



HAL
open science

Réseau bayésien dynamique étiqueté pour l'apprentissage de structure d'un réseau écologique

Etienne Auclair, David Bohan, Nathalie Dubois Peyrard, Régis Sabbadin

► To cite this version:

Etienne Auclair, David Bohan, Nathalie Dubois Peyrard, Régis Sabbadin. Réseau bayésien dynamique étiqueté pour l'apprentissage de structure d'un réseau écologique. *Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle*, 2018, 32 (1/2018), pp.1-9. 10.3166/RIA.32.1-9 . hal-02618147

HAL Id: hal-02618147

<https://hal.inrae.fr/hal-02618147>

Submitted on 25 May 2020

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

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

Réseau Bayésien Dynamique étiqueté pour l'apprentissage de réseau écologique d'arthropodes dans les cultures

Etienne Auclair¹, David Bohan², Nathalie Peyrard¹, Régis Sabbadin¹

¹MIAT, UR875, INRA. F-31320 Castanet-Tolosan, France

²UMR 1210 Biologie et Gestion des Adventices, INRA. BP 86510, Dijon, France

RÉSUMÉ. Un réseau écologique représente l'ensemble des interactions entre les différentes espèces d'un écosystème. Parmi les réseaux d'intérêt, on peut citer les réseaux écologiques renseignant sur les interactions entre les espèces vivantes de parcelles cultivées. La connaissance de tels réseaux peut être très utile dans la connaissance de la biodiversité et la gestion des parcelles cultivées. Cependant, la connaissance de la structure de ces réseaux est rarement complète. Dans cet article, nous décrivons l'apprentissage de réseau écologique d'arthropodes dans des parcelles expérimentales au Royaume-Uni à partir de l'information de la présence/absence des espèces au cours du temps. La dynamique de ces espèces est modélisée par un réseau bayésien dynamique dit "étiqueté", qui intègre des connaissances à priori sur le comportement écologique des espèces pour réduire le nombre de paramètres du modèle. L'apprentissage de la structure de ce réseau se fait par un algorithme à base de score alternant une phase d'estimation des paramètres et d'apprentissage de la structure. Un modèle à blocs stochastiques permet également d'intégrer la connaissance à priori du réseau afin de guider la recherche d'un réseau écologique.

MOTS-CLÉS : Réseau bayésien dynamique, réseau écologique d'arthropodes, modèle à blocs stochastiques

DOI:10.3166/RIA.32.1-9 © 2018 Lavoisier

1. Introduction

Connaître les interactions entre les différentes espèces vivantes d'un écosystème en permet une meilleure compréhension et une gestion plus efficace (McDonald-Madden *et al.*, 2016). L'ensemble de ces interactions est appelé un réseau écologique. Un réseau écologique bien connu est le réseau trophique, regroupant les relations proies/prédateurs entre les espèces d'un écosystème. La connaissance de la structure d'un réseau écologique permet de mieux comprendre la biodiversité d'un milieu naturel, et son impact sur l'écosystème (Pocock *et al.*, 2012). Cela permet aussi la prise en compte des interactions entre espèces pour la gestion de ressources naturelles vivantes dans des environnements comme des zones de pêche ou des parcelles cultivées (Sainsbury *et al.*, 2000). Cependant, connaître la structure exacte d'un réseau écologique est une tâche complexe. Cette tâche peut se résoudre par l'utilisation d'outils issus de la statistique et de l'intelligence artificielle (Faisal *et al.*, 2010). Il s'agit souvent d'utiliser des données génétiques ou d'abondance d'espèces et d'en étudier les corrélations et les dépendances. Ces données peuvent parfois être dynamiques, donnant une informations supplémentaire. Dans des parcelles expérimentales au Royaume-Uni, la présence de différentes espèces d'arthropodes a été relevée à deux dates différentes, avant et après récolte. Ces données permettent de mieux connaître la biodiversité à l'intérieur de ces champs et d'apprendre les relations trophiques entre ces espèces (Bell *et al.*, 2010; Bohan *et al.*, 2005). Cependant, les réseaux appris à l'aide de ces données ne comportent que des relations proie/prédateur et ne prennent pas en compte la dynamique temporelle des espèces observées. Nous souhaitons apprendre les diverses interactions entre ces espèces d'arthropodes en utilisant un modèle dynamique combiné à des connaissances expertes sur les espèces observées. L'apprentissage de ces interactions permet de construire plusieurs réseaux écologiques selon les cultures des parcelles afin d'en comparer les structures. Nous comparons également les différences entre les interactions apprises par notre modèle dynamique avec celles obtenues par d'autres méthodes sur ces données (Bohan *et al.*, 2011).

