Réseau bayésien dynamique étiqueté : cadre et apprentissage de structure pour application aux réseaux écologiques

Etienne Auclair

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Etienne AUCLAIR

Le 24 janvier 2019

Réseau bayésien dynamique étiqueté : cadre et apprentissage de structure pour application aux réseaux écologiques.

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Informatique et Télécommunications**

Unité de recherche :
MIAT-INRA : Unité de Mathématiques et Informatique Appliquées Toulouse

Thèse dirigée par
Régis SABBADIN et Nathalie PEYRARD

Jury

Mme Ghislaine GAYRAUD, Rapporteur
M. Christophe GONZALES, Rapporteur
Mme Gersande Fort, Examinateur
Mme Elisa Thebault, Examinateur
M. Régis SABBADIN, Directeur de thèse
Mme Nathalie PEYRARD, Directeur de thèse

Remerciements

Avant toute chose, je souhaite adresser des remerciements à toutes les personnes qui ont, d'une manière ou d'une autre, contribué au bon déroulé de cette thèse.

Tout d'abord, je remercie mes directeurs de thèse, Nathalie et Régis, pour leur patience, leurs conseils et leur disponibilité tout au long de cette thèse.

Je remercie également mes rapporteurs, Ghislaine Gayraud et Christophe Gonzales pour leur relecture attentive et leurs remarques sur ce manuscrit. Merci à vous, ainsi qu'à Élisa Thébault et Gersende Fort pour votre participation à mon jury de thèse.

J'adresse mes remerciements à l'ensemble de mon comité de thèse, Frédéric Koriche, Olivier Buffer et David Bohan pour leurs conseils avisés et leurs remarques pertinentes. Je souhaite d'ailleurs adresser des remerciements bien particuliers à David Bohan pour la mise à disposition de ses données expérimentales, son aide précieuse sur ces données et son accueil à Dijon.

Thanks to Laura Dee and Jennifer Caselle for their help on the Pisco dataset.

Je remercie bien entendu l'ensemble des membres l'unité MIAT qui m'ont entouré pendant ces années de thèse. Merci à Nathalie, Fabienne et Alain sans qui l'unité ne fonctionnerait pas aussi bien. Merci à tous ceux qui m'ont accompagné pendant ces années lors de pauses café, parties de carte, bars, soirées jeux et autres bons moment passés entre collègues et amis : Sebastian, Franck, Alyssa, Charlotte, Sara, Mohamed, Clément, Régis, Léo, David, Ludo, Floréal, Lise, Manon, Fulya, Faustine, Sylvain, Damien, Damien, Clémence, Malo, Émilien, Amélia, Marie-Anne, et ceux que j'ai pu oublier, ce dont je m'excuse.

Je remercie également ma famille pour leur soutien durant ces années de thèse, particulièrement mes parents, qui ont en plus apporté une contribution directe à cette thèse par leur relecture assidue.

Résumé

Un réseau écologique désigne l'ensemble des interactions entre les espèces vivantes d'un écosystème donné. En connaître la structure est un défi important dans le domaine de l'écologie. Cela peut se faire par des méthodes d'inférence, c'est à dire le fait d'utiliser des données d'observation écologique (l'abondance des espèce, leur présence/absence...) afin de reconstruire par des méthodes mathématiques les interactions en captant leur influence sur ces observations. Dans cette thèse, nous nous plaçons dans le cadre où les données écologiques dont on dispose sont des données de présence/absence d'espèces mesurées à différents pas de temps. Le but est de développer une méthode exploitant la dynamique de ces données pour apprendre les interactions entre les espèces. La difficulté réside dans le fait que des données binaires sont peu informatives. Des connaissances expertes sur le système étudié pourront aider à l'apprentissage.

Un cadre naturel pour apprendre une structure de réseau à partir de données binaires et dynamiques est celui des réseaux bayésiens dynamiques : les données de présence/absence temporelles sont modélisées comme des réalisations d'une série de variables aléatoires dynamiques dont les dépendances sont indiquées par un graphe orienté. Dans le cas où l'on n'a que peu de données, grâce à de la connaissance experte supplémentaire, il est possible de simplifier ce modèle. Cette thèse décrit un modèle particulier de réseau bayésien dynamique dit « étiqueté ». Ce modèle utilise un graphe dans lequel il existe un petit nombre de types d'interactions différentes, représentées par un petit nombre d'étiquettes attribuées à chaque arc. Ce modèle permet de décrire plusieurs phénomènes renseignant d'une information ou d'une perturbation pouvant se propager par contact (rumeur, maladie, feu de forêt...). Les probabilités de chaque variable sont calculées par une fonction dépendante du nombre d'interactions de chaque étiquette que cette variable subit. Ce modèle permet de décrire toutes les probabilités conditionnelles à l'aide d'un petit nombre de paramètres, indépendant de la structure du réseau. Ce cadre est utilisable pour modéliser la dynamique dans un réseau écologique : l'information diffusée est la présence ou l'absence d'une espèce, dépendant des interactions entre les espèces du réseau.

Nous décrivons ensuite une méthode permettant d'apprendre la structure d'un réseau bayésien dynamique étiqueté à l'aide d'observations de présence/absence d'espèces au cours du temps. Cet algorithme dit d' "estimation-restauration" alterne deux phases : une phase d'estimation de paramètres à structure fixée et une phase d'apprentissage de structure à paramètres fixés. Cette deuxième phase peut être complexe, et est résolue comme un problème de programmation linéaire en nombres entiers. Cela permet, en plus de l'utilisation d'outils efficaces pour la résolution de tels problèmes, d'ajouter de la connaissance experte sous forme de contraintes.

Ce procédé a été appliqué à un cas d'étude en particulier : l'observation d'espèces d'arthropodes piégés dans des champs expérimentaux au Royaume-Uni. Afin de constater les différences entre les cultures des parcelles, des réseaux différents ont été appris. Enfin, nous comparons ces réseaux à ceux obtenus par d'autres méthodes d'inférence de réseaux qui avaient été appliquées sur ces mêmes données.

Abstract

An ecological network represents the interactions between living species within an ecosystem. The knowledge of the structure of such a network is an important challenge in the field of ecology. This task can be realized by inference methods : a set of methods that uses ecological observations data (species abundance, presence or absence of species...) in order to learn the interactions mathematically, by the exploitation of the effect of these interactions on the observed data. This thesis describes a case where the ecological data we dispose of are only data of presence/absence of species observed at different moments. Le goal is to develop a method that exploits those kind of data in order to learn the interaction between these species. The main difficulty is that binary variables carry little information. Expert knowledge on the system is used to help learning the network's structure.

We use the framework of dynamic Bayesian network : temporal presence/absence data are modeled as the realization of a set of dynamic random variables whose dependencies are described by an oriented graph. Such a model can be simplified using expert knowledge. This thesis describes a particular model of "labelled" dynamic Bayesian network. In this model, the graph is defined by a small number of different types of interactions that constitute a set of labels attributed to the edges of the graph. This model can describe several phenomena where an information or a perturbation can be propagated by contact (rumour, disease, forest fire....) This model describes the presence or absence probabilities of each species as a function of the number of interactions of each label this species is subject to. This model allows to describe every presence/absence probability of species using a small number of parameters independent from the network's structure. This is the framework used for the modeling of species dynamics within an ecological network : the information propagated is the presence or the absence of a species, knowing the interaction between the species of the network.

Then, we describe the processes we use for learning the structure of a labelled dynamic Bayesian network using time series of binary variables. This 'Estimation-Restoration' algorithm alternates two steps : a phase of parameter estimation knowing the structure, and a phase of structure learning knowing the parameters. This last step can be complex. It is done by solving a integer linear programming problem. This allows to use efficient existing tools for solving those kind of problems. Moreover, we can easily add expert knowledge by the form of linear constraints.

This process has been used on a particular case study :the observation of arthropods species trapped in experimental fields in the united kingdom. In order to highlight the differences between the different crops, different networks have been learnt. Finally, we compare the learnt network with others, learnt with different learning methods on the same data.

Table des matières

Introduction

Contexte et objectif

L'écologie est un domaine d'étude très vaste, dont le but est la compréhension du milieu naturel, du monde vivant et des interactions entre eux. Toute cohabitation entre plusieurs êtres vivants, quelles que soient leurs espèces, ayant un minimum d'interaction peut donc faire l'objet d'une étude écologique, d'une forêt tropicale grouillante de biodiversité à un désert semblant hostile à presque toute forme de vie. Un champ d'étude aussi vaste est bien entendu divisé en plusieurs spécialités que l'on peut différencier par le niveau de détail du domaine d'étude [Schneider, 2001]. L'écologie peut s'étudier à l'échelle d'un individu, afin de comprendre l'impact de l'environnement sur un organisme ; d'une population, afin d'étudier les interactions dans un petit groupe d'individus d'une même espèce ; d'une communauté de plusieurs espèces ; et enfin à l'échelle du paysage. Les limites entre ces échelles sont floues et leurs définitions peuvent varier. Par exemple, entre l'échelle de la communauté et l'échelle du paysage, on peut définir l'échelle de l'écosystème, qui étudie les interactions entre toutes les espèces partageant un milieu relativement homogène.

Les interactions étudiées par l'écologie peuvent également être de plusieurs types. Une interaction essentielle pour le monde vivant est l'échange d'énergie et de nutriments à toutes les échelles écologiques. En effet, ce sont ces interactions, dites *trophiques*, qui permettent à la plupart des êtres vivants de survivre. L'étude de ces interactions se fait dans le cadre de l'écologie trophique [Garvey and Whiles, 2016]. Un outil important de cette discipline est le *réseau trophique*. Un réseau trophique décrit schématiquement l'ensemble des échanges d'énergie entre espèces. Autrement dit, c'est l'ensemble des relations proies/prédateurs entre des espèces vivantes au sein d'un écosystème. En pratique, l'exhaustivité est impossible, voire peu souhaitable : il existe énormément d'interactions trophiques possibles, et certaines se font dans des cas très particuliers. Un carnivore consommant de l'herbe pour se purger constitue-elle une interaction trophique ? Le fait qu'un prédateur consomme en période de disette une proie qu'il ne consomme pas habituellement constitue-t-il une relation trophique utile à étudier dans des conditions normales ? Un réseau trophique désigne donc davantage un outil utile dans le cadre de l'écologie trophique qu'un ensemble exhaustif de toutes les interactions trophiques.

La connaissance d'un réseau trophique est très utile dans de nombreux domaines de l'écologie [Palmer et al., 2016]. En effet, les réseaux trophiques permettent de mettre en évidence l'impact que peut avoir chaque espèce sur son écosystème, le moindre changement de population ou de régime d'une espèce se répercutant sur tout le réseau. L'étude des réseaux trophiques a permis par exemple d'identifier des espèces, notamment les prédateurs, considérées comme clé de voûte dans certains écosystèmes, c'est à dire que leur présence est garante de l'ensemble de la biodiversité [Pace et al., 1999, Soulé and Terborgh, 1999]. De telles espèces régulent les populations d'herbivores,

donc contribuent à maintenir un couvert végétal suffisant pour l'intégralité de l'écosystème. Connaître des réseaux trophiques peut ainsi aider à la gestion et à la régulation des écosystèmes, en guidant l'effort de conservation sur certaines espèces [Silliman and Bertness, 2002]. Les réseaux trophiques permettent également de mieux comprendre l'impact d'espèces introduites ou invasives sur un écosystème afin d'en réguler artificiellement la population [Knapp et al., 2001].

Bien que très utile, l'utilisation des réseaux trophiques a ses limites. Ses premières limitations viennent, comme pour tout le domaine de l'écologie, du manque de clarté du concept d'espèce : à partir de quel moment peut-on dire que deux individus sont d'une espèce différente ? La définition du concept d'écosystème n'est pas claire non plus : où commencent et ou finissent deux écosystèmes considérés comme différents ?

Et au delà de ces questions purement scientifiques se pose un problème pratique : comment peut-on réellement connaître la structure d'un réseau trophique ? Heureusement, il est possible d'identifier des réseaux trophiques plus ou moins fiables, à l'aide de méthodes plus ou moins coûteuses. Cependant, les interactions trophiques sont un type d'interaction précis que ces méthodes ne peuvent pas toujours identifier formellement. Dans ce cas, il peut être plus efficace de prendre en compte les relations biologiques autres que trophiques, comme les relations de symbiose, de parasitisme, de compétition... Un réseau d'interactions regroupant les interactions trophiques et d'autres types d'interactions biologiques est désigné comme un *réseau écologique*.

Les méthodes permettant de connaître ces interactions écologiques peuvent demander un grand nombre de données écologiques et nécessitent l'intervention de disciplines mathématiques, statistiques et informatiques. Les interactions entre les espèces ont en effet un impact sur divers indicateurs mesurables, comme l'abondance, la présence ou l'absence de diverses espèces d'un écosystème. À partir de l'observation de ces indicateurs, des méthodes statistiques, mathématiques ou informatiques cherchent à trouver la part de ces indicateurs expliquée par les interactions entre espèces. L'*inférence* de réseau désigne le fait de chercher les interactions les plus plausibles à partir de données écologiques. Le principe des méthodes d'inférence consiste généralement à construire un modèle théorique décrivant comment obtenir des indicateurs biologiques à partir des interactions écologiques. L'idée est de trouver une structure de graphe permettant d'expliquer au mieux les indicateurs biologiques observés d'après le modèle. Le principe d'inférence de réseau se base sur le fait que les interactions écologiques ont un impact suffisamment fort sur les indicateurs biologiques pour qu'elles soient captées par le modèle. Cette hypothèse semble crédible [Polis et al., 2000].

Cependant, la récolte de ces données écologiques a un certain coût, surtout lorsque l'on souhaite obtenir des données précises et détaillées. Lorsqu'un écosystème est très mal connu ou difficile à observer, ces données peuvent être peu informatives. Le but de cette thèse est de proposer une méthode permettant d'inférer un réseau écologique à partir de données peu informatives. Les données que l'on cherchera à exploiter sont des données de présence/absence d'espèces au cours du temps. Le principe est de construire un modèle théorique faisant intervenir les interactions entre espèces. Ce modèle permet d'approcher la manière dont les données temporelles de présence/absence peuvent être théoriquement obtenues, et d'utiliser des observations réelles afin d'évaluer ce modèle. En résumé, l'objectif de cette thèse est de **développer une méthode permettant d'inférer un réseau écologique à partir de données temporelles de présence/absence d'espèces**.

Méthodologie et contributions

La méthode développée dans cette thèse se base sur un modèle particulier, dit de réseau bayésien dynamique étiqueté développé pour cette thèse. C'est un cas particulier de réseau bayésien

[Pearl, 1986]. Un réseau bayésien est un modèle regroupant un ensemble de variables aléatoires dont les dépendances sont représentées sous la forme d'un graphe orienté sans circuit. Cette structure permet de faciliter l'expression de la loi de probabilité jointe de l'ensemble des variables en exploitant les indépendances conditionnelles. Les réseaux bayésiens dynamiques sont un cas particulier de réseaux bayésiens dont les variables aléatoires suivent une dynamique temporelle. Si l'on considère la présence ou l'absence d'une espèce à un moment donné comme le résultat d'une variable aléatoire dont la probabilité dépend des interactions écologiques avec les autres espèces, la dynamique des espèces d'un réseau écologique peut être modélisée par un réseau bayésien dynamique.

Les réseaux bayésiens sont souvent utilisés pour calculer la probabilité d'une ou de plusieurs variables aléatoires sachant d'autres, ou pour déterminer une structure de graphe méconnue. La procédure pour déterminer ou *apprendre* cette structure est souvent désignée comme l'inférence ou l'*apprentissage* de réseau bayésien. Deux familles de méthodes existent afin d'apprendre cette structure à partir d'un ensemble de données constituant les observations des variables aléatoires du réseau bayésien [Margaritis, 2003]. Une première famille de méthodes consiste à tester empiriquement les indépendances des variables conditionnellement aux autres afin d'obtenir une structure de graphe cohérente par rapport à un ensemble d'observations. Une autre famille de méthodes consiste à établir un score caractérisant la pertinence d'un graphe pour modéliser les données observées par un réseau bayésien. Ces méthodes utilisent des procédures permettant de trouver la structure maximisant ce score d'après les données observées. Cependant, pour un graphe donné, un grand nombre de probabilités conditionnelles différentes doivent être calculées. Le nombre de paramètres augmente avec le nombre d'arcs dans le graphe et leur estimation est difficile lorsque l'on a peu de données. Nous souhaiterions utiliser des connaissances expertes afin de diminuer ce nombre de paramètres.

Un réseau bayésien étiqueté [Auclair et al., 2017] est un réseau bayésien dans lequel les arcs du graphe associé portent une information qualitative appelée étiquette. L'étiquette d'un arc entre deux variables permet de caractériser la nature de l'interaction qu'elle représente, c'est à dire le type d'impact qu'une variable a sur l'autre et la force de cette interaction. Dans un réseau bayésien étiqueté, le nombre d'étiquettes est fixé à priori et les probabilités conditionnelles de chaque variable ne sont déterminées que par le nombre de ses parents et par les étiquettes des arcs correspondantes. Cela permet de rendre le modèle plus compact en réduisant le nombre de paramètres de ce modèle. En effet, le nombre de paramètres est dépendant du nombre d'étiquettes fixées au départ et pas de la structure du graphe. Ainsi, en connaissant le nombre d'étiquettes d'un réseau bayésien étiqueté, le nombre de paramètre est fixe, et indépendant du nombre de variables et d'arcs. Une version dynamique de ce modèle est aussi décrite dans cette thèse.

Ce cadre est pertinent avec l'objectif d'apprentissage de réseaux écologiques, car des connaissances expertes existent sur ce type de réseau. Il est possible, avec l'aide d'experts, de définir par avance les types d'interactions écologiques que l'on s'attend à trouver dans l'écosystème étudié, et de se limiter à ces types d'interactions dans le modèle théorique afin de simplifier le problème. Typiquement, il s'agit soit d'interactions qui favorisent la présence d'une espèce, soit qui la défavorisent. Un tel modèle n'est d'ailleurs pas spécifique aux réseaux écologiques et peut être utile dans de nombreuses applications.

Si un réseau bayésien étiqueté réduit le nombre de paramètres du modèle, cela se fait au prix d'une expression de la vraisemblance plus difficile à dériver que dans le cas classique. De ce fait, les méthodes d'apprentissage doivent être ajustées dans le cadre des réseaux bayésiens étiquetés. Cette thèse présente une méthode permettant d'apprendre la structure d'un réseau bayésien dynamique étiqueté. Cet apprentissage se fait par un algorithme en deux étapes successives : une étape d'estimation des paramètres à structure de graphe fixée et une étape d'apprentissage à valeurs de paramètres fixées.

L'étape d'apprentissage de structure se fait par la résolution d'un problème de programmation linéaire en nombres entiers. La résolution de ce problème par programme linéaire permet d'utiliser des outils déjà existants et d'ajouter de la connaissance experte sous forme de contraintes linéaires.

Parmi ces connaissances expertes, on peut intégrer par exemple la connaissance de certains arcs, ou des niveaux trophiques des différentes espèces lorsque l'on cherche à apprendre un réseau écologique. La connaissance de ces niveaux trophiques permettra alors de prioriser certains arcs, afin d'éviter par exemple qu'une espèce basale soit interprétée comme une espèce prédatrice.

Plan du manuscrit

Ce manuscrit commence par décrire les concepts de réseau trophique et de réseau écologique et par présenter un ensemble de méthodes permettant d'apprendre la structure de ces réseaux dans le chapitre 1. Ce chapitre décrit plus précisément les méthodes d'inférence, utilisant des outils mathématiques, statistiques et informatiques qui ont déjà été utilisés afin d'estimer les interactions écologiques à partir de données écologiques.

Le chapitre 2 décrit en particulier une famille de méthodes permettant l'inférence d'un réseau qui reposent sur le cadre des *réseaux bayésiens*. Il existe plusieurs méthodes d'apprentissage utilisant l'intelligence artificielle ou la statistique afin de trouver la structure de réseaux bayésiens la plus plausible. Ces méthodes sont présentées dans ce chapitre.

Le chapitre 3 propose un nouveau modèle pour la représentation de connaissances incertaines appelé réseau bayésien étiqueté. Ce chapitre montre également la généricité de ce cadre, notamment par des exemples de phénomènes modélisables par réseau bayésien étiqueté. Il est notamment possible de modéliser un réseau écologique, car au sein d'un écosystème, la présence d'une espèce peut avoir sur une autre un impact positif ou négatif. Par exemple, une proie sert de nourriture à son prédateur. Il s'agit d'un impact positif pour le prédateur, et négatif pour sa proie.

Le chapitre 4 décrit une procédure permettant d'apprendre la structure d'un réseau dans le cadre des réseaux bayésiens étiquetés dynamiques, cas particulier du modèle décrit précédemment utile pour représenter tout phénomène évoluant dans le temps. Cet apprentissage se fait en résolvant un programme linéaire en nombres entiers décrit dans ce chapitre. Il présente également une étude de performance de cette procédure sur des données simulées, puis sur un jeu de données réelles.

Le chapitre 5 présente un cas d'étude sur des données réelles. Ces données sont issues d'une étude sur des champs expérimentaux où des arthropodes ont été piégés. Nous chercherons alors à apprendre les interactions entre ces arthropodes. Nous souhaitons également étudier les différences entre les réseaux appris pour différentes cultures. L'objectif est également de comparer ces résultats avec ceux issus d'une précédente étude sur ces données.

Chapitre 1

Apprentissage de structure de réseau écologique

Il existe un si grand nombre d'interactions possibles entre les espèces vivantes qu'en avoir la connaissance exhaustive est très difficile. Plusieurs méthodes, que nous développerons dans ce chapitre, sont utilisées pour apprendre la structure d'un réseau écologique. Ces méthodes peuvent être coûteuses en matériel ou en temps. De ce fait, le réseau trophique de certains écosystèmes très bien étudiés peut être bien mieux connu que ceux d'autres écosystèmes. Existe-t-il cependant une méthode qui ne soit pas limitée à un écosystème particulier, donc qui puisse reconstruire un réseau trophique, y compris pour un écosystème pour lequel nous disposons de peu d'informations ? Il serait utile d'avoir une méthode permettant d'exploiter des informations pouvant être assez peu informatives, mais simples à obtenir, comme la présence ou l'absence d'espèces vivantes dans un écosystème. Les données binaires fournissent peu d'informations par elles-même, mais il est possible de les combiner avec d'autres types d'informations assez simples à obtenir. Une information de dynamique temporelle permettrait de jouer ce rôle. En effet, l'étude d'un écosystème se faisant généralement sur plusieurs périodes temporelles, nous souhaitons pouvoir exploiter la dynamique des espèces peuplant cet écosystème à l'aide de l'information sur leur présence ou leur absence au cours du temps. Il n'y a également aucune raison d'occulter les connaissances que peuvent déjà avoir les experts sur l'écosystème étudié et les espèces le peuplant. Pouvoir exploiter des données peu informatives combinées à des connaissances expertes peut être une approche intéressante pour obtenir des réseaux écologiques exploitables dans un contexte où peu de données sont disponibles. Passons en revue les principales méthodes de construction de réseau écologique dans le but de trouver une méthode de construction de réseau écologique permettant d'exploiter :

— Des données de présence/absence d'espèces

— Des données dynamiques

— Des connaissances expertes

Après avoir donné une définition plus détaillée du concept de réseau écologique, nous allons passer rapidement en revue un ensemble de méthodes d'observations directe d'interactions pour construire un réseau trophique. Nous présenterons ensuite un ensemble de méthodes permettant de reconstruire des réseaux trophiques à partir de données indirectes.

1.1 Réseau écologique

1.1.1 Définition

Un réseau trophique [Dunne, 2009] constitue un ensemble de relations proie/prédateur entre des espèces. La figure 1.1 est une illustration d'un réseau trophique entre un ensemble d'espèces vivantes, une flèche d'une espèce vers une autre représente une relation proie→prédateur.

Un réseau trophique est donc un ensemble de chaînes alimentaires, qui peuvent s'entrecroiser et former un réseau ayant une structure pouvant être complexe. Généralement, on définit un réseau trophique pour un groupe d'espèces au sein d'un écosystème. La forme de ces réseaux trophiques peut énormément varier selon les écosystèmes[Pimm et al., 1991], ou lors d'un changement dans un écosystème [Schmidt et al., 2009] même lorsque les mêmes espèces sont considérées. De ce fait, un réseau trophique ne peut jamais être exhaustif. Il est possible d'ajouter des informations supplémentaires à ces réseaux, comme des descriptions quantitatives ou qualitatives [Bersier et al., 2002, Banašek-Richter et al., 2004] sur les espèces ou les interactions. Bien que les interactions trophiques (les relations proie/prédateur) soient les plus étudiées, il existe d'autres types d'interactions [van Veen et al., 2009] qui peuvent avoir une importance capitale sur les espèces [Kéfi et al., 2012]. Il ne faut effectivement pas nier l'importance d'interactions comme les relations de facilitation entre espèce (symbioses, plantes ou algue servant d'habitat à une espèce vivante), les relations de compétition pour une ressource, ou les relations de parasitisme. Ce type d'information peut être ajouté au réseau trophique, afin d'obtenir un réseau d'interaction plus étendu. Un réseau renseignant n'importe quel type d'interaction écologique est désigné comme **réseau écologique** [Ings et al., 2009]. Un réseau trophique est donc un cas particulier de réseau écologique. La structure d'un réseau écologique fait qu'il est facilement manipulable mathématiquement, à l'aide d'une branche des mathématique appelée **théorie des graphes** dont nous allons présenter ici quelques principes élémentaires.

1.1.2 Notions de théorie des graphes

Définition

Un **graphe** est un objet mathématique utilisé pour représenter des entités homogènes liées par des relations binaires. Cette représentation permet de gérer efficacement tout type de réseau (routier, social, trophique...).

Un graphe $\mathcal{G}$ est constitué d'un ensemble de **nœuds** $\mathcal{V}$ représentant les entités que l'on souhaite étudier. Deux nœuds peuvent être liés ensemble par une relation r que l'on cherche à étudier. L'ensemble des relations r entre les nœuds $\mathcal{V}$ constitue un ensemble d'**arcs** $\mathcal{E}$. On note $\mathcal{G}(\mathcal{V}, \mathcal{E})$ un graphe constitué des nœuds $\mathcal{V}$ et des arcs $\mathcal{E}$.

Si la relation r que décrit le graphe est symétrique, c'est à dire que $\mathcal{V}_1 r \mathcal{V}_2 \Leftrightarrow \mathcal{V}_2 r \mathcal{V}_1$, le graphe est dit **non-orienté** [1]. Le vocabulaire et les exemples présentés dans cette section correspondent au cas **orienté**, où $\mathcal{G}$ décrit des relations non symétriques.

La figure 1.2 représente un graphe orienté $\mathcal{G}(\mathcal{V}, \mathcal{E})$ où $\mathcal{V} = \{1, 2, 3, 4, 5\}$ et $\mathcal{E} = \{(1 \to 3), (2 \to 3), (2 \to 4), (3 \to 5), (4 \to 5), (5 \to 1)\}$

1. Le vocabulaire change légèrement lorsque l'on traite un graphe non-orienté. Dans le cas non-orienté, on désigne souvent les entités comme des *sommets* et les relations comme des *arêtes*. En pratique, ces termes sont interchangeables et il est fréquent d'entendre parler d'arête dans le cas orienté, de nœud dans le cas non-orienté etc.

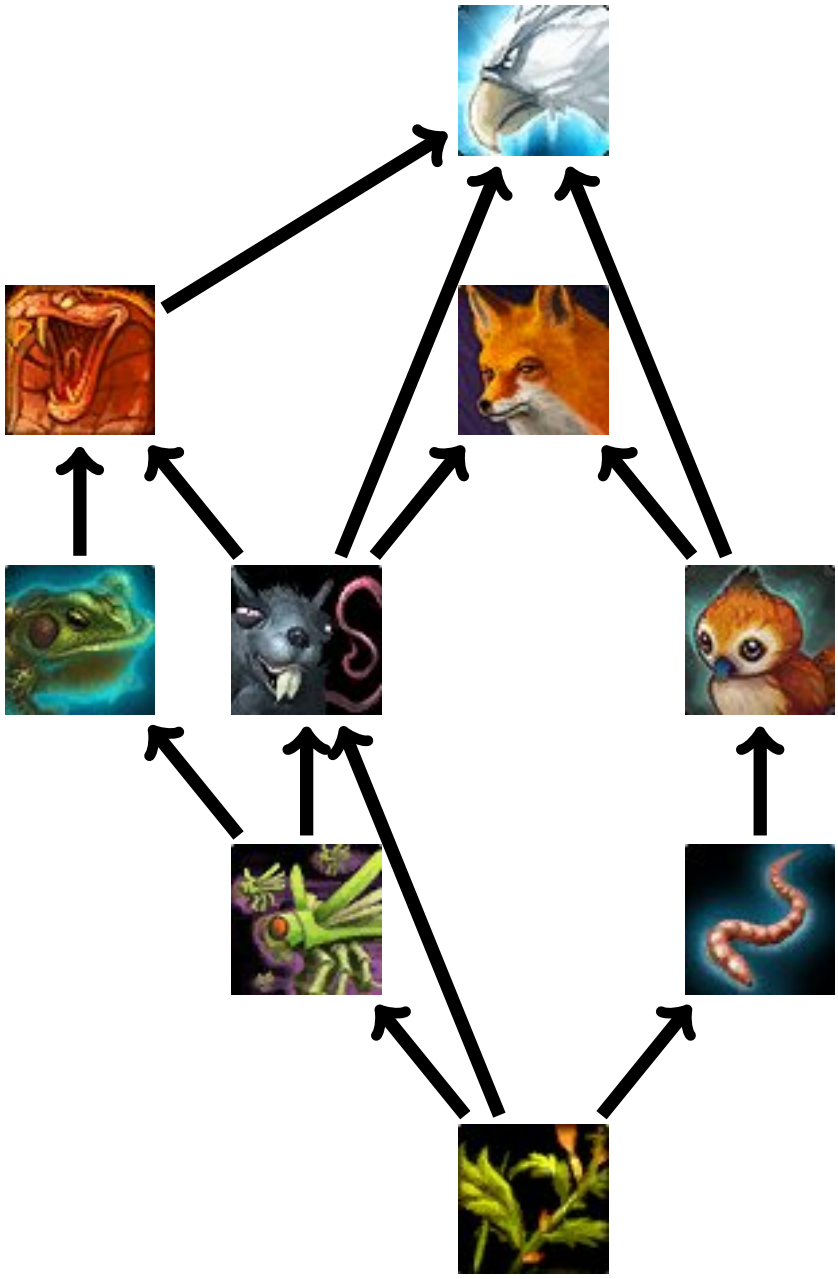


FIGURE 1.1 – Exemple simple de réseau trophique. Une flèche d'une espèce i vers une espèce j décrit le fait que i est une proie de j.

Éléments de vocabulaire

Un nœud V_1 est dit **parent** d'un nœud V_2 si il existe un arc $E(V_1, V_2)$ de V_1 vers V_2. L'ensemble des parents d'un nœud V est noté π_V. Sur le graphe $\mathcal{G}(\mathcal{V}, \mathcal{E})$ de la figure 1.2, les parents du nœud 3 constituent l'ensemble $\pi_3 = \{1, 2\}$.
Un **chemin** est une succession d'arcs distincts reliant une série de nœuds de façon continue. On peut par exemple définir dans le graphe $\mathcal{G}$ de la figure 1.2 le chemin $2 \to 4 \to 5 \to 1$ reliant le nœuds 2 au nœud 1.
Un **circuit** est un chemin reliant un nœud à lui-même. Dans le graphe $\mathcal{G}$ de la figure 1.2, le chemin $1 \to 3 \to 5 \to 1$ est un circuit.

Un graphe peut porter des informations supplémentaires. Les informations les plus courantes que l'on peut trouver dans des graphes sont des valeurs quantitatives ou qualitatives sur les nœuds ou les arcs. Un réseau routier, par exemple, peut se modéliser comme un graphe dont les nœuds sont des villes, les arcs des routes entre ces villes, chaque arc étant porteur d'une valeur correspondant à la longueur de la route. Un graphe portant un poids sur chaque arc est dit *valué*. La théorie des graphes est un ensemble de méthodes mathématiques et algorithmiques permettant de manipuler les graphes. Il existe un très grand nombre d'algorithmes pour un très grand nombre d'applications. On peut citer, parmi les plus courants, des algorithmes permettant de trouver le plus court chemin entre deux nœuds, de la détection de circuits dans un graphe, ou encore différentes manières d'étiqueter les nœuds selon leurs liens. Plusieurs de ces méthodes se basent sur la *matrice d'adjacence* d'un graphe. La *matrice d'adjacence* est une manière de représenter mathématiquement un graphe, à l'aide d'une matrice de taille $n * n$, n étant le nombre de nœuds. Un élément $A_{i,j} \in \{0, 1\}$ de cette matrice d'adjacence situé sur la ligne i et la colonne j décrit la présence ou l'absence d'un arc d'un nœud i vers un nœud j. La matrice d'adjacence d'un graphe non-orienté est symétrique. Un élément de la matrice d'adjacence d'un graphe valué peut prendre pour valeur le poids de l'arc qu'il représente au lieu de la simple information de sa présence, un arc inexistant ayant un poids de 0.

1.1.3 Réseau écologique et théorie des graphes

La théorie des graphes peut s'appliquer à tout type de problème assimilable à un réseau. Ainsi, un réseau écologique est simple à adapter dans ce cadre. On désigne alors un réseau écologique comme un graphe orienté dont les nœuds sont les espèces et les arcs sont les relations trophiques. Une relation écologique n'étant pas symétrique (dans le cas d'une relation trophique, il est rare qu'une proie soit également prédateur de son prédateur), un réseau écologique peut être représenté par un graphe orienté. La figure 1.3 représente le réseau trophique représenté dans la figure 1.1 sous la forme d'un graphe orienté à 9 nœuds, chaque nœud représentant une espèce.

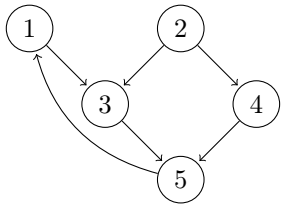


FIGURE 1.2 – Exemple de graphe orienté $\mathcal{G}(\mathcal{V}, \mathcal{E})$

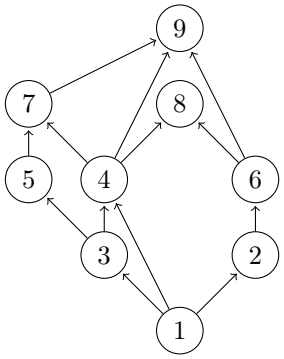


FIGURE 1.3 – Réseau trophique similaire à celui de la figure 1.1 modélisé sous forme de graphe orienté.

Ainsi, le problème de construction de réseau écologique peut se voir mathématiquement comme celui de trouver les arcs d'un graphe orienté dont les nœuds sont les espèces. La prochaine section va un peu plus loin dans la caractérisation d'un réseau écologique. Nous nous concentrerons sur les réseaux trophiques, en listant différentes formes qu'ils peuvent prendre.

1.2 Structure d'un réseau trophique

Si il est possible de définir des réseaux écologiques pour différents écosystèmes, différents milieux et différentes espèces, donc potentiellement très différents, il existe tout de même des régularités dans les structures de réseau écologique. Ces structures "type" de réseau écologique fournissent de l'information experte supplémentaire, que l'on peut intégrer à l'apprentissage de ces réseaux. Il est donc intéressant de caractériser la manière dont s'organise un réseau écologique. Cependant, les structures les plus étudiées sont celles des réseaux trophiques, c'est à dires des réseaux écologiques renseignant uniquement des interactions proie/prédateur. Il existe différentes manières de modéliser un réseau trophique [Allesina, 2008] et il est intéressant de connaître les règles qu'utilisent ces modèles afin de comprendre les mécanismes de ces réseaux. Ces modèles ont été établis par comparaison entre des réseaux simulés à l'aide de différents modèles et des réseaux connus. Le modèle de réseau trophique le plus simple est un graphe aléatoire, où tout lien entre espèce est équiprobable. Cependant, les relations trophiques sont plus complexes qu'une règle aléatoire aussi simple. Voyons donc quelques modèles plus complexes permettant de modéliser un réseau trophique.

1.2.1 Modèle de cascade

Le modèle de cascade [Pimm et al., 1991] permet de générer un profil de réseau trophique se basant sur un ordre hiérarchique entre toutes les espèces. À chaque espèce $i(i \in 1, ..., n)$ est attribuée une valeur v_i unique permettant d'établir un ordre hiérarchique entre les espèces permettant d'établir les règles suivantes :

— Une espèce i peut se nourrir sur des espèces dont la valeur est plus faible : Si $v_j < v_i$, alors i peur se nourrir sur l'espèce j.

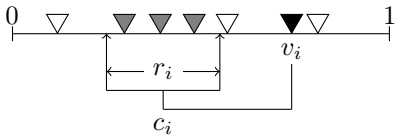


FIGURE 1.4 – Illustration du fonctionnement du modèle de niche. Chaque triangle représente une espèce positionnée sur un axe selon leur valeur de niche entre 0 et 1. Le triangle noir représente une espèce i de valeur de niche v_i ayant un intervalle de niche de centre c_i et de taille r_i. Toutes les espèces dans cet intervalle (en gris) sont ses proies.

— Une espèce i ne peut pas se nourrir sur des espèces de valeur aussi forte ou plus forte : Si $v_k >= v_i$, alors i ne peut pas se nourrir sur l'espèce k.

— Un lien trophique entre une espèce i et une espèce j a une probabilité d/n d'exister, d étant un paramètre à fixer ou à apprendre.

Le paramètre d est indépendant du nombre d'espèces n et ne varie pas avec i. La probabilité de présence d'une interaction trophique est donc équiprobable pour chaque espèce tant que les règles ci-dessus sont respectées. Excepté pour les espèces numérotées 1 et n, respectivement espèce basale et grand prédateur, ce modèle ne fixe pas les niveaux trophiques à priori : une espèce ayant une numérotation forte n'a pas nécessairement un niveau trophique plus élevé qu'une espèce avec une numérotation faible. Cependant, par construction, dans ce modèle, le nombre de proies probables augmente avec la numérotation des espèces. Si cela imite des comportement observés dans la nature, ce modèle occulte cependant plusieurs comportements trophiques. Il interdit notamment le cannibalisme ou tout cycle dans le réseau. Or, ce cas de figure n'est pas impossible dans la réalité [Polis, 1991]. En effet, plusieurs espèces peuvent pratiquer le cannibalisme ou consommer des espèces habituellement prédatrices.

1.2.2 Modèle de niche

Le modèle de niche [Williams and Martinez, 2000] propose un profil de réseaux trophiques plus affiné par rapport au modèle de cascade. Comme pour le modèle de cascade, à toute espèce i est attribuée une valeur. Cette "valeur de niche" v_i est unique et suit une loi uniforme. À toute espèce i est également attribué un intervalle de centre $c_i < v_i$ et de longueur r_i. L'espèce i a pour proie toute espèce j ayant sa valeur v_j dans cet intervalle, soit toute espèce j telle que $c_i - r_i/2 \geq v_j \geq c_i + r_i/2$. Le fonctionnement du modèle de niche est illustré par la figure 1.4.

Cette règle autorise les boucles et le cannibalisme, car la borne supérieure $c_i + r_i/2$ peut dépasser la valeur v_i.

1.2.3 Modèle de hiérarchie imbriquée

Le modèle de niche impose aux espèces des régimes alimentaires "continus" selon les valeurs de niche. C'est à dire qu'une espèce ayant pour proie deux espèces de valeur de niche v_j et v_k aura toujours pour proie toutes les espèces ayant une valeur de niche comprise entre v_j et v_k. Cette contrainte n'est pas toujours observée dans la réalité. Le modèle de hiérarchie imbriquée [Cattin et al., 2004] (Plus souvent trouvé sous le nom de *nested hierarchy* dans la littérature) se

base également sur des valeurs de niches v_i permettant d'ordonner les espèces.À chaque espèce est attribué un ensemble de proies, dans l'ordre croissant des valeurs de niches par la procédure suivante :

Pour chaque espèce i, dans l'ordre des valeurs de niche v_i :

— Appliquer la procédure du modèle de niche [Williams and Martinez, 2000] afin de déterminer un nombre de proies p_i.

— Une proie j est choisie aléatoirement parmi les espèces de valeur inférieure ($v_j < v_i$).

— Une fois que la proie j de i a été choisie, deux cas de figures sont possibles :

(1) j n'a pas de prédateurs. Dans ce cas, i est désigné comme prédateur de j et la prochaine proie de i est choisie aléatoirement parmi les espèces de valeur inférieure ($v_j < v_i$). Retour à l'étape précédente.

(2) j a au moins un prédateur. Dans ce cas, i rejoint l'ensemble des prédateurs de j Pr_j. Le reste des prochaines proies de i est choisi aléatoirement parmi l'ensemble des proies connues de Pr_j.

— L'espèce i ne peut pas choisir deux fois la même proie. Si, lors de l'étape (2), aucune proie n'est disponible, car l'ensemble des proies de Pr_j est trop petit, alors le reste de ses proies est choisi aléatoirement parmi les espèces n'ayant pas encore de prédateur et d'une valeur inférieure à v_i.

— Si i doit encore trouver des proies et qu'aucune proie ne respectant ces règles ne peut lui être attribuée, alors le reste de ses proies sont choisies aléatoirement. Ces proies peuvent avoir n'importe quelle valeur, y compris une valeur supérieure à v_i.

Ce modèle permet d'intégrer un comportement existant dans la nature qui n'était pas inclus dans les modèles précédents : le fait que des prédateurs partagent un même groupe de proies, ce qui est fréquent dans la nature.

Ces modèles permettent de modéliser n'importe quel réseau trophique, sans avoir besoin de connaissances particulières sur les espèces étudiées. Cependant, lorsque ces connaissances sont disponibles, il est possible de modéliser d'autres types de réseaux trophiques, plus précis car tenant compte des spécificités connus du milieu et des espèces étudiées.

1.2.4 Modèle à blocs latents

La classification croisée est une méthode de classification non-supervisée consistant à partitionner un ensemble de variables et d'individus en classes homogènes. Un bloc est un regroupement d'ensembles conjoints d'individus et de variables ayant un comportement similaire [Brault, 2014]. Un modèle à blocs latents se construit à la manière d'une classification croisée où les partitions de lignes et de colonnes à regrouper en classes sont considérées comme des variables latentes. À chaque individu est associée une probabilité d'appartenance à chaque classe d'individu, et à chaque variable est associée une probabilité d'appartenance à chaque classe de variable. Le croisement d'une classe de variable et d'une classe d'individu constitue une classe associée à l'observation. Chacune de ces probabilités est considérée comme indépendante. On considère également que la valeur d'une variable pour un individu est indépendante de toutes les autres conditionnellement à l'appartenance de cet individu à une classe. Un tel modèle peut être utilisé pour construire un réseau trophique particulier où l'on peut distinguer deux groupes d'espèces distinctes dont on souhaite connaître les interactions, sous forme de réseau trophique bipartide. On peut par exemple distinguer n_0 espèces

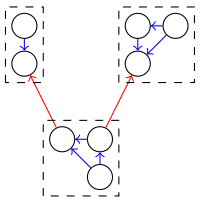


FIGURE 1.5 – Graphe organisé sous forme de modèle stochastique à bloc. Chaque bloc est représenté par un rectangle. La probabilité de présence d'un arc entre deux nœuds est dépendante des blocs auxquels appartiennent ces nœuds.

de plantes et n_1 espèces animales. Les observations des paires plante-animal sont alors une variable aléatoire binaire de co-occurence X_{ij} telle que $X_{ij} = 1$ si l'espèce animale j consomme la plante i, $X_{ij} = 0$ sinon. Construire un tel réseau trophique, sous forme de réseau bipartide par modèle à bloc latent [Donnet et al.,] consiste à affecter de manière indépendante chaque plante à une classe Z_0 et chaque animal à une classe Z_1 selon une distribution de probabilité donnée. Une fois chaque paire plante-animal affectée à une classe, sa valeur de co-occurence est modélisée conditionnellement à sa classe par une loi donnée.

1.2.5 Modèle stochastique à blocs

Les modèles stochastiques à blocs [Holland et al., 1983] fournissent un cadre pour modéliser des graphes dont les nœuds sont regroupés en communautés, et dont le comportement dépend de ces communautés. Un modèle stochastique à blocs contient n nœuds et r communautés. Chaque nœud appartient à une et une seule communauté. Ainsi, on peut définir un vecteur C de taille n contenant des valeurs entières comprises dans $[1, r]$ renseignant la communauté à laquelle appartient chaque nœud. Ainsi, C_i indique la communauté du nœud i. Un modèle à bloc stochastique contient également une matrice W de taille $r \times r$ contenant les probabilités de connectivité entre chaque communauté. Ainsi, $W_l^k \in [0, 1]$ désigne la probabilité qu'un nœud issu de la communauté k ait un arc vers un nœud issu de la communauté l. La probabilité qu'un arc E_j^i d'un nœud i vers un nœud j existe est définie par l'équation 1.1.

$$P(E_j^i) = W_{C_j}^{C_i} \tag{1.1}$$

La figure 1.5 représente un graphe organisé en communautés représentées par des rectangles pointillés. Chaque arc a une probabilité de présence dépendant de ces communautés. Un exemple de modèle stochastique à bloc pour ce graphe serait de définir une probabilité de présence d'une arête entre i et j égale à P_1 si i et j sont issus de la même communauté (arcs en bleu) et à P_2 si ils sont issus de communautés différentes (arcs en rouge).

Ce modèle est souvent utilisé pour détecter des communautés au sein d'un graphe [Abbe, 2017]. Il permet également de générer un réseau trophique lorsqu'on a connaissance de groupes fonctionnels d'espèces et de leur comportement trophique [Allesina and Pascual, 2009]. De plus, ce modèle permet d'intégrer d'autres informations, comme des interactions non trophiques ou des groupes d'espèces au comportement en partie connu.

1.2.6 Modèles de réseaux trophiques et apprentissage

Nous avons vu dans cette section plusieurs manières de modéliser un réseau trophique. Lorsque l'on n'a aucune indication sur les espèces dont on veut établir le réseau d'interaction, il semble que les modèles de niche et de hiérarchie imbriquée ressemblent à un comportement réaliste. Lorsque l'on a des informations sur les espèces, et que l'on peut les regrouper en communautés ou en groupes fonctionnels, un modèle de type bloc stochastique semble être efficace. Ce dernier n'est d'ailleurs pas incompatible avec les autres modèles de réseau trophique. Les modèles que nous avons vus ici décrivent des manières de concevoir des réseaux trophiques fictifs. Cependant, nous cherchons à apprendre des réseaux trophiques réels, à partir d'observations sur le terrain. Les réseaux générés par ces méthodes constituent une base de réseaux "cohérents", ainsi qu'une connaissance générale sur la forme de n'importe quel réseau trophique possible. Ces connaissances pourront être exploitées de diverses manières dans les méthodes d'apprentissage de réseau que nous allons voir dans la suite de ce chapitre. Il est par exemple possible d'utiliser les connaissances afin de vérifier la validité d'un réseau appris, ou même en intégrant la modélisation d'un tel réseau dans le processus d'apprentissage comme un apriori. Même lorsque le réseau que l'on cherche à apprendre est un réseau écologique, contenant également des interactions autres que trophiques, ces structures théoriques gardent leur intérêt, les interactions trophiques étant importantes parmi l'ensemble des relations biologiques entre espèces.

1.3 Apprentissage de réseau écologique par observation directe des interactions

Il existe plusieurs méthodes permettant de détecter directement des interactions écologiques entre des espèces. Basiquement, cela peut se faire par observation directe des espèces dans la nature ou dans des environnements contrôlés. Les technologies modernes facilitent ces observations, par l'utilisation de matériel optique de plus en plus performant comme des drones, des stations vidéos sous-marines ou des caméras cachées à haute définition [Linchant et al., 2015]. Une autre approche est l'observation de contenus d'estomacs ou de fèces de manière macroscopique [Hyslop, 1980] ou microscopique [King et al., 2008] afin d'y trouver des éléments et des traces d'ADN d'espèces proies, ou des **marqueurs trophiques**.

