Labeled DBN Learning with Community Structure Knowledge

E. Auclair, N. Peyrard[(✉)], and R. Sabbadin

MIAT, UR 875, Université de Toulouse, INRA,
31320 Toulouse, Castanet-Tolosan, France
`nathalie.peyrard@inra.fr`

Abstract. Learning interactions between dynamical processes is a widespread but difficult problem in ecological or human sciences. Unlike in other domains (bioinformatics, for example), data is often scarce, but expert knowledge is available. We consider the case where knowledge is about a limited number of interactions that drive the processes dynamics, and on a community structure in the interaction network. We propose an original framework, based on Dynamic Bayesian Networks with labeled-edge structure and parsimonious parameterization, and a Stochastic Block Model prior, to integrate this knowledge. Then we propose a restoration-estimation algorithm, based on 0-1 Linear Programing, that improves network learning when these two types of expert knowledge are available. The approach is illustrated on a problem of ecological interaction network learning.

Keywords: Labeled edge network learning
Dynamic Bayesian Network · Stochastic Block Model
0-1 linear programing · Trophic network

1 Introduction

Learning an interaction network between entities is a widespread problem in bioinformatics [29], ecology [19] or social sciences [17]. This problem is often formulated in the framework of Bayesian Networks (BN) [13]. When the state of variables changes through time, learning approaches based on Dynamic Bayesian Networks (DBN) have also been proposed [10]. Learning a DBN amounts to learning both its structure (i.e. the conditional independences between the variables) and its Transition Probability Tables (TPT). Several solution approaches to DBN learning exist. They generally extend the methods used for learning static BN (see [20] for a review). They often consist in defining a global score function on networks measuring their "goodness of fit" and in using search methods to find the DBN structure and TPT that jointly optimize the score function. While finding an optimal BN is NP-hard in general [4], it is not the case for DBN without synchronous edges where the global score function is decomposable into independent local scores (one per variable). This is because, without synchronous

arcs, a DBN structure is acyclic, so there is no need to check a global acyclicity constraint on the learned network, as opposed to the BN case. Under this assumption, [6] has provided polynomial time algorithms for learning DBN structure in the case of minimum description length (MDL) and Bayesian Dirichlet equivalence (BDe) scores. [27] have extended these results to the Mutual Information Tests (MIT) score.

Even with the hypothesis of no synchronous edges, learning DBN structure remains difficult since in many problems where interactions are to be learned, observed data are scarce. On the other hand, expert knowledge is often available for such problems, that could be taken into account in the learning process. In this paper, we consider two different types of expert knowledge and show how to use this knowledge to improve DBN structure learning.

First, we consider information about the mechanisms driving the process dynamics (e.g. facilitation, competition, cooperation. . .). This may be useful in order to constrain some elements of the TPT. For instance, equality constraints between some elements of distinct TPTs have been studied by [22]. Here, we derive such equalities in the case of generalized'per contact' processes where the dynamics of a variable is the result of a limited number of interaction types. This enables us to define a parsimonious parameterization of the TPT from a labeled-edge structure of the DBN, using one label and one parameter per interaction type. Then, variables submitted to the same influences share the same TPT. This defines the general framework of DBN with labeled-edge structure and parsimonious parameterization of the TPT (Sect. 2.1). We will refer to them as Labeled DBN (L-DBN). We consider in particular the case of only two types of interactions: impulsion and inhibition.

The idea of labeling the edges of a BN to model the positive or negative influence of a variable on another has already been considered in the framework of qualitative BN [28]. However, such a labeled network is usually given as an input of a BN parameter learning problem, in order to constrain the learned CPT [8]. In this article we tackle the question of learning the labels together with the structure and the parameters.

The second type of information we consider is knowledge about the structure of the interaction network. Structural constraints can be imposed on the network to simplify the learning task by reducing the search space, independently of the physical meaning of the network. They can be local constraints on node degree or forbidden edges [3]. Global features have also been considered. For instance, in [24] an upper bound of the treewidth of the BN graph is introduced in the learning procedure. In [23], the authors introduce a prior on the partial ordering of the nodes and show how to learn a BN in a Bayesian framework. As opposed to these kind of constraints, we consider structural constraints linked to expert knowledge, and we formalize the introduction of knowledge about a community structure of the network, during L-DBN learning. Namely we assume that the nodes of the interaction network are grouped into communities. Social networks, as well as food webs, are naturally structured in communities of individuals defined by jobs, schools, etc., or by trophic levels. . . Knowing them provides

some prior knowledge about the within and between communities interactions that can help the learning. We model this a priori as a Stochastic Block Model (SBM, [15]), in the spirit of [1] for static continuous data for learning Gaussian graphical models. In this paper, we extend SBM to multiple interaction types, in order to deal with the labeled edges of a L-DBN (Sect. 2.2).

In Sect. 3, we propose an iterative Restoration-Estimation (RE) algorithm for learning both the structure (edges and labels) and the parameters of a L-DBN model with SBM prior. In Sects. 4 and 5, we model and solve a problem of ecological interactions network learning by combining a L-DBN model related to causal independence BN models [14], the SBM prior and the RE algorithm. In Sect. 6, we evaluate, on synthetic ecological networks and on a real one, how the successive introductions of knowledge on interaction types and on network structure improve the quality of the restored network, compared to a learning approach based on a non-parameterized DBN.

2   Integrating Labeled Edges and Community Structure Knowledge in DBN

2.1   Labeled Dynamic Bayesian Networks

Let us consider a set $\{(X_1^t)_{t=1,\dots T}, \dots, (X_n^t)_{t=1,\dots T}\}$ of n coupled random processes over horizon T. Then, denoting $X^t = \{X_1^t, \dots, X_n^t\}$, a DBN allows to concisely represent the joint probability distribution $P(X^1, \dots, X^T)$ under Markovian and stationarity assumptions, by exploiting conditional independence between the variables. These independences can be represented by a bipartite graph $\mathcal{G}_\rightarrow = (V, E)$ between two sets of vertices, both indexed by $\{1, \dots, n\}$ and respectively representing the variables $\{X_1^t, \dots, X_n^t\}$ and $\{X_1^{t+1}, \dots, X_n^{t+1}\}$. In $\mathcal{G}_\rightarrow$, edges are directed from vertices at time t, to vertices at time $t+1$. The joint probability distribution writes $P(X^{t+1}|X^t) = \prod_{i=1}^n P(X_i^{t+1}|X_{Par(i,\mathcal{G}_\rightarrow)}^t)$, where $Par(i, \mathcal{G}_\rightarrow) = \{j, (j, i) \in E\}$.

