Labelled dynamic Bayesian network for learning arthropods ecological network

Étienne Auclair

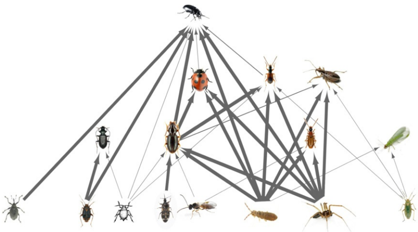
INRA - Unité MIAT

JFRB, 31 Mai 2018

Ecological context

Context

- Management of biodiversity within an ecological network
- Interactions are poorly known
- Few data, but expert knowledge



Objective

Case study and objective

- Arthropods species within experimental fields (Bohan et al., 2005)
- What are the interactions between arthropods species ?
- Are the interactions similar between different cultures ?

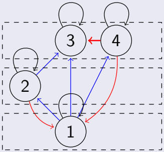
Objective and method

- Labelled DBN model combined to expert knowledge
- Learning interactions between species using temporal presence/absence data

Ecological network modelling

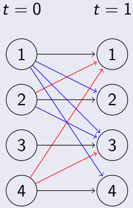
Ecological network

- Directed graph. Nodes represents species.
- Edges labeled according to the type of interaction :
  - + : Positive influence
  - - : Negative influence



Associated labelled Dynamic Bayesian Network model

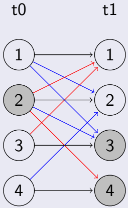
- Binary variables (presence/absence)
- Survival and recolonization depend on previous year



Labeled dynamic Bayesian network model

Definition

- Each edge is labelled
- The transition probability distribution of a variable X_i^t only depends on its number of parents of each state and label
- Two variables with the same numbers of parents of each state and each label have the same transition probability distributions
- Transition probability distribution : function of a small vector of parameters θ. The size of θ independent from the graph structure.



Transition Probabilities of a L-DBN

Notations

- Pa_i^+ and Pa_i^- : set of species with positive or negative influence on i
- N_i^+ and N_i^- : number of species in Pa_i^+ and Pa_i^- present at $t = 0$

Transition probabilities

- Apparition :

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

- Survival :

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

Restoration-Estimation procedure for L-DBN Structure Learning

Score-based method

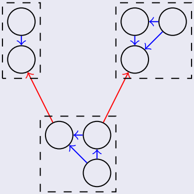
- Number of parameters independent from structure : likelihood as score
- Greedy algorithm[a]
  - Step 1 (Estimation) : Parameters estimation by likelihood maximization, graph structure known
  - Step 2 (Restoration) : Learning network structure maximizing likelihood, parameters known
  - Back to step 1 until convergence

[a] Auclair E., Peyrard N., Sabbadin R. (2017). Labeled DBN learning with community structure knowledge.

SBM model for prior on communities

Definition

- Known blocks (communities) of variables in the network
- The probability of presence of a labelled edge $i \to j$ is a function of its label and the blocks of i and j parameterized by ψ.

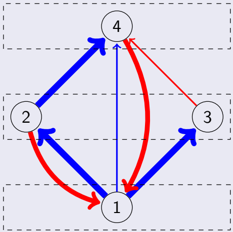


SBM model for prior on trophic levels

Hypothesis knowing trophic levels

- No top down positive edge
- Positive edges more likely on the closest superior trophic level
- Top down negative edges more likely than bottom-down or intra-level negative edges

Prior on edges



Probabilities of edges presence

Positive edges

Trophic levels $TL(i)$ and $TL(j)$ determine the probability of presence of the labelled edge G_{ij}^{l}.

- Top-down and intra-level : $P(G_{ij}^{+} | TL(i) \geq TL(j)) = 0$.
- Bottom-up :$P(G_{ij}^{+} | TL(i) \geq TL(j)) = \frac{e^{\alpha \Delta_{ij}}}{1+e^{\alpha \Delta_{ij}}}$

with $\Delta_{ij} = TL(i) - TL(j)$ and $\alpha > 0$

Negative edges

- $P(G_{ij}^{-} | TL(i) \leq TL(j)) = \beta_2$
- $P(G_{ij}^{-} | TL(i) > TL(j)) = \beta_1$

with $\beta_1 > \beta_2$

$\psi = (\alpha, \beta_1, \beta_2)$

Ecological network learning algorithm

Integer linear programming (ILP) 0-1

- Linearisation of the problem : addition of binary variables defined by linear constraints
- Optimization of the score using ILP
- One independent ILP per variable

Arthropods data

Arthropods dataset (Bohan et al, 2005)

Arthropods trapped in experimental fields

- 66 Beetroot
- 59 Maize
- 67 Summer rape
- 65 Winter rape
  - 2 sample dates : Early and Late.

Arthropods data

Data preprocessing

- 4 different cultures.
- We only use species presents at least once : 41 on beetroot, 29 on maize, 40 on summer rape, 29 on winter rape
- Abundance to presence/absence

Learnt network

- Size as trophic level
- Maximization of the join likelihood on every field
- One network per culture
- One "global" network mixing data from all cultures : 57 species

Network differences

Comparison method

- Subgraph of 20 species common to all cultures
- Number of common edges between each network

Common edges

Same label (Different labels)

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)

- Beetroot network is similar to global network
- Every other network have lots of differences

Network specificity

Likelihood differences

- Subgraph of 20 species common to all cultures
- Log-likelihood of presence/absence data for a culture with the graph learnt for another culture

Data	Graph B	Graph M	Graph SR	Graph WR	Graph 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**

The likelihood is higher for the graph corresponding to the data of its culture : an ecological network is specific to a culture.

Comparison to known network

Logic-based method

- Network learnt by logic-based method on count data [a]
- L-DBN method : loses abundance information (uses only presence/absence) and adds dynamic.
- Does we learn similar influences than logic-based ?

[a] Bohan D. A. et al. (2011).

Logic-based learnt network

- Trophic influence = Negative influence from predator to prey and positive influence from prey to predator
- 7 similar edges
- 2 edges of different label
- A few species have lots of influences on the other species (> 7 edges)

Conclusion

L-DBN for learning ecological network

- Few data available
- DBN with few parameters
- Inclusion of expert knowledge
- SBM prior by size of species

Ecological network for arthropods data

- Different networks for each cultures
- Some known interactions are learnt
- L-DBN graph is a "theoretic" graph of possible interactions
- Can be interpreted a posteriori, sorted by abundance