree distribution of interaction network and spatial network modularity jointly shapes the metacommunity capacity

In this section, we are interested in how the structure of the spatial network and the interaction network shapes the metacommunity capacity of a bipartite mutualistic system. In order to mimic a gradient of landscape fragmentation, we used a gradient of modularity while keeping constant the number of causal influences (i.e. the number of edges of the spatial network). To do so, we generated spatial random graphs using a stochastic block model, with 10 nodes, an increasing modularity and a constant expected connectance along the gradient made of 4 location ( $C = 0.25$ , more details in Supporting Information). We only kept connected spatial networks and used 50 replicates per

## Temporal dynamic

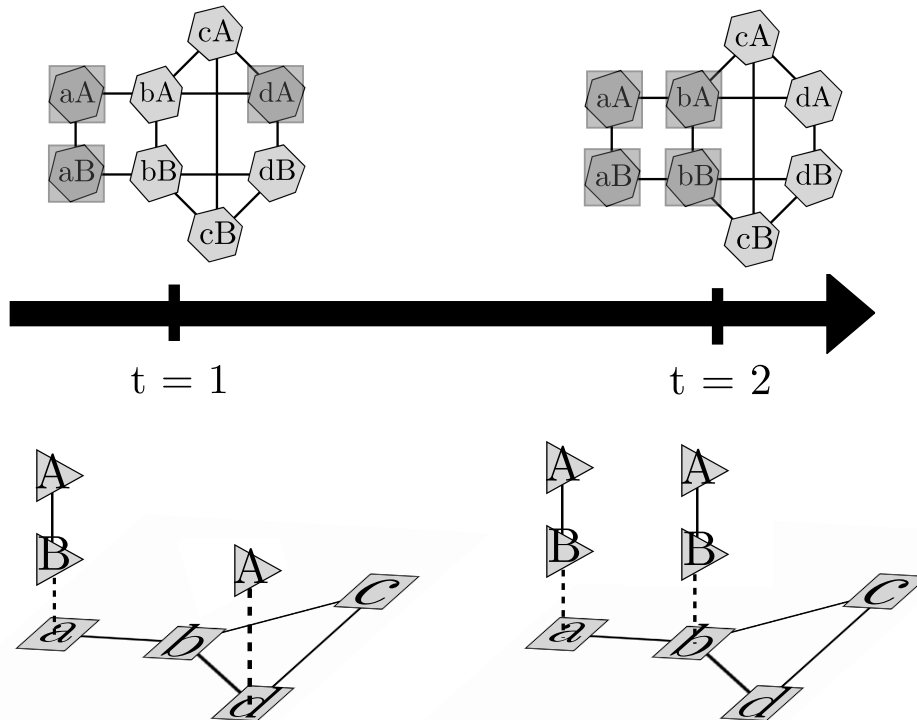


Figure 2: Simulating a dynamic in the combined effect model between two time steps. Upper: the nodes of the product network are either empty or occupied (grey squares). For the sake of simplicity, the model here is turned deterministic ( $c = 1$ ,  $e = 1$ ). To colonise a new node of the product network, species *A* and *B* must be present in the same site and can colonise adjacent site only. The population of species *A* originally present in site *d* goes extinct since it does not co-occur with *B* at *t* whereas species *A* and *B* co-occurring in site *a* colonise the site *b*.

Table 2: The different models using various parameterisations

Name of the model	Mechanisms and hypothesis	Metacommunities capacities
Levins type model	The spatial and the biotic network modulate the colonisation probability. The extinction probability is constant. The different sites and species acts independently on the probability of presence of a species.	$\lambda_I = \lambda_M = \Lambda_d + \Lambda_b$
Combined effect model	The spatial and the biotic network modulate the colonisation probability. The biotic interaction network modulates the extinction probability. The different sites and species acts independently on the probability of presence of a species.	$\lambda_I$ : to compute from jacobian matrix $\lambda_M$ : to compute numerically
Separated effect model	The spatial network modulates the colonisation probability. The biotic network modulates the extinction probability. The different sites and species acts independently on the probability of presence of a species.	$\lambda_I$ : to compute from jacobian matrix $\lambda_M$ : to compute numerically
Rescue effect model	The spatial network modulates the colonisation probability. The spatial and the biotic network modulates the extinction probability. The different sites and species acts independently on the probability of presence of a species.	$\lambda_I$ : to compute from jacobian matrix $\lambda_M$ : to compute numerically

location of the modularity gradient. Concerning the mutualistic network, we simulated different bipartite networks with various shapes of the degree distribution, from the shape generated by a random network (Erdős-Renyi) to power-law degree distributions with various parameters, using 16 nodes and 28 edges (more details in Supporting Information). We held constant the number of edges of the different biotic interaction networks. To do so, we used the function *sample\_fitness\_pl* implemented in the R package *igraph* (Csardi & Nepusz 2006). We only kept connected mutualistic networks and used 50 replicates per location of the gradient and 50 replicates also per power-

law constant for the interaction network. We computed then the colonisation and extinction networks for combinations of spatial and biotic interaction networks, generating so 50 different networks per location of the gradients. We first compute the metacommunity capacities for each combination of the spatial and the interaction networks, in order to assess the viability range of the metapopulation. Moreover, knowing that a metacommunity will persist given the parameters, we are interested in studying how equilibrium occupancies and aggregated quantities build from these occupancies (mean occupancy, species diversity) depend on the characteristics of the nodes of both networks.

## Computing metacommunity capacities

We computed the metacommunity persistence capacity  $\lambda_M$  along the gradients of network structure for the four different models (Fig. 3). For the Levins type model, the metacommunity capacity decreases when the spatial network becomes more modular and also decreases when the degree distribution gets away from a power-law. In this case, the modularity of the spatial network has a stronger impact on the metacommunity capacity compared to the structure of the biotic interaction network. For the combined effect model we observed a similar trend concerning the variation of the metacommunity capacity with the structure of the two networks and the relative impact of the modularity. In this case, the metapopulation capacities are higher compared to the Levins type model. Indeed, the colonisation network are the same in the Levins type model and the combined effect model whereas, in the combined effect model, mutualistic interactions tend to decrease the extinction probability contrary to the Levins type model. For the separated-effect model and the rescue effect model, we observed similar patterns except that metacommunity capacities are significantly lower than the Levins type and combined effect model. In this two latter cases, the interactions do not modulate the colonisation probability.

## A focus on occupancies at equilibrium for a given network combination

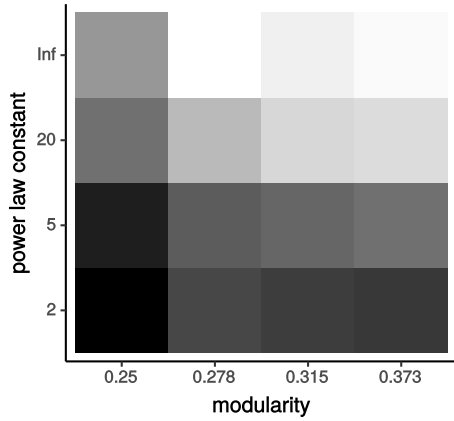
In this section, we focus on a given network combination and use extinction and colonisation parameters so that the metacommunity persists. Then, we are interested in simulating the dynamics and study how the occupancy at equilibrium of each node of the product network depends on its degree for the Levins type model (and for the three other models in Supporting Information). Moreover, we study the mean occupancy of species across sites and the species and link diversity in each site. We used a spatial network with modularity of 0.23 and a biotic interaction network with a degree distribution sampled in a power law of parameter 2. Moreover, we used a colonisation parameter  $c = 0.1$  and an extinction parameter  $e = 0.1$ . With these parameters, the extinction-colonisation ratio is above the metacommunity capacity and the metacommunity persists, leading to non-null occupancies at equilibrium of each species in each site.

We represented the occupancy of the nodes of the product network (that is the colonisation network in this combined effect model) in function of their degree (Fig. 4). The occupancy of the nodes of the product network (indexed by a species and a site) increases with the degree of these nodes. Moreover, in this model, at a fixed node degree of the product network, the occupancy decreases with the ratio of the degree of the site over the degree of the node of the product network. This means that nodes of the product network that combines a generalist species with a low-connected site will have a higher occupancy at equilibrium compared to nodes that combines a specialist species with a highly connected site. We observed the same patterns for the three other models (Supporting Information). Furthermore, we computed, from the occupancies at equilibrium, species  $\alpha$ -diversities in each site using the framework developed in (Ohlmann *et al.*, 2019) with  $\eta = 2$  (Fig. 5a). We observed positive relationship between species  $\alpha$ -diversity and the degree of the nodes of the spatial network (Fig. 5b). Mirroring the analysis on the spatial network, we represented the mean occupancy across of the sites of the different species (Fig. 5c) and observed a positive relationship between mean occupancy of a species and its degree in the biotic interaction network.

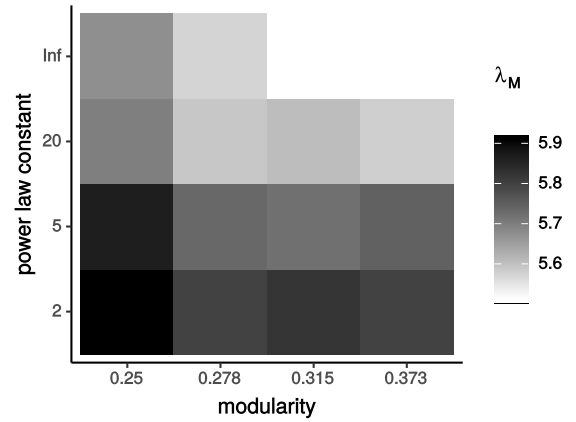
## Discussion

In this paper, we proposed a stochastic spatially explicit model of mutualistic metacommunities by integrating spatial and a biotic interaction networks. Using the mean-field approximation, we

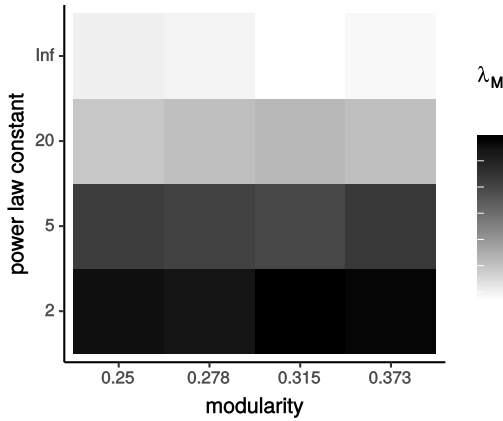
a. Levins type model



b. Combined effect model



c. Separated effect model



d. Rescue effect model

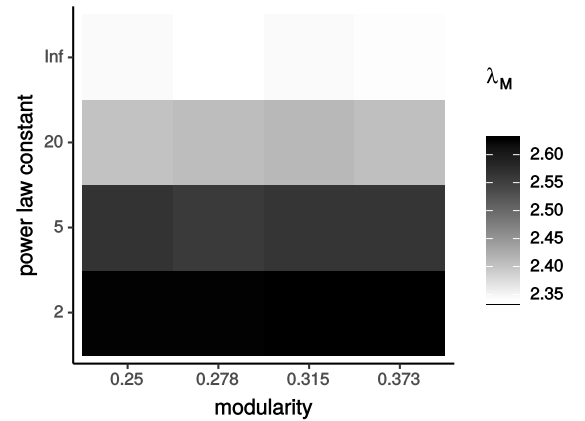


Figure 3: Assessing metacommunity persistence capacity in function of the structure of the spatial network and the biotic interaction network. a. Metacommunity persistence capacity for the Levins type model in function of the mean modularity of the sampled spatial networks and the power law constant of the degree distribution of the biotic b. Metacommunity persistence capacity for the combined effect model. c. Metacommunity persistence capacity for the separated effect model. d. Metacommunity persistence capacity for the rescue effect model.

derived a deterministic mutualistic metacommunity model and extend the metapopulation capacity *sensu* Ovaskainen & Hanski to metacommunity capacity. Metacommunity capacity allows to assess viability of the metacommunity from the parameters only (colonisation, extinction parameter, spatial and biotic interaction network), without having to simulate the dynamic of the system.

Importantly, as shown by the case study, the spatial network and the biotic interaction network shape the metacommunity capacity. It means that any viability statement on a metacommunity should be done with relevant representations of the spatial network and the biotic interaction network. Moreover, studying the dynamics of the occupancy of the nodes of the product network (made of combination of sites and species) allows to ultimately study aggregated quantities build from occupancies (e.g.  $\alpha$ -diversity of the sites or mean occupancy of the species). This recalls the duality diagram defined formally in multivariate analysis where sites can be seen in the space of species or species in the space of sites (Dray *et al.* 2007, Holmes *et al.* 2008).

There are two noticeable properties of our proposed spatially explicit model of mutualistic metacommunities. First, the use of dynamic Bayesian networks allow to build models from spatial

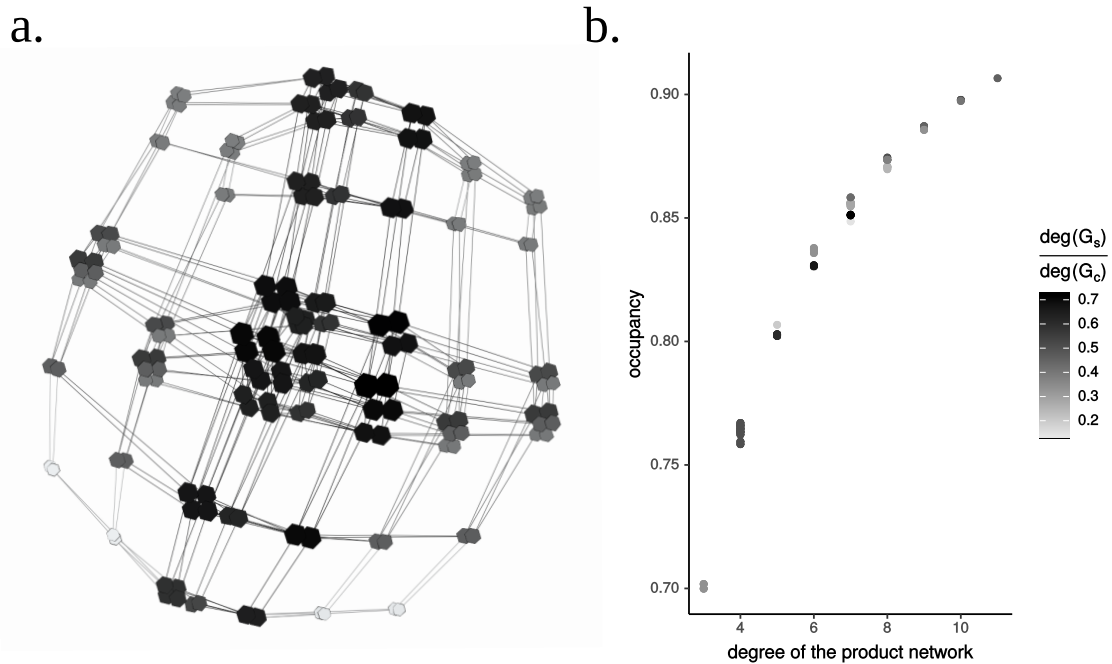


Figure 4: Simulating the dynamics for a given spatial and biotic interaction network with the Levins type model. a. colonisation network whose size of the nodes is proportional to their degree and color indicates the occupancy at equilibrium (grey: low occupancy, black: high occupancy). b. Relationship between the occupancy at equilibrium and the degree of the node of the product network. Each point of the relationship (corresponding to a node of the product graph) is colored according to the ratio of the degree of the site in the spatial network over the degree of the focal node in the colonisation network.

and interaction networks in an analogous ways, even though their nodes represent distinct entities (site or species) and distinct relationships (geographic proximity or biotic interaction). Second, our mathematical model unifies the well-studied metapopulation models and mainland-island interactions models, making these two models particular cases of the developed framework. For example, the spatially realistic Levins model is made of a combination of a spatial network and a biotic interaction network made of a single node.

Interestingly, while both spatial network and biotic interaction networks are somehow analogous in the dynamic Bayesian network formalism, they modulate the colonisation and extinction probabilities in different manners. Indeed, if the structure of the networks encode the causal relationships between variables (sites and species), the different parameterisations allow to encode the way interactions between species and proximity between sites influence the species' probability of presence/absence by using the colonisation and extinction networks. We propose here four different models where the spatial network and the biotic interaction network can affect species colonisation or extinction rates (Fig. 3). When modelling a given metacommunity, careful thoughts should be given to the way biotic interactions and spatial networks act on colonisation and extinction mechanisms. These hypothesis can drastically affect the persistence of a metapopulation/metacommunity leading for instance, to non-extinction in the case of the rescue effects (Huth *et al.* 2015). The proposed formalism of extinction and colonisation network allows to encode the different mechanisms through product graphs. Moreover, even in a given model, other shape of conditional probabilities might be proposed to encode several mechanisms, using logic-based rules as suggested in (Staniczenko *et al.*, 2017) and (Bohan *et al.*, 2011) for network reconstruction and turning so phenomenological models into mechanistic models.

The Trophic Theory of Island Biogeography embeds a stochastic mainland-island model of interaction with a different perspective compared to the proposed mainland-island model of species interaction. It is made of a drastically simple logic rule: having at least one prey to survive (Gravel *et al.* 2011, Massol *et al.* 2017). The causal assumption behind this rule is that interactions are perceived as a bottom-up dependency. Such formalism of the TTIB constraints the trophic

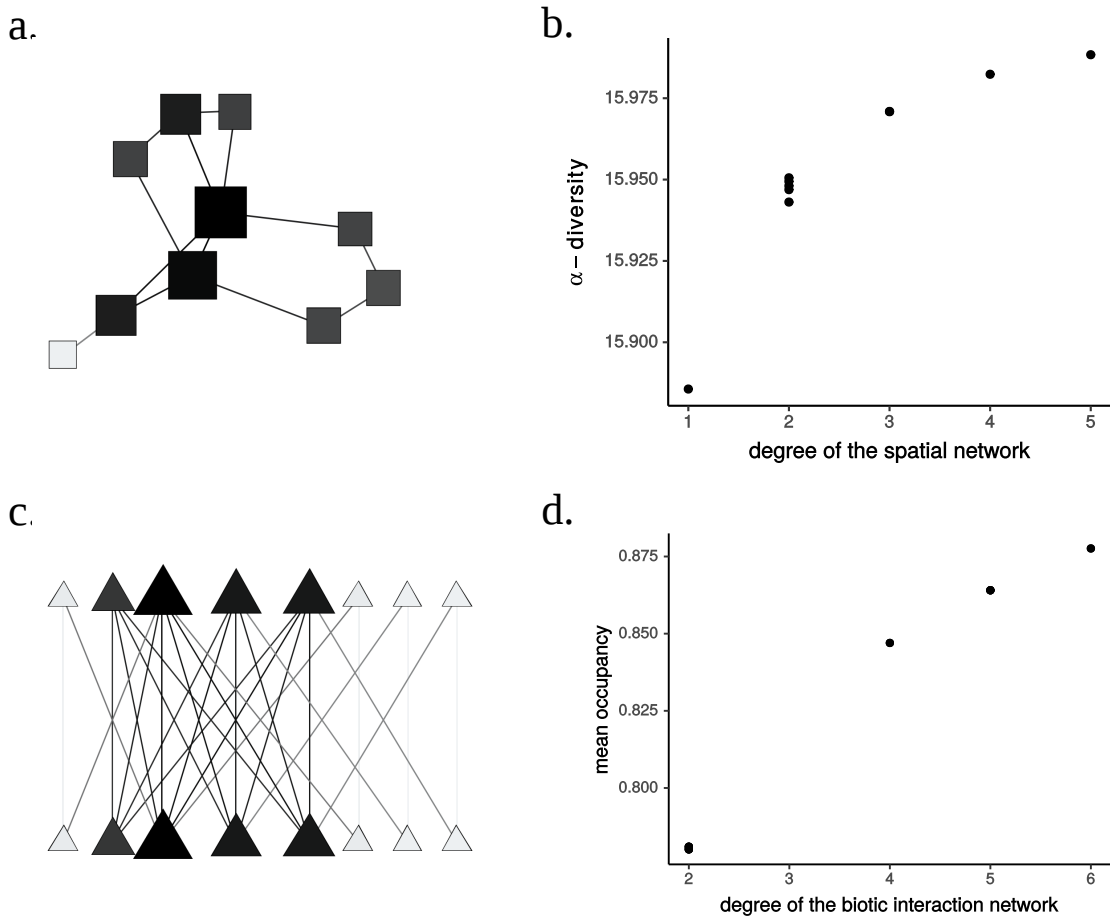


Figure 5: Aggregated statistics from occupancy at equilibrium for the Levins type model in the spatial network and the biotic interaction network. a. Spatial network whose size of the nodes is proportional to their degree and colour indicates the species  $\alpha$ -diversity at equilibrium (grey: low  $\alpha$ -diversity, black: high  $\alpha$ -diversity). b. Relationship between the species  $\alpha$ -diversity at equilibrium and the degree of the sites in the spatial network. c. Biotic interaction network whose size of the nodes is proportional to their degree and colour indicates the mean occupancy across the sites at equilibrium (grey: low  $\alpha$ -diversity, black: high  $\alpha$ -diversity). d. Relationship between the mean occupancy at equilibrium and the degree of the species in the biotic interaction network.

network to be directed acyclic. Interestingly, [Massol \*et al.\* \(2017\)](#) have also seen the model of the TTIB as a Markov chain). However, in the TTIB, the state space is not on a space that is not all possible species assemblages contrary to the proposed framework but only the assemblages where there exists at least a path between any species and a basal species (the TTIB-compatible species assemblages *sensu* [Massol \*et al.\*](#)). Still, as the proposed framework, the stochastic model is not tractable. Contrary to the TTIB, the proposed model can embed loops in the mutualistic interaction network. These loops can be found in trophic interactions networks ([Neutel \*et al.\* 2002](#)) and the proposed model of interaction, adapted from [Auclair \*et al.\* \(2017\)](#) can embed loops in the interaction network. Indeed, this last model describes the causal influences on species presence between  $t$  and  $t + 1$  whereas the TTIB describes the dependencies between the members of the community at a given time step.

Finally, the proposed metacommunity model is built using the same mathematical objects (dynamix Bayesian networks that belongs to graphical models) than those used as tools to reconstruct interaction network. Such task is a hard unsolved problem in statistical ecology ([Freilich \*et al.\* 2018](#)), even when using long term time-series ([Sander \*et al.\* 2017](#)). One conceptual hurdle is the polymorphism of the mathematical representation of biotic interactions. Indeed, the ecological theory describing the biotic interaction phenomenon like Lotka-Volterra models or the coexistence theory ([Chesson 2000, 2018](#)) do not use the same mathematical models of interaction compared



to those used for network inference, even if a network is a parameter of the model in both cases. This complicates the interpretation of reconstructed network since they cannot be read in the light of classic theoretical ecology. Interestingly, the proposed framework builds ecological theory using the same mathematical objects (graphical models) as the ones used for network reconstruction and might so reduce the gap between theoretical ecology and statistical ecology, aiming to build a unified ecological network science.

## References

- Amarasekare, P., Hoopes, M.F., Mouquet, N. & Holyoak, M. (2004). Mechanisms of coexistence in competitive metacommunities. *The American Naturalist* 164, 310–326.
- Araújo, S.B. & de Aguiar, M.A. (2007). Pattern formation, outbreaks, and synchronization in food chains with two and three species. *Physical Review E* 75, 061908.
- Auclair, E., Peyrard, N. & Sabbadin, R. (2017). Labeled dbn learning with community structure knowledge. In: *Joint european conference on machine learning and knowledge discovery in databases*. Springer pp. 158–174.
- Bianconi, G. (2018). *Multilayer Networks: Structure and Function*. Oxford University Press; p.60-62.
- Bohan, D.A., Caron-Lormier, G., Muggleton, S., Raybould, A. & Tamaddoni-Nezhad, A. (2011). Automated discovery of food webs from ecological data using logic-based machine learning. *PLoS One* 6, e29028.
- Calcagno, V., Mouquet, N., Jarne, P. & David, P. (2006). Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecology letters* 9, 897–907.
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics* 31, 343–366.
- Chesson, P. (2018). Updates on mechanisms of maintenance of species diversity. *Journal of Ecology* 106, 1773–1794.
- Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.
- Dale, M. & Fortin, M.J. (2010). From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics* 41.
- Dray, S., Dufour, A.B. *et al.* (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* 22, 1–20.
- Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002). Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences* 99, 12917–12922.
- Elton, C.S. (1927). *Animal ecology*. New York, Macmillan Co., p. 58. <https://www.biodiversitylibrary.org/bibliography/7435>.
- Etienne, R.S. & Nagelkerke, C. (2002). Non-equilibria in small metapopulations: comparing the deterministic levins model with its stochastic counterpart. *Journal of Theoretical Biology* 219, 463–478.
- Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual review of ecology, evolution, and systematics* 34, 487–515.
- Filotas, E., Grant, M., Parrott, L. & Rikvold, P.A. (2010). The effect of positive interactions on community structure in a multi-species metacommunity model along an environmental gradient. *Ecological Modelling* 221, 885–894.
- Fletcher Jr, R.J., Didham, R.K., Banks-Leite, C., Barlow, J., Ewers, R.M., Rosindell, J., Holt, R.D., Gonzalez, A., Pardini, R., Damschen, E.I. *et al.* (2018). Is habitat fragmentation good for biodiversity? *Biological conservation* 226, 9–15.
- Freilich, M.A., Wieters, E., Broitman, B.R., Marquet, P.A. & Navarrete, S.A. (2018). Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* 99, 690–699.
- Gilarranz, L.J. & Bascompte, J. (2012). Spatial network structure and metapopulation persistence. *Journal of Theoretical Biology* 297, 11–16.

- Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecology Letters* 14, 1010–1016.
- Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D. *et al.* (2015). Habitat fragmentation and its lasting impact on earth’s ecosystems. *Science advances* 1, e1500052.
- Hagen, M., Kissling, W.D., Rasmussen, C., De Aguiar, M.A., Brown, L.E., Carstensen, D.W., Alves-Dos-Santos, I., Dupont, Y.L., Edwards, F.K., Genini, J. *et al.* (2012). Biodiversity, species interactions and ecological networks in a fragmented world. 46, 89–210.
- Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature* 404, 755.
- Hanski, I. & Ovaskainen, O. (2003). Metapopulation theory for fragmented landscapes. *Theoretical population biology* 64, 119–127.
- Holmes, S. *et al.* (2008). Multivariate data analysis: the french way. In: *Probability and statistics: Essays in honor of David A. Freedman*. Institute of Mathematical Statistics pp. 219–233.
- Houchmandzadeh, B. (2014). The remarkable discreteness of being. *Journal of biosciences* 39, 249–258.
- Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press.
- Huth, G., Haegeman, B., Pitard, E. & Munoz, F. (2015). Long-distance rescue and slow extinction dynamics govern multiscale metapopulations. *The American Naturalist* 186, 460–469.
- Imrich, W. & Klavzar, S. (2000). *Product graphs: structure and recognition*. Wiley p.27.
- Lähdesmäki, H. & Shmulevich, I. (2008). Learning the structure of dynamic bayesian networks from time series and steady state measurements. *Machine Learning* 71, 185–217.
- 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, 319–345.
- Layman, C.A., Giery, S.T., Buhler, S., Rossi, R., Penland, T., Henson, M.N., Bogdanoff, A.K., Cove, M.V., Irizarry, A.D., Schalk, C.M. *et al.* (2015). A primer on the history of food web ecology: fundamental contributions of fourteen researchers. *Food Webs* 4, 14–24.
- Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist* 15, 237–240.
- Massol, F., Dubart, M., Calcagno, V., Cazelles, K., Jacquet, C., Kéfi, S. & Gravel, D. (2017). Island biogeography of food webs. *Advances in Ecological Research* 56.
- Mouquet, N., Matthiessen, B., Miller, T. & Gonzalez, A. (2011). Extinction debt in source-sink metacommunities. *PLoS One* 6, e17567.
- Nee, S., Hassell, M.P. & May, R.M. (1997). Two-species metapopulation models. In: *Metapopulation biology*. Elsevier pp. 123–147.
- Neutel, A.M., Heesterbeek, J.A. & de Ruiter, P.C. (2002). Stability in real food webs: weak links in long loops. *Science* 296, 1120–1123.
- 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, 737–747.
- Ovaskainen, O. & Hanski, I. (2001). Spatially structured metapopulation models: global and local assessment of metapopulation capacity. *Theoretical population biology* 60, 281–302.
- Pearl, J. *et al.* (2009). Causal inference in statistics: An overview. *Statistics surveys* 3, 96–146.

- Sander, E.L., Wootton, J.T. & Allesina, S. (2017). Ecological network inference from long-term presence-absence data. *Scientific reports* 7, 7154.
- 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, 693–707.
- Tilman, D., Lehman, C.L. & Yin, C. (1997). Habitat destruction, dispersal, and deterministic extinction in competitive communities. *The American Naturalist* 149, 407–435.
- Valdovinos, F.S. (2019). Mutualistic networks: moving closer to a predictive theory. *Ecology letters* 22, 1517–1534.

## **2.3 Annexes de l'article (en anglais)**

# Supporting Information: A spatial theory of mutualistic metacommunity

Marc Ohlmann, François Munoz, Fabien Laroche, Wilfried Thuiller

This document aims to give mathematical insights of the framework developed in Spatial theory of metacommunity. Moreover, it provides implementation of metacommunity computation and details on the simulation.

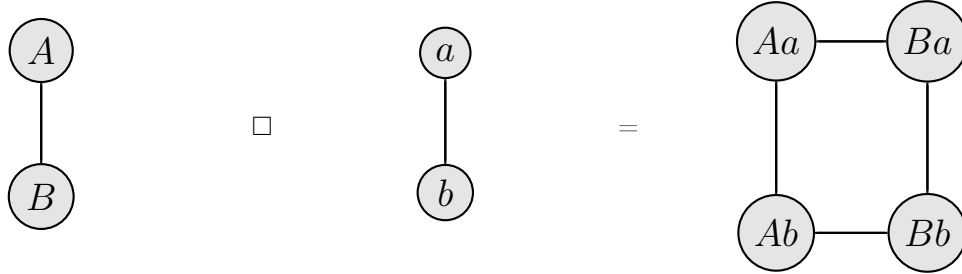
## 1 Graph product, Dynamic Bayesian Network and Markov chains

### 1.1 Graph products

Several products can be defined on graph. Here, we focus on tensor product and cartesian product. Several additional elements can be found in [Imrich & Klavzar \(2000\)](#).

Let  $G_s = (V_s, E_s)$  and  $G_b = (V_b, E_b)$  two graphs, we note  $\mathbf{A}_s$  and  $\mathbf{A}_b$  their adjacency matrices.

The cartesian product of  $G_s$  and  $G_b$ ,  $G_{s,d} = G_s \square G_b$  is the graph whose the set of nodes is  $V_d \times V_b$ . A node of this graph is identified by a pair of nodes of  $G_s$  and  $G_b$ . Moreover, there is an edge between  $(u_s, u_b)$  and  $(v_s, v_b)$  if  $(u_s = v_s \text{ and } (u_b, v_b) \in E_b)$  or  $(u_b = v_b \text{ and } (u_s, v_s) \in E_s)$ . An example of such operation is :



The tensor product of  $G_s$  and  $G_b$  is the graph whose adjacency matrix is  $\mathbf{A}_s \otimes \mathbf{A}_b$

### 1.2 Dynamic Bayesian Network

We introduce here key elements on Bayesian network and refer to [Koller & Friedman \(2009\)](#) for a general introduction on the subject.

Given a set of  $n$  random variables  $(X_1, \dots, X_n)$  (we note  $I = \{1, \dots, n\}$ )

Two random variables  $X_i$  and  $X_j$  are independent conditionally to  $\mathbf{X}_{I \setminus \{i,j\}}$  if :

$$Pr(X_i, X_j | \mathbf{X}_{I \setminus \{i,j\}}) = Pr(X_i | \mathbf{X}_{I \setminus \{i,j\}}) Pr(X_j | \mathbf{X}_{I \setminus \{i,j\}}) \quad (1)$$

Bayesian networks aim to represent conditional Independence statement using a Directed Acyclic Graph  $G$ . We note  $V = \{1, \dots, n\}$  the set of nodes of  $G$ . For a given node  $u$ , we note  $Pa_u(G)$  the set of nodes that are parents of  $u$ .

$$Pa_u(G) = \{v \in V, (v, u) \in E\} \quad (2)$$

The joint probability factorises over  $G$  as :

$$P(X_1, \dots, X_n) = \prod_i Pr(X_i | \mathbf{X}_{Pa_i(G)}) \quad (3)$$

The factorisation gives the independence conditional statement according to the structure of the DAG. A particular case of Bayesian network consists in Dynamic Bayesian Networks. The bayesian networks describe the dependencies between  $\{X_1^t, \dots, X_n^t\}$  and  $\{X_1^{t+1}, \dots, X_n^{t+1}\}$  using a directed bipartite network  $G_{bip}$  (we note  $\mathbf{A}_{bip}$  its adjacency matrix). Importantly, as the structure of  $G_{bip}$  does not depend of  $t$ , it can be built using an aggregated network  $G$  (we note  $\mathbf{A}$  its adjacency matrix) and a graph  $P_2$  (we note  $\mathbf{A}_2$  its adjacency matrix) whose set of nodes is  $\{t, t+1\}$  and set of edges is  $\{(t, t+1)\}$ . We have

$$\mathbf{A}_{bip} = \mathbf{A}_2 \otimes (\mathbf{A} + \mathbf{I}_n) \quad (4)$$

The joint probability factorises over  $G$  according to:

$$P(X_1^{t+1}, \dots, X_n^{t+1} | X_1^t, \dots, X_n^t) = \prod_i Pr(X_i | \mathbf{X}_{\mathbf{Pa}_i(\mathbf{G})}^t) \quad (5)$$

### 1.3 Markov chains from dynamic Bayesian networks

Following Lähdesmäki & Shmulevich (2008), we propose to build Markov Chains using Dynamic Bayesian Networks. In our case, the structure of the dynamic Bayesian network is given by the cartesian product of the spatial network and the interaction network. In the spatially realistic models of mutualistic metacommunities, the transition matrix of the chain is of dimension  $2^{mn} * 2^{mn}$ , encoding the probability of transition between a state  $\mathbf{s}_k = (x_{11}, \dots, x_{mn})$  of the metacommunity and a state  $\mathbf{s}_l = (x_{11}, \dots, x_{mn})$ , where  $x_{ij} \in \{0, 1\}$  describes the presence of a population of species  $i$  in site  $j$ . The probability of transition between  $\mathbf{s}_k$  and  $\mathbf{s}_l$  is :

$$P_{k,l} = \prod_{i,j} Pr(X_{ij}^{t+1} = x_{ij} | X_{11}^t = x_{11}, \dots, X_{mn}^t = x_{mn}) \quad (6)$$

Moreover, we have:

$$Pr(X_{i,j}^{t+1} = 1 | X_{i,j}^t = 0, \sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t) = \epsilon + (1 - \epsilon) * (1 - (1 - c)^{\sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t}) \quad (7)$$

with  $\epsilon \in ]0; 1[$  and  $c \in ]0; 1[$ . We have:

$$Pr(\epsilon < \epsilon + (1 - \epsilon) * (1 - (1 - c)^{\sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t}) < 1) = 1 \quad (8)$$

Moreover:

$$Pr(X_{i,j}^{t+1} = 0 | X_{i,j}^t = 1, \sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t) = e - (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t}) / M \quad (9)$$

where  $e \in ]0; 1[$  and:

$$M = \max(\deg(G_e)) + 1 \quad (10)$$

We have then:

$$Pr((1 - e)^{\max(\deg(G_e))} < (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t} < 1) = 1 \quad (11)$$

$$Pr(0 < (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t}) / M < (1 - (1 - e)^{M-1}) / M) = 1 \quad (12)$$

However:  $(1 - (1 - e)^{M-1}) / M < e / M$

and since :  $M \geq 1$

It follows that the probability of extinction is in  $]0; 1[$ .

Consequently :

$$\forall i \in \{1, \dots, m\}, \forall j \in \{1, \dots, m\}, Pr(X_{i,j}^{t+1} = x_{i,j} | X_{1,1}^t = x_{1,1}, \dots, X_{m,n}^t = x_{m,n}) > 0$$

The hypothesis of *Theorem 1* of Lähdesmäki & Shmulevich (2008) are verified and the chain  $(\mathbf{X}^t)^t$  converges towards a stationary distribution.

## 2 Deterministic models and metacommunities capacity

In this section, we derive the mean-field deterministic model from the stochastic metacommunity model and define metacommunity capacities.

## 2.1 Mean field approximation

The approximation is derived from [Bianconi \(2018\)](#). The aim is to study the dynamic of occupancy of each species  $j$  in each site  $i$ , i.e.  $p_{ij}(t) = \mathbb{E}(X_{ij}^t)$ . For all  $i$  and  $j$ , we have

$$X_{i,j}^{t+1} = (1 - X_{i,j}^t)(\epsilon + (1 - \epsilon)(1 - (1 - c)^{\sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t})) + (1 - e + (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t})/M)X_{i,j}^t \quad (13)$$

It follows :

$$p_{i,j}(t+1) = \mathbb{E}((1 - X_{i,j}^t)(\epsilon + (1 - \epsilon)(1 - (1 - c)^{\sum_{(k,l) \in N_{G_c}(i,j)} X_{k,l}^t}))) + \mathbb{E}((1 - e + (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} X_{k,l}^t})/M)X_{i,j}^t) \quad (14)$$

This approach leads to a hierarchy of equations that cannot be solved (i.e. we need to consider  $\mathbb{E}(X_{1,1}^t, \dots, X_{m,n}^t)$  to find a solution to the system). A drastic approximation consists in the mean field approximation, for any sequence of indices  $n(1), n'(1); \dots, n(r), n'(r)$ , we assume :

$$\mathbb{E}(X_{n(1),n'(1)}^t, \dots, X_{n(r),n'(r)}^t) \simeq \mathbb{E}(X_{n(1),n'(1)}^t) \dots \mathbb{E}(X_{n(r),n'(r)}^t) \quad (15)$$

We now get:

$$p_{i,j}(t+1) = (1 - p_{i,j}(t))(\epsilon + (1 - \epsilon)(1 - (1 - c)^{\sum_{(k,l) \in N_{G_c}(i,j)} p_{kl}(t)})) + (1 - e + (1 - (1 - e)^{\sum_{(k,l) \in N_{G_e}(i,j)} p_{kl}(t)})/M)p_{i,j}(t) \quad (16)$$

We assume that  $c \ll 1, e \ll 1$  and  $\epsilon \ll c$  and do a Taylor expansion at order 1:

$$p_{i,j}(t+1) = (1 - p_{i,j}(t))(\epsilon + (1 - \epsilon)(c \sum_{(k,l) \in N_{G_c}(i,j)} p_{k,l}(t) + (1 - e + e \sum_{(k,l) \in N_{G_e}(i,j)} p_{k,l}(t))/M)p_{i,j}(t) \quad (17)$$

$$p_{i,j}(t+1) = (1 - p_{i,j}(t))(c \sum_{(k,l) \in N_{G_c}(i,j)} p_{k,l}(t) + (1 - e + e \sum_{(k,l) \in N_{G_e}(i,j)} p_{kl}(t))/M)p_{i,j}(t) \quad (18)$$

We introduce a single index  $v$  for the nodes of the product networks and get:

$$p_v(t+1) = (1 - p_v(t))(c \sum_{u \in N_{G_c}(v)} p_u(t) + (1 - e + e \sum_{u \in N_{G_e}(v)} p_u(t))/M)p_v(t) \quad (19)$$

$$p_v(t+1) - p_v(t) = (1 - p_v(t))(c \sum_{u \in N_{G_c}(v)} p_u(t) - e(1 - \sum_{u \in N_{G_e}(v)} p_u(t))/M)p_v(t) \quad (20)$$

$$p_v(t+1) - p_v(t) = C_v(\mathbf{p}(t))(1 - p_v(t)) - E_v(\mathbf{p}(t))(p_v(t)) \quad (21)$$

where  $C_v(\mathbf{p}(t)) = c \sum_u [A_c]_{v,u} p_u(t)$  and  $E_v(\mathbf{p}(t)) = e(1 - \sum_u [A_e]_{v,u} p_u(t))/M$ .

## 2.2 Metapopulation capacities

In order to define metacommunity capacities in the next section, we first recall definition and results from [Ovaskainen & Hanski \(2001\)](#) on metapopulation capacities.

Following [Ovaskainen & Hanski \(2001\)](#) that defined metapopulation capacities for the model defined in Eq. 21 for a single species only, we aim to define metacommunity capacities for the model defined in Eq. 21 but whose parameters are colonisation and extinction networks made of products between spatial network and interaction network.

In the metapopulation case (i.e. with  $G_b$  made of a single node,  $m = 1$ ),  $\forall t \in \mathbb{N}^*$ ,  $\mathbf{p}(t) \in \bar{\Omega} = \{x \in \mathbb{R}^n, \forall i, 0 \leq x_i \leq 1\}$ , with the following assumptions on the colonisation functions,  $E_i(\cdot)$ , and extinction functions,  $C_i(\cdot)$ ,

- $C_i(0) = 0, \forall \mathbf{p} \in \Omega = \{x \in \mathbb{R}^n, \forall i, 0 < x_i < 1\}, C_i(\mathbf{p}) > 0, \frac{\partial C_i}{\partial p_j} \geq 0 (i \neq j), \frac{\partial C_i}{\partial p_i} = 0$
- $\forall \mathbf{p} \in \bar{\Omega}, E_i(\mathbf{p}) > 0, \frac{\partial E_i}{\partial p_j} \leq 0 (i \neq j), \frac{\partial E_i}{\partial p_i} = 0$
- $C_i \in \mathcal{C}^1(\bar{\Omega}), E_i \in \mathcal{C}^1(\bar{\Omega})$



Let :

$$g(\mathbf{p}) = \frac{C(\mathbf{p})}{E(\mathbf{p})} \quad (22)$$

The model is also assumed to be irreducible. Let  $\mathbf{D}$  be the matrix of dimension  $n \times n$  so that:

$$D_{ij} = \begin{cases} 1 & \text{if } \frac{\partial g_i}{\partial p_j}(\mathbf{p}) \geq 0, \mathbf{p} \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

The model is irreducible if  $\mathbf{D}$  is irreducible, i.e. the graph that has  $\mathbf{D}$  as adjacency matrix is strongly connected.

In the case of the spatially realistic Levins model :

- $C_i(\mathbf{p}) = c(\mathbf{A}_s \mathbf{p})_i$
- $E(\mathbf{p}) = e$

where  $\mathbf{A}_s$  is the adjacency matrix of the spatial network. Then,

$$g_i(\mathbf{p}) = \frac{c(\mathbf{A}_s \mathbf{p})_i}{e} \quad (23)$$

and the model is irreducible since  $\mathbf{A}_s$  is irreducible.

The metapopulation invasion capacity,  $\lambda_I$ , is defined as the dominant eigenvalue of the jacobian matrix of  $g$  evaluated in  $\mathbf{p} = 0$ . It measures the stability of the equilibrium  $\mathbf{p} = 0$ .

The metapopulation persistence capacity,  $\lambda_M$ , is defined as:

$$\lambda_M = \sup_{\mathbf{p} \in \Omega} h(\mathbf{p}) \quad (24)$$

where

$$h(\mathbf{p}) = \min_i h_i(\mathbf{p}) \quad (25)$$

and

$$h_i(\mathbf{p}) = g_i(\mathbf{p}) \frac{1 - p_i}{p_i} \quad (26)$$

*Theorem 2.1* of [Ovaskainen & Hanski \(2001\)](#) states that a non-trivial equilibrium that the dynamical system has a non trivial equilibrium if  $\lambda_M > 1$ .  $\lambda_M$  is then a threshold on the colonisation/extinction parameters that allows the metapopulation to persist. Importantly, the equilibrium point is interior (it belongs to  $\Omega$ ).

Moreover, for the spatially realistic Levins model :

$$\lambda_M = \lambda_I = \Lambda \quad (27)$$

where  $\Lambda$  is the dominant eigenvalue of  $\mathbf{A}_s$ .

### 2.3 Metacommunity capacities

We propose to extend the concept of metapopulation capacities to metacommunity capacities in the dynamical system defined by equation (17). In this case,  $\Omega = \{x \in \mathbb{R}^{n*m}, \forall v, 0 < x_v < 1\}$

We have:

- $C_v(\mathbf{p}(t)) = c \sum_u [A_c]_{v,u} p_u(t)$
- $E_v(\mathbf{p}(t)) = e(1 - \sum_u [A_e]_{v,u} p_u(t))/M$

By setting:

$$M = \max(\text{degree}(G_e)) + 1 \quad (28)$$

The hypothesis on colonisation and extinction functions are verified. Moreover,

$$g_v(\mathbf{p}) = \frac{c \sum_u [A_c]_{v,u} p_u}{e(1 - \sum_u [A_e]_{v,u} p_u)/M} \quad (29)$$

$$\frac{\partial g_v}{\partial p_u}(\mathbf{p}) = \frac{ce[A_c]_{v,u} \left( \frac{1 + \sum_k [A_e]_{v,k} p_k}{M} \right) + ce[A_e]_{v,u} \frac{\sum_k [A_c]_{v,k} p_k}{M}}{E_v(p)^2} \quad (30)$$

For the Levins type model:

$$\frac{\partial g_v}{\partial p_u}(\mathbf{p}) = \frac{c[A_c]_{v,u}}{e} \quad (31)$$

And since,  $G_s$  and  $G_b$  are both strongly connected and that  $G_c = G_s \square G_b$ ,  $G_c$  is strongly connected and  $\mathbf{D}$  (build from the jacobian matrix of  $g$ ) is irreducible.

For the separated effect model,  $E(G_e) \cap E(G_c) = \emptyset$ . So  $\langle A^e, A^c \rangle = 0$ . So,

- if  $[A_c]_{v,u} = 0$ , then  $\frac{\partial g_v}{\partial p_u}(\mathbf{p}) = -\frac{ce[A_e]_{v,u} \sum_k [A_c]_{v,k} p_k}{ME_v(\mathbf{p})^2}$   
since  $\forall(v, k), [A_c]_{v,k} \geq 0$  and  $[A_e]_{v,k} \geq 0$ , then:  
 $g_{v,u}(\mathbf{p}) > 0 \Leftrightarrow (v, u) \in E(G_e)$
- if  $[A_e]_{v,u} = 0$ , then:  $g_{v,u}(\mathbf{p}) > 0 \Leftrightarrow (v, u) \in E(G_c)$

So,

$$g_{v,u}(\mathbf{p}) > 0 \Leftrightarrow (v, u) \in E(G_c) \cup E(G_e)$$

and since  $E(G_c) \cup E(G_e)$  is the edge set of  $G_s \square G_b$ , that is strongly connected, it follows that  $\mathbf{D}$  is irreducible in this case.

Similarly, we show that the combined effect model and the rescue effect model are irreducible.

We then define metacommunity invasion capacity as the dominant eigenvalue of the jacobian matrix of  $g$  evaluated in  $\mathbf{p} = 0$ .

We define the metacommunity persistence capacity as

$$\lambda_M = \sup_{\mathbf{p} \in \Omega} h(\mathbf{p}) \quad (32)$$

where

$$h(\mathbf{p}) = \min_v h_v(\mathbf{p}) \quad (33)$$

and

$$h_v(\mathbf{p}) = g_v(p) \frac{1 - p_v}{p_v} \quad (34)$$

By applying *Theorem 2.1* of [Ovaskainen & Hanski \(2001\)](#), a non-trivial equilibrium that the dynamical system has a non trivial equilibrium if  $\lambda_M > 1$ .  $\lambda_M$  is then a threshold on the colonisation/extinction parameters that allows the metacommunity to persist. Importantly, non-trivial the equilibrium point is interior (it belongs to  $\Omega$ ), so each species in each site has a positive abundance at equilibrium.

Moreover, the Levins type model is actually the spatially realistic model with  $G_c = G_s \square G_b$  as spatial network. The dominant eigenvalue,  $\Lambda_c$ , of  $A_c$  is  $\Lambda_s + \Lambda_b$ . Consequently, for the Levins type model:

$$\lambda_M = \lambda_I = \Lambda_s + \Lambda_b \quad (35)$$

The metacommunity capacities for the three other models need to be computed numerically. ( $\lambda_I$  can be computed from the jacobian matrix evaluated in  $\mathbf{p} = 0$ )

## 2.4 Computing numerically $\lambda_M$

In order to compute  $\lambda_M$  for the combined effect model, the separated effect model and the rescue effect model, we used simulating annealing. The code to compute the metacommunity capacity in the different models is available at: [https://gitlab.com/marcohlmann/metacommunity\\_theory](https://gitlab.com/marcohlmann/metacommunity_theory).

We assessed the performance of the method on the Levins type model on the simulated data, since we know analytically the metacommunity capacity in this case. With 20000 time steps on the 800 networks, (15 hours of computation on 16 cores in parallel on a machine per model). The maximum is not reached (Fig. S1) but there is a strong correlation (0.988) between the estimated metacommunity capacities and the theoretical metacommunity capacities (Fig. S1), allowing so comparison of the metacommunity capacities among the different locations of the gradient.

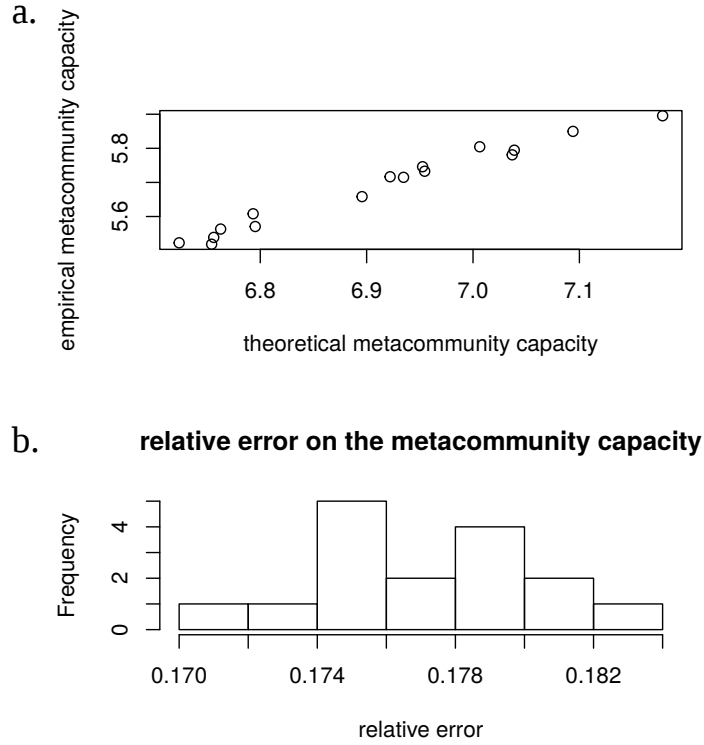


Figure S1 : a. Relation between the metacommunity persistence capacity computing using a simulating annealing algorithm and the theoretical metacommunity capacity for the Levins type model, averaged over the 50 replicates. b. Distribution of the relative error in the estimation of the metacommunity capacity

### 3 More details on the simulation

#### 3.1 The spatial networks

In order to mimic a gradient of fragmentation of landscape, we sampled spatial networks in a Stochastic block model (ref) for 4 different values of parameters (to represent 4 locations on the gradient). Each network had a 10 nodes that were partitioned in two groups,  $G_1$  and  $G_2$ , with a matrix of probability of connection,  $\mathbf{\Pi}_\alpha$ , given by :

$$\begin{pmatrix} G_1 & G_2 \\ C + \frac{\alpha}{5} & C - \frac{\alpha}{5} \\ C - \frac{\alpha}{5} & C + \frac{\alpha}{5} \end{pmatrix} \begin{matrix} G_1 \\ G_2 \end{matrix} \quad (36)$$

where  $C = 0.25$ . Moreover, we chose 4 different values for  $\alpha$  :  $\{0, \frac{1}{3}, \frac{2}{3}, 1\}$ . We sampled 50 networks per value of  $\alpha$ . The overall probability of connection in the network is :

$$Pr(i \leftrightarrow j) = \sum_{k \in \{0,1\}, l \in \{0,1\}} Pr(i \leftrightarrow j | i \in G_k, j \in G_l) Pr(i \in G_k) Pr(j \in G_l) \quad (37)$$

$$Pr(i \leftrightarrow j) = \frac{1}{2} \left( C + \frac{\alpha}{5} \right) + \frac{1}{2} \left( C - \frac{\alpha}{5} \right) \quad (38)$$

$$Pr(i \leftrightarrow j) = C \quad (39)$$

So the expected value of connectance for all spatial networks is the same.