The DBN framework enables a huge gain in space by representing individual tables $P_i(X_i^{t+1}|X_{Par(i,\mathcal{G}_\rightarrow)}^t)$ rather than directly the joint transition probability. However, when some domain-specific knowledge impose that some individual transition probabilities are identical, it is possible to save even more space. This will be the case, for instance, when a limited number L of interaction types between variables exists, and all interactions of same type have the same effect on a variable, regardless which parent variables are concerned. We can use these interaction types to define the TPT by a small number of parameters. This is the case for epidemic contact processes models [9], where there is only one interaction type (contamination) and the state of a variable X_i^t only depends on the number of infected parents (and not on the precise knowledge of which parents are infected). We generalize this idea with the L-DBN framework. To do so, we consider a labeled version of graph $\mathcal{G}_\rightarrow$, namely graph $\mathcal{LG}_\rightarrow = (V, E, \mathcal{L}, \lambda)$, where E is a set of edges, $\mathcal{L} = \{1, \dots, L\}$ a set of edge labels (interaction types) and $\lambda : E \rightarrow \mathcal{L}$ a labeling function.

Definition 1. *A Labeled DBN is a DBN such that:*

– *In the graphical representation of the conditional independences of the global transition probability, each edge is labeled by a label $l \in \mathcal{L}$ (except edges from X_i^t to X_i^{t+1} if present). The set of parents of a vertex i connected through an edge with label l is denoted $Par^l(i, \mathcal{LG}_\rightarrow)$.*
– *Two parents in $Par^l(i, \mathcal{LG}_\rightarrow)$ are assumed indistinguishable in their influence on i, and each labeled influence applies independently. This means that the transition probability distribution of X_i^{t+1} only depends on the number of parents in each possible state for each label (and the state of X_i^t if the edge exists).*
– *Two individuals i and j, such that $card(Par^l(i, \mathcal{LG}_\rightarrow)) = card(Par^l(j, \mathcal{LG}_\rightarrow))$ for all $l \in \mathcal{L}$, have the same TPT.*
– *This transition probability distribution is defined as a function of a vector of parameters θ, of low dimension (one per label and, possibly, a further one to model transitions independent of the $\{Par^l(i, \mathcal{LG}_\rightarrow)\}_l$).*

Once the form of the parameterized transition function is given, the TPTs of a L-DBN can be modeled in a very concise way, by specifying only the labeled graph (sets $Par^l(i, \mathcal{LG}_\rightarrow)$) and the parameters vector θ. One advantage of using a parameterized representation of a L-DBN is that it can be learnt more efficiently from small data sets than a non-parameterized representation.

The L-DBN framework is very general. A family of L-BDN of interest is that of binary *per contact propagation* processes. In this case, X_i^t is a binary random variable: $X_i^t = 1$ is for presence and $X_i^t = 0$ represents absence. Two types of interactions are possible ($\mathcal{L} = \{+, -\}$): an *impulsion interaction* ($+$) from a variable X_j^t to a variable X_i^{t+1}, increases the probability of presence of the process i at $t + 1$; an *inhibition interaction* ($-$) from a variable X_i^t to a variable X_i^{t+1} decreases the probability of presence of the process i at $t + 1$ (as in qualitative BN, [28]). All edges of identical label have the same impact on the transition probabilities of the affected variables. We associate a parameter to each label: ρ^+ is the probability of success of an impulsion; ρ^- is the probability of success of an inhibition. We also assume that the success or failure of the influence of all parents are independent (as in a causal independence BN model, [14]). The L-DBN model when the parents only have an impact on the survival (e.g. for species in interactions) is as follows. First the probability of apparition at vertex i is independent from the state of the other variables. We model this by a parameter ε, interpretable as the probability of spontaneous apparition. Then, the probability of survival of a process i is the probability of success of at least one impulsion interaction and of failure of all inhibition interactions. Therefore, the survival of i is the result of independent coin flips. Let $N_{i,l}^t = |\{j \in Par^l(i, \mathcal{LG}_\rightarrow), X_j^t = 1\}|$, for $l \in \{+, -\}$, then,

$$P(X_i^{t+1} = 1 | X_i^t = 0) = \varepsilon$$
$$P(X_i^{t+1} = 1 | X_i^t = 1, N_{i,+}^t, N_{i,-}^t) = \left(1 - (1 - \rho^+)^{N_{i,+}^t}\right) \cdot (1 - \rho^-)^{N_{i,-}^t}.$$

Similarly, the L-DBN model when interactions have only an impact on apparition (e.g. for disease spread) is defined by

$$P(X_i^{t+1} = 1 | X_i^t = 1) = \varepsilon$$
$$P(X_i^{t+1} = 1 | X_i^t = 0, N_{i,+}^t, N_{i,-}^t) = \left(1 - (1-\rho^+)^{N_{i,+}^t}\right) \cdot (1-\rho^-)^{N_{i,-}^t}.$$

The family of per contact propagation processes also includes processes where survival (or apparition) requires the success of all impulsion interactions and the failure of one inhibition interaction, and processes defined by any other AND/OR combination of independent events of impulsion and inhibition successes. In this family, the TPT are defined by three parameters only: $\theta = \{\rho^+, \rho^-, \varepsilon\}$.

Figure 1 (left) shows the graphical representation of an example L-DBN structure $\mathcal{LG}_\rightarrow$ with $n = 4$. In this example, $Par^+(1, \mathcal{LG}_\rightarrow) = \emptyset$ and $Par^-(1, \mathcal{LG}_\rightarrow) = \{2, 4\}$. Because the state of X_i^t determines whether the transition is a survival or an apparition, we also add an edge from X_i^t to X_i^{t+1} (without label) not associated to any parameter. This edge is known, it does not have to be learnt. Figure 1 (right) shows an equivalent static representation of $\mathcal{LG}_\rightarrow$, where nodes corresponding to variables X_i^t and X_i^{t+1} have been collapsed. This representation may have a natural meaning with respect to the represented process, as will be the case with the ecological network case study we will describe in Sect. 4. The meaning of the dashed boxes is related to this example (see Sect. 4.2).
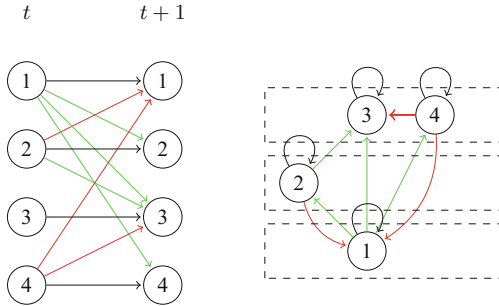

Fig. 1. The two graphical representations of the structure of a L-DBN with 4 variables and 2 labels. Green and red edges represent respectively label '+' and '−' in the case of ecological network. Black edges represent the unlabeled edges accounting for the dependence of X_i^{t+1} on X_i^t. Left: $\mathcal{LG}_\rightarrow$, dynamics representation. Right: equivalent static representation (dashed rectangles represent the blocks of the SBM) (Color figure online).

