International Journal of Approximate Reasoning 80 (2017) 460–474

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Hybrid time Bayesian networks ☆

Manxia Liu [a,1], Arjen Hommersom [c,a,2], Maarten van der Heijden [a,*,2], Peter J.F. Lucas [a,b]

[a] Radboud University, ICIS, Nijmegen, The Netherlands
[b] Leiden University, LIACS, Leiden, The Netherlands
[c] Open University, MST, Heerlen, The Netherlands

ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form 19 February 2016
Accepted 28 February 2016
Available online 9 March 2016

Keywords:
Continuous time Bayesian networks
Dynamic Bayesian networks
Dynamic systems

ABSTRACT

Capturing heterogeneous dynamic systems in a probabilistic model is a challenging problem. A single time granularity, such as employed by dynamic Bayesian networks, provides insufficient flexibility to capture the dynamics of many real-world processes. The alternative is to assume that time is continuous, giving rise to continuous time Bayesian networks. Here the problem is that the level of temporal detail is too precise to match available probabilistic knowledge. In this paper, we present a novel class of models, called hybrid time Bayesian networks, which combine discrete-time and continuous-time Bayesian networks. The new formalism allows us to more naturally model dynamic systems with regular and irregularly changing variables. We also present a mechanism to construct discrete-time versions of hybrid models and an EM-based algorithm to learn the parameters of the resulting BNs. Its usefulness is illustrated by means of a real-world medical problem.

1. Introduction

Many real-world systems exhibit complex and rich dynamic behavior. As a consequence, capturing these dynamics is an integral part of developing models of physical-world systems. Time granularity is an important parameter in characterizing dynamics as it determines the level of temporal detail in the model. In cases where one time granularity is coarser than another, dealing with multiple time granularities becomes significantly important, e.g., in the context of mining frequent patterns and temporal relationships in data streams and databases [2].

Bayesian networks (BNs) have been very successful in modeling complex situations involving uncertainty [3]. Dynamic Bayesian networks (DBNs) are part of the Bayesian network framework, supporting the modeling of dynamic probabilistic systems [4]. DBNs extend standard Bayesian networks by assuming that changes in a process can be captured by a sequence of states at discrete time points. Usually the assumption is made that the distribution of variables at a particular time point is conditional only on the state of the system at the previous time point. A problem occurs if temporal processes of a system are best described using different rates of change, e.g., one temporal part of the process changes much faster than another. In that case, the whole system has to be represented using the finest time granularity, which is undesirable from a

☆ This is an extended version of a conference paper with the same title [1].

* Corresponding author.
E-mail addresses: m.liu@cs.ru.nl (M. Liu), arjenh@cs.ru.nl (A. Hommersom), m.vanderheijden@cs.ru.nl (M. van der Heijden), peterl@cs.ru.nl (P.J.F. Lucas).

[1] ML is supported by the China Scholarship Council.
[2] AH and MvdH are supported by the ITEA2 MoSHCA project (ITEA2-ip11027).

http://dx.doi.org/10.1016/j.ijar.2016.02.009
0888-613X/

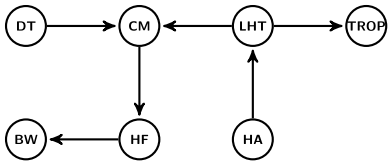


Fig. 1. Causal model for heart failure: CM = Contractility Myocardium, DT = Digitalis, LHT = Loss Heart Tissues, HA = Heart Attack, TROP = Troponin, HF = Heart Failure, BW = Body Weight.

modeling and learning perspective. In particular, if a variable is observed irregularly, much data on discrete-time points will be missing and conditional probabilities will be hard to estimate.

As an alternative to DBNs, temporal processes can be modeled as continuous time Bayesian networks (CTBNs), where time acts as a continuous parameter [5]. In these models, the time granularity is infinitely small by modeling transition rates rather than conditional probabilities. Thus, multiple time granularities, i.e., slow and fast transition rates, can easily be captured. A limitation from a modeling perspective is that all probabilistic knowledge, for example derived from expert knowledge, has to be mapped to transition rates which are hard to interpret. Moreover, it is assumed that the transition rates, governing the time until a transition occurs, are exponentially distributed, which may not always be appropriate.