#### 3.2 The biotic interaction networks

We first generated random undirected network with various shapes of the degree distribution using the function `sample_fitness_pl` implemented in the R package `igraph`. We only kept connected

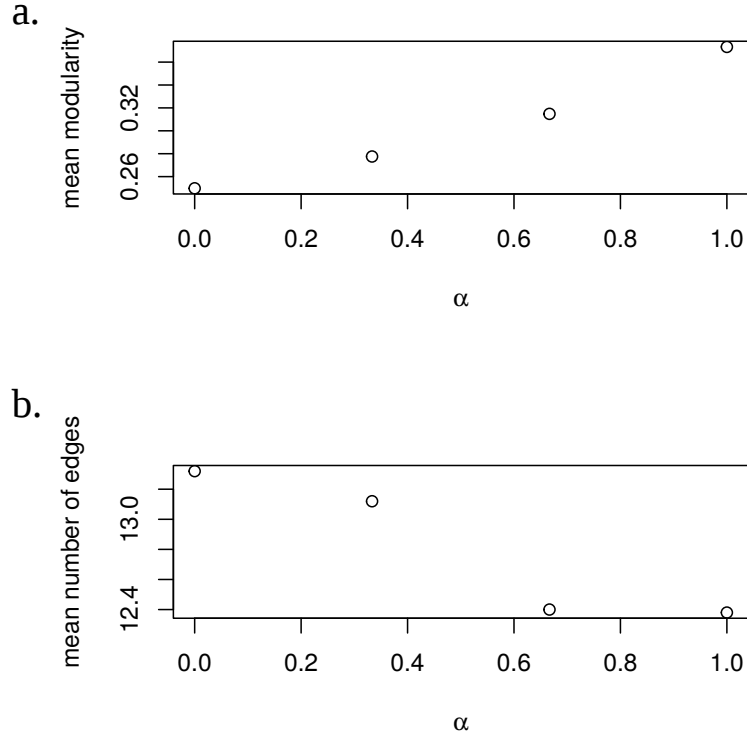


Figure S2 : a. Mean modularity averaged over the 50 replicates of each location of the spatial gradient versus the  $\alpha$  parameter. b. Mean number of edges averaged over the 50 replicates of each location of the spatial gradient versus the  $\alpha$  parameter.

networks. On the random network  $G$  sampled ( $\mathbf{A}$  is its adjacency matrix), we build a bipartite network  $G_{bip}$  with adjacency matrix  $\mathbf{A}_{bip}$  as:

$$\mathbf{A}_{bip} = \mathbf{A}_2 \otimes (\mathbf{A} + \mathbf{I}_n) \quad (40)$$

where  $\mathbf{A}_2$  is the adjacency matrix of an undirected graph made of two nodes and a single edge between these two nodes. By doing so, all the sampled undirected bipartite networks are strongly connected.

### 3.3 Results

We simulated the dynamic (as presented in the main text for the Levins type model) for the combined effect model (Fig. S3, Fig. S6), for the separated effect model (Fig. S4, Fig. S7) and for the rescue effect model (Fig. S5, Fig. S8).

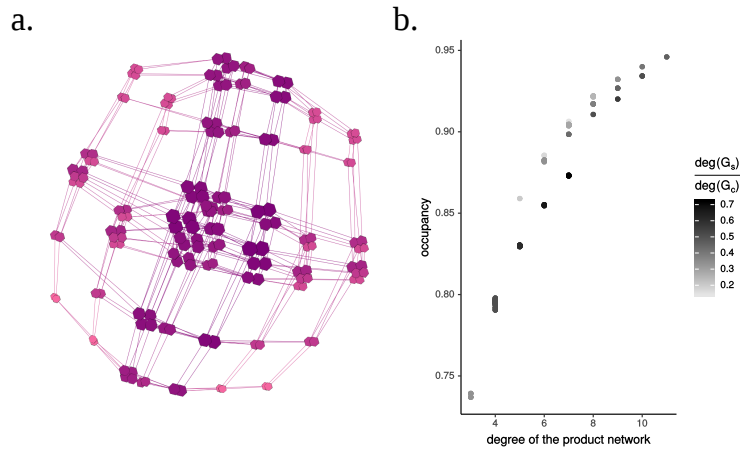


Figure S3: Simulating the dynamics for a given spatial and biotic interaction network with the combined effect model. a. colonisation network whose size of the nodes is proportional to their degree and color indicates the occupancy at equilibrium (pink: low occupancy, purple: high occupancy). b. Relationship between the occupancy at equilibrium and the degree of the node of the product network. Each point of the relationship (corresponding to a node of the product graph) is colored according to the ratio of the degree of the site in the spatial network over the degree of the focal node in the colonisation network.

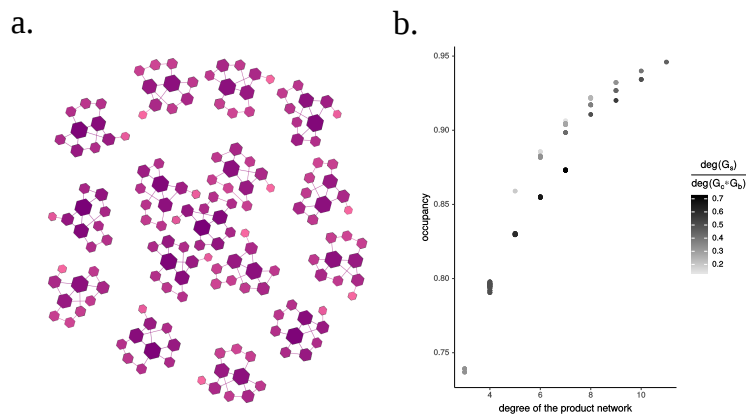


Figure S4: Simulating the dynamics for a given spatial and biotic interaction network with the separated effect model. a. colonisation network whose size of the nodes is proportional to their degree and color indicates the occupancy at equilibrium (pink: low occupancy, purple: high occupancy). b. Relationship between the occupancy at equilibrium and the degree of the node of the product network. Each point of the relationship (corresponding to a node of the product graph) is colored according to the ratio of the degree of the site in the spatial network over the degree of the focal node in the colonisation network.

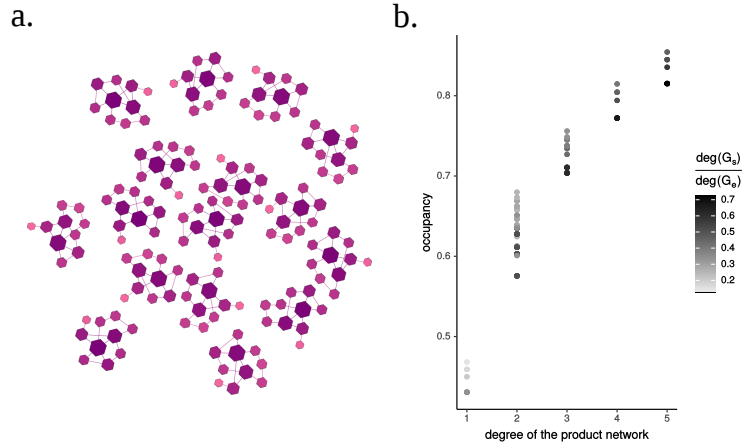


Figure S5: Simulating the dynamics for a given spatial and biotic interaction network with the rescue effect model. a. colonisation network whose size of the nodes is proportional to their degree and color indicates the occupancy at equilibrium (pink: low occupancy, purple: high occupancy). b. Relationship between the occupancy at equilibrium and the degree of the node of the product network. Each point of the relationship (corresponding to a node of the product graph) is colored according to the ratio of the degree of the site in the spatial network over the degree of the focal node in the colonisation network.

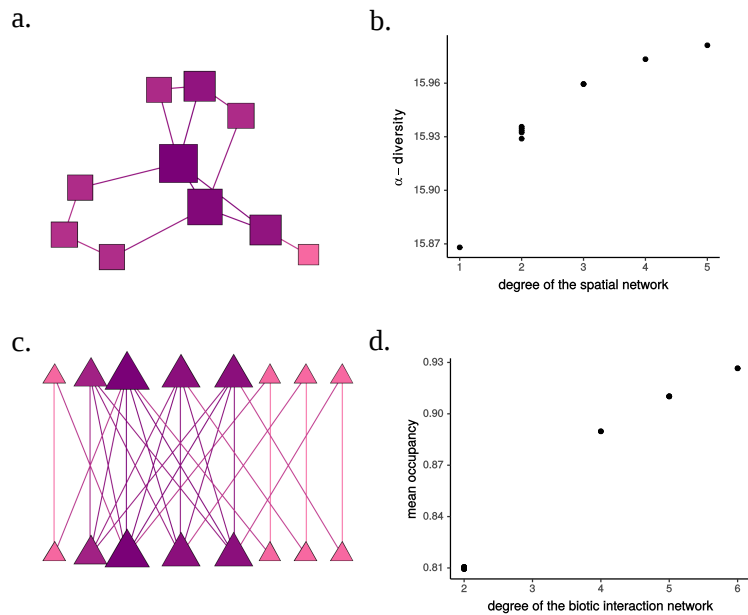


Figure S6: Aggregated statistics from occupancy at equilibrium for the combined effect model in the spatial network and the biotic interaction network. a. Spatial network whose size of the nodes is proportional to their degree and colour indicates the  $\alpha$ -diversity at equilibrium (pink: low  $\alpha$ -diversity, purple: high  $\alpha$ -diversity). b. Relationship between the  $\alpha$ -diversity at equilibrium and the degree of the sites in the spatial network. c. Biotic interaction network whose size of the nodes is proportional to their degree and colour indicates the mean occupancy across the sites at equilibrium (grey: low  $\alpha$ -diversity, black: high  $\alpha$ -diversity). d. Relationship between the mean occupancy at equilibrium and the degree of the species in the biotic interaction network.

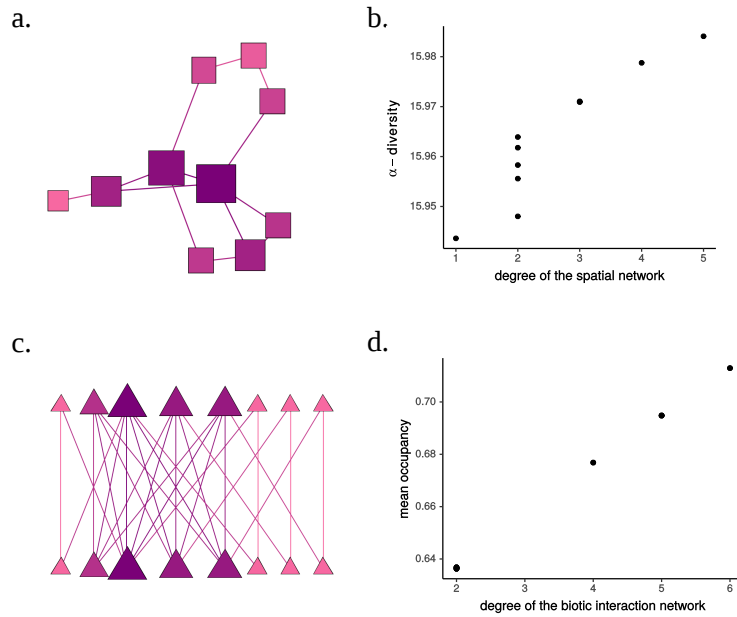


Figure S7: Aggregated statistics from occupancy at equilibrium for the separated effect model in the spatial network and the biotic interaction network. a. Spatial network whose size of the nodes is proportional to their degree and colour indicates the  $\alpha$ -diversity at equilibrium (pink: low  $\alpha$ -diversity, purple: high  $\alpha$ -diversity). b. Relationship between the  $\alpha$ -diversity at equilibrium and the degree of the sites in the spatial network. c. Biotic interaction network whose size of the nodes is proportional to their degree and colour indicates the mean occupancy across the sites at equilibrium (grey: low  $\alpha$ -diversity, black: high  $\alpha$ -diversity). d. Relationship between the mean occupancy at equilibrium and the degree of the species in the biotic interaction network.

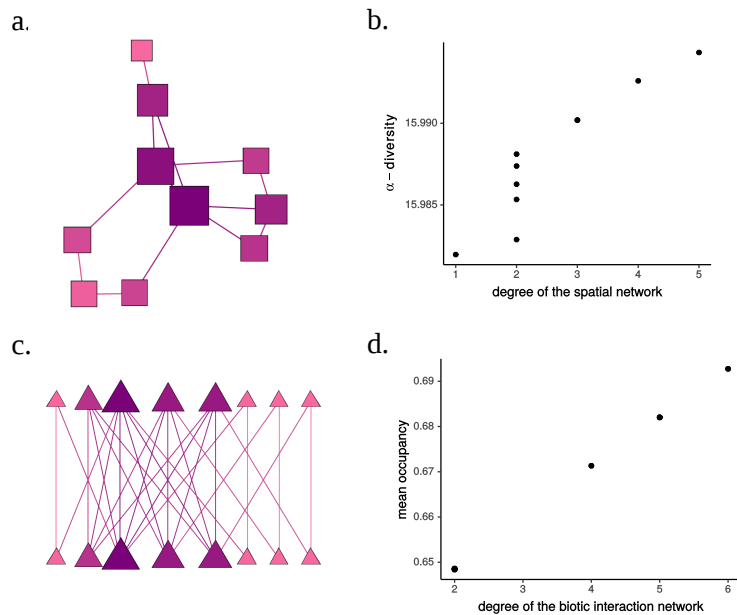


Figure S8: Aggregated statistics from occupancy at equilibrium for the rescue effect model in the spatial network and the biotic interaction network. a. Spatial network whose size of the nodes is proportional to their degree and colour indicates the  $\alpha$ -diversity at equilibrium (pink: low  $\alpha$ -diversity, purple: high  $\alpha$ -diversity). b. Relationship between the  $\alpha$ -diversity at equilibrium and the degree of the sites in the spatial network. c. Biotic interaction network whose size of the nodes is proportional to their degree and colour indicates the mean occupancy across the sites at equilibrium (grey: low  $\alpha$ -diversity, black: high  $\alpha$ -diversity). d. Relationship between the mean occupancy at equilibrium and the degree of the species in the biotic interaction network.

## References

- Bianconi, G. (2018). *Multilayer Networks: Structure and Function*. Oxford University Press; p.60-62.
- Imrich, W. & Klavzar, S. (2000). *Product graphs: structure and recognition*. Wiley p.27.
- Koller, D. & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- Lähdesmäki, H. & Shmulevich, I. (2008). Learning the structure of dynamic bayesian networks from time series and steady state measurements. *Machine Learning* 71, 185–217.
- Ovaskainen, O. & Hanski, I. (2001). Spatially structured metapopulation models: global and local assessment of metapopulation capacity. *Theoretical population biology* 60, 281–302.



## 2.4 Équivalence des espèces dans le réseau et distribution stationnaire

Dans le cadre du stage de Master 2 (Mathématiques en action) d’Alexis Teskrat que j’ai co-encadré, nous nous sommes intéressés à la façon dont les équivalences de nœuds dans les réseaux spatiaux et les réseaux d’interaction impactait la dynamique du modèle stochastique (*cf.* section 5.5).

Nous nous sommes penchés sur le modèle île continent avec interactions mutualistes et avons prouvé le résultat suivant :

**Proposition 2.4.1** *Si le réseau d’interaction est un  $K_{\underbrace{n_1, \dots, n_1}_{r_1 \text{ fois}}, \dots, \underbrace{n_p, \dots, n_p}_{r_p \text{ fois}}}$  (multiparti complet), alors le nombre d’états ayant des probabilités différentes dans la distribution stationnaire est inférieur à  $\prod_{i=1}^p \binom{n_i+r_i}{r_i}$*

De plus, nous pouvons calculer dans ce cas la distribution stationnaire en agrégeant le noyau de transition. Ainsi, en utilisant la structure du graphe initial et les équivalences des nœuds, on peut réduire la dimension du problème et calculer la distribution stationnaire. Outre un intérêt computationnel, cela montre, pour le modèle avec île-continent avec interaction, que les équivalences entre espèces (niche d’Elton) jusque là utilisée de façon statique, se retrouvent dans la dynamique. Certains assemblages d’espèces, faisant intervenir des espèces ayant même niche d’Elton, ont même probabilité dans l’état stationnaire.

On remarquera que si l’on prend un graphe complet  $K_{2n}$  comme réseau d’interaction ou un graphe biparti complet  $K_{n,n}$ , alors, dans les deux réseaux, toutes les espèces sont équivalentes (les deux graphes sont sommets-transitifs). Ainsi, on pourrait dire que dans les deux réseaux, il n’y a qu’une seule niche d’Elton. Par contre, le nombre d’assemblages d’espèces ayant des probabilités distinctes dans l’état stationnaire est différent dans les deux cas :  $2n + 1$  pour le  $K_{2n}$  et  $\binom{n+2}{2} = \frac{(n+1)(n+2)}{2}$  pour le  $K_{n,n}$ .

Pour plus de détails et des simulations dans le modèle stochastique, nous renvoyons au rapport de stage d’Alexis Teskrat en annexe.

# L’empreinte des interactions biotiques sur la $\beta$ -diversité

---

When we try to pick out anything by itself, we find that it is bound fast by a thousand invisible cords that cannot be broken, to everything in the universe.

---

John Muir

## 3.1 Synopsis

Reconstruire des réseaux d’interactions entre espèces à partir de communautés d’espèces échantillonnées en plusieurs points de l’espace est un problème complexe en écologie statistique dont la résolution permettrait d’obtenir des informations des interactions difficiles à échantillonner sur le terrain (par exemple sur les micro-organismes). Poursuivre un tel objectif sur des données de metabarcoding de sol permettrait d’avoir une meilleure image des organismes vivant dans les sol et des interactions qu’ils entretiennent entre eux. Toutefois, des problèmes de bioinformatique (relatifs à la technologie d’ADN environnemental), de statistique (relatifs à la méthode pour faire de l’inférence de réseau) et de méthode scientifique (interprétation des arêtes du modèle graphique comme interactions biologiques, validation du réseau inféré) se posent. Dans ce chapitre, nous nous intéresserons à la reconstruction de réseaux d’associations, en utilisant un modèle graphique, entre  $\beta$ -diversités de groupes trophiques du sol. Nous effectuerons d’abord une simulation pour tester la méthode statistique puis l’appliquerons sur des données échantillonnées le long d’un gradient d’altitude dans les Alpes françaises. Nous montrerons que les associations détectées sont cohérentes vis-à-vis des interactions entre groupes trophiques connues dans la littérature. Nous montrerons également que, conditionnellement à la connaissance de la diversité des autres groupes d’organismes, le lien entre l’environnement abiotique et la diversité d’un groupe cible est plus faible que le lien entre la diversité de ce groupe cible et la diversité des autres groupes. Ainsi, cela suggérerait que, même au sein d’un système à fort contraste environnemental comme un gradient altitudinal, les interactions biotiques jouent un rôle plus important dans la structuration de la communauté que l’environnement abiotique.


Ce chapitre a fait l’objet d’une publication dans Ecology Letters :

Ohlmann, M., Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L.,  
... & Thuiller, W. (2018). Mapping the imprint of biotic interactions on  
-diversity. Ecology letters, 21(11), 1660-1669.

## **3.2 Corps principal de l’article (en anglais)**

LETTER

## Mapping the imprint of biotic interactions on $\beta$ -diversity

Marc Ohlmann,<sup>1\*</sup>   
 Florent Mazel,<sup>2</sup>   
 Loïc Chalmandrier,<sup>3,4</sup>  
 Stéphane Bec,<sup>1</sup> Eric Coissac,<sup>1</sup>  
 Ludovic Gielly,<sup>1</sup> Johan Pansu,<sup>5</sup>  
 Vincent Schilling,<sup>6</sup> Pierre Taberlet,<sup>1</sup>  
 Lucie Zinger,<sup>7</sup>  Jérôme Chave<sup>6</sup>  
 and Wilfried Thuiller<sup>1</sup> 

### Abstract

Investigating how trophic interactions influence the  $\beta$ -diversity of meta-communities is of paramount importance to understanding the processes shaping biodiversity distribution. Here, we apply a statistical method for inferring the strength of spatial dependencies between pairs of species groups. Using simulated community data generated from a multi-trophic model, we showed that this method can approximate biotic interactions in multi-trophic communities based on  $\beta$ -diversity patterns across groups. When applied to soil multi-trophic communities along an elevational gradient in the French Alps, we found that fungi make a major contribution to the structuring of  $\beta$ -diversity across trophic groups. We also demonstrated that there were strong spatial dependencies between groups known to interact specifically (e.g. plant-symbiotic fungi, bacteria-nematodes) and that the influence of environment was less important than previously reported in the literature. Our method paves the way for a better understanding and mapping of multi-trophic communities through space and time.

### Keywords

$\beta$ -diversity, graphical lasso, graphical model, interaction network, meta-communities, partial correlation networks.

Ecology Letters (2018) 21: 1660–1669

### INTRODUCTION

Understanding the processes that determine the spatial structure of biodiversity is one of the overarching goals of ecology (Ricklefs 1987). In particular, the study of  $\beta$ -diversity, the change in species identities (or species groups) across sampled locations, sheds light on different ecological, evolutionary and biogeographic processes (e.g. Graham & Fine 2008; Anderson *et al.* 2011). For a given regional species pool, the processes responsible for  $\beta$ -diversity are usually assumed to be environmental filtering, dispersal limitations and biotic interactions (HilleRisLambers *et al.* 2012; Meynard *et al.* 2013).

Previous studies have sought to analyse  $\beta$ -diversity by teasing apart the effects of environmental filtering from biotic interactions along environmental gradients on a single group of species (e.g. Peay *et al.* 2016 for fungi; reviewed in Hanson *et al.* 2012 for bacteria; Mazel *et al.* 2017 for mammals; Chalmandrier *et al.* 2015 for plants). However, biotic interactions across groups are also expected to drive the structure and distribution of biodiversity. For example, plant-pollinator, trophic (e.g. prey–predator or plant-decomposers) and host-symbiont (including pathogens, mutualistic and commensal organisms) interactions strongly impact diversity distribution and ecosystem functioning (Brose & Hillebrand 2016). These interactions among groups of species, although not necessarily species-specific (Walker *et al.* 2011; Peay *et al.* 2015), harbour some degree of specificity due to trait constraints and shared

habitat preferences (e.g. Allesina *et al.* 2008; Gonzalez-Varo & Traveset 2016). This spatial interdependence among groups implies that the  $\beta$ -diversity of a single group is likely to be contingent to that of the other groups. Hence, studying how the  $\beta$ -diversity of multiple groups covaries along environmental gradients should help better understand their spatial distribution and uncover their interactions. In particular, soil systems provide several examples of coupled biological systems (Wardle 2006; Bardgett & Wardle 2010); for example plant-mycorrhiza associations (Smith *et al.* 2008), direct plant control on fungal communities (Broeckling *et al.* 2008), predator-prey relationships (Hedlund & Öhrn 2000) or plant-decomposer relationships (Hättenschwiler *et al.* 2005). Empirical evidence suggests that these biological couplings could indeed produce spatial dependencies between the  $\beta$ -diversity of the different groups (e.g. plants and fungi or bacterial turnover, Zinger *et al.* 2011; Prober *et al.* 2015; Geremia *et al.* 2016; aquatic macro/micro consumers and producers Matias *et al.* 2016). However, such studies remain seldom due to difficulties in compiling comprehensive multi-trophic inventories (but see De Bie *et al.* 2012; Matias *et al.* 2016; Kéfi *et al.* 2016). While, amplicon-based DNA analysis of environmental samples (i.e. environmental DNA, Taberlet *et al.* 2012; Kress *et al.* 2015) holds the promise to unlock this limit by enabling consistent all-biodiversity environmental surveys in soils (Tedersoo *et al.* 2016; Zinger *et al.* 2017) or aquatic environments (Lima-Mendez *et al.* 2015), we still critically miss appropriate

<sup>1</sup>University Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Écologie Alpine, F-38000, Grenoble, France

<sup>2</sup>Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

<sup>3</sup>Landscape Ecology, Institute of Terrestrial Ecosystems, ETH Zürich, Zürich, Switzerland

<sup>4</sup>Swiss Federal Research Institute WSL, 8903, Birmensdorf, Switzerland

<sup>5</sup>Princeton University, 110 Morrison Hall, Princeton, NJ 08544, USA

<sup>6</sup>Université Toulouse 3 Paul Sabatier, CNRS, IRD, UMR 5174 Evolution et Diversité Biologique (EDB), F-31062, Toulouse, France

<sup>7</sup>Ecole Normale Supérieure, PSL Research University, CNRS, Inserm, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), F-75005, Paris, France

\*Correspondence: E-mail: marc.ohlmann@univ-grenoble-alpes.fr

statistical tools to explore these multi-trophic  $\beta$ -diversity patterns while taking into account environmental variation.

Indeed, analysing how the  $\beta$ -diversity of a given trophic group depends not only on the  $\beta$ -diversity of the other groups but also on the spatial variation in environmental conditions is challenging as soon as the number of groups becomes large. Indeed, many groups will inevitably show multiple correlations leading to multi-collinearity in the analysis. Path analyses (Wright 1921; Shipley 2000; Schuldt *et al.* 2017) are one of the available tools to deal with this issue. However, they require assumptions on the overall structure of the network and they depict hierarchical dependences. Alternatively, probabilistic graphical models have been designed to account for conditional dependencies among multiple variables (Koller & Friedman 2009). For instance, partial correlation networks, a type of Markov networks, do not require any predefined structure, has also been used to model species interactions (Harris 2016) and can be inferred using graphical lasso (Friedman *et al.* 2007; Mazumder & Hastie 2011). Partial correlation networks could then be applied to multi-trophic systems in order to provide a map of dependencies between  $\beta$ -diversities of trophic groups without any *a priori* structure.

Here, we propose to document multi-group  $\beta$ -diversity patterns in soil communities using partial correlation networks. Specifically, we hypothesise that interrelated  $\beta$ -diversity patterns among groups are partly explained by biotic interactions along abiotic gradients, and that these spatial dependencies can be detected with the graphical lasso method to infer a partial correlation network. First, using a simulation model of multi-trophic communities with several trophic networks, we show that biotic interactions may indeed produce interrelated  $\beta$ -diversity patterns (i.e. non-zero partial correlation coefficients) that can be uncovered using a partial correlation network. Second, we apply the method to an empirical dataset including bacteria, micro-eukaryotes, meso/macrofauna, plants and abiotic factors along an elevation gradient in the French Alps. We jointly explore the co-variation between the  $\beta$ -diversity of multiple trophic groups to unravel known and unknown potential biotic interactions, while controlling for the relative role of the abiotic environment and the  $\beta$ -diversity of the other groups.

## MATERIAL AND METHODS

### The method: applying the Graphical lasso to multi-trophic $\beta$ -diversity patterns

*From species identity to (trophic) groups, and measuring  $\beta$ -diversity and environmental distances*

Species grouping can be defined by known trophic position (e.g. symbiotic fungi), or by taxonomy (e.g. bacteria) or function. This step is essential and has to be implemented in the light of the prior knowledge of the study system (see our case study for an example). Then, one possible measure of  $\beta$ -diversity between two local communities, A and B, is the Jaccard dissimilarity index, defined as one minus the ratio of the number of species present in both A and B over the number of species present in either A or B. It equals 0 when A and B share the same species, and 1 when they do not share

any. We used the R package ‘vegan’ (Oksanen *et al.* 2015) to compute the Jaccard dissimilarity matrix for multiple species groups (see below). Since the Jaccard index is sensitive to sample size, and since metabarcoding data often produce samples of heterogeneous sizes (Figs S1, S2), we partitioned the Jaccard index into the true turnover component and the nested component (Baselga 2010; the true turnover component reflects the turnover independently of richness variation while the nested component is deduced by subtracting true turnover component to total turnover), using the R package ‘betapart’ (Baselga & Orme 2012). Environmental distances between pairs of local communities were computed using Euclidean distances. For the sake of comparison, we also ran the analysis using the Sorensen dissimilarity index and its true turnover component.

### The graphical lasso method

The goal of our approach is to use a network to parsimoniously represent the partial correlations between the  $\beta$ -diversity matrix of each group as a function of the others and the environmental distances. Consequently, a suitable description of the system consists of using a class of models that (1) represent the conditional dependencies between random variables (here the  $\beta$ -diversity matrices of multiple species groups) using partial correlations, while (2) allowing for a parsimonious representation of the dependencies using a network. While the Lasso approach was created to produce this type of parsimonious set of variables (Tibshirani 1994), its multivariate form, the Graphical lasso (Glasso), allows representing the partial correlations among multiple variables in a network (here the  $\beta$ -diversity matrix of multiple groups and the environmental distances, Friedman *et al.* 2007; Mazumder & Hastie 2011).

In short, the Glasso uses the empirical variance-covariance matrix  $S$  to estimate a partial correlation matrix that quantifies the degree of association between pairs of variables conditional to the other variables. Here a variable is a  $n \times n$   $\beta$ -diversity matrix for a given group ( $n$  being the number of plots), or a  $n \times n$  environmental distance matrix. To estimate the partial correlation matrix in Glasso, the  $S$  matrix is inverted, and its inverse is called the precision matrix  $P$ . Moreover, a penalty term in the likelihood (modulated by a coefficient  $\lambda$ ) ensures the sparsity of the matrix  $P$  (i.e.  $P$  have many zeros, see Friedman *et al.* 2007 for mathematical details). The partial correlation matrix is then computed from  $P$  as follows (eqn 1):

$$\text{cor}(y_i, y_j | y_{\Gamma(i,j)}) = -\frac{P_{i,j}}{P_{i,i}P_{j,j}} \quad (1)$$

where  $\text{cor}(y_i, y_j | y_{\Gamma(i,j)})$  represents the partial correlation between the components  $i$  and  $j$  of a random variable  $Y$  given all the other components, and  $P_{i,j}$ ,  $P_{i,i}$ ,  $P_{j,j}$  are the elements of  $P$ . The elements of  $P$  consist in partial correlations between the  $\beta$ -diversity matrices of the different trophic groups and the environmental distances. As the precision matrices have been inverted with the constraint to ensure sparsity, it follows that the partial correlation matrix is also sparse. This sparse representation of the relationships between the  $\beta$ -diversity of multi-trophic groups and the environmental distances means it can

be represented using a network. In the Glasso, the number of coefficients equal to 0 in the partial correlation matrix depends on the coefficient  $\lambda$ . Here, we used the Extended Bayesian Information Criterion (Foygel & Drton 2010) to select an optimal  $\lambda$ . We used the R package 'qgraph' to estimate the partial correlation matrix with graphical lasso (Epskamp & Fried 2016). We expected partial correlations to be more informative than marginal correlations (Pearson correlations), because they avoid spurious correlations due to confounding effects.

#### *Representing conditional dependencies between $\beta$ -diversity using a network*

We assessed how the  $\beta$ -diversity of each group is influenced by environmental change and the  $\beta$ -diversity of the other groups by analysing the degree and the weighted degree of the correspondent node in the network (using Gephi, Bastian *et al.* 2009). The degree of a node in the network is its number of direct neighbours. If the  $\beta$ -diversity of two groups is conditionally independent (i.e. has a zero partial correlation coefficient), they cannot causally influence each other (Murphy 2012). Consequently, the more connected a group is, the more central it is to structuring the  $\beta$ -diversity of all groups. The weighted degree represents the total sum of partial correlations in  $\beta$ -diversity between a given group and the groups that are directly connected to this group. The higher the sum, the greater the interdependencies with other groups.

#### **Approximating the structure of a simulated trophic network**

We then tested whether the Glasso approach was able to recover known interactions among species groups from local community  $\beta$ -diversity patterns. We first built a set of simulated data by constructing a regional trophic web (step 1) from which local multi-trophic communities were sampled using a stochastic model (step 2). Then, we measured the partial correlation between the  $\beta$ -diversity of each trophic level (step 3) and tested whether these patterns matched the simulated regional trophic web.

Step 1 – The regional web was assumed to have six trophic groups and three trophic levels (three basal groups, two intermediate groups and one top group), containing 20 species each. We assumed some degree of specialisation in the relationships between trophic levels: each species from the intermediate and top trophic levels had a number of prey equals to one plus a random number drawn in a Poisson law of parameter three (Fig. 1a). Thus, each consumer species had one prey species at least, and on average four prey species. Once the number of prey species had been drawn for a given consumer species, prey were drawn randomly from the species belonging to the lower trophic levels (T4 had preys in T1 and T2, T5 in T3 and T4 and T6 in T4 and T5).

Step 2 – Based on this regional network, we generated 1000 local multi-trophic communities. Communities were simulated using a stochastic model of multi-trophic community assembly inspired by the Trophic Theory of Island Biogeography (TTIB, Gravel *et al.* 2011; Massol *et al.* 2017). The TTIB assumes bottom-up sequential dependencies (Holt 1997, 2009; Dunne *et al.* 2002) with two phases. In phase 1, each species

can colonise a local community if at least one of its prey species is present. In phase 2, a species which has lost its last prey species goes extinct. For the sake of clarity, we assumed a homogeneous environment. The probability of each basal species being present in the local community was assumed to be constant and set to  $p_0 = 0.5$ . The probability of each consumer species C being present in the local community is related to the fraction of its prey available through the relation  $p_C = (k/g)^r$  where  $g$  is the diet breadth of C (i.e. the number of potential prey species),  $k$  is the number of its prey species present in the community and  $r$  is a constant that controls the shape of the relation. In the TTIB, having more prey species present in the community does not increase the probability of consumer presence, and the probability of survival is either 0 (when  $k = 0$ ) or 1 (when  $k > 0$ ). This corresponds to the case  $r = 0$ . For the simulation, we used  $r = 1$ , assuming that  $p_C$  grows linearly with the number of prey species present in the community. We also studied the case  $r = 1/3$  presented in the appendices.

Step 3 – We then computed  $\beta$ -diversity matrices for each trophic group. We inferred then the partial correlations and computed the marginal correlations between these  $\beta$ -diversities using the Glasso method. We thus obtained a distribution of partial and marginal correlations between pairs of the  $\beta$ -diversity at the different trophic levels. We expected these partial correlations to be high between trophic groups which interact directly and low between trophic groups which do not interact.

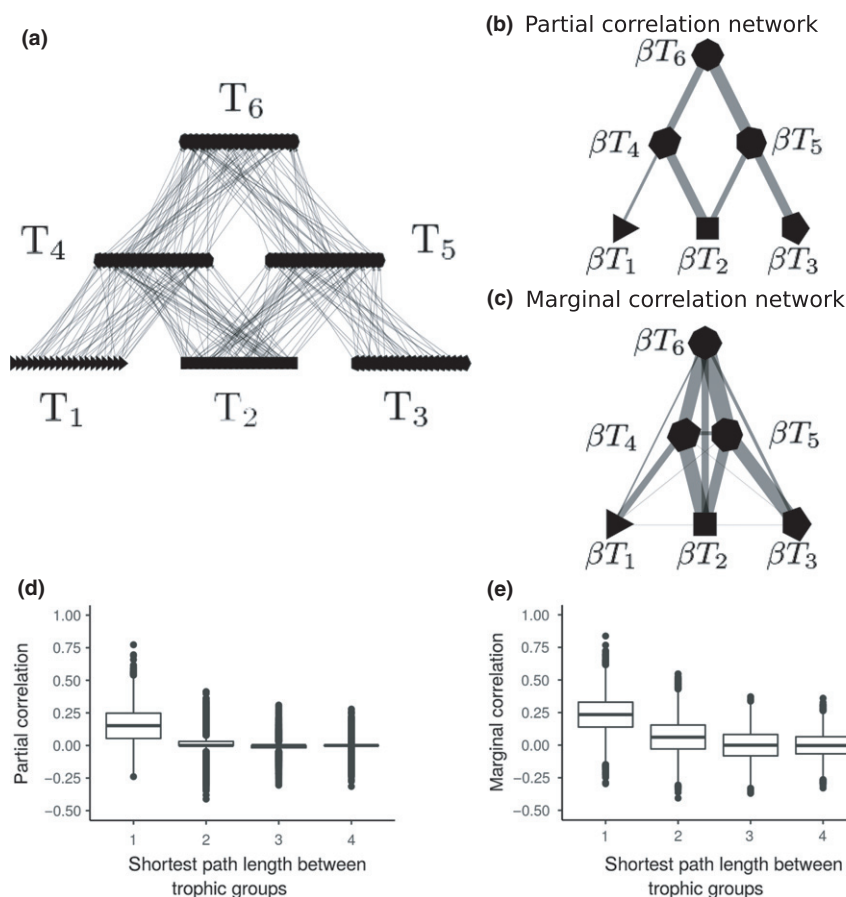
#### **Analysing multi-trophic patterns in soil ecosystems in the French Alps**

##### *Study site and soil sampling*

The study was conducted in the northern French Alps (Arves Massif, 45.12° N, 6.40° E) along a 977 m elevational gradient (1748 m to 2725 m a.s.l.) located in a single cow-grazed pasture, above the tree-line. The vegetation at the bottom of the gradient corresponds mainly to subalpine grasslands, while alpine meadows with sparse vegetation dominate at high elevation (Chalmandrier *et al.* 2017). Ten plots were established at 100 m altitude intervals along the gradient, each of them composed of two 10 × 10 m<sup>2</sup> subplots. All plots were placed on the same south-facing slope with a similar bedrock type and land-use to ensure a relatively homogeneous gradient. Mean annual temperature ranges between 8 °C at the bottom and 3 °C at the top, while mean annual rainfall is 473 mm over the period 2000–2012. The soil sampling field campaign was conducted in September 2012. We collected 21 soil samples per subplot. More details are presented in Appendix S1.

##### *Molecular analyses*

Soil biodiversity was estimated using four DNA markers. Universal markers such as 18S (amplifying all Eukaryotes, 18S nuclear rDNA) and 16S (amplifying all Bacteria, 16S rRNA) were used to obtain a general overview of the multi-trophic composition of the sites. Another two markers focus on Eukaryota diversity by targeting fungi (ITS1) and vascular plants (Chloroplast trnL-P6 loop) respectively. Molecular analysis and data curation are presented in Appendix S1. We pooled the samples together per subplot in order to obtain a single community per subplot and converted the data into



**Figure 1** Design and results of the simulation assessing the impact of biotic interactions on the marginal and partial correlation of the  $\beta$ -diversity values of different trophic groups. (a) the trophic network used for the simulation, composed of six trophic groups and three trophic levels. The indegree of each node is one plus an integer randomly drawn from a Poisson law of parameter three (except for the basal species). (b) Partial correlation network built using the median values of the partial correlations (inferred using the graphical lasso) between the  $\beta$ -diversity values of different trophic groups. (c) Marginal correlation networks built using the median values of the Pearson correlations between the  $\beta$ -diversity values of different trophic groups. (d) Partial correlation coefficient (inferred using the graphical lasso) between the  $\beta$ -diversity values of two trophic groups as a function of the shortest path length between these trophic groups. (e) Marginal Pearson correlation coefficient between the  $\beta$ -diversity values of two trophic groups as a function of the shortest path length between these trophic groups.

presence-absences. The raw and curated sequencing data as well as associated data and codes are available on the Dryad Digital Repository under accession <https://doi.org/10.5061/dryad.5b58400> and the summary statistics are presented in Table S1. As explained above, we computed  $\beta$ -diversity using the Jaccard dissimilarity index. Since this index is sensitive to sequencing depth difference between samples, we also used the true turnover component of the Jaccard index (Baselga 2010). The results obtained using the true turnover component of the Jaccard index are similar to those using the Jaccard index; so, we only refer to the Jaccard index in the main text, and present the results of the true turnover component of the Jaccard index in Appendix S2 (Figs S3–S7, Tables S2–S3). Moreover, we present the analysis using the Sorensen dissimilarity index and its true turnover component in Appendix S3 (Figs S8–S11).

#### Defining trophic groups of species

We selected *a priori* groups of soil taxa based on their distinctive role in the functioning of the soil ecosystem, namely: plants, fungi, bacteria, oribatid mites, nematodes and

springtails (Bardgett 2005). We included plants since they are the primary producers and their diversity and identity drive the functioning and the stability of most terrestrial ecosystems (Hooper 1997; van der Heijden *et al.* 1998). We also included litter feeders that contribute to dead material fragmentation, in particular oribatid mites and springtails. Oribatid mites form one of the most abundant groups of arthropods in soil, (Behan-Pelletier 1999) with up to several hundreds of thousands of individuals per square metre (Norton 1990). Springtails are the most numerous group of hexapods in most terrestrial ecosystems (Deharveng 2004). We also included a group of taxa that mineralise the fragmented litter, such as saprophytic fungi and bacteria (Bardgett 2005). Fungi are found in different compartments of the soil trophic web so we classified fungal OTUs into three main functional groups, namely, symbiotic fungi, saprophytic fungi and pathogenic fungi using the FUNguild database (Nguyen *et al.* 2016). Root associated symbiotic fungi rare found on over 90% of terrestrial plant families (Wang & Qiu 2006). Moreover, in arctic and alpine systems, 60–80% of the nitrogen available

for plants is supplied by mycorrhizal fungi (Bjorbækmo *et al.* 2010). Bacteria also contribute to this supply by fixing atmospheric nitrogen (Bonfante & Anca 2009; Haq *et al.* 2014). Finally, we included a group of predators, here nematodes, the most abundant belowground multicellular animals (Bardgett 2005). Nematode OTUs were divided into bacterivore nematodes and herbivore/fungivore nematodes using the NEMAGuild database (Nguyen *et al.* 2016). In summary, the nine groups included in this study were: plants, symbiotic fungi, pathogenic fungi, saprophytic fungi, bacterivore nematodes, herbi/fungivore nematodes, bacteria, springtails and oribatid mites. Most of these groups interact via trophic interactions, so that this study case matches the simulations described above, but not all of them. For the sake of clarity, we will hereafter refer to them as trophic groups.

#### *Environmental characteristics*

Mean annual soil temperature was estimated from field meteorological stations placed at the centre of each plot. We also estimated growing season length and number of frost days based on daily maps of snow cover at 15 m resolution for 5 years falling between 2000 and 2014 and air temperature values extracted from the SAFRAN meteorological model developed by Météo France for the French Alps (Durand *et al.* 2009). More methodological details and validation results for the snow cover model are available in Carlson *et al.* (2015). Fine-scale topography and associated parameters (topographic wetness index and slope) were inferred from airborne LIDAR data acquired the year of sampling. Mean soil pH over the gradient was 5.40 (SD 0.300), whereas mean soil temperature over the year was 6.06 °C (SD 2.84). The environmental distances between subplots were estimated with Euclidean distances from the first two axes of a principal component analysis run for all produced (and normalised) environmental variables (the variance captured by the first two axes was 34 and 25% respectively, the first axis roughly represented the climatic conditions, whereas the second axis was related to the soil conditions, see Appendix S4 (Figs S12–S14, Table S4) for more details).

## RESULTS

### **Approximating the structure of a simulated trophic network**

Our simulations showed that trophic interactions do produce non-zero marginal and partial correlations between  $\beta$ -diversity on consecutive trophic levels. Moreover, the Glasso method detected the conditional dependencies (i.e. an edge in the partial correlation network) between the  $\beta$ -diversity of the different trophic groups corresponding to their trophic position in the network contrary to the marginal correlations (Fig. 1b,c). The median of both marginal and partial correlation coefficients between pairs of trophic levels decreased with the shortest path length between these trophic levels. Nevertheless, while the median of the marginal correlation coefficients decreased slowly, the median of the partial correlation coefficients dropped to values close to 0 once the trophic level distance was higher than 1 (Fig. 1d,e). Changing the shape of the relationship linking the probability of presence of a

consumer species with its number of available prey species did not alter this conclusion (Fig. S15).

### **$\beta$ -diversity modelling of empirical soil communities**

In the French Alps, the partial correlations estimated between the  $\beta$ -diversity of each predefined trophic group and environmental distances were all positive (Figs 2, S16, S17, Table S5). The estimated partial correlation network had 11 nodes (9 trophic groups and 2 environmental variables), was composed of 34 undirected edges out of 55 possible edges and had a connectance of 0.618.

Saprophytic fungi were the most influential group in conditioning the  $\beta$ -diversity of the other groups (highest degree value, 8, and highest weighted degree value, 1.30, Fig. 3). Plants and oribatid mites also had a strong influence on the  $\beta$ -diversity of other groups, as did pathogenic and symbiotic fungi. In contrast, environmental variables had a relatively small direct impact on the  $\beta$ -diversity of the trophic groups.

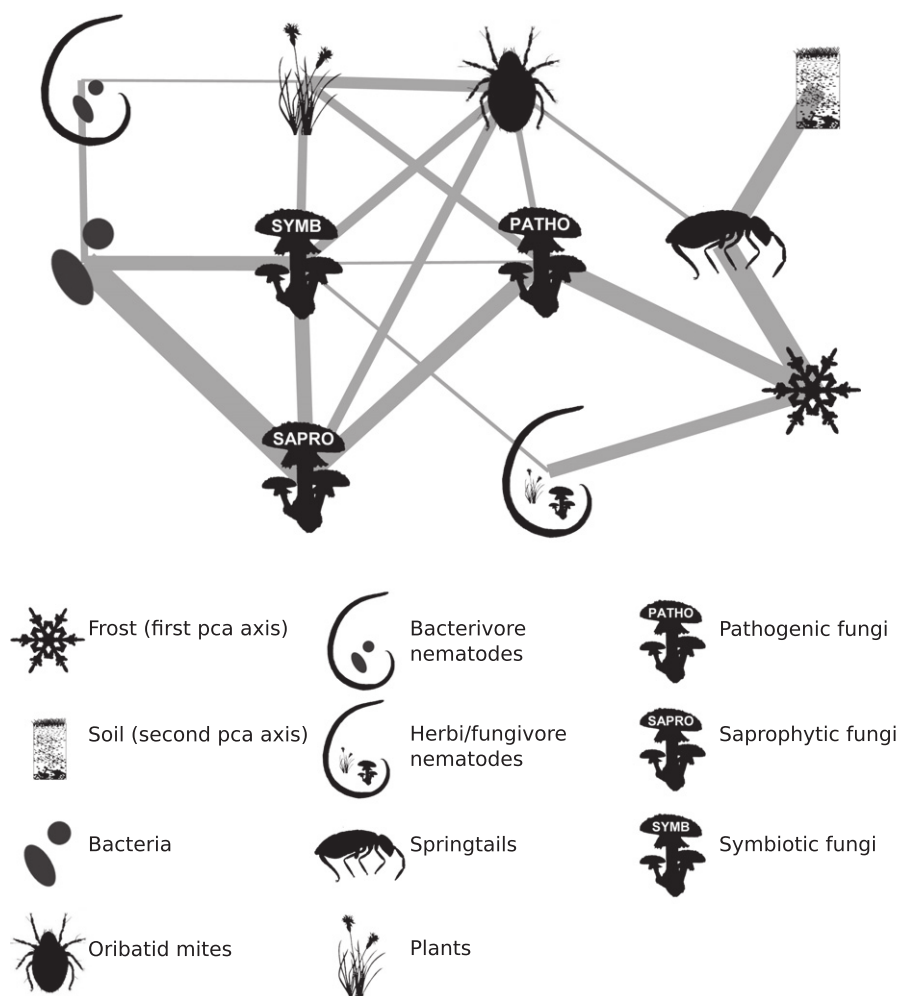
The probability of observing a non-null partial correlation between the  $\beta$ -diversity of a trophic group and the environmental distance was 0.44 (8 edges linking environmental nodes to the trophic group nodes and 18 potential edges), whereas the probability of observing a link between the  $\beta$ -diversity of any two trophic groups was 0.69 (25 edges and 36 potential edges). Since the variables associated with disconnected nodes were conditionally independent and that conditionally independent variables could not causally influence each other, this result demonstrates that, in general, environmental variables had a lower influence on the  $\beta$ -diversity of the trophic groups than the other trophic groups.

## DISCUSSION

In this study, we applied a method for dissecting the joint spatial structure of multiple trophic groups. This method builds on observed patterns of  $\beta$ -diversity in multiple trophic groups to infer the conditional dependencies between pairs of groups and with the environment, in order to pinpoint potential biotic interactions and influential effects of environmental variables on some specific groups. Simulations confirmed that our method is able to recover the overall structure of a trophic network using partial correlations of  $\beta$ -diversity between pairs of groups.

When applied to soil multi-trophic diversity along an elevation gradient of the French Alps, we were able to quantify the relative importance of biotic interactions and the environment in shaping the spatial structure of the meta-communities. Pairwise environmental distances displayed a low correlation with the  $\beta$ -diversity of each group (as measured with the few non-zero partial correlations). This result implies that the overall ecological community is primarily driven by biotic interactions and, to a lower extent, by environmental constraints (but still important on some groups like springtails or pathogenic fungi). This result is surprising, given the sharpness of the elevational gradient, and the expected importance of environmental filtering in shaping above and below-ground communities along elevation gradients (Meynard *et al.* 2013). However, many previous studies have focused on intraguild biotic interactions, or on environmental effect only (see Kraft *et al.*





**Figure 2** Undirected partial correlation network inferred using the graphical lasso method between the  $\beta$ -diversity of the major trophic groups constituting soil biodiversity and the environmental distances. Each node represents the  $\beta$ -diversity of a trophic group or an environmental distance. Here, only the partial correlations above the median value of the non-null partial correlation coefficients (0.106) are shown. The non-filtered network is presented in Fig. S1. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.

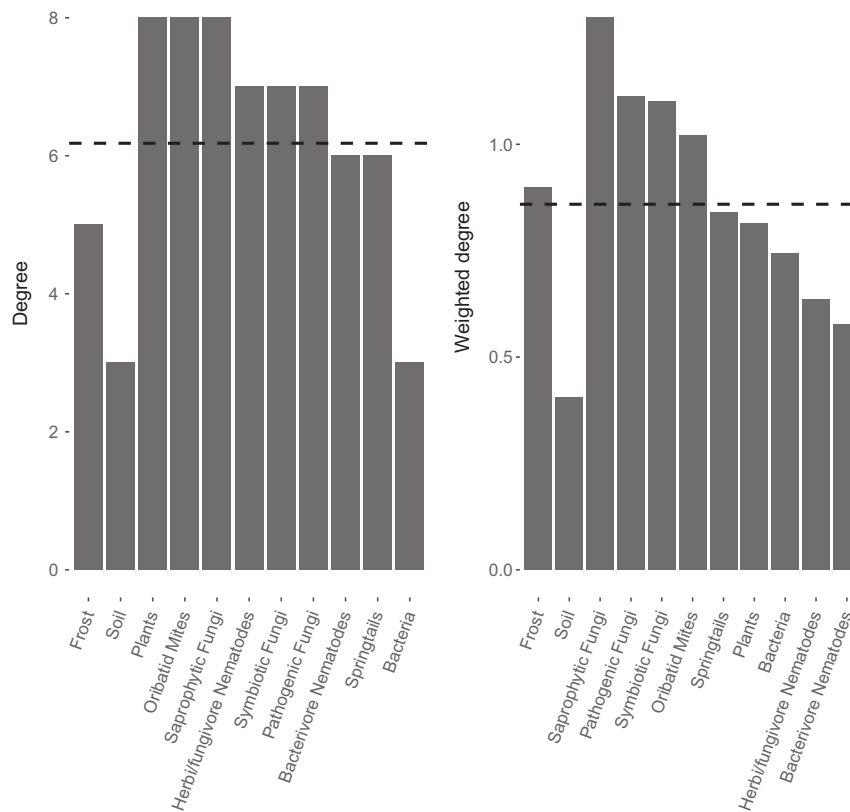
2015 for a critical synthesis), and they have ignored the importance of interguild interactions (e.g. Chalmandrier *et al.* 2015).

As our climatic and soil properties were sampled at plot level, they might not have been measured at the appropriate scale to reflect the fine conditions experienced by below-ground organisms (Falconer *et al.* 2015; Baveye *et al.* 2016; Matias *et al.* 2016; Zinger *et al.* 2017). Further studies should empirically investigate the  $\beta$ -diversity area relationships of the different groups (see Barton *et al.* 2013) and exhibit at which scale the spatial turnover of the different groups match together and with the environment (Barberán *et al.* 2015; Zinger *et al.* 2017). Ultimately, this will give an idea of the appropriate scale of sampling and collecting environmental information for such multi-trophic analyses.

Finally, we cannot rule out that some important missing environmental factors might explain the lower than expected predictive power of environmental distances (e.g. phosphorous for symbiotic mushrooms, Liu *et al.* 2012; Camenzind *et al.* 2014). Interestingly, focusing on the true turnover, instead of total turnover, reveals a stronger influence of environmental variation

in structuring the  $\beta$ -diversity of the different trophic groups. This means that the effects of the other groups are pivotal to explain the overall turnover and change in species richness (nestedness component) across space, while environmental variation is important in driving the pure turnover between groups.

The strong advantage of Graphical lasso over other related approaches (e.g. Bayesian network, path analyses) is that the partial correlation network can be inferred without assuming any *a priori* structure. This is important when the goal is of exploring and mapping co-variations between different groups and when little knowledge of the system is available. As explained in the introduction, the graphical lasso is expected to be sensitive to the effect of a missing predictor since the structure of the partial correlation network may be affected by the addition of a variable. In our case, we tested to what extent the addition of an environmental variable impacted the structure of the partial correlation network. We showed a moderate impact on the network topology, guaranteeing so the robustness of the analysis (Appendix S5, Figs S18–S20). This method does not directly infer the interaction network at



**Figure 3** Properties of the inferred network. The degree (left panel) is the number of neighbours of nodes in a graph (here, the undirected partial correlation network). It measures the number of variables that are conditionally dependent on the variable associated with this node. The weighted degree (right panel) is the sum of the partial correlation coefficients attached to the edges adjacent to this node. Dashed lines represent the mean values of degree and weighted degree.

the species level, and the inferred conditional dependencies at the group level do not imply causality. However, they shed light on potential interactions and pave the way for more analyses along identified co-variations.