Un **marqueur trophique** est un élément propre à une espèce vivante et qui ne peut se transmettre à une autre espèce que par sa consommation. C'est à dire que la présence d'un marqueur trophique propre à une espèce i retrouvé chez une espèce j indique que j est prédateur de i, car ce traceur n'a pu être acquis que par ingestion. Il existe des parasites servant de marqueur trophique [Marcogliese, 2004]. Par exemple, certains parasites ont pour stratégie la contamination d'un hôte primaire intermédiaire facile d'accès (crustacé, insecte...) qui sert de proie pour son hôte final (poisson, oiseau...). Une espèce identifiée comme hôte primaire d'un parasite peut alors être considérée comme proie de toute espèce dans lesquelles ce parasite a été trouvé. D'autres types de marqueurs trophiques sont les **traceurs de biomasse** [Gannes et al., 1997]. Un **traceur de biomasse** est un élément chimique propre à un organisme. Ces traceurs de biomasse peuvent prendre la forme d'isotopes stables de carbone ou d'azote [Fry, 2006], d'acides gras [Hebert et al., 2009], d'acides aminés [McClelland and Montoya, 2002] ou de contaminants [Ramos and González-Solís, 2012] propres à une espèce. La présence d'un traceur dans une autre espèce indique la consommation de cette espèce. Au delà de cette information qualitative, les traceurs de biomasse apportent des informations

quantitatives, permettant d'informer de la contribution de chaque élément à la biomasse d'une espèce.

Cependant, ces méthodes demandent des échantillons d'organismes, ce qui peut poser des problèmes pratiques : cela peut nécessiter du matériel coûteux, du personnel qualifié, des études à long terme et une logistique conséquente. Obtenir des échantillons d'organismes peut aussi poser des problèmes techniques (possibilité de contamination, problèmes d'identification d'espèces, faux positifs...) ou éthiques (dissections, prélèvements d'espèces...). De plus, le nombre d'interactions possibles entre les espèces est souvent trop grand pour les tester toutes. Toutes ces méthodes exploitent des données bien plus complexes à obtenir que des données de présence/absence d'espèces. Elles ne semblent donc pas pouvoir répondre à nos attentes. C'est cependant ce genre de méthodes qui permettent d'obtenir des connaissances fiables sur certaines interactions. Ces connaissances acquises par des méthodes d'observation directe peuvent être utilisées comme connaissance supplémentaire pour vérifier la pertinence de réseaux trophiques simulées ou estimés par d'autres méthodes.

1.4 Apprentissage de réseau écologique par inférence

Reconstruire des réseaux écologiques complets par observation directe des interactions semble illusoire, surtout pour des écosystèmes mal connus ou difficiles à observer. Lorsque l'on veut établir un réseau écologique complet sans tester toutes les combinaisons d'espèces possibles, il est possible d'inférer les interactions à partir de données écologiques plus simples à obtenir, centrées sur les populations d'espèces plutôt que sur les interactions. Ces données peuvent être obtenues par observation directe des espèces dans leur milieu, ou par des méthodes indirectes, à l'aide de traces spécifiques laissées par les espèces ou divers signes de présence. Ces données renseignent directement ou indirectement sur l'abondance ou l'occurrence des espèces étudiées. Cependant, elles ne renseignent pas directement sur les interactions entre les espèces. Il est tout de même raisonnable de penser que ces interactions ont une influence sur l'abondance ou l'occurrence des espèces. L'idée est donc d'extraire cette influence à partir de ces données, et d'en déduire les interactions entre espèces. Ce type de raisonnement, cherchant à déduire des informations à partir d'informations partielles est désignée comme étant de l'**inférence**. Cette inférence peut être guidée par d'autres types de connaissances. L'inférence de réseau écologique est donc le fait de combiner les données d'abondance, d'occurence ou de séquences de gènes des espèces et diverses connaissances expertes, par exemple sur leur comportement trophique, afin de reconstruire totalement ou partiellement un réseau écologique.

1.4.1 Programmation logique

Principe

La programmation logique [Muggleton, 1999] consiste à établir l'ensemble des **faits** déductible à partir d'un ensemble de **faits** observés et de **règles logiques**. En programmation logique, un **fait** est une observation combinant un ou plusieurs individus et une ou plusieurs valeurs. Une **règle logique** est une série d'affirmations permettant de déduire un ou plusieurs faits à partir d'autres. Une **règle logique** prend la forme d'une série de faits permettant de déduire d'autres faits. Par exemple, une règle logique permettant de vérifier l'égalité entre deux nombres peut prendre la forme :

```
SI :
A-B=0;
```

```
ALORS
A=B
```

Ce cadre permet de déduire un ensemble des faits à partir de faits observés et de règles définies. Cependant, cette logique, dite déductive, reste limitée. En effet, cette forme de programmation logique impose que les déductions découlent toujours uniquement des règles décrites à priori. Ce processus peut être très limité lorsqu'on ne peut pas déterminer précisément les règles logiques d'un problème. L'induction est le fait d'utiliser des faits observés afin d'établir automatiquement des règles. Par exemple, il est possible de générer une règle de cette forme :

```
SI tous les A observés sont B;
ALORS tous les A sont B.
```

Cette forme de programmation logique se rapproche de l'inférence statistique. Il s'agit donc d'utiliser les faits observés, de capter des corrélations ou des relations logiques entre ces faits et d'apprendre de nouvelles règles logiques probables à partir de ces informations. Cette méthode permet donc d'établir de nouvelles règles, mais les faits restent totalement inféodés aux règles décrites à priori ou apprises par induction.

L'abduction permet d'inférer des faits probables à partir des faits observés, même s'ils ne suivent pas strictement les règles de la logique. On établi alors des règles abductives pouvant prendre cette forme :

```
Tous les A sont B;
OR X est B;
ALORS X est probablement A.
```

Apprentissage de réseaux écologiques par programmation logique

Il est possible d'apprendre une structure de réseau écologique à l'aide de méthodes de programmation logique [Bohan et al., 2011]. Dans ce cas, les faits observés peuvent être des abondances d'espèces, des comportements trophiques ou des tailles moyennes d'espèces et des règles combinant ces informations afin de déduire des faits concernant les interactions entre les espèces. Un exemple de règles pour un système contenant un site d'observation S et deux espèces X et Y est le suivant :

```
X a une abondance élevée SI
X est un predateur ET
Il y a co-occurence de X et de Y ET
X est plus grand que Y ET
Y a une abondance élevée ET
X mange Y
```

De ces règles et de ces observations peuvent être déduits les faits $XmangeY$ dont l'ensemble forme un réseau trophique. [Tamaddoni-Nezhad et al., 2013] ont utilisé cette méthode pour reconstruire des réseaux trophiques pour des espèces d'arthropodes dans des champs expérimentaux. Les relations trouvées par programmation logique ont été comparées à des interactions connues de la littérature. Les réseaux trophiques appris par méthode logique semblent convaincants, beaucoup de liens appris correspondant à des interactions déjà étudiées. Les données nécessaires aux méthodes de programmation logique peuvent être très variables. En effet, à partir du moment où les règles régissant chaque fait sont correctement décrites, elles peuvent utiliser des données continues, discrètes,

qualitatives ou quantitatives. De la même façon, il est possible de décrire des règles régissant le comportement de données dynamiques si besoin. Cependant, la simple information sur l'abondance ou la présence des espèces ne permet pas d'établir facilement des règles précises. Cette méthode demande donc un certain nombre de données complémentaires afin de faciliter la définition des règles logiques.

1.4.2 Apprentissage par régression linéaire

Principe

La régression statistique consiste à étudier le comportement d'une variable dite endogène Y par rapport à une série de p variables dites exogènes $X = X_1, ..., X_p$. Pour cela, on modélise cette variable endogène comme une fonction de l'ensemble des variables exogènes, comme dans l'équation 1.2.

$$Y = f(X) + \varepsilon \tag{1.2}$$

f est alors une fonction pouvant prendre plusieurs formes et qui sert à approcher et à généraliser Y et ε est un ensemble d'erreurs empiriques traduisant la différence entre le modèle théorique et la réalité. La régression linéaire est un exemple classique de régression qui consiste à modéliser une variable Y inconnue à partir de variables X par une fonction $f(X)$ linéaire. On estime alors que Y se comporte comme une droite dans le plan formé par l'ensemble des variables de X. La régression linéaire consiste à trouver une droite approximant le mieux le comportement de Y à l'aide d'une formule de forme similaire à l'équation 1.3.

$$Y = aX + B + \varepsilon \tag{1.3}$$

Cette droite est définie par la valeur a qui exprime sa pente et un vecteur $B = \{b_1, ..., b_p\}$ son ordonnée à l'origine pour chacun des n axes du plan.

Une méthode simple pour trouver ces valeurs est la méthode dite des moindres carrés. Elle consiste à trouver, à partir d'un vecteur $\hat{Y} = \hat{y}_1, ..., \hat{y}_n$ de n valeurs observées, la fonction $Y = aX + B$ ($Y = y_1, ..., y_n$ étant un ensemble de valeurs théoriques) minimisant la quantité $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Cependant, la méthode des moindres carrés peut avoir tendance à faire du sur-apprentissage : toutes les variables exogènes ayant la moindre importance, des variables exogènes normalement très peu liées à Y peuvent avoir une influence sur la fonction f qui fausse l'estimation de Y. Pour remédier à cela, deux approches existent :

- Réduire le nombre de variables exogènes en cherchant les variables les moins utiles pour la régression
- Pénaliser la fonction d'erreur à minimiser afin de limiter le phénomène de sur-apprentissage.

La pénalisation de la fonction d'erreur a l'avantage de ne pas nécessiter d'étape supplémentaire. Cette pénalisation se fait le plus souvent sur les paramètres B associés aux variables exogènes. La régression ridge [Hoerl and Kennard, 1970] propose par exemple de pénaliser les moindres carrés en cherchant la fonction minimisant la quantité exprimée dans l'équation 1.4, où λ est un paramètre positif à choisir.

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} b_j^2 \tag{1.4}$$

Cependant, le paramètre de pénalisation de ridge est difficile à interpréter. Un autre paramètre de pénalisation est la pénalisation lasso, très proche de ridge [Tibshirani, 1996]. Il s'agit de chercher la fonction minimisant la quantité exprimée dans l'équation 1.5

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |b_j| \tag{1.5}$$

Ce paramètre de pénalisation est plus simple à interpréter, et lorsque le paramètre λ est bien choisi, certains paramètres B peuvent être égaux à 0, ce qui permet de faire très facilement de la sélection de variable, en enlevant du modèle les variables exogènes associées aux paramètres nuls. Les deux méthodes permettant de corriger le phénomène de surapprentissage sont donc combinées par l'utilisation de ce paramètre.

Si la régression linéaire est souvent utilisée sur des données continues, cette méthode est adaptable à des données discrètes. La régression logistique [Hosmer Jr et al., 2013] est un cas particulier de régression linéaire où la variable endogène Y est une variable qualitative.

Apprentissage de réseaux écologiques par régression linéaire

Il est possible d'utiliser un tel modèle afin d'apprendre des réseaux écologiques. L'ensemble des variables aléatoires X du modèle est constitué des abondance d'un ensemble d'espèces. Pour une espèce cible i d'abondance observée $\hat{y}_i$, il est possible d'utiliser la régression linéaire afin de chercher un ensemble d'espèces régulatrices r, dont les abondances constituent un ensemble de variables exogènes X_r prédisant bien l'abondance de i. On peut alors modéliser l'abondance de i par la formule donnée dans l'équation 1.6.

$$Y_g = \sum_{r} B_r^i X_r \tag{1.6}$$

Les paramètres B_r^i correspondent alors à la force de l'interaction entre l'espèce cible i et les espèces régulatrices r. L'utilisation d'un paramètre de type lasso permet de sélectionner facilement les espèces dont l'interaction est significative. En appliquant cette régression et en sélectionnant les espèces en interaction les unes avec les autres, il est alors possible de définir un réseau écologique.

[Faisal et al., 2010] ont cherché à dresser un réseau d'interactions entre des espèces d'oiseaux à partir de données de présence/absence d'espèces d'oiseaux d'habitats différents. Ils ont pour cela utilisé la régression lasso, et ont réussi à retrouver quelques interactions significatives entre ces espèces déjà connues de la littérature.

Apprentissage de réseaux écologiques par régression linéaire inverse

Une autre approche utilisant la régression linéaire consiste à considérer les relations écologiques comme un flux de biomasse circulant entre chaque espèce : une espèce basale apporte à son prédateur herbivore un flux direct correspondant à la quantité de matière ou d'énergie que peut apporter sa consommation. Ce flux inconnu peut être quantifié par modélisation inverse à partir de données d'observation. La modélisation inverse [Penland and Matrosova, 1998] consiste à estimer les composantes de la fonction f à partir de l'estimation du modèle Y qui est connu. Dans le cadre du modèle linéaire, cela revient à chercher l'ensemble des variables explicatives X inconnues à partir d'un ensemble de variables Y observées. Le modèle linéaire inverse est un modèle utilisable pour l'apprentissage de réseaux écologiques. Le flux de matière constitue l'ensemble des informations inconnues à apprendre

à l'aide d'observations biologiques. Ces observations biologiques sont modélisées comme une fonction linéaire des flux de matière à caractériser par modélisation linéaire inverse. C'est par ce biais que [van Oevelen et al., 2010] ont cherché à reconstruire un réseau trophique d'espèces présentes dans des prairies d'herbes basses à l'aide de divers indices d'abondance de ces espèces trouvés dans la littérature. Une approche par modèle linéaire inverse consistant à chercher l'ensemble de flux de matière par vraisemblance a été relativement efficace pour dresser un réseau trophique acceptable sur ces espèces.

Cependant la régression linéaire ou à la régression linéaire inverse sont difficilement adaptables au cas des données dynamiques.

1.4.3 Modèle graphique gaussien

Principe

Une variable aléatoire Z est dite *gaussienne* ou *normale* si elle suit une loi normale $Z \sim \mathcal{N}(\mu, \sigma^2)$ d'espérance μ et de variance σ^2. La fonction de densité d'une variable gaussienne prend pour forme :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.7}$$

Un *vecteur gaussien* est un vecteur $X = \{X_1, ..., X_p\}$ tel que toute combinaison linéaire de ses composants est une variable aléatoire gaussienne. On associe un vecteur gaussien à son vecteur moyenne défini par $E(X) = \{E(X_1, ..., E(X_p)\}$, soit l'espérance de chacun de ses éléments. On y associe également sa matrice de variance-covariance définie par $Var(X) = E\left((X - E(X))^T (X - E(X))\right)$. La matrice de variance-covariance d'un vecteur gaussien est une matrice symétrique dont la diagonale comporte les variances de chaque élément de X et le reste les covariances entre chaque paire d'éléments de X.
Soit Z un vecteur gaussien et $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un graphe où $\mathcal{V} = \{1, ...p\}$ désigne l'ensemble de ses nœuds et $\mathcal{E}$ l'ensemble de ses arcs.

On dit que Z respecte la propriété de Markov locale par rapport à un graphe orienté $\mathcal{G}$ si pour tout nœud i, Z_i est indépendant de l'ensemble des nœuds Z_{Np} conditionnellement à Z_P où Np est l'ensemble des nœuds non reliés à i et P est l'ensemble nœuds reliés à i.

Dans le cas d'un graphe orienté $\vec{\mathcal{G}}$, on distinguera, pour un nœud i l'ensemble de ses parents Z_P, c'est à dire l'ensemble de tout nœud j tel qu'il existe un arc de j vers i, de l'ensemble de ses descendants Z_D, c'est à dire l'ensemble de tout nœud k tel qu'il existe un chemin orienté allant de i vers k. Un vecteur gaussien Z respecte la propriété de Markov locale orientée par rapport à $\vec{\mathcal{G}}$ si pour tout nœud i, Z_i est indépendant de l'ensemble des nœuds Z_{Nd} conditionnellement à Z_P où Nd est l'ensemble des nœuds n'étant pas des descendants de i.

Un vecteur gaussien respectant la propriété de Markov locale par rapport à $\mathcal{G}$ est un modèle graphique gaussien par rapport à $\mathcal{G}$. Un vecteur gaussien respectant la propriété de Markov locale orientée par rapport à $\vec{\mathcal{G}}$ est un modèle graphique gaussien orienté par rapport à $\vec{\mathcal{G}}$. Un vecteur gaussien peut être un modèle graphique gaussien de plusieurs graphes. Par exemple, un vecteur gaussien de taille p est toujours un modèle graphique gaussien d'un graphe complet contenant p nœuds. Il existe cependant un graphe minimal $\mathcal{G}_{min}$ minimisant le nombre d'arcs tel que Z est un modèle graphique gaussien de $\mathcal{G}_{min}$ [Verzelen, 2008, Lauritzen, 1996]. Ce graphe minimal est déductible de la matrice de variance-covariance : il découle des 0 de la matrice inverse de $Var(X)$. L'inférence de réseaux dans les modèles graphiques gaussiens correspond à la recherche de graphe

minimal, ce qui peut se faire par l'estimation de la matrice de variance-covariance du vecteur gaussien associé.

Une première méthode consiste à estimer la matrice de covariance à partir de la matrice de corrélation empirique. Cette matrice est complétée par le calcul des coefficients de corrélation ρ_{ij} partiels décrivant les corrélations entre deux nœuds i et j conditionnellement à tous les autres nœuds. Ce coefficient de corrélation est calculable à partir de la matrice des covariances inverse. Un coefficient de corrélation partiel faible traduit une corrélation indirecte que l'on peut écarter de la matrice de corrélation pour établir le graphe minimum du modèle graphique gaussien. Pour estimer de manière plus précautionneuse la matrice de variance-covariance à partir d'une matrice de corrélation empirique, une bonne méthode est d'utiliser les méthodes de minimisation d'erreur utilisées en régression. Une régression de type lasso permet notamment d'estimer une matrice de variance-covariance en écartant efficacement des corrélations faibles par sélection de variable [Friedman et al., 2008].

L'apprentissage par contraction [Schäfer and Strimmer, 2005] consiste à proposer deux modèles d'estimation de la matrice de variance-covariance : un modèle complet cherchant à estimer une matrice de variance-covariance avec beaucoup de paramètres U^+ (correspondant à un graphe complet par exemple), et un modèle contraint cherchant à estimer une matrice de variance-covariance avec peu de paramètres U^- (correspondant à un graphe vide par exemple). La contraction consiste à chercher un modèle U^* correspondant au meilleur compromis entre ces deux modèles à l'aide d'un paramètre $\lambda \in [0, 1]$ d'intensité de contraction estimable par minimisation de l'erreur quadratique moyenne à l'aide de la formule décrite dans l'équation 1.8.

$$U^* = \lambda U^- + (1 - \lambda)U^+ \tag{1.8}$$

Apprentissage de réseaux écologiques par modèle graphique gaussien

Diverses données biologiques (abondance d'espèces, marqueurs trophiques...) peuvent être considérées comme une approximation de vecteur gaussien, le graphe minimal associé peut alors être considéré comme un réseau écologique. Plusieurs méthodes permettent l'inférence des interactions écologiques par modèles graphiques gaussiens. [Faisal et al., 2010] estiment la matrice de variance-covariance d'un modèle gaussien issu de données d'abondances d'espèces à partir de la matrice de covariance empirique. Les valeurs de covariance significativement faibles correspondent alors à des 0 dans cette matrice. Les covariances significativement supérieures à 0 permettent de reconstruire le réseau écologique correspondant. [Kurtz et al., 2015] utilisent des données d'abondances microbiennes afin d'inférer leurs interactions par modèle graphique gaussien. Ils sélectionnent les covariances empiriques suffisamment importantes pour être considérées comme des interactions entre espèces à l'aide d'un estimateur lasso. Ils comparent leurs résultats avec un réseau microbien connu [McDonald et al., 2018]. Ils parviennent à retrouver très convenablement des liens connus, en évitant même d'apprendre des liens indirects.

Cependant, l'apprentissage par modèle graphique gaussien s'applique sur des données prenant la forme de variables aléatoires gaussiennes. Si cette contrainte convient à des données d'abondance d'espèces, elle n'est plus tenable lorsque l'on utilise des données binaires. De plus, ce modèle est difficilement adaptable à des données dynamiques.

1.4.4 Réseau bayésien

Principe

Un réseau bayésien [Pearl, 1985] modélise la loi jointe d'un ensemble de variables aléatoires $X_1, ..., X_n$ à l'aide d'un graphe orienté sans circuit $\mathcal{G}(\mathcal{V}, \mathcal{E})$. L'ensemble des nœuds $\mathcal{V}$ du graphe représente les variables aléatoires $X_1, .., X_n$ et les arcs $\mathcal{E}$ sont déterminés par les indépendances conditionnelles entre ces variables. Une absence d'arc entre les nœuds représentés par deux variables aléatoires X_i et X_j indique qu'elles sont indépendantes conditionnellement aux autres variables aléatoires [2]. Cette structure en graphe permet de réduire la taille de la table de probabilité conditionnelle d'une variable i sachant toutes les autres. En effet, cette table ne dépend que de l'état des parents de i dans $\mathcal{G}$. L'ensemble des éléments de toutes ces tables de probabilités conditionnelles constituent les paramètres du modèle. Le principe de l'apprentissage de structure de réseau bayésien consiste à trouver la structure de graphe la plus probable à partir d'un ensemble d'observations des variables aléatoires du modèle. Plusieurs approches dont nous parlerons dans un chapitre dédié permettent cet apprentissage.

Apprentissage de réseaux écologiques par réseau bayésien

L'apprentissage de réseau écologique par réseau bayésien consiste à chercher, à partir de données biologiques sur les espèces, la structure de graphe la plus probable. Ces données permettent en effet de calculer ou d'estimer des relations entre les espèces, représentables par un réseau bayésien. La structure de réseau bayésien la plus probable sachant les données écologiques peut être considérée comme un réseau écologique, sous l'hypothèse que les relations causales entre les différentes espèces soient en grande partie expliquées par leurs interactions écologiques. Cette méthode a été utilisée par [Faisal et al., 2010], sur les même données que celles utilisées pour la méthode d'apprentissage par régression lasso, afin de chercher des interactions entre diverses espèces d'oiseaux. Ces deux méthodes ont des performances à peu près similaires pour reconstruire des interactions connues de la littérature.

Le paradigme des réseaux bayésiens convient très bien à des données binaires, les tables de probabilités conditionnelles entre des variables discrètes étant simples à manipuler. De plus, il existe une forme de réseau bayésien dit "dynamique", adaptée à des données évoluant de façon dynamique. Nous en parlerons dans le prochain chapitre. Ces méthodes d'apprentissage de réseaux semblent donc bien adaptées à notre problème.

1.5 Bilan

Pour rappel, notre objectif est d'apprendre la structure d'un réseau écologique en utilisant des données simples et peu coûteuses à obtenir. Nous souhaitons donc utiliser des données de présence/absence d'espèces au sein d'un écosystème sur plusieurs périodes de temps. Cela nous permet de nous baser sur la dynamique des espèces pour faciliter l'apprentissage du réseau. Cela exclut donc toute méthode de construction de réseau écologique par observation directe d'interactions écologiques. Il existe cependant plusieurs méthodes permettant d'inférer la structure de réseaux écologiques à partir de données indirectes traduisant souvent l'abondance ou la présence des espèces. Ces méthodes se basent principalement sur des modèles mathématiques permettant de prédire les

2. Nous reviendrons sur ces notions plus en détail dans le chapitre suivant

Méthode	Données nécessaires	Adaptable au cas dynamique
Programmation logique	Données discrètes ou continues	Oui, en définissant les bonnes règles
Régression linéaire	Données discrètes ou continues	Non
Modèle graphique gaussien	Données continues uniquement	Non
Réseaux bayésiens	Données discrètes	Oui : réseau bayésien dynamique

TABLE 1.1 – Comparaison de différentes méthodes d'inférence de réseaux écologiques selon leur compatibilité à des données binaires (discrètes) et dynamiques.

interactions à partir de ces données indirectes. Pour notre problème, l'idéal serait donc une méthode permettant d'exploiter des données discrètes et dynamiques. Le tableau 1.1 donne la compatibilité des méthodes vues précédemment à ces deux critères.

La programmation logique comme l'inférence de réseaux bayésiens sont des méthodes facilement adaptables à notre problème. Cependant, avec uniquement des données de présence/absence, la programmation logique est compliquée à mettre en place. En effet, elle demande d'établir des règles précises. Ces règles sont plus simples à établir et plus précises lorsqu'elles s'appuient également sur des données supplémentaires (le groupe fonctionnel des espèces, leur taille, leur régime alimentaire le plus courant...). Pour utiliser uniquement des données de présence/absence dynamique, l'inférence de réseaux bayésiens semble donc être la méthode la plus indiquée.

Un autre critère dont nous n'avons pas encore parlé avait été énoncé au début de ce chapitre. Il s'agit de la possibilité d'intégrer des connaissances expertes dans le modèle. En effet, nous pouvons avoir un certain nombre d'informations sur la structure générale d'un réseau écologique, et ces connaissances peuvent être encore plus importantes si les espèces dont on cherche à apprendre les interactions sont bien connues. Puisque les méthodes d'inférence de réseau écologique consistent toutes à chercher des interactions sous forme d'arcs d'un graphe, il est facile d'intégrer des règles ou des tendances sur ces interactions dans n'importe quelle méthode d'interaction.

Dans le cadre de la programmation logique, cela se fait sous forme de règles déjà connues. Dans le cadre de la régression linéaire, il est possible d'ajouter des contraintes aux paramètres que l'on souhaite apprendre, ou d'intégrer des connaissances sur les interactions dans une phase de sélection de variable. Dans le cadre des modèles graphiques gaussiens, il est possible d'ajouter des connaissances ou des priors sur la matrice de variance/covariance à estimer. Dans le cadre des réseaux bayésiens, les connaissances expertes peuvent s'ajouter sous forme de contraintes sur les tables de probabilités conditionnelles ou sur la structure du graphe.

Les différentes manières de modéliser un réseau écologique convainquant peuvent constituer cette base de connaissances expertes et être utilisées afin de guider l'apprentissage. En effet, il est possible de contraindre les réseaux appris afin qu'ils ressemblent à un réseau généré par l'un des modèles de réseaux trophiques décrits en début de chapitre. L'une des approches possibles est donc de tenir compte de l'un ou de plusieurs de ces modèles afin d'établir des contraintes ou un prior sur les arcs du graphe que l'on cherche à inférer, et à l'intégrer dans le processus d'apprentissage.

Les méthodes basées sur l'apprentissage de réseaux bayésiens semblent les mieux adaptées à notre problème. Il existe plusieurs méthodes différentes pour apprendre la structure de cette famille de modèles. Le chapitre suivant détaille donc le concept de réseau bayésien, et passe en revue des

méthodes courantes d'apprentissage de structure de réseaux bayésiens.

Chapitre 2

Apprentissage de la structure d'un réseau bayésien

Ce chapitre traite des méthodes d'apprentissage de structure de réseaux bayésiens. Si nous avons déjà parlé de la notion de réseau bayésien dans le chapitre 1, cela s'est fait de manière très succincte, et avec du vocabulaire spécifique pouvant être peu explicite. Nous allons ici expliquer plus en détail le concept de réseau bayésien. Pour cela, nous allons commencer par introduire des notions de probabilité, utiles pour définir explicitement les réseaux bayésiens.

2.1 Notions de probabilité

2.1.1 Variable aléatoire

Une **variable aléatoire** X est une fonction qui associe à une observation i une valeur parmi un ensemble possible d'événements aléatoires (domaine) selon une distribution de probabilité P. La réalisation x d'une variable aléatoire X désigne une observation de cette variable aléatoire.

Exemple. *Lancer un dé à 6 faces bien équilibré revient à observer la réalisation x d'une variable aléatoire X. Le domaine de cette variable est $\{1, 2, 3, 4, 5, 6\}$ (ensemble des valeurs possibles), et sa distribution est décrite dans le tableau suivant :*

x	*1*	*2*	*3*	*4*	*5*	*6*
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

2.1.2 Probabilités jointes et marginales

Si on considère deux variables aléatoires X et Y, s'intéresser à la **probabilité jointe** de ces deux variables $P(X \cap Y)$ consiste à considérer la probabilité des états simultanés de ces deux variables. Dans un contexte multivarié, les distributions de probabilité $P(X)$ et $P(Y)$ sont désignées comme les distributions **marginales**. Il est possible de calculer les distributions marginales à partir de la distribution jointe :

$$P(X = x) = \sum_{y \in D_Y} P(X = x \cap Y = y)$$

Il est à noter que, si on peut calculer les probabilités marginales à partir des probabilités jointes, on ne peut pas calculer la distribution de probabilité jointe à partir des distributions marginales.

2.1.3 Probabilités conditionnelles

Soient deux variables aléatoires X et Y. La probabilité **conditionnelle** de X sachant Y, notée $P(X|Y)$ décrit la distribution de la probabilité de X lorsque l'état de Y est connu. Il se calcule de la manière suivante :

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

Pour des variables aléatoires X et Y, le théorème de Bayes dresse une formule permettant de calculer une probabilité conditionnelle inverse :

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)}$$

Par extension, il est possible passer des probabilités conditionnelles aux probabilités jointes :

$$P(X \cap Y) = P(X|Y) \cdot P(Y) = P(Y|X) \cdot P(X)$$

2.1.4 Indépendance

Deux variables aléatoires X et Y sont dites indépendantes si la connaissance de l'une n'influe pas sur la distribution de l'autre. Cette indépendance se note $X \perp Y$ et se vérifie par la condition suivante :

$$X \perp Y \Leftrightarrow P(X|Y) = P(X)$$

2.1.5 Probabilité jointe multivariée

Lorsqu'il y a plus de 2 variables aléatoires, on peut étendre la définition de la probabilité conditionnelle. Pour 3 variables X, Y et Z, $P(X|Y,Z)$ désigne la distribution de probabilité de X lorsque les valeurs de Y et de Z sont connues. La formule exprimant les probabilités conditionnelles permet, à l'aide du théorème de Bayes, de calculer une probabilité jointe à partir des probabilités conditionnelles et marginales. Cette formule peut s'étendre pour calculer la probabilité jointe sur 3 variables aléatoires :

$$P(X \cap Y \cap Z) = P(X|Y,Z) \cdot P(Y \cap Z) = P(X|Y,Z) \cdot P(Y|Z) \cdot P(Z)$$

Il est d'ailleurs possible d'étendre cette formule pour un ensemble de n variables aléatoires par la **règle de probabilité en chaîne** (*chain rule*) :

$$P(X_1 \cap X_2 \cap ... \cap X_n) = \prod_{i=1}^{n} P(X_i|X_1, ..., X_{i-1})$$

2.1.6 Indépendance conditionnelle

Deux variables X et Y sont dites indépendantes conditionnellement à une troisième variable Z si $X \perp Y$ lorsque la valeur de Z est connue. L'indépendance conditionnelle se note $X \perp Y|Z$ et se vérifie par la condition suivante :

$$X \perp Y|Z \Rightarrow P(X|Y,Z) = P(X|Z)$$

La connaissance d'indépendances conditionnelles permet de simplifier la formule de la règle de probabilité en chaîne. Notons $\mathbb{X}\backslash\mathbb{Y}$ avec $\mathbb{Y} \in \mathbb{X}$ l'ensemble des éléments de $\mathbb{X}$ ne faisant pas partie de $\mathbb{Y}$. En considérant un ensemble $\mathbb{X}$ de n variables aléatoires $\mathbb{X} = \{X_1, X_2, ..., X_n\}$, l'ensemble d'indépendances conditionnelles $C(X_i) \subset \mathbb{X}$ associé à X_i est défini par $X_i \perp \{\mathbb{X}\backslash C(X_i)\}|C(X_i)\}$, c'est à dire que X_i est indépendant des autres variables aléatoires sachant l'ensemble $C(X_i)$. Dans ce cas, la règle de probabilité en chaîne peut s'écrire :

$$P(X_1 \cap X_2 \cap ... \cap X_n) = \prod_{i=1}^{n} P(X_i|C(X_i))$$

2.2 Modèle graphique probabiliste

Construire un modèle graphique probabiliste [Murphy, 2001] consiste à représenter les indépendances conditionnelles entre des variables aléatoires à l'aide d'un graphe. Un nœud du graphe représente une variable aléatoire, et un arc entre deux nœuds représente une dépendance conditionnelle entre ces deux variables. Il existe différents types de modèles graphiques probabilistes, selon le type de graphe représentant les indépendances conditionnelles.

— Les **Champs de Markov Aléatoires** [Rue and Held, 2005] sont des modèles graphiques non-orientés.

— Les **Réseaux Bayesiens** (parfois désigné par RB, *Bayesian Network* ou BN) [Pearl, 1985] sont des modèles graphiques orientés.

Nous ne nous intéresserons ici qu'aux réseaux bayésiens.

2.2.1 Réseau bayésien

Un réseau bayésien $\mathcal{B}$ représente un ensemble $\mathbb{X}$ de n variables aléatoires ($\mathbb{X} = \{X_1, X_2, ..., X_n\}$) sous la forme d'un graphe $\mathcal{G}(\mathcal{V}, \mathcal{E})$ orienté sans circuit ou DAG (*Directed Acyclic Graph*) où les nœuds $\mathcal{V}$ décrivent les variables aléatoires de $\mathbb{X}$ et les arcs $\mathcal{E}$ représentent des liens entre les variables aléatoires. La structure globale de ce graphe représente les indépendances conditionnelles entre les variables aléatoires.À chaque nœud est associé une table de probabilité conditionnelle $P(X_i|\Pi_i))$. La probabilité jointe de toutes les variables aléatoires d'un réseau bayésien peut donc être formulée d'après la règle de probabilité en chaîne :

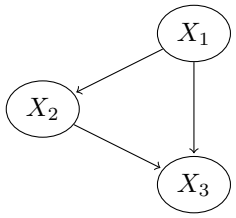
$$P(X_1 \cap X_2 \cap ... \cap X_n) = \prod_{i=1}^{n} P(X_i|\Pi_i)$$

Chaque élément d'une table de probabilité conditionnelle d'un nœud constitue un **paramètre** du réseau bayésien. La dimension $dim(\mathcal{B})$ d'un réseau bayésien $\mathcal{B}$ est son nombre de paramètres au

total, soit la taille d'un ensemble θ regroupant tous les paramètres de $\mathscr{B}$. Cette dimension augmente donc avec le nombre d'arcs qui composent le graphe $\mathscr{G}$ associé et avec le nombre d'états possibles de chaque variable aléatoire X_i. La figure 2.1 représente la structure $\mathscr{G}$ d'un réseau bayésien $\mathscr{B}$ à 3 variables, et des tables de probabilités conditionnelles à valeurs fictives pour ces variables. Le nombre d'éléments dans ces tables correspond à la structure $\mathscr{G}$. La dimension de ce réseau est $dim(\mathscr{B}) = 7$, car il y a 14 éléments dans les tables de probabilités conditionnelles, la moitié pouvant être déduite de l'autre moitié ($P(X = 1) = 1 - P(X = 0)$).

$P(X_1)$	
$P(X_1 = 0)$	$P(X_1 = 1)$
0.5	0.5

$P(X_2 \mid X_1)$	$P(X_2 = 0)$	$P(X_2 = 1)$
$X_1 = 0$	0.7	0.3
$X_1 = 1$	0.1	0.9



$P(X_3 \mid X_1, X_2)$		$P(X_3 = 0)$	$P(X_3 = 1)$
$X_1 = 0$	$X_2 = 0$	0.6	0.4
$X_1 = 0$	$X_2 = 1$	0.2	0.8
$X_1 = 1$	$X_2 = 0$	0.9	0.1
$X_1 = 1$	$X_2 = 1$	0.3	0.7

FIGURE 2.1 – Représentation d'un réseau bayésien $\mathscr{B}$ avec 3 variables et des tables de probabilités conditionnelles à valeurs illustratives.

Les réseaux bayésiens sont utilisés dans de nombreux domaines comme la bioinformatique [Yu et al., 2004] ,les sciences sociales [Liben-Nowell and Kleinberg, 2007], la médecine [Su et al., 2013] ou l'écologie [Milns et al., 2010] et de nombreuses extensions ont été développées pour répondre à des problèmes plus particuliers. Une extension qui nous intéressera particulièrement est celle des réseaux bayésiens dynamiques, qui permet de modéliser des variables aléatoires qui évoluent de façon dynamique (par exemple, dans le temps). Avant de parler des réseaux bayésiens dynamiques, voyons un exemple de modèle graphique dynamique plus simple qui nous permettra d'introduire des éléments de vocabulaire importants.

2.2.2 Chaîne de Markov

Une **Chaîne de Markov** est une suite de T variables aléatoires $\mathbb{X} = \{X^0, X^1, X^2, ..., X^T\}$ qui représentent la dynamique d'un événement. X^t représente l'état de l'événement à un instant $t \in 0, 1, ..., T$. Une chaîne de Markov a pour propriété d'être un processus stationnaire [1] vérifiant la propriété de Markov.

1. La stationnarité n'est pas forcément vraie, il existe des modèles de chaînes de Markov non stationnaires [Mallak, 1996]

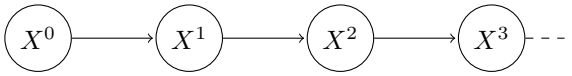


FIGURE 2.2 – Structure d'un graphe associé à une chaîne de Markov

Un processus vérifie la **propriété de Markov** lorsque sa probabilité à un instant t ne dépend que de son état à l'instant $t-1$, autrement dit, la variable aléatoire X^t est indépendante de tous les autres états antérieurs du processus conditionnellement à X^{t-1}, c'est à dire $\forall t \in \{1,,...,T\},\ P(X^t|X^0,...,X^{t-1}) = P(X^t|X^{t-1})$. Un **processus de Markov** est un processus vérifiant la propriété de Markov.
Un **processus stationnaire** est un processus dont la loi de transition est indépendante des instants. Les probabilités de transition de chaque variable aléatoire du processus sont les mêmes quel que soit l'instant t.

Ainsi, dans une chaîne de Markov décrivant la suite de variables aléatoires X, à chaque instant t du processus, on a :

$$\forall t \in \{1,...,T\}, P(X^t|X^0,...,X^{t-1}) = P(X^t|X^{t-1}) = P(X^1|X^0)$$

La probabilité $P(X^t|X^{t-1})$ est appelée **probabilité de transition**. Une chaîne de Markov peut être vue comme un réseau bayésien $\mathscr{B}$ dont la structure $\mathscr{G}$ est similaire à celle représentée en figure 2.2.

2.2.3 Réseau bayésien dynamique

Un réseau bayésien dynamique [Dean and Kanazawa, 1989] décrit le comportement d'un ensemble de variables qui évolue de façon dynamique. C'est une extension aux chaînes de Markov qui permet de prendre en compte plusieurs phénomènes dynamiques en interaction. Classiquement, un réseau bayésien dynamique décrit les états successifs d'un réseau bayésien dont les variables évoluent dynamiquement, de façon **stationnaire** en vérifiant la **propriété de Markov**. Il est possible de considérer des réseaux bayésiens dynamiques ne respectant pas la stationnarité [Gonzales et al., 2015], mais nous n'allons pas étudier ce cas dans cette thèse.
Un réseau bayésien dynamique permet donc de représenter un ensemble $\mathbb{X}$ évoluant sur T pas de temps, que l'on peut diviser en T sous ensembles $\mathbb{X}^1,...,\mathbb{X}^T$ de n variables aléatoires chacun. Pour un pas de temps t donné $\mathbb{X}^t$ est un ensemble de n variables aléatoires $\mathbb{X}^t = \{X_1^t,...,X_n^t\}$, où X_i^t représente l'état d'un phénomène i à l'instant t. On divise généralement un réseau bayésien dynamique en deux composantes $\mathscr{B}_0$ et $\mathscr{B}_\rightarrow$. La composante $\mathscr{B}_0$ décrit la structure $\mathscr{G}_0$ et les tables de probabilités jointes des variables aléatoires à l'état initial constituant un ensemble de paramètres θ_0. La composante $\mathscr{B}_\rightarrow$ décrit la structure dynamique $\mathscr{G}_\rightarrow$ du réseau et les probabilités de transition constituant un ensemble de paramètres $\theta_\rightarrow$ qui se répète à chaque transition. L'état des variables $\mathbb{X}^t$ de $\mathscr{B}_\rightarrow$ ne dépend que de l'état des variables $\mathbb{X}^t$ et $\mathbb{X}^{t-1}$. La figure 2.3 donne une représentation de la structure d'un réseau bayésien dynamique avec 3 groupes de variables dynamiques. Ce réseau bayésien dynamique a une structure sans arcs synchrones (d'un nœud i à un instant t à un nœud j à t). Dans ce cas particulier de réseau bayésien dynamique, la structure $\mathscr{G}_\rightarrow$ est la même pour toute transition entre t et $t+1$, une telle structure peut être vue comme une version dynamique d'un graphe "statique". Par exemple, si dans un graphe "statique", il existe un arc d'un nœud V_i vers un nœud V_j, la même structure "dynamique" décrira un arc de tout nœud V_i^t vers tout nœud

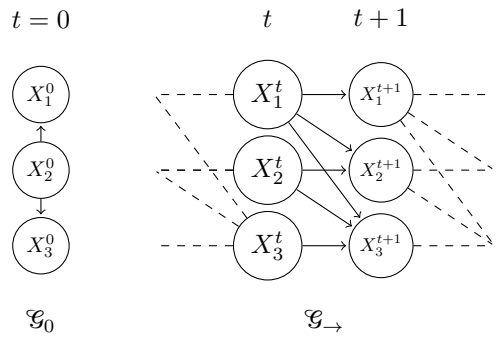


FIGURE 2.3 – Représentation de la structure d'un réseau bayésien dynamique composé d'un réseau initial $\mathcal{B}_0$ et d'un réseau de transition $\mathcal{B}_\rightarrow$. La structure initiale $\mathcal{G}_0$ à gauche est différente de la structure de transition $\mathcal{G}_\rightarrow$ à droite. Dans cet exemple, $\mathcal{G}_\rightarrow$ est similaire à la structure du réseau bayésien représenté en figure 2.1, mais en version dynamique.

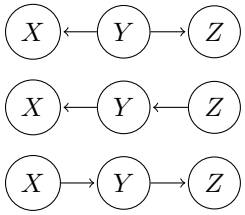


FIGURE 2.4 – Exemple de trois structures représentant la même indépendance conditionnelle $X_1 \perp X_2|Y$

V_j^{t+1} pour tout pas de temps t. Ainsi, la structure $\mathcal{G}_\rightarrow$ de la figure 2.3 est une version dynamique du graphe représenté en figure 2.1, à laquelle on a ajouté des arcs entre chaque variable X_i^t et X_i^{t+1}. Ces arcs ajoutés servent à représenter les dépendances de la mesure d'un phénomène à la mesure du même phénomène au pas de temps précédent, à la manière d'une chaîne de Markov. Cette "auto-dépendance" est très courante en pratique lorsque l'on modélise un réseau bayésien dynamique.

2.2.4 Équivalence au sens de Markov

L'ensemble des indépendances conditionnelles entre un ensemble de variables peut être représenté par un graphe. Cependant, plusieurs graphes peuvent correspondre au même ensemble d'indépendances conditionnelles, et donc, à la même loi jointe. Deux graphes sont dits équivalents au sens de Markov s'ils représentent la même loi jointe entre un ensemble de variables [Flesch and Lucas, 2007]. Les structures représentées en figure 2.4 sont équivalentes au sens de Markov. En effet, elles représentent la même distribution $P(X, Y, Z) = P(X)P(Z|X)P(Y|Z) = P(X|Z)P(Z)P(Y|Z) = P(X|Z)P(Z|Y)P(Y)$. Cependant, la structure $X \rightarrow Y \leftarrow Z$ n'est pas équivalente aux trois autres structures présentées dans cette figure. Ce type de structure est appelé V-structure. Deux graphes sont équivalents au sens de Markov s'ils ont le même squelette, c'est à dire la même structure

non-orientée, et les mêmes V-structures [François, 2006].

2.2.5 Apprentissage de structure de réseaux bayésiens

L'avantage de modéliser une distribution de probabilité par un réseau bayésien est qu'il existe un grand nombre de méthodes permettant de manipuler ces réseaux [Cowell et al., 2006]. L'utilisation des réseaux bayésiens se fait souvent dans le cas où l'on cherche à retrouver des informations manquantes dans le modèle. La plupart du temps, ces données manquantes sont les paramètres θ du réseau, c'est à dire les tables de probabilités conditionnelles, ou l'information sur la structure $\mathcal{G}$ du réseau. Ces techniques utilisent souvent une série de données observées. Ces données sont en fait des réalisations des variables aléatoires du modèle. Un jeu de données contenant plusieurs observations indépendantes de n phénomènes différents permet notamment d'estimer les tables de probabilités conditionnelles de chacune des variables aléatoires dont ils sont la réalisation. Par exemple, pour le problème d'apprentissage de structure d'un réseau trophique à partir d'observations de présence/absence d'espèces, ces observations constituent des réalisations de variables aléatoires. Pour modéliser un réseau trophique par un réseau bayésien, nous considérons l'observation de présence ou d'absence d'une espèce i comme une variable aléatoire X_i. $X_i = 0$ si l'espèce est absente et $X_i = 1$ si elle est présente. Le problème est alors d'apprendre la structure de ce réseau bayésien, cette structure représentant les dépendances entre chaque espèce, elle peut être interprétée comme un réseau d'interaction écologique. Concentrons nous alors sur des méthodes permettant d'apprendre la structure du graphe, soit les dépendances entre les variables aléatoires du réseau [Liu, 2015]. Cette tâche est complexe. En effet, le nombre de graphes possibles augmente de manière exponentielle en fonction du nombre de variables [Robinson, 1977] et [Chickering et al., 2004] ont montré que ce problème était NP-difficile [2]. Il existe majoritairement deux types de méthode permettant d'apprendre la structure d'un réseau bayésien à partir d'observations :

— Des méthodes cherchant les indépendances conditionnelles entre chaque paire de variables à l'aide de tests statistiques.

— Des méthodes utilisant une fonction de score décomposable afin de rechercher la structure optimale.

Passons en revue les différents outils de ces deux familles de méthodes. Nous verrons en fin de chapitre comment traiter spécifiquement le cas dynamique.

2.3 Algorithmes basés sur des tests d'indépendance conditionnelle

Un test d'indépendance entre deux variables couramment utilisé est le test du χ^2 d'indépendance. Un test du χ^2 consiste à comparer un vecteur $\hat{X}$ contenant des valeurs empiriques à un vecteur X de même taille contenant des valeurs "attendues" ou théoriques du phénomène mesuré. Ces valeurs théoriques peuvent être obtenues de différentes manières : dans le cas d'un test dit d'adéquation à une loi l donnée, le vecteur X est généré de telle sorte que $X \rightsquigarrow l$. Dans le cadre d'un test d'indépendance, le vecteur empirique $\hat{X}$ est constitué des valeurs conjointes de deux variables X_1

2. Un problème NP-Difficile peut être ramené à un problème de type NP-complet en un temps raisonnable (polynomial). Un problème NP-complet est un problème dont la solution peut être vérifiée facilement (en temps polynomial), mais dont la résolution peut être complexe. Tout problème NP-complet peut être ramené à un autre problème NP-complet en temps polynomial.

et X_2 observées et le vecteur théorique X est constitué des valeurs de la distribution conjointe entre ces deux variables si celles ci sont indépendantes. Un test de χ^2 d'indépendance conditionnelle fonctionne de la même manière, mais en comparant la table de contingence de chaque variable deux à deux conditionnellement à toutes les autres à leur distribution attendue en cas d'indépendance conditionnelle.