2. Données écologiques et modélisation

Les données utilisées recensent des arthropodes piégés dans différentes parcelles expérimentales au Royaume-Uni. Cette étude a été effectuée sur 66 parcelles de betterave, 59 parcelles de maïs, 67 parcelles de colza d'été et 65 parcelles de colza d'hiver. Une partie de ces parcelles est composée de cultures OGM résistantes aux herbicides. Des pièges sont installés dans chaque parcelle afin de prélever les arthropodes piégés. Deux prélèvements ont été effectués : le premier avant récolte, le second après récolte. Nous avons donc, pour chaque parcelle, des données d'abondance d'arthropodes sur deux périodes temporelles successives. Nous avons également accès à quelques informations sur les espèces d'arthropodes susceptibles de se trouver dans ces parcelles. Ces informations peuvent concerner leur taille ou leur comportement trophique. Notre objectif est d'utiliser ces données, ainsi que des informations génériques sur les arthropodes afin d'apprendre la structure du réseau d'interactions écologiques entre ces espèces.

Nous souhaitons utiliser une méthode prenant en compte la dimension temporelle des données et ne nécessitant pas forcément des données d'abondance, mais des simples informations sur la présence ou l'absence d'une espèce dans une parcelle. Nous établissons alors un modèle de dynamique des espèces basé sur le réseau écologique de ces espèces. Un réseau écologique est défini ici comme un graphe orienté dont les nœuds représentent les différentes espèces et une arête entre deux nœuds représente une influence d'une espèce vers une autre. Afin d'avoir un modèle simple, nous définissons uniquement deux types d'influences : les influences positives et les influences négatives. Une influence positive d'une espèce i vers une espèce j signifie que la présence de i augmente les chances de présence de l'espèce j au pas de temps suivant, ce qui peut s'interpréter comme le fait que i est une proie ou un facilitateur de j . Une influence négative d'une espèce i vers une espèce j signifie que la présence de i diminue les chances de présence de l'espèce j au pas de temps suivant, ce qui peut s'interpréter comme le fait que i est un prédateur, un parasite ou un compétiteur de j . Nous supposons qu'aucune autre distinction que le caractère négatif ou positif n'est faite entre les différentes influences.

Un modèle très courant de la littérature pour représenter des variables binaires en interaction au cours du temps est le modèle de réseau bayésien dynamique (Friedman *et al.*, 1998). Afin d'intégrer le fait qu'il n'existe que deux types d'influences différentes, nous modélisons la dynamique des espèces par un réseau bayésien dynamique étiqueté (Auclair *et al.*, 2017). Dans un réseau étiqueté, chaque arête est étiquetée selon la nature de l'influence représentée en utilisant un nombre restreint d'étiquettes. Dans ce modèle, la probabilité de présence d'une espèce i dépend uniquement du nombre d'espèces présentes influençant i pour chaque type d'influence. Une représentation graphique d'un réseau écologique, ainsi que la représentation dynamique de ce même réseau est montré dans la figure 1.

Nous modélisons la dynamique des espèces dans un réseau écologique par un réseau bayésien dynamique à deux étiquettes : une étiquette positive et une étiquette négative. Une espèce est présente lors du second échantillonnage si au moins une de ses influences positives réussit **et** toutes ses influences négatives échouent.

$X_i^t \in \{0, 1\}$ est une variable aléatoire renseignant sur la présence ou l'absence de l'espèce $i \in \{1, \dots, n\}$ lors de échantillonnage $t \in \{0, 1\}$ (si $t = 0$, il s'agit de l'échantillonnage avant récolte, si $t = 1$, il s'agit de l'échantillonnage après récolte). La probabilité de réussite d'une influence positive de i vers j est donnée par un paramètre ρ et la probabilité de réussite d'une influence négative de i vers j est donnée par un paramètre τ . Cette influence se fait de manière dynamique : l'état de l'espèce i à $t = 0$ influe sur l'état de l'espèce j à $t = 1$. Par ailleurs, nous doublons ces paramètres afin de différencier l'impact de l'influence sur l'apparition et la survie. Une apparition désigne le fait qu'une espèce i absente à $t = 0$ est présente à $t = 1$ et la survie désigne le fait qu'une espèce i présente à $t = 0$ est également présente à $t = 1$. Les paramètres ρ^{app} et τ^{app} désignent donc les probabilités de réussite des influences respectivement positives et négatives lorsque $X_i^0 = 0$ pour une espèce i tandis que ρ^{sur} et τ^{sur} désignent les probabilités de réussite de ces influences lorsque $X_i^0 = 1$. Soient respectivement