The approach we proposed here should help interpret a wide range of multi-trophic biodiversity patterns and yield new testable hypotheses. For example, in multiple pairs of trophic groups, we found strong spatial associations which can be interpreted in light of functional associations. In the soil, plants are directly affected by symbiotic/pathogenic fungi and root herbivores, whereas the complex network of detritivorous organisms affects them indirectly (Bardgett 2005; Wardle *et al.* 2011). Our method showed that plant  $\beta$ -diversity was strongly linked to the  $\beta$ -diversities of symbiotic fungi, pathogenic fungi, oribatid mites and bacterivore nematodes. While the links with the two groups of fungi most likely reflect the strong direct associations between these groups and the vegetation structure, the link with oribatid mites may be explained by the fact that this group feeds on plant litter fungi, especially saprophytic fungi (Schneider *et al.* 2004; Crowther *et al.* 2011; Crowther & A'Bear 2012). The graphical lasso method detected this link showing that changes in the composition of saprophytic fungi lead to a change in the oribatid mite assemblages. As expected, the  $\beta$ -diversity of bacteria correlated with that of bacterivore nematodes, which reflects this known trophic interaction (Ettema 1998; Wardle

2006). Moreover, the  $\beta$ -diversity of herbi/fungivore nematodes correlated with that of symbiotic fungi. The link between the  $\beta$ -diversity of oribatid mites and springtails might be explained by mite predation of springtails (Ferguson & Joly 2002). We also highlighted a link between the  $\beta$ -diversity of oribatid mites and symbiotic fungi. This result suggests that mite assemblages could be influenced by fungi spatial distribution through trophic interactions, which have been so far very poorly documented (Gange & Brown 2003). Interestingly, our analysis also showed a strong partial correlation (the strongest partial correlation: 0.365) between the  $\beta$ -diversities of saprophytic fungi and bacteria. Indeed, bacteria can rely on decomposition products from organic matter, which are provided by saprophytic fungi fully equipped from an enzymatic point of view (De Boer *et al.* 2005; Romani *et al.* 2006). The relationship uncovered between bacteria and symbiotic fungi could be attributed to the fact that bacteria can assist mycorrhiza by colonising the extraradical hyphae or by living in the cytoplasm of mycorrhizal fungi (Bonfante & Anca 2009; Haq *et al.* 2014). We did not observe direct relationship between bacteria and plants, probably because we lumped together bacterial taxa typical from the bulk soil (e.g. acidobacteria) with those interacting with plants (e.g. N-fixing bacteria or pathogens). A better functional assignment of bacterial OTUs would probably help to clarify this identified relationship likely due to mutualistic bacterial OTUs.

## CONCLUSIONS

The rise of environmental DNA metabarcoding and the ever-increasing availability of databases on species co-occurrence have opened up a new era in quantitative and predictive ecology. While comprehensive species lists of taxa are necessary, they do tell much on how species interact across space and time and how multi-trophic interactions shape community assembly. The method we proposed here addresses this challenge by revealing how trophic groups influence each other and respond to environmental variation. As such, our method is able to uncover the potential determinants of the compositional turnover of species groups from a multi-trophic interaction perspective. This paves the way for larger applications of this method to ecological data where comprehensive biodiversity assessments are becoming more and more available and where knowledge of the structure of the system is still limited.

## ACKNOWLEDGEMENTS

We would like to thank all the students and colleagues who participated in the intensive field season that generated the data used in the manuscript, as well as the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) and Pierre Solbes (EDB-Calc Cluster) for providing computing and storage resources. The research received funding from the French Agence Nationale de la Recherche (ANR) through the METABAR (ANR-11-BSV7-0020) and GlobNets (ANR-16-CE02-0009) projects, and from 'Investissement d'Avenir' grants managed by the ANR (Trajectories: ANR-15-IDEX-02; Montane: OSUG@2020: ANR-10-LAB-56; CEBA: ANR-10-LABX-25-01, TULIP: ANR-10-599-LABX-0041).

## AUTHORSHIP

WT, FM and MO conceived the ideas with the help of LC. LC, EC, LG, JP, FM, PT, LZ and WT collected the soil and environmental data in the field. LG, JP and PT ran the PCR and DNA extractions. LZ and VS ran the bioinformatic analyses. MO ran all statistical analyses and developed the method together with FM and WT. MO, FM and WT wrote the first version of the paper, and all authors contributed substantially to the revisions.

## COMPETING INTERESTS

LG and PT are co-inventors of patents related to the gh primers and the use of the P6 loop of the chloroplast trnL (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers.

## DATA ACCESSIBILITY STATEMENT

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.5b58400>

## REFERENCES

Allesina, S., Alonso, D. & Pascual, M. (2008). A general model for food web structure. *Science*, 320, 658–661.

- Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L. *et al.* (2011). Navigating the multiple meanings of  $\beta$  diversity: a roadmap for the practicing ecologist. *Ecol. Lett.*, 14, 19–28.
- Barberán, A., McGuire, K.L., Wolf, J.A., Jones, F.A., Wright, S.J., Turner, B.L. *et al.* (2015). Relating belowground microbial composition to the taxonomic, phylogenetic, and functional trait distributions of trees in a tropical forest. *Ecol. Lett.*, 18, 1397–1405.
- Bardgett, R.D. (2005). *The Biology of Soil: A Community and Ecosystem Approach*. Oxford University Press, Oxford Biology of Habitat Series.
- Bardgett, R.D. & Wardle, D.A. (2010). *Aboveground-Belowground Linkages: Biotic Interactions, Ecosystem Processes, and Global Change*. Oxford University Press, Oxford Series in Ecology and Evolution.
- Barton, P.S., Cunningham, S.A., Manning, A.D., Gibb, H., Lindenmayer, D.B. & Didham, R.K. (2013). Spatial scaling of beta diversity. *Global Ecol. Biogeogr.*, 22, 639–647.
- Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.*, 19, 134–143.
- Baselga, A. & Orme, C.D.L. (2012). betapart : an R package for the study of beta diversity. *Methods Ecol. Evol.*, 3, 808–812.
- Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.
- Baveye, P.C., Berthelin, J. & Munch, J.-C. (2016). Too much or not enough: reflection on two contrasting perspectives on soil biodiversity. *Soil Biol. Biochem.*, 103, 320–326.
- Behan-Pelletier, V.M. (1999). Oribatid mite biodiversity in agrosystems: role for bioindication. *Agric. Ecosyst. Environ.*, 74, 411–423.
- Bjorbækmo, M., Carlsen, T., Brysting, A., Vrålstad, T., Høiland, K., Ugland, K. *et al.* (2010). High diversity of root associated fungi in both alpine and arctic *Dryas octopetala*. *BMC Plant Biol.*, 10, 244.
- Bonfante, P. & Anca, I.-A. (2009). Plants, Mycorrhizal Fungi, and Bacteria: a network of interactions rhizosphere: the narrow zone of soil surrounding living roots. *Annu. Rev. Microbiol.*, 63, 363–383.
- Broeckling, C.D., Broz, A.K., Bergelson, J., Mamer, D.K. & Vivanco, J.M. (2008). Root exudates regulate soil fungal community composition and diversity. *Appl. Environ. Microbiol.*, 74, 738–744.
- Brose, U. & Hillebrand, H. (2016). Biodiversity and ecosystem functioning in dynamic landscapes. *Philos. Trans. R. Soc. B Biol. Sci.*, 371, 20150267.
- Camenizind, T., Hempel, S., Homeier, J., Horn, S., Velescu, A., Wilcke, W. *et al.* (2014). Nitrogen and phosphorus additions impact arbuscular mycorrhizal abundance and molecular diversity in a tropical montane forest. *Glob. Change Biol.*, 20, 3646–3659.
- Carlson, B.Z., Choler, P., Renaud, J., Dedieu, J.P. & Thuiller, W. (2015). Modelling snow cover duration improves predictions of functional and taxonomic diversity for alpine plant communities. *Ann. Bot.*, 116, 1023–1034.
- Chalmandrier, L., Münkemüller, T., Lavergne, S. & Thuiller, W. (2015). Effects of species' similarity and dominance on the functional and phylogenetic structure of a plant meta-community. *Ecology*, 96, 143–153.
- Chalmandrier, L., Münkemüller, T., Colace, M.P., Renaud, J., Aubert, S., Carlson, B.Z. *et al.* (2017). Spatial scale and intraspecific trait variability mediate assembly rules in alpine grasslands. *J. Ecol.*, 105, 277–287.
- Crowther, T.W. & A'Bear, A.D. (2012). Impacts of grazing soil fauna on decomposer fungi are species-specific and density-dependent. *Fungal Ecol.*, 5, 277–281.
- Crowther, T.W., Boddy, L. & Jones, T.H. (2011). Species-specific effects of soil fauna on fungal foraging and decomposition. *Oecologia*, 167, 535–545.
- De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel, H., *et al.* (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecol. Lett.*, 15, 740–747.
- De Boer, W., Folman, L.B., Summerbell, R.C. & Boddy, L. (2005). Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol. Rev.*, 29, 795–811.

- Deharveng, L. (2004). Recent advances in Collembola systematics. *Pedobiologia*, 48, 415–433.
- Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002). Network structure and biodiversity loss in food webs: robustness increase with connectance. *Ecol. Lett.*, 5, 558–567.
- Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B. & Mérindol, L. (2009). Reanalysis of 44 yr of climate in the French Alps (1958–2002): methodology, model validation, climatology, and trends for air temperature and precipitation. *J. Appl. Meteorol. Climatol.*, 48, 429–449.
- Epskamp, S. & Fried, E.I. (2016). A tutorial on regularized partial correlation networks. arXiv:1607.01367
- Ettema, C.H. (1998). Soil nematode diversity: species coexistence and ecosystem function. *J. Nematol.*, 30, 159–169.
- Falconer, R.E., Battaia, G., Schmidt, S., Baveye, P., Chenu, C. & Otten, W. (2015). Microscale heterogeneity explains experimental variability and non-linearity in soil organic matter mineralisation. *PLoS ONE*, 10, 1–12.
- Ferguson, S.H. & Joly, D.O. (2002). Dynamics of springtail and mite populations: the role of density dependence, predation, and weather. *Ecol. Entomol.*, 27, 565–573.
- Foygel, R. & Drton, M. (2010). Extended Bayesian information criteria for gaussian graphical models. *Adv. Neural. Inf. Process. Syst.*, 23, 604–612.
- Friedman, J., Hastie, T. & Tibshirani, R. (2007). Sparse inverse covariance estimation with the lasso. *Biostatistics*, 9, 432–441.
- Gange, A.C. & Brown, V.K. (2003). Actions and interactions of soil invertebrates and arbuscular mycorrhizal fungi in affecting the structure of plant communities. In: *Mycorrhizal Ecology* (eds van der Heijden, M.G.A. & Sanders, I.R.). Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 321–344.
- Geremia, R.A., Púscas, M., Zinger, L., Bonneville, J.M. & Choler, P. (2016). Contrasting microbial biogeographical patterns between anthropogenic subalpine grasslands and natural alpine grasslands. *New Phytol.*, 209, 1196–1207.
- Gonzalez-Varo, J.P. & Traveset, A. (2016). The labile limits of forbidden interactions. *Trends Ecol. Evol.*, 31, 700–710.
- Graham, C.H. & Fine, P.V.A. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecol. Lett.*, 11, 1265–1277.
- Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecol. Lett.*, 14, 1010–1016.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B.H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.*, 10, 497–506.
- Haq, I.U., Zhang, M., Yang, P. & van Elsas, J.D. (2014). The interactions of bacteria with fungi in soil: emerging concepts. *Adv. Appl. Microbiol.*, 89, 432–441.
- Harris, D.J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, 97, 3308–3314.
- Hättenschwiler, S., Tiunov, A.V. & Scheu, S. (2005). Biodiversity and litter decomposition in terrestrial ecosystems. *Annu. Rev. Ecol. Evol. Syst.*, 36, 191–218.
- Hedlund, K. & Öhrn, M.S. (2000). Tritrophic interactions in a soil community enhance decomposition rates. *Oikos*, 88, 585–591.
- van der Heijden, M.G.A., Klironomos, J.N., Ursic, M., Moutoglis, P., Streitwolf-Engel, R., Boller, T. *et al.* (1998). Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature*, 396, 69–72.
- HilleRisLambers, J., Adler, P.B., Harpole, W.S., Levine, J.M. & Mayfield, M.M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annu. Rev. Ecol. Evol. Syst.*, 43, 227–248.
- Holt, R.D. (1997). From metapopulation dynamics to community structure: some consequences of spatial heterogeneity. *Metapopulation Biology: Ecology, Genetics and Evolution* (eds Hanski, I.A. & Gilpin, M.E.). Academic Press, New York, pp. 149–164.
- Holt, R.D. (2009). Towards a trophic island biogeography: reflections on the interface of island biogeography and food web ecology. In: *The Theory of Island Biogeography Revisited* (eds Losos, J.B. & Ricklefs, R.E.). Princeton University Press, Princeton, pp. 143–185.
- Hooper, D.U. (1997). The effects of plant composition and diversity on ecosystem processes. *Science*, 277, 1302–1305.
- 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 Biol.*, 14, e1002527.
- Koller, D. & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques – Adaptive Computation and Machine Learning*. The MIT Press, Cambridge, USA.
- Kraft, N.J., Adler, P.B., Godoy, O., James, E.C., Fuller, S., Levine, J.M. *et al.* (2015). Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.*, 29, 592–599.
- Kress, W.J., Garcia-Robledo, C., Uriarte, M. & Erickson, D.L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends Ecol. Evol.*, 30, 25–35.
- Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F. *et al.* (2015). Ocean plankton: determinants of community structure in the global plankton interactome. *Science*, 348, 1262013.
- Liu, Y., Shi, G., Mao, L., Cheng, G., Jiang, S., Ma, X. *et al.* (2012). Direct and indirect influences of 8 yr of nitrogen and phosphorus fertilization on Glomeromycota in an alpine meadow ecosystem. *New Phytol.*, 194, 523–535.
- Massol, F., Dubart, M., Calcagno, V., Cazelles, K., Jacquet, C., Kéfi, S. *et al.* (2017). Chapter four – island biogeography of food webs. *Adv. Ecol. Res.*, 56, 183–262.
- Matias, M.G., Pereira, C.L., Raposeiro, P.M., Gonçalves, V., Cruz, A.M., Costa, A.C. *et al.* (2016). Divergent trophic responses to biogeographic and environmental gradients. *Oikos*, 126, 101–110.
- Mazel, F., Wüest, R.O., Gueguen, M., Renaud, J., Ficetola, G.F., Lavergne, S. *et al.* (2017). The geography of ecological niche evolution in mammals. *Curr. Biol.*, 27, 1369–1374.
- Mazumder, R. & Hastie, T. (2011). The graphical lasso: new insights and alternatives. arXiv:1111.5479
- Meynard, C.N., Lavergne, S., Boulangeat, I., Garraud, L., Van Es, J., Mouquet, N. *et al.* (2013). Disentangling the drivers of metacommunity structure across spatial scales. *J. Biogeogr.*, 40, 1560–1571.
- Murphy, K.P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, USA.
- Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J. *et al.* (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.*, 20, 241–248.
- Norton, R.A. (1990). Acarina: Oribatida. In: *Soil Biology Guide* (ed. Dindal, D.L.). Wiley, New York, pp. 779–803.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D. *et al.* (2015). vegan: community ecology package. R Package Version 2.3-2, <https://CRAN.R-project.org/package=vegan>.
- Peay, K.G., Russo, S.E., Mcguire, K.L., Lim, Z., Chan, J.P., Tan, S. *et al.* (2015). Lack of host specificity leads to independent assortment of dipterocarps and ectomycorrhizal fungi across a soil fertility gradient. *Ecol. Lett.*, 18, 807–816.
- Peay, K.G., Kennedy, P.G. & Talbot, J.M. (2016). Dimensions of biodiversity in the Earth mycobiome. *Nat. Rev. Microbiol.*, 14, 434–447.
- Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S. *et al.* (2015). Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecol. Lett.*, 18, 85–95.
- Ricklefs, R.E. (1987). Community diversity: relative roles of local and regional processes. *Science*, 235, 167–71.
- Romaní, A.M., Fischer, H., Mille-Lindblom, C. & Tranvik, L.J. (2006). Interactions of bacteria and fungi on decomposing litter: differential extracellular enzyme activities. *Ecology*, 87, 2559–2569.
- Schneider, K., Renker, C., Scheu, S. & Maraun, M. (2004). Feeding biology of oribatid mites: a minireview. In: *Phytophaga XIV: Acarine Biodiversity in the Natural and Human Sphere* (Proceedings of the Symposium of the European Association of Acarologists), Berlin, Germany, pp. 247–256.

- Shipley, B. (2000). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Ecology, Cambridge, UK: Cambridge University Press, pp. 48–55.
- Schuldt, A., Bruelheide, H., Buscot, F., Assmann, T., Erfmeier, A., Klein, A.M. *et al.* (2017). Belowground top-down and aboveground bottom-up effects structure multitrophic community relationships in a biodiverse forest. *Sci. Rep.*, 7(1).
- Smith, S.E., Read, D.J. & David, J. (2008). *Mycorrhizal Symbiosis*. Academic Press, Cambridge, USA.
- Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L.H. (2012). Environmental DNA. *Mol. Ecol.*, 21, 1789–1793.
- Tedersoo, L., Bahram, M., Cajthaml, T., Pölme, S., Hiiesalu, I., Anslan, S. *et al.* (2016). Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *ISME J.*, 10, 346–362.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc.*, 58, 267–288.
- Walker, J.F., Aldrich-Wolfe, L., Riffel, A., Barbare, H., Simpson, N.B., Trowbridge, J. *et al.* (2011). Diverse helotiales associated with the roots of three species of arctic ericaceae provide no evidence for host specificity. *New Phytol.*, 191, 515–527.
- Wang, B. & Qiu, Y.L. (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza*, 16, 299–363.
- Wardle, D.A. (2006). The influence of biotic interactions on soil biodiversity. *Ecol. Lett.*, 9, 870–886.
- Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., Wim, H., Putten, V.D. *et al.* (2011). Ecological linkages between aboveground and belowground biota. *Am. Assoc. Advancement Sci. Stable*, 304, 1629–1633.
- Wright, S. (1921). Correlation and causation. *J. Agric. Res.*, 20, 557–585.
- Zinger, L., Lejon, D.P.H., Baptist, F., Bouasria, A., Aubert, S., Geremia, R.A. *et al.* (2011). Contrasting diversity patterns of crenarchaeal, bacterial and fungal soil communities in an alpine landscape. *PLoS ONE*, 6, 1–7.
- Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M. *et al.* (2017). Soil community assembly varies across body sizes in a tropical forest. bioRxiv 154278; <https://doi.org/10.1101/154278>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, Jonathan Chase

Manuscript received 19 February 2018

First decision made 1 April 2018

Second decision made 11 July 2018

Manuscript accepted 25 July 2018

### **3.3 Annexes de l'article (en anglais)**

## Supplementary Information

### *Appendix 1 : Soil sampling, molecular and bioinformatics analyses*

We collected twenty-one soil samples per sub-plot following the two diagonals with the distance increasing exponentially from the corners to the centre on one diagonal and from the centre to the corners on the other. Central points were sampled only once. This sampling scheme was implemented to cover as much local biodiversity as possible at the sub-plot scale (Taberlet et al. 2012; Chalmandrier et al. 2017). Each sample contained 50g of soil from the uppermost 10 cm of soil and was placed in individual plastic bags to be processed individually. Corers were cleaned and sterilised using a blowtorch between each sample to prevent cross-contamination (Taberlet et al. 2012). Extracellular DNA was extracted from 15g of soil as previously described (Taberlet et al. 2012; Zinger et al. 2017) within four hours after sample collection to prevent microbial growth. DNA was extracted twice for each sample. Blank extraction controls were included in the extraction process (using solely phosphate buffer as the template).

For each DNA extract, PCRs were run in duplicate leading to four PCR replicates per core sample. For each marker, PCM products were purified (MinElute™ PCR purification kit, Qiagen) with extraction and PCR blank controls included in the mix. High-throughput sequencing of eukaryotes and plant amplicons was performed on an Illumina HiSeq 2500 platform (2x150 bp paired-end reads for Eukaryotes and 2\*200pb for plants) while fungal, archae and bacterial amplicons (200-350 bp) were sequenced on an Illumina MiSeq (2 x 250 bp paired-end reads) platform. Data were then curated using the OBITools software package (Boyer *et al.* 2016) together with custom R scripts (R Core Team 2015) following the procedure described in Zinger *et al.* (2017). Paired-end reads were assembled, assigned to their respective samples/marker and dereplicated. Low-quality sequences were excluded. For the remaining ones we computed pairwise dissimilarities between sequences (i.e. the number of mismatches, allowed to be 0-3) using the Sumatra algorithm (Mercier *et al.* 2013), and formed Operational Taxonomic Units OTUs using the Infomap community detection algorithm (Rosvall & Bergstrom 2008). Abundance of OTUs was defined as the sum of read abundances of sequences belonging to them. In subsequent analyses, each OTU was represented by its most abundant sequence. Each OTU was assigned a taxonomic clade using the *ecotag* program (Boyer *et al.* 2016) using a set of reference databases to refine taxonomic annotations. (e.g. GENBANK, UNITE database, Abarenkov *et al.* 2010) We paid particular attention to minimize PCR/sequencing errors, contaminant and false positive sequences as well as potential non-functional PCRs by using the procedure described in Zinger *et al.* (2017). PCR replicates were finally summed for each samples.

	<b>Bacteria</b>	<b>Eukaryota</b>	<b>Fungi</b>	<b>Viridiplantae</b>
<b>Primers set</b>				
Targeted region	V5-V6 regions, 16S rRNA gene	V7 region, 18S rRNA gene	ITS1	P6 loop of the chloroplast trnL intron
Forward primer (5'–3')	GGATTAGATACCC TGGTAGT <sup>‡</sup>	TTTGTCTGSTTAATTSCG <sup>§</sup>	CAAGAGATCCGTTGTT GAAAGTK <sup>†</sup>	GGGCAATCCTGAG CCAA <sup>¶</sup>
Reverse primer (5'–3')	CACGACACGAGCT GACG <sup>‡</sup>	TCACAGACCTGTTATTGC <sup>§</sup>	GGAAGTAAAAGTCGTA ACAAGG <sup>§§</sup>	CCATTGAGTCTCT GCACCTATC <sup>¶</sup>
<b>PCR conditions</b>				
Number of cycles	35	45	40	40
PCR cycles characteristics <sup>†</sup> (Denaturation, Annealing, Elongation – Final elongation)	95°C(30s), 57°C(30s), 72°C(90s) – 72°C(7min)	95°C(30s), 45°C(30s) 72°C(60s) – 72°C(7min)	95°C (30 sec), 55°C (30 sec), 72°C (60s) – 72°C (7 min)	95°C(30s), 50°C(30s), 72°C(60s) – 72°C(7min)
Illumina platform	MiSeq 2x250	HiSeq 2x150	MiSeq 2x250	HiSeq 2x100
<b>Reads and Otus statistics</b>				
Initial number of OTUs	25383	7242	4193	2153
Final number of OTUs	24058	5833	2987	555
Initial number of reads	5042640	68893592	4159074	37509301
Final number of reads	4357053	61818803	3626819	35950935

<sup>†</sup> this study

<sup>§§</sup> White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 In: PCR Protocols: A Guide to Methods and Applications, eds. Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White. Academic Press, Inc., New York.

<sup>‡</sup> Fliegerova, K., Tapio, I., Bonin, A., Mrazek, J., Callegari, M. L., Bani, P., *et al.* (2014). Effect of DNA extraction and sample preservation method on rumen bacterial population. *Anaerobe*, **29**, 80-84.

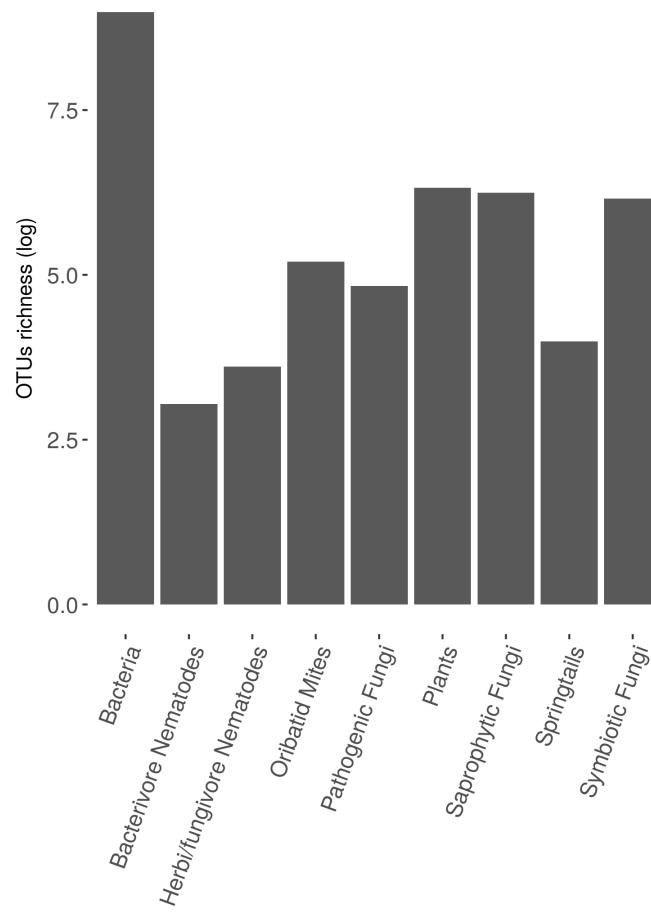
<sup>§</sup> Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangenstein, O. S., & Turon, X. (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. *PLoS One*, **10**, e0139633.

<sup>¶</sup> Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105-1109.

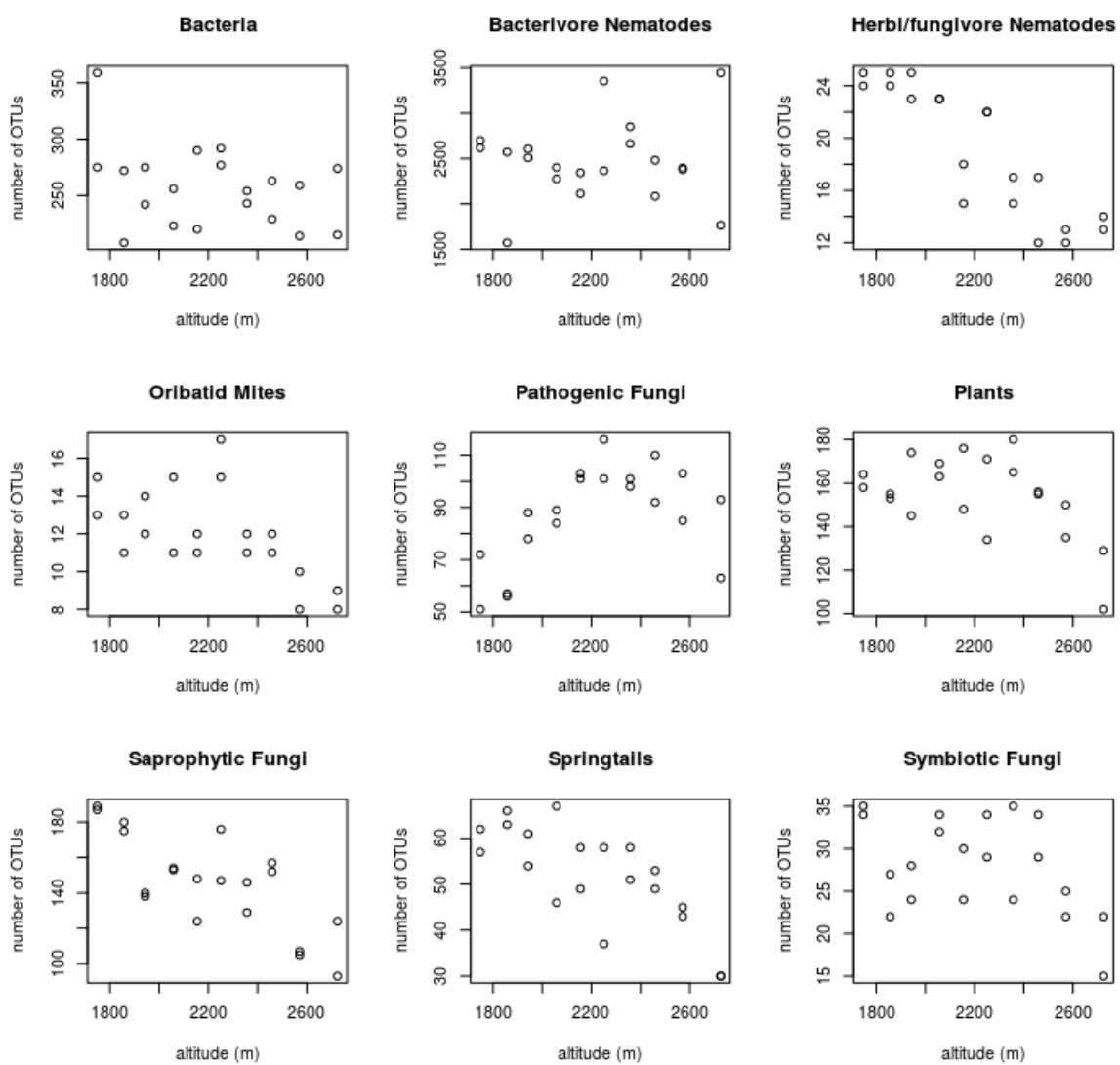


†† All PCRs were preceded by an initial denaturation at 95°C (10 min)

**Table S1** Markers characteristics, PCR conditions and sequencing technology



**Figure S1** Histogram of the natural logarithm of the numbers of Operational Taxonomic Units (OTUs) assigned to each trophic group.



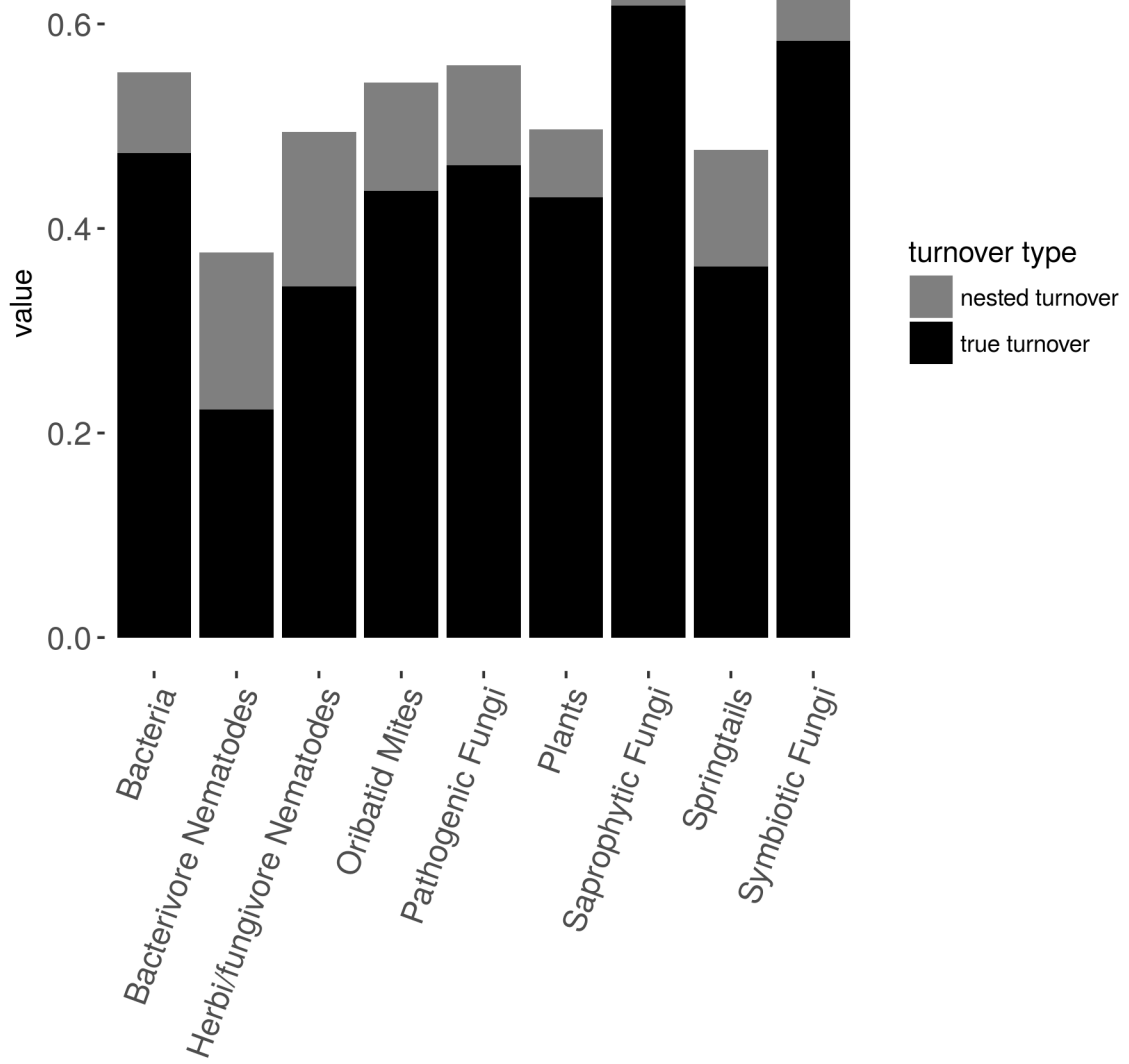
**Figure S2** Number of OTUs per trophic group in function of the altitude of the plots

*Appendix 2 : Statistical analysis using the true-turnover component of the  $\beta$ -diversity (Jaccard index)*

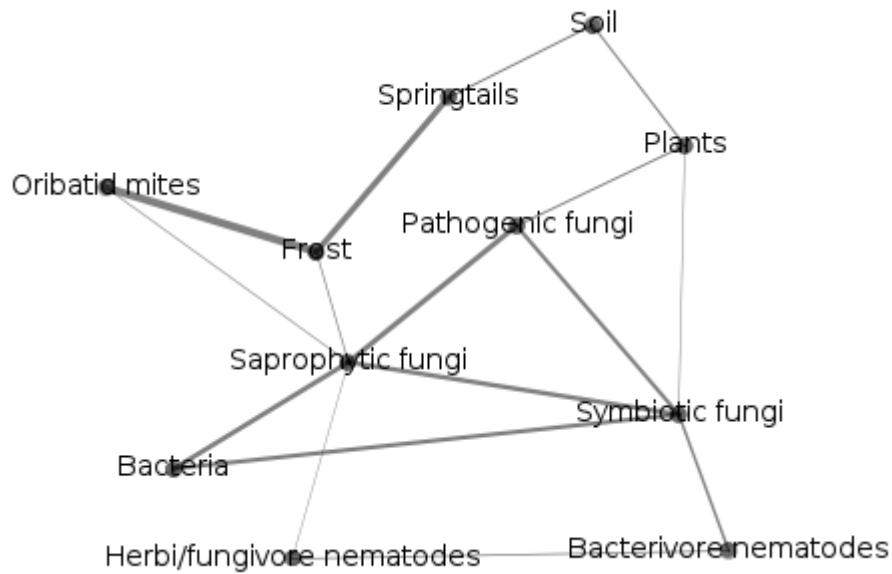
We ran the statistical analyses explained in the main text on the true turnover component of the  $\beta$ -diversity (Baselga 2010) of each predefined trophic group. The largest amount of  $\beta$ -diversity was due to true-turnover (Fig. 9) and on average, 80% of the  $\beta$ -diversity across trophic groups. The partial correlations estimated between the true turnover component of the  $\beta$ -diversity of each predefined trophic group and environmental distances were all positive except one coefficient (Fig. S10, Fig. S11, Fig. S12, table S4). The estimated partial correlation network had a connectance of 0.60 and was composed of 33 undirected edges (i.e. partial correlation coefficients > 0) out of 55 possible edges and 11 nodes (9 trophic groups and 2 environmental variables). The mean value of non null partial correlation coefficients was 0.133, the median was 0.102 and the maximum 0.459.

Saprophytic fungi seemed to be the most influential group in conditioning the  $\beta$ -diversity of the other groups (highest degree value, 9, and highest weighted degree value, 1.37, Fig. S12). Environmental variables had a moderate impact on the  $\beta$ -diversity of the trophic groups (degree of 6 for the first PCA axis, 4 for the second PCA axis and whereas the mean degree of the partial correlation network was 6; weighted degree of 1.13 for the first PCA-axis, 0.381 for the second PCA axis node whereas the mean value is 0.759).

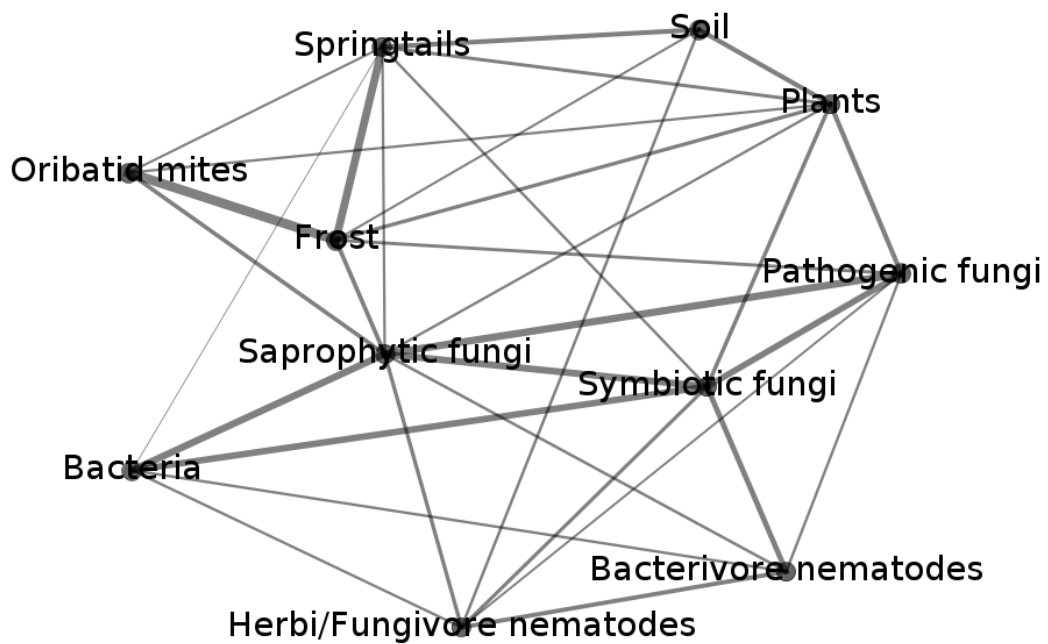
The probability of observing a non-null partial correlation between the  $\beta$ -diversity of a trophic group and the environmental distance was 0.44 (8 edges linking environmental nodes to the trophic groups nodes and 18 potential edges) whereas the probability of observing a link between the  $\beta$ -diversity of any two trophic groups was 0.67 (24 edges and 36 potential edges), highlighting the weaker influence of environmental variables on the true turnover component of the  $\beta$ -diversity of the trophic groups.



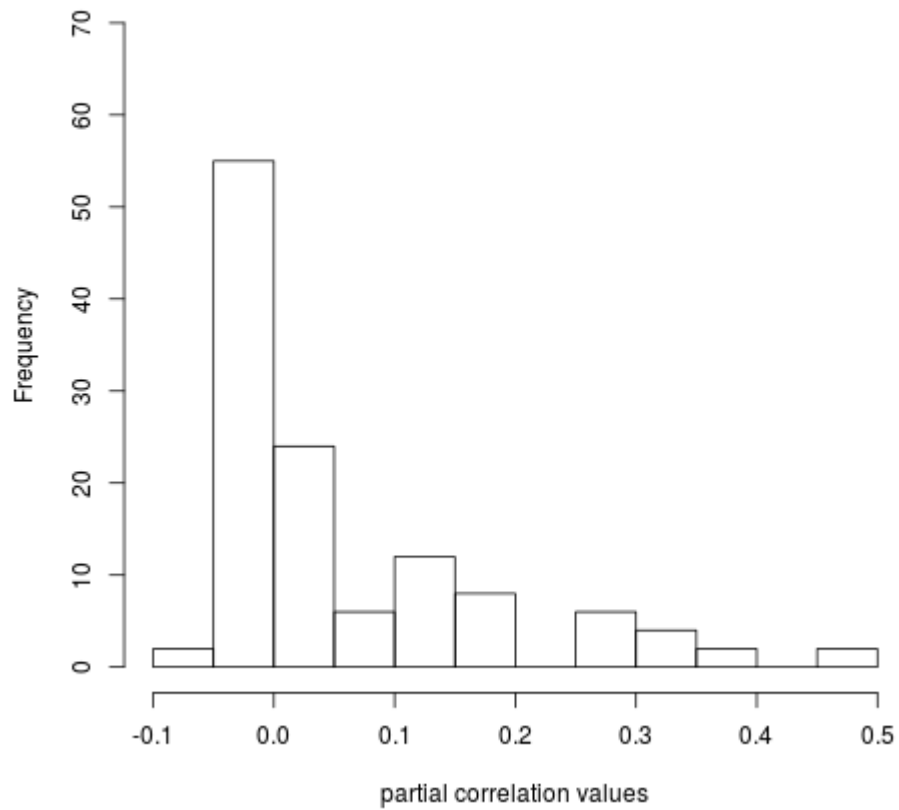
**Figure S3** Mean true turnover and nested turnover components of the Jaccard index for each trophic group.



**Figure S4** Undirected partial correlation network inferred by the graphical lasso method between the true turnover component of the  $\beta$ -diversity of the major trophic groups constituting of soil biodiversity and the environmental distances. Each node represent the  $\beta$ -diversity of a trophic group or an environmental distance. Here, are only shown the partial correlations above the median value of the non-null partial correlation coefficients (0.102). The non-filtered network is presented in Figure S11. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.

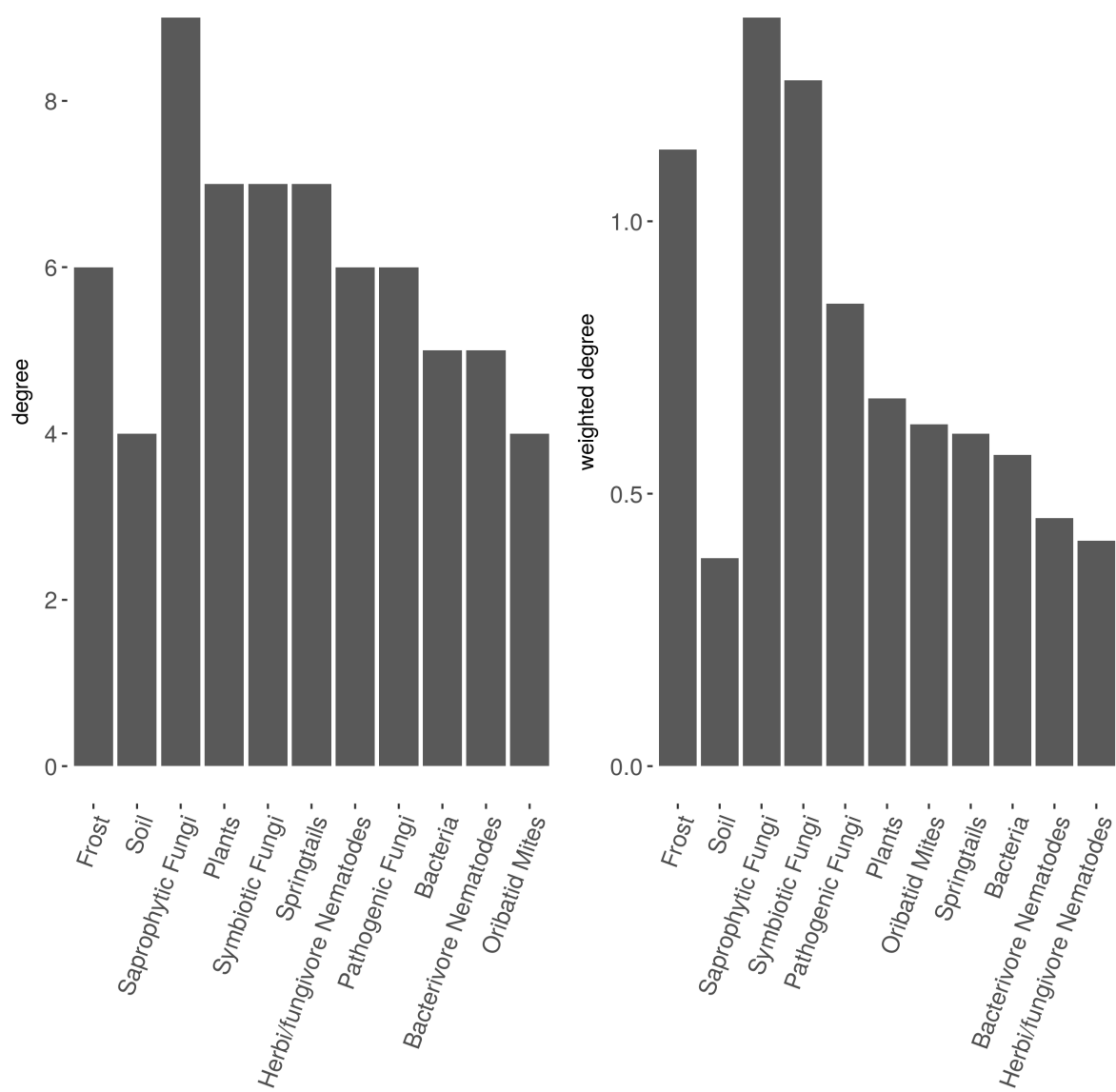


**Figure S5** Undirected partial correlation network inferred by the graphical lasso method between the true turnover component of the  $\beta$ -diversity of the major trophic groups constituting of soil biodiversity and the environmental distances. Each node represent the  $\beta$ -diversity of a trophic group or an environmental distance. The connectance of the graph is 0.600 . Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.



**Figure S6** Histogram of the partial correlation coefficients estimated between the true turnover component of the  $\beta$ -diversity of major trophic groups constituting of soil biodiversity and environmental distances. Partial correlation coefficients are all positive.





**Figure S17 Properties of the inferred network.** The mean degree (left panel) is the number of neighbors of nodes in a graph (here, the undirected partial correlation network). It measures the number of variables that are conditionally dependent to the variable associated to this node. The weighted degree (right panel) is the sum of the partial correlation coefficients attached to the edges adjacent this node.

	Frost	Soil	Plants	Bacteria	Herbi/Fun	Bacterivor	Oribatida	Saprophyt	Symbiotic	Pathogeni	Springtails
--	-------	------	--------	----------	-----------	------------	-----------	-----------	-----------	-----------	-------------

					givre nematodes	e nematodes		ic fungi	fungi	c fungi	
Frost	0	0.0122	0.1	0	0	0	0.459	0.129	0	0.0725	0.359
Soil	0.0122	0	0.0161	0	0.0442	0	0	0	0	0	0.164
Plants	0.1	0.161	0	0	0	0	0.0289	0.0314	0.112	0.162	0.0796
Bacteria	0	0	0	0	0.0327	0.0354	0	0.0302	0.275	0	-0.0744
Herbi/Fun givre nematodes	0	0.0422	0	0.0327	0	0.139	0	0.104	0.0896	0.00466	0
Bacterivor e nematodes	0	0	0	0.0354	0.139	0	0	0.0457	0.2	0.0353	0
Oribatida	0.459	0	0.0289	0	0	0	0	0.122	0	0	0.0181
Saprophyt ic fungi	0.129	0	0.0314	0.302	0.104	0.0457	0.122	0	0.292	0.321	0.0268
Symbiotic fungi	0	0	0.112	0.275	0.0896	0.2	0	0.292	0	0.254	0.0366
Pathogeni c fungi	0.0725	0	0.162	0	0.00466	0.0353	0	0.321	0.254	0	0
Springtail s	0.359	0.164	0.0796	-0.0744	0	0	0.0181	0.0268	0.0366	0	0

**Table S2** Partial correlation between the true turnover component of  $\beta$ -diversity values of trophic groups constituting of soil diversity and abiotic environmental distances inferred using Glasso.

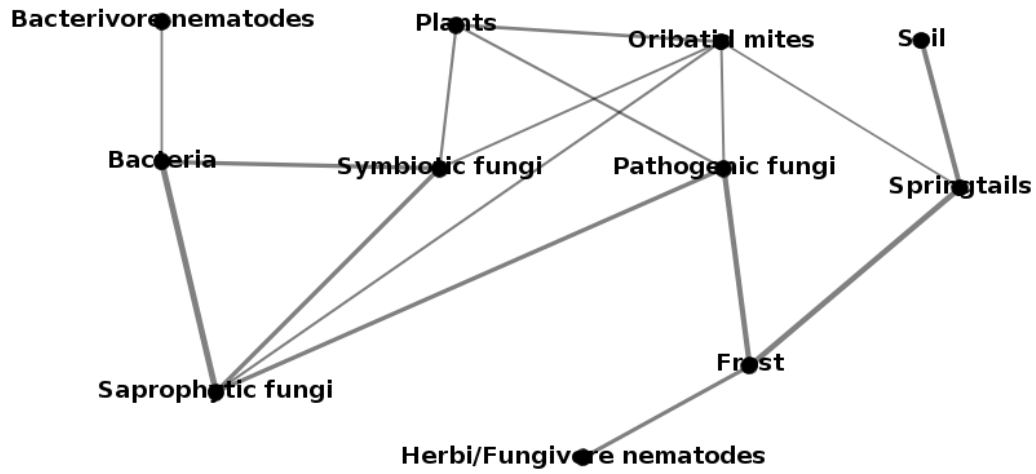
	Frost	Soil	Plants	Bacteria	Herbi/Fun givre Nematode s	Bacterivor e nematodes	Oribatida	Saprophyt ic fungi	Symbiotic fungi	Pathogeni c fungi	Springtails
Frost	1	0.305	0.503	0.313	0.102	0.117	0.734	0.619	0.471	0.551	0.637
Soil	0.305	1	0.363	0.0432	0.216	-0.0226	0.249	0.255	0.207	0.158	0.367
Plants	0.503	0.363	1	0.276	0.231	0.235	0.416	0.535	0.54	0.557	0.417
Bacteria	0.313	0.0432	0.276	1	0.395	0.439	0.346	0.743	0.732	0.593	0.0491
Herbi/Fun givre nematodes	0.102	0.216	0.231	0.395	1	0.398	0.0963	0.468	0.467	0.396	0.158
Bacterivor e nematodes	0.117	-0.0226	0.235	0.439	0.398	1	0.196	0.499	0.557	0.456	0.0921
Oribatida	0.734	0.249	0.416	0.346	0.0963	0.196	1	0.556	0.448	0.463	0.461
Saprophyt ic fungi	0.619	0.255	0.535	0.743	0.468	0.499	0.556	1	0.835	0.799	0.422
Symbiotic fungi	0.471	0.207	0.54	0.732	0.467	0.557	0.448	0.835	1	0.768	0.382
Pathogeni c fungi	0.551	0.158	0.557	0.593	0.396	0.456	0.463	0.799	0.768	1	0.331
Springtail s	0.637	0.367	0.417	0.0491	0.158	0.0921	0.461	0.422	0.382	0.331	1

**Table S3** Marginal (Pearson correlation) between the true turnover component of  $\beta$ -diversity values of trophic groups constituting of soil diversity and abiotic environmental distances.