2.2   Stochastic Block Models for L-DBN

In the above section, the labels and the parameterized TPT enable to encode knowledge about the mechanisms underlying the dynamics of X^t, for a given DBN structure. Now we present how to embed knowledge about the properties of the structure $\mathcal{LG}_\rightarrow$ itself in the L-DBN model.

Let $\left\{G_{ij}^l\right\}_{1 \leq i,j \leq n, 1 \leq l \leq L}$ be a random binary vector representing the presence or absence of each type of edge from i to j: $G_{ij}^l = 1$ if $i \in Par^l(j, \mathcal{LG}_\rightarrow)$ and 0 otherwise. A realization of $\{G_{ij}^l\}_{1 \leq i,j \leq n, 1 \leq l \leq L}$ defines the labeled graph $\mathcal{LG}_\rightarrow$ of a L-DBN. Without prior information, the variables G_{ij}^l could be modeled as independent variables with uniform distribution. Instead, we assume that the vertices of the static representation of $\mathcal{LG}_\rightarrow$ are organized into B disjoints blocks, or communities, and block membership is indicated by a function $f : \{1, \ldots, n\} \rightarrow \{1, \ldots, B\}$. In the example of Fig. 1, there are three such blocks: $\{1\}, \{2\}, \{3, 4\}$. Then, we model the distribution of the $\{G_{ij}^l\}$s in the Stochastic Block Model (SBM) framework [15]. The SBM model makes only two assumptions: (1) the presence of an edge with label l from vertex i towards j (variable $\{G_{ij}^l\}$) is independent of the presence of an edge of the same label from vertex u towards v (variable $\{G_{uv}^l\}$), $\forall (i, j, u, v)$ and (2), the probability distribution of G_{ij}^l only depends on l, $f(i)$ and $f(j)$ (and not on the specific vertices i and j directly). Therefore, in the case of two labels ($L = 2$) the joint distribution of the $\left\{G_{ij}^l\right\}_{1 \leq i,j \leq n, 1 \leq l \leq L}$ is fully determined by the two probabilities $P(G_{ij}^1 = 1 \mid f(i), f(j))$ and $P(G_{ij}^2 = 1 \mid f(i), f(j), G_{ij}^1)$. We will assume that these probabilities are parameterized by parameter ψ. Note that, unlike in most applications of SBM, we assume here that the block memberships are known while the edges are unknown and modeled by random variables G_{ij}^l. This is because our objective is to learn the network from the blocks and observations X_i^t, instead of learning the blocks from the network, as usual.

3   A Restoration-Estimation Procedure for L-DBN Structure Learning

L-DBN parameters and structure learning poses several difficulties. In the non-parameterized DBN learning case, when the structure of the DBN is known, analytic expressions for the estimators of the transition probabilities from counts on data are available [20]. An analytic expression of the solution for likelihood maximization is not available anymore in L-DBN since the tables are no longer independent. So we will have to rely on numerical solvers. Structure learning steps (for given model parameters) must also be handled differently: First, not only edges presence but also their labels must be learned. Then, usual score functions combine a term measuring how a network fits a dataset and a penalty term on the model complexity [26], to avoid over-fitting that occurs when increasing the number of edges in the learnt network. Such penalties are no longer relevant for a L-DBN, where the number of parameters is fixed and does not vary with $\mathcal{LG}_\rightarrow$: the model complexity does not increase with the number of edges. The BDe score [13] is not relevant either, due to the assumption of independence between parameters in the different tables, which does not hold in a L-DBN.

Therefore, we propose to maximize the non-penalized log-likelihood. Our restoration-estimation algorithm is an iterative algorithm which alternatively updates estimates of $\mathcal{LG}_\rightarrow^k$ and (θ^k, ψ^k) until a local maximum of $P(D, \mathcal{LG}_\rightarrow \mid \theta, \psi)$ is found, that is:

$$E \ step : \theta^{k+1}, \psi^{k+1} \leftarrow \underset{\theta, \psi}{\arg\max} \log P(D, \mathcal{LG}_{\rightarrow}^{k} | \theta, \psi),$$

$$R \ step : \mathcal{LG}_{\rightarrow}^{k+1} \leftarrow \underset{\mathcal{LG}_{\rightarrow}}{\arg\max} \log P(D, \mathcal{LG}_{\rightarrow} | \theta^{k+1}, \psi^{k+1}).$$

These two iterative steps can be rewritten as follows:

$$E : \theta^{k+1} \leftarrow \underset{\theta}{\arg\max} \log P(D | \mathcal{LG}_{\rightarrow}^{k}, \theta),$$

$$\psi^{k+1} \leftarrow \underset{\psi}{\arg\max} \log P(\mathcal{LG}_{\rightarrow}^{k} | \psi), \tag{1}$$

$$R : \mathcal{LG}_{\rightarrow}^{k+1} \leftarrow \underset{\mathcal{LG}_{\rightarrow}}{\arg\max} [\log P(D | \mathcal{LG}_{\rightarrow}, \theta^{k+1}) + \log P(\mathcal{LG}_{\rightarrow} | \psi^{k+1})]. \tag{2}$$

In the first step, given a fixed labeled graph $\mathcal{LG}_{\rightarrow}^{k}$, both the parameters θ^{k+1} of the L-DBN and the parameters ψ^{k+1} of the SBM are estimated by continuous optimization. In the second step, given fixed parameters values, the labeled graph $\mathcal{LG}_{\rightarrow}^{k+1}$ is updated by solving a 0-1 Integer Linear Program (ILP) [16]. In practice, since the log-likelihood is a decomposable score, it amounts to solving n 0/1 ILPs by defining for each vertex as many variables (i.e. an exponential number in k) as potential parents sets [5]. However, the structure of real problems often allows to decrease the number of introduced variables in the 0/1 ILP. In the next section, we illustrate this by instanciating the RE algorithm on a problem of ecological interaction network learning, in which the number of variables will only be quadratic in k.