2.3.1 Algorithmes basés sur la recherche de V-structures

Algorithme SGS

L'algorithme SGS [Spirtes et al., 2000] permet d'apprendre une structure de réseau bayésien. Pour cela, il considère un graphe non-orienté complet qu'il s'agit d'élaguer puis d'orienter. L'élagage consiste à tester l'intégralité des indépendances conditionnelles entre les nœuds du graphe, puis à supprimer les arcs entre tous les nœuds testés comme indépendants conditionnellement aux autres. Une fois cette première phase achevée, le graphe obtenu est un graphe élagué non orienté. Pour orienter le graphe, l'algorithme repère les "V-structures". Une V-structure est un triplé de nœuds $\{i, j, k\}$ pour lesquels il existe des arcs de la forme $i \rightarrow j \leftarrow k$, c'est à dire qu'il n'existe entre les nœuds de ce triplé que 2 arcs : un arc de i vers j et un arc que k vers j. Pour repérer ces V-structures, l'idée est de chercher tous les triplés de nœuds $\{i, j, k\}$ contenant deux arcs (i, j) et (j, k) dans le graphe non-orienté [Chickering, 2002]. Dans une telle configuration, X_i et X_k sont indépendants, mais, en connaissant la valeur de X_j, ils ne le sont plus. X_i et X_k ne sont donc pas indépendants conditionnellement à X_j. Une fois les V-structures trouvées, les autres arcs peuvent être orientés. L'orientation des arcs restants peut se faire dans n'importe quel sens, mais doit respecter une règle : les orientations de ces arcs ne doivent pas générer de V-structures supplémentaires, non détectés par les tests d'indépendances conditionnelles. Plusieurs structures différentes peuvent donc être obtenues par cette méthode, mais toutes sont équivalentes au sens de Markov. Cet algorithme nécessite cependant un très grand nombre de tests d'indépendances conditionnelles.

Algorithme PC

Afin d'accélérer la phase d'élagage du graphe non orienté, l'algorithme PC [Spirtes et al., 2000] exploite la propriété suivante : si deux nœuds i et j sont indépendants conditionnellement à un ensemble de nœuds M, alors i est indépendant de j conditionnellement au voisinage de i, ainsi que conditionnellement au voisinage de j.

Une indépendance entre deux variables X_i et X_j est dite indépendance conditionnelle d'ordre 0. Si deux variables X_i et X_j sont indépendantes conditionnellement à une seule variable X_k, il s'agit d'une indépendance conditionnelle d'ordre 1. Une indépendance conditionnelle d'ordre 2 est donc une indépendance entre deux variables conditionnellement à deux autres variables, et ainsi de suite. L'algorithme PC fonctionne comme l'algorithme SGS, mais sa phase d'élagage change : elle consiste à tester les indépendances d'ordre 0 entre chaque paire de variables afin de supprimer une partie des arcs. L'élagage continue en testant les indépendances conditionnelles d'ordre o entre chaque paire de variables conditionnellement à chaque ensemble de variables de taille o, en incrémentant o à chaque étape. Lorsque deux variables X_i et X_j ont déjà été testées comme indépendantes conditionnellement à un ensemble de taille o, les tests d'indépendance entre X_i et X_j conditionnellement à un ensemble de taille $o + 1$ sont inutiles.

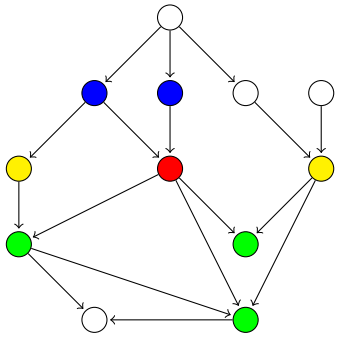


FIGURE 2.5 – Illustration de la couverture de markov. Dans ce graphe, la couverture de markov du nœud rouge est composée de ses nœuds parents (en bleu), enfants (en vert) et parents de ses enfants (en jaune).

2.3.2 Algorithmes basés sur la couverture de Markov

Principe

La couverture de Markov d'un nœud i est l'ensemble des nœuds suivants :

— Les nœuds *parents* de i, c'est à dire l'ensemble des nœuds k pour lesquels il existe un arc de k vers i

— Les nœuds *enfants* de i, c'est à dire l'ensemble des nœuds j pour lesquels il existe un arc de i vers j

— Les nœuds constituant les *parents* de l'ensemble des *enfants* de i. C'est à dire, pour chaque *enfant* j de i, l'ensemble des nœuds l tels qu'il existe un arc de l vers j.

Par exemple, dans la figure 2.5, la couverture de markov du nœud en rouge est constituée de tous les nœuds colorés.

La couverture de Markov d'un nœud permet de connaître une partie des arcs du graphe $\mathcal{G}$, et la connaissance de l'ensemble des couvertures de Markov de chaque nœud indique l'ensemble des arcs du graphe. Le principe des algorithmes basés sur la couverture de Markov est de trouver la couverture de Markov de toutes les variables [Yu et al., 2013].

Algorithme grow-shrink

Pour trouver les nœuds appartenant à la couverture de Markov MB d'un nœud i, en partant d'un ensemble $MB(i)$ vide, on peut utiliser un algorithme grow-shrink [Margaritis, 2003]. Cet algorithme applique à chaque nœud, 2 étapes permettant d'apprendre sa couverture de Markov.

Initialisation : La couverture de Markov de i est vide : $MB(i) = \emptyset$

Étape 1 - Croissance (Grow) : Ajout récursif à $MB(i)$ de tous les nœuds non indépendants de i conditionnellement à $MB(i)$.

Étape 2 - Contraction (Shrink) : Enlèvement récursif de $MB(i)$ de tous les nœuds indépendants de i conditionnellement à $MB(i)$.

La connaissance de la couverture de Markov de chaque nœud ne suffit pas à apprendre la structure complète du graphe. Afin d'apprendre la structure d'un graphe, cet algorithme doit être intégré à la procédure suivante :

1. Utiliser l'algorithme Grow-Shrink afin de trouver la couverture de Markov $MB(i)$ pour tout nœud i
2. Trouver un ensemble de voisins potentiels $N(i)$ à tout nœud i. Soit, $\forall j \in MB(i)$, l'ensemble T défini comme le plus petit ensemble entre $MB(i)\backslash\{j\}$ et $MB(j)\backslash\{i\}$, $j \in N(i)$ si $\forall S \subseteq T$, $X_i \perp X_j | S$.
3. Orienter $j \rightarrow i$ pour tout nœud i et tout nœud $j \in N(i)$ si il existe un nœud $k \in N(i)\backslash N(j)\backslash\{j\}$ tel que $X_i \not\perp X_k | S \cup \{i\}$ pour tout $S \subseteq T$ où T est le plus petit ensemble entre $MBj\backslash\{i,k\}$ et $MBk\backslash\{i,j\}$.
4. Tant qu'il existe encore des cycles dans le graphe, enlever itérativement l'arc faisant partie du plus de cycles et l'intégrer à une file R.
5. Insérer les arcs présents dans la file R en inversant leur orientation, dans l'ordre dans lequel ils ont été placé das R.
6. Pour tout nœud i et tout nœud $j \in N(i)$ sans arc orienté, orienter $i \rightarrow j$ de manière à ne pas former de cycle.

Cet algorithme permet d'apprendre efficacement la structure d'un réseau bayésien, au prix d'un grand nombre de calculs destinés à tester des indépendances conditionnelles, et des orientations d'arcs pouvant être assez arbitraires.

Afin de réduire ce nombre de calculs, l'algorithme IAMB [Tsamardinos et al., 2003] et ses variantes [Yaramakala and Margaritis, 2005] propose d'évaluer en avance si l'ensemble de nœuds trouvé lors de la phase de croissance a des chances d'être effectivement une couverture de Markov. Cette évaluation se fait par une fonction heuristique (par exemple, l'heuristique proposée par [Koller and Sahami, 1996]), et permet de réduire le nombre de tests d'indépendances conditionnelles en n'explorant pas les ensembles de nœuds étant, au sens de la fonction heuristique, mauvais candidats à être la couverture de Markov $MV(i)$ d'un nœud i.

2.4 Algorithmes basés sur un score

Il s'agit de définir une fonction de score que l'on cherche à optimiser afin de trouver les parents les plus vraisemblables de chaque variable.

2.4.1 Fonction de score

Une fonction de score $S(\mathbb{X}|\mathcal{G})$ associée à un ensemble de n variables $\mathbb{X} = \{X_1, ..., X_n\}$ d'un réseau bayésien $\mathcal{B} = (\mathcal{G}, \theta)$ est bien plus simple à manipuler si elle est *décomposable*, c'est à dire si elle peut s'écrire comme une somme de scores locaux dépendant uniquement d'une variable X_i et de l'ensemble des variables aléatoires Π_i des parents de i, comme dans l'équation 2.1.

$$S(X|\mathcal{G}) = \sum_{i=1}^{n} s(X_i|\Pi_i) \tag{2.1}$$

Un score simple et utilisé dans de nombreux modèles est la vraisemblance. La vraisemblance est la probabilité globale que les données observées $x = \{x_1, ..., x_n\}$ soient générées à partir d'un

modèle donné. On utilise plus fréquemment le logarithme de cette vraisemblance (log-vraisemblance). En effet, dans le cas des réseaux bayésiens, la transformation par logarithme permet de calculer cette probabilité par une somme de log-vraisemblance locale par variable. Il s'agit donc d'un score décomposable, plus simple à manipuler. La log-vraisemblance des données empiriques x pour une structure et un ensemble de paramètres de réseau bayésien $\mathcal{G}$ d'après la règle de probabilité en chaîne est la somme des probabilités de chaque variable conditionnellement à ses parents à la manière de l'équation 2.2.

$$logP(x|\mathcal{G},\theta) = \sum_{i=1}^{n} log(P(X_i|\Pi_i)) \tag{2.2}$$

Cependant, la log-vraisemblance possède un inconvénient majeur : la quantité $log(P(X_i|\Pi_i))$ dépend du nombre d'éléments dans la table de probabilités conditionnelles de i sachant ses parents. L'ajout d'un arc, donc d'un parent, augmente nécessairement cette probabilité. Ainsi, plus cette table compte d'éléments, plus cette quantité est grande. De ce fait, plus la structure d'un graphe est complexe et possède d'arcs, plus la vraisemblance est élevée. Un graphe complet, possédant un arc entre chaque nœud, sera donc le plus probable. C'est pourquoi, en pratique, il est souvent plus efficace de chercher une fonction plus parcimonieuse que la vraisemblance par la pénalisation de la complexité de la structure du réseau. Il existe de nombreux scores permettant cela [Naïm et al., 2011].

Les critères AIC [Akaike, 1970] et BIC [Schwarz et al., 1978] ajoutent à la vraisemblance un terme de pénalisation en fonction de la dimension du réseau bayésien $\mathcal{B}$. Le score AIC pour un réseau bayésien $\mathcal{B}$ est présenté dans l'équation 2.3 et le score BIC dans l'équation 2.4. Une structure de réseau bayésien optimale est une structure *minimisant* ces scores.

$$AIC(x|\mathcal{G},\theta) = logP(x|\mathcal{G},\theta) - dim(\mathcal{B}) \tag{2.3}$$

$$BIC(x|\mathcal{G},\theta) = logP(x|\mathcal{G},\theta) - \frac{log\ n\ \mathcal{B}}{2} \tag{2.4}$$

Un autre score très couramment utilisé est le score de Dirichlet bayésien (score BD) [Cooper and Herskovits, 1992]. Il s'agit d'attribuer une loi à priori aux structures possibles d'un réseau bayésien $\mathcal{B}$. Le score BD associé à une structure $\mathcal{G}$ est la probabilité à posteriori de cette structure sachant les observations x des variables aléatoires $\mathbb{X}$ du réseau. En choisissant comme loi à priori une distribution de Dirichlet sur les paramètres du réseau, le score BD s'exprime facilement [Chickering and Heckerman, 1996]. L'équation 2.5 montre l'expression d'un score BD d'un réseau bayésien $\mathcal{B}$ où r_i est l'ensemble des états possibles de la variable aléatoire X_i et q_i le nombre de configurations possibles de combinaisons d'états des variables aléatoires Π_i associées aux parents π_i de i. N_{ijk} représente le nombre de fois où X_i a été observée à l'état k et ses parents à la combinaison d'états j. Chaque paramètre α de la fonction gamme Γ est exprimé par la loi de Dirichlet à priori sur la structure $\mathcal{G}$.

$$CD(x|\mathcal{B}) = P(\mathcal{G}) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(N_{ij}+\alpha_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk}+\alpha_{ijk})}{\Gamma(\alpha_{ijk})} \tag{2.5}$$

Ce score dépend des distributions α sur les paramètres qui peuvent être difficiles à estimer. Le score de Dirichlet bayésien équivalent (score BDe) [Heckerman et al., 1995] donne une distribution sur ces paramètres exprimée dans l'équation où N' exprime une *taille d'échantillon équivalent*, c'est à dire un nombre d'états d'une variables et de ses parents estimé par l'utilisateur, servant de prior sur

cette distribution. $\mathcal{G}_c$ représente une structure de graphe complet, où tous les nœuds sont connectés à tous les autres.

$$\alpha_{ijk} = N' \times P\left(X_i = k, \pi_i = j \backslash \mathcal{G}_c\right) \tag{2.6}$$

D'autres distributions sont basées sur le même principe, comme le score BDeu définissant les paramètres de la loi Gamma par $\alpha_{ijk} = \frac{N'}{r_i q_i}$.

Tous ces scores pénalisés peuvent être décrits comme une somme de 2 termes : une mesure d'*entropie* désignant la capacité des données observées à expliquer le modèle (souvent mesuré par la vraisemblance ou son logarithme) et un terme de pénalisation, faisant baisser l'entropie avec la complexité du modèle.

2.4.2 Algorithme de recherche exhaustive

Un premier algorithme simple d'apprentissage de structure de réseau bayésien dynamique basé sur un score est l'algorithme de recherche exhaustive [Cooper and Herskovits, 1992]. Il s'agit de chercher, pour chaque variable, le meilleur ensemble de parents possibles par maximisation de son score local, en explorant l'intégralité des ensembles de parents possibles. La recherche exhaustive peut être décrite, pour toute variable X_i, par l'algorithme 1. En général, on utilise comme liste de parents de départ $\pi_1 = \emptyset$.

Procédure *RechExh*$(\pi_1) \rightarrow \pi_1^*, P_1$
  Entrées : π_1 : Liste de parents
  Output : π_1^* : Liste de parents appris, P_1 : score local de X_i pour π_1^*
  Calculer le score P_1 de X_i pour π_1;
  $\pi_1^* \leftarrow \pi_1$;
  si $\pi_1' = \emptyset$ **alors**
    $j \leftarrow 1$;
  sinon
    $j \leftarrow$ Dernier élément ajouté à π_1^* ;
  fin
  pour chaque *nœud* q *tel que* $j + 1 \leq q \leq n$ **faire**
    $\pi_2 \leftarrow \pi_1' \cup q$;
    $\pi_2^*, P_2 \leftarrow$RechExh(π_2);
    si $P_2 < P_1$ **alors**
      $\pi_1^* \leftarrow \pi_2^*$;
      $P_1 \leftarrow P_2$;
    fin
  fin
end

Algorithme 1 : Algorithme de recherche exhaustive

Cependant, cet algorithme demande de chercher dans tout l'espace des graphes possibles, ce qui peut prendre très longtemps lorsque le nombre de variables est élevé.

2.4.3 Algorithme Branch and Bound

L'algorithme Branch and Bound [Suzuki, 1996] améliore l'algorithme de recherche exhaustive en exploitant le critère de pénalisation de vraisemblance. En effet, un score local $S(i)$ peut se

voir comme une formule en deux parties : $S(i) = H(i) - d(i)$ où $H(i)$ est une mesure permettant d'expliquer les données par un réseau donné (par exemple, la vraisemblance), souvent désignée comme *entropie* et $d(i)$ est un paramètre de pénalisation de la complexité du réseau. C'est le cas du score de longueur de description minimal MDL (minimal description length), dont le score BIC est une approximation souvent utilisée. La pénalisation par ce type de critère a une propriété utile pour l'apprentissage de réseau bayésien. Lorsqu'un arc de j vers i est ajouté au réseau, le paramètre de complexité $d(i)$ est facilement calculable, car il ne dépend que des valeurs possibles de la variable aléatoire X_j associée au nœud j. Cette valeur peut alors constituer une borne inférieure pour $H(i)$ permettant d'éliminer la solution : $d(i)|J \in Pa(i) > H(i)$, c'est que l'ajout d'un arc de j vers i augmente davantage la pénalisation de la complexité que le pouvoir explicatif du réseau, alors on peut éliminer cet arc des solutions possibles. L'algorithme 2, applicable à un nœud i donné, prend en compte ce résultat afin de simplifier la recherche en ajoutant une *étape d'élagage* [Suzuki, 1999]. Généralement, cet algorithme prend en entrée une liste de parents de départ $\pi_1 = \emptyset$, la pénalisation p_1 étant calculable à partir de cet ensemble de parents.

Procédure $BnB(\pi_1, d_1) \rightarrow \pi_1^*, S_1$
- **Entrées :** π_1 : Liste de parents
- **Output :** π_1^* : Liste de parents appris, P_1 : score local avec Π_1
- Calculer l'entropie H_1 par rapport à π_1;
- $S_1 = H_1 + p_1$;
- $\pi_1' \leftarrow \pi_1$;
- **si** $\pi_1^* = \emptyset$ **alors**
  - $i \leftarrow 0$;
- **sinon**
  - $i \leftarrow$ Dernier élément ajouté à π_1' ;
- **fin**
- **pour chaque** *nœud* q *tel que* $i + 1 \leq q \leq n$ **faire**
  - $\pi_2 \leftarrow \pi_1 \cup q$;
  - Calculer la pénalisation d_2 par rapport à π_2;
  - `/* Étape d'élagage */`
  - **si** $d_2 < H_1$ **alors**
    - $\pi_2^*, S_2 \leftarrow$BnB(π_2, d_2);
    - **si** $S_2 < S_1$ **alors**
      - $\pi_1^* \leftarrow \pi_2^*$;
      - $S_1 \leftarrow S_2$;
    - **fin**
  - **fin**
- **fin**

end

Algorithme 2 : Algorithme Branch and Bound

Cet algorithme permet donc d'écarter des solutions qui feraient baisser le score global du réseau avant d'avoir besoin de calculer ce score, ce qui accélère la procédure. Certaines approches [Yuan and Malone, 2012, Fan et al., 2014] cherchent à affiner davantage cette phase d'élagage. Cet algorithme peut être encore plus performant lorsque l'on possède des connaissances à priori sur le réseau. L'espace des solutions à rechercher peut déjà être nettement réduit lorsque l'on connaît un ordre sur les variables aléatoires constituant le réseau bayésien. Si à toute variable i est associé un ordre v_i tel que :

si $v_i > v_j$ alors il ne peut pas exister d'arc de j vers i. Dans ce cas, lorsque l'on applique les algorithmes 1 et 2 à un nœud j, chaque nœud q que l'on cherche à ajouter à l'ensemble des parents de j peut être sélectionné tel que $v_{i+1} \leq v_q \leq v_{j-1}$. Si l'on a déjà des aprioris sur les arcs du graphe dont on cherche la structure, en connaissant le nombre de parents de certains nœuds, voire l'existence ou l'inexistante de certains arcs, il est possible d'ajouter à l'*étape d'élagage* de cet algorithme une condition prenant en compte cette connaissance [De Campos et al., 2009]. Cela permet d'éviter de calculer les scores correspondant à des structures de graphe connues comme étant impossibles.

Cependant, ces algorithmes de recherche ne garantissent pas l'acyclicité du graphe appris. Lorsque l'on ne connaît pas d'ordre sur les variables aléatoires du réseau bayésien, ces algorithmes nécessitent une étape de vérification d'acyclicité, par exemple en ajoutant des contraintes supplémentaires interdisant l'ajout de nœud induisant un cycle lors de l'*étape d'élagage* [Campos and Ji, 2011].

2.4.4 Algorithmes d'ordonnancement de variables

Principe

Puisqu'un réseau bayésien $\mathcal{B}$ se base sur un graphe orienté sans circuit $\mathcal{G}(\mathcal{V}, \mathcal{E})$, il contient au moins une *feuille*, c'est à dire un nœud qui n'est parent d'aucun nœud. Pour trouver la meilleure structure d'un réseau bayésien, une approche est de commencer par chercher ses feuilles, puis le meilleur choix de parents pour cette feuille afin d'obtenir un sous-graphe optimal. La notion d'"optimal" ou de "meilleur choix" de parent ou de feuille s'évalue à l'aide de score : on cherche le graphe maximisant un score global, somme de scores locaux associés à chaque variable. Un sous-graphe optimal associé à un nœud i est donc un sous graphe maximisant le score local associé à cette variable $max_{\pi_i \subseteq \mathcal{V}}\{Score(i|\pi_i)\}$. Un feuille optimale est un nœud maximisant le score global du réseau lorsqu'il constitue une feuille avec un ensemble de parents optimal. Pour un ensemble de nœuds $\mathcal{V}$, un score étant décomposable localement, maximiser le score global $S(\mathcal{V})$ sachant un nœud i revient à maximiser le score global du reste du réseau $S(\mathcal{V}\backslash\{i\})$. On cherche donc le score calculé par l'équation 2.7.

$$S(\mathcal{V}) = max_{i \in \mathcal{V}} \{\mathcal{V}\backslash\{i\} + max_{\pi_i \subseteq \mathcal{V}}\{S(i|\pi_i)\}\} \tag{2.7}$$

Trouver une feuille optimale i demande de trouver la structure de graphe optimale pour un sous graphe contenant les variables $\mathcal{V}\backslash\{i\}$. Cela peut de nouveau se faire par recherche de feuille optimale. Ce problème revient donc à ordonner chaque nœud et à chercher son ensemble de parents optimal dans cet ordre.

Algorithme des puits optimaux

L'algorithme des puits optimaux [Silander and Myllymaki, 2012] fonctionne sur le principe de recherche de feuille optimale et de son ensemble de parents optimal par ordonnancement. L'algorithme des puits optimaux cherche à trouver les variables les plus susceptible d'être des feuilles (aussi désignées comme *puits*, d'où le nom de l'algorithme), et à trouver récursivement les meilleurs parents des puits, de leur parents, puis des nouveaux parents jusqu'à avoir parcouru tous les nœuds du graphe. Cet algorithme fonctionne en 5 étapes :

1. Calcul des scores locaux pour chacun des $n2^{n-1}$ choix possibles d'une variable et de son ensemble de parents.
2. À l'aide des scores locaux, trouver l'ensemble de parents optimal pour chaque variable.

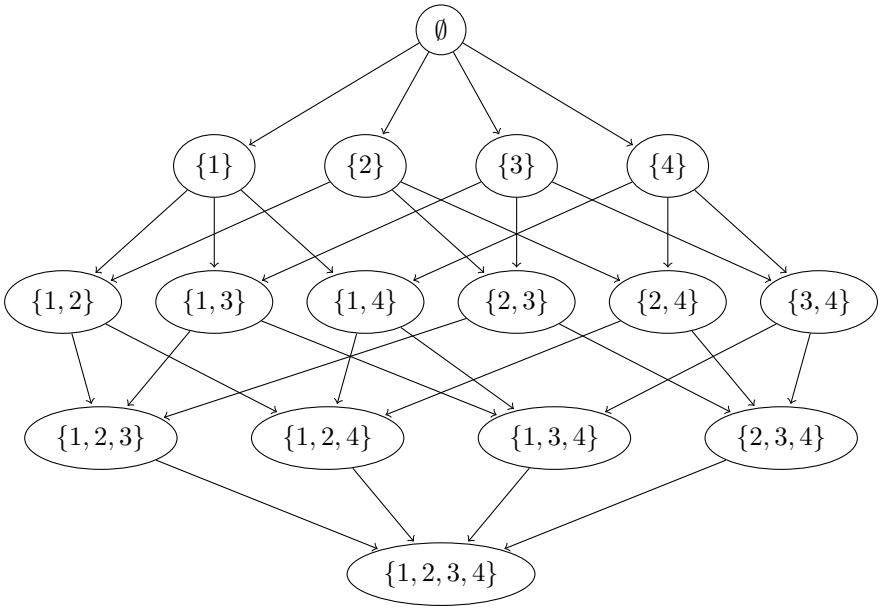


FIGURE 2.6 – Illustration d'un graphe d'ordonnancement pour un réseau bayésien à 4 nœuds.

3. Trouver la meilleure *feuille* possible $max_{\pi_i \subseteq \mathcal{V}}\{Score(i|\pi_i)\}$ sur tout le graphe.
4. À l'aide du résultat de l'étape 3, trouver le meilleur ordre possible pour les variables en évaluant récursivement les feuilles optimales sur le sous graphe contenant les variables $\mathcal{V}\backslash\{i\}$, i étant le puits optimal du graphe.
5. Trouver le meilleur réseau possible en attribuant les parents optimaux trouvés dans l'étape 2 respectant l'ordonnancement trouvé dans l'étape 4.

Cependant, cet algorithme demande de calculer un très grand nombre de scores locaux pour trouver les ensembles de parents optimaux pour chaque variable. En effet, pour n variables, il y a $n \cdot 2^{n-1}$ ensembles variable-parents. Cependant, il existe des méthodes permettant d'éviter de calculer tous ces scores.

Plus court chemin dans un graphe d'ordonnancement

Un graphe d'ordonnancement est un graphe orienté listant tous les sous ensembles de variables possibles d'un ensemble vide (constituant la racine) à un ensemble complet (constituant une feuille unique), 2 ensembles étant reliés par un arc si l'un contient l'autre. La figure 2.6 représente un graphe d'ordonnancement listant chaque ensemble possible de 4 variables. Dans ce graphe, un chemin depuis la racine vers la feuille constitue un ordonnancement de variables. Par exemple, pour un graphe d'ordonnancement à 4 variables comme dans la figure 2.6, le chemin $\emptyset, \{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}$ correspond à l'ordre $1, 2, 3, 4$.

Dans le cadre de l'apprentissage de structure de réseau bayésien par ordonnancement, un nœud de ce graphe correspond à un sous-ensemble de variables de $\mathcal{V}$ pour lequel on peut proposer un sous-graphe et en mesurer le score. Si l'on connaît un sous-ensemble $\mathcal{U} \subset \mathcal{V}$ dont la structure

optimale constitue un sous-graphe de la structure globale du réseau complet, trouver un ensemble de parents optimal pour une nouvelle variable i ordonnée de manière cohérente est alors plus simple, car cet ensemble est inclus dans ce sous ensemble : $\pi_i \subset \mathcal{U}$. En évaluant la structure d'un réseau par un score à minimiser, la meilleure structure $\mathcal{G}(\mathcal{V}), \mathcal{E}$ possible pour un réseau bayésien étant celle minimisant la quantité $\sum_{i \in \mathcal{V}} Score(i|\pi_i)$, on peut définir le *coût* $c(S_1, S_2)$ d'un arc entre deux ensembles de variables S_1 et S_2 voisines dans le graphe d'ordonnancement par l'équation 2.8. De par la construction d'un graphe d'ordonnancement, l'ensemble $S_2 \backslash S_1$ contient bien un élément i unique.

$$c(S_1, S_2) = min_{\pi_i \subseteq S_1} Score(i|\pi_i), \text{ où } i = S_2 \backslash S_1 \tag{2.8}$$

Le plus court chemin entre l'ensemble vide et l'ensemble complet dans ce graphe d'ordonnancement pondéré est par construction équivalent au graphe optimal. Il est alors possible d'utiliser n'importe quel algorithme de recherche du plus court chemin (ex : [Dijkstra, 1959, Floyd, 1962]) afin de trouver ce graphe. Un graphe d'ordonnancement a une structure particulière : chaque nœud correspond à un sous-ensemble de taille m et est relié à tout sous-ensemble de taille $m + 1$.Cette disposition hiérarchique peut être exploitée pour faciliter la recherche du plus court chemin en évitant de calculer les coûts de chaque arc. C'est ce qu'exploite l'algorithme A* [Yuan et al., 2011]. Cet algorithme fonctionne à l'aide d'une file de priorité appelée liste *ouverte*, contenant les nœuds de l'espace de recherche, initialisé avec le nœud racine i_0, correspondant à l'ensemble vide dans le graphe d'ordonnancement. À chaque itération, on calcule le *coût total* de chaque nœud i depuis la racine. Une fois ce *coût total* calculé, l'algorithme sélectionne le nœud i de la liste *ouverte* dont le *coût total* est minimal. Les nœuds suivants de i, c'est à dire tous les nœuds k pour lesquels il existe un arc de i vers k rejoignent alors la liste *ouverte* et i est ajouté à une file de nœuds appelée liste *fermée*. Si un nœud ajouté à la liste *ouverte* y est déjà présent, on ne conserve que le doublon dont le *coût total* est le plus faible en retenant le parent associé à ce coût. Si un nouveau nœud ajouté à la liste *fermée* existe déjà, ce nouveau nœud n'est pas retenu. Une fois le nœud final atteint, on peut extraire le plus court chemin à l'aide des nœuds contenus dans la liste *fermée* et de leur *coûts totaux*. Ce chemin contient alors un ordonnancement de variables permettant de trouver rapidement la structure de réseau bayésien optimale par optimisation locale d'ensembles de parents pour chaque nœud pris dans l'ordre.

L'ordonnancement des nœuds d'un réseau bayésien permet de s'assurer de l'acyclicité du graphe appris, ce qui évite une étape de recherche d'acyclicité pouvant être coûteuse.

2.4.5 Algorithmes basés sur la programmation linéaire en nombre entiers

Principe

La programmation linéaire consiste à optimiser une fonction linéaire d'un ensemble de variables liées entre elles par des relations linéaires. Pour une série de variables $X = \{X_1, .., X_n\}$, la fonction objectif à maximiser ou à minimiser prend pour forme $f(X) = a_1X_1 + a_2X_2 + ... + a_2X_n$, où $a_1, ..., a_n \in \mathbb{R}$ correspondent aux coefficients associés à chaque variable. Une relation linéaire ou contrainte entre deux variables X_1 et X_2 prend pour forme $\alpha_1X_1 + \alpha_2X_2 + ... + \alpha_nX_n \leq \beta$ où $\alpha_1, ..., \alpha_n \in \mathbb{R}$ correspondent aux coefficients associés à chaque variable pour une contrainte et $\beta \in \mathbb{R}$ à la valeur d'inégalité associée à cette contrainte.

Soit C un vecteur de taille n contenant les valeurs des coefficients $a_1, ..., a_n$ de la fonction objectif associés à chaque variable. Soit A une matrice de taille $n * p$, où p est le nombre de contraintes du problème, qui contient les valeurs des coefficients $\alpha_1, ..., \alpha_n$ de chaque contrainte, et B un vecteur

de taille p contenant les valeurs d'inégalité β de chaque contrainte. Résoudre un problème de programmation linéaire revient à résoudre l'équation 2.9 en accord avec les contrainte en 2.10 où x représente le vecteur des valeurs observées de chaque variable de X.

$$min(C^T x) \tag{2.9}$$

$$Ax \leq B \tag{2.10}$$

Un programme linéaire en nombres entiers fonctionne sur le même principe, mais une contrainte est ajoutée : Les valeurs possibles des X_i sont des nombres entiers. La fonction objectif, elle, peut toujours prendre des valeurs réelles.

Application aux réseaux bayésiens

Ce paradigme peut être utilisé afin d'apprendre la structure d'un réseau bayésien en maximisant le score associé à la structure d'un réseau bayésien sachant l'ensemble des observations des variables aléatoires qu'il contient [Bartlett and Cussens, 2013, Cussens, 2014]. Cependant, cette fonction doit pouvoir être exprimée comme une fonction linéaire d'un ensemble de variables décrivant la structure de réseau bayésien recherchée.

On peut définir pour chaque nœud un ensemble de 2^{n-1} *variables de famille*. Une *variable de famille* $W(i, \mathcal{U})$ est associée à i et à un sous-ensemble $\mathcal{U} \subseteq \mathcal{V}$ de nœuds. $W(i, \mathcal{U}) = 1$ si $\mathcal{U} = \pi_i$, c'est à dire si l'ensemble des parents de i est $\mathcal{U}$, $W(i, \mathcal{U}) = 0$ sinon. À chaque nœud i est associé une unique *variable de famille* $W(i, \mathcal{U})$ non-nulle. L'ensemble des variables d'un problème de programmation linéaire d'apprentissage de structure de réseau bayésien est donc l'ensemble des $n \cdot 2^{n-1}$ *variables de famille*, et les contraintes linéaires sur ces variables permettent d'imposer le fait que seule une de ces *variables de famille* associée à un nœud i soit non-nulle. Cela peut se faire par une série de contraintes sur le modèle de l'équation 2.11.

$$\forall i \in \mathcal{V} : \sum_{\mathcal{U} \subseteq \mathcal{V} \setminus \{i\}} W(i, \mathcal{U}) = 1 \tag{2.11}$$

Un autre groupe de contraintes permet de s'assurer que le graphe appris n'ait pas de circuit. Cela peut se faire à l'aide de contraintes [Jaakkola et al., 2010] comme décrit dans l'équation 2.12. Ces contraintes permettent de s'assurer que pour chaque sous ensemble $\mathcal{C}$ de $\mathcal{V}$, il existe au moins un nœud $i \in \mathcal{C}$ n'ayant aucun parent dans $\mathcal{C}$.

$$\forall \mathcal{C} \subseteq \mathcal{V} : \sum_{i \in \mathcal{C}} \sum_{\mathcal{U} : \mathcal{U} \cap \mathcal{C} = \emptyset} W(i, \mathcal{U}) \geq 1 \tag{2.12}$$

Une fois ces variables de famille et leurs contraintes associées définies, la résolution de ce problème donne directement l'ensemble de parents optimal pour chaque nœud i, constituant le graphe maximisant le score. Les variables de famille étant binaires, le problème est décrit dans le cadre de la programmation linéaire en nombres entiers.

La résolution d'un problème de programmation linéaire est un problème très bien décrit, dont il existe de nombreuses méthodes de résolution [Wolsey, 2000], et plusieurs outils comme *cplex* permettent de résoudre de tels problèmes. L'un des algorithmes les plus courants est l'algorithme du simplexe [Dantzig, 1990]. Il s'agit d'une résolution géométrique du problème : les contraintes

linéaires du problème décrivent un polyèdre convexe dans un espace à dimension $dim(x)$. La solution optimale est alors un des sommets de ce polyèdre. L'algorithme du simplexe consiste donc à partir d'un sommet de départ, à trouver à chaque itération en suivant une arête du polyèdre un sommet représentant une solution faisant décroître la fonction objectif. La contrainte supplémentaire induite par la programmation linéaire en nombres entiers complexifie cet algorithme. C'est pourquoi, dans ce cas, il est souvent réalisé de la *relaxation linéaire*, où la contrainte de valeurs entières de x est levée. Le résultat de ce problème constitue alors une borne supérieure de la fonction objectif, et son approximation peut être une solution acceptable, même si son optimalité n'est pas garantie. Il reste possible d'obtenir une solution binaire par arrondi.

2.5 Apprentissage de réseaux bayésiens dynamiques

Principe général

Pour rappel, un réseau bayésien dynamique décrit un ensemble de variables aléatoires X modélisant plusieurs phénomènes en interaction qui évoluent selon une dynamique stationnaire. X_i^t décrit une variable aléatoire X_i à un pas de temps t, $i \in \{1, .., n\}, t \in \{1, ..., T\}$. Cette dynamique ne change pas au cours du temps, et l'état d'une variable à un pas de temps t ne peut dépendre que de variables aléatoires au temps précédent. On décrit alors un réseau bayésien dynamique par deux composantes :

— $\mathscr{B}_0$ qui décrit un réseau bayésien "classique" statique modélisant l'ensemble des variables aléatoires X, c'est à dire à $t = 0$.

— $\mathscr{B}_{\rightarrow}$ qui décrit la composante dynamique du réseau. Cette structure décrit les indépendances de l'ensemble des variables à un temps $t + 1$ conditionnellement à l'ensemble des variables au temps t.

C'est en fait un réseau bayésien classique, mais dont la structure répond aux contraintes liées au caractère uniforme et markovien du phénomène étudié. De ce fait, apprendre la structure d'un réseau bayésien dynamique peut se faire par les mêmes méthodes que pour l'apprentissage d'un réseau bayésien classique [Murphy, 2001]. Il s'agit d'apprendre de manière indépendante la structure du graphe à l'état initial, puis la structure du graphe de transition. Cette structure se répétant au cours du temps, l'apprentissage de ceux deux parties est suffisant et n'importe quelle méthode d'apprentissage décrite dans ce chapitre peut convenir.

Exploitation de l'acyclicité d'un réseau bayésien dynamique

Malgré l'apparente complexité d'un réseau bayésien dynamique, notamment dû au fait qu'il contient souvent plus de variables qu'un réseau bayésien classique, sa régularité peut être exploitée afin d'optimiser l'apprentissage et même de proposer des méthodes mieux optimisées que dans le cas "statique". En effet, la structure d'un réseau bayésien prend la forme d'un graphe orienté acyclique. Cette contrainte oblige à vérifier l'acyclicité d'une structure apprise, par contraintes sur la structure ou par ordonnancement des nœuds. Cependant, il existe des cas particuliers de réseaux bayésiens dynamiques ne possédant aucun arc synchrone, c'est à dire que pour tout nœud i et j, il n'existe pas de lien direct entre X_i^t et X_j^t. Dans ce cas, la structure de la partie dynamique $\mathscr{B}_{\rightarrow}$ d'un réseau bayésien est naturellement acyclique. En effet, un nœud i à un temps t ne peut avoir pour parent que des nœuds provenant du pas de temps $t - 1$. Lorsque l'apprentissage de la structure initiale $\mathscr{B}_0$ n'est pas critique ou n'est pas nécessaire, l'apprentissage de la partie dynamique peut être facilitée

en ignorant l'étape de vérification d'acyclicité. [Dojer, 2006] propose un algorithme fonctionnant à la manière d'un algorithme de recherche exhaustive, permettant d'apprendre un graphe dont l'acyclicité n'a pas besoin d'être vérifiée. Cet algorithme est basé sur un score $S(x|\mathcal{G})$ ayant les propriétés suivantes :

1. Additivité : le score $S(X|\mathcal{G})$ peut être décomposé en scores locaux pour chaque nœud. $S(x|\mathcal{G}) = \sum_{i=1}^{n} s(x_i\pi_i)$
2. Décomposabilité : le score $s(x_i\pi_i)$ peut être divisé en deux termes : un terme d'entropie $H(x_i|\pi_i)$ et un terme de pénalisation $d(x_i|\pi_i)$ pénalisant le nombre de parents $|\pi_i|$. $s(x_i\pi_i) = H(x_i|\pi_i) + d(x_i|\pi_i)$ avec $\pi_i' \subseteq \pi_i \Rightarrow d(x_i|\pi_i) \leq d'(x_i|\pi_i)$
3. Uniformité : Le terme de pénalisation $d(x_i|\pi_i)$ ne dépend que du nombre de parents $|\pi_i|$. Ce terme est donc similaire pour deux sous-ensembles de parents π_i de même taille. $|\pi_i| = |\pi_i'| \Rightarrow d(x_i|\pi_i) = d(x_i|\pi_i')$

Si les propriétés 1. et 2. sont respectées, alors l'apprentissage de structure d'un réseau acyclique peut se faire par l'algorithme 3.

Entrées : x_i : Observation d'une variable aléatoire X_i, Q : Ensemble de parents potentiels de i
Output : π_1 : Liste de parents appris
$\pi_1 \leftarrow \emptyset$;
pour chaque $\pi_2 \subseteq Q$ *choisis dans l'ordre croissant de* $d(i|\pi_2)$ **faire**
  si $s(x_i, \pi_2) < s(x_i, \pi_1)$ **alors** $\pi_1 \leftarrow \pi_2$;
  si $g(x_i, \pi_2) > s(x_i, \pi_1)$ **alors**
    retourner π_1;
    stop;
  fin
fin

Algorithme 3 : Algorithme de Dojer pour un score ayant les propriétés d'additivité et de décomposabilité

Cependant, la recherche des sous ensembles $\pi_2 \subseteq Q$ peut s'avérer peu optimale, car il peut arriver de calculer des quantités $d(x_i|\pi_i)$ déjà calculées pour un autre ensemble $d(x_i|\pi_i')$. Si les propriétés 1., 2. et 3. sont respectées, alors une autre version, décrite par l'algorithme 4 permet encore d'optimiser la recherche en utilisant la quantité $\hat{d}(x_i|q) = d(x_i|\pi_i)$ avec $|\pi_i| = q$.

Entrées : x_i : Observation d'une variable aléatoire X_i, Q : Ensemble de parents potentiels de i
Output : π_1 : Liste de parents appris
$\pi_1 \leftarrow \emptyset$;
pour $q = \{1, ..., n\}$ **faire**
  si $g(x_i, q) > s(x_i, \pi_1)$ **alors**
    retourner π_1;
    stop;
  fin
  $\pi_2 \leftarrow argmin_{\{Q' \subseteq Q : |Q'| = q\}} s(x_i, Q)$;
  si $s(x_i, \pi_2) < s(x_i, \pi_1)$ **alors** $\pi_1 \leftarrow \pi_2$;
fin

Algorithme 4 : Algorithme de Dojer pour un score ayant les propriétés d'additivité, de décomposabilité et d'uniformité

Ces algorithmes permettent donc d'apprendre un réseau bayésien dynamique avec une complexité moins importante que les algorithmes d'apprentissage de structure de réseaux bayésiens classiques.

2.6 Bilan

Nous avons vu dans le chapitre 1 que pour apprendre la structure d'un réseau trophique à l'aide de données dynamiques d'observation de présence/absence d'espèces, la méthode d'apprentissage par réseau bayésien dynamique semble bien adaptée. En effet, de nombreuses méthodes permettent d'apprendre la structure d'un réseau bayésien, et aucune n'est incompatible avec des données binaires. Ce cadre est d'autant plus adapté qu'il existe une classe de réseaux bayésiens dits dynamiques convenant très bien aux données que l'on souhaite utiliser. Nous avons vu dans ce chapitre quelques unes des principales méthodes d'apprentissage de réseau bayésien (il en existe d'autres[Larrañaga et al., 1996, Tsamardinos et al., 2006, O'Gorman et al., 2015]). Toutes demandent beaucoup de calcul et peuvent demander beaucoup de temps, de données ou de mémoire si la structure que l'on cherche à apprendre est complexe [Chickering, 1996]. Cependant, cette complexité peut être réduite dans certains cas.

Pour modéliser un réseau trophique par un réseau bayésien dynamique, nous considérons l'observation de présence ou d'absence d'une espèce i à un temps t comme une variable aléatoire X_i^t, et les indépendances conditionnelles avec les variables représentant la présence ou l'absence des autres espèces à $t-1$ permet de constituer un graphe interprétable comme un réseau trophique. La dynamique des espèces au sein d'un écosystème n'est pas un phénomène dont on peut réellement établir un début. Ainsi, même lors de la première observation de cet écosystème, la dynamique entre les espèces de cet écosystème existait déjà avant. Ainsi, la composante initiale $\mathscr{B}_0$ du réseau bayésien dynamique associé ne peut pas être vraiment établie, et n'a pas vraiment d'utilité dans le problème d'apprentissage de structure du réseau trophique. Nous sommes alors tout à fait dans le cas décrit dans la section 2.5, et les algorithmes associés semblent donc applicables. Il s'agit d'algorithmes fonctionnant à base de score, ce qui présente un autre avantage par rapport aux méthodes d'apprentissage de réseau basé sur des tests d'indépendances conditionnelles : la structure du graphe est évaluable par ce score, et cette évaluation peut être utilisée lorsque l'on souhaite utiliser le réseau appris pour gérer l'écosystème correspondant [McDonald-Madden et al., 2016] comme mesure de la confiance que l'on peut attribuer aux relations trophiques correspondantes. Un autre avantage de cette modélisation par réseau bayésien est le fait qu'il est assez simple d'ajouter des connaissances expertes au processus d'apprentissage. En effet, les connaissances que l'on peut avoir sur le graphe peuvent être définies comme contraintes lors de l'apprentissage des parents de chaque nœud [De Campos et al., 2009, Bartlett and Cussens, 2017]. Le cadre des réseaux bayésiens permet également d'inclure des contraintes sur les paramètres, c'est à dire sur l'ensemble des tables de probabilités conditionnelles [Niculescu et al., 2006]. Ces contraintes peuvent refléter par exemple des valeurs connues de probabilités conditionnelles, ou encore des égalités connues entre plusieurs éléments de ces tables. Il s'agit là d'exemples de contraintes dont nous aurons besoin pour modéliser efficacement un problème d'apprentissage de réseau trophique par réseau bayésien dynamique, car les données de présence/absence dont nous disposons sont peu informatives par elles-même. Le prochain chapitre détaille donc la manière dont nous modéliserons un réseau bayésien dynamique afin de représenter la dynamique de présence et d'absence d'espèces influencées par leurs relations trophiques.

Chapitre 3

Modélisation d'un multi-processus de contact par réseau bayésien étiqueté

Introduction

Pour rappel, nous souhaitons apprendre la structure d'un réseau trophique à partir de données temporelles de présence/absence d'espèces. Pour cela, nous nous plaçons dans le cadre méthodologique des réseaux bayésiens dynamiques, dans lequel le problème d'apprentissage de structure d'un réseau est bien défini. Cependant, on trouve dans ces modèles un grand nombre de paramètres sous la forme de tables de probabilités conditionnelles. Or, c'est le nombre de ces paramètres qui détermine la complexité d'un algorithme d'apprentissage de structure de réseaux bayésiens. Réduire ce nombre de paramètres permettrait de réduire cette complexité. Il est possible d'utiliser des aprioris que l'on peut avoir sur les tables de probabilités conditionnelles dans la modélisation d'un réseau bayésien, sous forme de contraintes sur les paramètres par exemple [Niculescu et al., 2006]. Or, dans la modélisation même du comportement des espèces d'un réseau trophique, nous avons une idée de leur comportement global et nous souhaitons intégrer ces connaissances afin de simplifier le problème en fixant à l'avance un nombre de paramètres indépendant de la structure. Dans ce chapitre, nous allons :

— Décrire comment modéliser la dynamique d'espèces dans un réseau écologique à l'aide d'un modèle de réseau bayésien dynamique avec un nombre de paramètres réduit.

— Proposer le cadre général des réseaux bayésiens étiquetés, ainsi que sa version dynamique qui inclut le modèle de réseau écologique décrit avant.

— Passer en revue des exemples de modèles connus modélisables par un réseau bayésien dynamique étiqueté.

3.1 Dynamique d'espèces dans un réseau écologique

Nous avons vu dans le premier chapitre plusieurs manières de modéliser des indicateurs d'état d'espèces d'un réseau trophique. Nous souhaiterons aller plus loin en exploitant l'évolution temporelle de ces espèces et en intégrant des effets des interactions écologiques par des lois simple, mais valables

pour toute espèce de n'importe quel écosystème. Nous allons ici définir un modèle permettant de décrire le comportement de la dynamique de présence/absence de n'importe quelle espèce sachant ses interactions écologiques. Puisque notre méthode a vocation à décrire le réseau trophique de n'importe quel écosystème, nous ne souhaitons pas attribuer une grande importance à la nature des espèces en interaction. Nous souhaitons aussi proposer un modèle simple, afin qu'il soit applicable à tout écosystème et compréhensible par tous. Nous faisons alors quelques hypothèses sur le comportement de la dynamique des espèces sachant leurs interactions écologiques. Pour plus de clarté, exposons ces hypothèses dans le cas où nous n'avons que des relations proie/prédateur.

1. La dynamique d'une espèce est définie par le nombre de ses proies et le nombre de ses prédateurs. L'identification ou la nature de ses proies ou de ses prédateurs n'a pas d'importance. Pour n'importe quelle espèce, une proie est indistinguable d'un autre. Il en est de même pour les prédateurs.
2. Une espèce absente d'un écosystème peut le recoloniser à tout instant. Les interactions avec les autres espèces ne jouent aucun rôle dans cette recolonisation.
3. Si une espèce n'a aucune proie à sa disposition, elle s'éteindra (soit par mort des individus de cette espèce, soit parce que les individus de cette espèce quittent l'écosystème en question).
4. Si une espèce exerce une pression de prédation trop forte sur une proie, cette proie s'éteindra (soit par mort des individus de cette espèce, soit parce que les individus de cette espèce quittent l'écosystème en question).