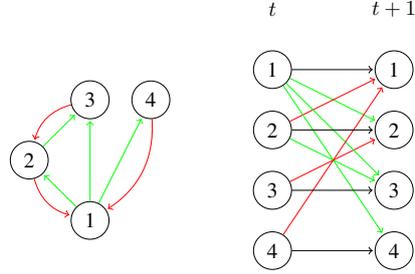


Figure 1. Représentation d'un réseau écologique modélisé par un RBD étiqueté. Gauche : Représentation de la structure d'un réseau écologique où les arêtes colorées en vert correspondent à une influence positive et les arêtes colorées en rouge à une influence négative. Droite : Représentation de la structure de transition du RBD étiqueté associé à ce réseau, dans laquelle les arêtes en noir décrivent les dépendances intrinsèques des espèces (survie, extinction ou apparition).

Pa_i^+ et Pa_i^- l'ensemble des espèces qui exercent une influence positive ou négative vers i ; N_i^+ et N_i^- le nombre d'espèces appartenant respectivement à Pa_i^+ et Pa_i^- présentes à $t = 0$, les probabilités de présence d'une espèce i à $t = 1$ sachant l'état du réseau au temps $t = 0$ sont définies dans les équations (1) et (2).

$$P(X_i^1 = 1 | N_i^+, N_i^-, X_i^0 = 0) = \left(1 - (1 - \rho^{app})^{N_i^+}\right) \cdot (1 - \tau^{app})^{N_i^-} \quad (1)$$

$$P(X_i^1 = 1 | N_i^+, N_i^-, X_i^0 = 1) = \left(1 - (1 - \rho^{sur})^{N_i^+}\right) \cdot (1 - \tau^{sur})^{N_i^-} \quad (2)$$

3. Apprentissage de structure du réseau

3.1. Algorithme d'apprentissage

Le modèle de réseau bayésien dynamique étiqueté a un nombre de paramètres fixé, qui ne dépend pas de la structure du graphe associé. La pénalisation de la vraisemblance par un critère BIC ou BDe n'est donc pas pertinente dans le cadre de ce modèle. L'apprentissage de la structure du réseau se fait donc par la maximisation de la vraisemblance non pénalisée. L'apprentissage de la structure d'un réseau écologique se fait à l'aide d'un algorithme glouton en deux étapes (Auclair *et al.*, 2017) :

- Une étape d'estimation, où l'on détermine, à graphe fixé, la valeur optimale des paramètres (ρ^{app} , τ^{app} , ρ^{sur} , τ^{sur}) au sens de la vraisemblance.
- Une étape de restauration, où l'on apprend, à paramètres fixés, une structure de réseau optimale au sens de la vraisemblance.

Ces deux étapes sont alternées jusqu'à convergence. La deuxième étape se résout à l'aide d'un problème de programmation linéaire en nombre entiers binaires 0/1 où une série de variables intermédiaires et de contraintes permettent de trouver la structure de réseau optimale. Le nombre de ces variables est polynomial en le nombre d'espèces

et de parcelles, et augmente exponentiellement en un nombre k correspondant à un nombre maximal d'influences que peut recevoir une espèce.