4   Ecological Network Modeling

An ecological network describes interactions between species in a given environment. The learning problem is that of learning this network from time series of observations of the species. Interactions can be trophic (prey/predator), parasitic, competitive, ... They can model positive or negative influence on the species survival. It is therefore possible to label the edges of an ecological network with a '+' or a '−' label (absence of interaction is represented by an absence of edge). In practice, the main interactions between species are trophic interactions. They structure the community into trophic levels that are often known [21]. We now show how to take these labels and trophic levels into account to model species dynamics in the L-DBN framework with SBM prior.

4.1   L-DBN Species Transition Probabilities

We assume that the species dynamics are observed at regular time steps, and that occurrence observations are available: dataset D corresponds to the observation of the absence $(x_i^t = 0)$[1] or presence $(x_i^t = 1)$ of every species at time $t \in \{1, \ldots, T\}$. Information is also available on whether the observed area is

---

[1] Here and in the following, upper case letters are used for random variables, and lower case letters for a realization.

protected ($a^t = 1$) or not ($a^t = 0$). Labels of the edges in the associated graph $\mathcal{LG}_\rightarrow$ can take 2 values: '+' or '−'. An example of labeled ecological network with four species is shown in Fig. 1 (right). Then, the definition of the TPT is based on the following assumptions:

(a) a species survives if at least one positive influence succeeds and all negative fail;
(b) a species with empty $Par^+(i, \mathcal{LG}_\rightarrow)$ (for instance a species at the bottom of the trophic chain) cannot disappear if it is protected and all the species in $Par^-(i, \mathcal{LG}_\rightarrow)$ are absent;
(c) a species with non empty $Par^+(i, \mathcal{LG}_\rightarrow)$ cannot survive if all species are absent in $Par^+(i, \mathcal{LG}_\rightarrow)$;
(d) If $i \in Par^+(j, \mathcal{LG}_\rightarrow)$, then $i \notin Par^-(j, \mathcal{LG}_\rightarrow)$.

These assumptions form "hard" knowledge, which limits the set of possible ecological interaction networks, for a given observed dataset D. Then, the TPT of the L-DBN are defined from the vector of parameters $\theta = (\varepsilon, \rho^+, \rho^-, \mu)$, where ε is a probability of recolonization, ρ^+, ρ^-, are probabilities of success of positive and negative influences. $\mu \in [0, 1]$ is a penalty factor applied to recolonization and survival probabilities of species when the area is unprotected. We describe only the transition probabilities towards presence $P(X_i^{t+1} = 1|X_i^t, a^t)$. All other transition probabilities are derived from these. Two situations are possible depending on whether species i is absent or present at time t:

(i) the probability for a species absent at t to colonize the observed area at $t+1$ is assumed fixed and independent of the presence of other species

$$P(X_i^{t+1} = 1|X_i^t = 0, a^t) = \mu^{(1-a^t)}\varepsilon. \tag{3}$$

(ii) The probability for a species present at t to survive at $t+1$ is the probability of success of at least one positive influence (if needed) and the probability of failure of all negative influences, and these interaction events are independent. It is expressed as follows: if $Par^+(i, \mathcal{LG}_\rightarrow) = \emptyset$

$$P(X_i^{t+1} = 1|X_i^t = 1, x_{Par^-(i,\mathcal{LG}_\rightarrow)\backslash i}^t, a^t) = \mu^{(1-a^t)}\left(1 - \rho^-\right)^{N_{i,-}^t}. \tag{4}$$

Else it is equal to $\mu^{(1-a^t)}\left(1 - \left(1 - \rho^+\right)^{N_{i,+}^t}\right)\left(1 - \rho^-\right)^{N_{i,-}^t}$. $\tag{5}$

When the area is unprotected ($a^t = 0$), the transition probabilities (3), (4) and (5) depend on parameter μ, to account for the loss in recolonization/survival probability.

4.2   SBM Model of the Prior on Ecological Links

The (known) trophic level of species i is denoted $TL(i)$[2]. By convention, top predators have the largest trophic level, while basal species have trophic level 0.

---

[2] Trophic levels are represented in Fig. 1, right: $TL(1) = 0$, $TL(2) = 1$, $TL(3) = TL(4) = 2$.

Species feed on species in lower trophic levels. So, it is more likely that there is a '+' edge from i to j if $TL(j) > TL(i)$, assuming that most '+' edges model a trophic relation. We will assume here that all positive influences are prey-to-predator ones and that, furthermore, the closer the trophic levels, the more likely i is a prey of j. This a priori knowledge can be modeled by the following SBM, where the blocks are the trophic levels and the block membership function $f(i)$ is defined by $TL(i)$:

$$P\left(G_{ij}^+ = 1\right) = 0 \text{ if } TL(i) \geq TL(j) \text{ and } \frac{e^{\alpha \Delta_{ij}}}{1 + e^{\alpha \Delta_{ij}}} \text{ if } TL(i) < TL(j).$$

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

Negative influences represent different phenomena (negative influence of the predator on its prey, but also parasitism, competition...). We consider a simple probability model for negative influences, only taking into account the relative position of trophic levels.

$$\text{If } TL(i) < TL(j), P\left(G_{ij}^- = 1 \mid G_{ij}^+ = 1\right) = 0 \text{ and } P\left(G_{ij}^- = 1 \mid G_{ij}^+ = 0\right) = \beta_2,$$
$$\text{If } TL(i) = TL(j), \qquad P\left(G_{ij}^- = 1\right) = \beta_2,$$
$$\text{If } TL(i) > TL(j), \qquad P\left(G_{ij}^- = 1\right) = \beta_1.$$

with $\beta_1 > \beta_2$, to represent the fact that predator-to-prey influences are the most frequent negative influences.

The vector $\psi = (\alpha, \beta_1, \beta_2)$ defines the prior on $\mathcal{LG}_\rightarrow$.

5   Ecological Network Learning Algorithm

In this section, we derive a version of the generic Restoration/Estimation algorithm of Sect. 3, specific to the L-DBN model of ecological network.