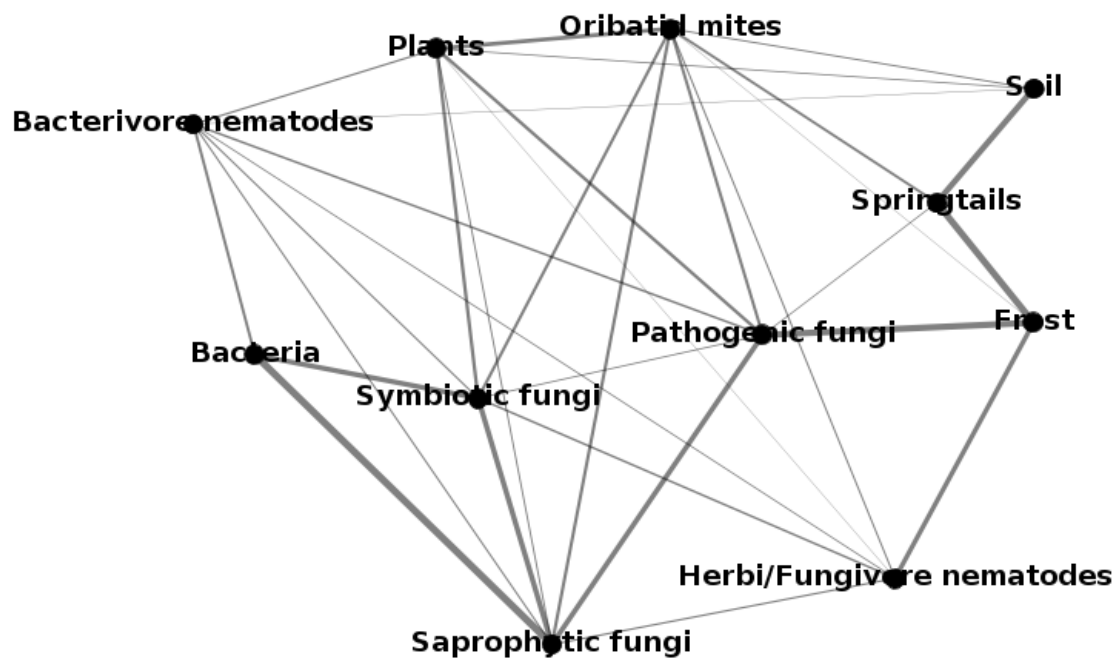
*Appendix 3 : Statistical analyses using the Sorensen dissimilarity index as  $\beta$ -diversity measure,*

*and Simpson dissimilarity index as true turnover measure. Comparison of the results using Jaccard and Sorensen dissimilarity index.*

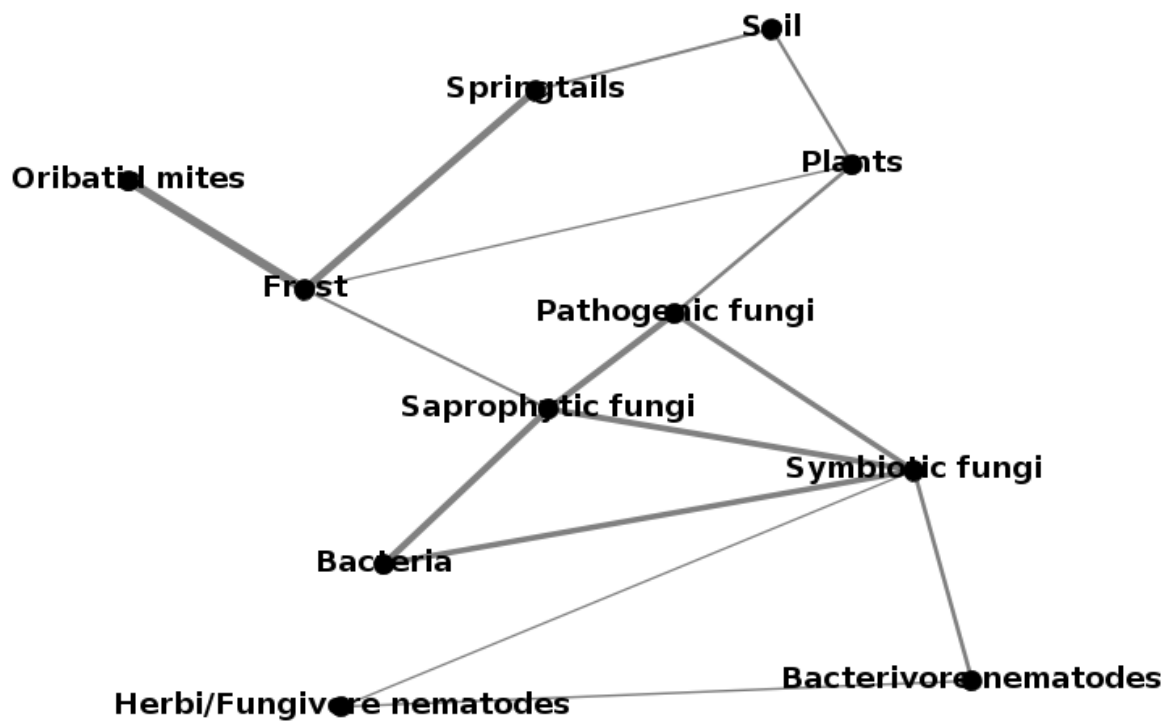
We ran the same statistical analyses than in the main text but using the Sorensen dissimilarity index (instead of Jaccard) as measure of  $\beta$ -diversity and the Simpson dissimilarity index as the true turnover component of the Sorensen index (Baselga 2010). We represented the inferred networks from  $\beta$ -diversity values computed from Sorensen index (Fig. S14, Fig. S15) as well as the networks built using Simpson dissimilarity index (Fig. S16, Fig. S17). Moreover, we compared the partial correlation networks associated to each dissimilarity index using the method described in Poisot et al. 2012. Briefly, the dissimilarity between two networks ( $\beta_{WN}$ ) is the Sorensen index on the set of edges of the two considered networks. If they have the same set of nodes, two networks are identical only if  $\beta_{WN}=0$ . They do not share any edges only if  $\beta_{WN}=1$ . We used the R package 'betalink'. The partial correlation networks built from Jaccard and Sorensen indices have a dissimilarity value equals to 0.061 (0.030 if only the edges which the weights are above the median value of partial correlation coefficients are considered). The partial correlation networks built from the true turnover part of the Jaccard and Sorensen indices (Simpson dissimilarity index) have a dissimilarity value equals to 0.031 (0.161 if only the edges which the weights are above the median value of partial correlation coefficients are considered).



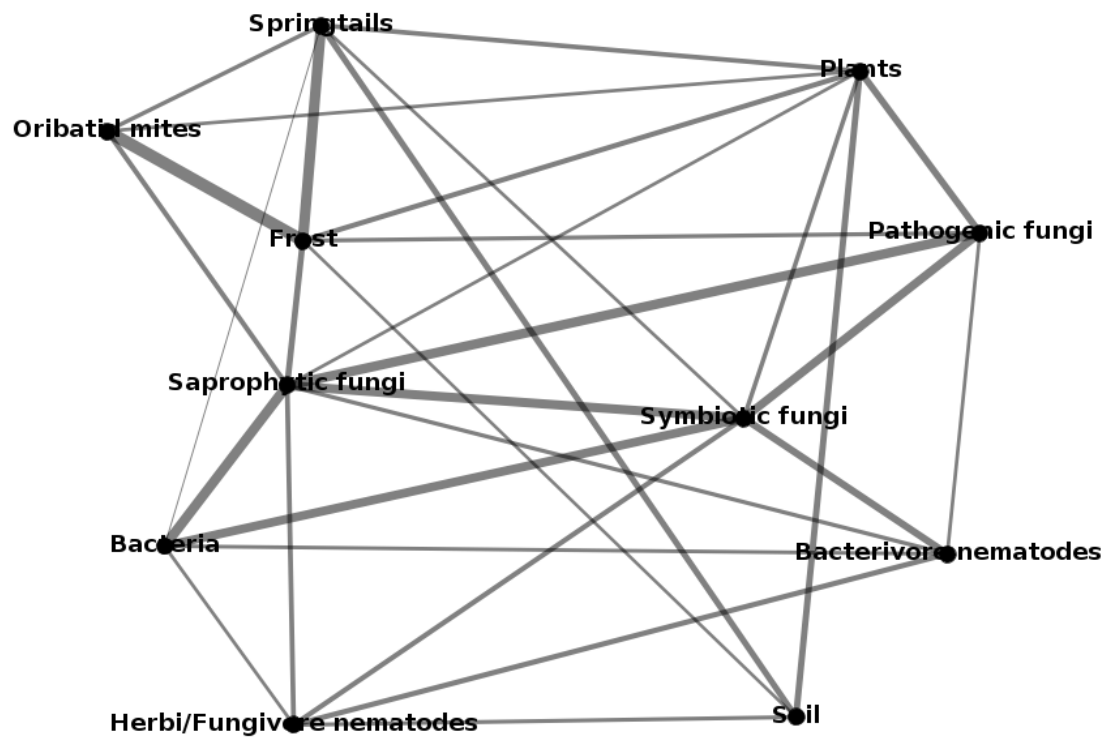
**Figure S8 Undirected partial correlation network inferred using the graphical lasso method between the  $\beta$ -diversity (measured using the Sorensen index) of the major trophic groups constituting soil biodiversity and the environmental distances.** Each node represents the  $\beta$ -diversity of a trophic group or an environmental distance. Here, only the partial correlations above the median value of the non-null partial correlation coefficients (0.116) are shown. The non-filtered network is presented in Figure S15. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.



**Fig. S9 Undirected partial correlation network between  $\beta$ -diversity (measured using the Sorensen index) of the major trophic groups constituting of soil biodiversity and environmental distances.** Each node represent the  $\beta$ -diversity of a trophic group or an abiotic environmental distance. Each edge represent a non-nul partial correlation between the variables associated to each node and was estimated using a graphical lasso. Consequently, if two nodes are disconnected, then the associated variables ( $\beta$ -diversity or environmental distances) are conditionally independent. The connectance of the network is 0.582.



**Figure S10** Undirected partial correlation network inferred by the graphical lasso method between the true turnover component of the  $\beta$ -diversity (measured using the Simpson dissimilarity index) of the major trophic groups constituting of soil biodiversity and the environmental distances. Each node represent the  $\beta$ -diversity of a trophic group or an environmental distance. Here, are only shown the partial correlations above the median value of the non-null partial correlation coefficients (0.101). The non-filtered network is presented in Figure S11. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.



**Figure S11** Undirected partial correlation network inferred by the graphical lasso method between the true turnover component of the  $\beta$ -diversity (measured using the Simpson dissimilarity index) of the major trophic groups constituting of soil biodiversity and the environmental distances. Each node represent the  $\beta$ -diversity of a trophic group or an environmental distance. The connectance of the graph is 0.563. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.

#### *Appendix 4 : Climatic and soil variables*

Climatic variables: We described the bioclimatic conditions of the gradient using four variables describing both thermal and nival regimes: soil temperature, solar radiations, Growing Season Length (hereafter “GSL”) and number of frost days. Mean annual soil temperature was recorded using HOBO stations located at the center of each sampling site. Solar radiation was averaged from simulations of solar radiation using a 2-m digital elevation model (DEM) (Carlson et al. 2015) derived from a LIDAR acquisition. The number of frost days, i.e. the number of snow-free days with air temperature below 0°C, was computed as described in Carlson et al. (2015). GSL, the number of snow-free days with a daily mean air temperature over 0°C between snow melt-out and August 15, was computed as described in Chalmandrier et al. 2016.

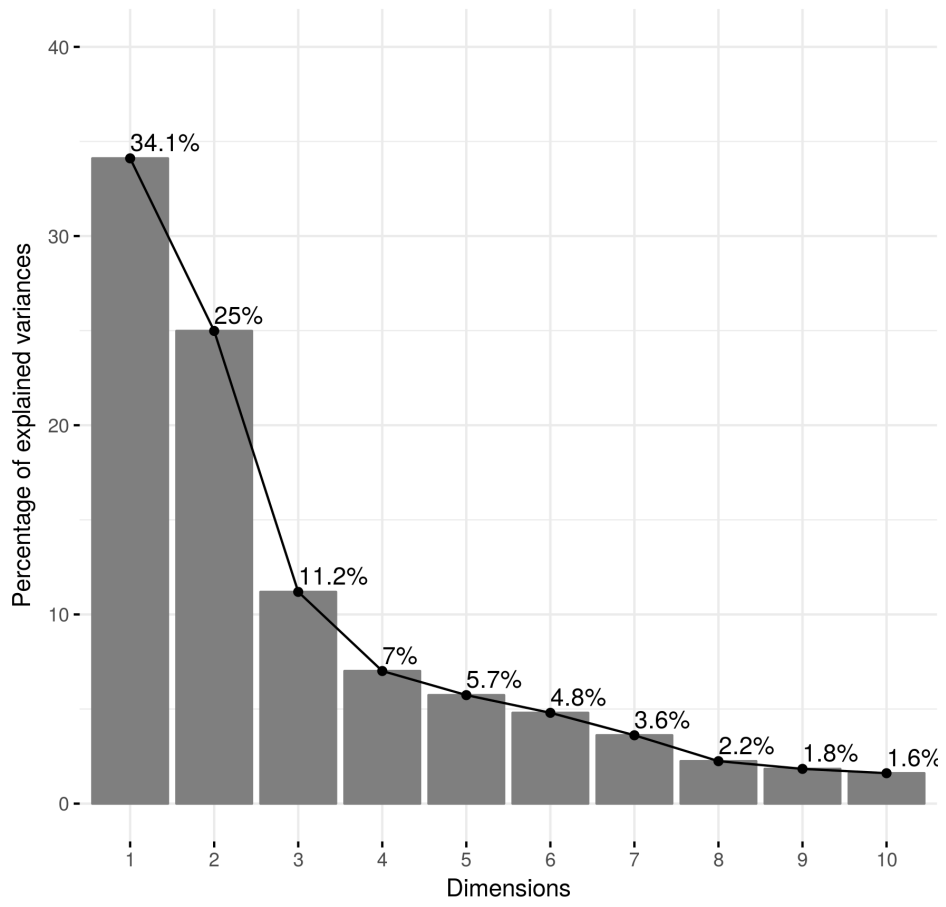
Topographical variables: Slope, Topographic Wetness Index (TWI) and Topographic Position Index (TPI) were inferred with the DEM derived from the airborne LiDAR data acquired the year of sampling.

Soil analyses were conducted in summer 2012 from soil cores taken at the center of each subplot (1m25x1m25). Each of them were weighed and 5mm-sieved (Legay et al. 2014) and processed for quantifying soil moisture, soil organic matter content, pH, soil texture (gravel mass and apparent density after 2mm sieving) and soil nutrients (ammonium (soil NH<sub>4</sub><sup>+</sup>), nitrates (soil NO<sub>3</sub><sup>-</sup>), total dissolved nitrogen (TDN), and dissolved organic nitrogen (DON)) from 0.5m K<sub>2</sub>SO<sub>4</sub> soil extracts (Jones et al. 2004; Legay et al. 2014). Soil subsamples were added 100 mL distilled water to saturate each soil core and calculate the waterfilled pore space (WFPS) and then air dried and ground to measure total soil C and N (Legay et al. 2014). Microbial biomass nitrogen (N) was measured using the chloroform-fumigation extraction technique of Vance et al. 1987). Potential N mineralization rates were estimated using anaerobic incubations of fresh soil subsamples (Legay et al. 2014). Potential ammonium and nitrates leaching was measured from the percolate of a site central supplementary core leached in a given volume of distilled water (Legay et al. 2016).

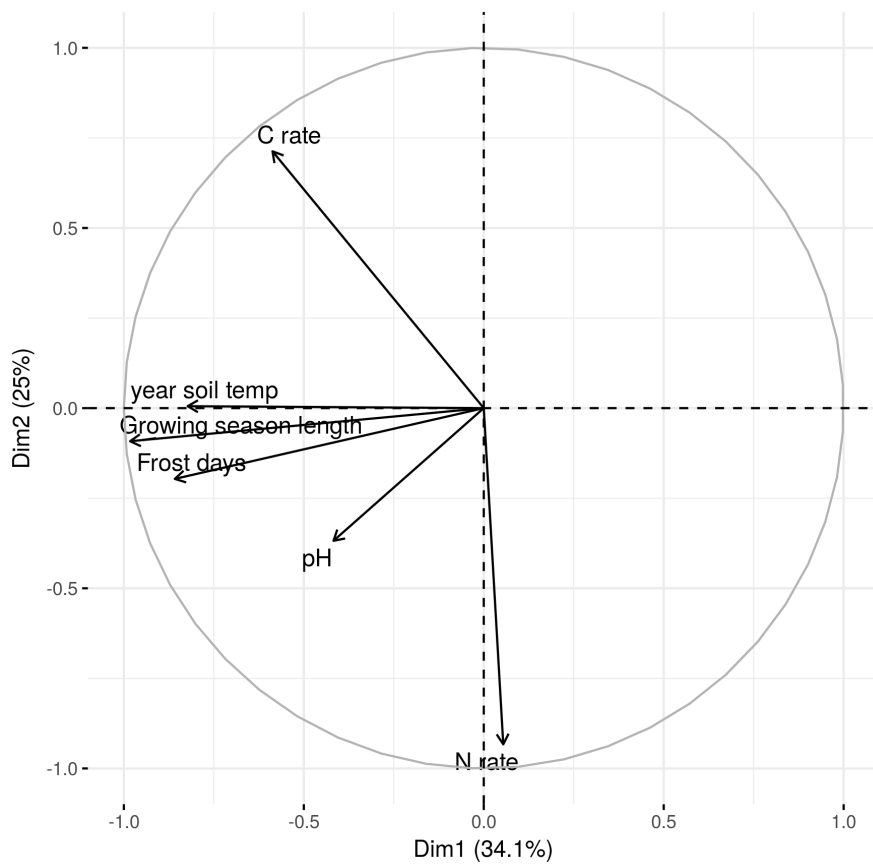


	PCA axis 1	PCA axis 2
pH	-0.417	-0.368
Gravel mass	0.791	-0.277
Apparent density	0.235	-0.781
Total porosity rate	0.264	0.789
Water Filled pore space	-0.76	-0.23
Leaching NO3	-0.398	-0.029
Leaching NH4	0.0901	0.791
N rate	0.0542	-0.933
C rate	-0.587	0.713
C/N ratio	-0.332	0.896
Rate organic matter	0.332	0.863
N NO3	-0.341	-0.368
N NH4	-0.512	0.264
N TDN	0.257	-0.106
N NH4 product	-0.158	-0.0632
Year soil temperature	-0.823	0.00538
FrostDays	-0.858	-0.196
Growing Season Length	-0.983	-0.0919
Topographic wetness index	-0.652	-0.0833
slope	0.28	-0.343
Topographic position indexPI10m_mean_topo	0.35	0.101
Altitude Digital Elevation Model	0.98	0.0867
Mean solar radiation	-0.964	-0.058
Variance Growing Season Length	0.834	0.347

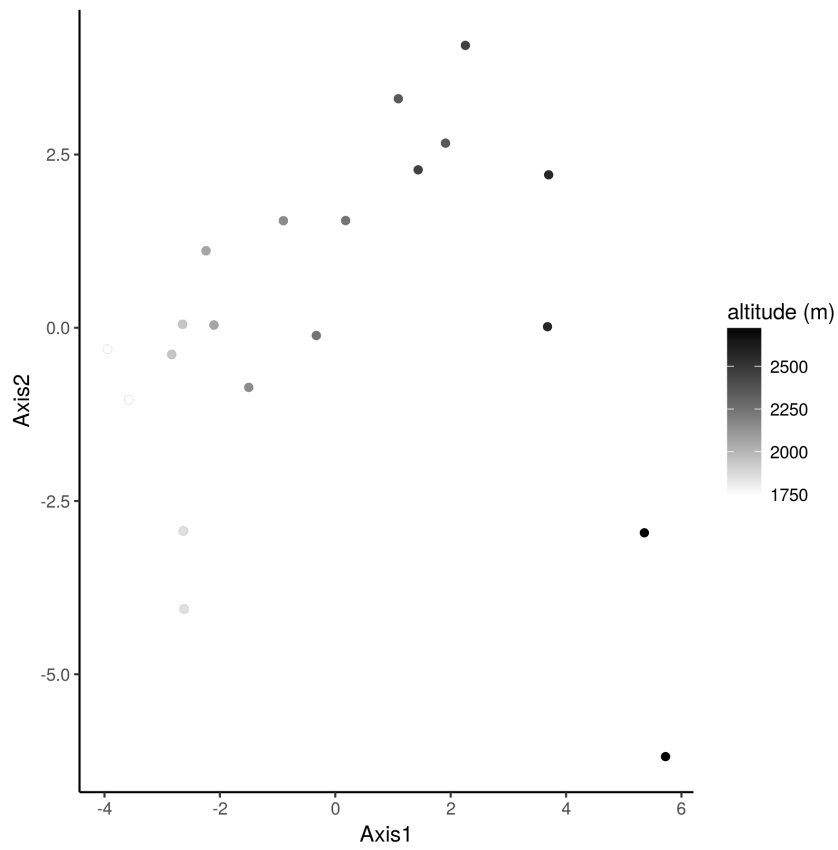
**Table S4** Correlations between the environmental variables and the first two axes of the Principal Component Analysis



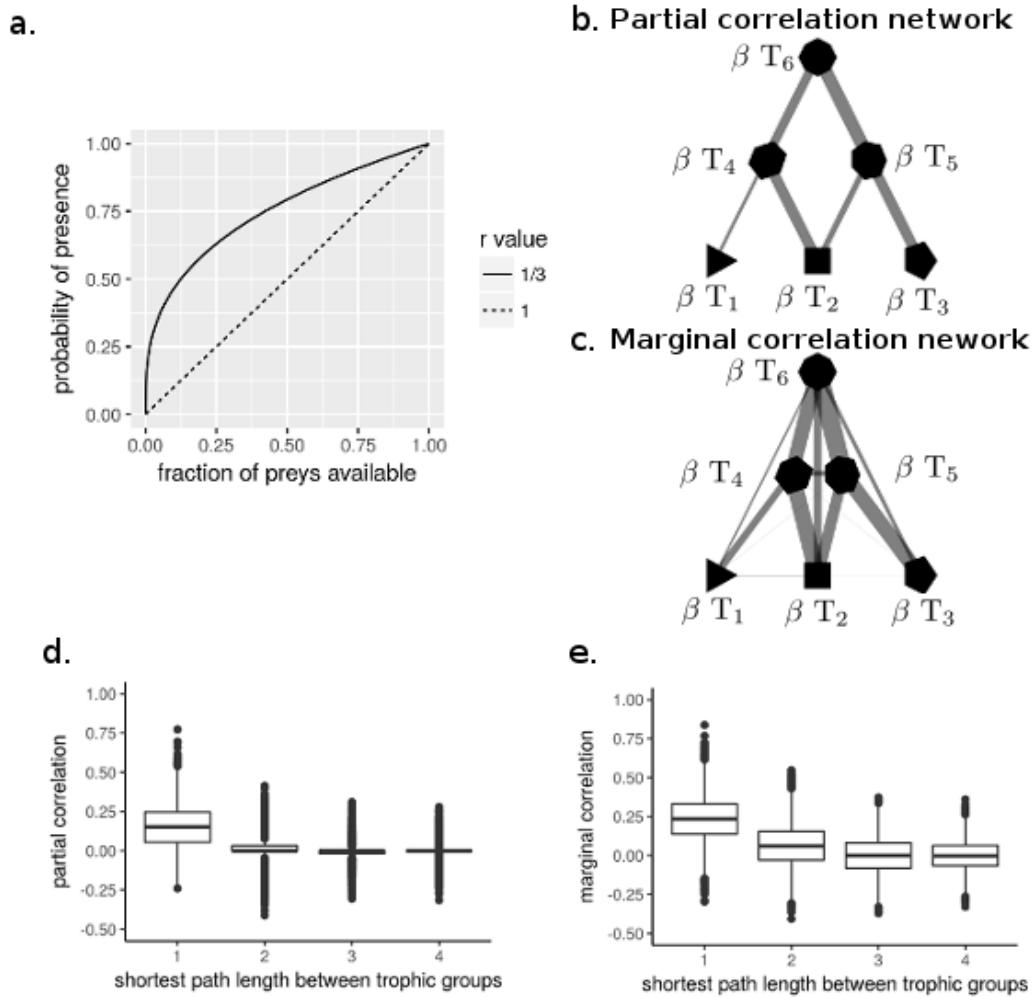
**Figure S11** Scree plot of eigenvalues on the principal component on environmental variables



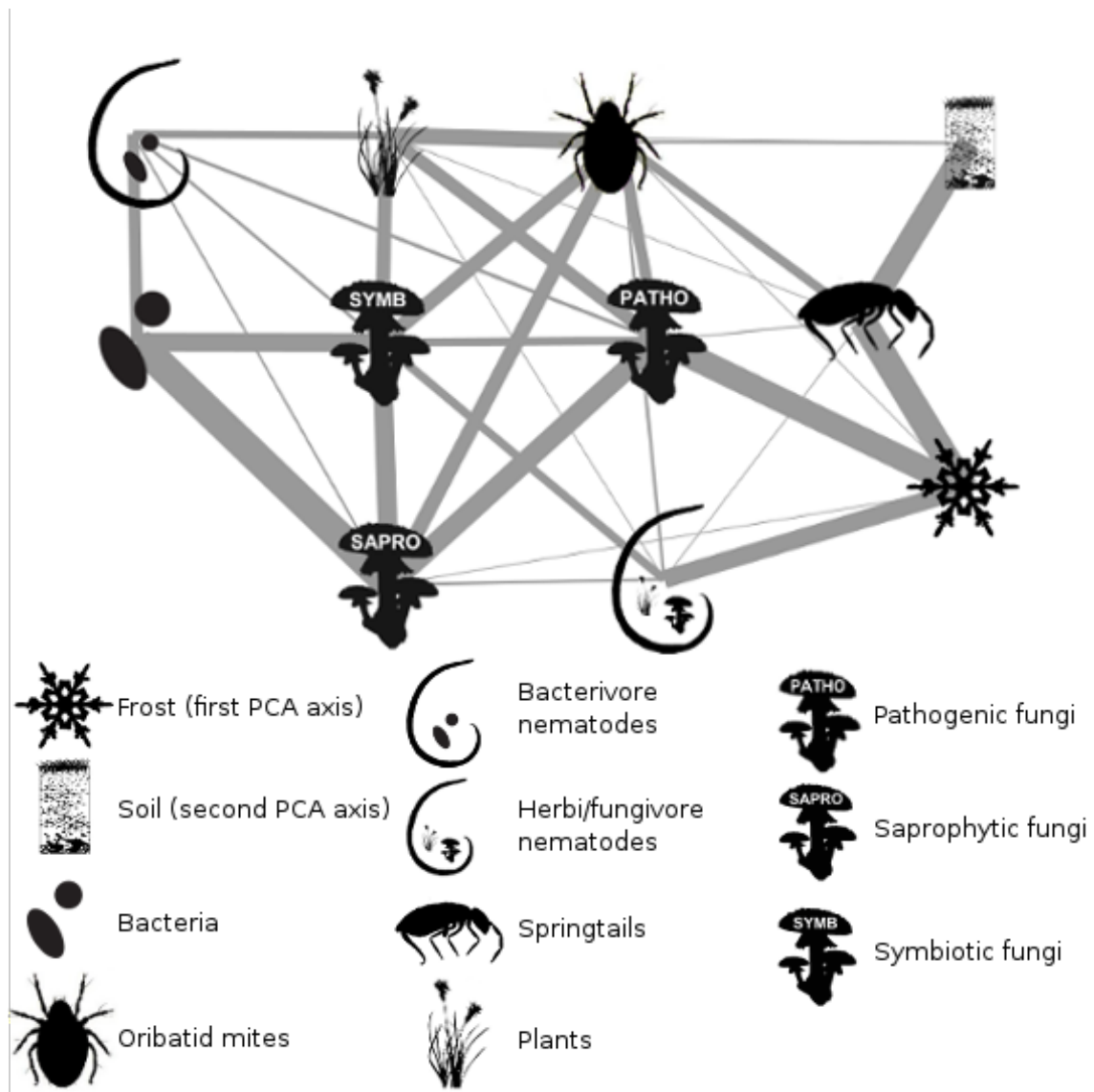
**Figure S13** Correlation circles of the principal component analysis. Only 6 variable over the 24 variables are represented here.



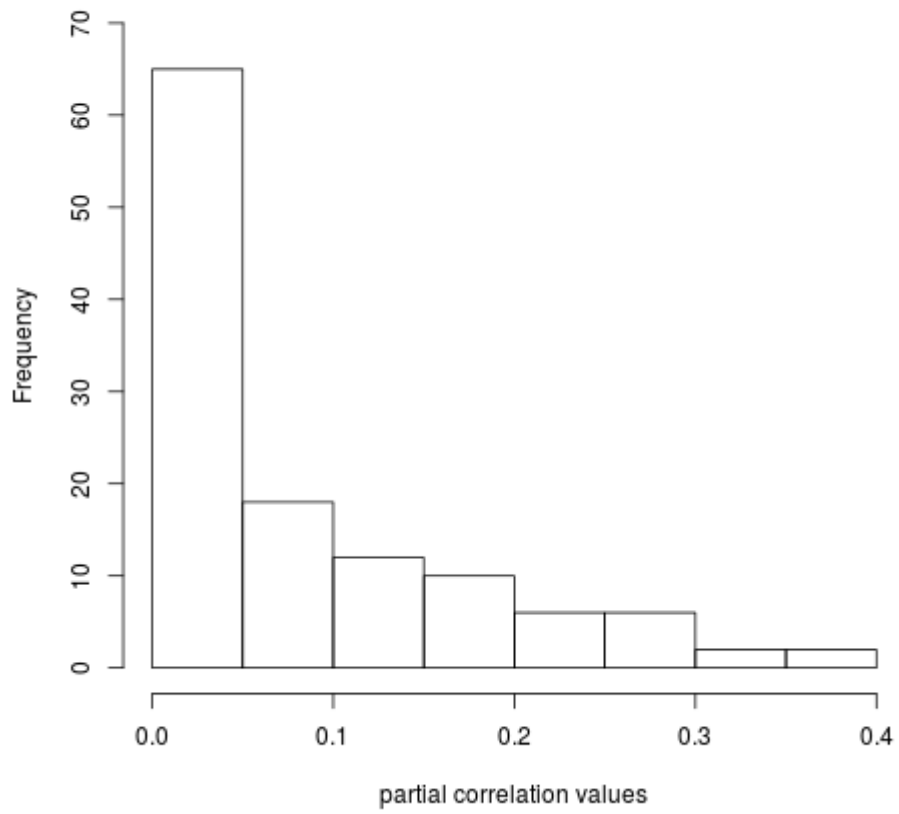
**Figure S14** Plots represented on the first factorial plan on the principal component analysis. The color of the points correspond to their altitude.



**Figure S15** Design and results of the simulation assessing the impact of biotic interactions on the marginal and partial correlation of the  $\beta$ -diversity values of different trophic groups for another shape of the relationship linking the probability of presence of a consumer species with its number of available preys. (a) Two different shapes of the relationship, the value used in the following plots is  $r=1/3$  (plain line), while the main text  $r=1$  (dashed line) (b) Partial correlation network built using the median values of partial correlations (inferred using Glasso) between the  $\beta$ -diversity values of different trophic groups. (c) Marginal correlation networks built using the median values of the Pearson correlations between the  $\beta$ -diversity values of different trophic groups. (d) Partial correlation coefficients (inferred using the graphical lasso) between the  $\beta$ -diversity values of two trophic levels in function of the shortest path length between these trophic groups. (e) Marginal Pearson correlation coefficients between the  $\beta$ -diversity values of two trophic groups in function of the shortest path length between these trophic groups.



**Figure S16** Undirected partial correlation network between  $\beta$ -diversity of the major trophic groups constituting of soil biodiversity and environmental distances. Each node represent the  $\beta$ -diversity of a trophic group or an abiotic environmental distance. Each edge represent a non-nul partial correlation between the variables associated to each node and was estimated using a graphical lasso. Consequently, if two nodes are disconnected, then the associated variables ( $\beta$ -diversity or environmental distances) are conditionally independent. The connectance of the network is 0.618.



**Figure S17** Histogram of the partial correlation coefficients estimated between  $\beta$ -diversity of major trophic groups constituting of soil biodiversity and environmental distances. Partial correlation coefficients are all positive.

	Frost	Soil	Plants	Bacteria	Herbi/Fun givre nematodes	Bacterivor e nematodes	Oribatida	Saprophyt ic fungi	Symbiotic fungi	Pathogeni c fungi	Springtails
Frost	0	0	0	0	0.222	0	0.0358	0.0149	0	0.3	0.327
Soil	0	0	0.0347	0	0	0	0.0807	0	0	0	0.29
Plants	0	0.0347	0	0	0.0274	0.101	0.197	0.0914	0.169	0.166	0.0284
Bacteria	0	0	0	0	0	0.149	0	0.365	0.23	0	0
Herbi/Fun givre nematodes	0.222	0	0.0274	0	0	0.0833	0.084	0.084	0.11	0	0.0238
Bacterivor e nematodes	0	0	0.101	0.149	0.0833	0	0	0.0708	0.0739	0.0994	0
Oribatida	0.0358	0.0807	0.197	0	0.084	0	0	0.185	0.182	0.14	0.116
Saprophyt ic fungi	0.0149	0	0.0914	0.365	0.084	0.0708	0.185	0	0.234	0.252	0
Symbiotic fungi	0	0	0.169	0.23	0.11	0.0739	0.182	0.234	0	0.102	0
Pathogeni c fungi	0.3	0	0.166	0	0	0.0994	0.14	0.252	0.102	0	0.0541
Springtail s	0.327	0.29	0.0284	0	0.0238	0	0.116	0	0	0.0541	0

**Table S5** Partial correlation between  $\beta$ -diversity values of trophic groups constituting of soil diversity and abiotic environmental distances inferred using Glasso.

	Frost	Soil	Plants	Bacteria	Herbi/Fun givre Nematode s	Bacterivor e nematodes	Oribatida	Saprophyt ic fungi	Symbiotic fungi	Pathogeni c fungi	Springtails
Frost	1	0.305	0.581	0.405	0.648	0.497	0.672	0.678	0.575	0.772	0.703
Soil	0.305	1	0.392	0.235	0.325	0.0994	0.443	0.33	0.27	0.347	0.545
Plants	0.581	0.392	1	0.644	0.585	0.611	0.775	0.778	0.78	0.774	0.546
Bacteria	0.405	0.235	0.644	1	0.47	0.642	0.688	0.857	0.808	0.687	0.34
Herbi/Fun givre nematodes	0.648	0.325	0.585	0.47	1	0.531	0.64	0.664	0.657	0.626	0.513
Bacterivor e nematodes	0.497	0.0994	0.611	0.642	0.531	1	0.573	0.675	0.656	0.645	0.346
Oribatida	0.672	0.443	0.775	0.688	0.64	0.573	1	0.836	0.818	0.808	0.622
Saprophyt ic fungi	0.678	0.33	0.778	0.857	0.664	0.675	0.836	1	0.888	0.864	0.553
Symbiotic fungi	0.575	0.27	0.78	0.808	0.657	0.656	0.818	0.888	1	0.81	0.476
Pathogeni c fungi	0.772	0.347	0.774	0.687	0.626	0.645	0.808	0.864	0.81	1	0.628
Springtail s	0.703	0.545	0.546	0.34	0.513	0.346	0.622	0.553	0.476	0.628	1

**Table S6** Marginal (Pearson correlation) between  $\beta$ -diversity values of trophic groups constituting of soil diversity and abiotic environmental distances.

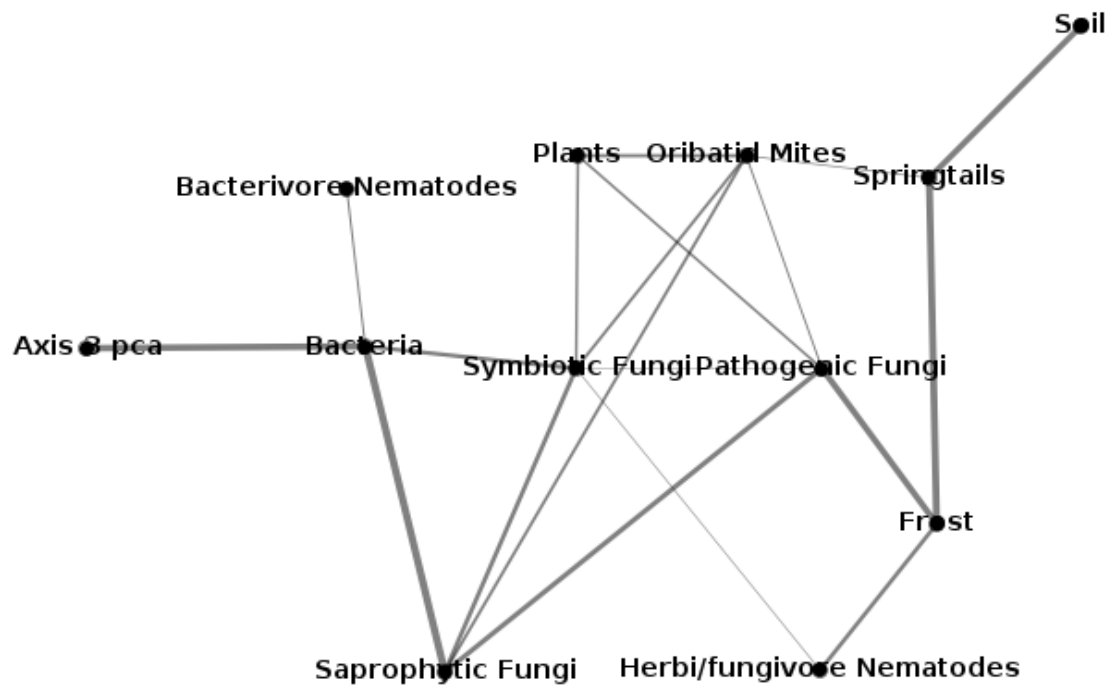


### *Appendix 5 : Assessing the sensitivity of the method to the addition of another abiotic predictor*

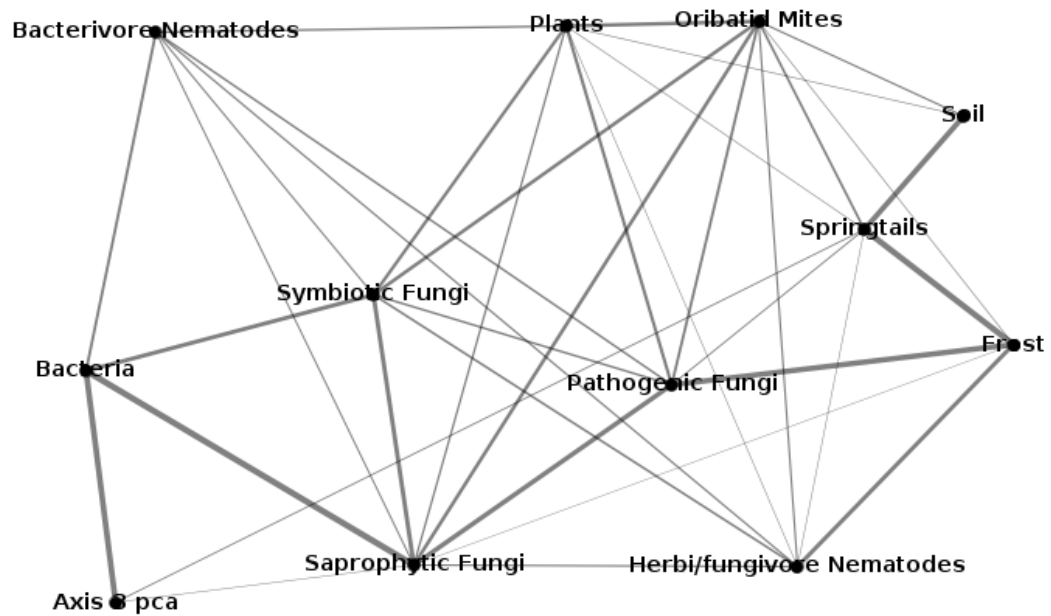
In order to assess the sensitivity of the method to the addition of new (initially missing) predictor, we ran the analysis using the eleven previously selected variables and added another abiotic distance, built from the third axis of the principal component analysis (it represents 11% of the total variance on environmental variables). We used the Jaccard index as measure of  $\beta$ -diversity. We present the inferred network (Fig. S18, Fig. S19) and the degree of each node (Fig. S20). Moreover, we compared the networks inferred with two or three environmental variables using Poiset's network dissimilarity (Poisot et al. 2012 ). Since the two networks do not share the same nodes, we used the additive partition of  $\beta_{WN}$  (total dissimilarity of edges) in  $\beta_{OS}$  (dissimilarity of edges due to turnover of edges in the shared parts of the two networks) and  $\beta_{ST}$  (dissimilarity of edges due to turnover of nodes).

The partial correlation networks built using two or three environmental variables have a  $\beta_{WN}$  dissimilarity value equals to 0.042 (0.056 if only the edges which the weights are above the median value of partial correlation coefficients are considered) and a  $\beta_{OS}$  dissimilarity value equals to 0 (0.029 if only the edges which the weights are above the median value of partial correlation coefficients are considered). This means the overall dissimilarity of edges is entirely due to the edges between the node representing the distance matrix built from the third axis of pca. In other words, if one removes the nodes 'pca3' (and the edges connected to this node), then the network is identical (omitting the weight on the edges) to the network inferred with two environmental variables.

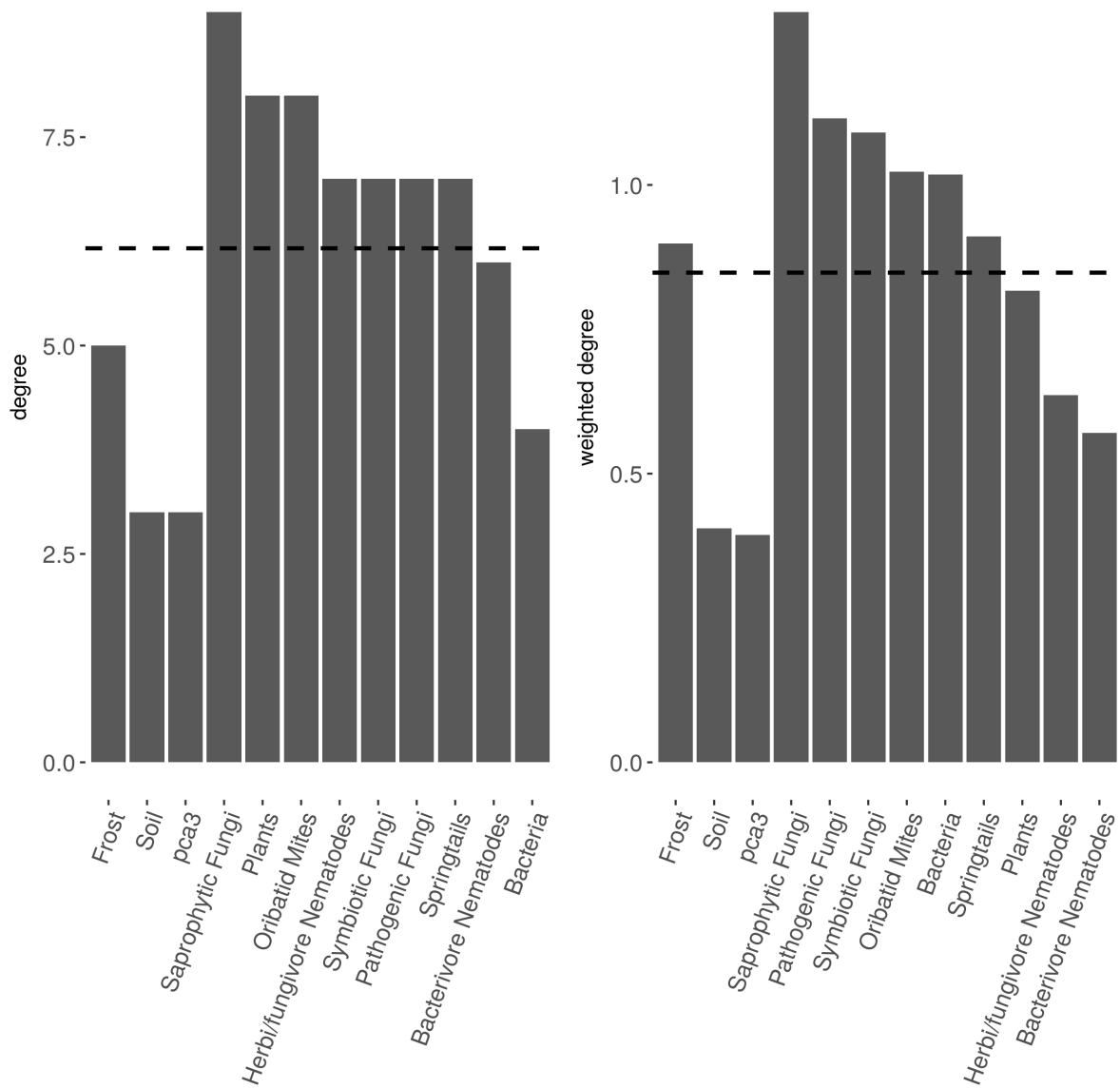
Adding an third axis of pca in the models does not change the edges between biotic groups. Moreover, even if the third axis of the principal component analysis represents a small part of total amount of variance on the environmental variables, the corresponding nodes connected to three biotic groups and presented a strong partial correlation with bacteria (0.313, mean value of non-nul partial correlation 0.138).



**Figure S18** Undirected partial correlation network inferred by the graphical lasso method between the  $\beta$ -diversity (measured using the Jaccard dissimilarity index) of the major trophic groups constituting of soil biodiversity and three environmental distances built from pca axis. Each node represent the  $\beta$ -diversity of a trophic group or an environmental distance. Here, are only shown the partial correlations above the median value of the non-null partial correlation coefficients (0.102). The non-filtered network is presented in Figure S19. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.



**Figure S19** Undirected partial correlation network inferred by the graphical lasso method between the  $\beta$ -diversity (measured using the Jaccard dissimilarity index) of the major trophic groups constituting of soil biodiversity and three environmental distances built from pca axis. Each node represent the  $\beta$ -diversity of a trophic group or an environmental distance. Edge thickness is proportional to the value of the partial correlation coefficient and the partial correlation coefficients are all positive.



**Figure S20 Properties of the inferred network.** The mean degree (left panel) is the number of neighbors of nodes in a graph (here, the undirected partial correlation network). It measures the number of variables that are conditionally dependent to the variable associated to this node. The weighted degree (right panel) is the sum of the partial correlation coefficients attached to the edges adjacent this node.

## References

- Abarenkov, K., Henrik Nilsson, R., Larsson, K.-H., Alexander, I. J., Eberhardt, U., Erland, S. et al. (2010). The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytologist*, *186*, 281–285.
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182.
- Mercier, C., Boyer, F., Bonin, A., & Coissac, E. (2013). SUMATRA and SUMACLUSt: fast and exact comparison and clustering of sequences. *Programs and Abstracts of the SeqBio 2013 Workshop*, 27-29.
- Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecology Letters*, *15*, 1353– 1361.
- Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 1118–23.
- Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M. et al. (2017). Soil community assembly varies across body sizes in a tropical forest. *bioRxiv* 154278; doi: <https://doi.org/10.1101/154278>

# Mélange de marqueurs pour l'ADN environnemental

---

Parfois la quantité fait office de qualité.

---

Iossif Vissarionovitch Djougachvili

## 4.1 Synopsis

L'ADN environnemental permet d'obtenir l'abondance de séquences d'ADN dégradées dans un milieu, assignées grâce à une base de données de référence à des taxons. Différents marqueurs permettent d'amplifier des régions différentes de l'ADN présents chez les organismes visés. Ils fournissent alors des données d'abondances différentes pour des communautés différentes se chevauchant néanmoins. Ainsi, certains taxons vont être vus par plusieurs marqueurs mais avec des comptages différents. On propose ici une méthode statistique pour évaluer la variabilité des comptages d'un même taxon d'un marqueur à l'autre et pour corriger les comptages du biais lié au marqueur. On montrera l'efficacité de la méthode proposée sur un jeu de données de plantes. De plus, on s'intéressera à l'impact sur la diversité de Shannon de l'incertitude sur les abondances. Ce travail a été initié en collaboration avec Stéphane Robin dans le cadre d'une visite au laboratoire UMR MIA-Paris AgroParisTech / INRA / Paris-Saclay permise grâce à une bourse du GdR Ecologie Statistique. Ce chapitre est encore en cours de développement mais nous présenterons ici quelques résultats préliminaires.

## 4.2 Analyses préliminaires

### 4.2.1 Introduction

L'avènement de l'ADN environmental permet d'obtenir des données d'abondances sur un large spectre du vivant (TABERLET et al., 2012). Il permet en particulier la quantification de l'abondance de micro-organismes en différents points de l'espace, donnée impossible à obtenir avec une autre technologie. Il permet également d'obtenir des données de co-occurrences sur des organismes ayant des échelles de taille différentes. Ainsi,

la mise en regard de patrons écologiques pour différents groupes d'organismes a pu être documentée et cette technologie peut permettre de suivre la biodiversité (BOHMANN et al., 2014). De même, en exploitant les données de co-occurrence, on peut alors obtenir des liens statistiques entre les présences ou les abondances de différents groupes (COINCE et al., 2014, OHLMANN et al., 2018, MARTINEZ-ALMOYNA et al., 2019). Ces relations statistiques entre distribution spatiales d'organismes ouvrent la voie à la reconstruction de réseaux d'interaction entre espèces de différents milieux (BOHAN et al., 2017).

Si des méthodes d'inférence de réseau à partir de données de comptage existent (CHIQUET, MARIADASSOU et ROBIN, 2018), cela requiert toutefois d'obtenir des données d'abondance d'ADN environnemental représentatives de l'abondance des taxons dans les échantillons. Or, un tel objectif est l'objet d'actives recherches (YATES, FRASER et DERRY, 2019). D'autant plus que, du fait de l'existence de multiples marqueurs (TABERLET et al., 2018), certains taxons sont vus avec différents marqueurs dans des abondances qui ne sont pas nécessairement en accord. Ainsi, d'un échantillon à l'autre, on peut distinguer deux sources de variabilité dans l'abondance d'un taxon : une variabilité biologique, liée au fait que l'abondance d'un taxon varie pour des raisons biologiques d'un site à l'autre et une variabilité technologique, liée au fait que l'accès à l'information d'abondance se fait par l'intermédiaire d'une technologie qui présente des biais.

Nous proposons ici une approche statistique utilisant un modèle Poisson-log-Normal pour estimer les variabilités biologiques et technologiques de l'abondance d'un taxon d'un échantillon à l'autre. De plus, cette méthode permet de corriger le biais lié au marqueur utilisé. On montrera l'efficacité de la méthode proposée sur un jeu de données de plantes diluées dans des proportions connues et amplifiées grâce à un marqueur plante et un marqueur eucaryote. De plus, on s'intéressera à l'impact sur la diversité de Shannon de l'incertitude sur les abondances.

## 4.2.2 Notations

Dans le cas général, on s'intéresse à l'abondance de Molecular Operational Taxonomy Units (MOTUs), un ensemble de séquences similaires assignées à un chemin dans la taxonomie  $\mathcal{T}$  amplifiées avec un marqueur  $p \in \{1, \dots, P\}$ . Soit  $\phi$  la fonction qui assigne un MOTU à un chemin de la taxonomie. Soit  $S_p$  l'ensemble des MOTUs qui ont été amplifiés en utilisant le marqueur  $p$ . On note  $V_p = \phi[S_p]$  l'image de  $S_p$  par  $\phi$ , *i.e.* l'ensemble des chemins taxonomiques auxquels ont été assignés des MOTUs de  $S_p$ . Plusieurs MOTUs peuvent avoir été assignés au même chemin taxonomique. De façon similaire, on définit  $V_{i,p}$ , l'ensemble des taxons auxquels des MOTUs amplifiés dans l'échantillon  $i$  avec le marqueur  $p$  ont été assignés.

Soit  $N_{i,j,p}$  l'abondance du MOTU  $j$  amplifié avec le marqueur  $p$  dans l'échantillon  $i$ . De

plus, soit  $N_{i,v,p}$  l'abondance du taxon  $v$  dans l'échantillon  $i$  vu avec le marqueur  $p$ , *i.e.*

$$N_{i,v,p} = \sum_{j' \in \phi^{-1}(\{v\})} N_{i,j',p} \quad (4.1)$$

La profondeur d'échantillonnage du marqueur  $p$  dans l'échantillon  $i$  est :

$$N_{i,+,p} = \sum_{v \in V_p} N_{i,v,p} = \sum_{j \in S_p} N_{i,j,p} \quad (4.2)$$

### 4.2.3 Modèle statistique

On s'intéresse alors d'une part à l'estimation de la variabilité dans l'abondance des taxons liée à la technologie d'observation (ici le marqueur) devant la variabilité intrinsèque des taxons d'un échantillon à l'autre et, d'autre part, à la correction du biais dans l'estimation des abondances introduits par le marqueur. On utilise alors un modèle Poisson-Log Normal pour modéliser les comptages :

$$N_{i,v,p} \sim \mathcal{P}(\exp(\mu_{i,p} + N_{i,v} + G_{v,p})) \quad (4.3)$$

où :  $\mu_{i,p} = \log(N_{i,+,p})$  est l'offset du modèle (terme fixe).

où :  $N_{i,v} \sim \mathcal{N}(0, \sigma^2)$

$N_{i,v}$  représente la variabilité biologique de l'abondance d'un taxon d'un échantillon à l'autre. et :  $G_{v,p} \sim \mathcal{N}(0, \theta^2)$ .

$G_{v,p}$  représentant la variabilité technologique induisant un biais dans l'abondance d'un taxon d'un marqueur à l'autre. On peut estimer les paramètres du modèle défini par l'équation (4.3) en utilisant le paquet R 'lme4'.

On peut également estimer l'abondance des taxons à travers les marqueurs en faisant une régression de Poisson (on considère alors les  $n_{i,v}$  et  $g_{v,p}$  comme des effets fixes) et en calculant des moyennes ajustées sur le facteur marqueur. Pour cela, on utilise le paquet R 'emmeans'.

### 4.2.4 Le jeu de données de plantes

Afin de tester la méthode proposée, on s'intéresse à un jeu de données dans lequel l'ADN de 16 espèces de plantes (appartenant à 14 familles distinctes) a été dilué en des proportions décroissant de façon exponentielle (Fig. 4.1). Le jeu de données est présenté dans l'annexe 5 de ZINGER et al., 2019. Les séquences ont été amplifiées dans 192 échantillons grâce aux marqueurs 'Sper01' (élaboré pour les spermaphytes) et 'Euka02' (élaboré pour les eucaryotes). Afin d'évaluer la qualité de la prédiction, nous avons calculé la corrélation de Spearman entre les abondances prédites par la régression et les abondances du mélange initial.