3.2. Loi à priori sur les arêtes par modèle à blocs stochastiques

Comme nous disposons d'informations sur les espèces, nous pouvons les intégrer au processus d'apprentissage. Nous pouvons ainsi utiliser la connaissance de la taille des espèces afin d'ajouter une priorité sur certaines arêtes retranscrivant le fait qu'il est rare qu'une espèce soit la proie d'une espèce de taille plus petite qu'elle. Nous partons pour cela du principe que les relations trophiques (proie/prédateur) sont suffisamment courantes dans les réseaux écologiques pour que toute relation positive décrive une relation d'une proie vers un prédateur. Nous avons accès, pour chaque espèce, à une information renseignant sur la taille et le poids réunie en une "classe" ordinale. Plus une espèce appartient à une classe élevée, plus son poids et/ou sa taille est grande. Nous pouvons alors organiser ces classes comme des niveaux trophiques afin de renseigner le fait que des espèces appartenant à une classe élevée ne soient pas des proies d'espèces de classe inférieure. Par ailleurs, nous souhaitons modéliser le fait qu'une espèce i a plus de chances d'être la proie d'une espèce j d'une classe proche. Pour cela, nous modélisons les différentes arêtes du réseau écologique comme un ensemble de variables aléatoires. $G_{i,j}^+$ et $G_{i,j}^-$ constituent un ensemble de variables aléatoires binaires ($G_{i,j}^+$ et $G_{i,j}^- \in \{0, 1\}$) définies comme suit : $G_{i,j}^+ = 1$ si il existe une influence positive de i vers j , $G_{i,j}^+ = 0$ sinon et $G_{i,j}^- = 1$ si il existe une influence négative de i vers j , $G_{i,j}^- = 0$ sinon. Les probabilités associées à ces variables aléatoires sont modélisées par un modèle à blocs stochastiques (Holland *et al.*, 1983). Tout nœud appartient à un (et un seul) bloc et une probabilité $P(G_{i,j}^+ = 1)$ ou $P(G_{i,j}^- = 1)$ est décrite par une fonction qui dépend uniquement du bloc de i et de celui de j . Pour l'apprentissage de réseau écologiques d'arthropodes, nous considérons un modèle à blocs stochastiques où les blocs correspondent aux classes de taille que nous interprétons comme un niveau trophique. Nous définissons le niveau trophique d'une espèce par la fonction notée $TL(i) \in \{1, \dots, maxTL\}$ dans un modèle avec $maxTL$ niveaux trophiques différents. Les probabilités de présence d'une arête positive sont définies par les équations (3) et (4) où une influence positive de i sur j est notée $G_{i,j}^+$ et une influence négative de i sur j est notée $G_{i,j}^-$.

$$P(G_{ij}^+ = 1) = 0 \quad \text{si} \quad TL(i) \geq TL(j) \quad (3)$$

$$P(G_{ij}^+ = 1) = \frac{e^{\alpha \Delta_{ij}}}{1 + e^{\alpha \Delta_{ij}}} \quad \text{si} \quad TL(i) < TL(j). \quad (4)$$

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

Nous considérons ici que les influences positives correspondent à des relations trophiques. Les relations trophiques ont un très fort impact sur la dynamique des espèces et sont les interactions les plus connues de la littérature. C'est pourquoi nous

contraignons davantage ce type d'interaction. Cependant, nous souhaitons également représenter des phénomènes divers (le parasitisme, les relations de compétition...). Nous utilisons pour cela les influences négatives, que nous représentons par un modèle de probabilité plus simple, prenant uniquement en compte les positions relatives des niveaux trophiques, comme décrit dans dans les équation (5) et (6).

$$\begin{aligned} \text{Si } TL(i) \leq TL(j), \quad & P(G_{ij}^- = 1 \mid G_{ij}^+ = 1) = 0 \\ & P(G_{ij}^- = 1 \mid G_{ij}^+ = 0) = \beta_2, \end{aligned} \quad (5)$$

$$\text{Si } TL(i) > TL(j), \quad P(G_{ij}^- = 1) = \beta_1. \quad (6)$$

avec $\beta_1 > \beta_2$ afin de représenter le fait que les relations proie/prédateurs sont les plus fréquentes.

Le vecteur $\psi = (\alpha, \beta_1, \beta_2)$ définit la loi à priori sur le graphe associé au modèle de réseau bayésien dynamique.

4. Étude des réseaux d'arthropodes appris

Les données concernent plusieurs parcelles réparties en 4 cultures différentes chacune ayant des parcelles OGM et non OGM. Nous nous focalisons dans cette étude sur les parcelles OGM et souhaitons comparer les différences entre les réseaux appris pour chaque culture. Les espèces d'arthropodes observées ne sont pas les mêmes dans toutes les cultures. Aussi, nous traitons les données afin de ne retenir, pour chaque culture, que les espèces présentes au moins une fois. Il y en a 41 dans les parcelles de betterave, 29 dans les parcelles de maïs, 40 dans les parcelles de colza d'été et 29 dans les parcelles de colza d'hiver. Nous ne nous intéressons pas à l'abondance, mais à la présence ou à l'absence de chaque espèce dans chaque parcelle. Nous avons appliqué cet algorithme en maximisant à chaque étape la vraisemblance sur l'ensemble des parcelles d'une culture donnée, afin d'obtenir 4 réseaux distincts : un pour chaque culture. Le même procédé a été appliqué sur l'intégralité des données relatives à toutes les parcelles sans distinction de culture afin d'apprendre un réseau "global".