5.1   Expression of $\log P(D|\mathcal{LG}_\rightarrow, \theta)$

To express the data log-likelihood, we distinguish the basal species (nb species) that have non-empty $Par^+(i, \mathcal{LG}_\rightarrow)$, from the other ones (the basal species b, which have no prey). We also define the quantity $R_{i,C}^{t,d^+,d^-}$ equal to 1 if the species i is of class $C \in \{nb, b\}$ and at time t, $N_{i,+}^t = d^+$, $N_{i,-}^t = d^-$ and 0 otherwise. By convention, for a species of type b, we set $N_{i,+}^t = 0$. We also assume that the maximum overall number of incoming edges of any node i is fixed, equal to k. The log-likelihood of a dataset $D = \{x^1, \dots, x^T\}$, for a given initial state x^0, can be computed as[3]:

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

---

[3] $P_{\mathcal{LG}_\rightarrow, \theta}(x^0)$ will not be estimated.

where $score(i)$ is the contribution of species i to the log-likelihood:

$$score(i) = \sum_{t=0}^{T-1}(1 - x_i^t)\log(P_0^t(x_i^{t+1})) + \sum_{t=0}^{T-1} x_i^t \sum_{0 \leq d^+ + d^- \leq k} \log\left(P_{1,+}^{t,d^+,d^-}(x_i^{t+1})\right) R_{i,nb}^{t,d^+,d^-}$$

$$+ \sum_{t=0}^{T-1} x_i^t \sum_{d^-=0}^{k} \log\left(P_{1,b}^{t,0,d^-}(x_i^{t+1})\right) R_{i,b}^{t,0,d^-} \tag{6}$$

At time t , there is only one term among the three which is non-zero: either the one corresponding to the probability of transition from $x_i^t = 0$ to x_i^{t+1} $(P_0^t(x_i^{t+1}))$ or from $x_i^t = 1$ to x_i^{t+1} for non-basal species $(P_{1,+}^{t,d^+,d^-}(x_i^{t+1}))$ or from $x_i^t = 1$ to x_i^{t+1} for basal species $(P_{1,b}^{t,0,d^-}(x_i^{t+1}))$. The probabilities in Eq. (6) are defined by Eqs. (7) and (8) :

$$\log\left(P_0^t(x_i^{t+1})\right) = x_i^{t+1}a^t\log\epsilon + (1 - x_i^{t+1})a^t\log(1 - \epsilon)$$
$$+ x_i^{t+1}(1 - a^t)\log(\mu\epsilon) + (1 - x_i^{t+1})(1 - a^t)\log(1 - \mu\epsilon). \tag{7}$$

$$\log\left(P_{1,C}^{t,d^+,d^-}(x_i^{t+1})\right) = x_i^{t+1}a^t\log\left(P_{1\to 1}^{1C}(d^+,d^-)\right) + (1 - x_i^{t+1})a^t\log\left(P_{1\to 0}^{1C}(d^+,d^-)\right)$$
$$+ x_i^{t+1}(1 - a^t)\log\left(P_{1\to 1}^{0C}(d^+,d^-)\right)$$
$$+ (1 - x_i^{t+1})(1 - a^t)\log\left(P_{1\to 0}^{0C}(d^+,d^-)\right), \tag{8}$$

where $P_{1\to x_i^{t+1}}^{a^t C}(d^+,d^-)$ is the probability to transition from $x_i^t = 1$ to x_i^{t+1} for species i of type C under action a^t, when it has d^+ favorable and d^- unfavorable species extant. Those probabilities are described in (4) and (5). Note that these expressions are linear functions of the variables $\{R_{i,C}^{t,d^+,d^-}\}$, given the data $\{x_i^t\}$, $\{a^t\}$ and parameters $(\varepsilon, \rho^+, \rho^-, \mu)$ of the model.

5.2 Restoration Step

Let us focus first on the graph update phase (2). If we ignore the SBM part for the moment, the maximization of the first term in (2) can be decomposed into n independent maximization problems (one per $score(i)$). Each maximization problem can be expressed as a 0-1 ILP by introducing auxiliary variables. The auxiliary variables and the linear constraints are provided in the appendix. The SBM term in expression (2) is also decomposable: $\log P(\mathcal{LG}_\to|\psi) = \sum_j score^{SBM}(j)$. The function $score^{SBM}$ writes (provided $\mathcal{LG}_\to$ only contains edges which are consistent with the SBM):

$$score^{SBM}(j) = \sum_{i,\Delta_{ij}=0} g_{ij}^-\log\beta_2 + (1 - g_{ij}^-)\log(1 - \beta_2)$$

$$+ \sum_{i,\Delta_{ij}<0} \alpha\Delta_{ij}g_{ij}^+ - \log(1 + \exp^{\alpha\Delta_{ij}}) + (1 - g_{ij}^+)(g_{ij}^-\log\beta_2 + (1 - g_{ij}^-)\log(1 - \beta_2))$$

$$+ \sum_{i,\Delta_{ij}>0} g_{ij}^-\log\beta_1 + (1 - g_{ij}^-)\log(1 - \beta_1).$$

This expression is not linear in the variables $\{g_{ij}^l\}$. We linearize it by adding an extra variable g_{ij}^{+-} equal to 1 if $g_{ij}^+ = 1$ and $g_{ij}^- = 1$ and 0 otherwise. So doing, the network optimization step (with or without SBM prior) can be performed by solving n independent 0-1 integer linear programs.

5.3 Parameters Estimation Step

Recall that in the parameters update phase (1), parameters vectors θ^{k+1} and ψ^{k+1} can be updated separately: The update of θ is performed using the interior point method for non-linear programming [2]. For β_1 and β_2 the solution of the update is analytic:

$$\beta_1^{k+1} = \frac{\sum_{(i,j),\Delta_{ij}>0} g_{ij}^-}{|\{(i,j),\Delta_{ij} > 0\}|},$$

$$\beta_2^{k+2} = \frac{\sum_{(i,j),\Delta_{ij}\leq 0} g_{ij}^-(1 - g_{ij}^+)}{\sum_{(i,j),\Delta_{ij}\leq 0}(1 - g_{ij}^+)}.$$

The updated α is obtained as a (numerical) solution of the moment-matching equation:

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

6 Experiments

We considered ecological network learning in situations where the sample size is small and we compared the behavior of 4 DBN learning methods corresponding to different levels of embedded a priori knowledge. First the Restoration-Estimation algorithm of Sect. 4 was applied to the L-DBN model of species dynamics (1) without additional knowledge (L-DBN-OK), (2) with a SBM prior (L-DBM-SBM), and (3) with 20% of variables G_{ij}^l known[4] and no SBM prior (L-DBM-20K). The restoration step was solved using the CPLEX solver. We also applied MIT [27] which optimizes a mutual information test score and works with a full (non-parameterized) representation of the TPT. For comparison purposes, we have enriched MIT with an edge-labeling method using the notion of qualitative influence from [28] In qualitative influence, positive and negative influences of a binary variable Y on a binary variable X is defined as follows:

$$Y \xrightarrow{+} X \; iff \; P(X = 1|Y = 1, Z) \geq P(X = 1|Y = 0, Z), \forall Z,$$

$$Y \xrightarrow{-} X \; iff \; P(X = 0|Y = 1, Z) \geq P(X = 0|Y = 0, Z), \forall Z,$$

where Z is the set of other variables influencing X. Replacing probabilities with data counts, we used these definitions to (partially) label the structure learned

---

[4] Known variables were selected uniformly at random.

by MIT (links between variables for which counts do not satisfy any of the above conditions remain unlabeled).

