Labeled Dynamic Bayesian Network model for learning network structure

Étienne Auclair

INRA - Unité MIAT

April 28, 2017

Directors : Nathalie Peyrard, Régis Sabbadin

Funding : UPS

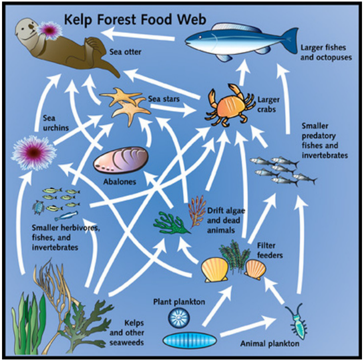
Ecological context and objective

Context and objective

- Biodiversity management using ecological network
- Objective : Definition of a method for learning the structure of an ecological network

Main steps

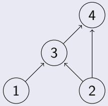
- **Labeled Dynamic Bayesian Network models for contact process**
- Learning the structure of L-DBN
- Application to ecological data



Framework

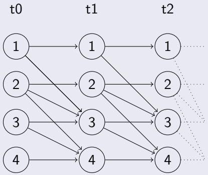
Bayesian network

- Directed acyclic graph
- Conditional probability tables



Dynamic Bayesian network

- Recurrent phenomenon
- Markov process : The state of the network at a moment only depend on the state of the network at the previous moment
- Stationary process : Same transition structure for each time step



Labeled Bayesian Network model

Main characteristics

- Binary process
- Labeled edges describing different interaction types
- Each edge of the same label have the same effect
- A fixed number of parameters for expressing the probabilities

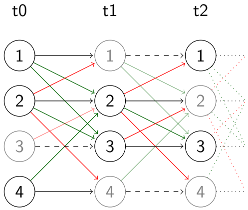
Labelled Dynamic Bayesian Network

- Spread of an infection/an information
- Differentiation between apparition and survival

Labeled DBN

Interaction labels

- Impulsion interactions $(+)$ improve the probabilities of presence.
- Inhibition interactions $(-)$ lowers the probabilities of presence.
- Each interaction is independent
- Number of present labeled neighbors describes probabilities



Parameters

- Probability independent of the interactions ε
- Probability of success of impulsion influence ρ
- Probability of success of inhibition influence τ

Transition probabilities

Notations

- $X_i^t \in \{1, 0\}$: presence or absence of the process $i(i \in \{1, ..., n\})$ at time $t(t \in \{1, ..., T\})$.
- $E(X_i^t)$: neighbors of X_i regardless the interactions labels.
- $E_+(X_i^t)$, $E_-(X_i^t)$: neighbors of X_i^t of label $+/-$.

Apparition probability

$$P(X_i^t = 1 | E(X_i^t), X_i^{t-1} = 0) =$$
$$(1-\tau)^{\sum_{j \in E_-(X_i^t)} X_j^t} \cdot \left(\varepsilon + (1-\varepsilon) \cdot \left(1 - (1-\rho)^{\sum_{j \in E_+(X_i^t)} X_j^t} \right) \right)$$

- No successful inhibition interaction
- Spontaneous apparition
- At least one successful impulsion interaction

L-DBN for additional information

Apparition probability with covariate

- $A(X_i^t) \in \{1, 0\}$: presence or absence of the covariate on variable X_i^t. If present, the probabilities are lowered.
- μ : probability of success of the weakening of the covariate
- $\mu^{A(X_i^t)} = \mu$ if $A(X_i^t) = 1$; $\mu^{A(X_i^t)} = 1$ otherwise
- $A(X_i^t)$ is known or deterministic

$$P(X_i^t = 1 | E(X_i^t), A(X_i^t), X_i^{t-1})0) =$$
$$\mu^{A(X_i^t)} \cdot (1-\tau)^{\sum_{j \in E_-(X_i^t)} X_j^t} \cdot \left(\varepsilon + (1-\varepsilon) \cdot \left(1 - (1-\rho)^{\sum_{j \in E_+(X_i^t)} X_j^t}\right)\right)$$

Multiple L-DBN

Characteristics

- Different forces of Impulsion and Inhibition interactions
- Several covariates and several fix probabilities ε