4.1. Comparaison des réseaux de différentes cultures

Nous cherchons à connaître les différences entre les réseaux des différentes cultures. Les réseaux sont-ils similaires pour chaque culture et si non, dans quelle mesure? Pour répondre à cette question, nous calculons tout d'abord la vraisemblance des données correspondant à des parcelles d'une culture par rapport aux réseaux appris pour les autres types de culture. Cela permet de savoir si le réseau appris sur une culture particulière explique bien les données d'une autre culture. Toutes les espèces ne sont pas présentes dans toutes les cultures, cependant, 20 espèces sont présentes au moins une fois dans au moins une parcelle de chaque culture. Pour comparer les cultures entre elles, nous utilisons pour chaque culture un extrait du réseau appris ne conservant que les nœuds correspondant aux 20 espèces communes à toutes les

cultures. Pour calculer la vraisemblance des séquences de présence/absence par rapport à ces extraits de graphes, nous utilisons un extrait des données ne conservant que les séquences de ces 20 espèces. Nous pouvons alors calculer la vraisemblance des données de présence/absence sur ces 20 espèces par rapport à chaque réseau appris pour les nœuds correspondants aux 20 espèces communes à toutes les cultures. Ces valeurs sont renseignées dans le tableau 1.

Tableau 1. Log-vraisemblance des données de présence/absences dans les différentes cultures selon le graphe utilisé. "Global" désigne les données de présence absence de chaque parcelle sans distinction de culture ainsi que le graphe appris par ces données.

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

Les vraisemblance sont toujours plus élevées pour les réseaux correspondant à la culture considérée. Un réseau écologique appris pour une culture ne peut donc pas modéliser efficacement la dynamique des espèces dans une autre culture. Cela est tout de même à relativiser pour la culture de betterave, dont la vraisemblance est proche si elle est calculée à partir du réseau appris à l'aide des données correspondant aux parcelles de maïs. Pour savoir à quel point ces graphes diffèrent et expliquer ces résultats, nous comparons pour chacun d'entre eux les arêtes qu'ils ont en commun et leurs étiquettes. Le tableau 2 symétrique donne pour le graphe de chaque culture le nombre d'arêtes en commun et de même étiquette avec les graphes des autres cultures. Comme pour les calculs de vraisemblance, la comparaison entre tous ces réseaux se fait sur un extrait de chaque graphe appris dans lesquels on ne retient que les 20 espèces communes à toutes les cultures. La diagonale de ce tableau comparant chaque réseau avec lui même, correspond au nombre d'arêtes total de ce réseau. Les nombres en rouge entre parenthèse indique le nombre d'arêtes existantes dans les deux graphes mais étiquetés différemment.

Tableau 2. Nombre d'arêtes communes entre les graphes appris pour chaque culture. En rouge, le nombre d'arêtes présentes dans les deux réseaux mais étiquetés différemment.

Graphe	B	M	CE	CH	Global
B	79 (0)	10 (14)	12 (8)	7 (13)	52 (0)
M	10 (14)	46 (0)	8 (3)	10 (7)	12 (24)
CE	12 (8)	8 (3)	29 (0)	9 (1)	13 (10)
CH	7 (13)	10 (7)	9 (1)	55 (0)	9 (21)
Global	52 (0)	12 (24)	13 (10)	9 (21)	83 (0)

On constate que le réseau appris correspondant aux cultures de betterave est assez similaire à celui appris sur l'ensemble des cultures car ils ont beaucoup d'arêtes en commun et aucune arête d'étiquette différente. Les autres réseaux n'ont que peu d'arêtes communes avec les autres, voire parfois plus d'arêtes étiquetées différemment que d'arêtes de même étiquettes, comme les réseaux associés aux cultures de maïs et de betterave. Cela laisse à penser que les interactions changent beaucoup d'une culture à l'autre.