Synthetic networks. We have generated ten synthetic networks of 20 species according to a SBM model with $\alpha = 1/\sqrt{20}, \beta_1 = \alpha/2, \beta_2 = \beta_1/2$. For each of these networks, we have generated 10 data sets, a data set corresponding to a simulated trajectory of length 30 of the species dynamics, with no protection action the first 12 years and protection after. Values of the L-DBN parameters $(\epsilon, \mu, \rho^+, \rho^-)$ were all set to 0.8. The RE algorithm was applied to each data set. So we obtained 10 restored graphs for a single synthetic one. We ordered learnt edges by their decreasing occurrence frequency in these 10 restored graphs, and defined the aggregated graph of size x as the restored graph composed of the x first edges in this ordering. Figure 2 shows the joint evolution of the precision and recall of '+' and '−' edges when the number of edges in the aggregated graph changes. Results for MIT are not reported because precision and recall were close to zero, showing the difficulty to learn both a DBN structure and its TPT in a non parameterized model, when data are scarce.
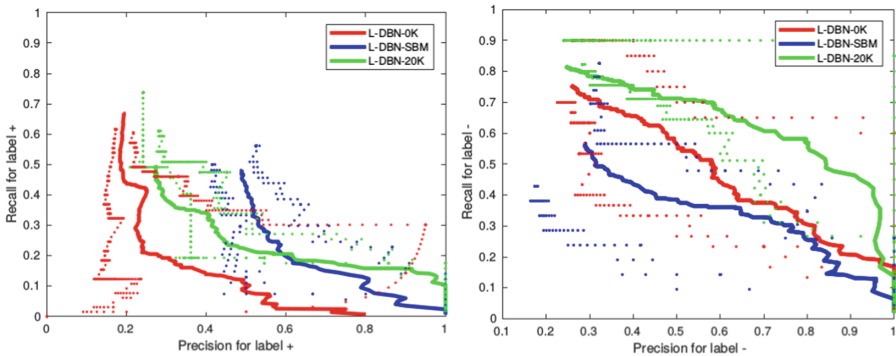

Fig. 2. Synthetic networks. Precision and recall for '+' (left) and '−'(right) edges, for the L-DBN-0K, L-DBN-SBM and L-DBN-20K learning methods. Plain lines are mean values: one dot in these lines is obtained by averaging results for a given value of edges in the aggregated graph, over the 10 data sets. Dotted lines are worse and best cases among the 10 data sets.

We observed that when incorporating a SBM prior in the learning procedure of a L-DBN, fewer edges are learnt. Let us denote by x_{SBM} the maximum number of edges in the aggregated graph built from the 10 L-DBN-SBM restored graphs. When comparing the aggregated graphs with x_{SBM} edges for the different methods, we observed that the one provided by L-DBN-SBM leads to the best precision and recall for '+' edges, and to the best precision and recall when the two labels are not distinguished. Here the SBM knowledge was more helpful than the knowledge of 20% of the edges in the learning process. However, the L-DBN-SBM method was less efficient for learning the '−' edges. This is not

surprising since the prior knowledge embedded in our SBM model is stronger for '+' edges than for '−' edges (it depends on the TL differences).

Real ecological network. We applied the MIT, L-DBN-0K and L-DBN-SBM learning methods on data generated with a L-DBN with the same parameter values as above, but for the real ecological network structure of the Alaskan food web [7]. This network is composed of 13 species, that can be grouped into 5 trophic levels, and contains 21 '+' edges and no '−' edges (see Fig. 3 top left). Here also 10 data sets were used to build an aggregated graph. The precision and recall reached for the aggregated graph composed of all edges learnt at least once were respectively (0.47, 0.33), (0.26, 0.86) and (0.49, 0.81) for MIT, L-DBN-0K and L-DBN-SBM. L-DBN-0K and L-DBN-SBM both learn fewer '−' edges than '+' edges. However, the L-DBN-SBM algorithm provided more parsimonious graphs (35 edges instead of 85 for L-DBN-0K). Figure 3 (bottom left and right) illustrate the gain in integrating the SBM prior: for instance, without SBM knowledge, the information that species do not feed on the same trophic level can not be recovered from the data alone.
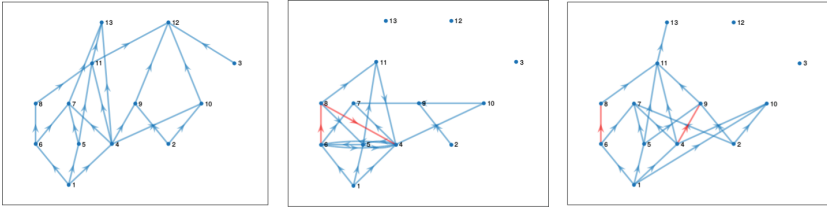

Fig. 3. Alaskan food web. Left: real network, with only '+' edges, Middle and Right: L-DBN-0K and L-DBN-SBM aggregated graphs with 21 edges. Blue edges are '+' edges, while red edges correspond to edges which are learnt both as '+' and '−' edges (Color figure online).

7   Conclusion

We proposed an approach to improve learning of a Dynamic Bayesian Network (DBN) structure *(without synchronous edges)* when data are scarce. The approach combines the definition of a family of parameterized DBN with labeled edges, an a priori Stochastic Block Model (SBM) on the DBN structure and a Restoration-Estimation (RE) learning algorithm. To define a parsimonious parameterization we make the assumption of identical transition probabilities tables for all variables submitted to the same number of each possible type of influence. This is a restrictive but necessary assumption in situations where there is not enough data to learn more complex models. The proposed modeling framework enables us to take into account expert knowledge to help the learning. Our experiments show that by limiting the number of parameters describing the DBN, and by introducing community structure knowledge via SBM,

we can improve learning quality compared to a method based on a full non-parameterized representation of the DBN.

The RE algorithm is a greedy iterative two-steps algorithm. It includes a structure improvement step modeled as n 0-1 integer linear programs, one per variable of the DBN. This procedure is generic since the log-likelihood for a DBN can always be decomposed as a linear function of variables describing the graph structure, as in [5], and as soon as additional constraints on these variables are linear, ILP can be applied. Still, for a specific L-DBN, it is worth deriving a specific ILP model, which will require fewer variables, as we examplified on the problem of learning an ecological interaction network from temporal data of presence/absence of species.

