# Labeled dynamic Bayesian network for learning ecological network

Étienne Auclair, Nathalie Peyrard, Régis Sabbadin

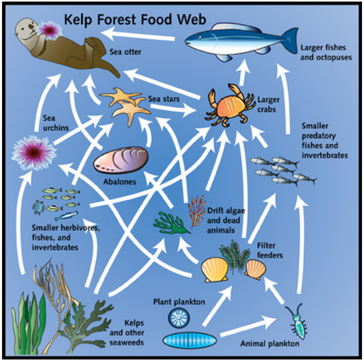
INRA - Unité MIAT

December 14, 2017

# Ecological context and objective

## Context

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



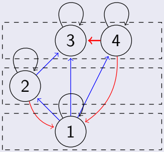

## Objective

Developing a method for learning the structure of an ecological network using presence/absence temporal data

# Ecological network modeling

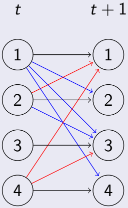
## Ecological network

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




## Associated labelled Dynamic Bayesian Network model

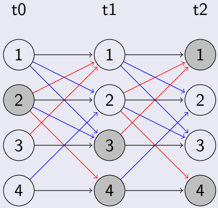
- Binary variables (presence/absence)
- Survival and recolonization depend on previous year



# Labeled dynamic Bayesian network model

## Definition

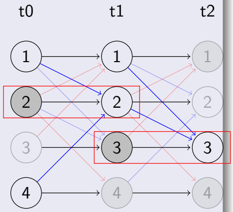
- Each edge is labelled
- The transition probability distribution of a variable $X_i^t$ only depends on its number of parents of each state and label
- Two variables with the same numbers of parents of each state and each label have the same transition probability distributions
- Transition probability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



# 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 $\theta$. The size of $\theta$ independent from the graph structure.



# Model of species dynamics

## Data

- $X_t^i \in \{1, 0\}$ presence or absence of the species $i$ ($i \in \{1, ..., n\}$) at year $t$ ($t \in \{1, ..., T\}$).
- $A^t \in \{1, 0\}$ protection or absence of protection at year $t$.

## Parameters

- Recolonization probability $\varepsilon$.
- Probability of success of each positive influence $\rho$.
- Probability of success of each negative influence $\tau$.

# Transition Probabilities

## Recolonization

Species absent at year $t-1$ : probability of recolonization at year $t$ :
$P(X_i^t = 1 | X_i^{t-1} = 0) = \varepsilon$

## Survival

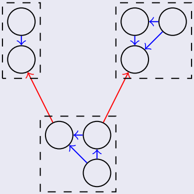
Species present at year $t-1$ : probability of survival at year $t$ :
$P(X_i^t = 1 | X_i^{t-1} = 1) = \left(1 - (1-\rho)^{N_{i,+}^t}\right)(1-\tau)^{N_{i,-}^t}$

$N_{i,l}^t$ : number of "$l$" labeled parents of the species $i$ present at year $t$.

# SBM model for prior on communities

## Definition

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

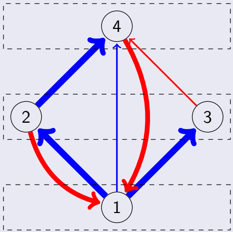


# SBM model for prior on trophic levels

## 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 $G^l_{ij}$.

- 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 $\beta_1 > \beta_2$

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

# Restoration-Estimation procedure for L-DBN Structure Learning

## Score-based method

- Number of parameters independent from structure : likelihood as score
- Greedy algorithm
  - 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

# Ecological network learning algorithm

## Decomposability of the likelihood

$$\log P(x^1, \ldots, x^T \mid x^0, a, \theta, \mathcal{LG}_\rightarrow) = \sum_{i=1}^n score(i)$$

$$\begin{aligned} score(i) &= \sum_{t=0}^{T-1} (1 - x_i^t) \log \varepsilon \\ &+ \sum_{t=0}^{T-1} x_i^t \sum_{0 \le d^+ + d^- \le k} \log \left( \left(1 - (1-\rho)^{d^+}\right) (1-\tau)^{d^-} \right) R_i^{t,d^+,d^-} \end{aligned}$$

with $R_i^{t,d^+,d^-} = 1$ iff the species $i$ has $d^+$ positive labelled parents and $d^-$ negative labelled parents present at year t

## Likelihood term for SBM

SBM term : $logP(\mathcal{LG}_\rightarrow|\psi) = \sum_j score^{SBM}(j)$

$$\begin{aligned} &score^{SBM}(j) = \sum_{i,\Delta_{ij}=0} g_{ij}^- \log \beta_2 + (1 - g_{ij}^-) \log (1 - \beta_2) \\ &+ \sum_{i,\Delta_{ij}<0} \alpha\Delta_{ij} g_{ij}^+ - \log(1 + \exp^{\alpha\Delta_{ij}}) + (1 - g_{ij}^+)(g_{ij}^- \log \beta_2 + (1 - g_{ij}^-) \log (1 - \beta_2)) \\ &+ \sum_{i,\Delta_{ij}>0} g_{ij}^- \log \beta_1 + (1 - g_{ij}^-) \log (1 - \beta_1) \end{aligned}$$

# 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

# Experiments

## Tested methods

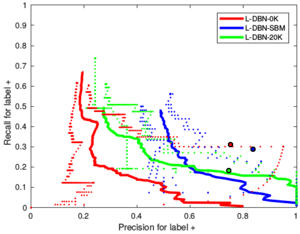
- L-DBN, no additionnal knowledge
- L-DBN, SBM prior
- L-DBN, 20% of known edges, no SBM prior
- MIT[a] score method and qualitative network[b]

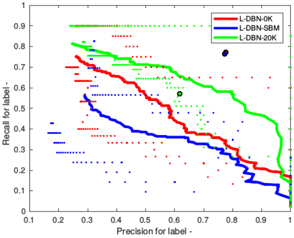
[a] Mutual Information Test. Vinh, 2011
[b] Wellman, 1990

## Simulations

- Synthetic networks built from SBM model ($\alpha = 1/\sqrt{20}, \beta_1 = \alpha/2, \beta_2 = \beta_1/2$)
- $\varepsilon = \rho^+ = \rho^- = \mu = 0.8$
- 10 networks simulated, 10 datasets per network, 20 species, 30 years, last 18 years protected

# Synthetic network results



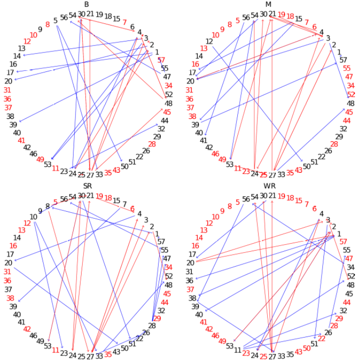



# Real network

## Arthropods dataset (Bohan et al, 2013)

Arthropods trapped in experimental fields

- 66 Beetroot (41 species)
- 59 Maize (29 species)
- 67 Summer rape (40 species)
- 65 Winter rape (29 species)



# Conclusion

## Labeled dynamic Bayesian network

- Few data available
- DBN with few parameters
- Structure learning using ILP
- Inclusion of expert knowledge
- SBM prior improves learning quality

## Application of LDBN and perspectives

- Adaptable for "propagation per contact" models (rumor propagation, network security, disease propagation, fire propagation...)
- Managing while learning : Factored reinforcment learning with MDP and LDBN transition structure