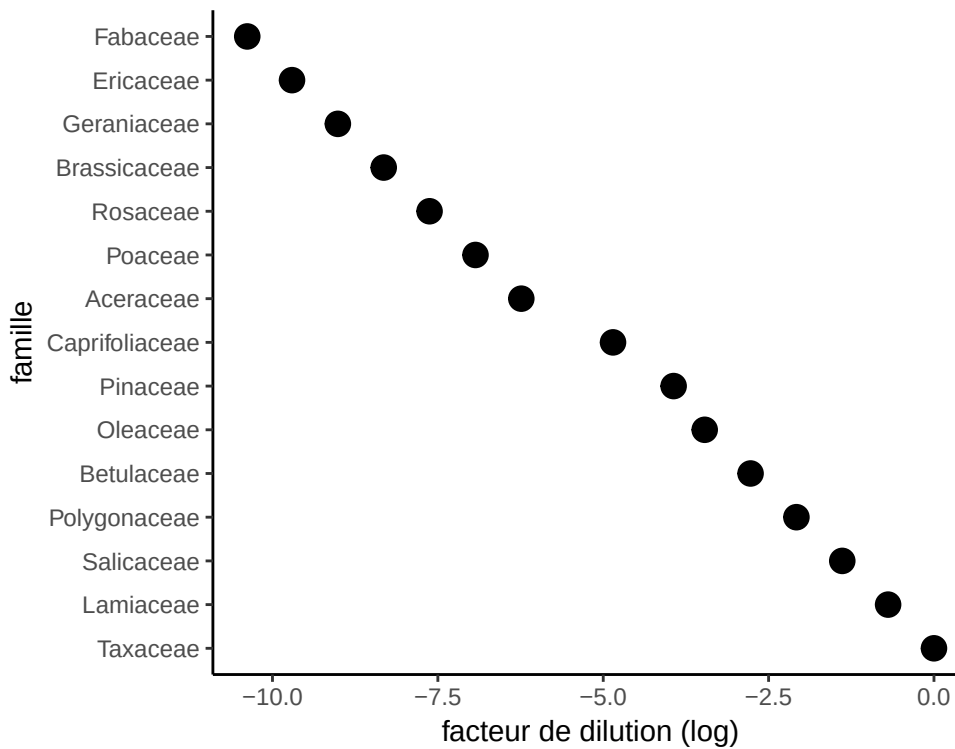


FIGURE 4.1 – Dilution de l’ADN des 14 familles introduites initialement dans les échantillons.

#### 4.2.5 Résultat sur les données de plantes

Dans le jeu de données, 13 des 14 familles initialement présentes ont été détectées avec le marqueur ‘Sper01’ alors que seulement 6 familles ont été détectées avec le marqueur ‘Euka02’. On restreint donc l’analyse aux 6 familles présentes dans les échantillons qui sont à l’intersection des deux marqueurs (*Taxaceae*, *Ericaceae*, *Polygonaceae*, *Brassicaceae*, *Geraniaceae*, *Fabaceae*). On obtient  $\sigma = 0.408$  et  $\theta = 3.98$ . Ainsi, on observe une plus grande variabilité technologique (variabilité de l’abondance des taxons vu avec différents marqueurs) que biologique (variabilité de l’abondance des taxons d’un échantillon à l’autre). La moyenne, à travers tous les échantillons, des corrélations de Spearman entre les abondances initiales des 6 familles et les abondances vues par le marqueur ‘Sper01’ est de 0.693 et de 0.611 pour le marqueur ‘Euka02’. Après avoir fait une régression et une moyenne ajustée à travers les deux marqueurs, la moyenne des corrélations de Spearman est de 0.803.

#### 4.2.6 Incertitude sur les abondances et estimation de la diversité

On s’intéresse ici à l’impact de l’incertitude des abondances sur l’estimation de la diversité. On se place dans un site avec des abondances (non relatives, des entiers) de  $n$  taxons données  $\nu = (\nu_v)_{1 \leq v \leq n}$  (on suppose que  $\forall v, \nu_v > 0$ ). On suppose alors que les abondances sont modulées de façon aléatoires ; cela peut représenter le bruit introduit par

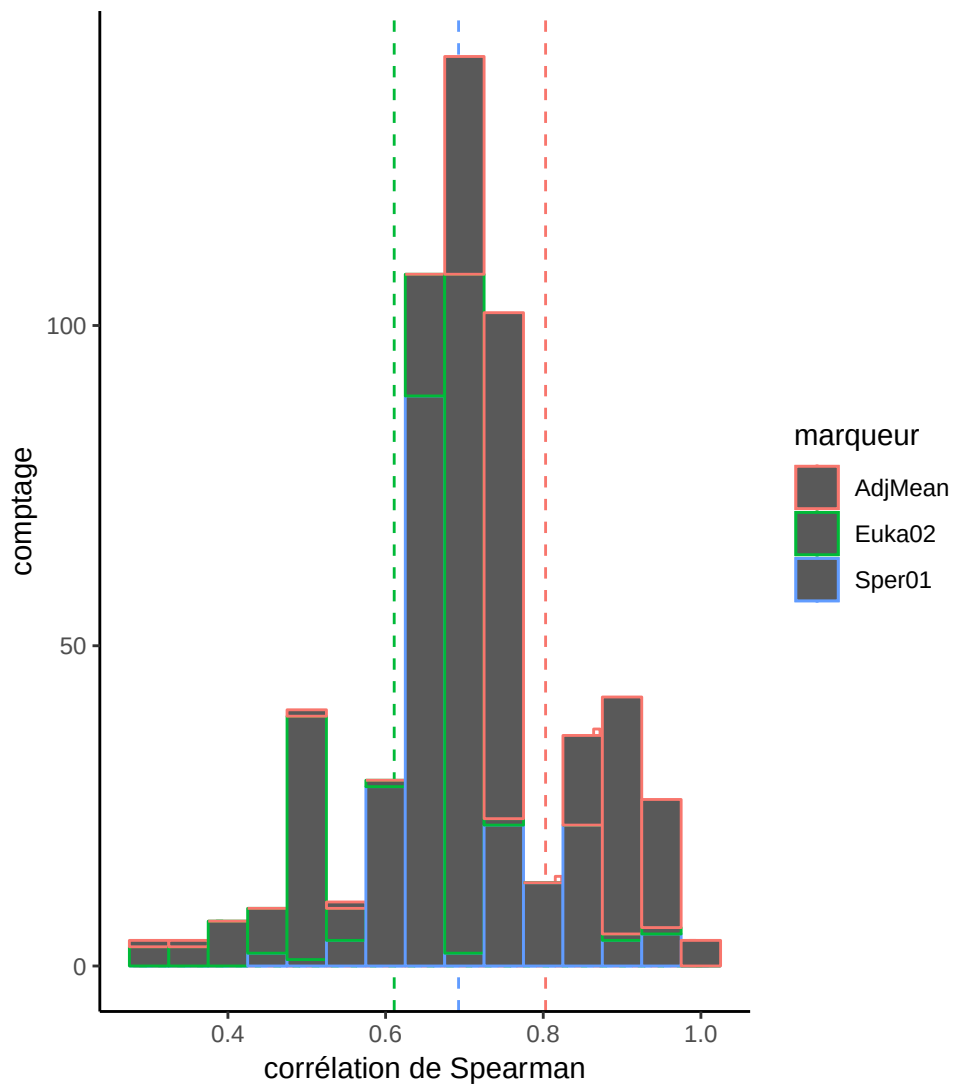


FIGURE 4.2 – Histogramme représentant la distribution des corrélations de Spearman entre l'abondance initiale des 6 familles de plantes et : (en bleu) l'abondance de familles de plantes amplifiées avec le marqueur 'Sper01', (en vert) l'abondance de familles de plantes amplifiées avec le marqueur 'Euka02', (en rose) les abondances prédites avec les moyennes ajustées. Les traits en pointillés correspondent aux moyennes de chacune des distributions.

les différentes étapes de traitement des données d'ADN environnemental.

Ainsi, on suppose que l'abondance de chaque taxon est une variable aléatoire,  $N_v$  :

$$N_v = \nu_v \Gamma_v \quad (4.4)$$

où  $\Gamma_v \sim \Gamma(\frac{1}{\theta^2}, \theta^2)$ . On suppose que les  $(\Gamma_v)_v$  sont tirés indépendamment et identiquement distribués.

On a ainsi  $\mathbb{E}(\Gamma_v) = 1$  et  $\mathbb{V}(\Gamma_v) = \theta^2$ . D'où  $\mathbb{E}(N_v) = \nu_v$  et  $\mathbb{V}(N_v) = \nu_v^2 \theta^2$ .

On s'intéresse à la diversité de Shannon des  $(N_v)_v$ , ramenées à des abondances relatives.

Ainsi, on considère la fonction de normalisation :

$$\pi : \begin{array}{l} \mathbb{R} \longrightarrow [0, 1] \\ n_v \longmapsto \frac{n_v}{\sum_v n_v} \end{array} \quad (4.5)$$

Et la diversité de Shannon :

$$D : \begin{array}{l} [0, 1]^n \longrightarrow [0, n] \\ \mathbf{p} \longmapsto \exp(\sum_{v=1}^n -p_v \log p_v) \end{array} \quad (4.6)$$

On utilise alors une méthode delta à l'ordre 2 pour approximer  $\mathbb{E}(D((\pi(N_v))_v))$  (OEHLERT, 1992).

$$D((\pi(N_v))_v) \approx D((\pi(\nu_v))_v) + \sum_v (N_v - \nu_v) \frac{\partial D}{\partial \nu_v}((\pi(\nu_v))_v) + \sum_{u,v} (N_u - \nu_u)(N_v - \nu_v) \frac{\partial^2 D}{\partial \nu_v \partial \nu_u}((\pi(\nu_v))_v)$$

$$\begin{aligned} \mathbb{E}(D((\pi(N_v))_v)) &\approx D((\pi(\nu_v))_v) + \sum_v (\mathbb{E}(N_v) - \nu_v) \frac{\partial D}{\partial \nu_v}((\pi(\nu_v))_v) + \sum_{u \neq v} \text{cov}(N_u, N_v) \frac{\partial^2 D}{\partial \nu_v \partial \nu_u}((\pi(\nu_v))_v) \\ &\quad + \sum_v \mathbb{V}(N_v) \frac{\partial^2 D}{\partial \nu_v^2}((\pi(\nu_v))_v) \end{aligned}$$

Or  $\mathbb{E}(N_v) = \nu_v$ ,  $\text{cov}(N_v, N_u) = 0$  et  $\mathbb{V}(N_v) = \nu_v^2 \theta^2$ .

D'où :

$$\mathbb{E}(D((\pi(N_v))_v)) \approx D((\pi(\nu_v))_v) + \sum_v \nu_v^2 \theta^2 \frac{\partial^2 D}{\partial \nu_v^2}((\pi(\nu_v))_v) \quad (4.7)$$

On va maintenant montrer que :  $\forall v, \frac{\partial^2 D}{\partial \nu_v^2}((\pi(\nu_v))_v) < 0$  à condition que  $\forall v, \pi(\nu_v) \leq \frac{1}{e} \approx$

0.37 On note  $\pi_v = \pi(\nu_v)$  et on a :

$$\frac{\partial^2 D}{\partial \nu_v^2}((\pi(\nu_v))_v) = \frac{\partial^2 \pi}{\partial \nu_v^2}(\nu_v) \frac{\partial D}{\partial \pi_v}((\pi_v)_v) + \left(\frac{\partial \pi}{\partial \nu_v}(\nu_v)\right)^2 \frac{\partial^2 D}{\partial \pi_v^2}((\pi(\nu_v))_v)$$

En notant  $\nu_+ = \sum_v \nu_v$ , on a :

$$\frac{\partial \pi}{\partial \nu_v}(\nu_v) = \frac{\nu_+ - \nu_v}{\nu_+^2}$$

$$\frac{\partial^2 \pi}{\partial \nu_v^2}(\nu_v) = -\frac{1}{\nu_+^2} - \frac{\nu_+^2 - 2\nu_v \nu_+}{\nu_+^4}$$

$$\frac{\partial^2 \pi}{\partial \nu_v^2}(\nu_v) = \frac{-2\nu_+^2 - 2\nu_v \nu_+}{\nu_+^4}$$

Donc :

$$\frac{\partial^2 \pi}{\partial \nu_v^2}(\nu_v) < 0$$

De plus :

$$\frac{\partial D}{\partial \pi_v}((\pi_v)_v) = -D((\pi_v)_v)(\log(\pi_v) + 1)$$

On a  $\frac{\partial D}{\partial \pi_v}((\pi_v)_v) \geq 0$  pour  $\pi_v \leq \frac{1}{e}$

$$\begin{aligned} \frac{\partial^2 D}{\partial \pi_v^2}((\pi_v)_v) &= -\frac{\partial D}{\partial \pi_v}((\pi_v)_v)(\log(\pi_v) + 1) - D((\pi_v)_v) \frac{1}{\pi_v} \\ \frac{\partial^2 D}{\partial \pi_v^2}((\pi_v)_v) &= D((\pi_v)_v) \left( (\log(\pi_v) + 1)^2 - \frac{1}{\pi_v} \right) \end{aligned}$$

On étudie alors la fonction :

$$f : \begin{array}{l} ]0, 1] \longrightarrow \mathbb{R} \\ \pi_v \longmapsto (\log(\pi_v) + 1)^2 - \frac{1}{\pi_v} \end{array}$$

On a :

$$f'(\pi_v) = \frac{2\pi_v \log(\pi_v) + 2\pi_v + 1}{\pi_v^2}$$

Or :

$$\begin{aligned} \forall \pi_v \in ]0, 1], \pi_v \log(\pi_v) &\geq -\frac{1}{e} \\ 2\pi_v \log(\pi_v) + 1 &\geq 1 - \frac{2}{e} > 0 \\ f'(\pi_v) &> 0 \end{aligned}$$

Donc  $f$  est croissante et  $f(1) = 0$ . D'où, pour  $\pi_v \in ]0, 1]$ ,

$$(\log(\pi_v) + 1)^2 - \frac{1}{\pi_v} \leq 0$$

On a finalement :

$$\frac{\partial^2 D}{\partial \nu_v^2}((\pi(\nu_v))_v) = \underbrace{\frac{\partial^2 \pi}{\partial \nu_v^2}(\nu_v)}_{<0} \underbrace{\frac{\partial D}{\partial \pi_v}((\pi_v)_v)}_{\geq 0} + \underbrace{\left(\frac{\partial \pi}{\partial \nu_v}(\nu_v)\right)^2}_{\geq 0} \underbrace{\frac{\partial^2 D}{\partial \pi_v^2}((\pi(\nu_v))_v)}_{\leq 0}$$

Ainsi, à condition qu'aucun taxon ne représente plus de 37% des abondances, on a :

$$\mathbb{E}(D((\pi(N_v))_v)) \leq D((\pi(\nu_v))_v) \quad (4.8)$$

Introduire un biais dans les abondances diminue donc la diversité attendue.

# Discussion

---

## 5.1 Conclusions des différents chapitres

Dans cette thèse, nous nous sommes intéressés à différents éléments permettant d'élaborer les prémisses d'une théorie spatiale des réseaux d'interactions en écologie. Dans la plus grande partie de cette thèse, nous avons représenté une communauté non pas par une collection d'espèces (avec des abondances) mais par une collection d'espèces (avec des abondances) et leurs interactions (les arêtes du réseau d'interaction qui peuvent être pondérées ou valuées).

### 5.1.1 Chapitre 1

Dans le premier chapitre, nous nous sommes intéressés à généraliser des indices de diversité jusque là développés pour quantifier la diversité d'une communauté d'espèces représentée par un vecteur d'abondance. Pour cela, nous avons étendu les nombres de Hill à une représentation de communauté fondée sur un modèle de graphe aléatoire, permettant de mesurer une diversité d'espèces, d'abondances de liens ou de probabilités d'interaction.

Nous avons montré que des indices de comparaison de réseaux préalablement utilisés en écologie étaient des cas particuliers du cadre statistique proposé. De plus, dans la mesure où différentes définitions de la  $\beta$ -diversité existent, nous avons formulé ces statistiques descriptives en utilisant les définitions de Jost, Chao et Reeve. Si cette pluralité des définitions d'indice de diversité peut paraître complexe voire rébarbative, cela permet à ces statistiques de vérifier des propriétés désirables pour mener des études quantitatives sur la biodiversité. Par exemple, le cadre proposé par REEVE et al., 2014 vérifie l'invariance par découpe. Cette propriété émet l'hypothèse implicite que chaque communauté représente une portion d'espace. Si une communauté est découpée en deux morceaux et que les deux nouvelles communautés sont identiques, alors la  $\beta$ -diversité de la métacommunauté doit être identique avant et après la découpe. Nous avons montré que le cadre proposé par CHAO et CHIU, 2016 ne vérifiait pas cette propriété. Ainsi, pour une étude de diversité où les sites sont amenés à être ré-échantillonnés dans le futur avec une meilleure résolution spatiale, cette propriété est désirable (car on ne veut pas que la  $\beta$ -diversité soit influencée par le nombre de sites) et l'on privilégiera le cadre de REEVE et al., 2014. Par contre, ce cadre statistique ne vérifie pas des propriétés de base des indices de diversité quand sont

pris en compte des traits ou des relations phylogénétiques. Ainsi, pour une extension des indices de diversité de réseau qui utilise ces informations, le cadre proposé par CHAO et CHIU, 2016 est souhaitable.

Mais outre un développement purement technique, la méthode proposée permet de calculer des diversités de réseau à différents niveaux d'agrégation des nœuds, c'est-à-dire en représentant la carte des relations qu'est le réseau à une résolution moindre. Cela permet notamment de documenter des patrons de diversité de réseau dont les nœuds représentent des entités de résolution croissante : individu, espèce, groupe fonctionnel. Reste à définir ces groupes ; en utilisant une représentation réseau (plutôt qu'abondance), on peut définir des groupes d'espèces en utilisant uniquement la topologie du réseau (sans covariables sur les espèces). Les méthodes d'agrégation des nœuds permettent de fournir un réseau à une résolution moindre qui reflète la structure du réseau initial.

Si les abondances relatives des espèces et les abondances de lien peuvent être estimées à partir d'observations, estimer les probabilités d'interaction peut être plus complexe. Elle peuvent se déduire à partir des abondances d'espèces et de liens. Toutefois, dans le cas où l'une des deux espèces impliquées dans l'interaction est absente localement, on ne peut estimer la probabilité d'interaction. Une solution alternative consisterait à calculer la probabilité d'interaction à partir des traits des espèces.

### 5.1.2 Chapitre 2

Dans un deuxième chapitre, nous avons développé une théorie des méta-communautés d'espèces en interaction sur un réseau spatial. Si, dans la proposition d'article, les interactions modélisées sont uniquement mutualistes, toutes les interactions encodées par le modèle de Lotka-Volterra peuvent être modélisées dans la version stochastique du modèle. Dans ce chapitre, le modèle de méta-population sur un réseau spatial et le modèle île-continent de communauté avec des interactions biotiques sont formulés avec des modèles graphiques, permettant ainsi de mélanger les modèles (puisque'ils ont le même formalisme) pour construire le modèle de méta-communauté spatialement explicite. Cela permet alors, dans le modèle déterministe, d'exprimer des caractéristiques de la dynamique de la méta-communauté (*e.g.* la persistance) en fonction de la structure du réseau d'interaction et du réseau spatial. On peut mesurer l'apport des interactions mutualistes pour la persistance de la méta-communauté. Imaginons que la dynamique d'une méta-communauté se déroule sur un réseau spatial de valeur propre dominante  $\Lambda_s$  et ait un réseau d'interaction de valeur propre dominante  $\Lambda_b$ , avec des taux d'extinctions  $c$  et  $e$ . On suppose, pour simplifier, que la dynamique a lieu selon un modèle type Levins. Alors, dans ce cas, la capacité de la méta-communauté est  $\lambda_M = \Lambda_s + \Lambda_b$ . Ainsi, la méta-communauté persiste si  $\Lambda_s + \Lambda_b > \frac{e}{c}$ . En l'absence d'interactions, chaque méta-population a une capacité de persistance égale à  $\Lambda_s$ . On a toujours  $\lambda_M > \Lambda_s$ . Cela signifie que, quelque soit la structure du

réseau spatial, les interactions mutualistes permettent à la méta-communauté de persister en supportant des taux d’extinction plus élevés et des taux de colonisation plus faibles. De plus, l’intensité de cet effet dépend de la structure du réseau d’interaction. Ainsi, dans les milieux où les taux d’extinction sont élevés, soit des interactions mutualistes permettent à la méta-communauté de se maintenir, soit l’absence d’interaction mutualistes conduit à l’extinction des différentes metapopulations. Ainsi, ce modèle pourrait servir à la compréhension de l’hypothèse de gradient de stress (BERTNESS et CALLAWAY, 1994). Dans ce chapitre, les réseaux ne sont plus seulement des descriptions statiques des relations entre les sites ou les espèces mais les paramètres de modèles dynamiques (stochastiques ou déterministes). Dans le cadre du stage de Master 2 (Maths en action) d’Alexis Teskrat, nous nous sommes intéressés à l’implication des équivalences entre les nœuds (sites où espèces) sur la dynamique du modèle. Là aussi, le concept sous-jacent de niche d’Elton (ici compris comme positions équivalentes occupées par des espèces dans un réseau d’interaction) n’est pas seulement un moyen de réduire la complexité d’un réseau statique mais un moyen de faciliter la compréhension de la dynamique. En effet, les espèces équivalentes dans le réseau d’interaction vont conduire à obtenir des assemblages avec des probabilités égales dans l’état stationnaire. On peut obtenir des résultats similaires dans les modèles de méta-population, en prenant en compte l’équivalence de certains sites. Reste à s’intéresser aux équivalences dans le réseau produit du réseau d’interaction et du réseau spatial et les exprimer à partir des équivalences des deux réseaux initiaux. Notons toutefois que si le concept de niche d’Elton a largement été représenté par des méthodes d’agrégation de nœuds ayant des positions similaires ou équivalentes dans un réseau (*e.g.* GAUZENS et al., 2015), il a aussi été interprété à l’aune des modèles dynamiques en terme de vecteur d’impact (LEIBOLD, 1995). Si le modèle proposé est théorique, il est construit avec des outils utilisés en écologie statistique (AUCLAIR, PEYRARD et SABBADIN, 2017). Ainsi, d’une part, développer une technique d’inférence pour les paramètres du modèle n’est pas inenvisageable et, d’autre part, il réduit les problèmes d’interprétation des modèles d’interaction présenté en Fig. 7. Une fois le réseau d’interaction inféré, on peut prédire la dynamique de la (méta-)communauté, en utilisant le même formalisme, grâce à ces développements théoriques.

### 5.1.3 Chapitre 3

Dans un troisième chapitre, nous nous sommes penchés sur la question de l’inférence d’interactions à partir de données d’ADN environnemental. Cette fois, le réseau d’interaction est le support d’un modèle statistique non-dynamique (où les communautés sont interchangeable sans que ça change la vraisemblance du modèle) mais il n’est pas connu. Soulignons que le modèle utilisé pour la simulation est un modèle stochastique non-dynamique d’interaction : un réseau bayésien dans lequel on peut échantillonner des



communautés de façon hiérarchique. La méthode d'inférence de réseau de diversité permet alors de retrouver un réseau, à une résolution moindre (c'est-à-dire en agrégeant les variables initiales), dont la structure est fidèle à celle du réseau initial. Pour ce faire, on choisit *a priori* les groupes d'espèces au sein desquels on calcule une  $\beta$ -diversité. Dans la simulation, le réseau d'interaction est construit en utilisant des niveaux trophiques. On a donc choisi ces groupes pour calculer une  $\beta$ -diversité. En ce qui concerne les données réelles, le choix des groupes de taxons est plus complexe dans la mesure où, contrairement à la simulation, on ne connaît pas le réseau d'interaction. Dans ce cas, nous avons constitué des groupes trophiques à partir des assignations taxonomiques des différents MOTUs, en utilisant des bases de données. Ainsi, on émet implicitement l'hypothèse que les organismes présents dans les groupes trophiques constitués ont la même niche d'Elton : on indentifie alors groupe trophique et niche d'Elton. Loin de documenter des interactions de façon aussi précise qu'une observation sur le terrain ou une expérience, nous avons inféré des associations entre diversité de groupes d'organismes ayant des fonctions similaires dans le sol. Si le nombre de nœuds du réseau inféré est modeste (9 groupes), cela permet toutefois de discuter les arêtes du réseau en utilisant la littérature documentant les interactions entre les groupes.

#### 5.1.4 Chapitre 4

Il n'y a pas de réseaux dans le quatrième chapitre mais une méthode pour corriger les abondances de données d'ADN environnemental issues de différents marqueurs. C'est une étape préalable nécessaire à la reconstruction de réseaux à partir de données de metabarcoding qui utilisent des méthodes d'inférence avec des abondances (*e.g.* CHIQUET, MARIADASSOU et ROBIN, 2018). Mais loin d'être seulement une méthode de combinaison d'information provenant de différentes technologies, on peut s'interroger sur l'image du taxon que donne les différents marqueurs. Dans ce chapitre, l'ADN de plantes a été amplifié avec le marqueur 'Sper01', qui amplifie l'ADN des chloroplastes, et le marqueur 'Euka02', qui amplifie l'ADN du noyau (TABERLET et al., 2018). Ainsi, les deux marqueurs donnent des images différentes des taxons. L'ADN de plante initialement introduit est un mélange, dans des proportions inconnues, d'ADN chloroplastique et d'ADN nucléaire. Ainsi, suivant les parties de l'individu qui ont été introduites (feuilles ou racines), ces proportions peuvent varier. Il est donc légitime de recombinaison l'abondance des séquences provenant de différentes unités cellulaires pour obtenir une abondance reflétant l'ensemble des éléments de l'individu. De plus, nous avons comparé la variance biologique (variabilité de l'abondance des taxons d'un échantillon à l'autre) et la variance technologique (variabilité de l'abondance d'un taxon d'un marqueur à l'autre) sur un jeu de données témoin dans lequel tous les échantillons constituent des répliquats. Ainsi, la variabilité biologique attendue est nulle. Mais ce n'est plus le cas si l'on s'intéresse à l'abondance de

taxons le long d'un gradient environnemental. Les premiers tests de la méthode sur des données d'insecte le long de gradients d'altitude montrent une plus grande variance technologique que biologique, alors que la variabilité biologique dans l'abondance des taxons le long d'un gradient d'altitude est élevée à cause des effets de niche abiotique. Cela justifie alors l'emploi de la méthode des moyennes ajustées pour corriger les abondances. On pourrait imaginer une hypothèse plus réaliste quant à la modélisation du bruit introduit par la technologie d'observation. On a pour le moment supposé que chaque abondance était déformée par des variables aléatoires indépendantes et identiquement distribuées. On pourrait alors introduire une structure de corrélation qui reflète la structure de la taxonomie ou la phylogénie.

## 5.2 Vers de l'inférence et de la comparaison de réseaux pour l'observatoire ORCHAMP

### 5.2.1 L'observatoire ORCHAMP

L'observatoire ORCHAMP (Observatoire spatio-temporel de la biodiversité et des écosystèmes de montagne) est un programme d'observation qui a démarré en 2016, porté par le LECA mais qui fédère également un grand nombre de partenaires scientifiques (Edytem, Irstea, IMBE) et d'acteurs de terrain (parcs nationaux, régionaux, réserves, AFB, Agence de l'eau). Il propose de mutualiser les sites et les techniques d'observation le long d'une multitude de gradients d'altitude répartis dans les Alpes Françaises. Au cours de ces trois années de thèse, j'ai participé à la mise en place de sites et l'échantillonnage de sol en vue d'analyses d'ADN environnemental (env. 50 jours de terrain). Le but étant de reconstruire des réseaux d'associations entre groupes du sol le long des différents gradients d'altitude. Chaque site est constitué d'un gradient d'environ 1000m de dénivelé dans lequel des placettes sont installées tous les 200m d'altitude. Ces observations consistent en un prélèvement de sol pour des analyses d'ADN environnemental, physico-chimiques et de fonctionnements (*e.g.* activités enzymatiques) mais aussi un transect de végétation fait par des experts botanistes, des mesures sur les bois morts et vivants, une fosse pédologique et une station de température. ORCHAMP peut ainsi fournir des données pour répondre à des questions fondamentales (*e.g.* impact du climat sur la diversité) ou pour la conservation (*e.g.* suivi des populations, impact du changement climatique). En ce qui concerne l'ADN environnemental, les marqueurs 'Acar01', 'Arth02', 'Bact01', 'Coll01', 'Euka02', 'Fung02', 'Inse01', 'Olig01' et 'Sper01' ont été utilisés.

FIGURE 5.1 – Les sites d'échantillonnage d'ORCHAMP pour les années 2016, 2017 et 2018

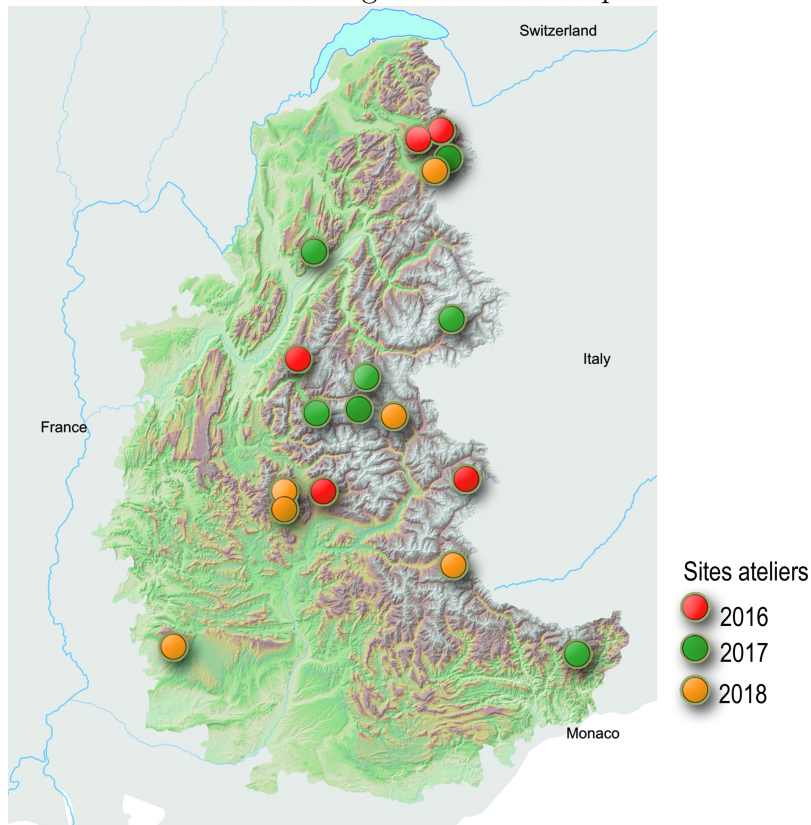


FIGURE 5.2 – Carte IGN de la partie supérieure (placettes 4, 5, 6, 7) du gradient du Lauvitel (Oisans).

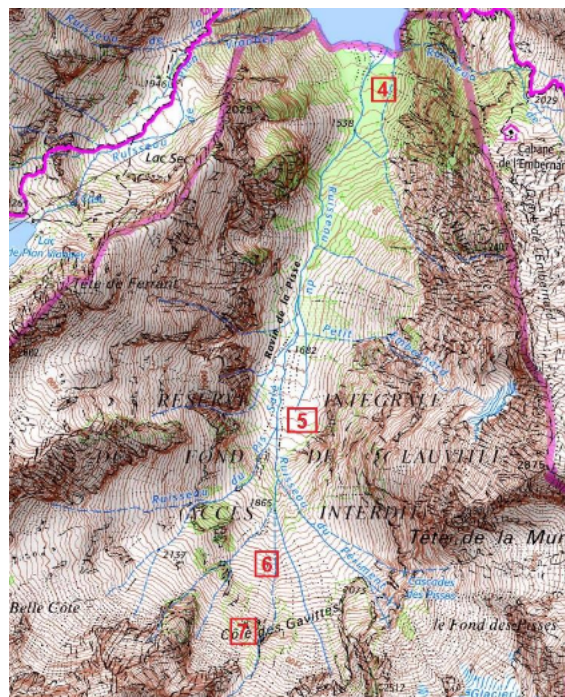


FIGURE 5.3 – Photographie de la partie supérieure du gradient du Lauvitel (Oisans)



### 5.2.2 Un chemin méthodologique escarpé

L'observatoire ORCHAMP permet d'obtenir des données de co-occurrences sur sol, des micro-organismes à la macrofaune, ainsi que des covariables pédo-climatiques. Ainsi, on peut espérer reconstruire des réseaux d'interaction impliquant une grande partie des organismes présents dans le sol tout en prenant en compte les variables abiotiques. On pourrait alors tâcher de comparer les réseaux inférés d'un gradient à l'autre. Une telle analyse permettrait de documenter les distributions spatiales d'espèces, leurs associations ainsi que la variabilité de ces associations dans le sol des Alpes en fonction de l'environnement. Des études expérimentales documentent déjà la variabilité des interactions entre plantes le long de gradient d'altitude dans les Alpes (CHOLER, MICHALET et CALLAWAY, 2001). Une documentation de la variabilité des interactions sur une grande partie des organismes présents dans le sol permettrait de tester l'hypothèse de gradient de stress à travers les groupes d'organismes présents dans les sols. Néanmoins, un tel objectif se heurte à d'importantes difficultés méthodologiques, qui mobilisent des connaissances dans plusieurs domaines : écologie, informatique, mathématiques.

#### Pré-traitement des données

Les données d'ADN environnemental permettent d'obtenir des données massives sur les organismes présents dans le sol. Toutefois, les difficultés liées à cette technologie sont mul-

tiples. D’abord, même si les marqueurs sont développées pour amplifier certains taxons, ils amplifient d’autres taxons que les taxons cibles (*cf.* Fig. 5.4). Ainsi, il y a des recouvrements entre les taxons amplifiés par les différents marqueurs et l’emploi de la méthode proposée au chapitre 4 est nécessaire. Notons toutefois que le fait d’obtenir des taxons qui ne soient pas visés par le marqueur n’est pas nécessairement lié au manque de spécificité du marqueur. Cela peut également venir d’erreurs lors de l’assignation des MOTUs à un nœud de la taxonomie (que l’on effectue en utilisant la fonction ‘ecotag’ du paquet ‘OBITools’, BOYER et al., 2016). Ainsi, le résultat de l’assignation dépend de la base de données et de sa complétude pour certains taxons.

### **Quelles entités dans le réseau ?**

Une fois le prétraitement des données effectué, reste à choisir quelles entités biologiques modéliser dans le réseau ; c’est-à-dire les nœuds du modèle graphique. Sur le plan technique, aux vues des succès contrastés de la reconstruction de réseau en écologie, choisir un ensemble de nœuds trop grand n’est probablement pas raisonnable. D’autant plus que les entités utilisées dans les données d’ADN environnemental dépendent de la technologie ; les MOTUs sont des groupes de séquences similaires et non des espèces. De plus, au vu du système modélisé (le sol), la notion même d’individu et d’espèce peut être difficile à manipuler pour certains groupes d’organismes comme les champignons (HARRINGTON et RIZZO, 1999) ou les bactéries (HANAGE, FRASER et SPRATT, 2005). Ces difficultés à appréhender la notion d’individualité sont évoquées par ce Haïku de Bashô Matsuo :

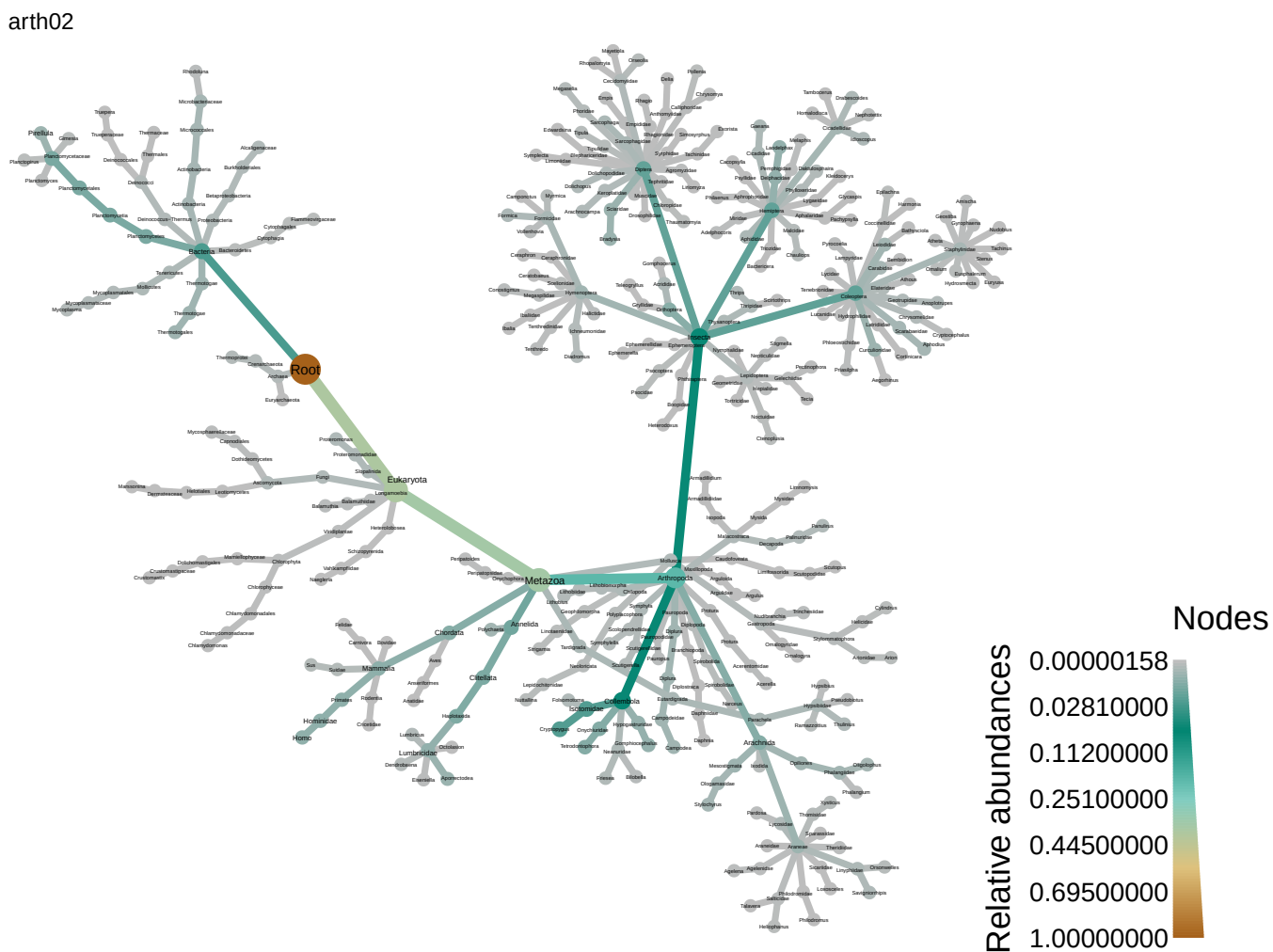
Collée sur un champignon une feuille d’on ne sait quel arbre.

Une solution consiste à faire des groupes trophiques d’organismes à différentes résolutions en utilisant des bases de données taxonomiques (*e.g.* FUNGUILD, NEMGUILD ; NGUYEN et al., 2016). Ainsi, on réduit la complexité de la carte en regroupant ensemble des organismes aux fonctions similaires. De plus, cela rend plus aisé la validation du réseau inféré.

### **Confusion dans le bestiaire des modèles**

Une fois les entités modélisées choisies, reste à choisir le modèle statistique utilisé pour reconstruire le réseau. Tous les modèles visent à reconstruire un réseau d’interaction ; pourtant la nature du modèle peut conditionner la structure du réseau ou le type d’interaction représentée. Par exemple, utiliser un réseau bayésien contraint le réseau à être un graphe acyclique orienté (DAG) et plusieurs réseaux peuvent représenter les mêmes indépendances conditionnelles. On obtient alors des graphes dirigés (ou partiellement dirigés) dont la direction des arêtes ne représente toutefois pas de la causalité. Ainsi, l’interprétation des arêtes (par exemple en terme de contrôle ‘top down’ ou ‘bottom up’) est sujette à caution. Il est toutefois possible de construire des réseaux qui représentent

FIGURE 5.4 – Arbre taxonomique pondéré par les abondances pour le marqueur Arth02 (Arthropodes) sur l'ensemble des données ORCHAMP 2016



de la causalité en utilisant des réseaux bayésiens causaux (PEARL, 1995). Toutefois, cela requiert alors de travailler avec des données interventionnelles (*i.e.* de contrôler les abondances des organismes modélisés) et plus seulement des données observationnelles. Or dans le cadre de l'observatoire ORCHAMP, cela paraît compliqué dans la mesure où les données d'abondance sont de l'ADN dégradé, qui peut provenir d'organismes non-vivants. Ainsi, si on contrôle l'abondance de la quantité d'ADN d'un groupe, cela n'aura pas les mêmes conséquences sur l'ADN des autres groupes que si l'on contrôlait l'abondance d'un groupe d'organismes vivants.

En utilisant des réseaux bayésiens dynamiques, on s’affranchit des problèmes de causalité car on peut raisonnablement supposer que c’est le présent qui influence le futur et pas l’inverse. De plus, si le réseau déroulé dans le temps est un DAG, le réseau d’interaction peut avoir n’importe quelle structure et donc en particulier représenter des interactions avec des cycles (*e.g.* compétition). L’application sur des données requiert toutefois des données temporelles, qui seront peut être disponibles dans le cadre d’ORCHAMP quand des sites seront ré-échantillonnés. De plus, dans l’approche de AUCLAIR, PEYRARD et SABBADIN, 2017, il est possible de définir un *a priori* sur la structure du réseau en utilisant un modèle stochastique à blocs. On pourrait alors combiner des données de traits sur les organismes (pour la structure *a priori*) et les données d’abondances. Ainsi, le réseau inféré aurait une structure influencée par les traits mais variable d’un site à l’autre, modulée par les données de co-occurrence utilisées dans l’inférence.

Pour s’affranchir des problèmes de causalité et d’équivalence de structure de réseau, on peut utiliser un modèle graphique non-dirigé. Les arêtes représentent alors des dépendances conditionnelles, qui peuvent modéliser plusieurs types d’interactions (même des interactions asymétriques). On parlera alors plutôt de réseaux d’associations dans ce cas, qui peuvent constituer une première étape dans la construction de réseaux d’interactions.

Obtenir des réseaux plus détaillés au niveau des interactions implique d’encoder différents mécanismes d’interaction (comme dans AUCLAIR, PEYRARD et SABBADIN, 2017) dans le modèle, cela prend alors la forme d’étiquette sur les arêtes qui se traduisent par des paramétrisations dans les probabilités conditionnelles.

Si l’on parvient à inférer des modèles qui représentent fidèlement le réseau d’interaction dans un site, reste à effectuer la comparaison des réseaux inférés entre les différents sites. S’il existe des méthodes pour comparer des réseaux dont l’une est proposée en chapitre 1, les méthodes de comparaison de structure de modèles graphiques sont plus rares. BACH, 2008 propose un kernel de graphe adapté à la structure des modèles graphiques. Néanmoins, ces méthodes de comparaison de structure ne prennent pas en compte l’incertitude des arêtes de modèles graphiques, qui peut être estimé grâce à des méthodes de ré-échantillonnage (LIU, ROEDER et WASSERMAN, 2010).

On notera ici également les ressemblances entre les modèles graphiques utilisés pour la reconstruction de réseau d’interaction et les modèles joints de distribution d’espèces. Les modèles joints de distribution se posent comme successeurs des modèles de distribution d’espèces, où les paramètres associés à chaque espèce sont estimés indépendamment (GUISAN, THUILLER et ZIMMERMANN, 2017). Le premier modèle proposé est une régression probit multivariée (POLLOCK et al., 2014).

En supposant qu’il y a  $n$  sites et  $S$  espèces différentes, on explique alors la présence des espèces au site  $i$  par  $\mathbf{Y}_i$ , un vecteur aléatoire binaire.

On a :

$$Pr(Y_{ij} = 1|Z_{ij}) = Pr(Z_{ij} > 0) \tag{5.1}$$

où  $\mathbf{Z}_i$  est une variable latente définie par :

$$\mathbf{Z}_i = \mathbf{B}\mathbf{x}_i + \epsilon_i \quad (5.2)$$

où  $\mathbf{x}_i$  est le vecteur des variables environnementales au site  $i$  (on a donc une régression sur les variables environnementales) et  $\epsilon_i \sim \mathcal{N}(\mathbf{0}_S, \Sigma)$ . Ainsi,  $\Sigma$  représente les associations marginales entre les espèces (elle est de taille  $S$ ).

Une façon de réduire la taille de l'espace des paramètres consiste à exprimer  $\Sigma$  comme :

$$\Sigma = \mathbf{A}\mathbf{A}^T + \sigma_\epsilon^2 \mathbf{I}_S \quad (5.3)$$

où  $\sigma_\epsilon^2 \in \mathbb{R}^{+*}$  et  $\mathbf{A}$  est une matrice de taille  $S \times r$ , avec  $r < S$ .

TAYLOR-RODRIGUEZ et al., 2017 propose alors de réduire la dimension du problème en regroupant des lignes de  $A$  avec une méthode bayésienne non-paramétrique. On remarquera que si deux lignes de  $A$  sont égales (elles sont associées à deux espèces  $i$  et  $j$ ), alors les espèces  $i$  et  $j$  sont dans la même orbite du réseau de corrélation. En d'autres termes, elles ont même niche d'Elton dans le réseau qui décrit les associations (marginales) entre les espèces. TIKHONOV et al., 2017 introduit une dépendance à l'environnement dans la matrice d'association  $\Sigma$  et propose de l'interpréter comme une matrice d'interaction. Il vérifie alors l'hypothèse de gradient de stress chez des plantes dans la toundra finlandaise.

### 5.3 Projets périphériques

En lien avec le chapitre 1, des analyses biogéographiques de réseau à différentes échelles d'agrégation ont été effectuées sur un réseau trophique de vertébrés à l'échelle de l'Europe (O'CONNOR et al., 2019). Le meta-réseau trophique a été construit à dire d'expert. A l'image du gradient latitudinal de la répartition des espèces, un gradient latitudinal dans la structure des réseaux a été mis en évidence, au niveau des espèces mais aussi des groupes d'espèces ayant des façon similaires de se connecter formé avec un modèle stochastique à blocs. En lien avec le chapitre 2, et en collaboration avec Charline SMADI (IRSTEA Grenoble) et dans le prolongement du stage d'Alexis TESKRAT, nous essayons de mieux comprendre l'implication des équivalences entre nœuds du réseau spatial (sites) et nœuds du réseau d'interaction (espèces) sur la dynamique du modèle de métacommunauté. Nous cherchons un cadre plus général pour la paramétrisation du réseau bayésien dynamique ainsi que des équivalences dans le modèle de métacommunauté qui seraient le produit d'équivalence d'espèces et de sites. En lien avec le chapitre 3, un article a été publié, sur les mêmes données, pour comprendre comment les variables abiotiques et la diversité des différents groupes influencent différentes fonctions écologiques du système (MARTINEZ-ALMOYNA et al., 2019). En utilisant les mêmes données et les mêmes groupes que dans



le chapitre 3, on s'est intéressé à l'explication, en utilisant un modèle de piste, de différentes mesures de fonctionnement de l'écosystème par les variables environnementales et les changements de composition des groupes trophiques. Il a été montré que la variabilité des propriétés physico-chimiques du sol le long du gradient et la variation de composition des groupes (en particulier les champignons) expliquaient la multi-fonctionnalité de l'écosystème. En lien avec le chapitre 4 et en collaboration avec Stéphane Robin, nous essayons d'améliorer les mélanges de marqueurs de metabarcoding. La méthode proposée pour le moment suppose l'indépendance des biais entre les différents taxons. On pourrait imaginer introduire une structure de dépendance entre les biais qui reflète la structure de l'arbre phylogénétique. De plus, nous essayons de prouver des résultats analytiques sur le biais introduit dans l'estimation de la diversité (mesurée avec des nombres de Hill) par l'hétérogénéité des moyens d'observation pour les taxons.

Enfin, en collaboration avec Catherine MATIAS (LPSM, Paris), Vincent MIELE (LBBE, Lyon) et Stéphane DRAY (LBBE), nous essayons non pas d'inférer les interactions entre organismes mais l'importance des interactions (à réseau d'interaction connu) devant les processus de niche abiotiques pour expliquer la répartition spatiale des organismes. Plus précisément, on suppose qu'à l'échelle d'une région composée de site  $i \in \{1, \dots, n\}$ , on connaît le metaréseau  $G^* = (V^*, E^*)$ , où  $V^* = \{1, \dots, S\}$  est l'ensemble des espèces à échelle régionale. On représente la présence des espèces en chaque site  $i$  par les variables aléatoire  $\{X_v^i\}_{v \in V^*}$ . Le metaréseau encode alors les dépendances entre présences d'espèce grâce à un champ de Markov. On suppose que les  $\{X_v^i\}_{v \in V^*}$  sont distribuées selon :

$$\mathbb{P}_{\psi_i}(\{X_v^i\}_{v \in V^*}) = \frac{1}{Z_{\psi_i}} \exp \left( \sum_{v \in V^*} \alpha_{v,i} X_v^i + \beta_i \sum_{(u,v) \in E^*} 1\{X_u^i = X_v^i\} \right) \quad (5.4)$$

où  $1\{E\}$  est la fonction indicatrice de l'évènement  $E$ ,  $\psi_i = (\alpha_{1,i}, \dots, \alpha_{S,i}, \beta_i) \in \mathbb{R}^S \times \mathbb{R}^+$  sont les paramètres de la distribution et  $Z_{\psi_i}$  est la fonction de partition :

$$Z_{\psi_i} = \sum_{\{x_v\}_{v \in V^*} \in \{0,1\}^S} \exp \left( \sum_{v \in V^*} \alpha_{v,i} x_v + \beta_i \sum_{(u,v) \in E^*} 1\{x_u = x_v\} \right) \quad (5.5)$$

La technique d'inférence permet d'obtenir  $\psi_i$ . Ainsi, on peut évaluer l'importance des interactions biotiques localement (comprise comme le degré de dépendance entre les variables,  $\beta_i$ ) devant l'importance des processus de niche (comprise comme la probabilité de présence d'une espèce en un site sans interaction,  $\alpha_{v,i}$ ). Il est possible d'introduire des covariables environnementales dont les  $\alpha_{v,i}$  sont fonctions. A noter que dans ce cas, on représente des interactions trophiques (non-symétriques) par un réseau non-dirigé. Ainsi, le modèle ne représente pas des interactions (causales) mais suppose que ces interactions produisent des associations. Plus précisément, on suppose dans ce modèle que si une proie et un prédateur sont co-absents ou co-présents, alors la probabilité de l'assemblage

d'espèce augmente.

## 5.4 Représentations de la nature en écologie mathématique

L'introduction de cette thèse mêle références scientifiques, littéraires et artistiques. La volonté des différents auteurs, quelque soit leur discipline, est de représenter la nature. Toutefois, ils ont des rapports à l'idée de nature assez différents les uns des autres. Dans *le voile d'Isis*<sup>1</sup>, Pierre Hadot propose une classification des attitudes vis-à-vis de la nature à la lumière de la mythologie grecque. Il se propose de commenter un fragment d'Héraclite :

La Nature aime à se cacher

Il distingue alors deux attitudes vis-à-vis de cette difficulté à appréhender la nature. D'une part l'attitude prométhéenne, attribuée à Prométhée, qui vola le feu aux Dieux pour en faire bénéficier les hommes. Cette attitude cherche à percer par la ruse et l'expérience les mystères de la Nature, à la soumettre à un interrogatoire technique. D'autre part, il distingue l'attitude orphique, attribuée à Orphée, qui descendit aux Enfers lors d'un voyage initiatique pour retrouver son épouse défunte Eurydice. Sa lyre et ses chants avaient un pouvoir de séduction irrésistible sur les êtres vivants. Orphée veut pénétrer les mystères de la Nature par la mélodie et la poésie. Dans *l'Hymne de l'Eternité*, Pierre de Ronsard écrit :

Rempli d'un feu divin qui m'a l'âme échauffée,  
Je veux, mieux que jamais, suivant les pas d'Orphée,  
Découvrir les secrets de Nature et des Cieux

On pourrait être tenté de dire, en première approximation, que l'attitude prométhéenne relève de la science alors que l'attitude orphique relève des arts. Toutefois, en écologie les deux attitudes peuvent être présentes, surtout si on les distingue par rapport au fait de simplement observer la nature ou d'interagir avec elle en faisant des expériences. Georges Cuvier écrit :

L'observateur écoute la Nature, l'expérimentateur la soumet à un interrogatoire et la force à se dévoiler

On peut souligner l'importance de l'observation en écologie par rapport à l'expérience en biologie. Cela relève peut être plus de contraintes techniques que de choix méthodologique, dans la mesure où les expériences *in natura* en écologie sont rares car complexes à mettre en oeuvre. Elles existent toutefois comme les expériences de suppression d'individus de communautés végétales ou les expériences transplantation de communauté au col du Lautaret auxquelles j'ai participé. Ainsi, l'écologie scientifique, ayant plus à trait

---

1. HADOT, 2004.

FIGURE 5.5 – La science dévoilant la Nature, Gerhard-Blasius, 1681



à l’observation pourrait avoir des traits orphiques. Mais c’est oublier que le façon de collecter des spécimens peut s’avérer être destructeur, comme nous l’avons vu avec la chasse au XIX<sup>ème</sup> siècle. De même, la collection massive de données sur la nature peut relever d’une attitude prométhéenne même s’il n’y a pas d’expérience ou d’intervention sur la nature : le but est tout de même de disséquer la nature par la surveillance massive. Mais cette dynamique a un objectif : documenter pour conserver la nature et répondre aux changements globaux, comme Prométhée pensait sauver l’humanité en lui confiant le feu. La documentation de la nature peut toutefois s’effectuer avec un soucis d’esthétique délibéré comme dans les planches de Haeckel, se soustrayant au devoir purement utilitaire de documentation.