In the ecological network application, the L-DBN transition function is merely an extension of a generic contact process model [12] to more than one influence type. It can also be seen as a DBN with a causal independence model [14] for each transition probabilities table, where each parent's influence is either positive or negative as in a qualitative BN model [18,28]. The proposed model may seem simple compared to the complexity on an ecological network. For instance, we assume identical strengths of positive and negative influences for all species and stationarity of the interaction network structure. The model could be straightforwardly extended to more than two labels, in order to relax the first assumption. Stationarity is more critical and cannot be relaxed without modifying the learning algorithm. Still, propagation by contact models are encountered in several other domains such as fire propagation, health management (disease propagation, [25]), social networks (rumor propagation, [11]), computer science (network security). Therefore (probably with an adaptation of the SBM prior model) the L-DBN model for ecological network and the associated RE algorithm could be useful for learning interaction networks in a wide range of applications.

Appendix: Maximization of the Data Log-likelihood as a 0-1 ILP

We describe how the problem of maximizing the log-likelihood $\log P(D|\mathcal{LG}_\rightarrow, \theta^{t+1})$ over the variables g_{ij}^l defining a DBN structure can be expressed as n independent 0-1 ILP. The expression of the likelihood is an expression of the variables $\left\{R_{i,C}^{t,d^+,d^-}\right\}$ which are themselves functions of the variables g_{ij}^l, as well as of the observed data D. In the following we show how to define the binary variables $\left\{R_{i,C}^{t,d^+,d^-}\right\}$ from linear constraints involving the binary variables $\{g_{ij}^l\}$, the data D and some other auxiliary binary variables. So doing, we will have defined one 0-1 integer linear program per species, which can be solved by classical solvers. The 0-1 integer linear program for a species i describes the vertices pointing to i that maximize the quantity $score(i)$. The following auxiliary binary variables are defined for a particular species i.

– *Non-basal species*: $\left\{h_i^{nb}\right\}$. $h_i^{nb} = 1$ iff $Par^+(i, \mathcal{LG}_\rightarrow) \neq \emptyset$. These variables are defined for all $i \in \{1, .., n\}$ by the constraints described in (9).

- *Lower bound on the number of extant parents:* $\left\{M_{i,l}^{t,d}\right\}$. $M_{i,l}^{t,d} = 1$ iff $N_{i,l}^t \geq d$ (the species i has at least d parents of label type l extant at time t). k is the maximum allowed number of parents of any label. These variables are defined for all $i \in \{1, .., n\}, t \in \{1, .., T\}, d \in \{0, .., k\}, l \in \{+, -\}$ by the constraints in (10).
- *Upper bound on the number of extant parents:* $\left\{\nu_{i,l}^{t,d}\right\}$. $\nu_{i,l}^{t,d} = 1$ iff $N_{i,l}^t \leq d$ (the species i has at most d parents of label type l extant at time t). These variables are defined for all $i \in \{1, .., n\}, t \in \{1, .., T\}, d \in \{0, .., k\}, l \in \{+, -\}$ by the constraints in (11).
- *Number of extant parents:* $\left\{\Lambda_{i,l}^{t,d}\right\}$. $\Lambda_{i,l}^{t,d} = 1$ iff $N_{i,l}^t = d$ (species i has exactly d parents of label type l extant at time t). These variables are defined for all $i \in \{1, .., n\}, t \in \{1, .., T\}, d \in \{0, .., k\}, l \in \{+, -\}$ by the set of constraints in (12).

Now, we are ready to write the linear constraints defining the binary variables $R_{i,C}^{t,d^+,d^-}$. Recall that $R_{i,C}^{t,d^+,d^-} = 1$ if and only if the species i is of type $C \in \{nb, b\}$ and has exactly d^+ parents of type $+$ and d^- parents of type $-$ extant at time t. Thus, $R_{i,nb}^{t,d^+,d^-} = 1$ iff $h_i^{nb} = \Lambda_{i,+}^{t,d^+} = \Lambda_{i,-}^{t,d^-} = 1$. $R_{i,b}^{t,d^+,d^-} = 1$ iff $h_i^{nb} = 0$, $\Lambda_{i,+}^{t,d^+} = \Lambda_{i,-}^{t,d^-} = 1$.

Variables $\left\{R_{i,nb}^{t,d^+,d^-}\right\}$ are defined by the set of constraints (13–14):
$\forall i, j \in \{1, .., n\}$,

$$h_i^{nb} \leq \sum_{j=1}^n g_{ji}^+ \qquad (9)$$

$\forall i \in \{1, .., n\}, t \in \{1, .., T\}, d \in \{0, .., k\}, l \in \{+, -\}$,

$$\begin{aligned} M_{i,l}^{t,d} \cdot (d+1) - \sum_{j=1}^n \left(g_{ji}^l \cdot x_j^t\right) &\leq 1 \\ M_{i,l}^{t,d} \cdot (k+1-d) - \sum_{j=1}^n \left(g_{ji}^l \cdot x_j^t\right) &> -d \end{aligned} \qquad (10)$$

$$\begin{aligned} \nu_{i,l}^{t,d} \cdot (k+1-d) + \sum_{j=1}^n \left(g_{ji}^l \cdot x_j^t\right) &\leq k+1 \\ \nu_{i,l}^{t,d} \cdot (d+1) + \sum_{j=1}^n \left(g_{ji}^l \cdot x_j^t\right) &> d \end{aligned} \qquad (11)$$

$$\Lambda_{i,l}^{t,d} - M_{i,l}^{t,d} \leq 0 \; ; \; \Lambda_{i,l}^{t,d} - \nu_{i,l}^{t,d} \leq 0 \; ; \; \Lambda_{i,l}^{t,d} - M_{i,l}^{t,d} - \nu_{i,l}^{t,d} \geq -1 \qquad (12)$$

$\forall i \in \{1, .., n\}, t \in \{1, .., T\}, d^+ \in \{0, .., k\}, d^- \in \{0, .., k - d^+\}$,

$$R_{i,b}^{t,d^-} \leq \Lambda_{i,l}^{t,d^-} \; ; \; R_{i,b}^{t,d^-} \leq 1 - h_i^{nb} \; ; \; R_{i,b}^{t,d^-} \geq -h_i^{nb} + \Lambda_{i,-}^{t,d^-} \qquad (13)$$

$$\begin{aligned} R_{i,nb}^{t,d^+,d^-} &\leq \Lambda_{i,l}^{t,d^+} \; ; \; R_{i,nb}^{t,d^+,d^-} \leq \Lambda_{i,l}^{t,d^-} \; ; \; R_{i,nb}^{t,d^+,d^-} \leq h_i^{nb} \\ R_{i,nb}^{t,d^+,d^-} &\geq h_i^{nb} + \Lambda_{i,+}^{t,d^+} + \Lambda_{i,-}^{t,d^-} - 2 \end{aligned} \qquad (14)$$