Soit un écosystème dans lequel on a observé n espèces pendant T instants, $X_i^t \in \{0,1\}$ est une variable aléatoire renseignant sur la présence ($X_i^t = 1$) ou l'absence ($X_i^t = 0$) d'une espèce i à un instant t. Le fait de "n'avoir aucune proie à disposition" décrit dans l'hypothèse 3, ainsi que le fait d'"exercer une forte pression de prédation" n'est pas forcément bien défini. Certains prédateurs peuvent se nourrir sur des espèces, de manière suffisamment parcimonieuse pour que cette prédation n'ait pas un impact très fort. Certaines espèces de proies peuvent être présentes dans l'écosystème, mais avec trop peu d'individus ou trop difficiles à attraper pour pouvoir véritablement constituer une proie disponible. Puisque nous n'avons pas d'indication sur l'abondance de ces espèces, le fait de réussir à se nourrir sur une proie disponible ou d'exercer une forte pression de prédation doit se traduire par des probabilités. Soit ρ la probabilité pour une proie présente de réussir à nourrir un prédateur définie comme "Réussite d'approvisionnement". Soit τ la probabilité d'exercer une pression de prédation suffisante pour éteindre localement une proie définie comme "Réussite de prédation". Définissons également ε comme la probabilité de recolonisation de l'écosystème par une espèce. Un modèle de réseau bayésien dynamique se définit par des probabilités de transitions pour une variable X_i^t conditionnellement à un ensemble de variables de l'instant précédent $\mathbb{X}^{t-1}$. La probabilité de présence d'une espèce i à un instant t est posée dans notre modèle comme égale à :

— La probabilité de recolonisation de l'écosystème si i est absente à $t-1$

— La probabilité d'**au moins une** réussite d'approvisionnement de ses proies présentes à $t-1$ **et** de l'échec de **toutes** les prédations qu'elle subit par ses prédateurs présents à $t-1$ si i est présente à $t-1$.

La probabilité de présence d'une espèce à un instant $t-1$ sachant l'ensemble des variables aléatoires du temps t est définie par l'équation 3.1. Dans cette équation, le terme N^{proies} désigne le nombre de proies de i présentes à t et $N^{predateurs}$ le nombre de prédateurs de i présents à t.

$$P(X_i^{t+1}|X^t) = \left((1-X_i^t)\varepsilon\right) + X_i^t \left(1-(1-\rho)^{N^{proies}}\right)(1-\tau)^{N^{predateurs}} \tag{3.1}$$

Les probabilités ρ, τ et ε sont les mêmes pour toutes les espèces. Ainsi, en connaissant le graphe, un état initial de chaque espèce et la valeur de ces trois probabilités, l'ensemble des probabilités conditionnelles de chaque variable aléatoire peut être déduites. De ce fait, il n'y a que 3 paramètres inconnus dans ce modèle. Une relation trophique est assez contrainte. En effet, elle décrit forcément une double influence entre la proie et la prédateur : une influence positive de la proie sur le prédateur et une influence négative du prédateur vers la proie. Ces deux relations sont indissociables, ce qui peut poser des problèmes pour un modèle aussi simple car certains comportements peuvent ne pas transparaître ou être identifiés comme des relations trophiques à tort. Nous voudrions pouvoir distinguer d'autres types d'influences afin de moins contraindre le modèle. Cela permettrait en outre de modéliser d'autres relations écologiques. L'idée est donc de distinguer deux différents types d'interactions : les interactions positives, de probabilité de réussite ρ et les interactions négatives, de probabilité de réussite τ. Ainsi, il est possible de modéliser des relations de facilitation (une interaction positive d'une espèce vers une autre, sans relation négative dans l'autre sens), de symbioses (deux interactions positives mutuelles entre deux espèces) ou de compétition (deux interactions négatives mutuelles entre deux espèces). Nous sortons alors légèrement du cadre strict des réseaux trophiques mais traiterons le cas de la modélisation et de l'apprentissage de *réseaux écologiques*.

Cependant, ces caractérisations d'arêtes, ainsi que ce nombre de paramètres réduit et indépendant du nombre d'arcs dans le réseau ne semble plus couvert par le cadre classique des réseaux bayésiens. Lorsque l'on veut apprendre les paramètres ou la structure d'un réseau bayésien dont les probabilités conditionnelles sont décrites par un tel modèle, il n'est pas nécessaire d'apprendre les tables de probabilité de chaque variable comme des paramètres distincts. Nous introduisons dans la section suivante le concept de Réseau Bayésien étiqueté [Auclair et al., 2017], permettant de représenter des variables binaires par un réseau bayésien dans lequel les dépendances conditionnelles entre les variables ont toutes un effet similaire. Le but du réseau bayésien étiqueté est de limiter le nombre de paramètres du réseau dont le nombre dépend d'informations qualitatives sur les arcs du graphe.

3.2 Réseau bayésien étiqueté

Nous allons dans cette section présenter le modèle de réseau bayésien étiqueté (RBE) étape par étape, en généralisant de plus en plus. Nous commencerons par le modèle de base non dynamique issu des réseaux bayésiens pas à pas, avant de décrire le modèle de réseau bayésien dynamique étiqueté permettant la modélisation d'un réseau écologique et qui conviendrait pour la modélisation d'autres phénomènes.

3.2.1 Modèle de base

Un réseau bayésien étiqueté cherche à décrire l'état de n processus stochastiques, chacun décrit par une variable aléatoire binaire $X_1, X_2, ..., X_n$. La valeur d'une variable X_i est binaire et peut être interprétée comme une présence ($X_i = 1$) ou une absence ($X_i = 0$) de la variable. Cette valeur dépend d'une composante fixe, mais aussi des liens orientés avec les autres variables dont l'ensemble permet d'établir les indépendances conditionnelles entre ces variables. Dans un modèle de réseau bayésien étiqueté, chaque lien est associé à une étiquette et chaque lien d'une étiquette donnée a le même impact sur la probabilité de présence de la variable vers laquelle le lien est dirigé. L'ensemble des n processus stochastiques considérés constituent alors les nœuds d'un graphe orienté dont les arcs représentent les liens entre les variables décrivant ces processus. Nous distinguons les liens

d'impulsion, qui améliorent les chances de présence de la variable, des liens d'inhibition qui baissent les chances de présence. Les arêtes du graphe correspondant au réseau bayésien sont alors étiquetées pour distinguer les liens d'impulsion des liens d'inhibition. Définissons alors :

- $i \rightsquigarrow j$ désigne une relation quelconque de i vers j. $i \overset{q}{\rightsquigarrow} j$ désigne une relation d'impulsion de i vers j et $i \overset{r}{\rightsquigarrow} j$ désigne une relation d'inhibition de i vers j.
- π_i désigne l'ensemble de tous les nœuds $j \in 1, ..., n; j \neq i$ tels que $j \rightsquigarrow i$. π_i est désigné comme l'ensemble des parents de i, et tout nœud j tel que $j \in \pi_i$ est désigné comme parent de i. Π_i désigne l'ensemble des variables aléatoires X_j telles que $j \in \pi_i$.
- $\overset{q}{\pi}_i$ désigne l'ensemble des parents de i tels que $\forall j \in \pi_i, j \overset{q}{\rightsquigarrow} i$. Tout nœud j tel que $X_j \in \overset{q}{\pi}_i$ est désigné comme *impulseur* de i. $\overset{q}{\Pi}_i$ désigne l'ensemble des variables aléatoires X_j telles que $j \in \overset{q}{\pi}_i$.
- $\overset{r}{\pi}_i$ désigne l'ensemble des parents de i tels que $\forall j \in \pi_i, j \overset{r}{\rightsquigarrow} i$. Toute variable j telle que $X_j \in \overset{r}{\pi}_i$ est désignée comme *inhibiteur* de i. $\overset{r}{\Pi}_i$ désigne l'ensemble des variables aléatoires X_j telles que $j \in \overset{r}{\pi}_i$.

Les arêtes $\mathscr{E}$ du graphe $\mathscr{G}$ possèdent donc une information qualitative appelée *étiquette* $l \in \{q, r\}$. Tout lien $i \rightsquigarrow j$ est *actif* si $X_i = 1$ mais n'a pas d'impact si $X_i = 0$. Un lien d'impulsion $i \overset{q}{\rightsquigarrow} j$ *actif* augmente la probabilité de présence de X_j et un lien d'inhibition $i \overset{r}{\rightsquigarrow} j$ *actif* baisse les chances de présence de X_j. Dans un modèle de réseau bayésien étiqueté, les liens d'un même type ont tous le même impact : la probabilité de présence ou d'absence dépend uniquement du nombre de voisins de chaque type présents. Cette probabilité ne dépend que d'un nombre de paramètres limité :

- ε représentant la partie indépendante des liens, c'est à dire la probabilité de présence inhérente à X_i
- ρ est interprétable comme la probabilité de réussite d'un lien d'impulsion
- τ est interprétable comme la probabilité de réussite d'un lien d'inhibition

Les probabilités de toute variable X_i peuvent être exprimées conditionnellement à l'ensemble des variables Π_i correspondant à ses voisins π_i, pour lesquelles on distingue les impulseurs $\overset{q}{\Pi}_i$ et les inhibiteurs $\overset{r}{\Pi}_i$. La probabilité de présence d'une variable X_i correspond à la probabilité de présence (réussite inhérente ou d'au moins un lien d'impulsion) et de l'échec de tous les liens d'inhibition. Ces quantités font intervenir le nombre de parents d'une certaine étiquette l présents, soit pour une variable X_i, $\sum_{j \in \overset{l}{\pi}_i} X_j$, notée $\overset{l}{N}_i$ (la somme des valeurs de la variable aléatoires X_j pour tout processus j tel que j est parent de i). La formule de cette probabilité est pour tout i dans l'équation (3.2).

$$\begin{aligned} P(X_i = 1 | \Pi_i) \quad = \quad & \left(\varepsilon + (1-\varepsilon) \cdot \left(1 - (1-\rho)^{\overset{q}{N}_i}\right)\right) \\ & \cdot \quad (1-\tau)^{\overset{r}{N}_i} \end{aligned} \tag{3.2}$$

3.2.2 Gestion de l'affaiblissement des probabilités par covariable

Afin de contrôler la diffusion du processus modélisé, nous introduisons la possibilité d'associer à chaque variable un élément permettant de freiner ou d'accélérer sa diffusion. Cet élément peut être contrôlé et donné à priori ou être déterminé par un ensemble de règles. Une covariable a_i est une

variable binaire, liée à une variable X_i. Cette covariable peut donc être active ou inactive sur X_i à chaque moment considéré. La covariable n'a pas de composante aléatoire : sa valeur n'est pas décrite par des liens étiquetés du graphe associé au réseau bayésien, et n'est déterminée par aucun paramètre θ. Dans le cas général, la covariable associée à une variable X_i est décrite comme un affaiblissement, baissant la probabilité que la variable X_i soit présente lorsque la covariable associée est active. a_i décrit la covariable associée à X_i. Sa valeur décrit le fait que la covariable est active ($a_i = 1$) ou non ($a_i = 0$) sur la variable X_i. Lorsque $a_i = 1$, la variable X_i a une probabilité de présence réduite. Le calcul de la vraisemblance utilisera, en plus de ceux définis dans la section précédente, le paramètre μ_A ($0 < \mu_A <= 1$) correspondant à la probabilité d'échec de l'affaiblissement provoqué par le phénomène A sur la présence de la variable (sa réussite empêchant cette présence). Ce paramètre n'est exprimé que si la covariable est active sur X_i. Lorsque la covariable n'est pas active sur X_i cela n'impacte pas la survie ou l'apparition de X_i.La probabilité de présence de la variable X_i sachant son voisinage et l'éventuelle covariable $P(X_i = 1|\Pi_i, a_i)$ est développée dans l'équation 3.3.

$$\begin{aligned} P(x_i = 1|\Pi_i, a_i) \quad = \quad & \mu_{a_i} \\ & \cdot \quad \left(\varepsilon + (1-\varepsilon) \cdot \left(1 - (1-\rho)^{\overset{q}{N_i}}\right)\right) \\ & \cdot \quad (1-\tau)^{\overset{r}{N_i}} \end{aligned} \tag{3.3}$$

Une covariable peut correspondre à tout phénomène indépendant des relations entre individus diminuant les probabilités d'apparition et de survie sous certaines conditions. On peut, avec ce concept, modéliser un affaiblissement, mais aussi un renforcement en définissant $a_i = 0$ comme le fait qu'une covariable de renforcement est appliquée à la variable X_i et $a_i = 1$ comme le fait qu'elle n'est pas appliquée. Ce paramètre est alors interprétable comme une pénalisation des probabilités associée au fait de ne pas renforcer la variable X_i.

3.2.3 Intégration d'étiquettes de force différente

Le modèle décrit ci-dessus correspond à un cas où le réseau est décrit par des variables aléatoires binaires, une covariable et un graphe avec 2 étiquettes distinctes. On veut étendre ce modèle à des cas où la transmission peut se faire à l'aide de liens de même nature, mais pouvant compter plusieurs types d'impulseurs ou d'inhibiteurs ayant des forces différentes.Dans ce cas, un paramètre sera associé à chacun de ces différent types de lien d'une étiquette donnée, et un autre sera associé à la réussite de chaque phénomène d'affaiblissement distinct.On définit alors :

— $q = \{q_1, q_2, ..., q_{max}\}$ les différentes types d'impulseurs ; ρ_q leur probabilités de réussite.

— $r = \{r_1, r_2, ..., r_{max}\}$ les différentes types d'inhibiteurs ; τ_r leur probabilités de réussite.

La probabilité de présence d'une variable aléatoire X_i sachant son voisinage et les éventuelles covariables $P_{X_i=0}(X_i = 1|E_i, a_i)$ est développée dans l'équation 3.4.

$$\begin{aligned} P(X_i = 1|\Pi_i, a_i) = & \\ \mu \quad \cdot \quad & \left(\varepsilon + (1-\varepsilon) \cdot \left(1 - \textstyle\prod_{q=q_1}^{q_{max}} (1-\rho_q)^{\overset{q}{N_i}}\right)\right) \\ \cdot \quad & \textstyle\prod_{r=r_1}^{r_{max}} (1-\tau_r)^{\overset{r}{N_i}} \end{aligned} \tag{3.4}$$

Cette extension permet de décrire de nombreux exemples de réseaux bayésiens. Cependant, de la même manière que pour les étiquettes, il semble utile de donner la possibilité au modèle d'ingérer plusieurs forces différentes sur les covariables et sur les types de comportements inhérents.

3.2.4 Intégration de plusieurs forces de probabilités indépendantes et d'affaiblissement par covariables

En plus de permettre de considérer des étiquettes de plusieurs forces, il est possible de considérer plus d'une seule probabilité indépendantes des voisins, et plusieurs types de covariables d'affaiblissements. Ainsi, à chaque nœud i du graphe est associée une classe de nœud u_i permettent de renseigner sur son comportement inhérent, et donc sa probabilité de présence indépendante de ses voisins. De plus, nous considérons que la covariable a_i associée au nœud i ne prend pas forcément des valeurs binaires. On définit alors :

- $a_i =\in \{1, ..., \overset{max}{a}\}$ la covariable associée à i pouvant prendre plusieurs valeurs ; $\mu_{\overset{c}{a}}$ la probabilité de réussite de l'affaiblissement d'une covariable dont la valeur est $\overset{c}{a}$.
- $u_i =\in \{1, ..., \overset{max}{u}\}$ le comportements inhérent à i ; $\varepsilon_{\overset{c}{u}}$ la probabilité de présence spontanée d'une variable associée au comportement $\overset{c}{u}$.

Ainsi, la probabilité de présence d'une variable aléatoire X_i sachant son voisinage et les éventuelles covariables $P_{X_i=0}(X_i = 1|E_i, a_i)$ est développé dans l'équation 3.5.

$$\begin{aligned} P(X_i = 1|\Pi_i, a_i, u_i) = & \\ \mu_{a_i} \quad \cdot \quad & \left(\varepsilon_{u_i} + (1 - \varepsilon_{u_i}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q)^{\overset{q}{N}_i}\right)\right) \\ \cdot \quad & \prod_{r=r_1}^{r_{max}} (1 - \tau_r)^{\overset{r}{N}_i} \end{aligned} \tag{3.5}$$

Cette extension permet alors de décrire de nombreux exemples de réseaux bayésiens. La modélisation de modèles complexes peut faire intervenir des contraintes sur la structure et sur tous les types d'étiquettes du graphe associé au réseau bayésien dynamique considéré ainsi que sur différentes covariables et types de comportements inhérents. La section suivante détaille le cas du réseau bayésien dynamique étiqueté.

3.3 Réseau bayésien dynamique étiqueté

Dans le cadre des réseaux bayésiens classiques, un modèle de réseau bayésien dynamique permet de décrire un modèle dont chaque variable X_i est décrite par sa dynamique au cours de T moments. On note alors $X_i^t, t = 1, ...T$ l'état d'une variable renseignant de l'état d'un phénomène i à un instant t. Dans un réseau bayésien dynamique, l'état des variables au premier pas de temps ($t = 0$) se décrit par un réseau bayésien classique. Sur les autres pas de temps, l'état d'une variable $X_i^t(t > 0)$ ne dépend que de l'état de l'ensemble des variables au pas de temps précédent $X_j^{t-1}, j = 1, ...n,$. Nous montrons ici comment modéliser la partie "dynamique" d'un réseau bayésien dynamique ($t > 0$) sous forme de réseau bayésien étiqueté. Dans un modèle de réseau bayésien dynamique étiqueté (RBDE), l'ensemble π_i correspond aux parents du nœud i au pas de temps t désigné par la variable X_i^t. Cet ensemble de parents est le même pour tout t. Π_i^t désigne l'ensemble des variables aléatoires X_j^t telles que $j \in \pi_i$ et des étiquettes $l \in \{q_1, ..., q_{max}, r_1, ..., r_{max}\}$ des arcs renseignant des interactions entre ces variables. Pour décrire un tel modèle, nous différencions les dynamiques d'apparition ($X_i^t = 0$ et $X_i^{t+1} = 1$) et de survie ($X_i^t = 1$ et $X_i^{t+1} = 1$). Chaque lien (d'impulsion comme d'inhibition) a 2 effets différents : on distingue son impact sur la survie de celui sur l'apparition. L'impact de chacun de ces liens se traduit donc par deux paramètres : ρ^{app} décrit la probabilité réussite d'un lien d'impulsion pour l'apparition et ρ^{sur} décrit la probabilité de réussite d'un lien d'impulsion pour

la survie. De la même façon, on distinguera les probabilités de réussite des liens d'inhibition sur l'apparition τ^{app} et sur la survie τ^{sur}. En pratique, dans plusieurs modèles de RBD-E, ces liens n'auront d'impact que sur une seule dynamique -l'apparition ou la survie-, le paramètre associé à l'autre type de dynamique étant nul. Le nombre de voisins ayant un lien d'une certaine étiquette l sur i présents, soit pour une variable X_i^t, $\sum_{j \in \pi_i^l} x_j^t$ est noté $N_i^{l,t}$ (la somme des valeurs de la variable aléatoire X_j^t pour tout processus j tel que X_j appartient au voisinage de X_i, le graphe ne changeant pas au cours du temps).

De même, la probabilité inhérente à chaque variable X_i^t est interprétable comme une probabilité de dynamique spontanée, c'est à dire indépendante des interactions. On distinguera la probabilité d'apparition spontanée ε_u^{app}, utilisée par toute variable X_i^{t+1} lorsque $X_i^t = 0$, de la probabilité d'apparition spontanée ε_u^{sur}, utilisée par toute variable X_i^{t+1} lorsque $X_i^t = 1$. L'impact d'une covariable se fait également de manière dynamique : une covariable A_i^t n'a d'effet que sur la probabilité de présence de la variable X_i^{t+1}. Si une covariable a un effet différent sur la survie et sur l'apparition, on peut la modéliser par des covariables différentes selon la valeur de X_i^t. Dans un tel modèle, l'information de la valeur d'une variable X_i^t a une influence sur X_i^{t+1} : cette information décrit l'activation ou non des impulseurs, inhibiteurs et dynamiques spontanées de survie ou d'apparition. Cette influence n'est pas directement interprétable comme un impulseur ou un inhibiteur dans ce modèle, car aucun paramètre n'y est associé, mais il est tout de même intégré implicitement et représenté par un lien entre X_i^t et X_i^{t+1} dans le modèle de réseau bayésien dynamique étiqueté comme "dynamique inter-variable".

La figure 3.1 représente un graphe décrivant un phénomène modélisable par réseau bayésien étiqueté. Un tel graphe est la version statique d'un phénomène dont on peut modéliser une dynamique par un réseau bayésien dynamique étiqueté représenté en figure 3.2. En effet, entre chaque pas de temps de la figure 3.2, on trouve des arcs entre les mêmes variables que sur la figure 3.1. Par exemple, on trouve un impulseur entre du nœud 1 vers 2 dans la figure 3.1 que l'on retrouve entre chaque pas de temps sur la figure 3.2. L'équation 3.6 décrit les probabilités de survie et d'apparition de variables dans un modèle de réseau bayésien dynamique étiqueté avec un impulseur, un inhibiteur, un type de survie et d'apparition spontanée, et une covariable.

$$
\begin{aligned}
&P\left(X_i^{t+1} = 1 | X_i^t = 0, \Pi_i^t, a_i^t, u_i\right) = \\
&\qquad \mu_{a_i^t}^{app} \cdot \left(\varepsilon_{u_i}^{app} + (1 - \varepsilon_{u_i}^{app}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q^{app})^{N_i^{q,t}}\right)\right) \\
&\qquad \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r^{app})^{N_i^{r,t}} \\
&P\left(X_i^{t+1} = 1 | X_i^t = 1, \Pi_i^t, a_i^t, u_i\right) = \\
&\qquad \mu_{a_i^t}^{sur} \cdot \left(\varepsilon^{sur} + (1 - \varepsilon^{sur}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q^{sur})^{N_i^{q,t}}\right)\right) \\
&\qquad \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r^{sur})^{N_i^{r,t}}
\end{aligned}
\tag{3.6}
$$

3.4 Comparaison à des modèles connus

Les réseaux bayésiens étiquetés peuvent modéliser plusieurs phénomènes différents, et il est légitime de se demander si ce cadre constitue un cadre particulier de réseaux bayésiens, qui permettent

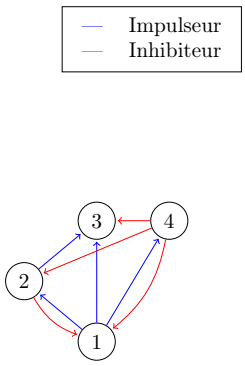


FIGURE 3.1 – Représentation étiquetée d'un graphe avec 2 types d'impulseurs et 2 types d'inhibiteurs pouvant modéliser un réseau bayésien étiqueté

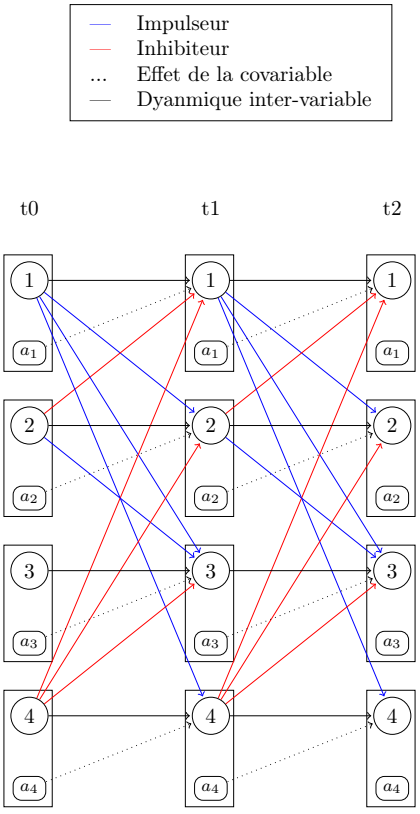


FIGURE 3.2 – Représentation des trois premiers pas de temps d'un phénomène modélisé par un réseau bayésien étiqueté avec 4 nœuds et une covariable par individu. Dans une telle représentation, on distingue des liens concernant la survie de ceux concernant l'apparition.

de faciliter la modélisation de certains phénomènes, ou s'il ne s'agit que d'une autre manière de concevoir les réseaux bayésiens, mais qu'ils permettent de modéliser n'importe quel phénomène de manière différente. En effet, un réseau bayésien étiqueté où chaque arc du graphe a une étiquette différente (le nombre d'étiquettes est alors égal au nombre d'arcs dans le graphe) est-il équivalent à un réseau bayésien "classique" ? Il est assez simple de se convaincre que ce n'est pas le cas lorsque l'on considère le nombre de paramètres d'un réseau bayésien "classique" et d'un réseau bayésien étiqueté. En effet, le nombre de paramètres d'un réseau bayésien étiqueté n'est pas dépendant de sa structure mais du nombre d'étiquettes. Prenons par exemple une variable binaire X_i dont la probabilité dépend d'une série d'autres variables $X_1, ..., X_p$. Cela se traduirait dans un réseau bayésien comme un nœud vers lequel pointent p arcs. La table de probabilité conditionnelle représentant les probabilités $P(X_i|X_1, ..., X_p)$ comporterait, dans le cas classique, 2^p paramètres si chaque variable $X_1, .., X_p$ est binaire. Dans le cas étiqueté où chaque arc comporte une étiquette différente, cette probabilité peut ne s'exprimer qu'à partir de $p + 2$ paramètres : 1 paramètre μ_i d'impact d'une éventuelle covariable associée à X_i, un paramètre ε_i de probabilité inhérente à i et p paramètres ρ et/ou τ associés aux étiquettes des p arcs pointant vers i. Dans ce cas, à partir de $p > 2$, $2^p > p + 2$. Cela se traduit par le fait que tous les éléments de la table de probabilité de $P(X_i|X_1, ..., X_p)$ d'un réseau bayésien quelconque ne peuvent pas toujours s'exprimer de manière indépendante comme dans le cas étiqueté. Un modèle de réseau bayésien étiqueté est donc un cadre particulier des réseaux bayésiens permettant de modéliser des dépendances entre certains éléments des tables de probabilités conditionnelles associées à chaque variable. Un modèle particulier de réseau bayésien appelé modèle Noisy-OR, et ses extensions, notamment le modèle Noisy-AND permettent une réduction du nombre de paramètre dans un cas particulier [Oniśko et al., 2001]. Il existe également un modèle appelé réseau bayésien qualitatif. Il s'agit d'un modèle de réseau bayésien dans lequel tout arc de i vers j porte une valeur qualitative renseignant sur une interaction augmentant ou diminuant les probabilités de j conditionnellement à j. Nous allons ici décrire ces deux modèles et expliquer les liens et les différences entre ces modèles et le réseau bayésien étiqueté.

3.4.1 Modèles Noisy-OR et Noisy-AND

Une porte Noisy-OR dans un réseau bayésien [Pearl, 1986] est une modélisation particulière de réseau bayésien où la distribution de probabilité d'une variable X_i dépend de l'état de ses voisins Π_i mais pas de leurs états conjoints. Dans le cas de variables binaires, si l'on considère une variable aléatoire X_i et son voisinage $\Pi_i = \{Y_1, ..., Y_m\}$, où chacune de ces variables a pour valeur 0 ou 1, considérons un modèle décrivant l'état de X_i sachant Π_i tel que $X_i = 1$ si $\sum_{j=1}^{m} Y_j > 0$, c'est à dire qu'au moins un de ses parents a pour valeur 1, $X_i = 0$ sinon. Une porte Noisy-OR suit une logique similaire, mais de manière moins contrainte, et considère que chaque parent Y_j de X_i tel que $Y_j = 1$ a une probabilité p_j de faire en sorte que $X_i = 1$. Pour modéliser cela, un ensemble de m variables binaires intermédiaires $Z_1, ..., Z_m$ est crée. Ces variables constituent les parents de X_i. $\forall j \in \{1, ..., m\}$, Y_j est l'unique parent de Z_j. La distribution de probabilité de chaque variable Z_j peut d'écrire : $P(Z_j = 1|Y_j) = Y_j * p_j$. L'état de X_i est alors déterminé par rapport aux états de Z_j comme précédemment : $X_i = 1$ si il existe $j \in \pi_i$ tel que $Z_j = 1$. Il est possible d'affiner ce modèle en intégrant une autre variable auxiliaire Z_0 avec $P(Z_0) = p_0$ parent de X_i afin d'ajouter une probabilité indépendante des variables Y_j pour modéliser l'état de X_i. La figure 3.3 illustre une porte Noisy-OR pour une variable X_i avec m parents. Un modèle Noisy-OR est un réseau bayésien dans lequel les probabilités conditionnelles de chaque variable X_i sont définies par des portes Noisy-OR. Ce modèle peut être généralisé à des variables non-binaires [Srinivas, 1993], au cas

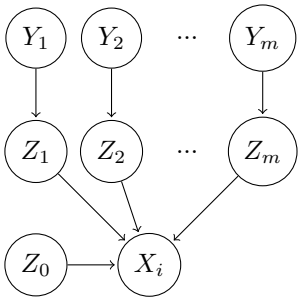


FIGURE 3.3 – Illustration d'une porte Noisy-OR pour une variable X_i.

où certaines variables Y_j ont un effet négatif sur X_i ou encore au cas où les probabilités p_i sont mal définies [Antonucci, 2011]. Une porte Noisy-AND est similaire à une Noisy-OR, mais la variable X_i vaudra 1 si toutes les variables Z_j valent 1, soit $X_i = \prod_{j=1}^{m} Z_j$. Un modèle Noisy-AND est un réseau bayésien dans lequel les probabilités conditionnelles de chaque variable X_i sont définies par des portes Noisy-AND. Voyons les ressemblances entre un tel modèle et un modèle de réseau bayésien étiqueté.

3.4.2 Lien entre modèle RBE et modèles Noisy-OR et Noisy-AND

Modélisation d'un RBE par Noisy-OR et Noisy-AND

La modélisation de la présence d'une espèce dans un réseau écologique est une probabilité modélisée par le couplage d'une porte Noisy-OR sur les proies et une porte Noisy-AND sur les prédateurs. Dans cet exemple, les probabilités p_j sont les mêmes pour toute proie j et toute probabilité p_k est la même pour tout prédateur k. Si les modèles Noisy-OR et Noisy-AND possèdent moins de paramètres qu'un modèle de réseau bayésien classique (les paramètres de ce modèle étant l'ensemble des probabilités p_j), il est encore possible d'en diminuer le nombre en considérant un nombre restreint de probabilités associés aux parents de chaque variable selon la nature de leurs interactions.

Ainsi, un réseau bayésien étiqueté est équivalent à un modèle mélangeant des portes Noisy-OR et Noisy-AND. Considérons une porte Noisy-OR pour modéliser une variable binaire Z_+ à partir d'un ensemble de variables $Y_1, Y_2, ..., Y_{m_1}$ où $p_1, ..., p_{m_1}$ sont les probabilités associées aux variables auxiliaires $Z_1, ..., Z_{m_1}$ de cette porte et Z_0 est une autre variable auxiliaire de probabilité $P(X_0 = 1) = p_0$ indépendante des variables Y_j. Considérons aussi une porte Noisy-AND pour décrire une variable Z_- à partir de m_2 variables $Y'_1, ..., Y'_{m_2}$ où $p'_1, ...,'_{m_2}$ sont les probabilités associées aux variables auxiliaires $Z'_1, ..., Z'_{m_2}$ de cette porte. Un modèle associant par une opération logique "ET" les variables Z_+ et Z_- générées par ces deux portes pour chaque variable X_i est équivalent à un modèle de réseau bayésien dynamique étiqueté où chaque lien possède une étiquette unique différente. En effet, on retrouve des paramètres équivalents :

— $\varepsilon_{u_i} = p_0$

— $\rho_{q_1} = p_1, ..., \rho_{q_{max}} = p_{m_1}$

— $\tau_{r_1} = 1 - p'_1, ..., \tau_{r_{max}} = 1 - p'_{m_2}$

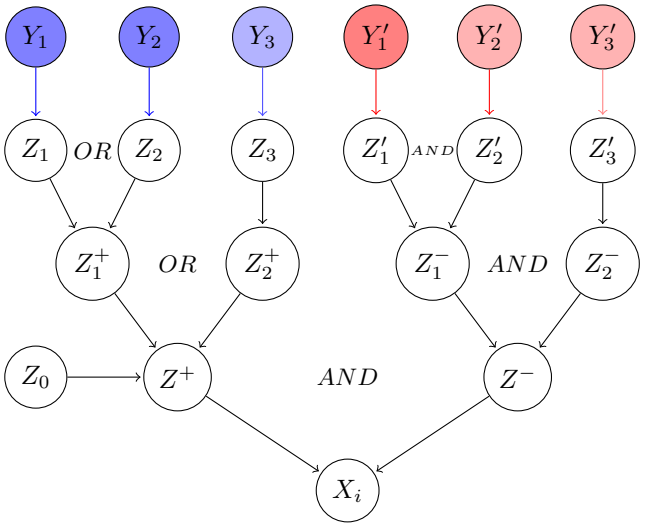


FIGURE 3.4 – Illustration d'un modèle de réseau bayésien dynamique étiqueté sous forme de portes Noisy-OR et Noisy-AND pour une variable X_i à 6 parents de 4 étiquettes différentes. Les nœuds colorés représentent les influences positives (en bleu) et négatives (en rouge) de i dont les étiquettes sont représentées par l'intensité des couleurs.

Cela correspond alors à un modèle de réseau bayésien étiqueté comportant m_1 impulseurs et m_2 inhibiteurs. Il est possible de modéliser tout modèle de réseau bayésien étiqueté avec un nombre quelconque d'impulseurs et d'inhibiteurs par un modèle comportant pour chaque variable X_i :

— Une variable Z_+ générée par un cas particulier de porte Noisy-OR où l'on connaît en avance des égalités entre les probabilités p_j pour tout nœud j impulseur, c'est à dire ayant une influence positive sur i.

— Une variable Z_- générée par un cas particulier de porte Noisy-AND où l'on connaît en avance des égalités entre les probabilités p_k pour tout nœud k inhibiteur, c'est à dire ayant une influence négative sur i.

— Une combinaison logique "ET" des variables Z_+ et Z_-

Un exemple de modélisation de réseau bayésien étiqueté sous forme de porte Noisy-OR et Noisy-AND est représenté en figure 3.4.

Dans un tel modèle, pour un nœud i, chaque variable auxiliaire $Z_q, q \in \{q_1, ..., q_{max}\}$ est générée par une porte Noisy-OR pour chaque variable $Y_j, j \in \overset{q}{\pi}_i$ avec une même probabilité $p_j = \rho^q$. De même, chaque variable auxiliaire $Z'_r, r \in \{r_1, ..., r_{max}\}$ est générée par une porte Noisy-AND pour chaque variable $Y'_j, j \in \overset{r}{\pi}_i$ avec une même probabilité $p_j = 1 - \tau^r$. Chaque variable Z'_r décrit alors l'échec de toutes les interactions d'inhibition d'étiquette r. La valeur de la variable auxiliaire Z_0 est décrite par une probabilité ε inhérente à X_i. La valeur de la variable auxiliaire Z_+ est déterminée par l'ensemble des variables $Z_q, q \in \{q_1, ..., q_{max}\}$ avec $Z_+ = 1$ si $\exists q$ tel que $Z_q = 1$, $Z_+ = 0$ sinon. La valeur de la variable auxiliaire Z_- est alors déterminée par l'ensemble des variables $Z'_r, r \in \{r_1, ..., r_{max}\}$ avec $Z_- = 1$ si $\forall r$, $Z'_r = 1$. $Z_- = 0$ sinon. La variable Z_+ décrit donc la réussite d'au moins une interaction d'impulsion sur i et la variable Z_- la réussite de toutes les interactions

d'impulsion. La valeur de X_i est déterminée par une opération logique "ET" entre Z_+ et Z_-avec $X_i = 1$ si $\forall q, r, Z_+ = 1$ et $Z_- = 1$, $X_i = 0$ sinon. Ainsi, on a $X_i = 1$ si au moins une interaction d'impulsion ou la probabilité inhérente réussit et toutes les interactions d'inhibition échouent, ce qui correspond bien au modèle de réseau bayésien étiqueté.

Modélisation de Noisy-OR et Noisy-AND par RBE

Un modèle Noisy-OR ou Noisy-AND peut également être modélisé par réseau bayésien étiqueté [1]. Commençons par le cas Noisy-OR en considérant un modèle de réseau bayésien étiqueté où :

— Tous les arcs du graphe correspondent à des relations d'impulsion (il n'y a pas d'inhibiteurs).

— Chaque arc du graphe a une étiquette différente.

— Chaque nœud appartient à une classe différente et a une probabilité inhérente différente.

Ainsi, pour tout nœud i, la probabilité de présence de X_i dépend de sa probabilité inhérente ε et des probabilités de réussite de chacun de ses parents $\rho_j \forall j \in \pi_i$, tous des impulseurs de force différente [2]. Cette probabilité est exprimée par l'équation 3.7.

$$P(X_i = 1|\Pi_i) = \varepsilon + (1 - \varepsilon) \cdot \left(1 - \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) \right) \tag{3.7}$$

Or, dans un modèle Noisy-OR, l'état de la variable aléatoire associée à chaque nœud i est exprimée par une porte Noisy-OR. C'est à dire une combinaison logique "ou" de variables binaires $Z_j \forall j \in \pi_i$ auxiliaires ainsi que d'une variable binaire auxiliaire Z_0. Sa probabilité peut donc être exprimée en fonction de la probabilité de chaque variable Z_j, qui s'exprime par une probabilité p_j si une variable aléatoire associée $Y_j = 1$, la variable auxiliaire Z_0 ayant une probabilité indépendante de toute autre variable. La probabilité de présence d'une variable X_i issue d'un modèle Noisy-OR est exprimée par l'équation 3.8.

$$P(X_i = 1|\Pi_i) = 1 - (1 - P(Z_0)) \cdot \prod_{j \in \pi_i \ tqY_j = 1} (1 - p_j) \tag{3.8}$$

Ce modèle est strictement équivalent au modèle de réseau bayésien dynamique étiqueté dont les probabilités sont exprimées par l'équation 3.7 si $p_j = \rho_j$ et $P(Z_0) = \varepsilon$, comme le montre l'équation 3.9.

$$\begin{aligned} P(X_i = 1|\Pi_i) &= \varepsilon + (1 - \varepsilon) \cdot \left(1 - \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) \right) \\ &= \varepsilon + 1 - \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) - \varepsilon + \varepsilon \cdot \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) \\ &= 1 - \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) + \varepsilon \cdot \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) \\ &= 1 - \prod_{j \in \pi_i \ tqX_j = 1} (1 - \rho_j) \cdot (1 - \varepsilon) \end{aligned} \tag{3.9}$$

De même, dans un modèle Noisy-AND, l'état d'une variable aléatoire X_i est exprimé par une porte Noisy-AND, c'est à dire une combinaison logique "et" de variables binaires $Z_j \forall j \in \pi_i$ et Z_0

1. Ce n'est le cas que pour des modèles Noisy-OR et Noisy-AND "classiques" où chaque variable est binaire et où une variable auxiliaire Z_j vaut 0 si la variable aléatoire Y_j associée vaut 0. Il existe par exemple des modèles Noisy-OR et Noisy-AND multivalués qui ne sont pas modélisables par réseau bayésien étiqueté.

2. Chaque nœud i a des valeurs différentes pour ε et ρ_j. En toute rigueur, il faudrait indicer ces quantités par i en désignant ε_i et $\rho_{j \to i}$, mais cela alourdirait la lecture.

auxiliaires générées de la même façon que dans le modèle Noisy-AND par des probabilités p_j lorsque les variables aléatoires Y_j associées valent 1. La probabilité de présence d'une variable X_i issue d'un modèle Noisy-AND est exprimée par l'équation 3.10.

$$P(X_i = 1|\Pi_i) = P(Z_0) \cdot \prod_{j \in \pi_i \ tqY_j = 1} (p_j) \tag{3.10}$$

Un tel modèle est strictement équivalent à un modèle de réseau bayésien étiqueté ne contenant aucun impulseur et où chaque inhibiteur est étiqueté différemment. Pour tout nœud i, la probabilité de présence de X_i dépend de sa probabilité inhérente ε et des probabilités de réussite de chacun de ses parents $\tau_j \forall j \in pi_i$, tous des inhibiteurs d'étiquettes différentes. Cette probabilité, exprimée par l'équation 3.11 est équivalente à l'équation 3.10 avec $p_j = 1 - \tau_j$ et $P(Z_0) = \varepsilon$.

$$P(X_i = 1|\Pi_i) = \varepsilon \cdot \prod_{j \in \pi_i \ tqX_j = 1} (1 - \tau_j) \tag{3.11}$$

Nous avons donc vu qu'un modèle Noisy-OR ou noisy-AND peut être modélisé directement dans le cadre des réseaux bayésiens étiquetés. Pour modéliser un réseau bayésien étiqueté par Noisy-OR et Noisy-AND, il faut forcer des égalités entre probabilités en fonction des étiquettes des arcs du graphe, ce qui n'est pas prévu par les modèles classique de réseaux bayésiens. Il existe cependant un cas particulier de réseau bayésien dans lequel les arcs du graphe portent une information qualitative.

3.4.3 Réseau bayésien qualitatif

Un réseau bayésien qualitatif [Wellman, 1990] est un modèle de réseau bayésien à variables binaires dans lequel plusieurs arcs du graphe associé représentent des influences qualitatives. Une influence qualitative exprime le fait qu'une variable puisse affecter la distribution de probabilité de présence d'une autre variable positivement ou négativement. Soient deux variables X et Y pour lesquelles il existe un arc de X vers Y. Une influence qualitative positive décrit le fait que X augmente la probabilité de Y comme décrit dans l'équation 3.12. Une influence qualitative négative décrit le fait que X baisse la probabilité de présence Y comme décrit dans l'équation 3.13.

$$P(Y = 1|X = 1) \geq P(Y = 1|X = 0) \tag{3.12}$$

$$P(Y = 1|X = 1) \leq P(Y = 1|X = 0) \tag{3.13}$$

Il existe des méthodes permettant d'apprendre les paramètres d'un réseau bayésien qualitatif en respectant les influences qualitatives [Feelders and van der Gaag, 2006]. Dans un modèle de réseau bayésien étiqueté, la nature positive ou négative des arcs étiquetés respecte cette définition. En effet, si X est un impulseur de Y et que toutes les interactions et les valeurs des autres variables sont connues, si $X = 1$, alors la quantité N dans l'expression (avec des notations simplifiées) $\left(1 - (1 - \rho)^N\right)$ augmentera de 1, ce qui augmentera la probabilité générale. De même, si X est un inhibiteur de Y et que toutes les interactions et les valeurs des autres variables sont connues, si $X = 1$, alors la quantité N dans l'expression (avec des notations simplifiées) $(1 - \tau)^N$ augmentera de 1, ce qui diminuera la probabilité générale. Cependant, un modèle de réseau bayésien étiqueté reste différent d'un modèle de réseau bayésien qualitatif. En effet, ce dernier ajoute juste un prior sur les probabilités conditionnelles à partir des caractérisations qualitatives des arêtes. Un réseau

	Valeur par défaut			
	Covariable	Disparition Spontanée	Impulsée	Inhibée
$P(X_i^{t+1} = 1 \mid X_i^t = 1, a(X_i^t), \Pi_i^t)$	$\mu = 1^*$	$\varepsilon = 0^*$	$\rho = 0^*$	$\tau = 0^*$

TABLE 3.1 – Valeurs par défaut de chaque type de paramètre

bayésien dynamique étiqueté modifie la manière dont sont générées ces probabilités en caractérisant les étiquettes des arcs par des paramètres associés. Cependant, le paradigme des réseaux bayésiens qualitatifs peut être un lien possible entre un réseau bayésien "classique" et un réseau bayésien étiqueté, car la structure qualitative du réseau est proche de la structure étiquetée décrite dans ce chapitre.

3.5 Exemples de processus modélisables par réseau bayésien dynamique étiqueté

Ce modèle peut être utile dans d'autres cadres que la modélisation d'un réseau écologique. Il s'applique notamment très bien au cas des processus de contact [Harris, 1974]. En effet, on considère qu'une variable aléatoire décrit l'état d'un individu i lorsqu'une information (une infection par exemple) contagieuse est présente dans le système. La propagation d'épidémies ou de rumeurs dans des réseaux sociaux en sont des exemples [Keeling and Eames, 2005]. Dans la section suivante, nous allons décrire quelques exemples de phénomènes de la littérature modélisables par réseau bayésien dynamique étiqueté. La plupart de ces modèles sont bien moins complexes que le modèle général, car ils n'impliquent pas tous les paramètres du modèle général, ou fixent une valeur pour certains paramètres. Lorsqu'une relation, une dynamique spontanée ou l'effet d'une covariable n'est pas impliquée dans la dynamique d'un processus, nous attribuons au paramètre associé une valeur *par défaut*, qui n'influe pas sur les probabilités de dynamique. Une valeur par défaut est indiquée par une astérisque, et la valeur par défaut de chaque paramètre est indiquée dans la table 3.1.

3.5.1 Processus de contact ou modèle SIS sur graphe

Un processus de contact classique[Harris, 1974, Franc, 2004] dit modèle SIS (sain-infecté-sain) en épidémiologie est un modèle qui permet d'étudier la propagation d'un agent infectieux dans une population regroupée au sein d'un réseau décrivant les individus amenés à être en contact, facilitant la propagation. Chaque individu est donc en contact avec un certain nombre de voisins. Les individus peuvent être dans un état infecté (I) ou sain (S), et on observe l'évolution de l'état des individus au cours d'un temps discret. L'évolution de l'état des individus d'un instant t à un instant $t+1$ se fait de la façon suivante [3] :

— Un individu infecté peut redevenir sain avec une probabilité ε

— Un individu sain peut être infecté par ses voisins. Chacun de ses voisins infectés $\pi_{Infectes}$ a une probabilité ρ de lui transmettre l'infection. La probabilité d'infection d'un individu est donc

3. Les notations utilisés dans ce paragraphe ne correspondent pas forcément aux notations utilisés dans la présentation du modèle de RBD étiqueté.

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Apparition	Spontanée	0^*	Pas d'apparition spontanée
	Impulsée	ρ^{app}	1 type d'impulseur : contact entre 2 individus permettant la transmission
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	1^*	Aucune covariable dans ce modèle
Survie	Spontanée	ε^{sur}	Disparition spontanée possible (guérison)
	Impulsée	0^*	Aucun effet des liens d'impulsion sur la survie
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	1^*	Aucune covariable dans ce modèle

TABLE 3.2 – Valeur des paramètres d'un RBDE modélisant un processus de contact en épidémiologie SIS
*Paramètre non utilisé : utilisation de la valeur par défaut

la probabilité qu'au moins l'un de ses voisins infectés réussisse à lui transmettre l'infection : $P(I|S, \pi_{Infectes}) = 1 - (1 - \rho)^{\#\pi_{Infectes}}$ ($\#\pi_{Infectes}$ désigne le nombre de parents infectés).

Il est assez simple de modéliser un modèle SIS par un réseau bayésien dynamique étiqueté. En reprenant les notations des sections précédentes pour un individu :

— Si ε^{sur}, $1 - \varepsilon^{sur}$ désigne une probabilité de guérison indépendante des contacts, on a alors $P(S|I) = 1 - \varepsilon^{sur}$

— ρ^{app} peut désigner la probabilité de réussite d'infection d'un voisin infecté. Dans ce cas, la probabilité qu'au moins l'un des voisins infectés réussisse à transmettre l'infection à l'individu peut s'écrire $P(I|S, \pi_{Infectes}) = 1 - (1 - \rho^{app})^{\#\pi_{Infectes}}$ ($\#\pi_{Infectes}$ désigne le nombre de parents infectés).