En ce qui concerne la modélisation mathématique et informatique de la nature, la collecte de données massives penche du côté prométhéen. En effet, elle vise, par le recours à la technique, à la connaissance des entrailles de la nature. Néanmoins, utiliser des modèles mathématiques complexes pour représenter la nature peut avoir attrait à l’orphisme. En effet, d’une part, cette représentation est formulée dans un langage abstrait qui requiert une initiation et, d’autre part, la modélisation met en exergue l’écart infranchissable qu’il existe entre la carte (le modèle) et le territoire (les phénomènes de la nature). Ainsi, cela accentue l’impression que la nature aime se cacher des modèles. Cette idée est symbolisée par le voile d’Isis (Isis est une déesse égyptienne, également vénérée chez les grecs, et symbolisant les mystères de la nature ; elle est parfois identifiée à Artémis). La science prométhéenne voudrait, par la technique, lever le voile d’Isis alors que l’Orphisme préfère chanter les mystères de la nature.

Dans cette thèse, l’accent a été mis sur les façons de représenter la nature, et, en

particulier, les interactions entre les organismes avec des modèles mathématiques. Si un tel travail peut se faire au détriment des conclusions écologiques que l'on peut tirer en utilisant ces représentations, il m'a paru essentiel car les représentations du mathématiques utilisées conditionnent tout discours sur la nature en écologie.



DEUXIÈME PARTIE

# Annexes

---



## **5.5 Rapport de stage M2 'Maths en action' d'Alexis Teskrat**



# Rapport de stage de M2

*Équivalence des noeuds dans les modèles stochastiques de colonisation-extinction sur des graphes en écologie : conséquences sur la dynamique*

ALEXIS TESKRAT

Master Maths en action, *filière* : Mathématiques pour la biologie et la médecine

Tuteurs :

- Marc Ohlmann, doctorant au Laboratoire d'Écologie Alpine.
- Charline Smadi, chercheuse à l'IRSTEA de Grenoble.
- Wilfried Thuiller, directeur de recherche au CNRS, Laboratoire d'Écologie Alpine.

Stage effectué au sein du Laboratoire d'Écologie Alpine du 2 avril 2019 au 27 septembre 2019.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Contexte mathématique</b>	<b>4</b>
<b>3</b>	<b>Modèles et simulations</b>	<b>9</b>
3.1	Modèle : île/continent . . . . .	9
3.2	Modèle méta-population . . . . .	12
3.3	Modèle méta-communauté . . . . .	14
<b>4</b>	<b>Réduction du problème</b>	<b>16</b>
4.1	Notions d'équivalence . . . . .	16
4.2	Agrégation matricielle . . . . .	19
<b>5</b>	<b>Conclusion, discussion</b>	<b>20</b>
<b>6</b>	<b>Références</b>	<b>21</b>
<b>7</b>	<b>Appendice</b>	<b>22</b>
7.1	Codes, simulations et exemples . . . . .	22
7.2	Preuves . . . . .	34

# 1 Introduction

En 1927, Charles Elton publie le premier réseau trophique connu en écologie, échantillonné dans des îles au large de la Norvège (Elton, 1927). Il suggère ainsi l'utilisation des graphes pour décrire les communautés d'espèces et leurs interactions. Depuis, les graphes sont utilisés en écologie pour représenter et analyser les communautés d'espèces en interaction. Ils permettent de représenter plusieurs types d'interactions en utilisant des étiquettes sur les arêtes du graphe : relation trophique (le prédateur a une influence négative sur la présence de sa proie et la proie a une influence positive sur la présence du prédateur), compétition (les deux partenaires ont une influence négative l'un sur l'autre) ou mutualisme (les deux partenaires ont une influence positive l'un sur l'autre). Elton propose le concept de niche d'espèces dans un réseau trophique en formulant une équivalence entre des espèces d'une communauté en fonction des interactions qu'elles entretiennent. D'autres se sont penchés sur la structure des réseaux pour trouver des similarités ou des équivalences entre espèces. (Luczkovich *et al.*, 2003), par exemple, utilise les notions d'équivalence structurel et d'équivalence régulière afin de créer des groupes d'espèces sur un réseau trophique : des classes isotrophiques. De tels concepts permettent, dans un réseau statique, de réduire la complexité des communautés étudiées en s'intéressant, par exemple, à la niche des espèces plutôt qu'à leur taxonomie. S'ouvre alors la voie de l'écologie fonctionnelle (avec les travaux de Raymond Lindemann, (Lindeman, 1942)), où un écosystème est vu comme un ensemble de compartiments d'espèces avec des fonctions particulières. En écologie, les graphes ne servent pas uniquement à représenter les réseaux d'interactions entre espèces mais aussi l'espace fragmenté dans lequel les populations d'espèces vivent (Dale & Fortin, 2010).

Mais les réseaux d'interactions ne sont pas seulement des descriptions statiques de communautés ou de l'espace, ils peuvent être utilisés dans des modèles dynamiques où ils sont alors paramètre. Si on considère un modèle dont le réseau d'interaction est un paramètre, nous allons chercher à comprendre la dynamique des espèces présentes dans le réseau. De même, si on considère un modèle dont le réseau spatial est un paramètre (il y a alors une seule espèce), nous allons chercher à comprendre la dynamique de la population d'espèce dans le réseau ; de tels modèles sont appelés modèles de méta-population en écologie (Hanski & Ovaskainen, 2003). Nous pouvons aussi nous pencher sur l'étude de modèles décrivant la dynamique de plusieurs espèces en interaction et évoluant sur un réseau spatial, ceux-ci combinent les informations des deux modèles précédents ; ils sont appelés modèles de méta-communauté en écologie.

Le problème étant que, généralement, le nombre d'espèces étudiées ainsi que le nombre d'interactions sont grands dans les systèmes réels. Ainsi, le temps et la mémoire nécessaire aux calculs permettant de calculer le résultat des dynamiques est colossale. Différentes méthodes de simulation sont alors mises en pratique afin de s'approcher du résultat initialement recherché ; une partie de mon travail porte sur ces méthodes de simulation. Une réflexion peut être portée, a priori, en étudiant le comportement de certaines espèces présentes dans les communautés étudiées, et leur place dans le réseau. En effet, certaines d'entre elles peuvent alors être équivalentes d'un point de vue écologique, se traduisant alors par une relation d'équivalence sur les noeuds du réseau d'interaction. Une fois les classes d'équivalences mises

en évidence, il est possible de trouver des destins communs à certaines espèces étudiées en évitant, ainsi, la redondance de certains calculs fastidieux.

Dans ce rapport, nous définissons les objets mathématiques employés tout du long, puis nous présentons un modèle stochastique de chaque type cité précédemment, ainsi que les simulations effectuées, et enfin nous résonnons sur des notions d'équivalence afin de résoudre le problème de trop grande dimension qui se pose. Les preuves des différents résultats que nous présentons sont en appendice, ainsi que des ensembles d'exemples effectués avec le langage R.

## 2 Contexte mathématique

Rappelons d'abord la définition de certains objets que nous utilisons tout au long de ce rapport. Tout d'abord les graphes et certaines de leurs particularités.

**Définition 2.1.** *Un graphe  $\mathcal{G}$  est un couple d'ensembles finis  $(V, E)$  tel que :*

-  $V$  est non vide et appelé ensemble des noeuds (ou sommets) du graphe. Le cardinal de  $V$  est appelé taille du graphe  $\mathcal{G}$ .

-  $E \subseteq \{(x, y) \mid (x, y) \in V^2, x \neq y\}$  est appelé ensemble des arêtes du graphe. Si  $(x, y) \in E$ , alors on dit que  $x$  et  $y$  sont voisins.

De plus un graphe  $(V, E)$  est dit orienté si on différencie les arêtes  $(x, y)$  et  $(y, x)$ . Dans ce cas, si  $(x, y) \in E$ , alors on dit que  $x$  est parent de  $y$ .

— **Remarque** Si le graphe que l'on considère est non-orienté, alors chaque noeud est parent du noeud dont il est voisin.

**Définition 2.2.** *Soit  $\mathcal{G} = (V, E)$  un graphe. On appelle fonction parent, noté  $Par$ , la fonction qui, à un noeud  $i \in V$ , associe l'ensemble des noeuds qui lui sont parents.*

**Définition 2.3.** *Un graphe pondéré sur ses arêtes est un graphe auquel on adjoint une fonction de valuation.*

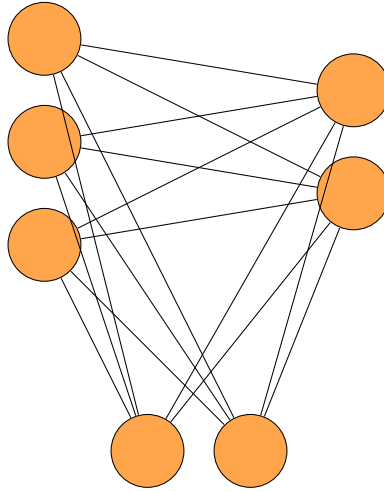
$$\text{Weight} : \begin{array}{l} E \longrightarrow \mathbb{R} \\ (x, y) \longmapsto \text{Weight}(x, y) \end{array}$$

Nous allons aussi avoir besoin de vocabulaire concernant les types de graphes dans la partie 4 de ce rapport.

**Définition 2.4.** *Un graphe est dit complet, dans le cas non-orienté, lorsque chaque noeud est voisin des autres noeuds. Dans le cas orienté, le graphe est dit complet lorsque chaque noeud est parent des autres noeuds. Un graphe complet de taille  $n$  se note  $\mathbf{K}_n$ .*

*Un graphe est dit biparti si son ensemble de sommets peut être divisé en deux sous-ensembles disjoints  $V_1$  et  $V_2$  tels que chaque arête ait une extrémité dans  $V_1$  l'autre dans  $V_2$ . L'ensemble des graphes biparties et complet dont l'ensemble des sommets peut être divisé en deux sous-ensemble de taille  $m$  et  $n$ , se note  $\mathbf{K}_{mn}$ .*

*Plus généralement, un graphe est dit multiparti si son ensemble de sommets peut être divisé en plusieurs sous-ensembles disjoints tels que chaque arête a une extrémité dans un de ces*

FIGURE 1 –  $K_{3,2,2}$ 

ensembles et l'autre dans un autre de ces ensembles. On note  $\mathbf{K}_{\underbrace{n_1 \dots n_1}_{r_1 \text{ fois}} \underbrace{n_2 \dots n_2}_{r_2 \text{ fois}} \dots \underbrace{n_p \dots n_p}_{r_p \text{ fois}}}$ , avec  $p, n_i, r_i \in \mathbb{N}^* \quad \forall i \in \{1, \dots, p\}$ , l'ensemble des graphes multipartis complets, dont la tailles des ensembles partitionnant le l'ensemble des noeuds est décrite par les indices.

**Exemple** Exemple d'un graphe multiparti complet dans  $\mathbf{K}_{3,2,2}$  en figure 1 .

Dans nos modèles les noeuds du graphe peuvent représenter des espèces ou des sites géographiques (suivant que l'on modélise un réseau d'interaction ou un réseau spatial); les arêtes, quant à elles, représentent les interactions entre espèces ou les relations de proximité géographiques entre les sites. En fonction du modèle ce graphe sera orienté et pondéré sur ses arêtes ou tout simplement non orienté et non pondéré. Ce graphe que nous venons de décrire, nous l'appellerons **graphe initial**, il résume l'information connue sur les espèces ou les lieux étudiés.

Désormais nous voulons construire un modèle dynamique à partir de ce graphe initial. Afin de mieux comprendre la suite du raisonnement, plaçons-nous dans le cas d'une étude sur  $n \in \mathbb{N}^*$  espèces vivantes interagissant entre elles pour le reste de cette partie de contextualisation (nous verrons par la suite comment il est possible d'appliquer notre démarche à une population d'une même espèce colonisant plusieurs sites géographiques).

Introduisons un pas de temps discret, noté  $t$ , et intéressons-nous à la présence ou à l'absence des espèces étudiées à chaque instant. Notre dynamique temporelle est aléatoire et les probabilités de colonisation et d'extinction sont conditionnées par la présence ou l'absence d'espèces voisines.

**Définition 2.5.** Soit  $n \in \mathbb{N}^*$ , on note  $Word_{bin}(n)$ , l'ensemble des mots de taille  $n$  composés uniquement de 0 et de 1.

De plus pour tout état (mot)  $S = \overline{x_1 x_2 \dots x_n}$  dans  $Word_{bin}(n)$  on définit la richesse de l'état  $S$ , notée  $|S|$ , un entier tel que  $|S| = \sum_{i=0}^n x_i$ .

On considère  $(\mathbf{X}^t)_t$  un processus aléatoire à valeur dans  $Word_{bin}(n)$ . À chaque mot correspond un assemblage des  $n$  espèces étudiées. On identifie les espèces en fonction de leur place dans le mot et leur présence en fonction de la valeur. Par exemple, l'état  $00110 \in Word_{bin}(n)$  signifie que les espèces 1,2 et 5 sont absentes, alors que les 3 et 4 sont présentes (dans le cas d'un modèle spatial, cela signifie que les sites 1,2 et 5 sont absents de vie et que les sites 3 et 4 sont occupés). S'intéresser à l'état de l'espèce numérotée  $i \in \llbracket 1, n \rrbracket$ , c'est chercher à connaître la valeur prise par la  $i^{\text{ème}}$  lettre du mot dans  $Word_{bin}(n)$ . On notera  $X_i^t$  la variable aléatoire dans  $\{0, 1\}$  donnant la valeur de la  $i^{\text{ème}}$  lettre du mot obtenue après tirage au temps  $t$ ; ainsi  $\mathbf{X}^t = (X_1^t, X_2^t, \dots, X_n^t)$ . Les probabilités de transition du processus  $\mathbf{X}^t$  dépendent du modèle choisi, de ses paramètres et du graphe initial. La valeur de  $X_i^{t+1}$  dépend de celle de  $X_i^t$  et des voisins de l'espèce  $i$  à l'instant  $t$  (voisins visibles sur le graphe initial). Dans le cadre d'un modèle spatial, c'est l'état des sites voisins qui influe sur les valeurs prises par les différents  $X_i^t$ .

Une manière de résumer les informations de dépendance conditionnelle, entre les variables aléatoires associées aux présences-absences des espèces, aux différents pas de temps, est d'utiliser un réseau bayésien dynamique (DBN) Auclair *et al.* (2017). Il s'obtient à partir du graphe initial avec ajout d'une boucle sur chaque noeud et du graphe linéaire à 2 noeuds,  $P_2$ , qui décrit le passage de  $t$  à  $t + 1$  (Figure 2).

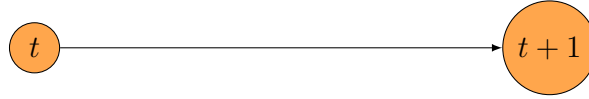


FIGURE 2 –  $P_2$ , graphe de passage de  $t$  à  $t + 1$

**Définition 2.6.** Une boucle d'un graphe  $\mathcal{G} = (V, E)$  est un couple  $(x, x)$  avec  $x \in V$ . On dit alors que le graphe admet des boucles.

Si le graphe est pondéré sur ses arêtes, il est possible d'étendre la définition du poids aux boucles si le graphe en contient. Ainsi, si  $x \in V$  et  $(x, x)$  une boucle de  $\mathcal{G}$  alors le poids de cette boucle est  $Weight(x, x)$ .

**Définition 2.7.** Soit  $\mathcal{G} = (V, E)$  un graphe de taille  $n \in \mathbb{N}^*$ . On note  $(v_1, v_2, \dots, v_n)$  les noeuds du graphe  $\mathcal{G}$ .

On appelle matrice d'adjacence, la matrice de taille  $n \times n$  donc les coefficients sont :

$$a_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon} \end{cases}$$

Si  $\mathcal{G}$  est un graphe pondéré, les coefficients de la matrice sont :

$$a_{ij} = \begin{cases} Weight(v_i, v_j) & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon} \end{cases}$$

On obtient la matrice d'adjacence du DBN recherché,  $\mathbf{A}_{\text{DBN}}$ , à partir de la matrice d'adjacence du graphe initial à  $n$  noeuds,  $\mathbf{A}$ , et de la matrice d'adjacence de  $P_2$ ,  $\mathbf{A}_2$  de la façon suivante :

$$\mathbf{A}_{\text{DBN}} = \mathbf{A}_2 \otimes (\mathbf{A} + \mathbf{I}_n) \quad (1)$$

Illustrons ceci à l'aide d'un exemple simple.

**Exemple** Considérons un graphe initiale entre deux espèces A et B, interagissant entre elles, avec ajout de boucles, présenté en figure 3. Le réseaux dynamique que l'on obtient est

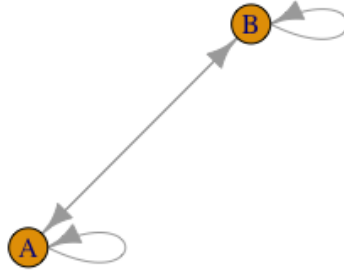


FIGURE 3 – Graphe initial avec boucle

présenté en figure 4.

Le réseaux bayésien dynamique constitue une carte des dépendances conditionnelles entre des variables aléatoires entre  $t$  et  $t + 1$ . Ces variables peuvent représenter soit la présence d'espèces ( $A$  et  $B$  dans l'exemple) sur une île à un pas de temps donné soit la présence dans deux sites géographiques d'une population d'espèce à un pas de temps donné. On peut construire une chaîne de Markov à partir d'un réseau bayésien dynamique (Lähdesmäki & Shmulevich (2008)). Dans notre cas, il s'agit de la chaîne  $(\mathbf{X}^t)_t$  présenté précédemment.

On peut appréhender la dynamique du modèle qui a comme paramètre le graphe initial en cherchant la distribution stationnaire ou quasi-stationnaire (en fonction du modèle) de la chaîne de Markov  $(\mathbf{X}^t)_t$ . Son graphe d'état est de taille  $\#Word_{bin}(n) = 2^n$  et sa matrice de transition de taille  $2^n \times 2^n$ . La mémoire de l'ordinateur et sa vitesse de calcul sont des obstacles au calcul de la matrice de transition, car la taille de la matrice augmente de façon

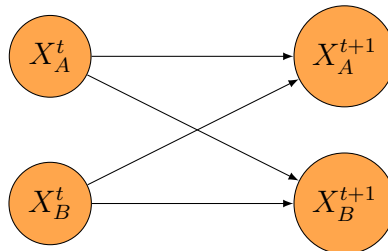


FIGURE 4 – Réseau bayésien dynamique. Ce graphe nous indique que la présence de l'espèce  $A$  à l'instant  $t$  influe sur sa propre présence à l'instant  $t + 1$  mais aussi sur celle de  $B$  à  $t + 1$ . Symétriquement, l'espèce  $B$  agit de la même manière sur elle même et sur  $A$ .

exponentielle avec le nombre d'espèces.

Dans la quatrième partie de ce travail nous cherchons un moyen de contourner le problème exprimé précédemment. Nous nous servons de la structure du graphe initial et du modèle utilisé afin de réduire la complexité des calculs. Nous avons besoin de rappeler les définitions, relatifs aux graphes, suivantes :

**Définition 2.8.** Soient  $\mathcal{G}_i = (V_i, E_i)$ ,  $i = 1, 2$  deux graphes.

$f : V_1 \Rightarrow V_2$  est un homomorphisme de  $\mathcal{G}_1$  dans  $\mathcal{G}_2$  si  $(x, x') \in E_1 \Rightarrow (f(x), f(x')) \in E_2$ .

Si  $f$  est bijective, alors  $f$  est un isomorphisme de  $\mathcal{G}_1$  dans  $\mathcal{G}_2$ .

Les isomorphismes de  $\mathcal{G}_1$  dans  $\mathcal{G}_1$  sont appelés automorphismes et on note,  $Aut(\mathcal{G}_1)$ , le groupe des automorphismes de  $\mathcal{G}_1$  muni de la loi de composition.

**Remarque** Si on considère un graphe  $\mathcal{G}$  de taille  $n$ , alors le groupe son groupe d'automorphismes,  $Aut(\mathcal{G})$ , est un sous groupe du groupe des permutations  $\mathfrak{S}_n$ .

Dans un soucis de clarification et de simplification du problème, nous introduisons une relation d'équivalence induite par les automorphismes du graphe. Cette relation crée ainsi des classes d'équivalence appelées orbites dont nous allons nous servir par la suite, notamment lors des preuves.

**Définition 2.9.** Soit  $\mathcal{G} = (V, E)$  un graphe. On note  $\mathcal{R}$  la relation d'équivalence sur les noeuds de  $G$  définie de la manière suivante :

$$\forall (x, y) \in V^2 \quad x\mathcal{R}y \Leftrightarrow \exists g \in Aut(\mathcal{G}) \quad g(x) = y$$

Les classes d'équivalence associées à la relation  $\mathcal{R}$  sont appelées "Orbites" du graphe  $\mathcal{G}$ , notées  $\mathcal{O}(x)$ , avec  $x \in V$  un représentant de la classe.

(Une preuve qu'il s'agit bien d'une relation d'équivalence est en appendice *démonstration 2*)

Nous avons rappelé les éléments mathématiques nécessaires à la compréhension des parties suivantes. Avant de nous attarder sur le premier modèle, nous allons présenter un outil utilisé dans le troisième modèle. Le graphe initial du troisième modèle dont nous parlons dans ce rapport résume les informations contenu par les graphes des deux premiers modèles à l'aide du produit cartésien de graphe.

**Définition 2.10.** Soient  $G_1 = (V_1, E_1)$  et  $G_2 = (V_2, E_2)$  deux graphes.

On définit le produit cartésien des graphes  $G_1$  et  $G_2$ , noté  $G_1 \square G_2$ , le graphe dont l'ensemble des noeuds est  $V_1 \times V_2$  et dont l'ensemble des arrêtes est tel que deux noeuds  $u_1u_2$  et  $v_1v_2$  de ce graphe sont adjacents si  $u_1 = v_1$  et  $(u_2, v_2) \in E(G_2)$  ou  $u_2 = v_2$  et  $(u_1, v_1) \in E(G_1)$ .

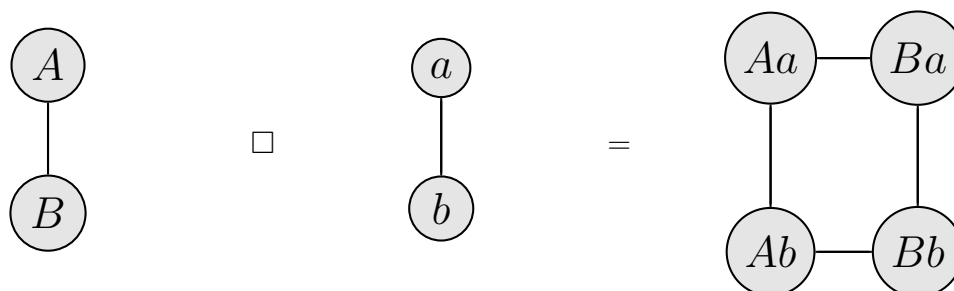
Ainsi,

$$V(G_1 \square G_2) = \{g_1g_2 \mid g_1 \in G_1, g_2 \in G_2\}$$

$$E(G_1 \square G_2) = \{(g_1g_2, g'_1g'_2) \mid g_1 = g'_1, (g_2, g'_2) \in E(G_2) \text{ ou } g_2 = g'_2, (g_1, g'_1) \in E(G_1)\}$$

**Exemple** Afin de mieux comprendre ce produit, voici le produit de deux graphes complets d'ordre deux :





La matrice d'adjacence de ce graphe peut s'obtenir à partir des matrices d'adjacences des graphes dont on fait le produit.

**Proposition 2.11.** *Soient  $G_1$  et  $G_2$  deux graphes de taille respective  $n_1$  et  $n_2$  et de matrice d'adjacence  $A_1$  et  $A_2$ . Alors la matrice d'adjacence du graphe produit  $G_1 \square G_2$ , noté  $A_{1 \square 2}$ , est donnée par  $A_{1 \square 2} = A_1 \otimes I_{n_2} + I_{n_1} \otimes A_2$ , avec  $\otimes$  le produit de Kronecker.*

### 3 Modèles et simulations

#### 3.1 Modèle : île/continent

Considérons  $n$  espèces différentes interagissant entre elles. Ces interactions peuvent être de différents types. Deux espèces  $A$  et  $B$  sont :

- en relation de compétition si la présence de  $A$  à  $t$  a une influence négative sur la présence  $B$  à  $t+1$  et si la présence de  $B$  à  $t$  a une influence négative sur la présence  $A$  à  $t+1$ . La compétition est présente chez les animaux (compétition pour l'espace) ou chez les plantes (à travers l'allélopathie, l'envoi de composés chimiques néfastes)
- en relation trophique si la présence de  $A$  à  $t$  a une influence négative sur la présence  $B$  à  $t+1$  et si la présence de  $B$  à  $t$  a une influence positive sur la présence  $A$  à  $t+1$ .  $A$  est alors le prédateur et  $B$  la proie. Ces relations sont très présentes dans le monde animal.
- en relation mutualiste si la présence de  $A$  à  $t$  a une influence positive sur la présence  $B$  à  $t+1$  et si la présence de  $B$  à  $t$  a une influence positive sur la présence  $A$  à  $t+1$ . Les interactions mutualistes sont présentes afin de fournir aux organismes les services qu'ils ne peuvent pas se procurer eux mêmes. Par exemple la reproduction (interaction plante-pollinisateur) ou l'acquisition de nutriments dans le sol (interaction plantes-champignon)

On cartographie alors les interactions entre espèces au sein d'une communauté en utilisant un graphe dirigé étiqueté; chaque arrête est étiquetée à l'aide d'une fonction valuation  $Weight$  définit de la sorte :  $Weight(u, v) = +1$  si l'interaction de l'espèce  $u$  sur l'espèce  $v$  est positive ou  $-1$  si elle est négative. On suppose de plus qu'une population d'espèce ne peut pas interagir avec elle même (pas de boucle sur le graphe initial).

**Exemple** Prenons l'exemple d'une relation trophique entre trois espèces : l'espèce  $X_1$  est prédateur, et les espèces  $X_2$  et  $X_3$  sont des proies. À l'aide du package `igraph` sur R (Csardi *et al.*, 2006), il est possible de représenter les graphes qui nous intéressent dans ce contexte (cf

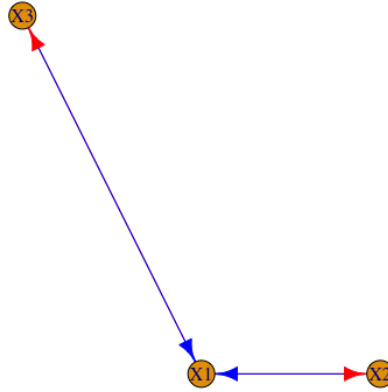


FIGURE 5 – Graphe d'interaction entre trois espèces. Les influences positives sont représentées en bleu et les négatives en rouge.

Figure 5).

Dans le cas du modèle île/continent, on s'intéresse à la présence d'espèces en interaction sur une île. Mais cette dernière n'est pas considérée comme isolée car un réservoir inépuisable d'espèces sur un continent adjacent permet la re-colonisation, avec un taux  $c \in ]0, 1[$ , d'une espèce éteinte sur l'île. Soit  $(\mathbf{X}^t)_{t \in \mathbb{N}}$  la chaîne de Markov en temps discret décrivant la diversité de l'île, son espace d'état est  $Word_{bin}(n)$  de taille  $2^n$ .

À l'aide de la formule de Bayes, nous pouvons calculer les probabilités de passage d'un état à un autre en utilisant les dépendances espèce par espèce. Soit  $S$  et  $S'$  deux états dans  $Word_{bin}(n)$ , tel que  $S = \overline{s_1 s_2 \dots s_n}$  et  $S' = \overline{s'_1 s'_2 \dots s'_n}$ , alors :

$$\mathbb{P}(\mathbf{X}^t = S' \mid \mathbf{X}^{t+1} = S) = \prod_i^n \mathbb{P}(X_i^t = s'_i \mid X_i^{t+1} = s_i, (X_j^t)_{j \in Par(i)})$$

Avec  $Par(i)$  l'ensemble des parents du noeud  $i$ . Ce produit se calcule en fonction des deux probabilités suivantes : la probabilité de réapparition au temps  $t + 1$  d'une espèce disparue au temps  $t$

$$\mathbb{P}(X_i^{t+1} = 1 \mid X_i^t = 0, (X_j^t)_{j \in Par(i)}) = \mathbb{P}(X_i^{t+1} = 1 \mid X_i^t = 0) = c$$

Dans ce modèle la probabilité de colonisation ne dépend pas de l'état des voisins. La probabilité de persistance d'une espèce du temps  $t$  au temps  $t + 1$

$$\mathbb{P}(X_i^{t+1} = 1 \mid X_i^t = 1, (X_j^t)_{j \in Par(i)}) = 1 - e + \frac{\left[1 - (1 - e)^{N_+^{\mathbf{X}^t(i)}}\right] (1 - e)^{N_-^{\mathbf{X}^t(i)}}}{M}$$

Avec  $e \in ]0, 1[$  le taux d'extinction,  $N_+^{\mathbf{X}^t}(i)$  la variable aléatoire indiquant le nombre des parents du noeud  $i$ , présent à l'instant  $t$ , dont l'influence est positive sur l'espèce  $i$ , et  $N_-^{\mathbf{X}^t}(i)$  celle donnant nombre de parents de  $i$ , présent à l'instant  $t$ , dont l'influence est négative et  $M = \frac{1-(1-e)^n}{e}$  une constante permettant à la probabilité de persistance d'être comprise entre 0 et 1.

Cette chaîne de Markov est **homogène**.

**Exemple** Revenons à l'exemple précédent (cf Figure 5), dans ce cas l'espace d'états est de taille  $2^n = 2^3 = 8$  et il est possible de le détailler :

$$X^t \in \{000, 001, 010, 011, 100, 101, 110, 111\}, \forall t \in \mathbb{N}$$

Si le nombre d'espèces étudiées est faible ( $n < 10$ ), alors le nombre d'états reste faible ( $2^n < 1000$ ) et il est possible d'obtenir le noyau de transition de la chaîne de Markov en moins de 10 minutes avec un programme approprié.

**Exemple** Dans le cas que nous étudions, le noyau de transition de cette chaîne de Markov est

	000	001	010	011	100	101	110	111
000	7.290000e-01	0.0810000000	0.0810000000	0.0090000000	0.0810000000	0.0090000000	0.0090000000	0.00100000
001	4.050000e-02	0.7695000000	0.0045000000	0.0855000000	0.0045000000	0.0855000000	0.0005000000	0.00950000
010	4.050000e-02	0.0045000000	0.7695000000	0.0855000000	0.0045000000	0.0005000000	0.0855000000	0.00950000
011	2.250000e-03	0.0427500000	0.0427500000	0.8122500000	0.0002500000	0.0047500000	0.0047500000	0.09025000
100	4.050000e-02	0.0045000000	0.0045000000	0.0005000000	0.7695000000	0.0855000000	0.0855000000	0.00950000
101	1.461218e-03	0.0277631464	0.0001623576	0.003084794	0.043538782	0.827236854	0.004837642	0.09191521
110	1.461218e-03	0.0001623576	0.0277631464	0.003084794	0.043538782	0.004837642	0.827236854	0.09191521
111	3.954864e-05	0.0007514242	0.0007514242	0.014277060	0.002460451	0.046748576	0.046748576	0.88822294

Avec  $c = 0, 1$  et  $e = 0, 05$ .

On remarque alors que notre cette matrice de transition est strictement positive.

Le fait qu'un continent puisse alimenter en espèce l'île voisine, empêche l'extinction définitive de la vie sur l'île. Ainsi l'état 000..00 n'est pas absorbant dans ce modèle. De plus les chaînes de Markov étudiées sont **irréductibles** et **apériodiques**; en effet tous les coefficients de la matrice de transition sont strictement positifs (*démonstration* 3.1 en appendice).

**Exemple** À présent, comme nous avons le noyau de transition de notre chaîne de Markov, il est possible d'obtenir sa mesure invariante.

000	001	010	011	100	101	110	111
0.02968645	0.04998088	0.04998088	0.08049733	0.08142466	0.1722413	0.1722413	0.3639471

On remarque que certaines valeurs de la stationnaire sont les mêmes. Nous verrons par la suite en quoi cette information peut nous permettre de réduire le temps de calcul lorsque la taille du graphe est grande.

La mesure invariante nous indique la probabilité de chaque assemblage d'espèce dans l'état stationnaire. Plusieurs autres exemples sont visibles dans la partie 7 de ce rapport. Lorsque la taille du graphe initial est grande, on passe par une simulation qui nous donne un résultat approximatif de la distribution stationnaire. Pour le modèle île/continent, on se contente de simuler un grand nombre de trajectoires durant un temps arbitrairement grand et on moyenne après avoir choisis un temps de mélange (arbitrairement grand).

### 3.2 Modèle méta-population

À la différence du premier modèle, ici ce ne sont pas  $n$  espèces que nous étudions mais une seule. Seulement, il est possible que des populations de cette espèce colonisent différents sites (lieux, emplacements) et s'éteignent. Considérons  $n$  sites géographiques et une espèce, alors les noeuds du graphe initial correspondent aux sites et les arêtes sont les chemins que peuvent emprunter l'espèce pour se rendre d'un site à l'autre. Le but sera de trouver, à partir du graphe expliquant la situation géographique, la dynamique en temps long. Le graphe initial est non-orienté, on considère qu'un chemin peut être pris dans les deux sens. Prenons l'exemple d'un graphe initial à quatre noeuds.

**Exemple** En figure 6, nous avons un exemple de graphe où les trois premiers noeuds communiquent entre eux, mais le quatrième est atteignable par le deuxième site seulement.

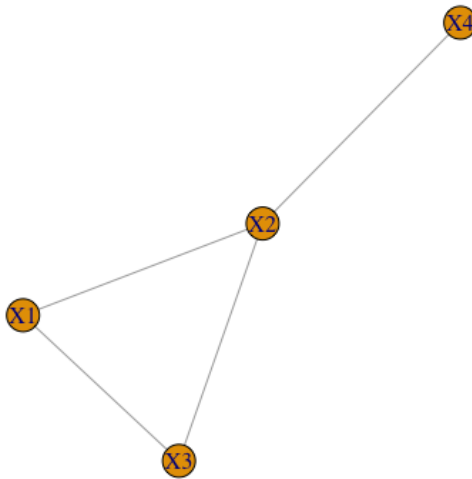


FIGURE 6 – Exemple de graphe initial, modèle méta-population

Introduisons la chaîne de Markov  $(\mathbf{X}^t)_t$  décrivant les présences et absences d'une population de l'espèce étudiée sur les différents sites. L'espace d'état de cette chaîne est  $Word_{bin}(n)$ , et si  $X_i^t = 1$  cela signifie que l'espèce est présente sur le  $i^{\text{ème}}$  site géographique, sinon c'est que l'espèce est absente.

La probabilité qu'un site reste occupé dépend d'une constante  $e \in ]0, 1[$ , et la probabilité qu'un site soit colonisé dépend du nombre de voisin et d'une constante de colonisation  $c \in ]0, 1[$ .

$$\mathbb{P}(X_i^{t+1} = 1 \mid X_i^t = 1, (X_j^t)_{j \in Par(i)}) = \mathbb{P}(X_i^{t+1} = 1 \mid X_i^t = 1) = 1 - e$$

$$\mathbb{P}(X_i^{t+1} = 1 \mid X_i^t = 0, (X_j^t)_{j \in Par(i)}) = 1 - (1 - c)^{N^{\mathbf{X}^t}(i)}$$

Avec  $Par(i)$  l'ensemble des noeuds parents du noeud  $i$ ; notons que, dans le cas de ce modèle, le graphe initial est non-orienté, cela signifie que tous les voisins du noeud  $i$  sont aussi des parents du noeud  $i$ .  $N^{\mathbf{X}^t}(i)$  est le nombre de voisins présent autour du site  $i$  sachant l'état de

la chaîne à l'instant  $t$ ,  $N^{\mathbf{X}^t}(i) \in \llbracket 0, n-1 \rrbracket \forall i \in \llbracket 1, n \rrbracket$ . La chaîne de Markov est **homogène**. On remarque que si le nombre de voisins d'un site est nul, alors il est impossible de coloniser le site ; cela implique que l'état 000..00 est absorbant. Les autres états communiquent entre eux, de plus ils sont susceptibles d'entraîner l'absorption avec probabilité non nul. Si nous cherchons à obtenir la distribution stationnaire de cette chaîne de Markov nous obtiendrons l'extinction de l'espèce presque sûrement. Nous allons plutôt concentrer nos recherches sur la distribution limite du processus conditionné par la non-absorption : la distribution quasi-stationnaire.

Pour l'obtenir, nous considérons le processus aléatoire sur l'espace  $Word_{bin}(n) \setminus \{000..00\}$ , dont les probabilités de passage d'un état à un autre sont les mêmes que pour la chaîne de Markov  $(\mathbf{X}^t)_t$ . Les coefficients de la matrice de passage, de taille  $2^n - 1 \times 2^n - 1$ , sont donc les mêmes que ceux du noyau de transition de la chaîne  $(\mathbf{X}^t)_t$  (privé de la ligne et de la colonne de l'état absorbant). Dans van Doorn & Pollett (2013), ce processus est appelé sous-chaîne de Markov et dans le cas irréductible (partie 3 du papier), le vecteur propre à gauche (renormalisé) de sa matrice de transition est une distribution quasi-stationnaire de la chaîne de Markov  $(\mathbf{X}^t)_t$  dont il est issu. Enfin l'espace d'état étant fini, cette quasi-stationnaire est unique Darroch & Seneta (1965)

**Exemple** Décrivons l'espace d'état  $Word_{bin}(4)$ , soit  $t \in \mathbb{N}^*$

$\mathbf{X}^t \in \{0000, 0001, 0010, 0100, 1000, 0011, 0101, 1001, 0110, 1010, 1100, 0111, 1011, 1101, 1110, 1111\}$

Le noyau de transition de cette chaîne de Markov est donc de taille  $16 \times 16$ , il est encore possible de l'afficher sans que cela soit trop encombrant :

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0001	0.050000	0.450000	0.000000	0.000000	0.050000	0.450000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0010	0.025000	0.000000	0.225000	0.000000	0.025000	0.000000	0.225000	0.000000	0.025000	0.000000	0.225000	0.000000	0.025000	0.000000	0.225000	0.000000
0011	0.001250	0.011250	0.011250	0.101250	0.003750	0.033750	0.033750	0.303750	0.001250	0.011250	0.011250	0.101250	0.003750	0.033750	0.033750	0.303750
0100	0.012500	0.012500	0.012500	0.012500	0.112500	0.112500	0.112500	0.112500	0.012500	0.012500	0.012500	0.012500	0.112500	0.112500	0.112500	0.112500
0101	0.002500	0.022500	0.002500	0.022500	0.022500	0.202500	0.022500	0.202500	0.002500	0.022500	0.002500	0.022500	0.022500	0.202500	0.022500	0.202500
0110	0.001250	0.001250	0.011250	0.011250	0.011250	0.011250	0.101250	0.101250	0.003750	0.003750	0.033750	0.033750	0.033750	0.033750	0.303750	0.303750
0111	0.000250	0.002250	0.002250	0.020250	0.002250	0.020250	0.020250	0.182250	0.000750	0.006750	0.006750	0.006750	0.006750	0.060750	0.060750	0.546750
1000	0.025000	0.000000	0.025000	0.000000	0.025000	0.000000	0.000000	0.225000	0.000000	0.000000	0.000000	0.000000	0.225000	0.000000	0.225000	0.000000
1001	0.001250	0.011250	0.001250	0.011250	0.003750	0.033750	0.033750	0.033750	0.011250	0.101250	0.011250	0.101250	0.033750	0.033750	0.033750	0.303750
1010	0.002500	0.000000	0.022500	0.000000	0.007500	0.000000	0.067500	0.000000	0.022500	0.000000	0.202500	0.000000	0.067500	0.000000	0.607500	0.000000
1011	0.000125	0.001125	0.001125	0.010125	0.000875	0.007875	0.007875	0.070875	0.001125	0.010125	0.010125	0.091125	0.007875	0.070875	0.070875	0.637875
1100	0.001250	0.001250	0.003750	0.003750	0.011250	0.011250	0.033750	0.033750	0.011250	0.011250	0.033750	0.033750	0.101250	0.101250	0.303750	0.303750
1101	0.000250	0.002250	0.000750	0.006750	0.002250	0.020250	0.006750	0.060750	0.002250	0.020250	0.006750	0.060750	0.020250	0.182250	0.060750	0.546750
1110	0.000500	0.000500	0.004500	0.004500	0.004500	0.004500	0.040500	0.040500	0.004500	0.004500	0.040500	0.040500	0.040500	0.040500	0.364500	0.364500
1111	0.000100	0.000900	0.000900	0.008100	0.000900	0.008100	0.008100	0.072900	0.000900	0.008100	0.008100	0.072900	0.008100	0.072900	0.072900	0.656100

La prochaine étape est de retirer la ligne et la colonne 0000, pour pouvoir trouver le vecteur propre à gauche. Puis il suffit de le renormaliser pour obtenir la distribution quasi-stationnaire de la chaîne :

	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101
0001	0.002924731	0.002959364	0.009537205	0.003023174	0.014353309	0.01747036	0.07808523	0.002959364	0.009537205	0.01762048	0.06454351	0.01747036	0.07808523
0010													
0011													
0100													
0101													
0110													
0111													
1000													
1001													
1010													
1011													
1100													
1101													
1110													
1111													
0.1239049													
0.5575258													

Pour simuler une quasi-stationnaire, il est possible de ré-utiliser une méthode de lancer de trajectoires classique comme dans la partie précédente. Malheureusement, dans le modèle méta-population, lorsqu'une trajectoire atteint l'état 000..00, elle est absorbée. Une solution

est de tirer beaucoup de trajectoires, d'attendre un certain temps de mélange, puis de garder les trajectoires qui ne se sont pas éteintes pour approximer la distribution quasi-stationnaire. Cette méthode est dépendante de la valeur du coefficient d'extinction, si de dernier est un peu trop élevé alors la plupart des trajectoires (voir toutes) vont être absorbées avant convergence. De ce fait, les résultats obtenus avec la méthode classique ne sont pas réguliers car ils dépendent beaucoup trop des paramètres.

Nous avons décidé d'utiliser une autre méthode de simulation : le système de Fleming-Viot (Thai (2015)). Il s'agit, ici encore, de lancer un grand nombre de trajectoires durant une longue durée. Mais si une de ces trajectoires simulées est absorbée à un instant  $t$  alors, on remplace sa valeur par celle d'une autre trajectoire non-absorbée à l'instant  $t$  (le choix de la trajectoire non-absorbée est aléatoire). Si la probabilité d'extinction est trop forte pour la méthode classique, utiliser un système de Fleming-Viot est une bonne alternative. Plus généralement, le système de Fleming-Viot donne des résultats plus réguliers que la méthode classique, car on est certain que les trajectoires ont continué leur vie jusqu'au temps de calcul souhaité.

Un ensemble d'exemples, concernant ce modèle, est présenté en partie 7 de ce rapport.

### 3.3 Modèle méta-communauté

On peut résumer les informations des deux modèles précédents en un seul. On étudie ainsi les dynamiques de communautés d'espèces (plusieurs populations en interaction) sur un graphe spatial : c'est le modèle méta-communauté. Le graphe initial est le produit cartésien de deux graphes : un graphe d'interaction (carte des relations biotiques entre les espèces) et un graphe spatial (carte géographique des lieux colonisables par les espèces).

Afin que la chaîne de Markov associé à ce problème soit irréductible, il faut considérer qu'il y a un réservoir d'individus inépuisable pouvant coloniser chaque lieu possible. Si ce n'est pas la cas, alors beaucoup d'états ne peuvent pas communiquer entre eux. Ce modèle est donc du type île/continent : on étudie la diversité sur tous les lieux géographiques d'une île avec possibilité de recolonisation en provenance d'un continent.

Soient  $(n_{int}, n_s) \in \mathbb{N}^2$ ,  $\mathcal{G}_{int} = (V_{int}, E_{int})$  le graphe d'interaction entre  $n_{int}$  espèces et  $\mathcal{G}_s = (V_s, E_s)$  le graphe spatial indiquant les  $n_s$  lieux colonisables et les chemins d'accès entre eux. Le graphe initial est le produit cartésien de ces deux graphes, noté  $\mathcal{G}_{prod}$ . On considère  $(\mathbf{X}^t)_t$  la chaîne de Markov issu de la dynamique étudiée. Les noeuds du graphe produit sont des éléments de  $V_{int} \times V_s$ , de ce fait, l'espace d'état n'est pas l'ensemble des mots de taille le nombre de noeuds du graphe, mais l'ensemble des matrices composées de 0 et de 1 de taille  $n_{int} * n_s$ , ainsi nous pouvons écrire  $\forall t \in \mathbb{N}^*$ ,  $\mathbf{X}^t = (X_{i,j}^t)_{i,j}$ , avec  $\forall (i, j) \in \llbracket 1, n_{int} \rrbracket \times \llbracket 1, n_s \rrbracket$ ,  $X_{i,j}^t$  une variable aléatoire dans  $\{0, 1\}$ . Cela implique que désormais, la taille de l'espace d'état est de  $2^{n_{int} * n_s}$ .

**Exemple** Prenons un exemple simple pour mieux comprendre le modèle. Choisissons un graphe d'interaction de deux espèces avec un rapport proie/prédateur (figure 7), puis un graphe spatial complet de taille trois (figure 8). Le graphe initial produit (figure 9) nous montre toutes les combinaisons possibles d'interaction entre ces deux espèces sur tous les sites géographiques.



FIGURE 7 – Graphe d'interaction type proie (espèce E2) / prédateur (espèce E1)

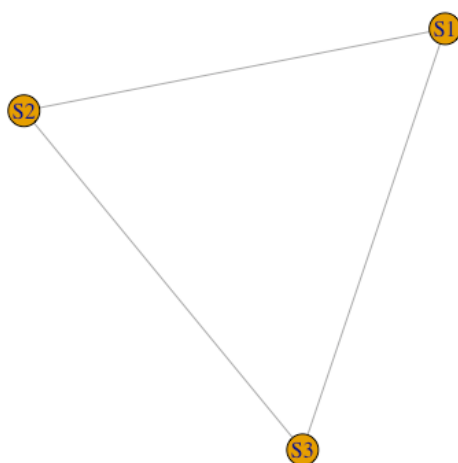


FIGURE 8 – Graphe spatial sur lequel les espèces peuvent se déplacer

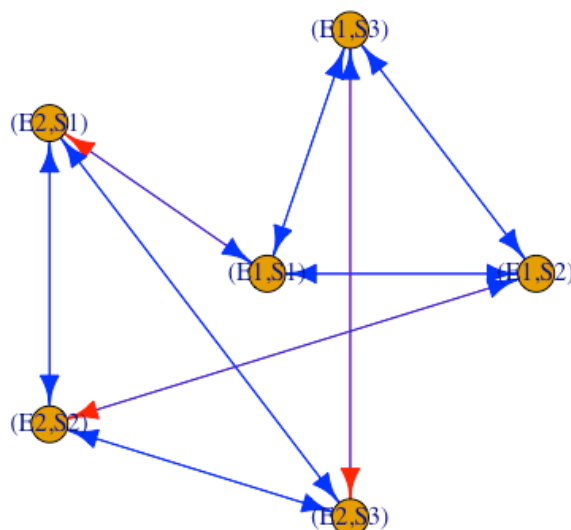


FIGURE 9 – Graphe initial produit des deux graphes précédents

Sur le graphe produit nous pouvons par exemple lire que, le fait que l'espèce E2 soit présente sur le site S1 a une influence positive sur la présence de l'espèce E1 sur le site S1. Dans ce cas les états de la chaîne de Markov sont tous les arrangements de ces deux espèces sur les trois sites. Si nous choisissons que les lignes représentent les espèces et les colonnes les sites, alors  $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$  est une état possible de la chaîne de Markov dans lequel l'espèce E1 est présente sur les sites S1 et S2, et l'espèce E2 est présente sur les site S2 et S3.

Soient  $t \in \mathbb{N}^*$ ,  $i \in \{1, 2, \dots, n_{int}\}$  et  $j \in \{1, 2, \dots, n_s\}$ , l'indice  $i$  correspond aux espèces et l'indice  $j$  aux sites. Considérons les trois paramètres du modèle  $(c, e, \epsilon) \in ]0, 1[^3$ , avec  $\epsilon$  la probabilité de colonisation d'une espèce sur un site depuis le continent extérieur,  $c$  la probabilité de colonisation d'un site par une espèce depuis l'intérieur de l'île et  $e$  la probabilité d'extinction d'une espèce sur un site. Les probabilités de passage pour chaque coefficient de la chaîne de Markov sont :

$$\mathbb{P}(X_{i,j}^{t+1} = 1 \mid X_{i,j}^t = 0, (X_{i,k})_{i,k \in Par(j)}) = \epsilon + (1 - \epsilon) \left[ 1 - (1 - c)^{\sum_{k \in Par(j)} X_{i,k}} \right]$$

$$\mathbb{P}(X_{i,j}^{t+1} = 1 \mid X_{i,j}^t = 1, (X_{k,j})_{k \in Par(i),j}) = 1 - e + \frac{\left[ 1 - (1 - e)^{N_{j,+}^{\mathbf{X}^t}(i)} \right] (1 - e)^{N_{j,-}^{\mathbf{X}^t}(i)}}{M}$$

Avec  $N_{j,+}^{\mathbf{X}^t}(i)$  la variable aléatoire indiquant le nombre des parents du noeud  $i$  sur le graphe d'interaction, présent à l'instant  $t$  sur le site  $j$ , dont l'influence est positive sur l'espèce  $i$ , et  $N_{j,-}^{\mathbf{X}^t}(i)$  celle donnant nombre de parents de  $i$ , présent à l'instant  $t$  sur le site  $j$ , dont l'influence est négative et  $M = \frac{1 - (1 - e)^{n_{int}}}{e}$  une constante permettant à la probabilité de persistance d'être comprise entre 0 et 1. La chaîne de Markov que nous étudions est homogène, irréductible et apériodique.

À stade nous comprenons qu'avec ce nouveau modèle, les problèmes rencontrés dans les parties précédentes sont accentués. Dans notre exemple de deux espèces sur un graphe spatial de taille trois, nous obtenons un espace d'état de taille  $2^{3*2} = 64$ . Nous avons trouvé des dynamiques, avec et sans l'utilisation de simulations sur R, et ces dernières sont de très grande taille. Il faut trouver une manière de réduire les calculs de manière drastique. Dans la partie suivante nous proposons de passer par des équivalences sur les graphes pour arriver à nos fins.

## 4 Réduction du problème

### 4.1 Notions d'équivalence

Nous avons pu observer, dans les dynamiques trouvées sur les exemples précédents, que certaines valeurs sur les stationnaires ou les quasi-stationnaires sont redondantes. Dans cette partie nous cherchons à trouver le nombre de valeurs distinctes sur la stationnaire seulement à partir du choix du modèle et du graphe initial. Le modèle méta-communauté étant bien



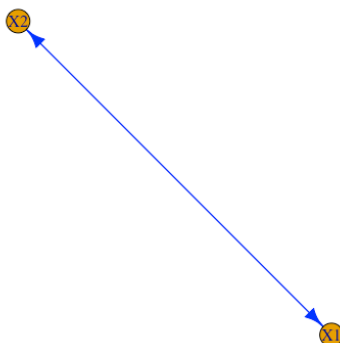


FIGURE 10 – Deux espèces en interaction mutualiste

plus complexe, nous raisonnons sur les deux premiers modèles uniquement (île/continent et méta-population) tout au long de la partie 4.

Nous allons détailler un exemple simple (deux espèces en interaction mutualistes) sur le modèle île/continent :

**Exemple** Le graphe initial est en figure 10. Le graphe est un graphe complet d'ordre 2, la permutation échangeant, les noeuds  $X1$  et  $X2$ , laisse le graphe inchangé ; ces deux espèces jouent un rôle équivalent dans le réseau d'interaction. Dans ce cas le résultat que l'on obtient est le suivant :

00	01	10	11
0.07441383	0.1530111	0.1530111	0.6195639

La valeur 10 et la valeur 01 sont les mêmes. Pour comprendre ce résultat, regardons le noyau de transition de la chaîne ainsi que le graphe d'état (figure 11) :

	00	01	10	11
00	0.810000000	0.09000000	0.09000000	0.0100000
01	0.045000000	0.85500000	0.00500000	0.0950000
10	0.045000000	0.00500000	0.85500000	0.0950000
11	0.0005933596	0.02376561	0.02376561	0.9518754

On remarque que le graphe d'état en figure 11 possède trois orbites :  $\{00\}$ ,  $\{11\}$ ,  $\{01, 10\}$  (cf définition 2.9). En effet la permutation, échangeant les noeuds 01 et 10, laisse le graphe inchangé.