Generic apparition probability

$$P(X_i^{t+1} = 1 | E(X_i^{t+1}), A_c(X_i^t), X_i^t = 0) =$$
$$\prod_{c=c_1}^{c_{max}} \mu^{A_c}(X_i^t)$$
$$\cdot \left(\sum_{u=u_1}^{u_{max}} \varepsilon_u + (1 - \sum_{u=u_1}^{u_{max}} \varepsilon_u \cdot \left(1 - \prod_{r=r_1}^{r_{max}} (1 - \rho_r)^{\sum_{j \in E_r^{app}(X_i^{t+1})} X_j} \right) \right)$$
$$\cdot \prod_{s=s_1}^{s_{max}} (1 - \tau_s)^{\sum_{j \in E_s^{app}(X_i^{t+1})} X_j}$$

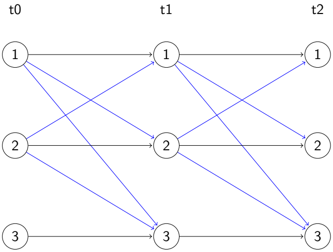
In practice, only a few parameters are used.

Per contact propagation process by L-DBN model

SIS model

- An individual i at a moment t is either not infected **S** ($X_i^t = 0$) or infected **I** ($X_i^t = 1$)
- Only one label : apparition impulsion interactions
- Only one ε : spontaneous disparition

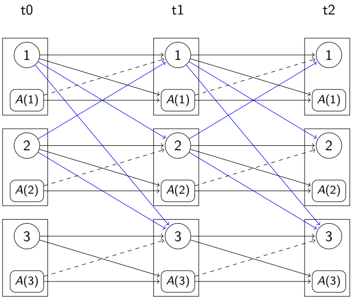
$$
\begin{aligned}
P\left(X_i^{t+1} = 1 | E_+^t(i), X_i^t = 0\right) &= 1 - (1-\rho)^{E_+^t(i)} \\
P\left(X_i^{t+1} = 1 | E_+^t(i), X_i^t = 1\right) &= \varepsilon
\end{aligned}
$$



Per contact propagation process by L-DBN model

SIR model

- SIS model, but a recovered infected individual cannot be infected anymore (immunization or death, **R**)
- Resistance is modeled by a covariate A
- $A(X_i^0) = 0$; $A(X_i^{t-1}) = 1 \rightarrow A(X_i^t) = 1$; $X_i^{t-1} = 1 \rightarrow A(X_i^t) = 1$
- $\mathbf{S} : \{X_i^t = 0, A(X_i^t) = 0\}$; $\mathbf{I} : X_i^t = 1$; $\mathbf{R} : \{X_i^t = 0, A(X_i^t) = 1\}$



Ecological network modelized as a L-DBN

Description

- A species i at a moment t can be absent ($X_i^t = 0$) or present ($X_i^t = 1$)
- Survival depend on positive (prey, facilitators...) or negative (predators, parasitics...) relations
- A species can spontaneously recolonize the observed area
- The observed area can be protected at some moments.

Ecological network modelized as a L-DBN

Modelization

- Spontaneous recolonization ε
- Survival : 1 impulsion label, 1 inhibition label
- Covariate : $A(X_i^t) = 1$ if the area is not protected
- Covariate only depend on the time step : $A(X_i^t) = A(X_j^t) \forall i, j$.

$$
\begin{aligned}
P\left(X_i^{t+1} = 1 | X_i^t = 0, E(X_i^t)\right) &= \mu^{A(X_i^t)} \cdot \varepsilon \\
P\left(X_i^{t+1} = 1 | X_i^t = 1, E(X_i^t)\right) &= \mu^{A(X_i^t)} \cdot \left(1 - (1-\rho)^{E_+^t(i)}\right) \cdot (1-\tau)^{E_-^t(i)}
\end{aligned}
$$

Other known models as L-DBN

Other examples

- Information spread within social network
- Spatial ecology management

M Gomez-Rodriguez, Inferring Networks of Diffusion and Influence, 2011.

R Durrett, Stochastic spatial model : a user's guide to ecological applications, 1994.

R Salathé, Dynamics and Control of Diseases in Networks with Community Structure, 2010.

S Nicol, Finding the best management policy for spatial ecological networks with simultaneous actions, 2016.

Conclusion

A general framework : L-(D)BN

- Labeled edges on the network
- Reduce set of parameters
- Usable for several known processes
- Static or dynamic version

Application of L-DBN model

Learning the structure of a L-DBN using greedy iterative algorithm

- Maximization of the likelihood
- Alternate parameters estimation and structure learning
- ILP for learning the structure
- Prior on expert knowledge by SBM

Application on ecological data

- Kelp forest species abundance[a]
- Arthropods species within experimental fields[b]

[a] J.Caselle, L.Dee
[b] D.Bohan