At this stage, all the variables needed for the calculation of $score(i)$, defined in (6), have been introduced. We incorporate further constraints in the ILP formulation to model hard expert knowledge about the ecological network. The species i has at most k parents (constraint (15)). There exists at most one edge from j to i (constraint (16)). If species i is non-basal, it will become extinct at time $t + 1$ if it has no prey at time t (constraint (17)). If the species is basal, it will remain extant at time $t + 1$ if it is extant at time t and has no negative influence (constraint (18)).

$$\sum_{j=1}^{n} \sum_{l \in \{+,-\}} g_{ij}^{l} \leq k, \forall i = 1, \ldots, n \tag{15}$$

$$g_{ji}^{+} + g_{ji}^{-} \leq 1, \forall i = 1, \ldots, n; j = 1, \ldots, n \tag{16}$$

$$R_{i,nb}^{t,d^{+}=0,d^{-}} \cdot x_{i}^{t+1} = 0, \forall i, t, d^{-} \tag{17}$$

$$R_{i,b}^{t,d^{+},d^{-}=0} \cdot x_{i}^{t} \cdot (1 - x_{i}^{t+1}) = 0, \forall i, t, d^{+}. \tag{18}$$

The problem of finding the ecological network structure which optimizes the log-likelihood is now modeled as a set of n 0-1 integer linear programs, whose variables are all the $\{g_{ij}^{l}, M_{i,l}^{t,d}, \nu_{i,l}^{t,d}, \Lambda_{i,l}^{t,d}, R_{i,V}^{t,d^{+},d^{-}}\}$ with constraints (9–18).

Note that the total number of variables and constraints of the 0-1 linear program for a species is linear in n and quadratic in k. Thus, the complexity of the graph update phase is "simply" exponential in n.

References

1. Ambroise, C., Chiquet, J., Matias, C.: Inferring sparse Gaussian graphical models with latent structure. Electron. J. Stat. **3**, 205–238 (2009)
2. Byrd, R.H., Hribar, M.E., Nocedal, J.: An interior point algorithm for large-scale nonlinear programming. SIAM J. Optim. **9**(4), 877–900 (1999)
3. de Campos, C.P., Ji, Q.: Efficient structure learning of Bayesian networks using constraints. J. Mach. Learn. Res. **12**, 663–689 (2011)
4. Chickering, D.M.: Learning Bayesian networks is NP-complete. In: Learning from data, pp. 121–130 (1996)
5. Cussens, J.: Bayesian network learning with cutting planes. In: UAI 2011, Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (2011)
6. Dojer, N.: Learning Bayesian networks does not have to be NP-hard. In: Královič, R., Urzyczyn, P. (eds.) MFCS 2006. LNCS, vol. 4162, pp. 305–314. Springer, Heidelberg (2006). https://doi.org/10.1007/11821069_27
7. Estes, J.A., Doak, D.F., Springer, A.M., Williams, T.M.: Causes and consequences of marine mammal population declines in southwest Alaska: a food-web perspective. Philos. Trans. R. Soc. Lond. B: Biol. Sci. **364**, 1647–1658 (2009)

8. Feelders, A., van der Gaag, L.: Learning Bayesian network parameters under order constraints. Int. J. Approx. Reason. **42**, 37–53 (2006)
9. Franc, A.: Metapopulation dynamics as a contact process on a graph. Ecol. Complex. **1**(1), 49–63 (2004)
10. Ghahramani, Z.: Learning dynamic Bayesian networks. In: Giles, C.L., Gori, M. (eds.) NN 1997. LNCS, vol. 1387, pp. 168–197. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0053999
11. Gomez Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1019–1028. ACM (2010)
12. Harris, T.E.: Contact interactions on a lattice. Ann. Probab. **2**, 969–988 (1974)
13. Heckerman, D., Geiger, D., Chickering, D.M.: Learning Bayesian networks: the combination of knowledge and statistical data. Mach. Learn. **20**(3), 197–243 (1995)
14. Heckerman, D., Breese, J.S.: Causal independence for probability assessment and inference using Bayesian networks. IEEE Trans. Syst. Man Cybern. Part A **26**(6), 826–831 (1996)
15. Holland, P., Laskey, K., Leinhardt, S.: Stochastic blockmodels: first steps. Soc. Netw. **5**(2), 109–137 (1983)
16. Korhonen, J.H., Parviainen, P.: Tractable Bayesian network structure learning with bounded vertex cover number. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
17. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. **58**(7), 1019–1031 (2007)
18. Lucas, P.J.F.: Bayesian network modelling through qualitative patterns. Artif. Intell. **163**(2), 233–263 (2005)
19. Milns, I., Beale, C.M., Smith, V.A.: Revealing ecological networks using Bayesian network inference algorithms. Ecology **91**(7), 1892–1899 (2010)
20. Murphy, K.: Dynamic Bayesian Networks: representation, inference and learning. Ph.D. thesis, University of California (2002)
21. Newman, M.E.J., Clauset, A.: Structure and inference in annotated networks. Nat. Commun., 7 (2016)
22. Niculescu, R.S., Mitchell, T.M., Rao, R.B.: Bayesian network learning with parameter constraints. J. Mach. Learn. Res. **7**(July), 1357–1383 (2006)
23. Oyen, D., Anderson, B., Anderson-Cook, C.: Bayesian networks with prior knowledge for malware phylogenetics. In: AAAI-2016 Workshop on Artificial Intelligence and Cybersecurity (2016)
24. Parviainen, P., Farahani, H.S., Lagergren, J.: Learning bounded tree-width Bayesian networks using integer linear programming. In: Seventeenth International Conference on Articial Intelligence and Statistics (AISTAT 2014) (2014)
25. Salathé, M., Jones, J.H.: Dynamics and control of diseases in networks with community structure. PLOS Comput. Biol. **6**(4), e1000736 (2010)
26. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)
27. Vinh, N.X., Chetty, M., Coppel, R., Wangikar, P.P.: Polynomial time algorithm for learning globally optimal dynamic Bayesian network. In: Lu, B.-L., Zhang, L., Kwok, J. (eds.) ICONIP 2011. LNCS, vol. 7064, pp. 719–729. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24965-5_81
28. Wellman, M.P.: Fundamental concepts of qualitative probabilistic networks. Artif. Intell. **44**, 257–303 (1990)
29. Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., Jarvis, E.D.: Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics **20**(18), 3594–3603 (2004)