C'est le nombre d'orbites sur le graphe d'état qui caractérise le nombre de valeurs différentes sur la stationnaire (ou la quasi-stationnaire). La démonstration de ce résultat est visible en appendice (*demonstration* 4.1.1). La difficulté principale est de trouver ce nombre d'orbites lorsque le graphe est grand. Le trouver revient à chercher tous les isomorphismes d'un graphe, or le problème de l'isomorphisme de graphe est l'un des problèmes de NP pour lequel nous n'avons pas d'algorithme pouvant nous aider en temps raisonnable (en

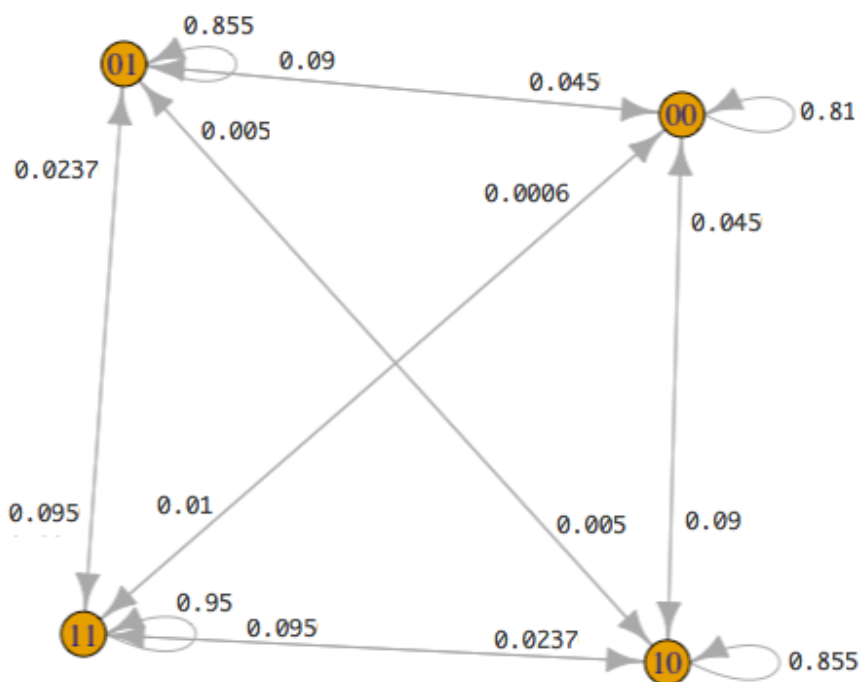


FIGURE 11 –

temps polynomial). Néanmoins, nous avons réussi à trouver le nombre d'orbites du graphe d'état pour certains types de graphes initiaux et dans le cas des deux premiers modèles présentés. Avant de présenter nos résultats, rappelons que la richesse d'un état correspond au nombre de 1 composant le mot (c'est le nombre d'espèces présentes dans l'arrangement d'espèces choisi, ou le nombre de sites occupés).

Si le graphe initial est complet, de taille  $n$ , avec interaction mutualiste uniquement (pour le modèle île/continent), alors pour nos deux premiers modèles, les états présents dans une même orbite du graphe d'état sont tous de même richesse (cette propriété caractérise les orbites du graphe d'état dans ce cas, *démonstration* 4.1.2 en appendice). Or le graphe initial est de taille  $n$ , il y a donc  $n + 1$  richesses d'état possibles. Nous pouvons en déduire que dans le cas du modèle île/continent le nombre de valeurs différentes sur la distribution stationnaire est :  $n + 1$ . Si on garde les mêmes hypothèses sur le graphe initial mais qu'il s'agit désormais du modèle méta-population, le nombre de valeurs différentes sur la quasi-stationnaire est :  $n$ . Nous passons, dans ces cas là, de  $2^n$  valeurs inconnues à  $n$  ou  $n + 1$ .

Dans un graphe complet, tous les noeuds sont dans la même orbite, cela signifie que les espèces placées dans le graphe initial sont interchangeables. Une conséquence de ce phénomène se retrouve dans le résultat précédent, c'est uniquement le nombre de ces espèces (ou sites) présentes qui influe sur le nombre de valeurs différentes dans la stationnaire (ou la quasi-stationnaire).

Maintenant, considérons que le graphe initial dans  $\mathbf{K}_{\underbrace{n_1 \dots n_1}_{r_1 \text{ fois}} \underbrace{n_2 \dots n_2}_{r_2 \text{ fois}} \dots \underbrace{n_p \dots n_p}_{r_p \text{ fois}}}$ , l'ensemble des graphes multipartis complets, avec  $p \in \mathbb{N}^*$  et  $\forall i \in \{1, \dots, p\}$ ,  $(n_i, r_i) \in (\mathbb{N}^*)^2$ . De plus, considérons qu'il s'agit d'un réseau d'interaction mutualiste de taille  $n$ . Le nombre de valeurs différentes sur la stationnaire est toujours égale au nombre d'orbites du graphe d'état, seulement cette fois-ci ces orbites sont caractérisées par le nombre de richesses différentes que l'on peut obtenir avec les états de chaque sous-graphe appartenant à  $\mathbf{K}_{\underbrace{n_i \dots n_i}_{r_i \text{ fois}}} \forall i \in \{1, \dots, p\}$ , du

graphe initial (ce sous graphe est de taille  $r_i \times n_i$ ). Or ce nombre est calculable (*démonstration* 4.1.3 en appendice), nous obtenons le résultat suivant :

Dans le cadre du modèle île/continent, le nombre de valeurs différentes sur la stationnaire est  $\prod_{i=1}^p \binom{n_i+r_i}{r_i}$ . Avec les mêmes hypothèses sur le graphe initial, dans le cadre du modèle méta-

population, le nombre de valeurs différentes sur la quasi-stationnaire est  $\left[ \prod_{i=1}^p \binom{n_i+r_i}{r_i} \right] - 1$ .

Le nombre d'orbites d'un graphe multiparti complet dans  $\mathbf{K}_{\underbrace{n_1 \dots n_1}_{r_1 \text{ fois}} \underbrace{n_2 \dots n_2}_{r_2 \text{ fois}} \dots \underbrace{n_p \dots n_p}_{r_p \text{ fois}}}$  est  $p$ , ici

encore cela a une influence direct sur le résultat précédent. Les présences et absences des espèces (ou le fait que des sites soient occupés ou non) dans une même orbite du graphe initial vont influencer le résultat sur la dynamique de la même manière, car ces espèces (ou sites) sont interchangeable sur le graphe initial.

On peut questionner le réalisme de cette hypothèse concernant la structure du réseau d'interaction. Si elle est probablement peu réaliste chez les vertébrés, elle est plus appropriée dans les communautés présentes dans le sol. En effet, en ce qui concerne les interactions mutualistes-plantes champignons, les interactions sont non-spécifiques ((Peay *et al.*, 2015)). Ainsi, des groupes de champignons peuvent interagir de façon équivalente avec les plantes, ce qui justifie l'emploi d'un graphe multiparti complet pour modéliser ce réseau d'interaction. Maintenant que nous arrivons à prévoir le nombre d'inconnues sur la dynamique dans certains cas, nous présentons une manière de les trouver efficacement dans la partie suivante.

## 4.2 Agrégation matricielle

Dans cette partie, nous nous concentrons sur le modèle île/continent et méta-population dont le graphe initial est mutualiste (dans le cas du modèle île/continent), complet ou multiparti complet.

Nous connaissons le nombre de valeurs distinctes sur la stationnaire pour ces types de graphe. Lorsque  $n$  est grand et que la matrice de transition, de taille  $2^n \times 2^n$ , n'est pas calculable, nous aimerions passer par une matrice plus petite dont les lignes et les colonnes ne correspondent plus aux états mais aux orbites du graphe d'états. Dans l'idéal, si nous avons la liste de toutes les orbites, nous créons une matrice dont les colonnes sont la somme des colonnes des états dans la même orbite, et dont les lignes sont les moyennes des lignes des états appartenant à la même orbite. Cette matrice donne les probabilité de passage d'une orbite à une autre. Les valeurs de la dynamique issu de cette nouvelle matrice stochastique, dite

”matrice agrégée”, sont uniques et donne le résultat de la dynamique orbite par orbite. Les valeurs sont donc toute unique dans cette nouvelle dynamiques (sauf dans le cas de certains paramétrages pouvant entraîner quelques égalités entre valeurs de probabilité pour certaines orbites, cf. *démonstration* 4.2).

Dans le cas d’un graphe initial complet, lister les orbites est aisé car elles sont caractérisées par la richesse des états qu’elles contiennent. Dans le cas d’un multipartite c’est plus compliqué, surtout si plusieurs des sous-ensembles partitionnant l’ensemble des noeuds ont le même cardinal.

Nous nous sommes concentrés sur le cas du modèle île/continent pour appliquer le résultat précédent. Nous avons codé une première fonction pour le cas complet prenant en argument le graphe initial et les constantes  $c$  et  $e$  et donnant en sortie la distribution stationnaire ”réduite”. La méthode de la matrice agrégée implique l’utilisation du noyau de transition de taille  $2^n \times 2^n$ , pour passer outre ce problème, nous calculons orbite par orbite les probabilités de passage de tous les états, qui les compose, vers les autres états. Deux exemples concernant le cas complet sont visibles en appendice.

Dans le cas multipartite nous avons du coder une autre fonction car les orbites ne sont pas caractérisées de la même manière que dans le cas complet. Lorsque le graphe initial multipartite est dans  $K_{n_1 n_2 \dots n_p}$ , avec  $p \in \mathbb{N}^*$  et  $\forall i \in \{1, 2, \dots, p\} n_i \in \mathbb{N}^*$  tel que les  $n_i$  soient tous différents entre eux ; alors les orbites sont caractérisées par les différentes richesses que l’on peut obtenir sur chaque sous-ensemble de noeuds. Mais si au moins deux des ces sous-ensembles sont de même taille, alors le nombre d’orbite est réduit car caractérisé par les différentes richesses que l’on peut obtenir sur les sous-ensembles de taille différente. Ces orbites sont plus difficiles à déterminer et la fonction codé ne prends pas en compte cette particularité. Elle définit les orbites uniquement en fonction des automorphismes échangeant seulement deux sommets du graphe d’état, or dans le cas où deux sous-ensembles de noeuds ont la même taille, la permutation échangeant tous les noeuds de ces deux sous-ensembles est un automorphisme définissant une orbite bien particulière. Deux exemples concernant le cas multipartite sont visibles en appendice pour présenter le phénomène.

## 5 Conclusion, discussion

Nous nous sommes intéressés à l’étude de la dynamique stochastique de plusieurs modèles dont les paramètres étaient des graphes, pouvant représenter un réseau spatial, un réseau d’interaction ou une combinaison des deux. À cause de la trop grande taille de l’espace d’état de la chaîne, calculer la distribution stationnaire pour un nombre de sites ou d’espèces réaliste est impossible. Néanmoins, en faisant des hypothèses sur la structure du graphe initial, on peut, en exploitant des symétries, réduire la dimension du problème. Puis, grâce à ce que nous appelons la méthode de la matrice agrégée, nous pouvons avoir accès à des dynamiques qui nous sont inaccessibles. Nous nous sommes rendu compte que les espèces équivalentes (en terme d’orbite du graphe initial), influent de la même manière sur la dynamique, leur destin en temps long est le même. Un tel résultat a des implications en écologie dans la mesure où quelque soit la distribution initiale d’assemblage d’espèces, en temps long, les assemblages équivalents auront même probabilité d’apparition.

Reste à s'intéresser au modèle de méta-communauté dont la complexité dépasse celle des deux premiers modèles, et trouver la aussi, des équivalences pour permettre une réduction du problème. Une première manière de procéder serait d'étudier le produit de deux graphes complets, nous savons que le résultat sera ce qu'on appelle un rook's graphe (graphe des déplacements possibles d'une tour sur un échiquier) dont la régularité doit permettre de trouver des équivalences.

Tout au long de ce travail, nous supposons qu'il y a une hétérogénéité dans la façon dont les espèces ou les sites se connectent (pas nécessairement un graphe complet) mais pas dans la valeur des paramètres  $c$  et  $e$  d'un site ou d'une espèce à l'autre. Une telle hypothèse conduit à des dynamiques de type source-puits (e.g. Bansaye & Lambert 2013). Enfin pour aller plus loin, il est aussi possible de considérer que les graphes d'interactions sont tirés de manière aléatoire (graphes d'Erdős Renyi), et nous pouvons alors étudier les résultats obtenu sur la moyenne de ces graphes. Si les arêtes de ces graphes ont toutes la même probabilité d'existence, alors le graphe moyen sera un graphe complet et il est alors possible d'utiliser les résultats que nous avons obtenu dans ce rapport.

## 6 Références

### Références

- Auclair, E., Peyrard, N. & Sabbadin, R. (2017). Labeled dbn learning with community structure knowledge. In : *Joint european conference on machine learning and knowledge discovery in databases*. Springer pp. 158–174.
- Bansaye, V. & Lambert, A. (2013). New approaches to source–sink metapopulations decoupling demography and dispersal. *Theoretical population biology* 88, 31–46.
- Csardi, G., Nepusz, T. *et al.* (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695, 1–9.
- Dale, M. & Fortin, M.J. (2010). From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics* 41.
- Darroch, J.N. & Seneta, E. (1965). On quasi-stationary distributions in absorbing discrete-time finite markov chains. *Journal of Applied Probability* 2, 88–100.
- Elton, C.S. (1927). *Animal ecology*. New York, Macmillan Co.,  
<https://www.biodiversitylibrary.org/bibliography/7435>.
- Hanski, I. & Ovaskainen, O. (2003). Metapopulation theory for fragmented landscapes. *Theoretical population biology* 64, 119–127.
- Lähdesmäki, H. & Shmulevich, I. (2008). Learning the structure of dynamic bayesian networks from time series and steady state measurements. *Machine Learning* 71, 185–217.

- Lauri, J. & Scapellato, R. (2016). *Topics in graph automorphisms and reconstruction*. vol. 432. Cambridge University Press.
- Lindeman, R.L. (1942). The trophic-dynamic aspect of ecology. *Ecology* 23, 399–417.
- Luczkovich, J.J., Borgatti, S.P., Johnson, J.C. & Everett, M.G. (2003). Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of Theoretical Biology* 220, 303 – 321.
- Peay, K.G., Russo, S.E., McGuire, K.L., Lim, Z., Chan, J.P., Tan, S. & Davies, S.J. (2015). Lack of host specificity leads to independent assortment of dipterocarps and ectomycorrhizal fungi across a soil fertility gradient. *Ecology Letters* 18, 807–816.
- Thai, A.T.M.N. (2015). *Processus de Fleming-Viot, distributions quasi-stationnaires et marches aléatoires en interaction de type champ moyen*. Ph.D. thesis Université Paris-Est.
- van Doorn, E. & Pollett, P. (2013). Quasi-stationary distributions for discrete-state models. *European journal of operational research* 230, 1–14. Eemcs-eprint-23360.

## 7 Appendice

### 7.1 Codes, simulations et exemples

Tous les exemples présentés dans cette partie ont été réalisés sur R avec les packages `igraph` et `R.utils` et les fonctions codées présentes sur les fichiers R joints avec le rapport. Il y a trois fichiers, un par modèle. La méthode de la matrice agrégée est employée dans le fichier du modèle île/continent. Dans chaque fichier il y a une partie pour obtenir la dynamique avec la fonction : `get_invar_nomdumodèle` (`invar` pour la mesure invariante et `qs` dans le cas de la mesure quasi-stationnaire). Il y a aussi une partie simulation où il est possible d’obtenir les dynamiques avec les fonctions `get_invar_simu_nomdumodèle`.

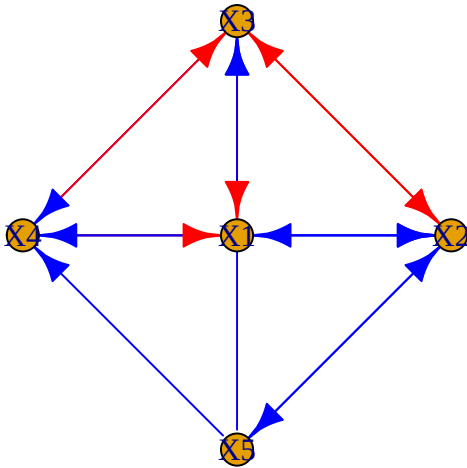
# Ensemble d'exemples sur le modèle île/continent

## Exemple 1

Les paramètres :

## [1] "c = 0.2"

## [1] "e = 0.05"



La matrice de transition faisant 4 pages, nous indiquons uniquement la distribution stationnaire :

```
##          00000      00001      00010      00011      00100
## [1,] 0.0001177335 0.000386169 0.0006156091 0.002269687 0.0005181597
##          00101      00110      00111      01000      01001
## [1,] 0.001887094 0.00312412 0.01296828 0.0005833761 0.002411449
##          01010      01011      01100      01101      01110      01111
## [1,] 0.003465391 0.0165486 0.002545579 0.01156673 0.01752426 0.09336976
##          10000      10001      10010      10011      10100
## [1,] 0.0003829171 0.00120574 0.002246811 0.007997112 0.001867846
##          10101      10110      10111      11000      11001      11010
## [1,] 0.006582258 0.01281608 0.05204326 0.002380209 0.009631529 0.01624586
##          11011      11100      11101      11110      11111
## [1,] 0.07712113 0.01135941 0.05064948 0.09115404 0.4864143
```

Regardons ce que nous donne le résultat à partir d'une simulation et comparons le à la valeur obtenue précédemment. Les choix des paramètres pour la simulation sont les suivants : 10000 itérations au total et un temps de mélange de 1000 itérations.

```
##          00000      00001      00010      00011      00100
## [1,] 0.0001110988 0.0002221975 0.0004443951 0.002110877 0.001999778
##          00101      00110      00111      01000      01001      01010
## [1,] 0.001222086 0.00266637 0.01322075 0.001444284 0.002333074 0.005666037
##          01011      01100      01101      01110      01111      10000
## [1,] 0.01966448 0.003999556 0.01277636 0.01688701 0.09498945 0.0004443951
##          10001      10010      10011      10100      10101
## [1,] 0.0007776914 0.001999778 0.006221531 0.00266637 0.007332519
##          10110      10111      11000      11001      11010      11011
## [1,] 0.008332408 0.04932785 0.003888457 0.01510943 0.02066437 0.07965782
```

```
##          11100          11101          11110          11111
## [1,] 0.01122098 0.05143873 0.09776691 0.463393
```

Calcul de l'erreur quadratique :

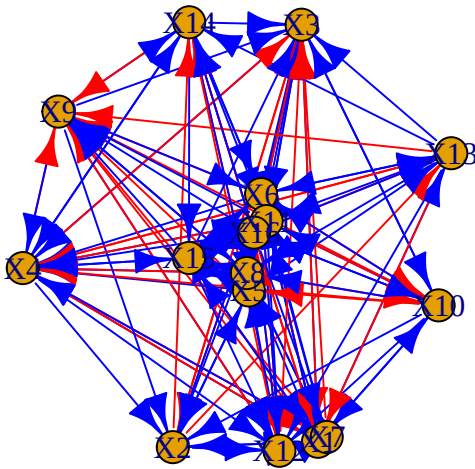
```
## [1] 0.02625971
```

## Exemple 2

Prenons cette fois ci un graphe initial trop grand pour que l'on puisse calculer la dynamique induite par théorie spectrale (trop grand car une matrice de taille  $2^{16} \times 2^{16}$  pèse plus de 10 Gigabytes):

```
## [1] "c = 0.2"
```

```
## [1] "e = 0.05"
```



Même si l'on veut simuler le résultat, plusieurs problèmes se posent. Le vecteur que nous cherchons possède  $2^{16} = 65536$  composantes, l'afficher dans le rapport prend environ 600 pages, de plus même si le calcul devient possible, le nombre d'itérations à faire afin d'obtenir un résultat satisfaisant doit être au moins de l'ordre de 1000000. C'est faisable mais cela reste long.

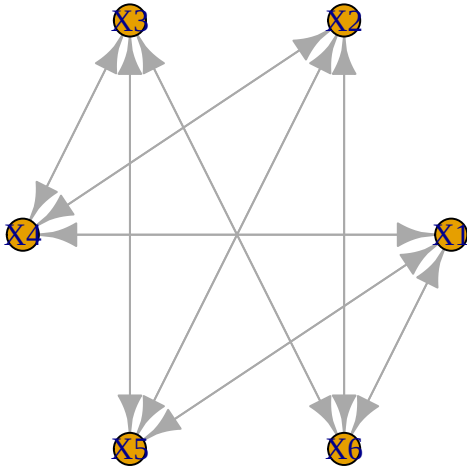
## Exemple 3

Dans cette exemple, le graphe initial est un biparti(3,3) complet et mutualiste (deux groupes d'espèces s'entraident, il n'y pas de compétition)

```
## [1] "c = 0.2"
```

```
## [1] "e = 0.05"
```





Voici la distribution stationnaire (il y a 64 valeurs) :

```
##          000000      000001      000010      000011      000100
## [1,] 1.178307e-05 5.03256e-05 5.03256e-05 0.0002089898 5.03256e-05
##          000101      000110      000111      001000      001001
## [1,] 0.0002089898 0.0002089898 0.0008323518 5.03256e-05 0.0002643773
##          001010      001011      001100      001101      001110
## [1,] 0.0002643773 0.001368669 0.0002643773 0.001368669 0.001368669
##          001111      010000      010001      010010      010011
## [1,] 0.006912988 5.03256e-05 0.0002643773 0.0002643773 0.001368669
##          010100      010101      010110      010111      011000
## [1,] 0.0002643773 0.001368669 0.001368669 0.006912988 0.0002089898
##          011001      011010      011011      011100      011101
## [1,] 0.001368669 0.001368669 0.008960764 0.001368669 0.008960764
##          011110      011111      100000      100001      100010
## [1,] 0.008960764 0.0582916 5.03256e-05 0.0002643773 0.0002643773
##          100011      100100      100101      100110      100111
## [1,] 0.001368669 0.0002643773 0.001368669 0.001368669 0.006912988
##          101000      101001      101010      101011      101100
## [1,] 0.0002089898 0.001368669 0.001368669 0.008960764 0.001368669
##          101101      101110      101111      110000      110001
## [1,] 0.008960764 0.008960764 0.0582916 0.0002089898 0.001368669
##          110010      110011      110100      110101      110110      110111
## [1,] 0.001368669 0.008960764 0.001368669 0.008960764 0.008960764 0.0582916
##          111000      111001      111010      111011      111100      111101
## [1,] 0.0008323518 0.006912988 0.006912988 0.0582916 0.006912988 0.0582916
##          111110      111111
## [1,] 0.0582916 0.4978778
```

On remarque que beaucoup de valeurs sont les mêmes ; isolons les valeurs identiques :

```
## [1] 1.178307e-05 5.032560e-05 2.089898e-04 8.323518e-04 2.643773e-04
## [6] 1.368669e-03 6.912988e-03 8.960764e-03 5.829160e-02 4.978778e-01
```

Au final, sur les 64 valeurs calculées, seulement 10 sont différentes. Cela est dû à la forme typique du graphe initial choisi et au fait que les interactions sont les mêmes.

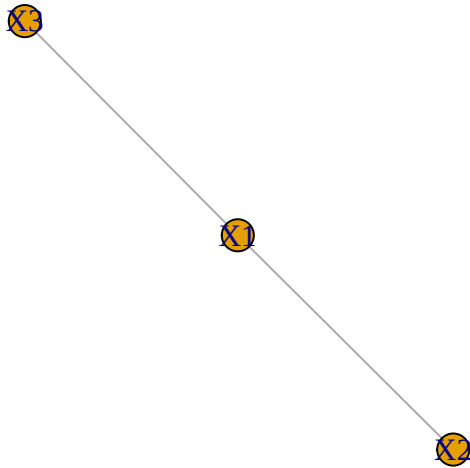
# Ensemble d'exemples sur le modèle méta-population

## Exemple 1

Prenons un premier exemple simple :

```
## [1] "c = 0.25"
```

```
## [1] "e = 0.05"
```



Sans surprise, la distribution stationnaire nous indique l'extinction de la population étudiée en temps long :

```
##      000 001 010 011 100 101 110 111
## [1,]  1  0  0  0  0  0  0  0
```

Intéressons nous donc à la distribution quasi-stationnaire :

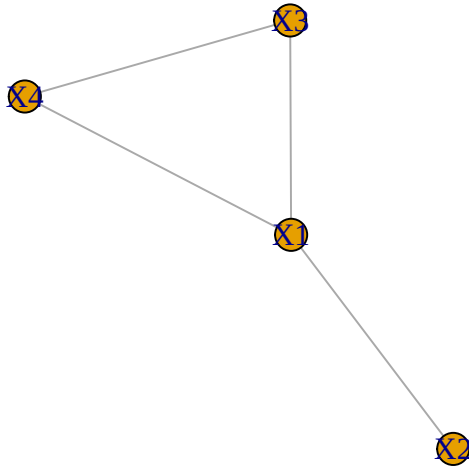
```
##           001           010           011           100           101           110
## [1,] 0.02800676 0.02800676 0.06188373 0.0245023 0.1281222 0.1281222
##           111
## [1,] 0.601356
```

Les sites X2 et X3 jouant les mêmes rôles (la permutation inversant ces deux noeuds dans le graphe laisse le graphe inchangé), on observe que, à X1 fixé, les valeurs sur la quasi-stationnaire impliquant X2 et X3 sont les mêmes. Par exemple pour 010 et 001, ou bien 101 et 110.

## Exemple 2

```
## [1] "c = 0.25"
```

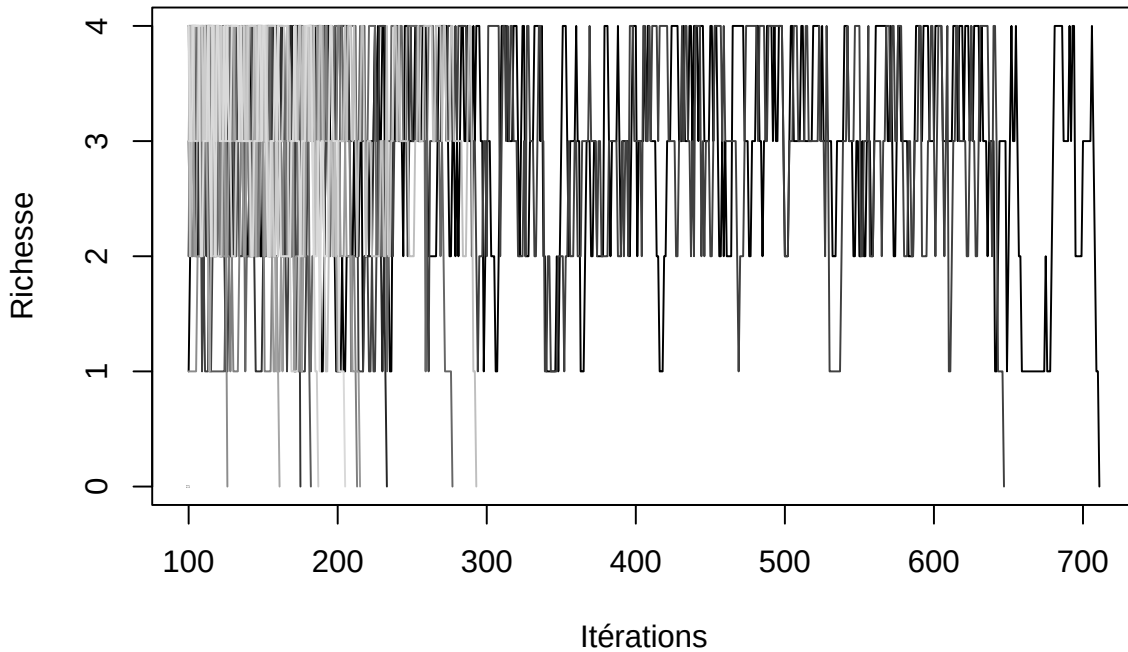
```
## [1] "e = 0.1"
```



Distribution quasi-stationnaire :

```
##          0001          0010          0011          0100          0101          0110
## [1,] 0.0145477 0.0145477 0.04317794 0.01483966 0.01671499 0.01671499
##          0111          1000          1001          1010          1011          1100
## [1,] 0.05861623 0.01341892 0.03924199 0.03924199 0.1645237 0.03416281
##          1101          1110          1111
## [1,] 0.08615981 0.08615981 0.3579317
```

Il y a plusieurs manières d'approximer ce résultat par simulation. Dans un premier temps nous simulons  $N = 20$  trajectoires durant 1000 itérations et fixons un temps de mélange à 100 itérations. Nous pouvons alors regarder la quasi-stationnaire obtenu par la trajectoire qui a fait le plus de chemin sans être absorbée par 0000 (elles peuvent être plusieurs). Ou bien, obtenir la quasi-stationnaire à partir de toutes les trajectoires non absorbées avant le temps de mélange. Le graphique, ci-dessous, présente la richesse de chaque trajectoire (le nombre de 1) qui n'a pas encore été absorbé au temps de mélange, en fonction du nombre d'itérations. Cela nous permet d'avoir un visuel des trajectoires simulées.



## \$Moyenne

```

##           0001           0010           0011           0100           0101           0110
## [1,] 0.01787453 0.01467571 0.04569131 0.03582524 0.01930177 0.01225214
##           0111           1000           1001           1010           1011           1100
## [1,] 0.05854375 0.01971581 0.05195565 0.05264476 0.1380508 0.02647808
##           1101           1110           1111
## [1,] 0.0934706 0.09196249 0.3215573
##
## $Maximum
##           0001           0010           0011           0100           0101           0110
## [1,] 0.008183306 0.02618658 0.07364975 0.0294599 0.003273322 0.01309329
##           0111           1000           1001           1010           1011           1100
## [1,] 0.05237316 0.02291326 0.03600655 0.05237316 0.1898527 0.03436989
##           1101           1110           1111
## [1,] 0.07037643 0.08510638 0.3027823
##
## $`Temps max`
## [1] 712

```

Le vecteur appelé Moyenne est issu de toutes les trajectoires non absorbées avant le temps de mélange. Le vecteur appelé Maximum est issu de la dernière trajectoire non absorbée uniquement.

Les erreurs quadratiques obtenues sont les suivantes, avec dans l'ordre, par rapport au vecteur Moyenne, puis au vecteur Maximum et enfin avec le système de Fleming-Viot.

```

## [1] 0.05509462
## [1] 0.07587717
## [1] 0.0086601

```

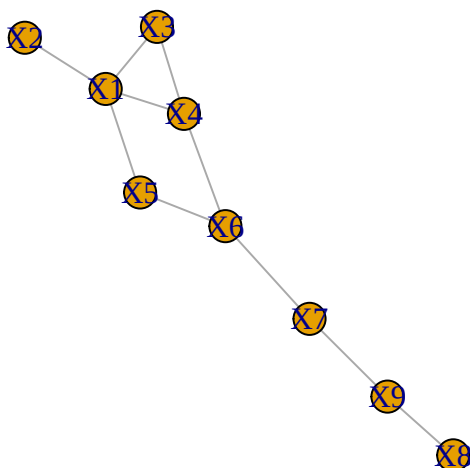
### Exemple 3

Prenons pour exemple un graphe initial assez grand pour que le calcul de la quasi-stationnaire soit fastidieux mais faisable ; il nous permettra de voir quel système de simulation est le plus performant.

```

## [1] "c = 0.2"
## [1] "e = 0.1"

```



Les vecteurs faisant approximativement 41 Kbytes, nous allons uniquement afficher les erreurs quadratiques entre les résultats simulés et la quasi-stationnaire réelle. En lançant 20 trajectoires pendant 10000 itérations

avec la méthode classique (temps d'attente de 1000 itérations), nous obtenons les valeurs d'erreurs suivantes :

```
## [1] "Avec la trajcetoire moyenne :" "0.033579190667996"
```

```
## [1] "Avec la trajectoire maximal" "0.0185054544586092"
```

A savoir que, la dernière trajectoire n'a pas atteint les 10000 itérations fixées mais seulement :

```
## [1] 6858
```

Le taux d'extinction était trop élevé pour la méthode classque, il faut dans ces cas là lancer beaucoup plus de trajectoires. Ou bien nous pouvons passer par un système de Fleming-Viot. Voici l'erreur quadratique avec un tel système (20 trajectoires, 10000 itérations, temps de mélange = 1000 itérations) :

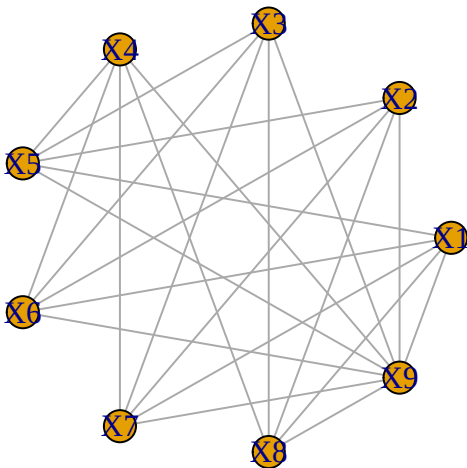
```
## [1] 0.003018503
```

Le résultat obtenu avec le système de Fleming-Viot est plus proche de la quasi-stationnaire théorique.

## Exemple 4

```
## [1] "c = 0.2"
```

```
## [1] "e = 0.1"
```



Ce graphe initial est un multiparti(4,4,1) complet, son nombre de valeurs sur la quasi-stationnaire est  $2^9 - 1 = 511$ . Il s'avère que le nombre de valeurs différentes est de : 29. Cela est dû à la forme particulière du graphe initial.

```
## [1] 8.869004e-07 1.011802e-06 2.548548e-06 1.885631e-06 8.720233e-06
```

```
## [6] 5.783405e-06 3.534271e-05 2.261229e-05 1.602768e-04 2.491468e-06
```

```
## [11] 1.004107e-05 7.895615e-06 4.355978e-05 3.107561e-05 2.051942e-04
```

```
## [16] 1.401989e-04 1.030517e-03 3.519582e-05 2.247766e-04 1.693482e-04
```

```
## [21] 1.199958e-03 8.679801e-04 6.608142e-03 9.437329e-04 7.087319e-03
```

```
## [26] 5.383936e-03 4.232382e-02 3.352980e-02 2.710640e-01
```

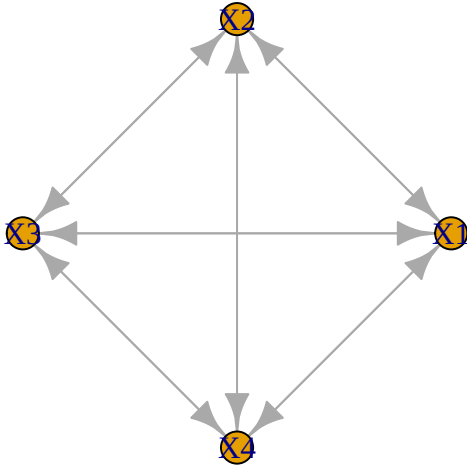
# Ensemble d'exemple sur les problèmes agrégés

## Exemple 1

Pour ce premier exemple prenons un graphe initial complet de taille 4 avec interactions mutualistes.

```
## [1] "c = 0.1"
```

```
## [1] "e = 0.05"
```



Regardons la distribution la dynamique qui résulte de ces choix de paramètres :

```
##          0000          0001          0010          0011          0100          0101
## [1,] 0.002782769 0.005811372 0.005811372 0.01640016 0.005811372 0.01640016
##          0110          0111          1000          1001          1010          1011
## [1,] 0.01640016 0.07011595 0.005811372 0.01640016 0.01640016 0.07011595
##          1100          1101          1110          1111
## [1,] 0.01640016 0.07011595 0.07011595 0.595107
```

On observe toujours des valeurs identiques, il s'agit des valeurs de probabilité des états dans une même orbite du graphe d'état. Appliquons la méthode de la matrice agrégé expliquée en partie 4.2 (nous utilisons ici la fonction codée suivante : `get_invar_agreg_comp_ile_cont`).

```
##          0          1          2          3          4
## [1,] 0.002782769 0.02324549 0.09840098 0.2804638 0.595107
```

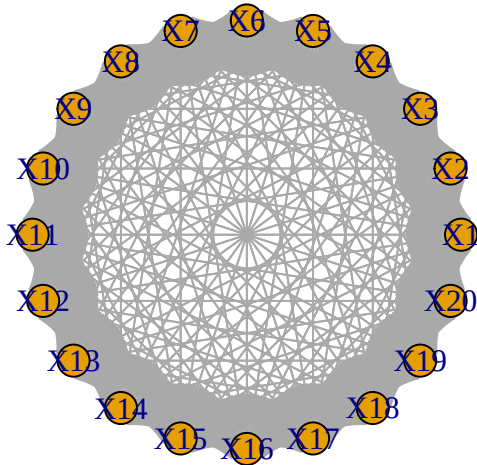
Ici ce ne sont pas les probabilités d'obtention de chaque états à l'équilibre que l'on retrouve sur les composantes de la distribution, mais les probabilités d'être dans l'orbite de richesse indiqué. Ainsi il y a 28% de chance d'être dans l'orbite de richesse 3 à l'équilibre, c'est à dire d'être dans un des états suivants : 1110, 1101, 1011, 0111 (On remarque que si on divise ce résultat par le nombre d'éléments de l'orbite, ici 4, on obtient la probabilité d'obtention de ces états en particulier).

## Exemple 2

Cas d'un graphe initial complet de taille 20.

```
## [1] "c = 0.1"
```

```
## [1] "e = 0.05"
```



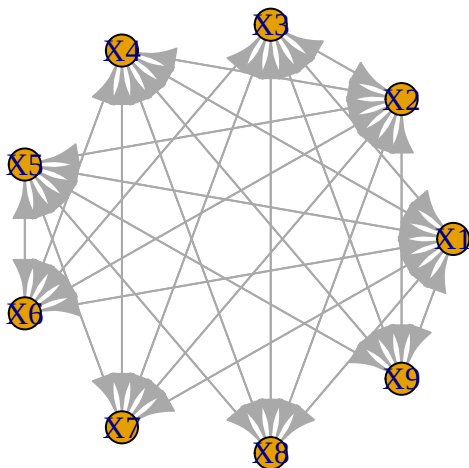
Cette fois-ci le graphe initial est trop grand pour obtenir la distribution stationnaire sur un ordinateur "classique", avec la méthode de matrice agrégée, le calcul est faisable en temps raisonnable, et on obtient :

```
##          0          1          2          3          4
## [1,] 8.877796e-18 1.313745e-15 2.663341e-14 4.043157e-13 4.724404e-12
##          5          6          7          8          9
## [1,] 4.531954e-11 3.717657e-10 2.682226e-09 1.737314e-08 1.026144e-07
##          10         11         12         13         14
## [1,] 5.59528e-07 2.844543e-06 1.359247e-05 6.146103e-05 0.0002644591
##          15         16         17         18         19         20
## [1,] 0.001087962 0.004295724 0.01632696 0.05983283 0.2111751 0.7069384
```

### Exemple 3

Prenons l'exemple d'un graphe initial multiparti(2,3,4) complet (les sous-ensembles partitionnant l'ensemble des noeuds sont de taille respective : 2, 3 et 4).

```
## [1] "c = 0.1"
## [1] "e = 0.05"
```



La fonction `get_invar_agreg_multi_ile_cont` codée, nous donne le résultat agrégé suivant pour cet exemple :

```

##          000          100          200          010          110
## [1,] 3.437762e-06 7.44753e-06 1.572385e-05 7.203125e-06 1.810979e-05
##          210          020          120          220          030
## [1,] 4.450431e-05 1.473522e-05 4.312119e-05 0.0001236004 2.92194e-05
##          130          230          001          101          201
## [1,] 9.972102e-05 0.000333381 7.01277e-06 1.759461e-05 4.315005e-05
##          011          111          211          021          121
## [1,] 1.69865e-05 5.047569e-05 0.0001475014 4.029979e-05 0.0001426986
##          221          031          131          231          002
## [1,] 0.0004998469 9.290467e-05 0.0003943872 0.001664573 1.399536e-05
##          102          202          012          112          212
## [1,] 4.078895e-05 0.0001164597 3.931651e-05 0.0001389393 0.0004857927
##          022          122          222          032          132
## [1,] 0.00010847 0.0004694961 0.002032245 0.0002911133 0.001561089
##          232          003          103          203          013
## [1,] 0.008466427 2.716239e-05 9.216775e-05 0.0003065401 8.87326e-05
##          113          213          023          123          223
## [1,] 0.00037532 0.00157946 0.0002850663 0.001526307 0.008269832
##          033          133          233          004          104
## [1,] 0.0008892973 0.006142052 0.04376873 5.08866e-05 0.0002012169
##          204          014          114          214          024
## [1,] 0.0007775742 0.0001936131 0.0009867098 0.005023107 0.000723551
##          124          224          034          134          234
## [1,] 0.004861812 0.03348491 0.002604968 0.02377803 0.2294937

```

Ici ce sont les différentes richesses par sous-ensemble de noeuds qui caractérisent les orbites du graphe d'état. Par exemple la composante 114 concernent les états dont la richesse est de 1 sur le premier ensemble de taille 2, puis 1 sur le deuxième ensemble de taille 3 et enfin 4 sur le dernier ensemble de taille 4 ; c'est à dire les états suivants : 010011111, 010101111, 011001111, 100011111, 100101111 et 101001111.

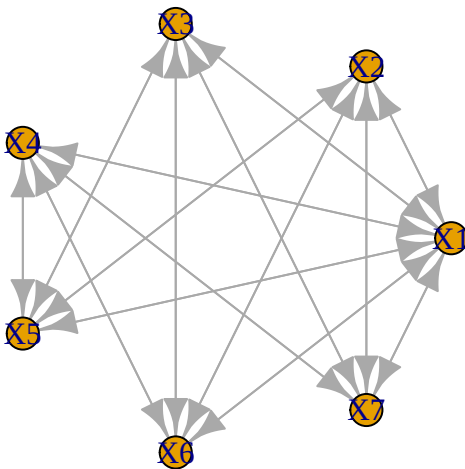
## Exemple 4

Désormais prenons un graphe multiparti(1,3,3), cette fois-ci deux des sous-ensembles indiqués sont de même taille.

```

## [1] "c = 0.1"
## [1] "e = 0.05"

```





Regardons le résultat obtenu avec la même fonction utilisée précédemment :

```
##          000          100          010          110          020
## [1,] 6.866836e-05 0.0001540381 0.0004226806 0.001143544 0.0008420257
##          120          030          130          001          101
## [1,] 0.00276211 0.0005368643 0.002141932 0.0004226806 0.001143544
##          011          111          021          121          031          131
## [1,] 0.003116757 0.01054465 0.007482565 0.03208639 0.005776873 0.03188638
##          002          102          012          112          022          122
## [1,] 0.0008420257 0.00276211 0.007482565 0.03208639 0.02175017 0.1250174
##          032          132          003          103          013          113
## [1,] 0.02034776 0.162241 0.0005368643 0.002141932 0.005776873 0.03188638
##          023          123          033          133
## [1,] 0.02034776 0.162241 0.02281736 0.2811908
```

On remarque que certaines valeurs sont redondantes, par exemple celles concernant 132 et 123, cela s'explique par le fait que la fonction codée prend en compte uniquement les automorphismes échangeant seulement deux noeuds du graphe d'état. Or avec ce type d'automorphisme uniquement, les états appartenant aux orbites 132 et 123, par exemple, ne peuvent pas être dans la même orbite. Pour qu'ils le soient, il aurait fallu que la fonction considère l'automorphisme pouvant échanger tous les noeuds des sous-ensembles de taille trois.

## 7.2 Preuves

*Démonstration. 2* ( $\mathcal{R}$  est une relation d'équivalence)

Soient  $\mathcal{G} = (V, E)$  et  $(x, y) \in V^2$  tel que  $x\mathcal{R}y$ .

$Id \in Aut(\mathcal{G})$ , ainsi  $Id(x) = x$ . nous avons donc montrer que  $x\mathcal{R}x$ , c'est à dire que  $\mathcal{R}$  est réflexive.

On suppose  $x\mathcal{R}y$ , alors il existe  $\phi \in Aut(\mathcal{G})$  tel que  $\phi(x) = y$ . Or  $\phi \in Aut(\mathcal{G})$  donc  $\phi^{-1} \in Aut(\mathcal{G})$  et ainsi  $\phi^{-1}(y) = \phi^{-1}(\phi(x)) = x$ . Nous avons montré que  $y\mathcal{R}x$ , c'est à dire que  $\mathcal{R}$  est symétrique.

Soit  $z \in V$  tel que  $y\mathcal{R}z$ . Il existe alors  $\phi' \in Aut(\mathcal{G})$  tel que  $\phi'(y) = z$ . Or comme  $x\mathcal{R}y$ , on peut écrire  $\phi'(y) = \phi'(\phi(x))z$ , ( $Aut(\mathcal{G}), \circ$ ) étant un groupe,  $\phi' \circ \phi \in Aut(\mathcal{G})$ , cela signifie que  $x\mathcal{R}z$  et que  $\mathcal{R}$  est transitive.

$\mathcal{R}$  est donc une relation d'équivalence sur les noeud du graphe  $\mathcal{G}$  □

*Démonstration. 3.1* (Les matrices de transition des chaînes de Markov du modèle île/continent sont strictement positives)

Soit  $(\mathbf{X}^t)_t$  chaîne de Markov d'espace d'état  $Word_{bin}(n)$  et dont les probabilités de transition sont données par le modèle île/continent présenté dans ce rapport. Soient  $S$  et  $S'$  deux états dans  $Word_{bin}(n)$ . Montrons que la probabilité de transition de  $S'$  à  $S$  est strictement positive.

D'après la formule de Bayes on peut écrire :

$$\mathbb{P}(\mathbf{X}^{t+1} = S \mid \mathbf{X}^t = S') = \prod_{j=1}^n \mathbb{P}(X_j^{t+1} = S_j \mid X_j^t = S'_j)$$

Or, on connaît les valeurs des termes dans le produit ; elles dépendent  $S_i$  et  $S'_i$  et sont données par le modèle. Par exemple  $S_i = S'_i = 0$  alors  $\mathbb{P}(\mathbf{X}_i^{t+1} = S_i \mid \mathbf{X}_i^t = S'_i) = 1 - c$ .

Soit  $i \in \llbracket 1, n \rrbracket$ , Posons  $p(i) = 1 - e + \frac{\left[1 - (1-e)^{N_+^{\mathbf{X}^t(i)}}\right] (1-e)^{N_-^{\mathbf{X}^t(i)}}}{M}$  la probabilité de persistance (avec  $M = \frac{1 - (1-e)^n}{e}$ ). Il y a quatre cas de figures différents, résumons les dans un tableau, soit .

$S'_i \setminus S_i$	0	1
0	$1 - c$	$c$
1	$1 - p(i)$	$p(i)$

Supposons que chacun de ces cas soit répété respectivement  $n_1, n_2, n_3$  et  $n_4$  fois, de telle sorte que  $n_1 + n_2 + n_3 + n_4 = n$ . On peut désormais écrire

$$\mathbb{P}(\mathbf{X}^{t+1} = S \mid \mathbf{X}^t = S') = (1 - c)^{n_1} c^{n_2} \prod_{j=1}^{n_3} (1 - p(j)) \prod_{j=1}^{n_4} p(j)$$

On rappelle que  $c \in ]0, 1[$ , donc  $(1 - c)^{n_1} c^{n_2} > 0$ .

De plus  $e \in ]0, 1[$  et  $n > n - 1 \geq N_+^{\mathbf{X}^t(i)} \geq 0$  p.s., car un noeud ne peut pas être parent de lui même dans notre modèle ; de la même manière  $n > N_-^{\mathbf{X}^t(i)}$ . Nous pouvons en déduire que

$$(1 - e)^n < (1 - e)^{N_+^{\mathbf{X}^t(i)}} \leq 1$$

$$\begin{aligned}
1 - (1 - e)^n &> \left[1 - (1 - e)^{N_+^{\mathbf{X}^t(i)}}\right] (1 - e)^{N_-^{\mathbf{X}^t(i)}} \geq 0 \\
1 &> \frac{\left[1 - (1 - e)^{N_+^{\mathbf{X}^t(i)}}\right] (1 - e)^{N_-^{\mathbf{X}^t(i)}}}{1 - (1 - e)^n} \geq 0 \\
e &> \frac{\left[1 - (1 - e)^{N_+^{\mathbf{X}^t(i)}}\right] (1 - e)^{N_-^{\mathbf{X}^t(i)}}}{\frac{1 - (1 - e)^n}{e}} \geq 0 \\
1 &> 1 - e + \frac{\left[1 - (1 - e)^{N_+^{\mathbf{X}^t(i)}}\right] (1 - e)^{N_-^{\mathbf{X}^t(i)}}}{M} \geq 1 - e > 0
\end{aligned}$$

Nous avons montré que  $1 > p(i) > 0$ , ainsi  $\prod_{j=1}^{n_3} (1 - p(j)) \prod_{j=1}^{n_4} p(j) > 0$ . Cela prouve que  $\mathbb{P}(\mathbf{X}^{t+1} = S \mid \mathbf{X}^t = S') > 0 \forall (S, S') \in \text{Word}_{bin}(n)^2$ , la matrice de transition de la chaîne de Markov  $(\mathbf{X}^t)_t$  est strictement positive.  $\square$

*Démonstration. 4.1.1* (Deux états dans la même orbite du graphe d'état ont la même valeur sur la stationnaire)

Soient  $n \in \mathbb{N}$ ,  $\mathcal{G} = (V, E)$  et  $(\mathbf{X}^t)_t$  la chaîne de Markov sur les arrangements possibles des noeuds du graphe  $\mathcal{G}$ , d'espace d'état  $\{0, 1\}^n$ . On note  $A$  le noyau de transition de cette chaîne de Markov.

On notera  $\mathcal{G}_{états} = (V_{états}, E_{états})$  le graphe orienté et pondéré sur les états de cette chaîne de Markov (rappelons que  $V_{états} = \{0, 1\}^n$ ), ce graphe est probabiliste et la matrice  $A$  est la matrice d'adjacence pondéré du graphe  $\mathcal{G}_{états}$ .

Cette chaîne de Markov étant irréductible et apériodique, d'après le théorème de Perron-Frobenius, elle admet une distribution stationnaire, notée  $\pi$ .

Soit  $\varphi$  un automorphisme du graphe d'états et  $P$  sa matrice de permutation alors d'après le corollaire 2.9 de Lauri & Scapellato (2016),  $P\pi$  est un vecteur propre associé à la valeur propre 1. Or le théorème de Perron-Frobenius assure l'unicité de la valeur propre associée à la valeur propre 1.

Donc

$$P\pi = \pi \tag{2}$$

Cela signifie que  $\pi(\varphi(v)) = \pi(v) \forall v \in V_{états}$

Si on considère un richesse  $k \in \llbracket 0, n \rrbracket$  et  $(u, v) \in \mathcal{O}(k)^2$  l'orbite des états de richesse  $k$ . Alors, étant dans la même orbite, il existe  $\phi \in \text{Aut}(G_{états})$  tes que  $\phi(u) = v$ . D'après (2) la valeur de probabilité sur la distribution stationnaire concernant l'état  $\phi(u) = v$  est la même que celle de l'état  $u$ . Ainsi tous les états d'une même orbite ont la même valeur de probabilité sur la distribution stationnaire.  $\square$

*Démonstration. 4.1.2* (Deux états de même richesse sont dans la même orbite du graphe d'état, modèle spatial ou interaction mutualiste avec graphe initial complet)

Soient  $n \in \mathbb{N}$ ,  $\mathcal{G} = (V, E) \in \mathbf{K}_n$  le graphe complet à  $n$  noeuds et  $(\mathbf{X}^t)_t$  la chaîne de Markov sur les arrangements possibles des noeuds du graphe  $\mathcal{G}$ , d'espace d'état  $\text{Word}_{bin}(n)$ . On note  $P$  le noyau de transition de cette chaîne de Markov.

On notera  $\mathcal{G}_{\text{états}} = (V_{\text{états}}, E_{\text{états}})$  le graphe d'états de cette chaîne de Markov (rappelons que  $V_{\text{états}} = \text{Word}_{\text{bin}}(n)$ ).

On veut montrer que :

$$\forall (S_1, S_2) \in (\text{Word}_{\text{bin}}(n))^2 \quad |S_1| = |S_2| \iff \exists \phi \in \text{Aut}(\mathcal{G}_{\text{état}}) \quad \phi(S_1) = S_2$$

C'est à dire que tous les états présents dans une orbite ont la même richesse et que cette dernière caractérise l'orbite.

Montrons le premier sens.

Soient  $S_1 = \overline{s_1^1 s_2^1 \dots s_n^1}$ ,  $S_2 = \overline{s_1^2 s_2^2 \dots s_n^2}$  deux noeuds du graphes d'états dans  $\text{Word}_{\text{bin}}(n)$  de même richesse ( $|S_1| = |S_2|$ ). On définit  $\text{Diff}_0(S_1, S_2) = \{i \in \llbracket 1, n \rrbracket \mid s_i^1 = 0, s_i^2 = 1\}$  et  $\text{Diff}_1(S_1, S_2) = \{i \in \llbracket 1, n \rrbracket \mid s_i^1 = 1, s_i^2 = 0\}$ . On remarque alors que  $\text{card}(\text{Diff}_0(S_1, S_2)) = \text{card}(\text{Diff}_1(S_1, S_2))$  car  $S_1$  et  $S_2$  ont la même richesse et sont de la même longueur  $n$  (ils ont donc le même nombre de 0 et de 1).