Un modèle de type SIS n'utilise pas de covariables et le réseau considéré ne contient qu'une seule sorte d'impulseur. Si l'apparition de l'infection est due uniquement au voisinage des individus, sa disparition ne peut être que spontanée. Les liens d'impulsion n'ont donc pas d'effet sur la disparition. Dans ce cadre, un modèle SIS simple n'utilise donc que les paramètres ε^{sur} (disparition spontanée) et ρ^{app} (réussite d'un lien d'impulsion sur l'apparition) du modèle général. Les autres paramètres prennent les valeurs par défaut. Les valeurs des paramètres sont données dans la table 3.2. L'expression simplifiée des probabilités de transitions d'un modèle SIS à partir du modèle de RBD-E multicontact est donnée dans l'équation (3.14).

$$\begin{array}{lcl} P\left(X_i^{t+1} = 1 | X_i^t = 0, \Pi_i^t\right) & = & 1 - (1 - \rho^{app})^{\overset{q}{N_i^t}} \\ P\left(X_i^{t+1} = 1 | X_i^t = 1\right) & = & \varepsilon^{sur} \end{array} \tag{3.14}$$

3.5.2 Processus d'épidémiologie SIR

Un processus épidémiologique SIR [Salathé and Jones, 2010] est un processus de contact dont la dynamique est similaire à un modèle SIS, mais où les individus peuvent avoir, en plus de l'état sain (S) et infecté (I), un état résistant (R). Le principe est qu'un individu infecté ne peut plus redevenir sain. L'infection peut toujours disparaître spontanément, mais dans ce cas, l'individu passe

de l'état infecté à l'état résistant (correspondant à une immunité à l'infection ou à la mort). Un individu résistant ne peut plus être infecté. L'évolution de l'état des individus se fait alors de la façon suivante [4] :

— Comme dans le modèle SIS, un individu sain peut être infecté par ses voisins. Chacun de ses voisins E infectés a une probabilité ρ de lui transmettre l'infection. La probabilité d'infection est donc la probabilité qu'au moins l'un de ses voisins infectés réussisse à lui transmettre l'infection : $P(I|S,\{E\}) = 1-(1-\rho)^{\#\pi_{infectes}}$

— Un individu infecté peut devenir résistant avec une probabilité ε de guérison (indépendante des contacts) : $P(R|I) = \varepsilon$

— Un individu infecté ne peut plus redevenir sain ($P(S|I) = 0$) et un individu résistant ne peut plus être infecté ($P(I|R) = 0$)

Un modèle RBDE ne compte que des variables binaires, ce qui pose un problème pour un modèle SIR, où 3 états sont possibles chez un individu. En revanche, il est possible d'ajouter à un RBDE des covariables associées à des individus afin de modifier leurs probabilités de survie et d'apparition. L'état résistant d'un individu i peut donc être défini par une telle covariable. Définissons une quantité a_i^t comme covariable de X_i^t définie de matière déterministe à partir de X_i^{t-1} et a_i^{t-1}. Cette quantité se construit de la façon suivante pour chaque individu i et chaque instant t :

— Au départ, aucun individu n'est résistant : $a_i^0 = 0$.

— Un individu infecté voit sa covariable a_i prendre la valeur 1 au pas de temps suivant : Si $X_i^t = 1$, alors $a_i^{t+1} = 1$.

— Un individu résistant restera résistant au pas de temps suivant : Si $a_i^t = 1$, alors $a_i^{t+1} = 1$.

Cette quantité peut être interprétée de la même manière qu'une covariable dans le modèle général car sa valeur ne dépend d'aucun paramètre et n'est pas aléatoire sachant l'ensemble des valeurs de X_i^t. La combinaison des variables X_i^t et a_i^t peut décrire l'état des variables :

— Si $X_i^t = 0$ et $a_i^t = 0$, alors l'individu i est sain (S).

— Si $X_i^t = 1$ alors l'individu est infecté (I), peu importe la valeur de a_i^t.

— Si $X_i^t = 0$ et $a_i^t = 1$, alors l'individu i est résistant (R).

Lorsque $a_i^t = 1$, alors l'individu i est résistant (à l'instant t). L'individu désigné comme résistant dans ce modèle pourra rester infecté jusqu'à guérison ou mort (a_i^t n'a pas d'impact sur la probabilité de présence X_i^{t+1} si $X_i^t = 1$), mais un individu résistant non infecté ne pourra plus être infecté. Cette covariable n'a donc aucun effet sur la survie de l'infection et la probabilité d'apparition de l'infection chez un individu résistant est nulle. Il est possible de modéliser un modèle SIR par réseau bayésien dynamique étiqueté, en considérant un modèle similaire au SIS, auquel on ajoute les contraintes suivantes :

— $\mu_1^{app} = 0$: Lorsque l'individu est résistant, l'infection ne peut plus apparaître. La résistance de l'individu n'a en revanche pas d'effet sur la survie de l'infection.

L'expression des probabilités de chaque dynamique est exprimée en (3.15).

$$\begin{aligned} P\left(X_i^{t+1} = 1 | X_i^t = 0, \Pi_i^t, a_i^t\right) &= \mu_{a_i^t}^{app} \cdot \left(1-(1-\rho^{app})^{N_i^t}\right) \\ P\left(X_i^{t+1} = 1 | X_i^t = 1, a_i^t\right) &= \varepsilon^{sur} \end{aligned} \tag{3.15}$$

4. Les notations utilisées dans ce paragraphe ne correspondent pas forcément aux notations utilisées dans la présentation du modèle de RBD étiqueté.

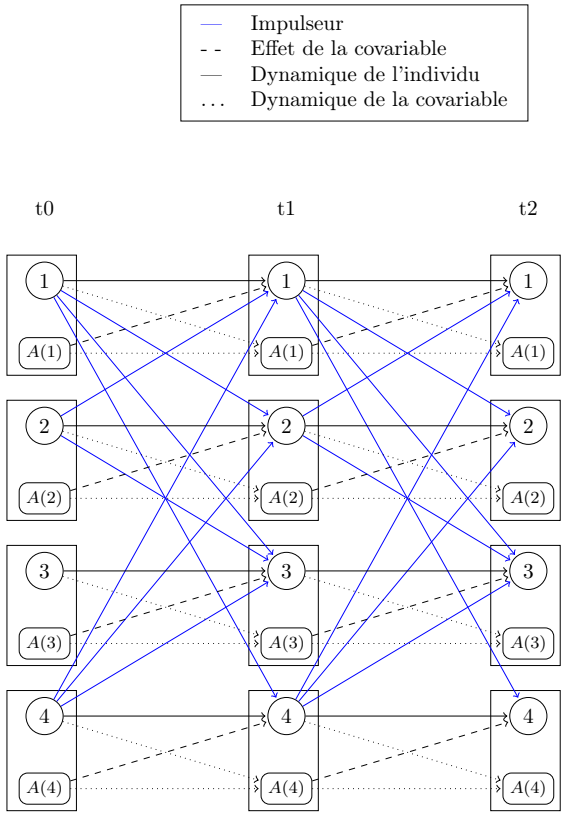


FIGURE 3.5 – Représentation des trois premiers pas de temps d'un processus de contact SIR modélisé par un réseau bayésien étiqueté

Toutes les contraintes sur les paramètres de ce modèle, que l'on peut représenter sous forme graphique par la figure 3.5 sont décrites dans la table 3.3.

3.5.3 Gestion de réseaux écologiques spatiaux avec des actions simultanées

Le cadre des réseaux bayésiens étiquetés multi-contacts peut décrire des modèles plus complexes, utilisés dans la gestion de réseaux écologiques spatiaux [Nicol et al., 2017]. Dans ce problème, la variable aléatoire correspond à la présence d'une espèce dans une mare. Afin d'éviter la prolifération de l'espèce invasive, une action d'éradication peut être appliquée à une mare peuplée par cette espèce pour donner la possibilité à l'espèce endémique de repeupler la mare. La zone dans laquelle se trouvent ces mares est régulièrement inondée. Lors de cette inondation, les espèces peuvent aller coloniser de nouvelles mares. La présence de l'espèce invasive dans une mare diminue les chances de voir l'espèce endémique survivre ou recoloniser cette mare. Chaque mare a donc un ensemble de mares voisines, reliées lors de l'inondation. On peut décrire ce modèle de la manière suivante :

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Apparition	Spontanée	0^*	Pas d'apparition spontanée
	Impulsée	ρ^{app}	1 type d'impulseur : contact entre 2 individus permettant la transmission
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	$\mu_1^{app} = 0$	Si $a_i^t = 1$ alors la probabilité d'apparition est nulle pour l'individu i au temps $t+1$
Survie	Spontanée	ε^{sur}	Disparition spontanée possible (guérison)
	Impulsée	0^*	Aucun effet des liens d'impulsion sur la survie
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	1^*	La covariable n'influe pas sur la survie

TABLE 3.3 – Résumé des paramètres utilisés dans le modèle de contact en épidémiologie SIR
*Paramètre non utilisé : utilisation de la valeur par défaut

— Chaque mare i peut contenir des espèces endémiques, dont la présence est décrite par une variable aléatoire X_i ; et des espèces invasives, dont la présence est décrite par une variable aléatoire Y_i. Toute mare i est reliée à un ensemble de mares π_i

— Chaque mare peut recevoir des individus de l'espèce endémique de ses mares voisines en contenant. La probabilité de transmission d'une espèce endémique d'une mare à l'autre est décrite par ρ^{end}. La probabilité pour qu'une mare acquière des espèces endémiques est donc de $1-(1-\rho^{end})^{\#\pi^{end}}$ ($\#\pi^{end}$ désigne le nombre de mares voisines contenant des espèces endémiques)

— Chaque mare peut recevoir des individus d'espèces invasives de ses mares voisines en contenant. La probabilité de transmission d'une espèce invasive d'une mare à l'autre est décrite par ρ^{inv}. La probabilité pour qu'une mare acquière des espèces invasives est donc de $1-(1-\rho^{inv})^{\#\pi^{inv}}$ ($\#\pi^{inv}$ désigne le nombre de mares voisines contenant des espèces invasives)

— La présence d'une espèce invasive dans une mare i lors d'une année t baisse les chances de présence de l'espèce endémique dans cette mare lors de l'année $t+1$. La présence de l'espèce invasive à t a une probabilité τ^{ext} de provoquer l'extinction et une probabilité τ^{col} d'empêcher la colonisation de l'espèce endémique à $t+1$ dans la mare i.

— Toute mare peut être empoisonnée afin d'éradiquer l'espèce invasive. Cet empoisonnement affecte également l'espèce endémique. Une mare empoisonnée a une probabilité μ de provoquer l'extinction de l'espèce endémique et/ou de l'espèce invasive.

La modélisation du fait que deux espèces se trouvent dans une même mare se fait par une connaissance à priori d'une partie du réseau associé à ce processus. En effet, on sait qu'une interaction entre deux mares correspond à une interaction entre chacune des deux espèces de chacune de ces deux mares. En outre, on connaît les interactions entre les espèces d'une même mare : l'espèce invasive dans une mare limite l'apparition et la survie de l'espèce invasive dans la même mare. Les interactions et la dynamique de ces espèces sont décrites par un même réseau bayésien dynamique étiqueté multicontact. Décrivons pas à pas un tel modèle.

Modélisation de la dynamique des espèces par inondation des mares

Décrivons tout d'abord la dynamique commune aux deux espèces, qui décrit la colonisation des mares reliées entre elles lors des inondations. La présence d'une espèce dans une mare i permet la colonisation dans les mares en relation décrites dans le cadre des RBD-E multicontact comme des impulseurs. Il y en a de deux types différents : les impulseurs de l'espèce endémique q_{end} et ceux de l'espèce invasive q_{inv}. ρ_{inv}^{app} correspond à la probabilité de colonisation d'une espèce invasive et ρ_{end}^{app} correspond à la probabilité de colonisation d'une espèce endémique. L'espèce invasive a également la possibilité de coloniser spontanément une mare avec une probabilité $\varepsilon_{u_i^{inv}}^{app}$ où u_i^{inv} renseigne si l'espèce i est invasive, l'inondation ouvrant le système de mares à l'extérieur. L'espèce endémique (u_i^{end} renseigne si espèce i est endémique) n'a pas de possibilité de colonisation spontanée($\varepsilon_{u_i^{end}}^{app} = 0$, cette quantité n'apparaît pas dans les calculs de probabilités). Chaque individu du modèle ne pouvant correspondre qu'à un seul type d'espèce (endémique ou invasive), il ne peut recevoir des impulseurs d'apparition que d'un type unique. En revanche, les relations entre les mares n'ont aucun effet sur la survie des espèces les peuplant. Les liens d'impulsions n'ont donc pas d'effet sur la survie. Chaque espèce peut s'éteindre spontanément avec une probabilité $1 - \varepsilon_{u_i^{end}}^{sur}$ si cette espèce est endémique et $1 - \varepsilon_{inv}^{sur}$ si cette espèce est invasive. Chaque individu du modèle ne pouvant correspondre qu'à un seul type d'espèce (endémique ou invasive), il ne peut être sujet qu'à un seul type d'apparition spontanée. Les différentes probabilités de dynamique sont décrites dans les équations en (3.16). Ces quantités sont bien des probabilités, car nous rappelons qu'un individu correspond soit à une espèce endémique, soit à une espèce invasive. Dans ce cas, lorsque i est une espèce endémique, $\varepsilon_{u_i^{inv}}^{app} = 0$, $\varepsilon_{u_i^{inv}}^{sur} = 0$ et $\#\pi_{inv}^{q} = 0$; et lorsque i est une espèce invasive, $\varepsilon_{u_i^{end}}^{sur} = 0$ et $\#\pi_{end}^{q} = 0$. Par souci de clarté, le nombre de voisins d'une étiquette l d'une espèce endémique est noté $\overset{l}{N}{}_{i\,end}^{t}$ et le nombre de voisins étiquetés l d'une espèce invasive est noté $\overset{l}{N}{}_{i\,inv}^{t}$.

$$
\begin{aligned}
P\left(X_i^{t+1} = 1 | X_i^t = 0, \Pi_i^t\right) = \quad & \varepsilon_{u_i^{inv}}^{app} + (1 - \varepsilon_{u_i^{inv}}^{app}) \cdot \left(1 - (1 - \rho_{end}^{app})^{\overset{q}{N}{}_{i\,end}^{t}} \cdot (1 - \rho_{inv}^{app})^{\overset{q}{N}{}_{i\,inv}^{t}}\right) \\
P\left(X_i^{t+1} = 1 | X_i^t = 1\right) = \quad & \varepsilon_{u_i^{end}}^{sur} + \varepsilon_{u_i^{inv}}^{sur}
\end{aligned}
\tag{3.16}
$$

Modélisation de l'impact de l'espèce invasive sur la dynamique de l'espèce endémique

La présence de l'espèce invasive dans une mare réduit les chances de survie et de colonisation de l'espèce endémique dans cette mare. Dans le modèle décrit par [Nicol et al., 2017], cela se traduit par des probabilités de colonisation $\rho_{end|inv}$ et de survie $\varepsilon_{u_{end}|uinv}$ plus faibles lorsque $Y_i^t = 1$. Dans le cadre des RBDE , on veut éviter que la valeur d'un paramètre change selon l'état des individus. Il est préférable de modéliser l'impact de l'espèce invasive sur l'espèce endémique comme des inhibiteurs ayant à la fois un effet sur l'apparition et sur la survie. Cette inhibition se fait au sein d'une même mare. On a donc pour tout moment t et tous individus i et j, tel que j correspond à l'espèce invasive de la même mare que i, un inhibiteur r de X_j^t vers X_i^{t+1} dont les probabilités de réussite sont de τ_{app} pour l'apparition et de τ_{sur} pour la survie. Avec cette modélisation les probabilités d'apparition et de survie des espèces sont décrites dans les équations en 3.17. Ces quantités restent des probabilités

car lorsque i est une espèce invasive, $\#\overset{r}{E}{}^t_{i\,i}nv = 0$.

$$
\begin{aligned}
P\left(X_i^{t+1}=1|X_i^t=0,\Pi_i^t\right) &= \left(\varepsilon^{app}_{u_i^{inv}} + (1-\varepsilon^{app}_{u_i^{inv}})\cdot\left(1-(1-\rho^{app}_{end})^{\overset{q}{N^t_{i\,end}}}\cdot(1-\rho^{app}_{inv})^{\overset{q}{N^t_{i\,inv}}}\right)\right)\cdot(1-\tau^{app})^{\overset{r}{N^t_{i\,inv}}}\\
P\left(X_i^{t+1}=1|X_i^t=1,\Pi_i^t\right) &= \left(\varepsilon^{sur}_{u_i^{inv}} + \varepsilon^{sur}_{u_i^{end}}\right)\cdot(1-\tau^{sur})^{\overset{r}{N^t_{i\,end}}}
\end{aligned}
\tag{3.17}
$$

Covariables sur les mares

Afin d'éviter la prolifération de l'espèce invasive, on peut tenter d'éradiquer les espèces présentes dans une mare en l'empoisonnant. Ceci se traduit dans un modèle de RBD-E multicontact par l'application d'une covariable A associé à un individu i. Cette action s'applique de la même manière sur deux individus d'une même mare : pour deux espèces i et j issues d'une même mare, $a_i^t = a_j^t$. On admettra que l'effet de la covariable est le même sur la colonisation et sur la survie. Un unique paramètre μ décrit alors la probabilité de réussite de la covariable (action d'éradication). Les probabilités de survie et d'apparition pour les deux espèces sont décrites dans les équations en (3.18).

$$
\begin{aligned}
P\left(X_i^{t+1}=1|X_i^t=0,\Pi_i^t,a_i^t\right) &= \mu^{app}_{a_i^t}\cdot\left(\varepsilon^{app}_{u_i^{inv}} + (1-\varepsilon^{app}_{u_i^{inv}})\cdot\left(1-(1-\rho^{app}_{inv})^{\overset{q}{N^t_{i\,end}}}\cdot(1-\rho^{app}_{end})^{\overset{q}{N^t_{i\,inv}}}\right)\right)\cdot(1-\tau^{app})^{\overset{r}{N^t_{i\,end}}}\\
P\left(X_i^{t+1}=1|X_i^t=1,\Pi_i^t,a_i^t\right) &= \mu^{sur}_{a_i^t}\cdot\left(\varepsilon^{sur}_{u_i^{inv}} + \varepsilon^{sur}_{u_i^{end}}\right)\cdot(1-\tau^{sur})^{\overset{r}{N^t_{i\,end}}}
\end{aligned}
\tag{3.18}
$$

La définition de ce modèle par réseau bayésien dynamique étiqueté est illustrée en figure 3.6 et la valeur des paramètres utilisés est listée dans la table 3.4 pour l'espèce invasive et dans les tables 3.5 et 3.6 pour l'espèce endémique.

3.5.4 Modèle de diffusion et de propagation d'influence interne et externe

Dans un modèle de diffusion d'information au sein d'un réseau (comme un réseau social), la diffusion se fait majoritairement entre les individus liés directement entre eux au sein du réseau, à la manière d'un processus de contact classique. Néanmoins, il arrive que l'on veuille modéliser également l'influence externe des individus non liés au sein du réseau, qui ont une faible chance de transmettre l'information. Une modélisation de cette influence extérieure au sein d'un réseau social a été proposée [Gomez Rodriguez et al., 2010]. Dans un tel réseau social, les individus peuvent être liés entre eux classiquement, et peuvent se transmettre une information avec une certaine probabilité. Pour modéliser l'influence extérieure, on relie les nœuds qui ne sont pas reliés au sein de ce réseau par un lien d'une nature différente, dont la probabilité de transmission d'information est plus faible. Un tel réseau correspond donc à une clique dans laquelle deux types d'arêtes cohabitent : des arêtes "du graphe" correspondant au réseau social classique, et des arêtes "extérieures" ajoutées entre chaque nœud non relié par une arête du graphe. L'intérêt de ces approches est d'étudier la diffusion de l'information. Sa disparition ou son apparition spontanée n'est généralement pas étudiée dans ce cadre. Sous forme d'un RBD-E multicontact, cela se traduit sous la forme d'un réseau contenant deux types d'impulseurs : un type q_1 correspondant aux arêtes du graphe, avec une probabilité ρ_{q_1} de transmettre l'information, et un type q_2 correspondant aux arêtes extérieures, avec une faible probabilité ρ_{q_2} de transmettre l'information. Ces impulseurs n'ont d'effet que sur l'apparition, donc

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Apparition de l'espèce invasive	Spontanée	ε_{inv}^{app}	Apparition spontanée possible : colonisation extérieure
	Impulsée	ρ_{inv}^{app}	1 type d'impulseur pour l'espèce invasive : colonisation à partir d'une mare
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	μ	La covariable correspond à une action d'affaiblissement : éradication des espèces dans une mare
Survie de l'espèce invasive	Spontanée	ε_{inv}^{sur}	Disparition spontanée possible : extinction de l'espèce dans la mare
	Impulsée	0^*	Aucun effet de l'impulseur sur la survie
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	μ	La covariable correspond à une action d'affaiblissement : éradication des espèces dans une mare. L'effet est similaire sur l'apparition et sur la survie

TABLE 3.4 – Résumé des paramètres utilisés pour la dynamique de l'espèce invasive dans le modèle de gestion de réseaux écologiques spatiaux
*Paramètre non utilisé : utilisation de la valeur par défaut

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Apparition de l'espèce endémique	Spontanée	0^*	Pas d'apparition spontanée de l'espèce endémique
	Impulsée	ρ_{end}^{app}	1 type d'impulseur pour l'espèce endémique : colonisation à partir d'une mare
	Inhibée	τ_{end}^{app}	1 type d'inhibiteur pour l'espèce endémique : compétition de l'espèce invasive sur la survie et l'apparition. Au maximum 1 type de ce lien est actif, celui de l'espèce invasive lorsqu'elle est présente dans la même mare
	Effet de la covariable	μ	La covariable correspond à une action d'affaiblissement : éradication des espèces dans une mare.

TABLE 3.5 – Résumé des paramètres utilisés pour la dynamique d'apparition de l'espèce éndémique dans le modèle de gestion de réseaux écologiques spatiaux
*Paramètre non utilisé : utilisation de la valeur par défaut

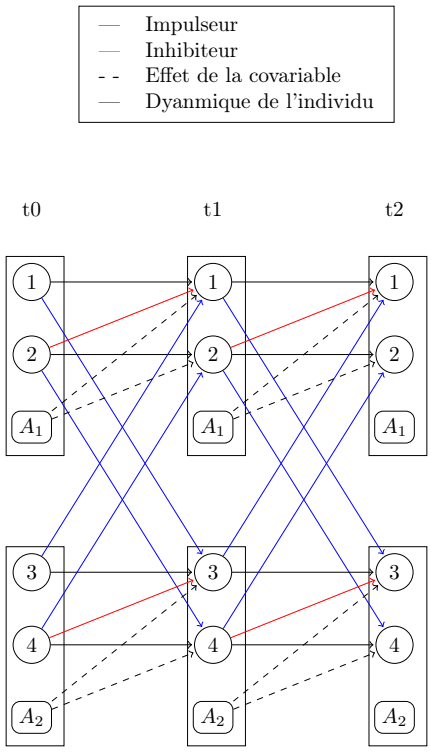


FIGURE 3.6 – Représentation des premiers pas de temps d'un modèle de gestion d'espèces invasives dans un réseau spatial à 2 mares. Pour une meilleure visibilité du phénomène, les espèces d'une même mare sont encadrées ensemble et une seule covariable est représentée pour les deux espèces de chaque mare. Les nœuds 1 et 3 désignent les espèces endémiques et les nœuds 2 et 4 les espèces invasives.

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Survie de l'espèce endémique	Spontanée	ε_{end}^{sur}	Disparition spontanée possible : extinction de l'espèce dans la mare
	Impulsée	0^*	Aucun effet de l'impulseur sur la survie
	Inhibée	τ_{end}^{sur}	1 type d'inhibiteur pour l'espèce endémique sur la survie et l'apparition : compétition de l'espèce invasive. Au maximum 1 type de ce lien est actif, celui de l'espèce invasive lorsqu'elle est présente dans la même mare
	Effet de la covariable	μ	La covariable correspond à une action d'affaiblissement : éradication des espèces dans une mare. L'effet est similaire sur l'apparition et sur la survie

TABLE 3.6 – Résumé des paramètres utilisés pour la dynamique de survie de l'espèce endémique dans le modèle de gestion de réseaux écologiques spatiaux
*Paramètre non utilisé : utilisation de la valeur par défaut

aucun effet sur la survie. La probabilité d'apparition de l'information chez un individu i est décrite par l'équation (3.19).

$$P\left(x_i^{t+1}=1|x_i^t=0,\Pi_i^t\right)=1-(1-\rho_{q_1}^{app})^{q_1 N_{i\,end}^t}\cdot(1-\rho_{q_2}^{app})^{q_2 N_{i\,end}^t} \tag{3.19}$$

Les différents paramètres utilisés dans un tel modèle sont résumés dans la table 3.7.

3.5.5 Dynamique d'espèces dans un réseau écologique

Un réseau écologique traduit les relations écologiques entre les espèces vivantes au sein d'un écosystème. Un réseau bayésien dynamique paramétré permet de modéliser la dynamique de ces espèces, c'est à dire leur présence ou leur absence au cours du temps, cette dynamique étant décrite par les différentes relations écologiques. La relation écologique la plus étudiée est la relation trophique entre deux espèces (relation proie/prédateur). Une telle relation est à double sens : une relation de type "proie" aidera au maintien de la présence de ses prédateurs, et une relation de type "prédateur" limite la possibilité de maintien de la présence de ses proies. La recolonisation de l'écosystème observé est décrite dans ce modèle comme indépendante du réseau, et l'écosystème considéré peut être, au cours du temps, transformé en une zone protégée, ce qui améliorera les chances de survie et de recolonisation des espèces. L'évolution de l'état des espèces se fait de la façon suivante :

— Une espèce absente de l'écosystème peut le recoloniser avec une probabilité ε

— Une espèce j peut avoir une influence positive (proie/facilitateur...) sur une espèce j. Chacune de ces influences a une probabilité ρ de réussir à influencer positivement la survie i.

— Une espèce j peut avoir une influence négative (prédateur/parasite/compétiteur...) sur une espèce j. Chacune de ces influences a une probabilité τ de réussir à influencer positivement la survie i.

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Apparition	Spontanée	0^*	L'apparition spontanée n'est pas modélisée
	Impulsée	ρ_{q_1} ρ_{q_2}	2 types d'impulseurs . Les arêtes q_1 décrivent les arêtes du graphe permettant de transmettre classiquement l'information. Les arêtes q_2 décrivent les arêtes depuis les nœuds non reliés dans le graphe, qui transmettent une information moindre.
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	1^*	Pas de covariable
Survie	Spontanée	1^*	Pas de disparition spontanée
	Impulsée	0^*	Aucun effet des impulseurs sur la survie
	Inhibée	0^*	Aucun inhibiteur
	Effet de la covariable	1^*	Pas de covariable

TABLE 3.7 – Résumé des paramètres utilisés dans le modèle de réseau d'interactions écologiques
*Paramètre non utilisé : utilisation de la valeur par défaut

— La survie d'une espèce présente dans l'écosystème correspond à la probabilité de réussite d'au moins une influence positive et d'échec de toutes les influences négatives, soit $P(survie(i)|positif(i), negatif(i)) = \left(1-(1-\rho)^{\#\{positif(i)\}}\right)\cdot(1-\tau)^{\#\{negatif(i)\}}$ ou $\#\{psitif(i)\}$ et $\#\{negatif(i)\}$ renseignent respectivement sur le nombre d'influences positives et négatives subies par i.

— Une espèce *basale* est une espèce n'ayant aucune influence positive. Cela peut être le cas pour des espèces en bas de la chaîne alimentaire comme les plantes, qui n'ont pas besoin de proies pour survivre. Une espèce basale ne s'éteindra que si au moins une de ses influences positives réussit, et survivra dans le cas contraire.

— Toutes ces probabilités sont multipliées par μ, paramètre diminuant les probabilités, lorsque la zone considérée n'est pas protégée. L'influence de la protection est la même sur l'apparition et sur la survie.

Un tel problème peut se modéliser dans le cadre des réseaux bayésiens dynamiques étiquetés. Pour cela, nous avons besoin d'un réseau contenant un type de lien d'impulsion et un type de lien d'inhibition. Chacun de ces liens n'a qu'un effet sur la survie. Les paramètres d'un tel modèle sont les probabilités de réussite de ces deux types de liens, un paramètre de probabilité d'apparition spontanée, et un paramètre pénalisant l'absence de protection, la protection ne pouvant être appliquée qu'à toutes les espèces en même temps à un pas de temps donné. Cette protection peut se modéliser par une covariable, qui s'interprétera en fait comme l'absence de protection de la zone. Lorsque la covariable est active, la zone n'est pas protégée et les chances de survie et de colonisation sont diminuées. On pourra également distinguer deux types de comportements inhérents différents : pour une espèce i, $u_i = 0$ si i est basale et $u_i = 1$ si i n'est pas basale. Les contraintes et les valeurs de paramètres sont résumées dans la table 3.8. Les probabilités de transition d'un individu i sont

Type de dynamique		Valeur du (des) paramètre(s)	Remarques
Apparition	Spontanée	ε^{app}	Apparition spontanée possible
	Impulsée	0^*	Aucun effet des impulseurs sur l'apparition
	Inhibée	0^*	Aucun effet des inhibiteurs sur l'apparition
	Effet de la covariable	μ	Si $A^t = 1$ (la zone n'est pas protégée) alors les probabilités d'apparition et de survie sont réduites pour tous les individus
Survie	Spontanée (espèce basale)	$\varepsilon_1^{sur} = 0^*$	Pas de disparition spontanée
	Spontanée (espèce basale)	$\varepsilon_0^{sur} = 1$	La survie ne dépend pas des impulseurs d'une espèce basale
	Impulsée	ρ^{sur}	1 type d'impulseur : les facilitateurs ou les proies permettent la survie
	Inhibée	τ^{sur}	1 type d'inhibiteur : les compétiteurs ou les prédateurs inhibent la survie
	Effet de la covariable	μ	La covariable A_i^t a le même effet sur l'apparition et sur la survie de i

TABLE 3.8 – Résumé des paramètres utilisés dans le modèle de réseau d'interactions écologiques
*Paramètre non utilisé : utilisation de la valeur par défaut

décrites par les équations (3.20).

$$
\begin{aligned}
&P\left(X_i^{t+1}=1|X_i^t=0,\Pi_i^t,a_i^t,u_i\right)=\mu_{a_i^t}^{app}\cdot\varepsilon^{app}\\
&P\left(X_i^{t+1}=1|X_i^t=1,\Pi_i^t,a_i^t,u_i\right)=\mu_{a_i^t}\cdot\left(\varepsilon_{u_i}^{sur}+(1-\varepsilon_{u_i}^{sur})\cdot\left(1-(1-\rho^{sur})^{\overset{q}{N_i^t}}\right)\right)\cdot(1-\tau^{sur})^{\overset{r}{N_i^t}}
\end{aligned}
\tag{3.20}
$$

Conclusion

Nous avons défini dans ce chapitre un modèle de réseau bayésien dynamique permettant de modéliser des phénomènes dans lesquels l'impact et la nature des interactions possibles entre les variables sont connus à l'avance. Ce modèle, appelé réseau bayésien étiqueté, considère des interactions pouvant être soit positives soit négatives. Chacun de ces types d'interaction peut avoir des forces différentes, le nombre de ces forces étant connu à l'avance. Chaque interaction est alors étiquetée selon sa nature et sa force. Un tel modèle contraint la forme des probabilités conditionnelles de chaque variable, ce qui permet de diminuer le nombre de paramètres, qui est alors indépendant de la structure du graphe associé à ce modèle. Cette classe de modèles est n'est adaptée qu'à certains problèmes, mais est applicable à de nombreux domaines. Son principal intérêt est la diminution du nombre de paramètres, ce qui facilite grandement l'estimation de leur valeur.

Cette classe de modèles de réseau bayésien à pour ambition de pouvoir utiliser les différents algorithmes déjà utilisés dans le cadre des réseaux bayésiens classiques, tant pour l'estimation des paramètres que pour l'apprentissage de la structure. La principale motivation pour définir ce modèle est qu'il définit un cadre théorique pour l'apprentissage de la structure de réseaux dans des phénomènes modélisables par RBD-E. C'est notamment le cas de la dynamique d'espèces dans un réseau écologique, dont la modélisation par RBD-E permet de développer une méthode d'apprentissage de la structure du réseau écologique en alternant des phases d'estimation des paramètres à structure fixée et d'apprentissage de structure à paramètres fixés. Dans la section suivante, nous allons présenter une méthode permettant d'apprendre la structure d'un réseau bayésien étiqueté.

Chapitre 4

Apprentissage de la structure d'un réseau bayésien étiqueté

Nous avons vu dans le chapitre 2 les deux familles de méthodes les plus courantes pour apprendre la structure d'un réseau bayésien : l'apprentissage par tests de probabilités conditionnelles ou par optimisation d'un score. Ces algorithmes sont-ils compatibles avec le modèle de réseau bayésien étiqueté présenté dans le chapitre 3 ? Pour rappel, les différences entre un réseau bayésien étiqueté et un réseau bayésien classique sont les suivantes :

— Les arcs du graphe associé au réseau bayésien portent une étiquette, renseignant sur le type d'interaction (positive ou négative) et sa force.

— La probabilité associée à une variable ne dépend que des étiquettes des arcs que la variable reçoit et du nombre de ses parents présents.

— Les probabilités associées à chaque variable sont fonction d'un petit nombre de paramètres (ce nombre dépend du nombre d'étiquettes dans le graphe) partagés par l'ensemble du réseau.

Quelles méthodes peut-on utiliser si l'on veut apprendre la structure d'un tel réseau ? La modélisation d'un réseau bayésien dynamique change la manière dont les probabilités de chaque variable aléatoire sont calculées, mais les indépendances conditionnelles entre les variables ne sont pas affectées en elles-même. Les algorithmes d'apprentissage de structure par tests d'indépendances conditionnelles pourraient être utilisés. Cependant, la structure étiquetée du réseau ne peut pas être apprise par cette méthode, et il serait alors nécessaire d'ajouter une étape supplémentaire d'étiquetage du graphe. Cela complexifiait alors cette méthode. Les méthodes par score pourraient alors convenir, car les probabilités de chaque variable aléatoire constituant un score tiennent compte des étiquettes des parents de chaque nœud. Cependant, si dans un modèle de réseau bayésien classique, la vraisemblance d'un réseau a une formule explicite et son optimisation est facile, en est-il de même pour un réseau bayésien étiqueté ? La motivation principale cette thèse concerne l'apprentissage de réseau écologique à partir de données dynamiques, nous ne nous intéresserons dans ce chapitre qu'au cas dynamique. Nous nous intéresserons ici à la manière d'apprendre la structure d'un réseau bayésien dynamique étiqueté. Nous verrons dans un premier temps comment exprimer la vraisemblance d'un tel modèle afin d'établir un score efficace. Dans un deuxième temps, nous proposerons un algorithme permettant d'apprendre une structure de graphe maximisant cette vraisemblance, en décrivant plus particulièrement une étape de cet algorithme consistant à résoudre un programme linéaire en

nombres entiers. Nous nous intéresserons ensuite à une manière d'inclure de la connaissance experte dans le processus d'apprentissage en ajoutant des aprioris sur certains arcs. Nous illustrons cela par un exemple écologique de connaissances expertes : la connaissance des niveaux trophiques des espèces du réseau. Nous évaluerons ensuite cette méthode en l'appliquant à des données simulées, puis à des données réelles.

4.1 Vraisemblance d'un RBDE

Voici pour rappel les formules des probabilités de transition d'une variable X_i^t vers X_i^{t+1} sachant les variables aléatoires Π_i^t associées aux parents π_i de i de chaque étiquette, l'action a_i^t sur le phénomène i au temps t, et le comportement u_i inhérent à i pour l'apparition et la survie spontanée. Notons $\overset{l}{N}_i^t$ la quantité $\sum_{j \in \pi_i^l} X_j^t$, c'est à dire le nombre de parents de i d'étiquette l présents au temps t.

- $\mu_{a_i^t}^{app}$ et $\mu_{a_i^t}^{sur}$ sont les impacts de l'action a_i^t sur les probabilités d'apparition et de survie de la variable X_i^t.
- $\varepsilon_{u_i}^{app}$ et $\varepsilon_{u_i}^{sur}$ sont les probabilités d'apparition et de survie spontanée associées à X_i^t.
- ρ_q^{app} et ρ_q^{sur} sont les probabilités de réussite d'une action d'impulsion d'étiquette q pour l'apparition et pour la survie
- τ_r^{app} et τ_r^{sur} sont les probabilités de réussite d'une action d'inhibition d'étiquette r pour l'apparition et pour la survie

Les probabilités d'apparition et de survie d'une variable X_i^t sont indiquées dans l'équation 4.1.

$$
\begin{aligned}
P\left(X_i^{t+1}=1|X_i^t=0,\Pi_i^t,a_i^t,u_i\right) &= \mu_{a_i^t}^{app}\cdot\left(\varepsilon_{u_i}^{app}+(1-\varepsilon_{u_i}^{app})\cdot\left(1-\textstyle\prod_{q=q_1}^{q_{max}}(1-\rho_q^{app})^{\overset{q}{N}_i^t}\right)\right)\cdot\textstyle\prod_{r=r_1}^{r_{max}}(1-\tau^{app})^{\overset{r}{N}_i^t}\\
P\left(X_i^{t+1}=1|X_i^t=1,\Pi_i^t,u_i\right) &= \mu_{a_i^t}^{sur}\cdot\left(\varepsilon_{u_i}^{sur}+(1-\varepsilon_{u_i}^{sur})\cdot\left(1-\textstyle\prod_{q=q_1}^{q_{max}}(1-\rho_q^{sur})^{\overset{q}{N}_i^t}\right)\right)\cdot\textstyle\prod_{r=r_1}^{r_{max}}(1-\tau_s^{sur})^{\overset{r}{N}_i^t}
\end{aligned}
\tag{4.1}
$$

Pour un jeu de données représentant n phénomènes binaires observés sur T pas de temps, la vraisemblance de ces données pour un modèle de réseau bayésien étiqueté donné est la probabilité que ces observations soient générées par ce modèle de réseau bayésien dynamique étiqueté. C'est à dire que ces observations constituent la réalisation d'un ensemble de variables aléatoires X_i^t générées à partir des probabilités indiquées en équation 4.1. Dans un tel modèle, le nombre de paramètres dans θ ne dépend pas directement de la structure du graphe $\mathcal{G}$, mais du nombre d'étiquettes, de classes d'individus et de valeurs possibles pour les covariables associées aux individus. Ainsi, contrairement à un modèle de réseau bayésien classique, l'ajout d'un arc dans le graphe n'ajoute pas de paramètre, et n'augmente pas nécessairement la vraisemblance. Ainsi, maximiser la vraisemblance seule est un score valide dans le cadre des réseaux bayésiens étiquetés. Il est possible d'exprimer cette vraisemblance de plusieurs manières, mais il est intéressant de pouvoir exprimer cette vraisemblance de manière indépendante pour chaque individu, afin de conserver les propriétés intéressantes d'un score. Ainsi, le logarithme de cette vraisemblance, plus simple à exprimer et à manipuler que la vraisemblance, prendrait la forme de l'équation 4.2. Dans ce cas, la quantité $score(\mathbb{X}_i|\pi_i,a_i,u_i)$ avec $\mathbb{X}_i=\{X_i^1,...,X_i^T\}$ et $a_i=\{a_i^1,...,a_i^T\}$ est un score local de contribution non pénalisée à la

vraisemblance d'un individu i.

$$\log P\left(X_1^1, ..., X_n^1, ..., X_1^T, ..., X_n^T | \mathcal{G}, \theta\right) = \sum_{i=1}^{n} score(x_i^{t+1} | \Pi_i^t, x_i^t, a_i^t, u_i) \tag{4.2}$$

Voyons comment il est possible d'exprimer cette quantité. Afin d'avoir une expression lisible de cette vraisemblance, introduisons quelques notations. Notons tout d'abord le logarithme de la probabilité d'apparition d'une espèce i à un temps $t+1$ présenté en équation 3.6 de la façon suivante :

$$L^{app}(X_i^{t+1}) = logP\left(X_i^{t+1} = 1 | X_i^t = 0, \Pi_i^t, a_i^t, u_i\right)$$

et, de la même manière, notons la probabilité de survie comme suit :

$$L^{sur}(X_i^{t+1}) = logP\left(X_i^{t+1} = 1 | X_i^t = 1, \Pi_i^t, u_i\right).$$

Notons également les quantités suivantes :

- $D_{imp} = (d_{q_1}, d_{q_2}, ..., d_{q_{max}} \in \{0, 1, ..., n\})$ un vecteur de nombres entiers où d_q désigne un nombre d'impulseurs d'étiquette q présents.
- $D_{inh} = (d_{r_1}, d_{r_2}, ..., d_{r_{max}} \in \{0, 1, ..., n\})$ un ensemble de nombres entiers où d_r désigne un nombre d'inhibiteurs d'étiquette r présents.
- $\underset{D_{imp},D_{inh}}{R_i^{t,u}} = 1$ si, pour un individu i, son comportement inhérent sur sa présence spontanée $u_i = u$, et au temps t, i a exactement : d_{q_1} impulseurs d'étiquette 1 présents , d_{q_2} impulseurs d'étiquette 2 présents,...,d_{r_1} inhibiteurs d'étiquette 1 présents, d_{r_2} inhibiteurs d'étiquette 2 présents... ($\overset{q_1}{N_i^t} = d_{q_1}$, $\overset{q_2}{N_i^t} = d_{q_2}$,...,$\overset{r_1}{N_i^t} = d_{r_1}$,$\overset{r_2}{N_i^t} = d_{r_2}$).

$$\begin{aligned} score(i) = \sum_{t=1}^{T} & (1 - x_i^t) \cdot \sum_{\substack{D_{imp},D_{inh} \\ tq:\ d_{q_1}+...+d_{q_{max}}+d_{r_1},...,d_{r_{max}}<n}} \sum_{u=u_1}^{u_{max}} \underset{D_{imp},D_{inh}}{R_i^{t,u}} \cdot L^{app}(X_i^{t+1}) \\ & + x_i^t \cdot \sum_{\substack{D_{imp},D_{inh} \\ tq:\ d_{q_1}+...+d_{q_{max}}+d_{r_1},...,d_{r_{max}}<n}} \sum_{u=u_1}^{u_{max}} \underset{D_{imp},D_{inh}}{R_i^{t,u}} \cdot L^{sur}(X_i^{t+1}) \end{aligned} \tag{4.3}$$

Ce score s'écrit donc comme une somme sur tous les pas de temps de toutes les probabilités des combinaisons d'ensembles possibles de nombre de voisins présents de chaque étiquette. Une seule de ces combinaisons étant vraie à un pas de temps t, ces probabilités sont multipliées par une quantité $\underset{D_{imp},D_{inh}}{R_i^{t,u}}$ valant 1 uniquement si le nombre de parents présents de chaque étiquette correspondent à D_{imp} et D_{inh}. Seuls les ensembles D_{imp} et D_{inh} tels que $d_{q_1} + ... + d_{q_{max}} + d_{r_1}, ..., d_{r_{max}} < n$ sont pris en compte, car tout individu ne peut avoir plus de n autres individus pour parents. Si l'on connaît ou que l'on impose le nombre maximal k de parents par individu, l'expression de la vraisemblance peut être simplifiée en ne prenant pas en compte les ensembles D_{imp} et D_{inh} dont la somme des éléments dépassent ce nombre k.

4.2 Apprentissage dans un RBDE

Apprendre un réseau bayésien dynamique étiqueté consiste à apprendre à la fois sa structure $\mathcal{G}_{\rightarrow}$ et ses paramètres $\theta_{\rightarrow}$ à partir de données D correspondant à un ensemble de trajectoires,

$D = \{(X_i^t)\} \forall i = 1, \ldots, n; t = 1, \ldots, T$ et à partir des valeurs des covariables a_i^t et des comportements u_i inhérents à chaque phénomène i. Il est à noter que la structure $\mathcal{G}_{\rightarrow}$ est, dans ce cadre, étiquetée, c'est à dire que les arcs de ce graphe comportent une étiquette renseignant sur leur nature (inhibition ou impulsion) et leur force. Contrairement à des méthodes d'apprentissage de réseaux bayésiens classiques, le score local d'un individu n'est pas compliqué à calculer en connaissant la structure du graphe associé, mais lorsque cette structure n'est pas intégralement connue, tout ajout d'arc demande de recalculer entièrement ce score. En effet, l'ajout d'un arc dans un réseau bayésien étiqueté modifie la valeur d'un exposant dans une probabilité. Les algorithmes d'apprentissage de réseaux bayésiens classiques n'ont pas cette difficulté. Nous présentons ici une procédure dite de *Restauration-Estimation* permettant d'apprendre à la fois les paramètres et la structure du modèle, afin de trouver un modèle maximisant sa vraisemblance.

Le problème d'apprentissage revient à trouver $\hat{\mathcal{G}}_{\rightarrow}$ et $\hat{\theta}_{\rightarrow}$ qui maximisent conjointement $\log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}}(D)\right)$.

4.2.1 Algorithme d'apprentissage de RBDE

Considérons une procédure itérative générale définie par l'algorithme 5 pour obtenir un maximum local de $\log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}}(D)\right)$.

$s \leftarrow 0$;
Choisir un graphe $\mathcal{G}_{\rightarrow}^{(0)}$ arbitraire ;
répéter
  Étape E (Estimation) : $\theta_{\rightarrow}^{(s)} \leftarrow \arg\sup_{\theta\rightarrow} \log\left(P_{\mathcal{G}_{\rightarrow}^{(s)},\theta_{\rightarrow}}(D)\right)$;
  Étape R (Restauration) : $\mathcal{G}_{\rightarrow}^{(s+1)} \leftarrow \arg\max_{\mathcal{G}_{\rightarrow}} \log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}^{(s)}}(D)\right)$;
  $s \leftarrow s + 1$;
jusqu'à *Convergence*;

Algorithme 5 : Procédure d'apprentissage par restauration-estimation de la structure d'un graphe par fonction de score

Cet algorithme est très général. Les deux étapes doivent être spécifiées pour un problème d'apprentissage de RBD étiqueté donné pour pouvoir implémenter cet algorithme. Toutefois, pour n'importe quelle implémentation de cet algorithme, la proposition suivante est vérifiée :

Proposition 1. *La procédure d'apprentissage d'un RBD étiqueté converge vers un maximum local de* $\log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}}(D)\right)$.

Preuve :

— Les étapes E et R augmentent conjointement la log-vraisemblance :
  — L'étape E se construit de telle manière à ce que les paramètres $\theta_{\rightarrow}^{(s)}$ maximisent $\log\left(P_{\mathcal{G}_{\rightarrow}^{(s)},\theta_{\rightarrow}}(D)\right)$ pour un réseau $\mathcal{G}_{\rightarrow}^{(s)}$. Il n'existe pas de jeu de paramètres $\theta_{\rightarrow}^{(x)} \neq \theta_{\rightarrow}^{(s)}$ tel que $\log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}^{(s)}}(D)\right) > \log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}^{(s)}}(D)\right)$.
  — L'étape R se construit de telle manière qu'un graphe $\mathcal{G}_{\rightarrow}^{(s+1)}$ ait une vraisemblance $\mathcal{G}_{\rightarrow}^{(s+1)} \geq \log\left(P_{\mathcal{G}_{\rightarrow},\theta_{\rightarrow}^{(s)}}(D)\right)$
  — Ainsi,

$$\log\left(P_{\mathcal{G}_{\rightarrow}^{(s+1)},\theta\rightarrow(s+1)}(D)\right) \geq \log\left(P_{\mathcal{G}_{\rightarrow}^{(s)},\theta_{\rightarrow}^{(s)}}(D)\right), \forall s$$

— Si il existe un k tel que $\mathcal{G}_{\rightarrow}^{(s+1)} = \mathcal{G}_{\rightarrow}^{(s)}$, alors $\theta_{\rightarrow}^{(s+1)} = \theta_{\rightarrow}^{(s)}$. Dans ce cas, l'algorithme a convergé.

— Si $\mathcal{G}_{\rightarrow}^{(s+1)} \neq \mathcal{G}_{\rightarrow}^{(s)}$, $\mathcal{G}_{\rightarrow}^{(s)}$ ne peut pas à nouveau être une solution d'une itération suivant $s' > s$. Puisque l'espace des graphes possibles est fini, il existe forcément un s tel que on a $\mathcal{G}_{\rightarrow}^{(s+1)} = \mathcal{G}_{\rightarrow}^{(s)}$.