4.2. Comparaison avec un réseau connu

Un réseau trophique a déjà été appris par une méthode logique et vérifiée par des méthodes de text-mining sur les publications écologiques et des connaissances expertes (Bohan *et al.*, 2011). Nous avons donc comparé le "graphe logique" obtenu par cette méthode avec le "graphe appris" par notre méthode sur l'ensemble des données sans distinction de culture. Si l'on considère une relation trophique comme une influence positive d'une proie vers son prédateur et une influence négative du prédateur vers sa proie, certaines influences imputables à des relations trophiques sont bien retrouvées chez certaines espèces (*Aphidoidea*, *Entomobryidae*, *Smithuridae*). En revanche, beaucoup d'autres influences ne sont pas communes aux deux réseaux. Dans le réseau appris, on remarque par ailleurs que quelques espèces (*Araneae*, *Bembidion guttula*, *Diptera adults*, *Entomobryidae*, *Isotomidae*) ont énormément d'influence négative (plus de 7) vers d'autres espèces, alors que la plupart des autres n'ont que peu d'influences. Ce phénomène n'apparaît pas dans le réseau appris par méthode logique. Ces résultats doivent encore être analysés avec l'aide d'experts.

5. Conclusion

Nous avons cherché à apprendre la structure du réseau écologique décrivant les interactions entre des espèces d'arthropodes dans des parcelles cultivées. Nous avons appliqué une méthode utilisant des données de présence/absence des espèces au cours du temps basée sur un modèle de réseau bayésien dynamique étiqueté. Cette méthode permet un apprentissage sur très peu de données, et peut prendre en compte plusieurs types de connaissances expertes. Cependant, l'algorithme utilisé demande la résolution de plusieurs problèmes de programmation linéaire en nombre entier sur des matrices de taille importante. Les résultats obtenus sur les données de présence/absence d'arthropodes dans les parcelles cultivées montrent des liens entre les réseaux associés aux différentes cultures. Si cette méthode permet de retrouver des influences déjà connues ou apprises par d'autres méthodes, elle apprend tout de même beaucoup d'influences inédites dans la littérature, dont la pertinence reste à évaluer. La prochaine étape est donc une interprétation plus en profondeur de ces résultats avec l'aide d'experts. La connaissance qu'ils peuvent nous apporter, notamment sur l'existence ou l'inexistence d'arêtes désignées par des experts peut être intégré au processus d'apprentissage de manière itérative.

Bibliographie

- Auclair E., Peyrard N., Sabbadin R. (2017). Labeled DBN learning with community structure knowledge. In *Joint european conference on machine learning and knowledge discovery in databases*, p. 158–174.
- Bell J. R., Andrew King R., Bohan D. A., Symondson W. O. (2010). Spatial co-occurrence networks predict the feeding histories of polyphagous arthropod predators at field scales. *Ecography*, vol. 33, n° 1, p. 64–72.
- Bohan D. A., Boffey C. W., Brooks D. R., Clark S. J., Dewar A. M., Firbank L. G. *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*, vol. 272, n° 1562, p. 463–474.
- Bohan D. A., Caron-Lormier G., Muggleton S., Raybould A., Tamaddoni-Nezhad A. (2011). Automated discovery of food webs from ecological data using logic-based machine learning. *PLoS One*, vol. 6, n° 12, p. e29028.
- Faisal A., Dondelinger F., Husmeier D., Beale C. M. (2010). Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods. *Ecological Informatics*, vol. 5, n° 6, p. 451–464.
- Friedman N., Murphy K., Russell S. (1998). Learning the structure of dynamic probabilistic networks. , p. 139–147.
- Holland P. W., Laskey K. B., Leinhardt S. (1983). Stochastic block models: First steps. *Social networks*, vol. 5, n° 2, p. 109–137.
- McDonald-Madden E., Sabbadin R., Game E., Baxter P., Chadès I., Possingham H. (2016). Using food-web theory to conserve ecosystems. *Nature communications*, vol. 7, p. 10245.
- Pocock M. J. O., Evans D. M., Memmott J. (2012). The robustness and restoration of a network of ecological networks. *Science*, vol. 335, n° 6071, p. 973–977. Consulté sur <http://science.sciencemag.org/content/335/6071/973>
- Sainsbury K. J., Punt A. E., Smith A. D. (2000). Design of operational management strategies for achieving fishery ecosystem objectives. *ICES Journal of Marine Science*, vol. 57, n° 3, p. 731–741.