Soient  $\phi : \text{Word}_{\text{bin}}(n) \rightarrow \text{Word}_{\text{bin}}(n)$  une permutation sur les noeuds du graphe d'états tel que  $\phi(S_1) = S_2$  et  $\varphi : \llbracket 1, n \rrbracket \rightarrow \llbracket 1, n \rrbracket$  une permutation sur les indices des mots des noeuds du graphe d'états tel que  $\phi(S) = \overline{s_{\varphi(1)} s_{\varphi(2)} \dots s_{\varphi(n)}} \quad \forall S \in \text{Word}_{\text{bin}}(n)$ . De plus  $\varphi$  est défini de la manière suivante : pour tout  $i$  tel que  $s_i^1 = s_i^2$  alors  $\varphi(i) = i$ , mais pour tout  $i \in \text{Diff}_1(S_1, S_2)$  il existe  $k \in \text{Diff}_0(S_1, S_2)$  tel que  $\varphi(i) = k$  et  $\varphi(k) = i$ . Définit ainsi,  $\varphi$  est un produit de transposition à support disjoint. Si on note  $d = \text{card}(\text{Diff}_1(S_1, S_2))$  alors  $\varphi = (i_1, k_1) \circ (i_2, k_2) \circ \dots \circ (i_d, k_d)$  avec  $i_1, i_2, \dots, i_d \in \text{Diff}_1(S_1, S_2)$  et  $k_1, k_2, \dots, k_d \in \text{Diff}_0(S_1, S_2)$ .

Soit  $(S, S') \in (\text{Word}_{\text{bin}}(n))^2$ , tel que  $S = \overline{s_1 s_2 \dots s_n}$  et  $S' = \overline{s'_1 s'_2 \dots s'_n}$ , vérifions que

$$\text{Weight}(\phi(S), \phi(S')) = \text{Weight}(S, S')$$

Si c'est le cas, alors  $\phi$  est un automorphisme.

Grâce à la formule de Bayes on peut faire apparaître le premier produit suivant :

$$\begin{aligned} \text{Weight}(\phi(S), \phi(S')) &= \mathbb{P}(\phi(S) \mid \phi(S')) \\ &= \prod_{i=1}^n \mathbb{P}(X_i^{t+1} = s_{\varphi(i)} \mid X_i^t = s'_{\varphi(i)}, X_{N(i)} = s'_{N(\varphi(i))}) \end{aligned}$$

Le modèle considéré distingue deux premiers cas : y avait-il de la vie ou non sur le site (*resp* de l'espèce) observé? Si il n'y avait pas de vie alors la probabilité d'apparition de vie est une fonction de colonisation, notée  $C$ , du nombre de voisin et si il y avait de la vie alors la probabilité d'extinction est une autre fonction, notée  $E$ , du nombre de voisin (ces fonctions n'ont pas la même expression et dépendent du paramètre  $c$  ou  $e$ ). Le graphe  $\mathcal{G}$  étant complet, alors pour chacun des sites (*resp* espèces) étudiés, le nombre de voisin dans le cas de la colonisation est  $|S|$  et dans le cas de la persistance ce nombre est  $|S| - 1$  car tous les autres sites (*resp* espèces) sont voisins.

De ce fait, posons l'ensemble  $\text{Null}(S) = \{i \in \llbracket 1, n \rrbracket \mid s_i = 0\}$  et  $\mathfrak{C}(\text{Null}(S))$  son complémentaire.

On peut alors écrire :

$$\begin{aligned} \text{Weight}(\phi(S), \phi(S')) &= \prod_{i \in \text{Null}(\phi(S'))} [C(|\phi(S')|) \mathbf{1}_{s_{\varphi(i)}=1} + (1 - C(|\phi(S')|)) \mathbf{1}_{s_{\varphi(i)}=0}] \times \\ &\quad \prod_{i \in \mathcal{L}(\text{Null}(\phi(S')))} [(1 - E(|\phi(S')| - 1)) \mathbf{1}_{s_{\varphi(i)}=1} + E(|\phi(S')| - 1) \mathbf{1}_{s_{\varphi(i)}=0}] \end{aligned}$$

Il est intéressant de noter que  $|\phi(S)| = |S| \forall S \in \{0, 1\}^n$  car  $\phi$  ne fait que permuter les lettres du mot  $S$ , la somme des 0 et des 1 reste la même.

$$\begin{aligned} \text{Weight}(\phi(S), \phi(S')) &= \prod_{i \in \text{Null}(\phi(S'))} [C(|S'|) \mathbf{1}_{s_{\varphi(i)}=1} + (1 - C(|S'|)) \mathbf{1}_{s_{\varphi(i)}=0}] \times \\ &\quad \prod_{i \in \mathcal{L}(\text{Null}(S'))} [(1 - E(|S'| - 1)) \mathbf{1}_{s_{\varphi(i)}=1} + E(|S'| - 1) \mathbf{1}_{s_{\varphi(i)}=0}] \end{aligned}$$

Penchons nous sur l'ensemble  $\text{Null}(\phi(S'))$

$$\begin{aligned} \text{Null}(\phi(S')) &= \{i \in \llbracket 1, n \rrbracket \mid \phi(S')_i = 0\} \\ &= \{\varphi(i) \in \llbracket 1, n \rrbracket \mid s'_{\varphi(i)} = 0\} \\ &= \{\varphi(i) \in \llbracket 1, n \rrbracket \mid s'_i = 0\} \end{aligned}$$

Cette dernière égalité nous permet de réécrire les produits précédents car pour tout  $i \in \text{Null}(\phi(S'))$  il existe un unique  $j \in \text{Null}(S')$  tel que  $\varphi(j) = i$  (l'unicité provient de la définition de  $\varphi$  et du fait que  $\text{card}(\text{Null}(\phi(S'))) = \text{card}(\text{Null}(S'))$ ). Ainsi :

$$\begin{aligned} \text{Weight}(\phi(S), \phi(S')) &= \prod_{j \in \text{Null}(S')} [C(|S'|) \mathbf{1}_{s_{\varphi(j)}=1} + (1 - C(|S'|)) \mathbf{1}_{s_{\varphi(j)}=0}] \times \\ &\quad \prod_{j \in \mathcal{L}(\text{Null}(S'))} [(1 - E(|S'| - 1)) \mathbf{1}_{s_{\varphi(j)}=1} + E(|S'| - 1) \mathbf{1}_{s_{\varphi(j)}=0}] \end{aligned}$$

Lors de la définition de  $\varphi$ , nous avons précisé que cette permutation est un produit de transposition à support distinct. Ainsi  $\varphi \circ \varphi = Id$ , cela nous permet de terminer le calcul du poids d'une arête dont les noeuds sont l'images de la permutation  $\phi$ .

$$\begin{aligned} \text{Weight}(\phi(S), \phi(S')) &= \prod_{j \in \text{Null}(S')} [C(|S'|) \mathbf{1}_{s_j=1} + (1 - C(|S'|)) \mathbf{1}_{s_j=0}] \times \\ &\quad \prod_{j \in \mathcal{L}(\text{Null}(S'))} [(1 - E(|S'| - 1)) \mathbf{1}_{s_j=1} + E(|S'| - 1) \mathbf{1}_{s_j=0}] \\ &= \prod_{j=1}^n \mathbb{P}(X_j^{t+1} = s_j \mid X_i^t = s'_j, X_{N(j)} = s'_{N(j)}) \\ &= \mathbb{P}(S \mid S') \\ &= \text{Weight}(S, S') \end{aligned}$$

Nous avons montré que l'application  $\phi$  sur les noeuds du graphe d'états, préserve le poids de toutes les arêtes du graphe. Cette application est donc un automorphisme du graphe, ainsi  $S_1$  et  $S_2$ , de même richesse, sont dans une même orbite du graphe d'états.

Montrons le sens inverse du théorème.

Dans un premier temps, nous allons montrer que l'état 111...11 de richesse  $n$  (l'état le plus riche) est seul dans son orbite. Il est important de comprendre la condition nécessaire suivante :

$$\forall (x, y) \in (V_{\text{états}})^2, \quad (x, y) \in \mathcal{O}^2 \Rightarrow \mathbb{P}(x|x) = \mathbb{P}(y|y)$$

Deux noeuds dans la même orbite ont le même poids sur leur boucle respective. Alors si on considère  $u \in V_{\text{états}} \setminus \{111..11\}$

$$\mathbb{P}(u|u) = (1 - E(|u| - 1))^{|u|} \times (1 - C(|u|))^{n-|u|}$$

de plus,

$$\mathbb{P}(111..11|111..11) = (1 - E(n - 1))^n$$

Si on montre que  $\mathbb{P}(111..11|111..11) \neq \mathbb{P}(u|u)$  alors 111..11 est seul dans son orbite.

$$\mathbb{P}(111..11|111..11) = \mathbb{P}(u|u)$$

$$(1 - E(n - 1))^n = (1 - E(|u| - 1))^{|u|} \times (1 - C(|u|))^{n-|u|} \quad (3)$$

Distinguons les cas en fonction des deux modèles en remplaçant les fonctions par leurs expressions en fonction de  $e$  et  $c$  :

Dans le cas du modèle île/continent, (3) devient :

$$\left(1 - e + \frac{1 - (1 - e)^{n-1}}{M}\right)^n = \left(1 - e + \frac{1 - (1 - e)^{|u|-1}}{M}\right)^{|u|} \times (1 - c)^{n-|u|}$$

On rappelle que  $M = \frac{1-(1-e)^n}{e}$ , d'autre part pour le cas du modèle méta-population, (3) devient :

$$(1 - e)^n = (1 - e)^{|u|} \times (1 - c)^{|u|(n-|u|)}$$

Si de plus on pose  $k \in \mathbb{Q}$  tel que  $c = ke$  alors dans le cas du modèle île/continent, (3) devient :

$$\left(1 - e + e \frac{1 - (1 - e)^{n-1}}{1 - (1 - e)^n}\right)^n - \left(1 - e + e \frac{1 - (1 - e)^{|u|-1}}{1 - (1 - e)^n}\right)^{|u|} \times (1 - ke)^{n-|u|} = 0$$

$$\begin{aligned} & ((1 - (1 - e)^n)(1 - e) + e - e(1 - e)^{n-1})^n - (1 - (1 - e)^n)^{n-|u|} ((1 - (1 - e)^n)(1 - e) + e - e(1 - e)^{|u|-1})^{|u|} \\ & \times (1 - ke)^{n-|u|} = 0 \end{aligned}$$

Et dans le cas du modèle méta-population, (3) devient :

$$(1 - e)^n - (1 - e)^{|u|} \times (1 - ke)^{|u|(n-|u|)} = 0$$

Résoudre ces équations, c'est trouver les racines de deux polynômes non nuls. Il y a donc un nombre fini de valeurs solutions de ces équations. Les racines sont contenues dans l'ensemble suivant  $R_k = \{c = ke \in ]0, 1[ \mid (1 - E(n-1))^n = (1 - E(|u| - 1))^{|u|} \times (1 - C(|u|))^{n-|u|}\}$ , avec  $k \in \mathbb{Q}$ . L'ensemble  $R_k$  est fini et de mesure nul dans  $]0, 1[$ . Toutes les solutions au problème 3 sont donc dans l'ensemble  $R = \bigcup_{k \in \mathbb{Q}} R_k$ , une réunion dénombrable d'ensembles de mesure nulle.  $R$  est donc de mesure nulle dans  $]0, 1[$ .

Cela signifie que  $\mathbb{P}(111..11 \mid 111..11) = \mathbb{P}(u \mid u)$ , uniquement pour certains couples de paramètres  $(e, c)$  dans un ensemble de mesure nulle sur  $]0, 1[^2$ . Si nous ne choisissons pas nos paramètres dans cette ensemble remplissant cette condition, alors il n'existe aucun état  $u \in \text{Word}_{bin}(n)$  dont la poids de la boucle sur le graphe d'état est le même que celui de l'état 111..111, ainsi 11..111 est seul dans son orbite.

Désormais considérons que  $S_1$  et  $S_2$  sont dans la même orbite  $\mathcal{O}(\mathcal{G}_{états})$  du graphe d'états. Alors il existe  $\phi \in \text{Aut}(\mathcal{G}_{états})$  tel que  $\phi(S_1) = S_2$ .

Montrons que  $|S_1| = |S_2|$ .

D'après la définition de  $\phi$ , on peut écrire que :

$$\text{Weight}(\phi(S_1), \phi(111..11)) = \text{Weight}(S_2, 111..11)$$

car 111...11 est seul dans son orbite et que  $\phi(111..11)$  est dans l'orbite de 111..11.

De plus  $\phi$  étant un automorphisme du graphe, il conserve les poids des arêtes. Ainsi :

$$\text{Weight}(\phi(S_1), \phi(111..11)) = \text{Weight}(S_1, 111..11)$$

Servons nous de l'égalité suivante

$$\text{Weight}(S_1, 111..11) = \text{Weight}(S_2, 111..11)$$

$$\begin{aligned} e(n-1)^{n-|S_1|} \times (1-e(n-1))^{|S_1|} &= e(n-1)^{n-|S_2|} \times (1-e(n-1))^{|S_2|} \\ (1-e(n-1))^{|S_1|-|S_2|} &= e(n-1)^{|S_1|-|S_2|} \end{aligned}$$

En supposant  $1 - E(n-1) \neq E(n-1) \Leftrightarrow E(n-1) \neq \frac{1}{2}$  alors  $|S_1| = |S_2|$

Dans le cas du modèle méta-population, cette condition devient :

$$e \neq \frac{1}{2}$$

En écologie, nous étudions des dynamiques en temps long, les taux d'extinction sont choisis assez faible :  $e < \frac{1}{2}$ .

Dans le cas du modèle île/continent, cette condition devient :

$$e - \frac{1 - (1 - e)^{n-1}}{M} \neq \frac{1}{2}$$

Or  $e \in ]0, 1[$  et  $n$  grand, on rappelle que  $M = \frac{1-(1-e)^n}{e}$ , donc :

$$e - \frac{1 - (1 - e)^{n-1}}{M} \ll \frac{1}{2}$$

Ainsi nous avons montré que, peu importe le modèle, que la condition  $E(n-1) \neq \frac{1}{2}$  est respecté, c'est à dire,  $|S_1| = |S_2|$ . □

*Démonstration. 4.1.3* (Nombre de manières qu'il y a de choisir une richesse différente dans les sous-ensembles de noeuds d'un graphe multiparti)

Soient  $(n, p) \in \mathbb{N}^2$   $\mathcal{G} \in \mathbf{K}_{\underbrace{nn\dots n}_p \text{ fois}}$ , un multiparti complet, les noeuds de ce graphe sont dans un ensemble partitionné en  $p$  sous-ensembles de taille  $n$ , ces sous-ensembles seront appelés paquet pour le reste de la preuve (Les noeuds de ce graphe sont dans une partition de  $p$  paquets de taille  $n$ ).

Nous allons chercher le nombre de paquets différents pouvant avoir la même richesse (c'est à dire le même nombre de 1). On introduit  $A(k, p) = (a_1, a_2, \dots, a_k)$  un  $k$ -uplet d'entier strictement positif tel que  $a_i$  correspond au nombre de paquets de richesse  $r_i$  et avec  $\sum_i a_i = p$ .  $A(k, p)$  décrit plusieurs états possibles, ceux dont la richesse est de  $r_i$  sur  $a_i$  paquets différents. Par exemple si  $A(k, p) = (a_1, \dots, a_k) = (1, \dots, 1)$  ça veut dire que tous les paquets ont des nombres d'occurrence de 1 différents (des richesses différentes), et dans ce cas, on a nécessairement  $k = p$ . Autre exemple, si  $p = 5$ , alors  $\mathcal{G} \in \mathbf{K}_{nnnnn}$  et  $A(2, 5) = (2, 3)$  correspond à tous les états dont deux paquets sont de richesse  $r_1$  et dont trois autres sont de richesse  $r_2$ . On va avoir besoin de dénombrer le nombre de ces compositions. Utilisons le résultat suivant :

Soit  $1 \leq k \leq p$ , alors

$$\{(a_1, a_2, \dots, a_k) \mid 1 \leq a_i, 1 \leq i \leq k, \sum_j a_j = p\} = \binom{p-1}{k-1}$$

Comme l'ordre des  $a_i$  dans  $A(k, p)$  n'est pas important (le graphe est multiparti donc  $(1, 1, 2)$ ,  $(1, 2, 1)$  et  $(2, 1, 1)$ , par exemple, sont équivalents pour nous), cela réduit le nombre de configurations différentes pour la répartition de la richesse entre les paquets. Nous divisons donc le résultat précédent par  $k!$ , le nombre de permutation possible sur  $A(k, p)$ . D'une autre part, pour connaître le nombre de configuration possible au total il faut choisir une richesse  $r_i$  pour chaque groupe de paquets de même richesse. Par exemple si  $A(2, 5) = (2, 3)$ , on a alors  $(n+1)n$  choix de richesse possibles dans  $[0, n]$ .

Plus généralement pour  $A(k, \cdot)$  on a  $(n+1)n\dots((n+1)-(k-1)) = (n+1)\dots(n+2k)$  ensembles de valeurs possibles. Au final on obtient :



$$\begin{aligned}
\sum_{k=1}^{\min(p,n+1)} \binom{p-1}{k-1} \frac{(n+1)\dots(n+2k)}{k!} &= \sum_{k=1}^{\min(p,n+1)} \binom{p-1}{k-1} \binom{n+1}{k} \\
&= \sum_{k=1}^{\min(p,n+1)} \binom{p-1}{k-1} \binom{n+1}{n+1-k} \\
&= \binom{n+p}{p}
\end{aligned}$$

Pour un graphe dans  $\mathbf{K}_{\underbrace{nn\dots n}_p \text{ fois}}$ , le nombre de manières qu'il y a de choisir une richesse différentes dans les sous-ensembles de noeuds est différent par paquet par paquet est  $\binom{n+p}{p}$ .  $\square$

*Démonstration.* **4.2** (Dans le vecteur propre associé à la valeur propre 1 de la matrice agrégée, nous retrouvons toutes les différentes valeurs du vecteur propre associé à la valeur propre 1 du noyau de transition)

Soit  $A$  la matrice d'adjacence d'un graphe pondéré  $\mathcal{G}$ , d'une chaîne de Markov, homogène, apériodique et irréductible dans l'espace d'état  $Word_{bin}(n)$ ,  $n \in \mathbb{N}^*$ .  $A = (a_{ij})_{1 \leq i, j \leq 2^n}$  est donc le noyau de transition de cette chaîne, on notera  $N = 2^n$ ,  $A$  est de taille  $N \times N$ . Soit  $K \in \mathbb{N}^*$  le nombre d'orbites du graphe d'état  $\mathcal{G}$ ,  $(\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_K)$  le  $K$ -uplet de ces orbites ( $K$  ensembles d'états dans  $Word_{bin}(n)$ ) et  $n_i \geq 1$  la taille de l'orbite  $\mathcal{O}_i$ ,  $\forall i \in \{1, 2, \dots, K\}$ . On note  $\pi$  le vecteur propre à gauche pour la valeur propre 1, renormalisé, de la matrice  $M$ , en réarrangeant l'ordre des états en fonction des orbites, (on rappelle que les état dans la même orbite on la même valeur de probabilité sur la stationnaire) on peut écrire :

$$\pi = (\pi_1, \dots, \pi_N) = \left( \underbrace{\pi_{\mathcal{O}_1}, \dots, \pi_{\mathcal{O}_1}}_{N_1 \text{ fois}}, \underbrace{\pi_{\mathcal{O}_2}, \dots, \pi_{\mathcal{O}_2}}_{N_2 \text{ fois}}, \dots, \underbrace{\pi_{\mathcal{O}_K}, \dots, \pi_{\mathcal{O}_K}}_{N_K \text{ fois}} \right)$$

Enfin on note  $S_0 = 0$ ,  $S_1 = N_1$ ,  $S_2 = N_1 + N_2$ ,  $S_i = N_1 + N_2 + \dots + N_i$ ,  $\forall i \in \{1, 2, \dots, K\}$ . // Introduisons désormais la matrice agrégée  $B = (b_{ij})_{1 \leq i, j \leq K}$ , avec

$$b_{ij} = \frac{1}{N_i} \sum_{k=S_{i-1}+1}^{S_i} \sum_{l=S_{j-1}+1}^{S_j} a_{kl} \quad \forall i, j \in \{1, 2, \dots, K\}$$

On sommes les coefficients des colonnes d'un même cycle et on moyenne sur les lignes. La matrice  $B$  est stochastique, vérifions que le vecteur  $\tilde{\pi}$  est bien un vecteur propre associé à la valeur propre 1 pour la matrice  $B$  où

$$\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_k) = (N_1 \pi_{\mathcal{O}_1}, N_2 \pi_{\mathcal{O}_2}, \dots, N_K \pi_{\mathcal{O}_K})$$

Soit  $j$  dans  $\{1, 2, \dots, K\}$

$$\begin{aligned}
\sum_{i=1}^K \tilde{\pi}_i b_{ij} &= \sum_{i=1}^K N_i \pi_{\mathcal{O}_i} b_{ij} \\
&= \sum_{i=1}^K N_i \pi_{\mathcal{O}_i} \frac{1}{N_i} \sum_{k=S_{i-1}+1}^{S_i} \sum_{l=S_{j-1}+1}^{S_j} a_{kl} \\
&= \sum_{i=1}^K \pi_{\mathcal{O}_i} \sum_{k=S_{i-1}+1}^{S_i} \sum_{l=S_{j-1}+1}^{S_j} a_{kl} \\
&= \sum_{i=1}^K \sum_{k=S_{i-1}+1}^{S_i} \sum_{l=S_{j-1}+1}^{S_j} \pi_k a_{kl} \\
&= \sum_{k=1}^N \sum_{l=S_{j-1}+1}^{S_j} \pi_k a_{kl} \\
&= \sum_{l=S_{j-1}+1}^{S_j} \sum_{k=1}^N \pi_k a_{kl} \\
&= \sum_{l=S_{j-1}+1}^{S_j} \pi_l \\
&= N_j \pi_{C_j} = \tilde{\pi}_j
\end{aligned}$$

Ainsi chaque valeur du vecteur propre  $\tilde{\pi}$  correspond à une valeur que l'on retrouve pour une orbite dans le vecteur  $\pi$ .

□

## 5.6 Vignette du paquet R 'econetwork'

# Package ‘econetwork’

July 8, 2019

**Type** Package

**Title** Analyzing Ecological Networks

**Version** 0.2

**Date** 2019-07-08

**Author** Stephane Dray [ctb], Catherine Matias [ctb], Vincent Miele [aut, cre], Marc Ohlmann [aut], Wilfried Thuiller [ctb]

**Maintainer** Vincent Miele <vincent.miele@univ-lyon1.fr>

**Description** A collection of advanced tools, methods and models specifically designed for analyzing different types of ecological networks - especially antagonistic (food webs, host-parasite), mutualistic (plant-pollinator, plant-fungus, etc) and competitive networks, as well as their variability in time and space. Statistical models are developed to describe and understand the mechanisms that determine species interactions, and to decipher the organization of these (multi-layer) ecological networks.

**Imports** stats, igraph, rdiversity, Matrix.utils

**LinkingTo**

**License** GPL-3

**URL** <https://plmlab.math.cnrs.fr/econetproject/econetwork>

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2019-07-08 13:10:03 UTC

## R topics documented:

econetwork-package . . . . .	2
disPairwise . . . . .	2
divPartition . . . . .	4
getMetaweb . . . . .	6
<b>Index</b>	<b>7</b>

---

econetwork-package      *Analyzing Ecological Networks*

---

### Description

A collection of advanced tools, methods and models specifically designed for analyzing different types of ecological networks.

### Details

econetwork is designed for analyzing different types of ecological networks - especially antagonistic (food webs, host-parasite), mutualistic (plant-pollinator, plant-fungus, etc.) and competitive networks, as well as their variability in time and space. Statistical models are developed to describe and understand the mechanisms that determine species interactions, and to decipher the organization of these (multi-layer) ecological networks.

### Author(s)

Authors: Stephane Dray, Catherine Matias, Vincent Miele, Marc Ohlmann, Wilfried Thuiller  
 Maintainer: Vincent Miele <vincent.miele@univ-lyon1.fr>

### References

Marc Ohlmann, Vincent Miele, Stephane Dray, Loic Chalmandrier, Louise O'Connor & Wilfried Thuiller, Diversity indices for ecological networks: a unifying framework using Hill numbers. *Ecology Letters* (2019) <doi:10.1111/ele.13221>

### See Also

[igraph](#) [bipartite](#) [vegan](#) [sna](#) [statnet](#) [enaR](#)

---

disPairwise      *Computation of the dissimilarity matrix (pairwise beta-diversity) for a set of networks*

---

### Description

Computation of the dissimilarity matrix for a set of networks. Each value of the matrix is the pairwise beta-diversity, computed using Hill numbers. It measures the dissimilarity in terms of groups, links, or probability of links.

### Usage

```
disPairwise(gList, groups=NULL, eta=1,
            type=c('P','L','Pi'), abTable=NULL)
```

**Arguments**

<code>gList</code>	A list of graph objects of class <code>igraph</code> . The nodes must have a name.
<code>groups</code>	A named vector of class <code>character</code> indicating the group to which each node belongs to. The length of <code>groups</code> must correspond to the number of different nodes present in <code>gList</code> . The names <code>names(groups)</code> must correspond to the nodes names in <code>gList</code> . If <code>NULL</code> , the groups are the initial nodes.
<code>eta</code>	A positive number that controls the weight given to abundant groups/links. Default value is 1.
<code>type</code>	The type of diversity used to measure dissimilarity. It can be groups diversity ('P'), links diversity ('L') or probability of links diversity ('Pi').
<code>abTable</code>	A matrix of size the number of nodes of the metanetwork times the number of networks. The rownames of this matrix must be the node names of metanetwork and the columns must be in an order corresponding to <code>gList</code> . The element (i,j) of this matrix is the abundance of species i in network j. Importantly, the non-null elements in each column of <code>abTable</code> must correspond to the nodes present in each element of <code>gList</code>

**Value**

Return a matrix whose elements are the pairwise dissimilarities.

**Author(s)**

Authors: Stephane Dray, Vincent Miele, Marc Ohlmann, Wilfried Thuiller Maintainer: Wilfried Thuiller <wilfried.thuiller@univ-grenoble-alpes.fr>

**References**

Marc Ohlmann, Vincent Miele, Stephane Dray, Loic Chalmandrier, Louise O'Connor & Wilfried Thuiller, Diversity indices for ecological networks: a unifying framework using Hill numbers. *Ecology Letters* (2019) <doi:10.1111/ele.13221>

**Examples**

```
# Generating a set of Erdos-Renyi graphs and give node names.
library(econetwork)
library(igraph)
nbGraph <- 10
gList <- c()
n <- 57 # number of nodes of each graph
C <- 0.1 # connectance of each graph
for(i in 1:nbGraph){
  graphLocal <- erdos.renyi.game(n, type='gnp', p.or.m=C, directed=TRUE)
  V(graphLocal)$name <- as.character(1:57)
  gList = c(gList,list(graphLocal))
}

# vector that gives the group of each node
groups <- c(rep("a",23),rep("b",34))
```

```

names(groups) <- as.character(1:57)
#generating random (non-nul) abundances data
abTable <- sapply(1:nbGraph,function(x) rpois(n,1)+1)
rownames(abTable) = unlist(unique(lapply(gList,function(g) V(g)$name)))

# Dissimilarity matrices based on links beta-diversity
# at a node level
disPairwise(gList, type = 'L')
# at a node level while taking into account node abundances
disPairwise(gList, type = 'L', abTable = abTable)
# at a group level
disPairwise(gList, groups, type = 'L')
# at a group level while taking into account node abundances
disPairwise(gList, groups, type = 'L', abTable = abTable)

```

---

divPartition

*Partitionning network diversity in alpha, beta and gamma diversity*


---

### Description

This function computes alpha, beta and gamma diversity of a list of networks. It measures either group, links, or probability of links diversity.

### Usage

```

divPartition(gList, groups, eta=1, framework=c('RLC','Chao'),
             type=c('P','L','Pi'), abTable=NULL)

```

### Arguments

gList	A list of graph objects of class igraph.
groups	A named vector of class character indicating the group to which each node belongs to. The length of groups must correspond to the number of different nodes present in gList. The names names(groups) must correspond to the nodes names in gList. If NULL, the groups are the initial nodes.
eta	A positive number that controls the weight given to abundant groups/links. Default value is 1.
framework	The framework used to partitionate diversity, either Reeve Leinster Cobbold ('RLC') or Chao ('Chao')
type	The type of diversity to measure and partitionate. It can be groups diversity ('P'), link diversity ('L') or probability of link diversity ('Pi').
abTable	A matrix of size the number of nodes of the metanetwork times the number of networks. The rownames of this matrix must be the node names of metanetwork and the columns must be in an order corresponding to gList. The element (i,j) of this matrix is the abundance of species i in network j. Importantly, the non-nul elements in each column of abTable must correspond to the nodes present in each element of gList

**Value**

Returns a list the following components:

mAlpha	The mean value of alpha-diversity across all networks.
Alphas	A vector of numeric containing the local alpha-diversities (i.e. the alpha-diversity value for each network).
Beta	The value of the overall beta-diversity
Gamma	The value of the gamma-diversity

**Author(s)**

Authors: Stephane Dray, Vincent Miele, Marc Ohlmann, Wilfried Thuiller  
Maintainer: Wilfried Thuiller <wilfried.thuiller@univ-grenoble-alpes.fr>

**References**

Marc Ohlmann, Vincent Miele, Stephane Dray, Loic Chalmandrier, Louise O'Connor & Wilfried Thuiller, Diversity indices for ecological networks: a unifying framework using Hill numbers. *Ecology Letters* (2019) <doi:10.1111/ele.13221>

**Examples**

```
# Generating a set of Erdos-Renyi graphs and give node names.
library(econetwork)
library(igraph)
nbGraph <- 10
gList <- c()
n <- 57 # number of nodes of each graph
C <- 0.1 # connectance of each graph
for(i in 1:nbGraph){
  graphLocal <- erdos.renyi.game(n, type='gnp', p.or.m=C, directed=TRUE)
  V(graphLocal)$name <- as.character(1:57)
  gList = c(gList,list(graphLocal))
}

# vector that gives the group of each node
groups <- c(rep("a",23),rep("b",34))
names(groups) <- as.character(1:57)
# generating random (non-nul) abundances data
abTable <- sapply(1:nbGraph,function(x) rpois(n,1)+1)
rownames(abTable) = unlist(unique(lapply(gList,function(g) V(g)$name)))

# Diversities in link abundances
# at a node level
divPartition(gList, framework='Chao', type = 'L')
# at a node level while taking into account node abundances
divPartition(gList, framework='Chao', type = 'L', abTable = abTable)
# at a group level
divPartition(gList, framework='Chao', groups, type = 'L')
# at a group level while taking into account node abundances
divPartition(gList, framework='Chao', groups, type = 'L', abTable = abTable)
```



---

`getMetaweb`*Get metaweb*

---

**Description**

Computation of the binary metaweb from a list of graph

**Usage**

```
getMetaweb(gList)
```

**Arguments**

`gList`            A list of graph objects of class `igraph`.

**Details**

This function computes the metaweb from a list of graph. It computes the union (in the sense of graph theory) of the set of graphs.

**Value**

`getMetaweb` returns an object of class `igraph`.

**Author(s)**

Authors: Stephane Dray, Vincent Miele, Marc Ohlmann, Wilfried Thuiller Maintainer: Wilfried Thuiller <wilfried.thuiller@univ-grenoble-alpes.fr>

**References**

Marc Ohlmann, Vincent Miele, Stephane Dray, Loic Chalmandrier, Louise O'Connor & Wilfried Thuiller, Diversity indices for ecological networks: a unifying framework using Hill numbers. *Ecology Letters* (2019) <doi:10.1111/ele.13221>

**Examples**

```
# Generating a set of Erdos-Renyi graphs
library(igraph)
gList <- c()
for(i in 1:4){
  graphLocal <- erdos.renyi.game(60, type = 'gnp', p.or.m=0.1, directed=TRUE)
  V(graphLocal)$name <- as.character(1:60)
  gList <- c(gList, list(graphLocal))
}
names(gList) <- c("A", "B", "C", "D")

## building the metaweb
graphMetaweb <- getMetaweb(gList)
```

# Index

- \*Topic **diversity**
  - econetwork-package, [2](#)
- \*Topic **ecology**
  - econetwork-package, [2](#)
- \*Topic **network**
  - econetwork-package, [2](#)
- \*Topic **package**
  - econetwork-package, [2](#)
  
- bipartite, [2](#)
  
- disPairwise, [2](#)
- divPartition, [4](#)
  
- econetwork (econetwork-package), [2](#)
- econetwork-package, [2](#)
- enaR, [2](#)
  
- getMetaweb, [6](#)
  
- igraph, [2](#)
  
- sna, [2](#)
- statnet, [2](#)
  
- vegan, [2](#)



# Bibliographie

---

## Sources primaires

- ALLESINA, Stefano et Mercedes PASCUAL, « Food web models : a plea for groups », *in* : *Ecology Letters* 12.7 (2009), p. 652-662.
- AMARASEKARE, Priyanga et al., « Mechanisms of coexistence in competitive metacommunities », *in* : *The American Naturalist* 164.3 (2004), p. 310-326.
- ARAÚJO, Miguel B et Alejandro ROZENFELD, « The geographic scaling of biotic interactions », *in* : *Ecography* 37.5 (2014), p. 406-415.
- ARRIBAS, Paula et al., « Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil », *in* : *Methods in Ecology and Evolution* 7.9 (2016), p. 1071-1081.
- AUCLAIR, Etienne, Nathalie PEYRARD et Régis SABBADIN, « Labeled DBN learning with community structure knowledge », *in* : *Joint european conference on machine learning and knowledge discovery in databases*, Springer, 2017, p. 158-174.
- AUSTEN, Punyasloke Bhadury Melanie C et al., « Molecular detection of marine nematodes from environmental samples : overcoming eukaryotic interference », *in* : *Aquatic Microbial Ecology* 44.1 (2006), p. 97-103.
- BACH, Francis R, « Graph kernels between point clouds », *in* : *Proceedings of the 25th international conference on Machine learning*, ACM, 2008, p. 25-32.
- BARNES, Andrew D et al., « Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning », *in* : *Nature communications* 5 (2014), p. 5351.
- BELLO, Carolina et al., « Atlantic frugivory : a plant–frugivore interaction data set for the Atlantic Forest », *in* : *Ecology* 98.6 (2017), p. 1729-1729, DOI : 10.1002/ecy.1818.
- BERLOW, Eric L et al., « Interaction strengths in food webs : issues and opportunities », *in* : *Journal of animal ecology* 73.3 (2004), p. 585-598.
- BERTNESS, Mark D et Ragan CALLAWAY, « Positive interactions in communities », *in* : *Trends in ecology & evolution* 9.5 (1994), p. 191-193.
- BOHAN, David A et al., « Next-generation global biomonitoring : large-scale, automated reconstruction of ecological networks », *in* : *Trends in Ecology & Evolution* 32.7 (2017), p. 477-487.
- BOYER, Frédéric et al., « obitools : A unix-inspired software package for DNA metabarcoding », *in* : *Molecular ecology resources* 16.1 (2016), p. 176-182.
- BRAGA, João et al., « Spatial analyses of multi-trophic terrestrial vertebrate assemblages in Europe », *in* : *Global Ecology and Biogeography* (2019), DOI : 10.1111/geb.12981.

- 
- CALCAGNO, Vincent et al., « Coexistence in a metacommunity : the competition–colonization trade-off is not dead », *in* : *Ecology letters* 9.8 (2006), p. 897-907.
- CANARD, Elsa et al., « Emergence of structural patterns in neutral trophic networks », *in* : *PloS one* 7.8 (2012), e38295.
- CARADONNA, Paul J et al., « Interaction rewiring and the rapid turnover of plant–pollinator networks », *in* : *Ecology letters* 20.3 (2017), p. 385-394.
- CHAO, Anne et Chun-Huo CHIU, « Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures », *in* : *Methods in Ecology and Evolution* 7 (2016).
- CHENG, Jie et al., « A sparse Ising model with covariates », *in* : *Biometrics* 70.4 (2014), p. 943-953.
- CHIQUET, Julien, Mahendra MARIADASSOU et Stéphane ROBIN, « Variational inference for sparse network reconstruction from count data », *in* : *arXiv preprint arXiv :1806.03120* (2018).
- CHOLER, Philippe, Richard MICHALET et Ragan M CALLAWAY, « Facilitation and competition on gradients in alpine plant communities », *in* : *Ecology* 82.12 (2001), p. 3295-3308.
- CLARK, Nicholas J, Konstans WELLS et Oscar LINDBERG, « Unravelling changing interspecific interactions across environmental gradients using Markov random fields », *in* : *Ecology* 99.6 (2018), p. 1277-1283.
- COHEN, Joel E, Frédéric BRIAND et Charles M NEWMAN, *Community food webs : data and theory*, t. 20, Springer, Berlin, 1990.
- COHEN, Joel E, Tomas JONSSON et Stephen R CARPENTER, « Ecological community description using the food web, species abundance, and body size », *in* : *Proceedings of the National Academy of Sciences* 100.4 (2003), p. 1781-1786.
- COINCE, Aurore et al., « Leaf and root-associated fungal assemblages do not follow similar elevational diversity patterns », *in* : *PloS one* 9.6 (2014), e100668.
- CORDIER, Tristan et al., « The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient », *in* : *New Phytologist* 196.2 (2012), p. 510-519.
- COUSINS, Steve, « The decline of the trophic level concept », *in* : *Trends in Ecology & Evolution* 2.10 (1987), p. 312-316.
- DALE, MRT et M-J FORTIN, « From graphs to spatial graphs », *in* : *Annual Review of Ecology, Evolution, and Systematics* 41 (2010).
- DAUDIN, J.-J., F. PICARD et S. ROBIN, « A mixture model for random graphs », *in* : *Statistics and Computing* 18.2 (2008), p. 173-183.
- DIGEL, Christoph et al., « Unravelling the complex structure of forest soil food webs : higher omnivory and more trophic levels », *in* : *Oikos* 123.10 (2014), p. 1157-1172.

- 
- DORMANN, Carsten F, Jochen FRÜND et H Martin SCHAEFER, « Identifying causes of patterns in ecological networks : opportunities and limitations », *in* : *Annual Review of Ecology, Evolution, and Systematics* 48 (2017), p. 559-584.
- DRAY, Stéphane, « A New Perspective about Moran's Coefficient : Spatial Autocorrelation as a Linear Regression Problem », *in* : *Geographical Analysis* 43.2 (2011), p. 127-141.
- DUNNE, Jennifer, *The network structure of food webs. In : Ecological Networks : Linking Structure to Dynamics in Food Webs*. Oxford : Oxford University Press, jan. 2006, p. 27-86.
- DVORSKÝ, Miroslav et al., « Testing the stress-gradient hypothesis at the roof of the world : effects of the cushion plant *Thylacospermum caespitosum* on species assemblages », *in* : *PLoS One* 8.1 (2013), e53514.
- ELTON, Charles S., *Animal ecology*, <https://www.biodiversitylibrary.org/bibliography/7435>, New York, Macmillan Co., 1927, URL : <https://www.biodiversitylibrary.org/item/31642>.
- EMMERT-STREIB, Frank, Matthias DEHMER et Yongtang SHI, « Fifty years of graph matching, network alignment and network comparison », *in* : *Information Sciences* 346 (2016), p. 180-197.
- ETIENNE, Rampal S et CJ NAGELKERKE, « Non-equilibria in small metapopulations : comparing the deterministic Levins model with its stochastic counterpart », *in* : *Journal of Theoretical Biology* 219.4 (2002), p. 463-478.
- FILOTAS, Elise et al., « The effect of positive interactions on community structure in a multi-species metacommunity model along an environmental gradient », *in* : *Ecological Modelling* 221.6 (2010), p. 885-894.
- FREILICH, Mara A et al., « Species co-occurrence networks : Can they reveal trophic and non-trophic interactions in ecological communities? », *in* : *Ecology* 99.3 (2018), p. 690-699.
- FRIEDMAN, Jerome, Trevor HASTIE et Robert TIBSHIRANI, « Sparse inverse covariance estimation with the graphical lasso », *in* : *Biostatistics* 9.3 (2008), p. 432-441.
- GAUZENS, Benoit et al., « Trophic groups and modules : Two levels of group detection in food webs », *in* : *J R Soc Interface* 12 (fév. 2015).
- GRAVEL, Dominique, Camille ALBOUY et Wilfried THULLER, « The meaning of functional trait composition of food webs for ecosystem functioning », *in* : *Philosophical Transactions of the Royal Society B : Biological Sciences* 371.1694 (2016), p. 20150268.
- GRAVEL, Dominique et al., « Trophic theory of island biogeography », *in* : *Ecology Letters* 14.10 (2011), p. 1010-1016.
- HAIRSTON, Nelson G, Frederick E SMITH et Lawrence B SLOBODKIN, « Community structure, population control, and competition », *in* : *The american naturalist* 94.879 (1960), p. 421-425.

- 
- HANAGE, William P, Christophe FRASER et Brian G SPRATT, « Fuzzy species among recombinogenic bacteria », *in* : *BMC biology* 3.1 (2005), p. 6.
- HANSKI, Ilkka et Otso OVASKAINEN, « Metapopulation theory for fragmented landscapes », *in* : *Theoretical population biology* 64.1 (2003), p. 119-127.
- HARRINGTON, Thomas C et David M RIZZO, « Defining species in the fungi », *in* : *Structure and dynamics of fungal populations*, Springer, 1999, p. 43-71.
- HARRIS, David J, « Inferring species interactions from co-occurrence data with Markov networks », *in* : *Ecology* 97.12 (2016), p. 3308-3314.
- HÉRAULT, Bruno, « Reconciling niche and neutrality through the Emergent Group approach », *in* : *Perspectives in Plant Ecology, Evolution and Systematics* 9.2 (2007), p. 71-78.
- HILL, M. O., « Diversity and Evenness : A Unifying Notation and Its Consequences », *in* : *Ecology* 54.2 (1973), p. 427-432, DOI : 10.2307/1934352.
- HUBBELL, Stephen P, *The unified neutral theory of biodiversity and biogeography (MPB-32)*, Princeton University Press, 2001.
- IBANEZ, Sébastien, « Optimizing size thresholds in a plant–pollinator interaction web : towards a mechanistic understanding of ecological networks », *in* : *Oecologia* 170.1 (2012), p. 233-242.
- IMRICH, Wilfried et Sandi KLAUZAR, *Product graphs : structure and recognition*, Wiley p.27, 2000.
- ISTAS, Jacques, *Mathematical modeling for the life sciences*, Springer Science & Business Media, 2006.
- JAKUSCHKIN, Boris et al., « Deciphering the pathobiome : intra-and interkingdom interactions involving the pathogen *Erysiphe alphitoides* », *in* : *Microbial ecology* 72.4 (2016), p. 870-880.
- KARST, Justine et al., « The mutualism–parasitism continuum in ectomycorrhizas : a quantitative assessment using meta-analysis », *in* : *Ecology* 89.4 (2008), p. 1032-1042.
- KÉFI, Sonia et al., « How structured is the entangled bank ? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience », *in* : *PLoS biology* 14.8 (2016), e1002527.
- KIVLIN, Stephanie N et al., « Biogeography of plant-associated fungal symbionts in mountain ecosystems : A meta-analysis », *in* : *Diversity and Distributions* 23.9 (2017), p. 1067-1077.
- KLECKA, Jan, « Modelling size structured food webs using a modified niche model with two predator traits », *in* : *PloS one* 9.8 (2014), e99355.
- LEVINS, Richard, « Some demographic and genetic consequences of environmental heterogeneity for biological control », *in* : *American Entomologist* 15.3 (1969), p. 237-240.
- LINDEMAN, Raymond L., « The Trophic-Dynamic Aspect of Ecology », *in* : *Ecology* 23.4 (1942), p. 399-417, DOI : 10.2307/1930126.

- 
- LIU, Han, Kathryn ROEDER et Larry WASSERMAN, « Stability approach to regularization selection (stars) for high dimensional graphical models », *in* : *Advances in neural information processing systems*, 2010, p. 1432-1440.
- LUCZKOVICH, Joseph J. et al., « Defining and Measuring Trophic Role Similarity in Food Webs Using Regular Equivalence », *in* : *Journal of Theoretical Biology* 220.3 (2003), p. 303 -321.
- MACARTHUR, Robert H et Edward O WILSON, *The theory of island biogeography*, t. 1, 2001, Princeton university press, 1965.
- MANDAKOVIC, Dinka et al., « Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience », *in* : *Scientific reports* 8.1 (2018), p. 5875.
- MARTINEZ-ALMOYNA, Camille et al., « Multi-trophic  $\beta$ -diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions », *in* : *Functional Ecology* (2019).
- MASSOL, François et al., « Island Biogeography of Food Webs », *in* : *Advances in Ecological Research* 56 (fév. 2017).
- MAY, Robert M, « Will a large complex system be stable? », *in* : *Nature* 238.5364 (1972), p. 413.
- MAYNARD, Daniel S, Carlos A SERVÁN et Stefano ALLESINA, « Network spandrels reflect ecological assembly », *in* : *Ecology letters* 21.3 (2018), p. 324-334.
- MILNS, Isobel, Colin M BEALE et V Anne SMITH, « Revealing ecological networks using Bayesian network inference algorithms », *in* : *Ecology* 91.7 (2010), p. 1892-1899.
- MOUQUET, Nicolas et al., « Extinction debt in source-sink metacommunities », *in* : *PLoS One* 6.3 (2011), e17567.
- MULLER, CB et al., « The structure of an aphid–parasitoid community », *in* : *Journal of Animal Ecology* 68.2 (1999), p. 346-370.
- NEWMAN, M.E.J. et E.A. LEICHT, « Mixture models and exploratory analysis in networks », *in* : *Proceedings of the National Academy of Science* 104 (juin 2007), p. 9564-9569.
- NGUYEN, Nhu H et al., « FUNGuild : an open annotation tool for parsing fungal community datasets by ecological guild », *in* : *Fungal Ecology* 20 (2016), p. 241-248.
- O’CONNOR, L. et al., « unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche », *in* : *Journal of Biogeography*, *in press* (2019).
- OGRAM, Andrew, Gary S SAYLER et Tamar BARKAY, « The extraction and purification of microbial DNA from sediments », *in* : *Journal of microbiological methods* 7.2-3 (1987), p. 57-66.
- OVASKAINEN, Otso, « The quasistationary distribution of the stochastic logistic model », *in* : *Journal of Applied Probability* 38.4 (2001), p. 898-907.



- 
- OVASKAINEN, Otso et Ilkka HANSKI, « Spatially structured metapopulation models : global and local assessment of metapopulation capacity », *in* : *Theoretical population biology* 60.4 (2001), p. 281-302.
- PEARL, Judea, « From Bayesian networks to causal networks », *in* : *Mathematical models for handling partial knowledge in artificial intelligence*, Springer, 1995, p. 157-182.
- PEAY, Kabir G et al., « Lack of host specificity leads to independent assortment of dipterocarps and ectomycorrhizal fungi across a soil fertility gradient », *in* : *Ecology Letters* 18.8 (2015), p. 807-816.
- PENNISI, Elizabeth, « A lichen ménage à trois », *in* : *Science* 353.6297 (2016), p. 337-337.
- PIÑOL, J et al., « A pragmatic approach to the analysis of diets of generalist predators : The use of next-generation sequencing with no blocking probes », *in* : *Molecular Ecology Resources* 14.1 (2014), p. 18-26.
- POISOT, Timothee et al., « Hosts, parasites and their interactions respond to different climatic variables », *in* : *Global ecology and biogeography* 26.8 (2017), p. 942-951.
- POISOT, Timothée et al., « The dissimilarity of species interaction networks », *in* : *Ecology Letters* 15.12 (2012), p. 1353-1361.
- POLIS, Gary A. et Donald R. STRONG, « Food Web Complexity and Community Dynamics », *in* : *The American Naturalist* 147.5 (1996), p. 813-846.
- REEVE, R. et al., « How to partition diversity », *in* : *arXiv :1404.6520* (avr. 2014), arXiv : 1404.6520 [q-bio.QM].
- RÉNYI, Alfréd, « On Measures of Entropy and Information », *in* : *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1 : Contributions to the Theory of Statistics*, Berkeley, Calif. : University of California Press, 1961, p. 547-561.
- ROSLIN, Tomas et al., « Higher predation risk for insect prey at low latitudes and elevations », *in* : *Science* 356.6339 (2017), p. 742-744.
- SALA, Enric et Michael H GRAHAM, « Community-wide distribution of predator-prey interaction strength in kelp forests », *in* : *Proceedings of the National Academy of Sciences* 99.6 (2002), p. 3678-3683.
- SALO, Pälvi et al., « Predator manipulation experiments : impacts on populations of terrestrial vertebrate prey », *in* : *Ecological monographs* 80.4 (2010), p. 531-546.
- SANDER, Elizabeth L, J Timothy WOOTTON et Stefano ALLESINA, « Ecological network inference from long-term presence-absence data », *in* : *Scientific reports* 7.1 (2017), p. 7154.
- SCHWALLER, Loic, S ROBIN et M STUMPF, « Bayesian inference of graphical model structures using trees », *in* : *arXiv preprint arXiv :1504.02723* 37 (2015).
- STANICZENKO, Phillip PA et al., « Linking macroecology and community ecology : refining predictions of species distributions using biotic interaction networks », *in* : *Ecology letters* 20.6 (2017), p. 693-707.

- 
- TABERLET, Pierre et al., *Environmental DNA : For biodiversity research and monitoring*, Oxford University Press, 2018.
- TABERLET, Pierre et al., « Towards next-generation biodiversity assessment using DNA metabarcoding », *in : Molecular ecology* 21.8 (2012), p. 2045-2050.
- TAUDIERE, Adrien et al., « Beyond ectomycorrhizal bipartite networks : projected networks demonstrate contrasted patterns between early-and late-successional plants in Corsica », *in : Frontiers in plant science* 6 (2015), p. 881.
- THOMPSON, Ross M. et al., « Food webs : Reconciling the structure and function of biodiversity », English (US), *in : Trends in Ecology and Evolution* 27.12 (déc. 2012), p. 689-697, ISSN : 0169-5347.
- THULLER, Wilfried et al., « Stochastic species turnover and stable coexistence in a species-rich, fire-prone plant community », *in : PloS One* 2.9 (2007), e938.
- TILMAN, David, Clarence L LEHMAN et Chengjun YIN, « Habitat destruction, dispersal, and deterministic extinction in competitive communities », *in : The American Naturalist* 149.3 (1997), p. 407-435.
- VANO, JA et al., « Chaos in low-dimensional Lotka–Volterra models of competition », *in : Nonlinearity* 19.10 (2006), p. 2391.
- VEECH, Joseph A, « A probabilistic model for analysing species co-occurrence », *in : Global Ecology and Biogeography* 22.2 (2013), p. 252-260.
- WEINSTEIN, Ben G et Catherine H GRAHAM, « Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions », *in : Ecology letters* 20.3 (2017), p. 326-335.
- WHITTAKER, R. H., « Vegetation of the Siskiyou Mountains, Oregon and California », *in : Ecological Monographs* 30.3 (1960), p. 279-338.
- WILLIAMS, Richard J et Neo D MARTINEZ, « Simple rules yield complex food webs », *in : Nature* 404.6774 (2000), p. 180.

---

## Sources secondaires

- BARABÁSI, Albert-László et al., *Network science*, Cambridge university press, 2016.
- BERGE, Claude, *Graphs and hypergraphs*, North-Holland Pub. Co., 1973.
- BERTRAND, Romain, *Le détail du monde*, L'univers historique, Seuil, 2019.
- GOETZMANN, William, *Les indiens. Aquarelles de Karl Bodmer*, Bibliothèque de l'Image, 1996.
- HADOT, Pierre, *Le voile d'Isis. Essai sur l'histoire de l'idée de nature*, 2004.
- HAECKEL, Ernst, *Kunstformen der Natur*, Bibliographisches Institut, 1904.
- HUMBOLDT, Alexandre von, *Essai sur la géographie des plantes ; accompagné d'un tableau physique des régions équinoxiales*. Levrault, Paris., 1805.
- *Lettre d'A. von Humboldt à V. von Ense*, citées in Marie Orr, *Le cosmos d'Alexandre von Humboldt*, 1834.
- ISTAS, Jacques, *Mathematical modeling for the life sciences*, Springer Science & Business Media, 2006.
- KOLLER, Daphne et Nir FRIEDMAN, *Probabilistic graphical models : principles and techniques*, MIT press, 2009.
- KORZYBSKI, Alfred, *Une carte n'est pas le territoire : prolégomènes aux systèmes non- aristotéliens et à la sémantique générale*, Éditions de l'Éclat, 1998.
- MUIR, John, « The Treasures of the Yosemite. », in : *Holt-Atherton Special Collections, University of the Pacific* (1890).
- MURPHY, Kevin P, *Machine learning : a probabilistic perspective*, MIT press, 2012.
- PLINE, l'Ancien, *Histoire naturelle. Livre XXIII (Remèdes tirés des arbres cultivés)* (traduction Jacques André), Collection des universités de France Série latine - Collection Budé, Paris, vers 77.
- PONGE, Francis, *Faune et Flore in Le parti pris des choses*, NRF-Gallimard, 1942.
- RECLUS, Élisée, « Du sentiment de la nature dans les sociétés modernes », in : *La Revue des deux Mondes* 63 (1866).
- SAINT-PIERRE, Jacques-Henri Bernardin de, *Etudes de la nature. Nouvelle édition, revue et corrigée*, Paris : Crapelet-Deterville, Paris, 1804.
- TAKEUCHI, Yasuhiro, *Global dynamical properties of Lotka-Volterra systems*, World Scientific, 1996.