4.2.2 Étape d'estimation

L'étape E consiste à estimer la valeur des paramètres à partir d'une structure de graphe connue. Contrairement au cas classique, le nombre de paramètres à estimer est connu à l'avance. Soit, pour un réseau bayésien dynamique étiqueté, a_{max} le nombre d'états possibles de la covariable, u_{max} le nombre de comportements inhérents possibles, q_{max} le nombre d'étiquettes d'impulseurs et r_{max} le nombre d'étiquettes d'inhibiteurs, ce réseau bayésien étiqueté a au plus $a_{max} + u_{max} + q_{max} + r_{max}$ paramètres inconnus, ou $2 * (a_{max} + u_{max} + q_{max} + r_{max})$ s'il s'agit d'un réseau bayésien dynamique étiqueté (chacun ayant un effet différent sur la survie et sur l'apparition). De ce fait, l'étape d'estimation se fait en recherchant les valeurs des paramètres maximisant la vraisemblance du réseau. Les méthodes classiques de maximum de vraisemblance peuvent s'utiliser facilement, car il y a un nombre restreint de paramètres. Dans le reste de ce manuscrit, nous utilisons la méthode des points intérieurs dans le cadre de la programmation non linéaire [Byrd et al., 1999] pour cette étape.

4.2.3 Étape de restauration

L'étape R consiste à trouver la meilleure structure de réseau à partir d'un ensemble de paramètres connus. La vraisemblance ne possédant pas de composante limitant le nombre de paramètres, cela limite les possibilités d'algorithmes (l'algorithme Branch and Bound, par exemple, exploite une composante de pénalité). Une difficulté lors de l'apprentissage d'un réseau bayésien étiqueté par un algorithme basé sur un score vient du fait que l'espace des graphes à explorer est plus vaste : en plus des arêtes, il faut apprendre leur étiquette. On devra donc apprendre, pour chaque nœud, un ensemble de parents ainsi que leurs étiquettes. Dans le cadre des réseaux bayésiens dynamiques étiquetés tels qu'ils ont été décrits, tout arc est dirigé d'un pas de temps t vers un pas de temps $t+1$. Il n'y a donc pas d'arc synchrone, ce qui permet d'ignorer les étapes de recherche de boucles, comme décrit dans l'algorithme de [Dojer, 2006]. Cependant, cet algorithme ne peut pas être utilisé en tant que tel car toutes les conditions de cet algorithme ne sont pas respectées dans le cadre des réseaux bayésiens dynamiques étiquetés. En effet, si la vraisemblance peut être considérée comme additive (décomposable en scores locaux pour chaque nœud), il n'y a pas de paramètres de pénalisation permettant la décomposabilité de ce score. La vraisemblance, présentée en section 4.1, peut prendre la forme d'une somme de variables binaires, exprimée à partir d'autres variables binaires. Une telle configuration est idéale pour utiliser des méthodes de programmation linéaire en nombre entiers, car toute variable binaire peut être exprimée dans ce cadre par des contraintes linéaires en fonction d'autres variables binaires [Williams et al., 2009]. Il est possible de décomposer cette vraisemblance en n problèmes indépendants, chacun cherchant à maximiser le score d'un phénomène i présenté dans l'équation 4.3. Les variables R de cette équation sont des variables binaires définies par des nombres de parents présents à un pas de temps t, c'est à dire par une série de présence ou d'absence d'arcs dans la structure graphique $\mathcal{G}$. Elles peuvent donc être exprimées par des contraintes linéaires sur ces variables. Un algorithme d'apprentissage utilisant la programmation linéaire en nombres entiers est faisable dans le cadre des réseaux bayésiens dynamiques étiquetés. Voyons de quelle manière il est possible d'exprimer ce problème.

4.3 Expression de l'étape de restauration comme un programme linéaire en nombres entiers

Apprendre une structure de graphe à paramètres connus consiste à trouver la structure de graphe maximisant la vraisemblance. Cela équivaut à chercher pour chaque phénomène i l'ensemble π_i de ses parents maximisant le score local de i exprimé dans l'équation 4.3. Dans cette expression, les variables R peuvent être définies à partir des arêtes du graphe $\mathcal{G}$ à l'aide de contraintes linéaires. Pour cela, il faut ajouter un ensemble de variables binaires intermédiaires. Ce problème revient donc à chercher les valeurs de ces variables binaires, définie par contrainte à partir de variables renseignant des parents d'un nœud i maximisant la vraisemblance. La résolution d'un programme linéaire en nombre entiers (PLNE) 0/1 permet cette maximisation. Cette section décrit les variables et les contraintes linéaires permettant de résoudre ce programme linéaire.

4.3.1 Définition du PLNE

Pour définir un programme linéaire cohérent, définissons des variables intermédiaires définies par des contraintes linéaires. Pour établir un tel problème, nous connaissons l'ensemble des observations X_i^t. Chaque interaction $g_{i,j}^l$ d'une variable i vers une variable j d'une étiquette l parmi l'ensemble de toutes les étiquettes d'impulsion comme d'inhibition possibles constitue une variable de ce problème de programmation linéaire. Chaque variable X_i est associée à un type de comportement inhérent $u_i \in \{u_1, ..., u_{max}\}$. Dans le cadre de la programmation linéaire en nombre entiers, nous considérons ici un nombre u_{max} de variables binaires définies comme $U_i = 1$ si le comportement inhérent à i a pour valeur u_i, $U_i = 0$ sinon. Afin de faciliter l'expression de ces variables, nous définissons à l'avance un nombre k indiquant le nombre de parents maximal que peut avoir un nœud du graphe. Par défaut, il est possible d'établir $k = n$, mais une telle valeur pour k peut mener à un problème de programmation linéaire très complexe. Les variables supplémentaires du problème de programmation linéaire pour une variable X_i donnée se décrivent de la façon suivante :

Borne inférieure du nombre de parents présents d'une étiquette l donnée : L'ensemble des variables $\underset{d,l}{M_i^t}$ est défini pour toute étiquette l, d'impulsion comme d'inhibition, ainsi que pour $t = \{1..T\}, d = \{0..k\}, l \in \{q_1, .., q_{max}, r_1, ..., r_{max}\}$, où k est le nombre de parents maximum, comme suit : $\underset{d,l}{M_i^t} = 1$ ssi $N_i^t \geq d$. $\underset{d,l}{M_i^t}$ vérifie si l'espèce i a au moins d parents étiquetés de type l présents au temps t.

Ces variables binaires sont définies par les contraintes linéaires définies en (4.4).

Borne supérieure du nombre de parents présents d'une étiquette l donnée : De la même façon, l'ensemble de variables binaires $\underset{d,l}{\nu_i^t}$ se définit pour toute étiquette l, d'implusion comme d'inhibition, ainsi que pour $t = \{1, ..., T\}, d = \{0, ..., k\}, l \in \{q_1, .., q_{max}, r_1, ..., r_{max}\}$ comme suit : $\underset{d,l}{\nu_i^t} = 1$ ssi $\underset{l}{N_i^t} \leq d$. $\underset{d,l}{\nu_i^t}$ vérifie si l'espèce i a au plus d parents étiquetés de type l présents au temps t. Ces variables binaires sont définies par les contraintes linéaires définies en (4.5).

Test du nombre de parents présents d'une étiquette l donnée : L'ensemble des variables binaires 0-1 $\underset{d,l}{\Lambda_i^t}$ se définit pour toute étiquette l, d'implusion comme d'inhibition, ainsi que pour $l \in \{q_1, .., q_{max}, r_1, ..., r_{max}\}, i = \{1, ..., n\}, t = \{1, ..., T\}, d = \{0, ..., k\}$ comme suit :

$\underset{d,l}{\Lambda_i^t} = 1$ ssi $\overset{l}{N_i^t} = d$. $\underset{d,l}{\Lambda_i^t}$ vérifie si l'espèce i a exactement d parents étiquetés de type l au temps t. Ces variables binaires sont définies par les contraintes linéaires définies en (4.6)

Test d'initialisation et du nombre de parents présents de chaque étiquette : Pour rappel, nous avons défini $\underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} = 1$ ssi au temps t, l' individu i a exactement : d_{q_1} impulseurs d'étiquette q_1 présents, d_{q_2} impulseurs d'étiquette q_2 présents,...,d_{r_1} inhibiteurs d'étiquette r_1 présents, d_{r_2} inhibiteurs d'étiquette r_2 présents et a un comportement $u_i = u$. Ainsi, $\underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} = 1$ ssi : $\underset{d_{q_1},q_1}{\Lambda_i^t} = 1, ..., \underset{d_{q_{max}},q_{max}}{\Lambda_i^t} = 1, \underset{d_{r_1},r_1}{\Lambda_i^t} = 1, ..., \underset{d_{r_{max}},r_{max}}{\Lambda_i^t} = 1, u_i = u$ Enfin, ces variables sont définies pour tout $t \in \{1, .., T\}$ $d_{imp_{1,...,max}}, d_{inh_{1,...,max}} \in \{0, .., k\}$ où $d_{q_1} + ... + d_{q_{max}} + d_{r_1} + ... + d_{r_{max}} <= k$ par l'ensemble des contraintes linéaires dans (4.7).

$$\begin{array}{l} \underset{d,l}{M_i^t} \cdot (d+1) - \sum_{j=1}^{n} \left(g_{ji}^l \cdot x_j^t \right) \leq 1 \\ \underset{d,l}{M_i^t} \cdot (k+1-d) - \sum_{j=1}^{n} \left(g_{ji}^l \cdot x_j^t \right) > -d \end{array} \tag{4.4}$$

$$\begin{array}{l} \underset{d,l}{\nu_i^t} \cdot (k+1-d) + \sum_{j=1}^{n} \left(g_{ji}^l \cdot x_j^t \right) \leq k+1 \\ \underset{d,l}{\nu_i^t} \cdot (d+1) + \sum_{j=1}^{n} \left(g_{ji}^l \cdot x_j^t \right) > d \end{array} \tag{4.5}$$

$$\begin{array}{l} \underset{d,l}{\Lambda_i^t} - \underset{d,l}{M_i^t} \leq 0 \\ \underset{d,l}{\Lambda_i^t} - \underset{d,l}{\nu_i^t} \leq 0 \\ \underset{d,l}{\Lambda_i^t} - \underset{d,l}{M_i^t} - \underset{d,l}{\nu_i^t} \geq -1 \end{array} \tag{4.6}$$

$$\begin{array}{rcl} \underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} & \leq & U_i \\ \underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} & \leq & \underset{d_{imp_1},imp_1}{\Lambda_i^t} \\ ... & & \\ \underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} & \leq & \underset{d_{imp_{max}},imp_{max}}{\Lambda_i^t} \\ \underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} & \leq & \underset{d_{inh_1},inh_1}{\Lambda_i^t} \\ ... & & \\ \underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} & \leq & \underset{d_{inh_{max}},inh_{max}}{\Lambda_i^t} \\ \underset{D_{imp},D_{inh}}{\overset{u}{R_i^t}} & \geq & \underset{d_{imp_1},imp_1}{\Lambda_i^t} + ... + \underset{d_{imp_{max}},imp_{max}}{\Lambda_i^t} \\ & + & \underset{d_{inh_1},inh_1}{\Lambda_i^t} + ... + \underset{d_{inh_{max}},inh_{max}}{\Lambda_i^t} \\ & + & U_i - imp_{max} - inh_{max} \end{array} \tag{4.7}$$

Ce problème de programmation linéaire en nombres entiers peut être utilisé pour trouver les voisins G_{ji} d'un individu i qui maximise le score de i. Maximiser ce score consiste à maximiser une fonction

objectif, définie comme n vecteurs de coefficients sur l'ensemble des variables binaires du programme linéaire. Les coefficients associés à toute variable $\underset{D_{imp},D_{inh}}{R_i^{t+1}}$, $\forall t \in \{1,..,T\}$ $d_{imp_{1,...,max}}, d_{inh_{1,...,max}} \in \{0,..,k\}$ où $d_{q_1} + ... + d_{q_{max}} + d_{r_1} + ... + d_{r_{max}} <= k$ sont indiqués dans l'équation 4.8. Seules les variables $\underset{D_{imp},D_{inh}}{R_i^{t+1}}$ sont directement utiles pour le calcul de la fonction objectif. Les autres variables du programme servent uniquement à définir les variables $\underset{D_{imp},D_{inh}}{R_i^{t+1}}$ par des contraintes linéaires, le coefficient associé à ces variables vaut donc 0.

$$\begin{array}{lllll} \underset{D_{imp},D_{inh}}{R_i^{t+1}} & \cdot & (1-X_i^t) & \cdot & \mu_{a_i^t}^{app} \cdot \left(\varepsilon_u^{app} + (1-\varepsilon_u^{app}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1-\rho_q^{app})^{d_{imp_q}}\right)\right) \cdot \prod_{r=r_1}^{r_{max}} (1-\tau^{app})^{d_{inh_r}} \\ + & & X_i^t & \cdot & \mu_{a_i^t}^{sur} \cdot \left(\varepsilon_u^{sur} + (1-\varepsilon_u^{sur}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1-\rho_q^{sur})^{d_{imp_q}}\right)\right) \cdot \prod_{r=r_1}^{r_{max}} (1-\tau_s^{sur})^{d_{inh_r}} \end{array} \tag{4.8}$$

Ainsi, si l'on note x le vecteur définissant la valeur de ces variables binaires, A une matrice définissant les contraintes linéaires indiquées dans cette section sous forme de coefficient sur l'ensemble des variables binaires, B un vecteur contenant les termes fixes des inégalités définissant ces contraintes et C le vecteur contenant les coefficients sur l'ensemble des variables du programme définissant la fonction objectif, l'étape de restauration revient à résoudre l'équation :

$$min(C^T x)$$

en respectant la contrainte :

$$Ax \leq B$$

4.3.2 Nombre de variables et de contraintes

Pour connaître la complexité de l'étape de restauration de l'algorithme, le facteur déterminant est le nombre de variables et de contraintes dans le problème de programmation linéaire en nombres entiers. Soit un réseau bayésien dynamique étiqueté avec n variables, T pas de temps, q_{max} impulseurs, r_{max} inhibiteurs et chaque nœud a k parents au maximum. Définissons la quantité $L = q_{max} + r_{max}$. Donnons le nombre de variables type par type :

- Une variable g_{ij}^l vaut 1 ssi il existe un arc d'étiquette l de i vers j. Il y en a une par étiquette et par paire d'individus, soit $n \times n \times L$.
- Une variable U_{u_i} vaut 1 ssi i a un comportement inhérent de type u_i. Si $u_i \in \{u_1, ..., u_{max}\}$, il y a u_{max} variables de ce type pour chaque individu i, soit $n \times u_{max}$ au total.
- Une variable $\underset{d,l}{M_i^t}$ vaut 1 ssi i a au plus d parents d'étiquette l au temps t. Il y en a une par individu, par pas de temps, par nombre de parents et par étiquette, soit $n \times T \times L \times k$.
- Une variable $\underset{d,l}{\nu_i^t}$ vaut 1 ssi i a au moins d parents d'étiquette l au temps t. Il y en a une par individu, par pas de temps, par nombre de parents et par étiquette, soit $n \times T \times L \times k$.
- Une variable $\underset{d,l}{\Lambda_i^t}$ vaut 1 ssi i a exactement d parents d'étiquette l au temps t. Il y en a une par individu, par pas de temps, par nombre de parents et par étiquette, soit $n \times T \times L \times k$.

— Une variable $\underset{D_{imp},D_{inh}}{R_i^{t\,u}}$, où $D_{imp} = \{d_{q_1}, ..., d_{q_{max}}\}$ et $D_{inh} = \{d_{r_1}, ..., d_{r_{max}}\}$ vaut 1 ssi i a, au temps t, exactement d_{q_1} impulseurs d'étiquette q_1,..., $d_{q_{max}}$ impulseurs d'étiquette q_{max}, d_{r_1} inhibiteurs d'étiquette r_1,..., $d_{r_{max}}$ inhibiteurs d'étiquette r_{max}. Il y en a une par variable, par pas de temps, et pour chaque vecteur D_{imp} et D_{inh} tel que la somme de leurs éléments ne dépasse pas k. Ce nombre est un peu plus complexe à calculer et fait intervenir du calcul combinatoire.

Voyons de quelle manière nous pouvons calculer le nombre décrit ci-dessus, c'est à dire le nombre de vecteurs possibles D_{imp} et D_{inh} tels que la somme des éléments de ces deux vecteurs ne dépasse pas k. Soit le vecteur D défini comme suit : $D = D_{imp} \cup D_{inh}$, c'est à dire le vecteur d'un nombre pour chaque étiquette différente, impulseurs comme inhibiteurs. Nous cherchons le nombre de vecteurs D possibles tels que la somme de tous les éléments de D est inférieure à k. Notons ce nombre Ω_D.

Définition. *Pour cette définition, nous allons utiliser des notations courantes du domaine de la combinatoire. Ainsi, les quantités n et k présentées ici ne correspondent pas nécessairement aux notations n et k utilisée dans le reste de ce manuscrit.*

Une combinaison avec répétition *est le résultat d'un tirage de k objets parmi n objets discernables, dont chacun peut être sélectionné plusieurs fois. L'ordre de ce groupement n'a pas d'importance. Par exemple, si l'on s'intéresse au nombre de "pile" et de "face" obtenus après dix lancers d'une pièce équilibrée, son résultat est une combinaison avec répétition de 10 objets parmi 2. Le nombre de combinaisons avec répétition de k objets parmi n se note* Γ_n^k *ou* $\left(\!\binom{n}{k}\!\right)$*. Cette quantité est égale à une combinaison sans remise de k éléments parmi* $n + k - 1$ *objets notée* C_{n+k-1}^k *ou* $\binom{n+k-1}{k}$*. On peut la définir comme suit :*

$$\Gamma_n^k = \left(\!\binom{n}{k}\!\right) = C_{n+k-1}^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Dans le programme linéaire en nombres entiers, chaque vecteur D est le résultat d'un tirage de 0 à k parents, chacun de ces éléments ayant une étiquette possible parmi $q_{max} + r_{max}$. Le nombre de vecteurs D possibles tels que la somme des éléments de D vaut exactement k est équivalent à une combinaison avec remise de k éléments parmi L. Ce nombre n'est pas équivalent au nombre de vecteurs D possibles, car tous les vecteurs D tels que la somme de ses éléments est inférieure à k ne sont pas comptabilisés. Le nombre de vecteurs D respectant cette définition peut donc s'écrire de la façon suivante :

$$\Omega_D = \sum_{h=0}^{k} \Gamma_L^h$$

Ce problème de programmation linéaire en nombres entiers peut être décomposé en n problèmes contenant chacun $n \cdot L + u_{max} + T\left(L \cdot k + \sum_{h=0}^{k} \Gamma_L^h\right)$ variables. Ce nombre augmente donc de manière factorielle avec le nombre d'étiquettes et le nombre maximal k de parents.

4.4 Modèle stochastique à blocs pour l'apprentissage de RBDE

4.4.1 Intégration dans un RBDE

L'algorithme d'apprentissage actuel considère que la présence ou l'absence de chaque arc est équiprobable. Seul l'état des variables au pas de temps précédent permet d'apprendre ces arcs. Nous

souhaitons ajouter la possibilité de prioriser certains arcs en fonction de connaissances expertes. Jusqu'à présent, seul l'état des variables était considéré comme issu d'un modèle particulier. Afin d'ajouter des informations expertes sur la structure du graphe, nous ajoutons l'hypothèse que le graphe d'interaction associé à un modèle de réseau bayésien étiqueté suit un modèle à blocs stochastiques [Holland et al., 1983] (désigné par l'acronyme SBM pour *Sochastic Block Model*) adapté à un graphe d'interaction étiqueté. Ce modèle a déjà été présenté dans la section 1.2.5. Pour l'apprentissage de structure de RBDE, nous faisons l'hypothèse que chaque nœud du graphe appartient à un et un seul bloc parmi B. Le bloc d'un nœud i est noté $f(i) \in \{1, ..., B\}$. Ces blocs sont représentés par des rectangles en pointillés dans la figure 4.1. Soit $G^l_{ij} \in \{0, 1\}$ la présence ou l'absence d'un arc orienté de i vers j et d'étiquette l. Le modèle SBM fait les hypothèses suivantes :

— La présence d'un arc G^l_{ij} d'étiquette l d'un nœud i vers un nœud j est indépendant de la présence d'un arc de même étiquette G^l_{uv} d'un nœud u vers un nœud v.

— La distribution de probabilité de G^l_{ij} ne dépend que de l, de $f(i)$ et de $f(j)$. Par exemple, dans le cas d'un modèle de réseau à 2 étiquettes ($L = 2$), la distribution jointe de $G^l_{i,j}$ est entièrement déterminée par les probabilités $P(G^1_{i,j}|f(i), f(j))$ et $P(G^2_{i,j}|f(i), f(j))\forall i, j, f(i), f(j), l$.

Nous supposons que les distributions de probabilités associées aux arcs du réseau sont déterminées par un jeu de paramètres ψ. Si les SBM sont généralement utilisés dans le but de détecter des blocs au sein d'un graphe connu, nous l'utilisons ici au contraire afin d'aider l'apprentissage d'un graphe inconnu en exploitant la connaissance à priori des blocs.

4.4.2 Exemple : niveaux trophiques dans un réseau écologique

Pour appliquer ce modèle à l'apprentissage de réseau écologique, nous considérons que les blocs connus sont les niveaux trophiques des espèces. Le niveau trophique d'une espèce indique sa place dans la chaîne alimentaire. Les organismes autotrophes comme les plantes ou les algues sont considérés comme des espèces basales, de niveau trophique 1. Les espèces ne consommant que des espèces basales ont un niveau trophique de niveau 2, et une espèce de niveau trophique k peut consommer des espèces de niveau trophique compris entre 1 et $k - 1$. La figure 4.2 montre un réseau écologique organisé en réseau trophique (les niveaux trophiques sont encadrés en pointillés). Nous faisons également l'hypothèse qu'une espèce d'un niveau trophique k consomme plus fréquemment des espèces du niveau trophique $k - 1$ que des espèces d'un niveau trophique inférieur. Une relation trophique est représentée par une relation positive dans un sens, et une relation négative dans l'autre. Modélisons un réseau écologique par un réseau bayésien dynamique étiqueté dans lequel on trouve deux étiquettes : une étiquette d'impulsion notée + et une étiquette d'inhibition notée −. Nous considérons que ces étiquettes n'ont pas d'effet sur l'apparition (seulement sur la survie). Nous

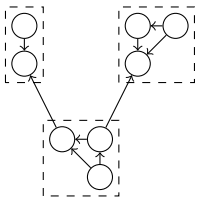


FIGURE 4.1 – Communautés dans un graphe orienté

choisissons d'appliquer ces hypothèses de manière stricte sur les interactions positives, tandis que les probabilités de présence des interactions négatives suivra une version relâchée de ces hypothèses. Dans un tel réseau, les impulsions ne décrivent que des interactions trophiques "positives" (l'impact d'une proie sur son prédateur) et les inhibitions peuvent décrire n'importe quelle interaction écologique négative (l'impact d'un prédateur sur sa proie, d'un parasite sur son hôte, la compétition entre deux espèces...). Nous intégrons également un bloc indépendant, qui comporte des espèces dont le niveau trophique n'est pas connu à priori. On ne peut donc appliquer aucune priorité sur les arcs qui concernent un individu de ce bloc, dit bloc indéterminé.

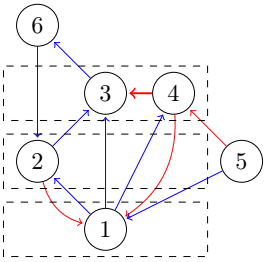


FIGURE 4.2 – Représentation d'un réseau écologique séparé en trois niveaux trophiques et deux espèces de bloc indéterminé

.

Nous modélisons ce comportement par un SBM. Le bloc $f(i)$ d'une espèce i correspond à son niveau trophique. $f(i) = 0$ si i appartient au bloc indéterminé, $f(i) = 1$ si i décrit une espèce basale, tandis que les grands prédateurs ont le niveau trophique le plus élevé.

Dans le cas où aucune des espèces i ou j n'appartient au bloc indéterminé ($f(i) \times f(j) > 0$) , nous décrivons les probabilités de présence d'un arc $G^l_{i,j}$ de la manière suivante pour les influences positives :

$$P\left(G^+_{ij} = 1\right) = 0 \text{ si } f(i) \geq f(j) \text{ et } \frac{e^{\alpha\Delta_{ij}}}{1 + e^{\alpha\Delta_{ij}}} \text{ si } f(i) < f(j).$$

où $\Delta_{ij} = f(i) - f(j)$ et $\alpha > 0$.

Les influences négatives représentent différents phénomènes (influence négative du prédateur sur sa proie, mais aussi parasitisme, compétition....). La distribution de probabilité de ce modèle pour les arcs représentant les influences positives ne prennent en compte que les positions relatives des niveaux trophiques lorsqu'aucune des deux espèces en relation n'appartient au bloc indéterminé. Ces probabilités sont les suivantes :

— Si $f(i) < f(j), P\left(G^-_{ij} = 1, G^+_{ij} = 1\right) = 0$ et $P\left(G^-_{ij} = 1 G^+_{ij} = 0\right) = \beta_2$

— Si $f(i) = f(j), P\left(G^-_{ij} = 1\right) = \beta_2$

— Si $f(i) > f(j), P\left(G^-_{ij} = 1\right) = \beta_1$

avec $\beta_1 > \beta_2$. Cela représente le fait que les relations proie/prédateurs sont les influences négatives les plus fréquentes.

Lorsqu'une espèce appartient au bloc indéterminé, il est impossible de définir une priorité sur les interactions qu'elle a avec les autres espèces quels que soit les sens ou les étiquettes de ces relations. Ainsi, lorsque $f(i) \times f(j) = 0$ (c'est à dire que i et/ou j appartient au bloc indéterminé),

les probabilités de présence d'un arc entre ces deux espèces sont les suivantes :

$$P\left(G_{ij}^{+}=1\right)=\alpha_0 \tag{4.9}$$

$$P\left(G_{ij}^{-}=1\right)=\beta_0 \tag{4.10}$$

α_0 et β_0 ont une valeur indépendant de α, β_1 et β_2. Le vecteur $\psi = (\alpha_0, \beta_0, \alpha, \beta_1, \beta_2)$ définit les priorités sur $\mathscr{L}G_{\rightarrow}$.

En pratique, l'estimation des paramètres β_1 et β_2 a une solution analytique, qui ne s'exprime que lorsque $TL(i) * TL(j) > 0$:

$$\begin{array}{rcl} \beta_1^{k+1} & = & \frac{\sum_{(i,j),\Delta_{ij}>0} g_{ij}^{-}}{|\{(i,j),\Delta_{ij}>0\}|} \\ \beta_2^{k+2} & = & \frac{\sum_{(i,j),\Delta_{ij}\leq 0} g_{ij}^{-}(1-g_{ij}^{+})}{\sum_{(i,j),\Delta_{ij}\leq 0}(1-g_{ij}^{+})} \end{array}$$

L'estimation de α est obtenue lorsque $TL(i) * TL(j) > 0$ par la solution numérique de l'équation suivante :

$$\sum_{(i,j),\Delta_{ij}<0} \Delta_{ij} g_{ij}^{+} = \sum_{(i,j),\Delta_{ij}<0} \Delta_{ij} \frac{\exp^{\alpha\Delta_{ij}}}{1+\exp^{\alpha\Delta_{ij}}}.$$

L'estimation des paramètres α_0 et β_0 a également une solution analytique :

$$\begin{array}{rcl} \alpha_0^{k+1} & = & \frac{\sum_{(i,j),TL(i)\times TL(j)=0} g_{ij}^{+}}{|\{(i,j),Tl(i)\times TL(j)=0\}|} \\ \beta_0^{k+2} & = & \frac{\sum_{(i,j),TL(i)\times TL(j)=0} g_{ij}^{-}(1-g_{ij}^{+})}{\sum_{(i,j),TL(i)\times TL(j)=0}(1-g_{ij}^{+})} \end{array}$$

L'étape E d'estimation des paramètres à graphe connu aura, dans ce cas, 5 nouveaux paramètres $\theta^{SBM} = \{\alpha_0, \beta_0, \alpha, \beta_1, \beta_2\}$ à estimer à l'aide de la structure $\mathcal{G}$ du graphe et des fonctions $f(i)$ renseignant sur les niveaux trophiques de chaque espèce.

L'étape R d'apprentissage de structure à paramètre connu ajoutera au vecteur C décrivant la fonction objectif un coefficient à chaque variable g_{ij}^{l}. Les coefficients associés aux influences positives sont indiqués dans l'équation 4.11 et ceux associés aux influences négatives dans l'équation 4.12.

$$\begin{array}{cl} \log(\alpha_0) & si\ f(i)\times f(j)=0 \\ \alpha\Delta_{ij} - \log(1+\exp^{\alpha\Delta_{ij}}) & si\ f(i)\times f(j)>0 et\ \Delta_{ij}<0 \\ 0 & si\ f(i)\times f(j)>0 et\ \Delta_{ij}>=0 \end{array} \tag{4.11}$$

$$\begin{array}{ll} \log(\beta_0) & si\ f(i)\times f(j)=0 \\ \log(\beta_1) & si\ f(i)\times f(j)>0\ et\ \Delta_{ij}>0 \\ \log(\beta_2) & si\ f(i)\times f(j)>0\ et\ \Delta_{ij}<=0 \end{array} \tag{4.12}$$

4.5 Évaluation de l'étape d'estimation des paramètres

Testons d'abord les performances de l'étape E, c'est à dire l'estimation des paramètres à graphe connu. Le modèle de RBDE utilisé pour les expériences est le suivant : L'apparition des variables est régie par un seul type de comportement inhérent qui n'a aucun impact sur la survie (donc un unique

cfg.	ε	ρ	τ	μ	cgf.	ε	ρ^+	ρ^-	μ
1	0.2	0.2	0.2	0.2	9	0.8	0.2	0.2	0.2
2	0.2	0.2	0.2	0.8	10	0.8	0.2	0.2	0.8
3	0.2	0.2	0.8	0.2	11	0.8	0.2	0.8	0.2
4	0.2	0.2	0.8	0.8	12	0.8	0.2	0.8	0.8
5	0.2	0.8	0.2	0.2	13	0.8	0.8	0.2	0.2
6	0.2	0.8	0.2	0.8	14	0.8	0.8	0.2	0.8
7	0.2	0.8	0.8	0.2	15	0.8	0.8	0.8	0.2
8	0.2	0.8	0.8	0.8	16	0.8	0.8	0.8	0.8

TABLE 4.1 – Description des différentes configurations de paramètres

paramètre ε, uniquement pour l'apparition). Ce modèle comporte une seule étiquette d'impulsion, une seule étiquette d'inhibition. L'impact de ces interactions ne se fait que sur la survie et pas sur l'apparition (nous avons donc un seul paramètre ρ et un seul paramètre τ). À chaque variable X_i^t est associée une covariable binaire $a_i^t \in \{0,1\}$ n'ayant aucun impact si $a_i^t = 0$ et pénalisant d'une quantité $\mu \in \{0,1\}$ les probabilités de survie et d'apparition de X_i^t.

L'estimation des paramètres est testée sur des données simulées à partir d'un modèle de réseau bayésien dynamique étiqueté à $n = 18$ variables et $T = 30$ pas de temps. Différentes configurations sont testées pour $\theta = \{\varepsilon, \rho, \tau, \mu\}$: chaque paramètre prend la valeur 0,2 (influence faible) ou 0,8 (influence forte) pour un total de 2^4 configurations possibles. Chaque configuration de $(\varepsilon, \rho^+, \rho^-, \mu)$ est indexée comme dans le tableau 4.1.

Pour chacune des 16 configurations de paramètres, nous réalisons 150 simulations de dynamique d'espèces et les paramètres sont estimés par maximisation de la vraisemblance sachant le graphe utilisé pour la génération des données à l'aide de la fonction fmincon de Matlab. Des indicateurs de performance (moyenne et écart-type des estimateurs) sur les 150 simulations de chaque configuration de paramètres sont représentés en Figure 4.3. En moyenne, la différence (en valeur absolue) entre la valeur des paramètres estimés et celle des paramètres réels est faible.

4.6 Évaluation de l'étape d'apprentissage de structure d'un RBDE

Étudions maintenant les performances de l'étape R, soit l'apprentissage de structure de graphe à paramètres connus avec le même modèle 4.5. Le réseau utilisé pour la simulation des données est un petit graphe de 4 nœuds. Les paramètres utilisés pour générer ces données sont les mêmes que dans la section précédente. Pour chacune des 16 configurations de paramètres décrites dans la section précédente, nous avons lancé 150 simulations de variables dynamiques. L'optimisation de la structure du graphe par programmation linéaire 0-1 est résolue à l'aide du logiciel cplex. Pour une configuration donnée de $\theta_{\rightarrow}$, nous utilisons différents indicateurs de performance. Nous relevons la précision et le rappel moyen pour chaque étiquette d'arc. Ces indicateurs sont calculés à partir d'une matrice de confusion entre le graphe appris $\hat{\mathcal{G}}$ et le graphe "réel" utilisé pour simuler les données. Une matrice de confusion entre deux graphes est un tableau à deux entrées affichant pour chacun

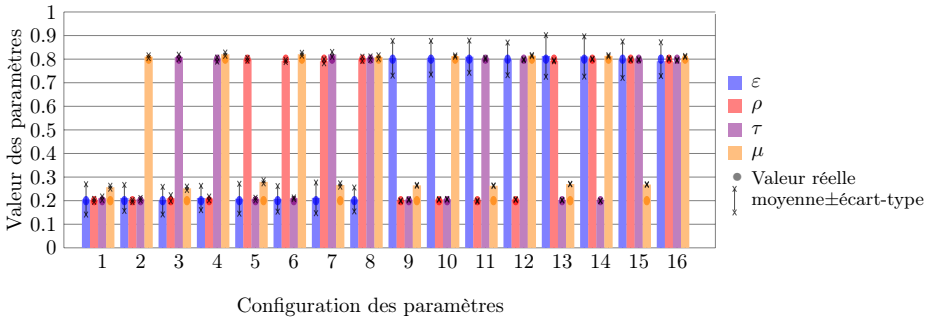


FIGURE 4.3 – Diagramme des valeurs moyennes des paramètres estimés pour 150 observations simulées pour les 16 configurations de paramètres. Les cercles représentent la valeur réelle des paramètres, et l'intervalle tracé entre deux croix représente la différence entre la valeur moyenne des paramètres estimés et leur écart-type.

Graphe	G			
$\hat{G}$	Étiquette	$\emptyset$	q	r
	$\emptyset$	Arcs absents des deux graphes	Arcs absents dans $\hat{G}$ mais d'étiquette q dans G	Arcs absents dans $\hat{G}$ mais d'étiquette r dans G
	q	Arcs absents dans G mais d'étiquette q dans $\hat{G}$	Arcs d'étiquette q dans les deux graphes	Arcs d'étiquette q dans $\hat{G}$ mais d'étiquette r dans G
	r	Arcs absents dans G mais d'étiquette r dans $\hat{G}$	Arcs d'étiquette r dans $\hat{G}$ mais d'étiquette q dans G	Arcs d'étiquette r dans les deux graphes

TABLE 4.2 – Définition d'une matrice de confusion entre deux graphes G et $\hat{G}$ avec deux étiquettes. Dans chaque cellule se trouve le nombre d'arcs correspondant à la description du tableau.

des deux graphes les arêtes similaires et différentes en matière de présence et d'étiquette. Le tableau 4.2 montre comment se présente une matrice de confusion entre deux graphes, pour deux étiquettes différentes. Pour qu'une matrice de confusion soit valide, les deux graphes doivent avoir le même nombre de nœuds et chaque nœud doit représenter la même chose pour les deux graphes. La précision pour une étiquette d'arc (un type d'interaction, + ou -) correspond à la quantité $\frac{\text{arcs correctement appris}}{\text{arcs appris}}$. Le rappel pour une étiquette d'arête correspond à la quantité $\frac{\text{arcs correctement appris}}{\text{arcs présents dans le } \textit{graphe original}}$. Pour un graphe de cette taille, il faut environ 35.6 secondes pour apprendre la structure d'un graphe à paramètres $\theta_{\rightarrow}$ fixé. La précision et le rappel sont représentés en Figure 4.4.

Pour ce problème de petite taille, les résultats dépendent de la configuration des paramètres. La qualité de l'apprentissage, mesurée par la précision ou le rappel, est bien meilleure lorsque les paramètres ρ et ε augmentent, et est légèrement meilleure lorsque les paramètres τ et μ augmentent.

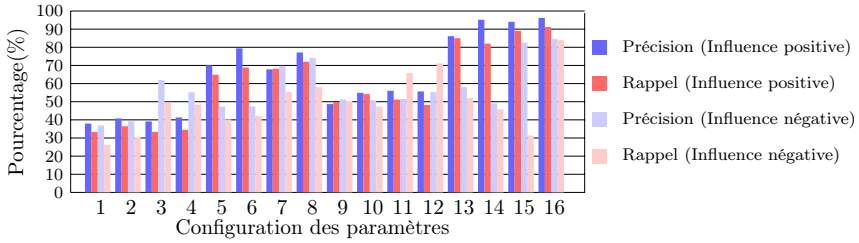


FIGURE 4.4 – Précision et rappel moyen (à partir de 150 données simulées) pour un graphe à 4 espèces pour l'étape d'apprentissage de structure pour chaque configuration de paramètres.

4.7 Évaluation de l'algorithme d'apprentissage

Pour l'apprentissage de la structure du graphe conjointement à l'estimation des paramètres, nous considérons un graphe avec 45 variables avec le modèle de RBDE et avec les caractéristiques de paramètres similaires aux sections précédentes. Le nombre de parents maximum pour chaque nœud est fixé à $k = 4$ et les paramètres prennent la valeur de la configuration 16 du tableau 4.1. 40 simulations de variables dynamiques sont réalisées avec ce graphe, et pour chacune de ces simulations, l'algorithme d'apprentissage par restauration-estimation est appliqué. L'algorithme itère en moyenne 6,4 fois. La précision moyenne est de moins de 20 % pour les deux étiquettes d'arcs, et le rappel moyen est inférieur à 30%, ce qui n'est pas un résultat satisfaisant. Toutefois, si l'on considère un graphe "consensus" composé des x arcs les plus souvent appris parmi les 40 graphes appris, et que l'on compare ce graphe avec l'original (utilisé pour la simulation des données), il est possible de contrôler un niveau de précision et de rappel en sélectionnant une valeur raisonnable pour x. La figure 4.5 illustre la manière de construire ces graphes consensus.

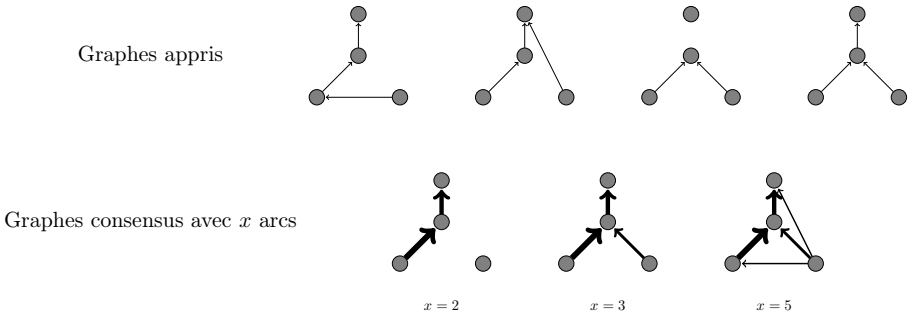


FIGURE 4.5 – Illustration de la construction d'un graphe consensus. En haut sont représentés 4 graphes appris avec 4 différents jeux de données. En bas sont représentés les graphes consensus des x arcs les plus souvent appris. La largeur de chaque arc est proportionnelle au nombre de fois où il a été appris.

Le rappel augmente nécessairement avec x alors que la précision peut baisser avec x, comme représenté en Figure 4.6. Par exemple, si l'on considère le graphe des 100 arcs les plus souvent représentés, la précision est de 65,67% pour les arcs étiquetés + et de 93,94% pour les arcs étiquetés

−, tandis que le rappel moyen est de 83,02% pour les arcs étiquetés + et de 59,62% pour les arêtes étiquetées −. Peu importe la valeur de x, le rappel et la précision sont bien meilleurs sur un graphe consensus appris avec notre méthode d'apprentissage que sur un graphe généré aléatoirement arête par arête (voir Figure 4.6). Sans ajouter de connaissances expertes, l'apprentissage est donc difficile

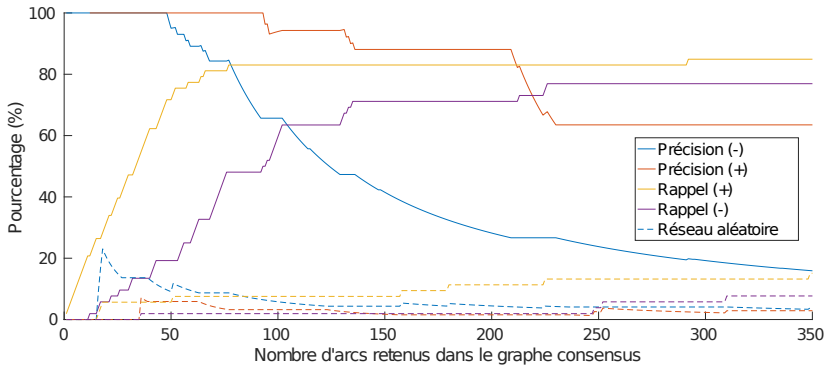


FIGURE 4.6 – Évolution de la précision et du rappel pour l'algorithme d'apprentissage complet de la structure d'un P-DBN pour un graphe à 42 espèces. L'axe x représente le nombre d'arêtes dans le graphe consensus construit à partir de 40 jeux de données.

et nécessite beaucoup de données pour être efficace. Heureusement, de telles connaissances sont souvent disponibles.

4.8 Performance de l'algorithme d'apprentissage avec ajout de connaissances

Nous souhaitons comparer les performances de l'algorithme d'apprentissage lorsque l'on connaît plus ou moins d'information à priori. Ces données sont générées à l'aide d'un réseau bayésien étiqueté ayant les mêmes caractéristiques que précédemment (1 type d'impulseur, 1 type d'inhibiteur, 2 types de comportements inhérents, 2 valeurs de covariables possibles, les seuls paramètres inconnus étant l'impact de l'impulseur sur la survie, celui de l'inhibiteur sur la survie, celui d'un des comportement sur l'apparition, et l'impact d'une des valeurs de la covariable similaire sur la survie et sur l'apparition). Nous avons alors réalisé, à partir des mêmes données, 4 apprentissages différents :

— Un apprentissage de réseau bayésien dynamique étiqueté "classique" sans aucune information

— Un apprentissage de réseau bayésien dynamique étiqueté dans lequel 20% des variables $G^l_{i,j}$, c'est à dire des arcs du graphe sont connus (Parmi l'ensemble des arcs possibles entre deux nœuds, on connaît, pour 20% leur absence ou leur présence, leur étiquette et leur sens.)

— Un apprentissage de réseau bayésien dynamique étiqueté dans lequel on connaît les niveaux trophiques et dont on utilise le prior décrit dans la section 4.4.2.

— Un apprentissage de réseau bayésien dynamique non étiqueté appris par l'algorithme de [Dojer, 2006] optimisant un score de test d'information mutuel (MIT) [Vinh et al., 2011]. Le graphe appris n'étant pas étiqueté, il est étiqueté à posteriori, à l'aide des formules définissant un réseau bayésien qualitatif [Wellman, 1990] décrites en 3.4.3 en utilisant les fréquences

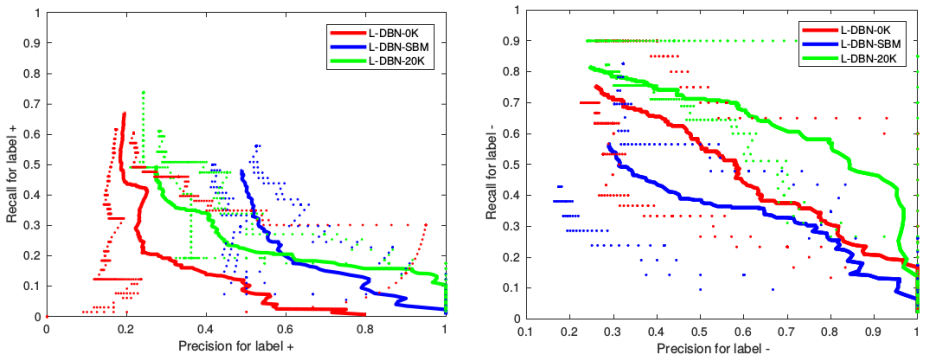


FIGURE 4.7 – Précision et rappel pour les impulseurs '+' (à gauche) et les inhibiteurs '−' (à droite) pour l'apprentissage de réseau bayésien étiqueté sans information (L-DBN-0K), avec 20% des arcs connus (DBN-20K) et avec un prior mis sur les nœuds par modèle stochastique à blocs (L-DBN-SBM). Les lignes continues relient les points décrivant la moyenne des indicateurs de précision et de rappel. Les points isolés représentent les meilleurs et les pires cas de graphes consensus sur les 10 jeux de données générés par un graphe synthétique.

associées aux données comme probabilités. Puisqu'il n'y a qu'un seul type d'inhibiteur et d'impulseur, cet étiquetage en intéractions "positives" et "négatives" convient.

Ces essais ont tout d'abord été effectués à partir de données simulées à partir de 10 graphes synthétiques de 20 nœuds générés à l'aide d'un modèle stochastique à blocs tel que décrit en 4.4.2 avec $\alpha = 1/\sqrt{20}, \beta_1 = \alpha/2, \beta_2 = \beta_1/2$. 10 jeux de données ont été générés pour chacun de ces graphes. Un jeu de données correspond à des trajectoires de présence/absence simulées sur 30 pas de temps, où $a_i^t = 0$ pour $t \in \{1, ..., 12\}$ et $a_i^t = 1$ pour $t \in \{13, ..., 30\}$ (dans ce modèle, $\mu_{a_i^t=1} = 1$ pour l'apparition comme pour la survie, on a donc un unique μ inconnu pour $a_i^t = 0$ ayant le même effet sur l'apparition et la survie). Les paramètres $\theta = \{\varepsilon, \mu, \rho, \tau\}$ ont chacun une valeur de 0.8. L'algorithme de restauration-estimation a été appliqué à chacun de ces jeux de données, donnant 10 graphes appris pour chaque graphe synthétique. Les arcs appris sont ordonnés par ordre décroissant d'occurrence (de nombre de fois où l'arc a été appris) pour les 10 graphes appris afin de définir un graphe consensus agrégé contenant les x premières arêtes de cet ordre. La figure 4.7 montre l'évolution de la précision et du rappel des impulseurs ('+') et des inhibiteurs ('−') lorsque le nombre de fois où ils ont été appris dans le graphe consensus change. Ainsi, chaque point dans ce graphique représente un graphe agrégé des x arcs les plus souvent appris. L'apprentissage de réseau bayésien classique par MIT apprend des graphes dont la précision et le rappel sont négligeables. Cette méthode n'a donc pas été représentée dans cette figure.

Moins d'arcs sont appris lorsqu'un prior SBM est utilisé pour guider l'apprentissage. L'utilisation de ce prior améliore la précision et le rappel pour les arcs d'impulseurs par rapport aux autres méthodes d'apprentissage, y compris l'apprentissage lorsque des arêtes sont déjà connues (le calcul du rappel et de la précision se faisant dans ce cas uniquement sur les arcs inconnus). En revanche, l'utilisation du prior SBM est moins efficace que les autres méthodes pour les inhibiteurs. Cela peut s'expliquer par le fait que ce prior contraint bien plus les impulseurs que les inhibiteurs.

Ces méthodes d'apprentissage ont également été testées sur des données simulées à partir

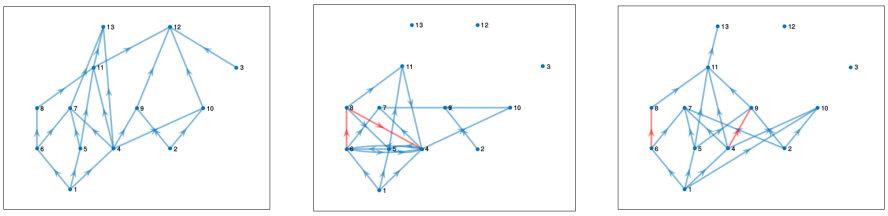


FIGURE 4.8 – Réseau "*Alaskan food web*" connu. Gauche : réseau réel (uniquement des influences positives). Milieu : Graphe appris par restauration-estimation sans aucune connaissance à priori. Droite : Graphe appris par restauration-estimation avec un prior SBM. Les arcs représentés en bleu représentent des influences positives (impulseurs) et les arcs représentés en rouge représentent les arcs ayant été appris parfois comme impulseurs et parfois comme inhibiteurs sur les différents jeux de données simulés.

d'un graphe réel correspondant à la structure d'un réseau écologique connu : "*Alaskan food web*" [Estes et al., 2009]. La méthode d'apprentissage avec certaines arêtes déjà connues n'a pas été testée, son intérêt par rapport aux autres méthodes semblant moindre. Il y a 13 nœuds dans ce graphe, chaque nœud correspondant à une espèce. Chaque espèce peut être associée à un niveau trophique parmi 5 différents, permettant d'utiliser un prior SBM. Ce graphe ne contient aucun inhibiteur en pratique (voir la figure 4.8). Cependant, l'apprentissage de réseau sur les données simulées à partir de ce graphe se fait sur un modèle pouvant contenir un inhibiteur. 10 jeux de données ont été simulés avec les mêmes caractéristiques que pour les essais sur graphes synthétiques. Ces 10 jeux de données permettent de construire un graphe consensus. La précision et le rappel maximal atteints pour les impulseurs ('+') pour un graphe consensus sur ces données sont les suivants :

— Une précision 0,47 et un rappel de 0,33 pour la méthode d'apprentissage par MIT

— Une précision 0,26 et un rappel de 0,86 pour la méthode d'apprentissage de RBDE sans aucune connaissance

— Une précision 0,49 et un rappel de 0,81 pour la méthode d'apprentissage de RBE avec un prior sur les niveaux trophiques

Si toutes ces méthodes apprennent très peu d'arêtes d'inhibiteurs, la prior SBM permet d'apprendre des graphes plus parcimonieux, avec moins d'arêtes (35 arêtes apprises avec un prior SBM contre 85 sans prior). Les figures 4.8 au milieu et à droite illustrent l'apport du prior SBM sur la qualité de l'apprentissage. Par exemple, sans ce prior, l'information selon laquelle une espèce ne se nourrit pas sur des espèces d'un même niveau trophique (comme c'est le cas pour le graphe réel) ne peut pas être déduite uniquement des données.

4.9 Choisir un nombre d'arcs dans le graphe consensus

Dans la pratique, nous ne connaissons pas le réseau réel. Il est difficile d'évaluer directement la qualité du graphe appris, et donc d'en définir un nombre d'arcs à retenir dans le graphe consensus. Pour obtenir un graphe unique après une série d'apprentissages sur plusieurs jeux de données, plusieurs solutions sont possibles. Une première est de considérer directement un graphe dans lequel chaque arc est pondéré par le nombre de fois où il a été appris sur l'ensemble des jeux de

données considérés. La valeur appliquée à chaque arc peut être interprétée comme la probabilité que l'interaction associée fasse partie du réseau écologique. Si l'on préfère considérer un réseau écologique dans lequel les arcs ne sont pas pondérés, on peut établir un seuil tel que les arcs retenus dans le graphe consensus soient appris dans au moins $x\%$ des jeux de données considérés. La figure (4.9) ordonne chaque arc selon le pourcentage de fois où il a été appris sur les 40 simulations de la section 4.7. Cette figure peut aider à choisir un nombre d'arcs x en cherchant un décrochage dans l'histogramme. Cette figure peut être mise en relation avec la figure (4.6), l'abscisse représentant un arc supplémentaire dans les deux cas. Si, à partir de la figure (4.9), on choisit 50 arcs, on peut lire la précision et le rappel sur la figure (4.6) à la même abscisse. Il est à noter que dans la pratique, la précision et le rappel ne sont pas disponibles, mais à la lecture de cet histogramme comme des probabilités de présence, on peut s'en servir pour choisir un seuil.

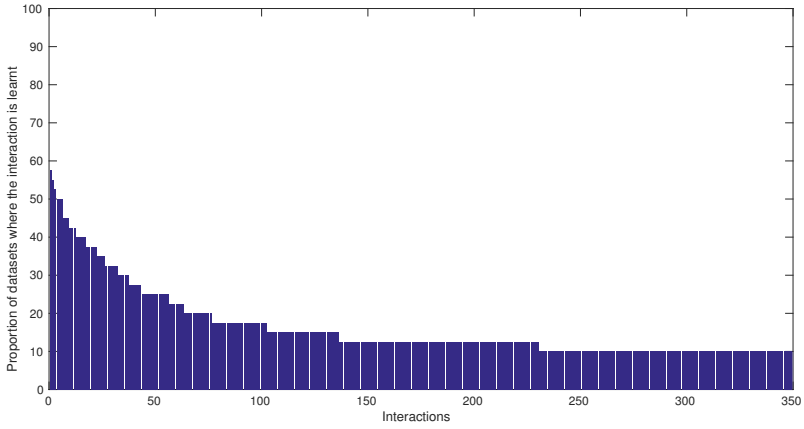


FIGURE 4.9 – Diagramme représentant le pourcentage de fois où chaque arc a été appris sur les 40 jeux de données utilisés dans l'expérimentation de la section 5.3.

4.10 Apprentissage sur des données réelles : espèces marines dans des forêts de kelp

Le programme PISCO (*Partnership for Interdisciplinarity Studies of Coastal Oceans*) mène des recherches sur les écosystèmes marins au large de la Californie afin de mieux comprendre et gérer ces écosystèmes [Caselle and Lowe, 2010, Hamilton et al., 2010]. Parmi ces écosystèmes se trouvent les forêts de kelp. Le kelp est une algue géante formant des forêts sous-marines abritant une biodiversité importante. La durabilité et la survie de ces forêts étant très dépendantes de leurs interactions écologiques, notamment avec les loutres de mer, cet écosystème est très étudié dans le domaine de l'écologie trophique [Graham, 2004, Estes and Duggins, 1995]. Le programme PISCO conduit régulièrement des observations d'espèces d'algues et d'animaux marins des forêts de kelp au large de la californie [Syms and Caselle, 2003]. Nous avons utilisé les données issues de cette étude afin d'apprendre les interactions écologiques entre les espèces de ce milieu. Ces données contiennent des mesures d'abondances d'un grand nombre d'espèces de poissons, d'invertébrés et d'algues. Ces

données proviennent de différents sites dont certains peuvent avoir le statut de réserve marine et sont donc protégés de la pêche et de l'exploitation humaine. Il est possible, en utilisant un ou plusieurs seuils sur les abondances de ces données, d'obtenir des jeux de données renseignent sur l'absence ($X_i^t = 0$) ou sur la présence ($X_i^t = 1$) de chaque espèce i à chaque année t $\forall i \in \{1, \ldots, n\}$ et $t \in \{1, \ldots, T\}$. Il est également possible de renseigner pour chaque site son statut de protection chaque année : Pour un site donné, $a^t = 1$ si le site est protégé à l'année t et $a^t = 0$ si ce site n'est pas protégé.

Nous modélisons ces données par un réseau bayésien dynamique étiqueté prenant en compte les informations de protection. Ce modèle est le même que celui présenté dans la section 3.5.5. Pour rappel, pour un site donné, la probabilité de recolonisation d'une espèce i est décrite dans l'équation 4.13 et la probabilité de survie d'une espèce i est décrite dans l'équation 4.14. Le paramètre ε décrit la probabilité d'apparition ou de survie spontanée, ρ décrit la probabilité de réussite d'une interaction positive, τ la probabilité de réussite d'une interaction négative et μ_{a^t} décrit l'impact de la protection décrite par a^t sur ces probabilités. $\pi_i = \overset{q}{\pi}_i \cup \overset{r}{\pi}_i$ décrit l'ensemble des parents de i, c'est à dire l'ensemble des espèces ayant une influence positive q ou négative r sur i. Pi_i^t Désigne l'ensemble des variables aléatoires renseignant sur la présence ou sur l'absence des espèces de π_i lors de l'année t. $\overset{q}{N}_i^t$ et $\overset{r}{N}_i^t$ décrit le nombre de parents à influence respectivement positive et négative sur i présents à t. u_i différencie les espèces basales des espèces non basales : si i est basale, c'est à dire qu'elle n'a pas de parents à influence positive, soit $\overset{q}{\pi}_i = \emptyset$, alors $u_i = 0$; sinon, $u_i = 1$. Ce modèle établit la quantité $\varepsilon_0^{sur} = 1$ et $\varepsilon_1^{sur} = 0$ afin de gérer le fait qu'une espèce basale n'a pas besoin d'influences positives pour survivre. Le modèle établit également que lorsqu'un site est protégé ou non, cela a le même impact sur toutes les espèces ($a^t = a_1^t = ... = a_n^t$). De plus, la protection d'un site n'a aucun impact sur les probabilités de survie et de recolonisation ($\mu_1^{app} = \mu_1^{sur} = \mu_1 = 1$) alors que l'absence de protection baisse les probabilités d'apparition et de survie de la même manière ($\mu_0^{app} = \mu_0^{sur} = \mu_0$).

$$P\left(X_i^{t+1} = 1 | X_i^t = 0, a_i^t\right) = \mu_{a_i^t}^{app} \cdot \varepsilon^{app} \tag{4.13}$$

$$P\left(x_i^{t+1} = 1 | x_i^t = 1, \Pi_i^t, a_i^t\right) = \mu_{a_i^t} \cdot \left(\varepsilon_{u_i}^{sur} + (1 - \varepsilon_{u_i}^{sur}) \cdot \left(1 - (1 - \rho^{sur})^{\overset{q}{N}_i^t}\right)\right) \cdot (1 - \tau^{sur})^{\overset{r}{N}_i^t} \tag{4.14}$$

Le but est d'utiliser cette modélisation ainsi que l'algorithme d'apprentissage de réseau bayésien dynamique étiqueté afin d'apprendre les interactions entre les espèces de ces jeux de données. Il existe une base de données regroupant un ensemble d'interactions entre ces espèces connues des écologues. Cette base de données peut servir de vérification des interactions apprises par réseau bayésien dynamique étiqueté. Cependant, l'apprentissage des interactions sur ces données ne permet d'apprendre que très peu d'interactions connues des écologues. Ces interactions connues ne semblent même pas correspondre aux corrélations des abondances entre ces espèces.

Cette étude pose question sur l'utilisation possible de la méthode d'apprentissage de structure de réseau par réseau bayésien dynamique étiqueté. Comme nous l'avons vu précédemment, cette méthode marche mieux lorsque l'on a plusieurs jeux de données différents. Or, les différents jeux de données utilisés ne sont obtenus que par seuillage d'un même jeu de données, et ne constituent donc pas des données provenant de différentes observations.

Il est possible que les données de cette étude ne soient pas adaptées à notre méthode d'apprentissage. Ces données proviennent effectivement d'un milieu ouvert, peu contrôlé, dont la dynamique des espèces est sans doute surtout expliquée par d'autres facteurs que les interactions entre ces

espèces (la météo, la date de relevé, la méthode d'échantillonnage, la profondeur...). En l'absence de connaissances sur ces facteurs, il est possible qu'ils aient bien plus d'influence sur la dynamique des espèces que les interactions écologiques.

Peut-être obtiendrons-nous de meilleurs résultats sur des données issues de milieux plus contrôlés, et où il est possible d'obtenir plusieurs jeux de données indépendants mais représentant les mêmes phénomènes. Ce genre de données peut se trouver dans des études d'espèces microbiennes dans des milieux similaires, mais séparés, ou dans des études sur des écosystèmes locaux comme des champs expérimentaux que nous décrirons dans un chapitre dédié.

4.11 Bilan

Nous avons décrit dans cette section un algorithme permettant d'apprendre la structure d'un réseau bayésien dynamique étiqueté. Il fonctionne en deux étapes répétées, une étape d'estimation des paramètres et une étape de restauration, apprenant la structure du graphe. Cette dernière étape peut se faire à l'aide de la résolution d'un problème de programmation linéaire en nombres entiers, dont la taille peut s'avérer très importante, grandissant de manière factorielle par rapport au nombre d'étiquettes et de parents maximum par nœud. Il a été testé sur plusieurs jeux de données simulées et réelles, en laissant la possibilité d'intégrer un prior sur l'apprentissage des arcs à l'aide d'un modèle de type SBM. Ces différents essais montrent que l'apprentissage d'une structure conforme à celle d'un réseau connu, est difficile lorsque l'on dispose de peu de jeux de données. Il semble nécessaire d'utiliser plusieurs jeux de données et d'agréger les structures apprises indépendamment par ces jeux de données. Une méthode d'agrégation de graphes appris par "graphe consensus" a été proposée, et semble donner des résultats satisfaisants sur des données simulées. Cependant, cette méthode n'est guère naturelle, et une meilleure méthode si l'on dispose d'une série de plusieurs jeux de données indépendants serait d'apprendre directement une seule structure pour l'ensemble des données, qui maximiserait globalement la vraisemblance sur tous ces jeux de données. Des essais supplémentaires avec cette méthode d'apprentissage sur des données simulées sont prévus, mais ils n'apparaîtront pas dans ce manuscrit. La section suivante détaillera l'application de cette méthode d'apprentissage de réseau sur un autre problème réel d'apprentissage de réseau écologique dans lequel on dispose de plusieurs jeux de données indépendants.

Afin d'alléger l'algorithme d'apprentissage, il est possible d'utiliser des méthodes de relaxation continue de programme linéaire [Agmon, 1954]. Il s'agit de ne pas contraindre les variables du problème à prendre des valeurs entières afin de réduire les contraintes. Si l'apprentissage est facilité, chaque paire de nœuds aura un poids, interprétable comme la probabilité de présence d'un arc. Cela demandera donc une étape supplémentaire de discrétisation des valeurs d'arête, ou l'interprétation d'un autre type de graphe dans lequel les arcs ne sont pas certains, mais associés à un poids.

Chapitre 5

Cas d'étude : réseau d'espèces d'arthropodes dans des cultures

Un apprentissage de réseau écologique à partir de données d'espèces marines a déjà été effectué. Cependant, les interactions apprises par réseau bayésien dynamique étiqueté ne correspondaient pas à des interactions connues par les écologues étudiant ce système. Il est possible que l'apprentissage par réseau bayésien dynamique étiqueté soit difficile sur des données en environnement ouvert et peu contrôlé dont les observations peuvent être en majeure partie dépendantes de facteurs extérieurs aux interactions écologiques des espèces.

L'étude *UK Farm Scale Evaluations* [Bohan et al., 2005] a été effectuée sur des champs expérimentaux au royaume-uni dans lesquels ont été installé des pièges aspirant les invertébrés fréquentant ces champs. Ce dispositif a été installé sur plusieurs parcelles différentes et 2 relevés ont été effectués à deux périodes différentes. Ces données sont de plus grande taille que les données PISCO, car elles contiennent plusieurs jeux de données indépendants (un par parcelle), mais une dynamique moins importante (uniquement deux dates de relevés). Des méthodes d'apprentissage de réseau trophique ont déjà été appliquées sur ces données [Tamaddoni-Nezhad et al., 2012], mais ne prenaient pas en compte l'information de la dynamique temporelle des espèces induite par les deux dates d'observations différentes pour chaque parcelle. Il semble possible d'utiliser ces données afin d'apprendre la structure du réseau écologique renseignant sur les interactions écologiques entre les espèces d'arthropodes piégés.

5.1 Description des données

Cette étude a été menée sur des champs expérimentaux de différentes cultures. Dans ces champs ont été placés des pièges aspirant les arthropodes passant dans le champ. Ces pièges permettent de recueillir des données d'abondance de chaque espèce d'arthropodes piégés. Cette étude porte sur 257 parcelles dont :

— 66 parcelles de betterave (B)

— 59 parcelles de maïs (M)

— 67 parcelles de colza d'été (SR)

— 65 parcelles de colza d'hiver (WR)

Chaque parcelle est séparée en deux, correspondant à deux traitements différents : une partie contient des plantes « traditionnelles » et une partie contient des plantes génétiquement modifiées, résistantes à l'herbicide. Chaque parcelle contient 3 points d'échantillonnage (pièges) par traitement : Un en bordure du champ (0m), un proche de la bordure (2m) et un plus éloigné (32m). Dans chaque point d'échantillonnage de chaque traitement de chaque parcelle sont relevés les nombres de chaque arthropode piégé. Les relevés ont lieu à deux dates distinctes (différentes selon le type de culture) , choisies d'après le cycle de vie des arthropodes et la pose d'herbicide : E (early) et L (late). Nous disposons en outre d'une base de données contenant la description de 5095 espèces d'arthropodes. L'apprentissage du réseau écologique représentant les interactions entre les espèces d'arthropodes présents dans ces champs a déjà été effectué. Une méthode de programmation logique sur ces données a permit d'apprendre une structure de réseau trophique à partir des données d'abondance des espèces. L'utilisation d'apprentissage automatique de fouille de texte sur la littérature scientifique a permis de repérer des interactions connues entre les espèces d'arthropodes en question. Ces interactions apprises par fouille de texte servent de vérification de la cohérence des interactions apprises par programmation logique. Cependant, dans ces études, la question de la dynamique n'a pas été abordée. Or, cette dynamique a été observée en partie, car deux relevés d'abondances ont eu lieu à deux pas de temps différents (avant et après récolte). Nous souhaitons contribuer à cette étude en exploitant cette information dynamique à l'aide de notre méthode d'apprentissage de réseau bayésien dynamique étiqueté. Notre méthode utilise des données binaires. Ainsi, pour pouvoir l'utiliser, nous devrons nous passer de l'information de l'abondance des espèces et n'en retenir que l'information de présence/absence. Le but est d'apprendre à partir de ces données le réseau trophique correspondant aux arthropodes relevés dans ces champs à l'aide d'un modèle de réseau bayésien dynamique étiqueté (RBDE). Puisque des études visant à apprendre la structure d'un réseau trophique ont déjà été effectuées, il est intéressant de comparer les résultats déjà obtenus avec ceux obtenus par apprentissage de structure de RBDE. Cela nous permettra de vérifier si la perte d'une information (abondance des espèces) et l'ajout d'une information inédite (dynamique) permet de retrouver des liens similaires. La nature des données, séparées en différentes cultures nous permet par ailleurs de pousser l'analyse. En effet, il peut être intéressant d'apprendre plusieurs réseaux écologiques différents : un par culture. Cela nous permettra de vérifier si notre méthode apprend des interactions similaires entre les différentes cultures, ou si, au contraire, chaque culture a son propre réseau écologique. Cette étude a donc 3 objectifs :

— Apprendre les interactions entre les espèces observées dans cette étude par notre méthode d'apprentissage de réseau bayésien dynamique étiqueté.

— Comparer les interactions apprises par apprentissage de réseau bayésien dynamique étiqueté aux interactions apprises par d'autres méthodes.

— Observer les différences entre les réseaux trophiques appris dans des cultures différentes.

5.2 Analyse des données

5.2.1 Étude de corrélations

Avant d'apprendre le réseau bayésien, étudions le comportement des espèces d'arthropodes relevés dans les champs. En effet, l'apprentissage du réseau par réseau bayésien dynamique étiqueté a un

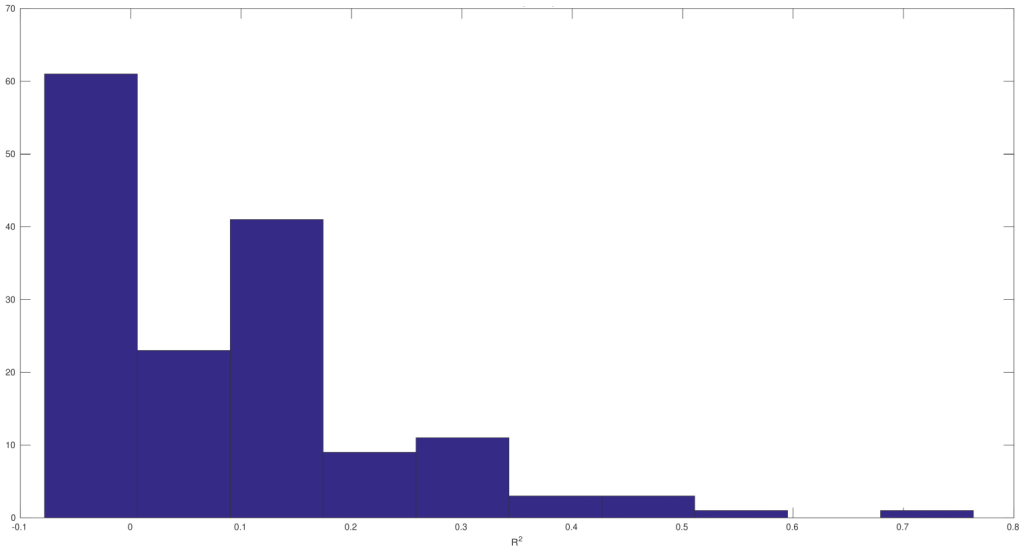


FIGURE 5.1 – Distribution des coefficients de corrélation entre les abondances avant récolte et après récolte pour chaque espèce

sens si les indicateurs de présence ou d'absence d'espèces sont corrélées entre elles. Le coefficient de corrélation entre deux variables aléatoires X et Y est défini par

$$r = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

où $cov(X, Y)$ désigne la covariance des variables X et Y, σ_X et σ_Y leurs écarts-types respectifs. Ce coefficient r est une quantité comprise entre -1 et 1, les valeurs négatives correspondant à des corrélations négatives (la valeur de X baisse lorsque celle de Y augmente et inversement) et les valeurs positives à des corrélations positives (la valeur de X augmente lorsque celle de Y augmente et inversement). les valeurs $r = -1$ et $r = 1$ correspondent à des corrélations parfaites négatives ou positives (par exemple entre une variable X et $-X$ ou entre X et elle-même respectivement), et $r = 0$ indique qu'il n'existe aucune corrélation. Observons tout d'abord les corrélations des abondances de chaque espèce entre les deux dates d'observation, dont la distribution est représentée en figure 5.1. Peu d'espèces sont corrélées entre elles. Les différences entre les deux dates d'observation ne peut donc pas expliquer le comportement de chaque espèce indépendamment. Cette différence pourrait en revanche s'expliquer par les autres espèces. La heatmap en figure 5.2 représente les corrélations entre les abondances de chaque espèce lors de la première date d'observation (lignes) et de la dernière (colonnes). La couleur rouge représente une corrélation faible, la couleur jaune une corrélation forte, le noir un manque de données.

Ce résultat semble encourageant : les fortes corrélations sont assez éparses, et dispersées entre beaucoup d'espèces. Ceci est la forme attendue d'un réseau d'interactions écologiques que l'on cherche à apprendre : un réseau dans lequel chaque espèce n'est liée qu'à un nombre restreint d'espèces.

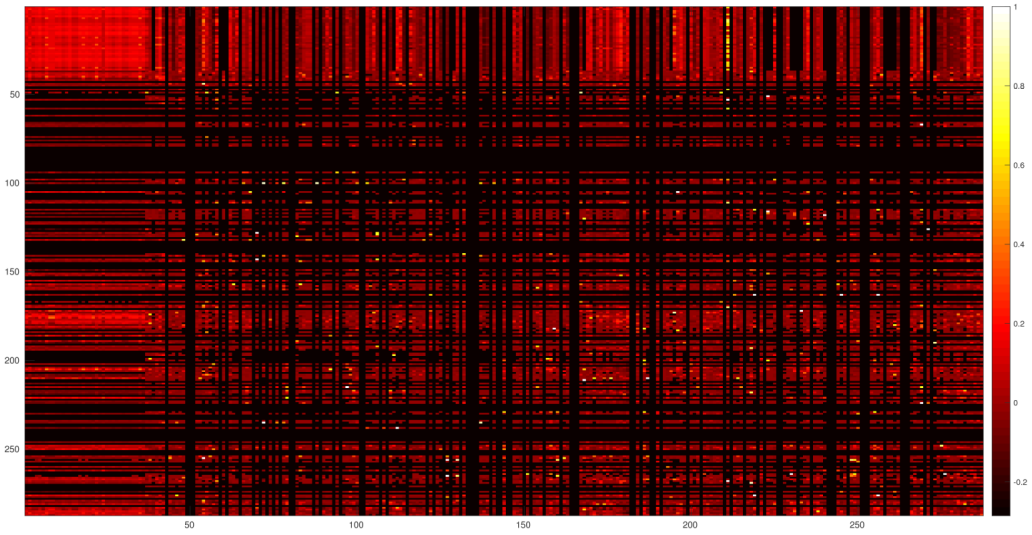


FIGURE 5.2 – Distribution des coefficients de corrélations entre les abondances de chaque espèce avant récolte et chaque espèce après époque

	E	L
0	14,2	20,85
2	3,76	10,51
32	3,14	8,59

TABLE 5.1 – Abondances moyennes de chaque piège pour chaque pas de temps

5.2.2 Abondances par pièges

Dans chaque parcelle sont relevés 3 pièges : un à l'extérieur de la parcelle (0m), un en bordure (2m) et un à l'intérieur (32m). Ces 3 pièges sont difficilement interprétables comme des observations indépendantes, et il est tentant de sommer les abondances de chaque espèce afin d'avoir une mesure par pas de temps, par espèce et par parcelle. Étudions la corrélation des abondances des espèces entre chacun de ces pièges pour savoir dans quel cas cette agrégation est pertinente. Le tableau 5.1 donne les moyennes des abondances des espèces pour chaque piège et chaque pas de temps.

Il semble que les pièges situés à 2m et à 32m représentent des abondances similaires, tandis que le piège situé à 0m semble à part. Une représentation de ces corrélations est donnée en figure 5.3. Les coordonnées en ligne et en colonne représentent un piège (la première ligne/colonne représente le piège 0m, la seconde 2m et la troisième 23m), le triangle supérieur représente la première date d'observation, le triangle inférieur la seconde. Il semble que la distribution des abondances des espèces issues des pièges 2m et 32m soient similaires, ce qui avait été suggéré par les moyennes

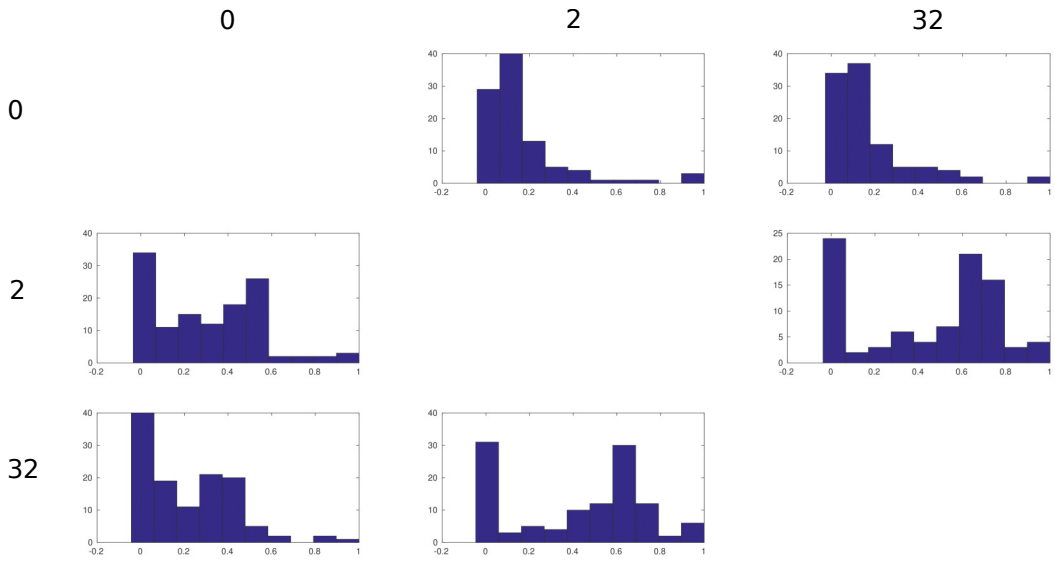


FIGURE 5.3 – Distribution des corrélations des abondances des espèces par piège. La corrélation se calcule pour chaque espèce entre les abondances avant récolte et les abondances après récolte.

proches. On peut penser qu'il est raisonnable de réunir les abondances des espèces observées dans ces deux pièges. Ce n'est cependant pas le cas pour les espèces recueillies dans le piège 0m. Ces différences peuvent s'expliquer par le fait que ce piège est à l'extérieur de la parcelle, contrairement aux pièges de 2m et de 32m. Les données concernant ce piège ne seront pas forcément utiles pour l'apprentissage du réseau car il peut avoir piégé des espèces ne peuplant pas réellement les champs. Nous ne considérerons donc pas ces données.

5.2.3 Distribution des abondances et seuil de présence

La méthode d'apprentissage demande des données de présence/absence des espèces et non d'abondance. Pour transformer les données d'abondance disponibles en données de présence/absence, une solution possible est de considérer une espèce comme présente dès que son abondance est supérieure à 0 lors d'une observation dans une parcelle donnée. Il est également possible de monter le seuil, afin de considérer comme absente une espèce dont l'abondance est très faible à un moment d'observation dans une parcelle donnée alors que son abondance est très importante d'habitude. Une représentation des abondances de chaque espèce aiderait à définir une méthode. Le diagramme représenté en figure 5.4 est un histogramme des comptages individuels (sans distinction d'espèce, de champs ou de piège, à l'aide d'un vecteur contenant l'ensemble de ces données) supérieurs à 0. La distribution des abondances semble classique pour des données de ce type : il y a plus d'abondances faibles que d'abondances élevées. Il semble y avoir assez peu d'abondances très élevées. Choisir un seuil supérieur à 0 pour changer les données d'abondance en données de présence/absence risque de conduire à beaucoup d'absences, et pourrait ne pas être très pertinent. Le problème des

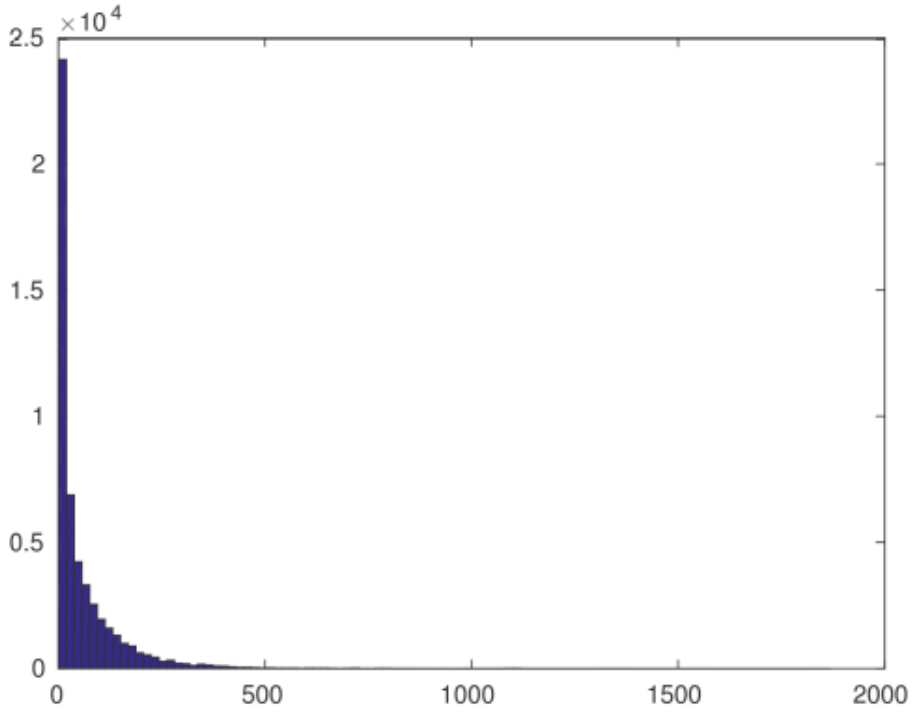


FIGURE 5.4 – Distribution des abondances supérieures à 0, sans distinction d'espèce, de champ ou de piège.

données homogènes se pose moins que dans les données PISCO. En effet, les nombreux jeux de données différents (un par parcelle) atténuent les risques d'homogénéité des données. Les données de présence/absence des espèces d'arthropodes sont donc converties de la façon suivante : si au moins un individu de l'espèce i a été observée dans une parcelle s à un moment t, $\overset{s}{X}{}_i^t = 1$, $\overset{s}{X}{}_i^t = 0$ sinon.

5.2.4 Données de présence/absence utilisées pour l'apprentissage

Les données que nous utiliserons pour apprendre la structure du réseau écologique associé à ces données sont donc les données de présence/absence obtenues à partir des sommes des abondances des pièges situés à 2m et à 32m pour chaque espèce et chaque parcelle. Dans ces données se trouvaient également des informations sur des espèces d'oiseaux (et non d'arthropodes). Ces données concernant les oiseaux sont en fait des données construites à partir des autres données d'abondance d'arthropodes et non pas des données observées. Les données concernant les oiseaux ont donc été écartées.

Les autres espèces ne sont pas non plus toutes observées partout. Afin d'avoir des données cohérentes, nous souhaitons éviter d'inclure les espèces qui n'ont jamais été observés. Nous faisons le choix de considérer un jeu de données par type de culture, et de n'inclure dans ces jeux de données que les espèces ayant été observées au moins une fois dans au moins une parcelle de cette culture. Cela donne 4 différents jeux de données, tous ayant 2 pas de temps observés, mais chacun ayant un nombre d'espèces différents :

— 41 espèces différentes sont observées dans les cultures de betterave.

— 29 espèces différentes sont observées dans les cultures de maïs.

— 40 espèces différentes sont observées dans les cultures de colza d'été.

— 29 espèces différentes sont observées dans les cultures de colza d'hiver.

Décrivons alors la procédure d'apprentissage des réseaux trophiques associés à ces données.

5.3 Apprentissage de réseau

5.3.1 Modélisation

L'apprentissage des réseaux écologiques associés aux arthropodes observés se fait par un modèle de réseau bayésien dynamique étiqueté, la dynamique est mesurée sur deux pas de temps : early ($t = 0$) et late ($t = 1$). La mesure de la présence ou de l'absence d'une espèce i à un instant t sur une parcelle s est considérée comme la réalisation d'une variable aléatoire $\overset{s}{X}{}_i^t$ dont la probabilité découle d'un modèle de réseau bayésien dynamique étiqueté où ε décrit la probabilité d'apparition ou de survie spontanée, ρ décrit la probabilité de réussite d'une interaction positive (impulsion), τ la probabilité de réussite d'une interaction négative (inhibition). $\pi_i = \overset{q}{\pi}_i \cup \overset{r}{\pi}_i$ décrit l'ensemble des parents de i, c'est à dire l'ensemble des espèces ayant une influence positive q ou négative r sur i. Ces parents sont les mêmes pour toutes les parcelles. Π_i^t désigne l'ensemble des variables aléatoires X_j^t telles que $j \in \pi_i$. $\overset{q}{N}{}_i^t$ et $\overset{r}{N}{}_i^t$ décrivent le nombre de parents dont l'influence est respectivement positive et négative sur i présents à t. u_i différencie les espèces basales des espèces non basales : si i est basale, c'est à dire qu'elle n'a pas de parents à influence positive, soit $\overset{q}{\pi}_i = \emptyset$, alors $u_i = 0$; sinon, $u_i = 1$. Ce modèle fixe la quantité $\varepsilon_0^{sur} = 1$ et $\varepsilon_1^{sur} = 0$ afin de gérer le fait qu'une espèce basale n'a pas besoin d'influences positives pour survivre. La probabilité d'apparition d'une espèce

est indépendante de ses parents et décrite par $P\left(\overset{s}{X}_i^{t+1} = 1 | \overset{s}{X}_i^t = 0\right) = \varepsilon$ et la probabilité de survie est décrite dans l'équation 5.1.

$$P\left(\overset{s}{X}_i^{t+1} = 1 | \overset{s}{X}_i^t = 1, \Pi_i^t\right) = \left(\varepsilon_{u_i}^{sur} + (1 - \varepsilon_{u_i}^{sur}) \cdot \left(1 - (1 - \rho^{sur})^{\overset{q}{N}_i^t}\right)\right) \cdot (1 - \tau^{sur})^{\overset{r}{N}_i^t} \tag{5.1}$$

L'apprentissage de la structure de ce réseau bayésien dynamique étiqueté consiste à trouver pour chaque espèce son ensemble de parents le plus vraisemblable sachant les observations de sa présence/absence sur les deux pas de temps, à l'aide de la procédure décrite dans le chapitre 4.

5.3.2 Apprentissage par graphe consensus

Un graphe dans lequel chaque espèce a 12 parents au maximum est appris par parcelle, et l'agrégation des graphes de toutes les parcelles d'une certaine culture donne un graphe consensus relatif à cette culture. Ces graphes sont représentés dans la figure 5.5 sous forme de heatmap, chaque lien étant coloré selon le nombre de parcelles dans lequel il a été appris. On remarque que certaines espèces ont des interactions avec beaucoup d'autres espèces. Le tableau 5.2 identifie les espèces dont le plus d'arcs ont été reconstitués. Par exemple, en moyenne, 19,13 interactions négatives de *Diptera adults* vers une autre espèce dans les cultures de betterave ont été apprises. Cette espèce a donc été identifiée comme prédatrice ou compétitrice de 19,13 espèces en moyenne pour cette culture, et elle a été identifiée comme proie ou facilitatrice de 26,15 espèces en moyenne pour cette culture.

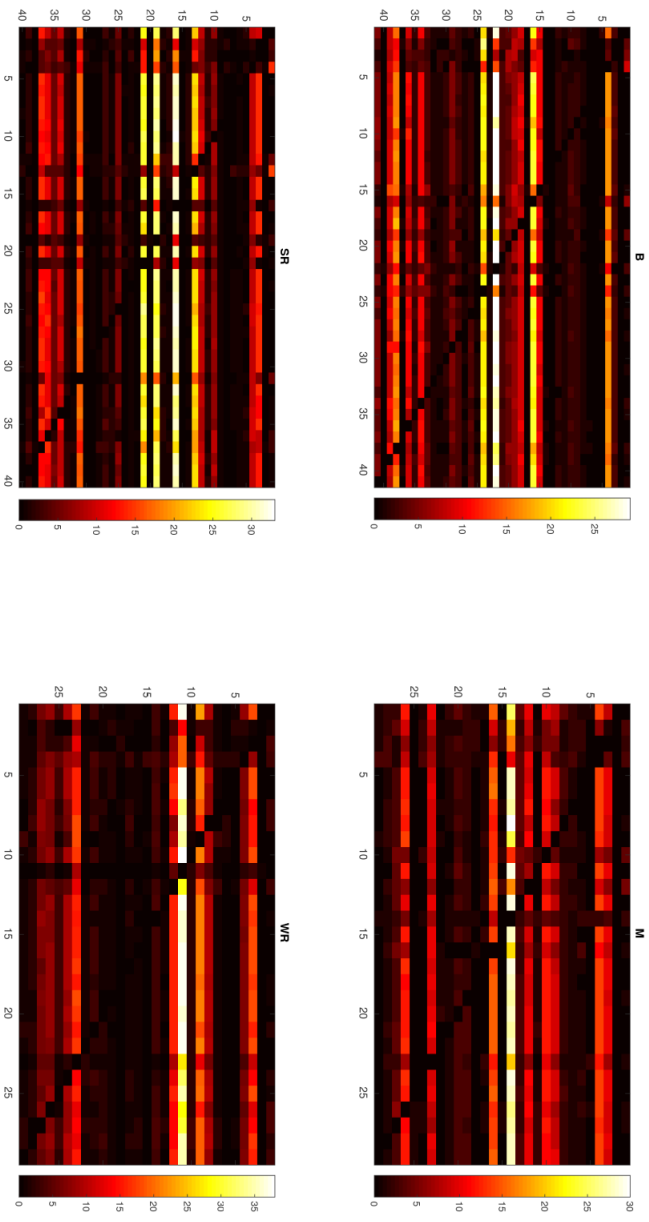


FIGURE 5.5 – Représentation par heatmap des graphes consensus appris pour chaque culture

Espèce	Moyenne du nombre d'arcs partant de l'espèce nommée reconstituées								Total
	B		M		SR		WR		
	négative	positive	négative	positive	négative	positive	négative	positive	
Diptera adults	19,13	26,15	18,11	24,89	25,67	25,03	16,50	32,93	23,61
Parasitica	33,68	20,38	23,57	13,43	26,92	22,72	28,82	13,43	23,40
Other Coleoptera	18,55	6,18	18,71	8,93	21,62	27,31	22,71	18,04	17,82
Aphidoidea	14,00	20,50	13,54	10,39	18,90	19,77	19,32	8,36	16,05
Sminthuridae	19,45	15,45	14,75	12,56	25,31	8,49	24,25	6,07	16,04
Entomobryidae	25,30	3,05	21,64	8,43	21,77	11,63	17,57	14,93	15,62
Staphylinidae adult	9,05	9,08	14,93	8,86	23,26	15,13	23,93	14,89	14,72
Araneae	16,15	14,95	16,57	10,54	17,87	10,10	10,82	6,64	13,27
Auchenorhyncha	16,43	9,18	10,64	7,46	17,77	8,33	10,18	1,26	10,66
Linyphiidae	12,73	8,68	11,61	3,29	8,33	11,28	13,04	5,89	9,51
Curculionidae	7,03	8,65	2,04	1,00	9,00	8,72	10,43	7,59	7,66
Isotomidae	6,10		7,07		11,23	1,00	2,54	1,00	6,75
Coleoptera	9,05	8,53			2,34	1,00			6,62
Cimicidae nymphs	8,30	2,33	5,46	1,32	6,38	6,51			5,24
Heteroptera	5,03	1,20							4,60
Diptera	9,05	4,38			2,21	1,41			4,40

TABLE 5.2 – Tableau renseignant le nombre moyen d'arcs dans les graphes appris dans chaque parcelle en excluant les espèces d'oiseaux.

5.3.3 Intégration des tailles des individus comme niveau trophique et apprentissage par vraisemblance généralisée

Afin d'affiner l'apprentissage, nous souhaitons utiliser les informations sur la taille des espèces. Les mesures de longueur et de poids moyen de plusieurs espèces permettent d'établir une « classe » de taille de ces espèces, c'est à dire de regrouper les espèces ayant des tailles similaires (petites, grandes etc.). Nous considérons 4 classes de taille différentes, de 1 (petites) à 4 (grandes). Plusieurs espèces n'ont pas de taille bien identifiée. L'apprentissage de structure de réseau par réseau Bayésien étiqueté permet d'intégrer des connaissances sur une structure du graphe en communautés distinctes. La connaissance de la structure des communautés permet de prioriser certains arcs par rapport à d'autres. Dans le cas d'un réseau trophique, ces communautés peuvent s'appliquer à la taille des individus, que l'on peut interpréter sous la forme de niveau trophique. On peut alors changer les probabilités $P(Gij|ci, cj)$ de présence de certains arcs de la façon suivante :

— Aucune espèce ne peut avoir d'influence positive sur une espèce de classe de taille inférieure, afin d'éviter qu'une espèce puisse être la proie d'une espèce plus petite : $P(G^+_{ij}|c_i > c_j) = 0$.

— Une espèce a plus de chances d'avoir une influence positive vers une espèce d'une classe de taille proche de la sienne. L'intérêt est de voir apparaître des niveaux trophiques dans la structure du graphe, et une espèce aura plus souvent pour proie des espèces du niveau trophique directement inférieur. Par exemple, une espèce de classe 1 aura plus de chance d'avoir une influence positive pour une espèce de classe 2 que pour une espèce de classe 3, et de classe 4 : $P(G^+_{ij}|c_j - c_i = 1) > P(G^+_{ij}|c_j - c_i = 2) > P(G^+_{ij}|c_j - c_i = 3)\ldots$

— Une espèce a plus de chance d'avoir une influence négative sur une espèce d'une classe de taille directement inférieure à la sienne. Cela priorise les relations trophiques vers des niveaux trophiques proches. Une espèce peut avoir une influence négative sur n'importe quelle autre espèce, mais avec une probabilité moindre : $P(G_{ij} + |c_i - c_j = 1) > P(G_{ij} + |c_j - c_i \neq 1)$.

5.3.4 Apprentissage par vraisemblance généralisée et connaissance des niveaux trophiques

L'agrégation de plusieurs graphes pose problème. Il est en effet difficile de construire un seul graphe à partir d'une agrégation de plusieurs graphes. De plus, cette méthode d'agrégation n'est pas très solide d'un point de vue théorique. L'apprentissage de la structure du réseau en prenant en compte les tailles des espèces se fait donc en intégrant les données d'observation d'espèces de tous les champs d'une même culture. Un seul graphe est appris : celui qui maximise la somme des vraisemblances de tous les champs. La forme des graphes peut être visualisée dans la figure 5.6, où chaque point représente une espèce, placé au même endroit pour toutes les cultures. Les espèces absentes d'un type de culture ont un numéro coloré en rouge dans la représentation du graphe correspondant à cette culture. Les réseaux obtenus semblent très différents pour les 4 cultures différentes, bien que certains arcs soient communs à plusieurs cultures.

5.4 Différences entre cultures

Puisque nous avions des données séparées par culture, nous avons décidé d'apprendre un réseau par culture. Il est intéressant de comparer les réseaux appris pour les différentes cultures afin de savoir s'il existe des interactions propres à chaque culture ou si les interactions entre espèces sont

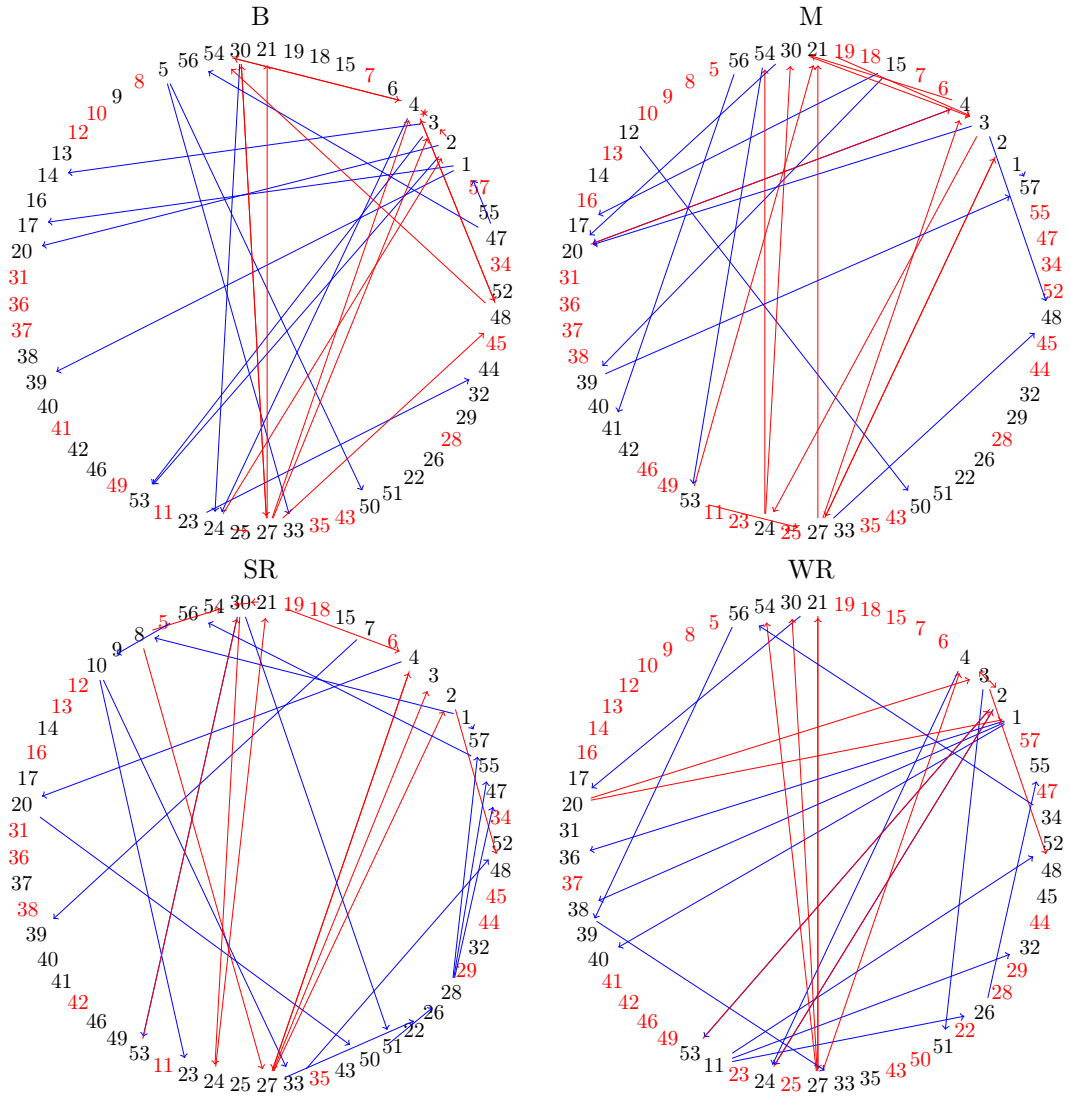


FIGURE 5.6 – Représentation des graphes appris par la méthode d'apprentissage de réseau bayésien dynamique étiqueté pour chaque culture. Les arcs en rouge représentent des influences positives et les arcs en bleus des influences négatives. Le numéro associé à chaque nœud correspond aux numéros associés aux espèces du Tableau 3. Les nœuds colorés en rouge correspondent aux espèces n'ayant été observées dans aucune parcelle de la culture associée au graphe.

Graph	B	M	SR	WR	Global
B	79 (0)	10 (14)	12 (8)	7 (13)	52 (0)
M	10 (14)	46 (0)	8 (3)	10 (7)	12 (24)
SR	12 (8)	8 (3)	29 (0)	9 (1)	13 (10)
WR	7 (13)	10 (7)	9 (1)	55 (0)	9 (21)
Global	52 (0)	12 (24)	13 (10)	9 (21)	83 (0)

TABLE 5.3 – Nombre d'arcs similaires entre les réseau appris pour chaque culture. Les chiffres en noir indiquent le nombre d'arcs présents entre une même paires d'espèces, de même orientation et de même étiquette entre les graphes appris pour deux cultures. Les chiffres en rouge indiquent le nombre d'arcs présents entre une même paires d'espèces, de même orientation mais d'étiquette différente entre les graphes appris pour deux cultures.

universelles. Pour répondre à cette question, un graphe appris à partir de toutes les parcelles sans distinction de culture est intéressant à apprendre car ce *graphe global* correspond à un graphe moyen dans le cas où les interactions étaient communes à toutes les cultures. À l'œil, sur la figure 5.6, les réseaux semblent différents. Ces différences sont confirmées par le tableau 5.3, où les chiffres en noir indiquent le nombre d'arcs similaires (entre une même paire d'espèces et de même orientation) et de même étiquette entre les graphes de deux cultures, et les chiffres en rouge indiquent le nombre d'arcs similaires mais d'étiquette différente. Dans ce tableau, la culture "Global" indique les données correspondantes à l'ensemble des parcelles sans distinction de culture et au graphe appris à l'aide de ces données. La diagonale de ce tableau compare chaque graphe avec lui-même, et indique donc le nombre d'arcs dans ce graphe. On remarque que les graphes associés à chaque culture sont très différents en terme d'arcs, car entre deux cultures différentes, on ne trouve jamais plus de 15 arcs similaires de même étiquette, et souvent un nombre équivalent d'arcs similaires d'étiquette différente. Par exemple, on trouve 24 arcs entre les mêmes paires d'espèces et de même orientation entre les graphes appris pour la betterave et le maïs, mais seuls 10 ont la même étiquette, les 14 autres ont une étiquette différente. Il y a tout de même une exception : le graphe correspondant à la culture de betterave compte 52 arcs entre les mêmes paires d'espèces, de même orientation et de même étiquette que le graphe correspondant à toutes les cultures. Il est possible que la culture de betterave contienne des espèces d'arthropodes dont les interactions peuvent se retrouver dans n'importe quelle autre culture, mais que dans chacune des autres cultures, il existe des interactions ou des espèces très spécifiques. Une autre interprétation possible est le fait que les interactions entre les espèces dans les parcelles de betteraves soient plus "fortes", plus déterminantes sur les dynamiques temporelles des espèces et plus stables que pour les autres cultures. De ce fait, ces interactions sont les mieux captées dans le graphe global. Ces différences sont-elles dues à de vraies différences d'interactions entre les cultures ou au fait que chacun de ces graphes est suffisamment générique pour pouvoir expliquer les données observées pour n'importe quelle culture ? Pour cela, nous avons calculé la vraisemblance des données correspondant à chaque culture par rapport à un modèle de réseau bayésien étiqueté dont la structure du graphe correspond au graphe appris pour une autre culture. Par exemple, pour une culture C_1, un graphe G_1 et des paramètres θ_1 ont été appris. Nous calculons la vraisemblance des données issues de C_1 sachant les paramètres θ_1 et un graphe G_2 correspondant à un autre culture C_2. Le tableau 5.4 indique les résultats de ces calculs de vraisemblance pour les données de chaque culture (en ligne) par rapport aux graphes de chaque culture (en colonne). Il est attendu que les valeurs de vraisemblance situées sur la diagonale de ce tableau soient les plus élevées, car elles

	Graphe				
Data	B	M	SR	WR	Global
B	**-2526.23**	**-2555.66**	-3476.53	-3858.50	-2706.41
M	-2177.56	**-1063.42**	-2184.20	-1983.35	-1898.98
SR	-2168.26	-2228.02	**-1659.00**	-2415.70	-2196.71
WR	-2099.49	-1421.53	-1680.65	**-768.55**	-1701.80
Global	-5170.31	-5148.71	-5622.41	-5807.81	**-4890.77**

TABLE 5.4 – Log-vraisemblance des données d'une culture (en ligne) sachant le graphe appris pour une autre culture (en colonne). En gras : vraisemblances les plus élevées pour chaque ligne.

correspondent aux vraisemblances d'une culture par rapport au graphe appris pour la même culture. C'est ce que l'on constate. Cependant, la vraisemblance des données correspondant aux parcelles de betterave par rapport au graphe appris sur ces données est assez proche de celle calculée par rapport au graphe appris avec les données correspondant aux parcelles de maïs. Cela pourrait s'expliquer par la possible généricité des interactions présentes dans les cultures de betterave.

5.5 Comparaison avec les précédents travaux

[Tamaddoni-Nezhad et al., 2013] ont utilisé les abondances renseignées dans ce jeu de données, ainsi que d'autres informations sur les espèces d'arthropodes afin d'apprendre les interactions entre ces espèces à l'aide d'une méthode basée sur la programmation logique (voir section 1.4.1). Le but est de chercher les faits de type "*L'espèce i mange l'espèce j*" à partir des informations de l'abondance de i et de j pour une parcelle donnée, les tailles relatives de i et de j, et l'information sur le régime alimentaire de i (ex : détritivore, herbivore, carnivore...) à l'aide des règles exposées dans la section 1.4.1. L'apprentissage de ces faits par abduction se fait par validation croisée à l'aide de programmation linéaire en nombres entiers, et l'évaluation de la qualité d'une interaction apprise de la sorte est faite par une *Estimation de fréquence d'hypothèse (HFE)* [Tamaddoni-Nezhad et al., 2012]. La validation croisée divise les données écologiques en deux parties : des données d'apprentissage servant à définir les règles, et des données de tests, sur lesquelles on applique ces règles [1]. La fréquence *HFE* d'une interaction est basée sur le nombre de fois où cette interaction a été apprise sur plusieurs essais d'apprentissage où des permutations ont eu lieu sur les données d'apprentissage. Afin de vérifier la cohérence des interactions apprises entre ces espèces, une vérification par fouille de texte (*text mining*) sur la littérature scientifique a été réalisée. Le principe est de rechercher dans la littérature scientifique issue de bases de données locales ou de moteurs de recherches (ex : google scholar) des indices de présence de chaque interaction apprise par méthode logique. Un indice de la présence d'une interaction trophique d'une espèce i sur une espèce j dans une publication scientifique est l'existence, au sein de cette publication de :

— La mention de l'espèce i

— La mention de l'espèce j

— L'existence de termes issu du lexique des interactions trophiques ("manger", "se nourrir", "proie", "consommer"...)

1. Plus précisément, la validation croisée consiste à effectuer plusieurs apprentissages, en changeant à chaque essai les données d'apprentissage et de test.

Le nombre de publications vérifiant ces critères (désigné comme *Litterature hit* ou LH) pour une interaction apprise entre deux espèces donne un score indiquant le degré de confiance que l'on peut accorder à l'existence de ce lien. Ce score LH reste peu précis, et il est plus efficace de mesurer le degré de confiance par un autre score dit ratio de fouille de texte TMR (*Text mining ratio*) calculé comme le score LH divisé par la somme des mentions de toute autre espèce dans les sources utilisées. Plus le score TMR est proche de 1, plus ces espèces semblent être spécifiquement décrites dans la littérature comme ayant une interaction trophique. Ces méthodes apprennent des interactions trophiques uniquement. Cependant, il est possible de comparer avec nos résultats, si on considère qu'une interaction trophique correspond à une influence positive de la proie vers le prédateur et une influence négative du prédateur vers la proie. Cette comparaison se fait sur les graphes appris pour chaque culture par réseau bayésien étiqueté. En réalité, assez peu d'interactions "doubles" entre deux espèces i et j ont été apprises par réseau bayésien dynamique étiqueté, c'est à dire des interactions représentées par un arc de i vers j et un autre arc de j vers i. Pour comparer le graphe appris par réseau bayésien dynamique étiqueté et celui appris par programmation logique, il est donc possible de repérer les interactions trophiques apprises entre une proie i et un prédateur j par programmation logique et les interactions positives de i vers j et/ou négatives de j vers i apprises par réseau bayésien dynamique étiqueté. Il y a cependant assez peu d'interactions similaires apprises par les deux méthodes. Les interactions apprises par les deux méthodes sont les suivantes :

— *Aphidoidea* est désigné comme proie de *Bembidion guttula* et de *Araneae* dans le graphe appris par programmation logique, avec dans les deux cas un score HFE faible (<10%). Des influences négatives de *Bembidion guttula* et *Araneae* vers *Aphidoidea* ont été apprises par réseau bayésien étiqueté.

— *Entombryidae* est désigné comme proie de *Coccinelid larvae* ($HFE = 100\%, LH = 2, TMR < 10^{-3}$) et de *Bembidion guttula* ($HFE \simeq 70\%, LH = 2, TMR < 10^{-3}$) dans le graphe appris par programmation logique. Une influence positive de *Entombryidae* vers *Bembidion guttula* et une influence positive de *Entombryidae* vers *Coccinelid larvae* ont été apprises par réseau bayésien étiqueté.

— *Sminthuridae* est désigné comme proie de *Bembidion guttula* ($HFE = 100\%, LH = 1, TMR < 10^{-3}$) et de *Notiophilus biguttatus* ($HFE = 100\%, LH = 7, TMR < 10^{-2}$) dans le graphe appris par programmation logique. Des influences positives de *Sminthuridae* vers *Bembidion guttula* et *Notiophilus biguttatus* ainsi qu'une influence négative de *Bembidion guttula* vers *Sminthuridae* ont été apprises par réseau bayésien étiqueté.

Dans le réseau appris, on remarque par ailleurs que quelques espèces (*Araneae, Bembidion guttula, Diptera adults, Entomobryidae, Isotomidae*) ont énormément d'influences négatives (plus de 7) vers d'autres espèces, alors que la plupart des autres n'ont que peu d'influences. Ce sont d'ailleurs des interactions issues de ces espèces qui sont apprises par les deux méthodes (méthode logique et RBDE). Ce phénomène n'apparaît pas dans le réseau appris par méthode logique. Vu les différences que l'on peut trouver entre les interactions apprises par réseau bayésien étiqueté et celles apprises par méthode logique ou par text-mining, la question de la qualité de l'apprentissage par réseau bayésien dynamique étiqueté se pose. Cette méthode d'apprentissage est-elle peu efficace, ou est-ce que ces deux réseaux ne sont pas simplement incomparables car ils n'apprennent pas la même chose ? La section suivante présente un bilan de ces résultats, et un début d'interprétation biologique.

5.6 Bilan et interprétation biologique

Il est possible d'apprendre des réseaux écologiques à partir de données de présence/absence d'espèces en exploitant la dynamique temporelle des espèces. Si cela permet de repérer d'importantes différences entre les espèces de chaque culture et leurs interactions probables, les réseaux appris sont très différents de ceux obtenus par d'autres méthodes issues des mêmes données. Cependant, même si le jeu de données est le même au départ, on trouve une différence importante entre celles utilisées pour la programmation logique et pour le réseau bayésien dynamique étiqueté. En effet, la programmation logique utilise des données d'abondances d'espèces combinées au régime alimentaire et aux tailles de chaque espèce et ne tient compte ni des différentes cultures ni des différentes dates d'observation. L'apprentissage par réseau bayésien dynamique étiqueté n'utilise que l'information de présence ou d'absence des espèces, leur taille et l'information de dynamique entre les deux dates d'observations. Le fait que l'on utilise si peu d'information peut faire que l'on n'apprend pas vraiment directement la structure d'un réseau écologique, mais qu'il faut plutôt interpréter le réseau appris comme un "réseau probable", montrant des interactions écologiques soupçonnées.

Un tel réseau pourrait être utile si l'on étudie un écosystème peu connu, pour lequel on a peu de données. Ce réseau probable faciliterait alors le travail des écologues trophiques qui souhaiteraient connaître les interactions entre les espèces de cet écosystème. Ils pourraient en effet dans un premier temps se concentrer sur les interactions probables apprises par réseau bayésien étiqueté, et en vérifier la validité.

Si il est possible d'intégrer des connaissances expertes à priori, pendant l'apprentissage du réseau, il peut également être utile d'en intégrer d'autres à posteriori, une fois le réseau probable appris. Ces connaissances pourraient servir à classer les arcs selon la confiance que l'on peut accorder à l'existence des interactions qu'ils représentent. Cela est déjà possible à l'aide des données dont on dispose pour cette étude sur les arthropodes. En effet, si nous n'avions pas inclus les informations d'abondance des espèces pour l'apprentissage du réseau, il est possible de les utiliser à posteriori afin de trouver les arcs correspondant aux interactions entre les espèces les plus abondantes. L'hypothèse étant que les espèces les plus abondantes sont celles qui influent le plus sur la dynamique des autres espèces de l'écosystème. Ainsi, il est probable que les arcs entre deux espèces très abondantes représentent bien leurs interactions, alors que les interactions entre des espèces moins abondantes pourraient être perturbées par les espèces très abondantes. Il est donc moins probable que les arcs incluant des espèces peu abondantes correspondent à des interactions réalistes.

Conclusion

Cette thèse présente des travaux dans le domaine de l'intelligence artificielle motivés par une application en écologie trophique. L'objectif est d'apprendre la structure d'un réseau écologique, c'est à dire un ensemble d'interactions entre les espèces vivantes d'un écosystème donné, en utilisant des données temporelles de présence/absence d'espèces et quelques connaissances expertes.

Bilan

Notre approche pour inférer un réseau écologique à partir de données temporelles de présence/absence d'espèces a été de définir un nouveau cadre méthodologique : les réseaux bayésiens étiquetés. Ce modèle est un cas particulier de réseau bayésien dans lequel les arcs du graphe portent une information qualitative appelée étiquette. Dans un tel modèle, le nombre de paramètres est nettement réduit, chaque probabilité étant fonction d'un petit nombre de paramètres partagés entre les variables. La probabilité associée à une variable dépend uniquement de son nombre de parents présents et des étiquettes des arcs provenant de ces parents. Il est possible de caractériser un réseau bayésien étiqueté comme un modèle utilisant des portes Noisy-OR [Diez, 1993] et Noisy-AND contraintes. Un réseau bayésien étiqueté est utile pour modéliser tout phénomène interprétable comme un processus de contact, comme la transmission d'une information, d'une maladie ou d'une perturbation au sein d'un réseau. L'intérêt principal d'un tel modèle est qu'il ne compte qu'un nombre restreint de paramètres, et que ce nombre de paramètres est indépendant de la structure du graphe lorsque l'on connaît son nombre d'étiquettes. Un autre intérêt de ce modèle est la possibilité d'intégrer facilement de la connaissance experte en fixant des valeurs de paramètres ou d'autres éléments de ce modèle afin de réduire encore le nombre de paramètres inconnus. Cependant, cette baisse du nombre de paramètres se fait au détriment d'une formule de la vraisemblance facilement dérivable.

Nous avons également défini une méthode permettant d'apprendre la structure étiquetée du graphe associé à un tel modèle. Il s'agit d'un algorithme glouton alternant deux étapes jusqu'à convergence. L'étape dite d'estimation estime la valeur des paramètres à structure de graphe fixée par maximisation de vraisemblance. Le nombre restreint de paramètres rend cette étape très peu coûteuse. L'étape dite de restauration apprend la structure de graphe la plus vraisemblable à paramètres fixés. Cette étape se fait par la résolution d'un programme linéaire en nombres entiers. La structure est complexe à apprendre, ce programme linéaire pouvant être de grande taille. Les avantages de résoudre cette étape par programmation linéaire sont principalement la possibilité d'utiliser des d'outils de résolution existants [Schrijver, 1998], et la possibilité d'exprimer de la connaissance experte sous forme de contraintes linéaires. Un type de connaissance que nous avons intégré à ce problème de programmation linéaire se fait sous forme de priors sur les arcs du graphe à

l'aide d'un modèle stochastique à blocs afin de distinguer, dans le cadre de l'apprentissage de réseaux écologiques, les niveaux trophiques des espèces. Les résultats sur des données simulées montrent que l'apprentissage de cette structure est plus efficace lorsque l'on dispose de plusieurs jeux de données indépendants. Intégrer des connaissances expertes dans le processus d'apprentissage améliore nettement sa qualité. En revanche, sans jeux de données indépendants ni connaissances expertes, il est rare que les données de présence/absence temporelle contiennent suffisamment d'information pour apprendre un réseau écologique cohérent.

Cette méthode a été testée pour apprendre la structure de réseaux écologiques à partir de données réelles de présence/absence d'espèces d'arthropodes dans des champs expérimentaux. Cette méthode permet de retrouver des interactions semblant cohérentes pour les écologues, et de constater des différences entre les réseaux écologiques de différentes cultures. Cependant, plusieurs interactions connues ne sont pas apprises, et certaines interactions apprises ne semblent pas correspondre à des interactions réelles. L'utilisation de données binaires est sans doute un facteur limitant, les informations que l'on peut extraire de ces données étant limitées. Il est probable que cette méthode capte davantage des corrélations entre espèces que des causalités, rendant difficile de distinguer des dynamiques d'espèces dues à des interactions directes de celles dues à des interactions indirectes ou à des facteurs externes aux interactions entre espèces. Un réseau appris par cette méthode à partir de données binaires serait alors un réseau d'interactions probables. Ce réseau permettrait aux écologues d'avoir une idée globale, bien qu'imprécise, des espèces semblant évoluer conjointement. Cela pourrait servir de base pour guider une étude plus approfondie des interactions écologiques, en indiquant des espèces pouvant avoir un impact sur l'écosystème. Ce travail de facilitation pourrait être utile en cas d'étude d'un écosystème mal connu, dans lequel les observations précises sont difficiles ou coûteuses.

Perspectives

Si le cadre méthodologique des réseaux bayésiens étiqueté a été posé, il y a encore eu assez peu d'essais sur l'efficacité de la méthode d'apprentissage. Une première étape sera donc d'effectuer des essais sur données simulées pour d'autres modèles de réseaux bayésiens étiquetés que celui des réseaux écologiques. À court terme, un package Matlab sera proposé afin de diffuser cette méthode et de permettre à d'autres chercheurs de s'en emparer sur d'autres applications. Il serait notamment intéressant d'utiliser ce type de modèles dans des cadres différents de l'apprentissage de réseaux écologiques. Par exemple, un réseau bayésien étiqueté pourrait convenir pour modéliser la diffusion d'une information ou d'une rumeur au sein d'un réseau social [Nekovee et al., 2007]. Dans ce cas, le cadre des réseaux bayésiens étiquetés pourrait aider à comprendre le mode de diffusion de l'information, ou à apprendre un réseau social mal connu à partir de la connaissance des individus ayant reçu l'information.

Le cadre des réseaux bayésiens étiqueté pourrait également à moyen terme être davantage généralisé. En effet, ce cadre n'est défini pour l'instant qu'avec des variables binaires et dans sa définition actuelle pour une variable i, un parent j de i ($X_j = 0$) est absent, l'arc correspondant n'est pas actif, et n'a pas d'impact sur X_i. Il serait intéressant d'intégrer la possibilité de modéliser l'impact d'un arc correspondant à un parent absent. Plus généralement, cela pourrait revenir à considérer le cas où les variables du modèle ne sont pas binaires, mais peuvent prendre plusieurs valeurs. Dans un réseau bayésien étiqueté multivarié, l'étiquette d'un arc de j vers i aurait un impact différent pour chaque valeur possible de X_j. Dans le cadre d'un réseau bayésien dynamique étiqueté multivarié, l'étiquette d'un arc de j vers i aurait un impact différent sur X_i^{t+1} pour chaque valeur

différente de X_j^t et de X_i^t.

Par ailleurs, la méthode d'apprentissage par programmation linéaire en nombres entiers n'est peut-être pas la plus optimale. D'autres méthodes d'apprentissage pourraient voir le jour sur ce type de réseau bayésien. Le premier ajustement possible pour l'apprentissage de réseau serait d'utiliser la relaxation du programme linéaire en nombres entiers [Jaakkola et al., 2010], c'est à dire autoriser les variables constituant le programme linéaire à prendre des valeurs réelles. De plus, si les scores pénalisant la vraisemblance par le nombre de paramètres sont inutiles dans le cadre des réseaux bayésiens étiquetés (car le nombre de paramètres est fixe et indépendant de la structure), la vraisemblance n'est pas nécessairement le score le plus approprié car il ne privilégie pas nécessairement une structure de graphe simple. Établir un score permettant de pénaliser la complexité du graphe sans se baser sur le nombre de paramètres pourrait être utile. En effet, la vraisemblance non pénalisée pour l'apprentissage de structure de réseau bayésien étiqueté n'a pas toutes les propriétés d'un score classique, notamment le fait d'avoir un paramètre de pénalisation sur lequel se base l'algorithme branch and bound [Suzuki, 1996] ou l'algorithme de [Dojer, 2006]. Parmi les autres contraintes induites par la méthode d'apprentissage présentée ici, il y a le fait que l'on considère des données complètes. Il serait utile d'adapter au cas étiqueté des méthodes classiques adaptées au cas de données incomplètes ou de variables non observables, comme l'algorithme E-M [Dempster et al., 1977]. D'autres types de connaissances expertes pourraient être ajoutées dans le processus d'apprentissage. Une approche possible serait de rendre l'algorithme d'apprentissage plus interactif [Cano et al., 2011], en ajoutant la connaissance de certains arcs pendant le processus d'apprentissage [Masegosa and Moral, 2013]. Ainsi, à chaque graphe proposé, un expert pourra en évaluer la pertinence et éventuellement relancer la procédure d'apprentissage en tenant compte de cette analyse en fixant l'absence ou la présence de certains arcs.

L'utilisation des réseaux bayésiens étiquetés pour l'inférence de réseaux écologiques gagnerait à être encore discutée et testée. Le réseau obtenu semble pouvoir être un réseau d'interactions probables entre espèces. Il serait intéressant de vérifier si, dans un écosystème mal connu, les interactions apprises par apprentissage de réseau bayésien étiqueté aident à l'établissement d'un réseau écologique cohérent. De plus, les interactions actuellement apprises semblent surtout être des mesures de corrélation, plus que de causalité. Pour l'instant, cette méthode ne permet pas d'identifier clairement la nature des interactions apprises. Des informations écologiques ou des données supplémentaires pourraient aider à mieux identifier la nature des arcs appris. Selon la nature des interactions apprises, il serait intéressant de discuter sur l'utilisation que l'on peut faire des réseaux écologiques obtenus.

Une idée d'utilisation de l'apprentissage de réseau écologique par réseau bayésien dynamique étiqueté est son utilisation dans le cadre de la gestion adaptative [Williams and Brown, 2016]. La gestion adaptative consiste à trouver un moyen optimal de gérer des ressources à l'aide d'informations initialement mal connues, que l'on cherche à apprendre conjointement avec la gestion. La gestion de ressources naturelles comme la biodiversité peut se faire à l'aide de la connaissance des réseaux écologiques [Kitchell, 2012]. Cependant, puisque le réseau écologique est souvent mal connu, une idée serait d'utiliser les techniques de gestion adaptative afin de conjointement gérer les ressources naturelles tout en apprenant la structure de réseau la plus plausible. Le but est de rechercher une stratégie permettant de distribuer les ressources alloués aux différentes actions de gestion et d'apprentissage afin d'avoir un compromis entre le bon apprentissage, la gestion de la biodiversité, et un coût raisonnable. Dans ce cas, l'apprentissage par renforcement [Sutton et al., 1998] dans le cadre des processus décisionnels de Markov [Puterman, 2014], dont le but est d'étudier les conséquences probables d'un ensemble de décisions afin trouver une politique de décision optimale pourrait être intéressant à explorer.

Bibliographie

[Abbe, 2017] Abbe, E. (2017). Community detection and stochastic block models : recent developments. *arXiv preprint arXiv :1703.10146*.

[Agmon, 1954] Agmon, S. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6(3) :382–392.

[Akaike, 1970] Akaike, H. (1970). Statistical predictor identification. *Annals of the institute of Statistical Mathematics*, 22(1) :203–217.

[Allesina, 2008] Allesina, S. (2008). A general model for food web structure. *science*, 1156269(658) :320.

[Allesina and Pascual, 2009] Allesina, S. and Pascual, M. (2009). Food web models : a plea for groups. *Ecology letters*, 12(7) :652–662.

[Antonucci, 2011] Antonucci, A. (2011). The imprecise noisy-or gate. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–7. IEEE.

[Auclair et al., 2017] Auclair, E., Peyrard, N., and Sabbadin, R. (2017). Labeled dbn learning with community structure knowledge. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 158–174. Springer.

[Banašek-Richter et al., 2004] Banašek-Richter, C., Cattin, M.-F., and Bersier, L.-F. (2004). Sampling effects and the robustness of quantitative and qualitative food-web descriptors. *Journal of Theoretical Biology*, 226(1) :23–32.

[Bartlett and Cussens, 2013] Bartlett, M. and Cussens, J. (2013). Advances in bayesian network learning using integer programming. *arXiv preprint arXiv :1309.6825*.

[Bartlett and Cussens, 2017] Bartlett, M. and Cussens, J. (2017). Integer linear programming for the bayesian network structure learning problem. *Artificial Intelligence*, 244 :258–271.

[Bersier et al., 2002] Bersier, L.-F., Banašek-Richter, C., and Cattin, M.-F. (2002). Quantitative descriptors of food-web matrices. *Ecology*, 83(9) :2394–2407.

[Bohan et al., 2005] Bohan, D. A., Boffey, C. W., Brooks, D. R., Clark, S. J., Dewar, A. M., Firbank, L. G., Haughton, A. J., Hawes, C., Heard, M. S., May, M. J., et al. (2005). Effects on weed and invertebrate abundance and diversity of herbicide management in genetically modified herbicide-tolerant winter-sown oilseed rape. *Proceedings of the Royal Society of London B : Biological Sciences*, 272(1562) :463–474.

[Bohan et al., 2011] Bohan, D. A., Caron-Lormier, G., Muggleton, S., Raybould, A., and Tamaddoni-Nezhad, A. (2011). Automated discovery of food webs from ecological data using logic-based machine learning. *PLoS One*, 6(12) :e29028.

[Brault, 2014] Brault, V. (2014). *Estimation et sélection de modèle pour le modèle des blocs latents.* PhD thesis, Université Paris Sud-Paris XI.

[Byrd et al., 1999] Byrd, R. H., Hribar, M. E., and Nocedal, J. (1999). An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4) :877–900.

[Campos and Ji, 2011] Campos, C. P. d. and Ji, Q. (2011). Efficient structure learning of bayesian networks using constraints. *Journal of Machine Learning Research*, 12(Mar) :663–689.

[Cano et al., 2011] Cano, A., Masegosa, A. R., and Moral, S. (2011). A method for integrating expert knowledge when learning bayesian networks from data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(5) :1382–1394.

[Caselle and Lowe, 2010] Caselle, J. E. and Lowe, C. (2010). Assessing changes in life history traits and reproductive function of ca sheephead across its range : historical comparisons and the effects of fishing.

[Cattin et al., 2004] Cattin, M.-F., Bersier, L.-F., Banašek-Richter, C., Baltensperger, R., and Gabriel, J.-P. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature*, 427(6977) :835.

[Chickering, 1996] Chickering, D. M. (1996). Learning Bayesian networks is NP-complete. In *Learning from data*, pages 121–130.

[Chickering, 2002] Chickering, D. M. (2002). Learning equivalence classes of bayesian-network structures. *Journal of machine learning research*, 2(Feb) :445–498.

[Chickering and Heckerman, 1996] Chickering, D. M. and Heckerman, D. (1996). Efficient approximations for the marginal likelihood of incomplete data given a bayesian network. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 158–168. Morgan Kaufmann Publishers Inc.

[Chickering et al., 2004] Chickering, D. M., Heckerman, D., and Meek, C. (2004). Large-sample learning of bayesian networks is np-hard. *Journal of Machine Learning Research*, 5(Oct) :1287–1330.

[Cooper and Herskovits, 1992] Cooper, G. F. and Herskovits, E. (1992). A bayesian method for the induction of probabilistic networks from data. *Machine learning*, 9(4) :309–347.

[Cowell et al., 2006] Cowell, R. G., Dawid, P., Lauritzen, S. L., and Spiegelhalter, D. J. (2006). *Probabilistic networks and expert systems : Exact computational methods for Bayesian networks.* Springer Science & Business Media.

[Cussens, 2014] Cussens, J. (2014). Integer programming for bayesian network structure learning. *Quality Technology & Quantitative Management*, 11(1) :99–110.

[Dantzig, 1990] Dantzig, G. B. (1990). *Origins of the simplex method.* ACM.

[De Campos et al., 2009] De Campos, C. P., Zeng, Z., and Ji, Q. (2009). Structure learning of bayesian networks using constraints. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 113–120. ACM.

[Dean and Kanazawa, 1989] Dean, T. and Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational intelligence*, 5(2) :142–150.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

[Diez, 1993] Diez, F. J. (1993). Parameter adjustment in bayes networks. the generalized noisy or–gate. In *Uncertainty in Artificial Intelligence, 1993*, pages 99–105. Elsevier.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1) :269–271.

[Dojer, 2006] Dojer, N. (2006). Learning bayesian networks does not have to be np-hard. In *International Symposium on Mathematical Foundations of Computer Science*, pages 305–314. Springer.

[Donnet et al.,] Donnet, S., Bar-Hen, A., and Barbillon, P. Modèles à blocs latents pour graphe multipartite.

[Dunne, 2009] Dunne, J. A. (2009). Food webs. In *Encyclopedia of complexity and systems science*, pages 3661–3682. Springer.

[Estes et al., 2009] Estes, J., Doak, D., Springer, A., and Williams, T. (2009). Causes and consequences of marine mammal population declines in southwest alaska : a food-web perspective. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 364(1524) :1647–1658.

[Estes and Duggins, 1995] Estes, J. A. and Duggins, D. O. (1995). Sea otters and kelp forests in alaska : generality and variation in a community ecological paradigm. *Ecological Monographs*, 65(1) :75–100.

[Faisal et al., 2010] Faisal, A., Dondelinger, F., Husmeier, D., and Beale, C. M. (2010). Inferring species interaction networks from species abundance data : A comparative evaluation of various statistical and machine learning methods. *Ecological Informatics*, 5(6) :451–464.

[Fan et al., 2014] Fan, X., Yuan, C., and Malone, B. M. (2014). Tightening bounds for bayesian network structure learning. In *AAAI*, volume 4, pages 2439–2445.

[Feelders and van der Gaag, 2006] Feelders, A. and van der Gaag, L. (2006). Learning Bayesian network parameters under order constraints. *International Journal of Approximate Reasoning*, 42 :37–53.

[Flesch and Lucas, 2007] Flesch, I. and Lucas, P. J. (2007). Markov equivalence in bayesian networks. In *Advances in probabilistic graphical models*, pages 3–38. Springer.

[Floyd, 1962] Floyd, R. W. (1962). Algorithm 97 : shortest path. *Communications of the ACM*, 5(6) :345.

[Franc, 2004] Franc, A. (2004). Metapopulation dynamics as a contact process on a graph. *Ecological Complexity*, 1(1) :49–63.

[François, 2006] François, O. (2006). *De l'identification de structure de réseaux bayésiens à la reconnaissance de formes à partir d'informations complètes ou incomplètes.* PhD thesis, INSA de Rouen.

[Friedman et al., 2008] Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3) :432–441.

[Fry, 2006] Fry, B. (2006). *Stable isotope ecology*, volume 521. Springer.

[Gannes et al., 1997] Gannes, L. Z., O'Brien, D. M., and Del Rio, C. M. (1997). Stable isotopes in animal ecology : assumptions, caveats, and a call for more laboratory experiments. *Ecology*, 78(4) :1271–1276.

[Garvey and Whiles, 2016] Garvey, J. E. and Whiles, M. (2016). *Trophic ecology*. CRC Press.

[Gomez Rodriguez et al., 2010] Gomez Rodriguez, M., Leskovec, J., and Krause, A. (2010). Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1019–1028. ACM.

[Gonzales et al., 2015] Gonzales, C., Dubuisson, S., and Manfredotti, C. E. (2015). A new algorithm for learning non-stationary dynamic bayesian networks with application to event detection. In *FLAIRS Conference*, pages 564–569.

[Graham, 2004] Graham, M. H. (2004). Effects of local deforestation on the diversity and structure of southern california giant kelp forest food webs. *Ecosystems*, 7(4) :341–357.

[Hamilton et al., 2010] Hamilton, S. L., Caselle, J. E., Malone, D. P., and Carr, M. H. (2010). Incorporating biogeography into evaluations of the channel islands marine reserve network. *Proceedings of the National Academy of Sciences*, 107(43) :18272–18277.

[Harris, 1974] Harris, T. E. (1974). Contact interactions on a lattice. *The Annals of Probability*, pages 969–988.

[Hebert et al., 2009] Hebert, C. E., Weseloh, D. C., Gauthier, L. T., Arts, M. T., and Letcher, R. J. (2009). Biochemical tracers reveal intra-specific differences in the food webs utilized by individual seabirds. *Oecologia*, 160(1) :15–23.

[Heckerman et al., 1995] Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning bayesian networks : The combination of knowledge and statistical data. *Machine learning*, 20(3) :197–243.

[Hoerl and Kennard, 1970] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression : Biased estimation for nonorthogonal problems. *Technometrics*, 12(1) :55–67.

[Holland et al., 1983] Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels : First steps. *Social networks*, 5(2) :109–137.

[Hosmer Jr et al., 2013] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

[Hyslop, 1980] Hyslop, E. (1980). Stomach contents analysis—a review of methods and their application. *Journal of fish biology*, 17(4) :411–429.

[Ings et al., 2009] Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., et al. (2009). Ecological networks–beyond food webs. *Journal of Animal Ecology*, 78(1) :253–269.

[Jaakkola et al., 2010] Jaakkola, T., Sontag, D., Globerson, A., and Meila, M. (2010). Learning bayesian network structure using lp relaxations. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 358–365.

[Keeling and Eames, 2005] Keeling, M. J. and Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4) :295–307.

[Kéfi et al., 2012] Kéfi, S., Berlow, E. L., Wieters, E. A., Navarrete, S. A., Petchey, O. L., Wood, S. A., Boit, A., Joppa, L. N., Lafferty, K. D., Williams, R. J., et al. (2012). More than a meal... integrating non-feeding interactions into food webs. *Ecology letters*, 15(4) :291–300.

[King et al., 2008] King, R., Read, D., Traugott, M., and Symondson, W. (2008). Invited review : Molecular analysis of predation : a review of best practice for dna-based approaches. *Molecular ecology*, 17(4) :947–963.

[Kitchell, 2012] Kitchell, J. F. (2012). *Food web management : a case study of Lake Mendota*. Springer Science & Business Media.

[Knapp et al., 2001] Knapp, R. A., Corn, P. S., and Schindler, D. E. (2001). The introduction of nonnative fish into wilderness lakes : good intentions, conflicting mandates, and unintended consequences. *Ecosystems*, 4(4) :275–278.

[Koller and Sahami, 1996] Koller, D. and Sahami, M. (1996). Toward optimal feature selection. Technical report, Stanford InfoLab.

[Kurtz et al., 2015] Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS computational biology*, 11(5) :e1004226.

[Larrañaga et al., 1996] Larrañaga, P., Poza, M., Yurramendi, Y., Murga, R. H., and Kuijpers, C. M. H. (1996). Structure learning of bayesian networks by genetic algorithms : A performance analysis of control parameters. *IEEE transactions on pattern analysis and machine intelligence*, 18(9) :912–926.

[Lauritzen, 1996] Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

[Liben-Nowell and Kleinberg, 2007] Liben-Nowell, D. and Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7) :1019–1031.

[Linchant et al., 2015] Linchant, J., Lisein, J., Semeki, J., Lejeune, P., and Vermeulen, C. (2015). Are unmanned aircraft systems (uass) the future of wildlife monitoring ? a review of accomplishments and challenges. *Mammal Review*, 45(4) :239–252.

[Liu, 2015] Liu, L. (2015). Survey of contemporary bayesian network structure learning methods.

[Mallak, 1996] Mallak, S. (1996). *Non-stationary Markov chains*. PhD thesis, Bilkent University.

[Marcogliese, 2004] Marcogliese, D. (2004). Parasites : small players with crucial roles in the ecological theater. *EcoHealth*, 1(2) :151–164.

[Margaritis, 2003] Margaritis, D. (2003). Learning bayesian network model structure from data. Technical report, Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

[Masegosa and Moral, 2013] Masegosa, A. R. and Moral, S. (2013). An interactive approach for bayesian network learning using domain/expert knowledge. *International Journal of Approximate Reasoning*, 54(8) :1168–1181.

[McClelland and Montoya, 2002] McClelland, J. W. and Montoya, J. P. (2002). Trophic relationships and the nitrogen isotopic composition of amino acids in plankton. *Ecology*, 83(8) :2173–2180.

[McDonald et al., 2018] McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., Aksenov, A. A., Behsaz, B., Brennan, C., Chen, Y., et al. (2018). American gut : an open platform for citizen science microbiome research. *mSystems*, 3(3) :e00031–18.

[McDonald-Madden et al., 2016] McDonald-Madden, E., Sabbadin, R., Game, E., Baxter, P., Chadès, I., and Possingham, H. (2016). Using food-web theory to conserve ecosystems. *Nature communications*, 7 :10245.

[Milns et al., 2010] Milns, I., Beale, C. M., and Smith, V. A. (2010). Revealing ecological networks using bayesian network inference algorithms. *Ecology*, 91(7) :1892–1899.

[Muggleton, 1999] Muggleton, S. (1999). Inductive logic programming : Issues, results and the challenge of learning language in logic. *Artificial Intelligence*, 114(1) :283 – 296.

[Murphy, 2001] Murphy, K. (2001). An introduction to graphical models.

[Naïm et al., 2011] Naïm, P., Wuillemin, P.-H., Leray, P., Pourret, O., and Becker, A. (2011). *Réseaux bayésiens*. Editions Eyrolles.

[Nekovee et al., 2007] Nekovee, M., Moreno, Y., Bianconi, G., and Marsili, M. (2007). Theory of rumour spreading in complex social networks. *Physica A : Statistical Mechanics and its Applications*, 374(1) :457–470.

[Nicol et al., 2017] Nicol, S., Sabbadin, R., Peyrard, N., and Chadès, I. (2017). Finding the best management policy to eradicate invasive species from spatial ecological networks with simultaneous actions. *Journal of Applied Ecology*.

[Niculescu et al., 2006] Niculescu, R. S., Mitchell, T. M., and Rao, R. B. (2006). Bayesian network learning with parameter constraints. *Journal of Machine Learning Research*, 7(Jul) :1357–1383.

[O'Gorman et al., 2015] O'Gorman, B., Babbush, R., Perdomo-Ortiz, A., Aspuru-Guzik, A., and Smelyanskiy, V. (2015). Bayesian network structure learning using quantum annealing. *The European Physical Journal Special Topics*, 224(1) :163–188.

[Oniśko et al., 2001] Oniśko, A., Druzdzel, M. J., and Wasyluk, H. (2001). Learning bayesian network parameters from small data sets : Application of noisy-or gates. *International Journal of Approximate Reasoning*, 27(2) :165–182.

[Pace et al., 1999] Pace, M. L., Cole, J. J., Carpenter, S. R., and Kitchell, J. F. (1999). Trophic cascades revealed in diverse ecosystems. *Trends in ecology & evolution*, 14(12) :483–488.

[Palmer et al., 2016] Palmer, M. A., Zedler, J. B., and Falk, D. A. (2016). Ecological theory and restoration ecology. In *Foundations of restoration ecology*, pages 3–26. Springer.

[Pearl, 1985] Pearl, J. (1985). Bayesian networks : A model of self-activated memory for evidential reasoning.

[Pearl, 1986] Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, 29(3) :241–288.

[Penland and Matrosova, 1998] Penland, C. and Matrosova, L. (1998). Prediction of tropical atlantic sea surface temperatures using linear inverse modeling. *Journal of Climate*, 11(3) :483–496.

[Pimm et al., 1991] Pimm, S. L., Lawton, J. H., and Cohen, J. E. (1991). Food web patterns and their consequences. *Nature*, 350(6320) :669.

[Polis, 1991] Polis, G. A. (1991). Complex trophic interactions in deserts : an empirical critique of food-web theory. *The American Naturalist*, 138(1) :123–155.

[Polis et al., 2000] Polis, G. A., Sears, A. L., Huxel, G. R., Strong, D. R., and Maron, J. (2000). When is a trophic cascade a trophic cascade ? *Trends in Ecology & Evolution*, 15(11) :473–475.

[Puterman, 2014] Puterman, M. L. (2014). *Markov decision processes : discrete stochastic dynamic programming*. John Wiley & Sons.

[Ramos and González-Solís, 2012] Ramos, R. and González-Solís, J. (2012). Trace me if you can : the use of intrinsic biogeochemical markers in marine top predators. *Frontiers in Ecology and the Environment*, 10(5) :258–266.

[Robinson, 1977] Robinson, R. W. (1977). Counting unlabeled acyclic digraphs. In *Combinatorial mathematics V*, pages 28–43. Springer.

[Rue and Held, 2005] Rue, H. and Held, L. (2005). *Gaussian Markov random fields : theory and applications*. CRC press.

[Salathé and Jones, 2010] Salathé, M. and Jones, J. H. (2010). Dynamics and control of diseases in networks with community structure. *PLoS Comput Biol*, 6(4) :e1000736.

[Schäfer and Strimmer, 2005] Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).

[Schmidt et al., 2009] Schmidt, S. N., Vander Zanden, M. J., and Kitchell, J. F. (2009). Long-term food web change in lake superior. *Canadian Journal of Fisheries and Aquatic Sciences*, 66(12) :2118–2129.

[Schneider, 2001] Schneider, D. C. (2001). The rise of the concept of scale in ecology : The concept of scale is evolving from verbal expression to quantitative expression. *AIBS Bulletin*, 51(7) :545–553.

[Schrijver, 1998] Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.

[Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2) :461–464.

[Silander and Myllymaki, 2012] Silander, T. and Myllymaki, P. (2012). A simple approach for finding the globally optimal bayesian network structure. *arXiv preprint arXiv :1206.6875*.

[Silliman and Bertness, 2002] Silliman, B. R. and Bertness, M. D. (2002). A trophic cascade regulates salt marsh primary production. *Proceedings of the National Academy of Sciences*, 99(16) :10500–10505.

[Soulé and Terborgh, 1999] Soulé, M. E. and Terborgh, J. (1999). Conserving nature at regional and continental scales—a scientific program for north america. *BioScience*, 49(10) :809–817.

[Spirtes et al., 2000] Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. MIT press.

[Srinivas, 1993] Srinivas, S. (1993). A generalization of the noisy-or model. In *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, pages 208–215. Morgan Kaufmann Publishers Inc.

[Su et al., 2013] Su, C., Andrew, A., Karagas, M. R., and Borsuk, M. E. (2013). Using bayesian networks to discover relations between genes, environment, and disease. *BioData mining*, 6(1) :6.

[Sutton et al., 1998] Sutton, R. S., Barto, A. G., et al. (1998). *Reinforcement learning : An introduction*. MIT press.

[Suzuki, 1996] Suzuki, J. (1996). Learning bayesian belief networks based on the mdl principle : an efficient algorithm using the branch and bound technique.

[Suzuki, 1999] Suzuki, J. (1999). Learning bayesian belief networks based on the mdl principle : An efficient algorithm using the branch and bound technique. *IEICE TRANSACTIONS on Information and Systems*, 82(2) :356–367.

[Syms and Caselle, 2003] Syms, C. and Caselle, J. (2003). Pisco annual subtidal survey training plan. *University of California Santa Barbara and Santa Cruz*.

[Tamaddoni-Nezhad et al., 2012] Tamaddoni-Nezhad, A., Bohan, D., Raybould, A., and Muggleton, S. H. (2012). Machine learning a probabilistic network of ecological interactions. In Muggleton, S. H., Tamaddoni-Nezhad, A., and Lisi, F. A., editors, *Inductive Logic Programming*, pages 332–346, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Tamaddoni-Nezhad et al., 2013] Tamaddoni-Nezhad, A., Milani, G. A., Raybould, A., Muggleton, S., and Bohan, D. A. (2013). Construction and validation of food webs using logic-based machine learning and text mining. *Advances in Ecological Research*, 49 :225–289.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

[Tsamardinos et al., 2003] Tsamardinos, I., Aliferis, C. F., Statnikov, A. R., and Statnikov, E. (2003). Algorithms for large scale markov blanket discovery. In *FLAIRS conference*, volume 2, pages 376–380.

[Tsamardinos et al., 2006] Tsamardinos, I., Brown, L. E., and Aliferis, C. F. (2006). The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1) :31–78.

[van Oevelen et al., 2010] van Oevelen, D., Van den Meersche, K., Meysman, F. J. R., Soetaert, K., Middelburg, J. J., and Vézina, A. F. (2010). Quantifying food web flows using linear inverse models. *Ecosystems*, 13(1) :32–45.

[van Veen et al., 2009] van Veen, F. F., Brandon, C. E., and Godfray, H. C. J. (2009). A positive trait-mediated indirect effect involving the natural enemies of competing herbivores. *Oecologia*, 160(1) :195–205.

[Verzelen, 2008] Verzelen, N. (2008). *Gaussian Graphical Models and Model Selection*. Theses, Université Paris Sud - Paris XI.

[Vinh et al., 2011] Vinh, N. X., Chetty, M., Coppel, R., and Wangikar, P. P. (2011). Polynomial time algorithm for learning globally optimal dynamic bayesian network. In *International Conference on Neural Information Processing*, pages 719–729. Springer.

[Wellman, 1990] Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44 :257–303.

[Williams and Brown, 2016] Williams, B. K. and Brown, E. D. (2016). Technical challenges in the application of adaptive management. *Biological Conservation*, 195 :255–263.

[Williams et al., 2009] Williams, H. P. et al. (2009). *Logic and integer programming*, volume 130. Springer.

[Williams and Martinez, 2000] Williams, R. J. and Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, 404(6774) :180.

[Wolsey, 2000] Wolsey, L. A. (2000). Integer programming. *IIE Transactions*, 32(273-285) :2–58.

[Yaramakala and Margaritis, 2005] Yaramakala, S. and Margaritis, D. (2005). Speculative markov blanket discovery for optimal feature selection. In *Data mining, fifth IEEE international conference on*, pages 4–pp. IEEE.

[Yu et al., 2004] Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J., and Jarvis, E. D. (2004). Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, 20(18) :3594–3603.

[Yu et al., 2013] Yu, K., Wu, X., Zhang, Z., Mu, Y., Wang, H., and Ding, W. (2013). Markov blanket feature selection with non-faithful data distributions. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 857–866. IEEE.

[Yuan and Malone, 2012] Yuan, C. and Malone, B. (2012). An improved admissible heuristic for learning optimal bayesian networks. *arXiv preprint arXiv :1210.4913*.

[Yuan et al., 2011] Yuan, C., Malone, B., and Wu, X. (2011). Learning optimal bayesian networks using a* search. In *IJCAI proceedings-international joint conference on artificial intelligence*, volume 22, page 2186. Citeseer.