In this paper, we propose a new formalism, which we call *hybrid time Bayesian networks* (HTBNs), inspired by discrete-time and continuous-time Bayesian networks. We develop the theoretical properties of HTBNs and show their practical use by means of a medical example. HTBNs facilitate modeling the dynamics of both irregularly-timed random variables and random variables whose evolution is naturally described by discrete time. As a result, the new formalism increases the modeling and analysis capabilities for dynamic systems.

This paper is an extended version of a conference paper with the same title [1]. In the next section we introduce the running, clinical example for the rest of this paper, followed by preliminaries on DBNs and CTBNs to fix the notation. Then, in Section 4, we define HTBNs with their associated factorization, followed by a construction that allows transforming an HTBN into an equivalent BN. Subsequently we return to our running example and demonstrate how the equivalent BN can be used to obtain a meaningful clinical simulation. The paper is concluded by a discussion.

2. Motivating example

To illustrate the usefulness of the proposed theory, we consider the medical problem of heart failure and, in particular, one possible cause of heart failure: heart attack (myocardial infarction). This usually occurs as the result of coronary artery disease giving rise to reduced blood supply to the heart muscle (myocardium). One consequence is that part of the heart muscle will die, which is revealed later in a blood sample analysis in the lab by an increased level of particular heart muscle proteins, in particular troponin. Loss of heart muscle will inevitably have an impact on the contractility of the myocardium, and thus heart function will be negatively affected. This is known as *heart failure*. In particular, the heart fails with respect to its function as a pump. This will enforce an increase in the amount of extracellular fluid (the patient is flooded with water), which can be measured quite simply by means of the body weight. With regard to treatment, digitalis is considered as one of the drugs to improve contractility. This causal knowledge is formalized as a directed graph in Fig. 1.

Heart attacks can occur repeatedly in patients, although after some interval of time, and this may negatively affect heart function. After administration of digitalis it will take some time before the drug has a diminishing effect on heart failure. Thus, the course of heart failure will likely depend on various factors, and how they interact. Of particular importance here is the dynamic over time of the probability distributions.

In modeling processes such as heart failure, it is essential to notice the existence of different time granularities. There are *discrete*, *regular* variables which are observed regularly such as a routine checkup for body weight and a regular intake of a drug. On the other hand, some variables are observed *irregularly*, such as the indicator troponin which is elevated after about half an hour after damage to the heart muscle is obtained; however its measurement is repeated with time intervals that increase after the patient's condition has been stabilized. Clearly, it is not possible to obtain a satisfactory representation of the clinical evolution of heart failure using only discrete time, regular or irregular, or continuous time. In the remainder of this paper we propose a method to deal with these heterogeneous time aspects.

3. Preliminaries

We start by introducing Bayesian networks, dynamic Bayesian networks and continuous time Bayesian networks. In the following, upper-case letters, e.g. X, Y, or upper-case strings, e.g. HA, denote random variables. We denote the values of a variable by lower-case letters, e.g. x, or by T or F, short for *true* and *false*. Note that all variables considered here have a finite domain of values. Continuous-time variables are variables that have a finite domain of values over infinite trajectories. For discrete-time variables time changes regularly, evenly, whereas continuous-time variables are irregularly spaced over time. In what follows we will make use of a successor function s, which is defined on a countable, linearly ordered set of numbers Z in which every element $z_i \in Z$ with index i is mapped to element $s(z_i) = z_{i+1} \in Z$ (with the potentially greater

than last element z_{i+1} excluded). If the set Z consists of natural numbers, then we also assume that $s(z_i) = z_i + 1 = z_{i+1}$, with $z_i \in Z$.

3.1. Bayesian networks

A Bayesian network is a probabilistic graphical model which represents a joint probability distribution of a set of random variables. A *Bayesian network* $\mathcal{B}$ is defined as a pair $\mathcal{B} = (G, P)$, where G is an acyclic directed graph with $G = (V(G), E(G))$, where $V(G)$ is a set of nodes, and $E(G) \subseteq V(G) \times V(G)$ a set of directed edges or arcs. A joint probability distribution P is defined by a set of conditional probabilities of each random variable X given its immediate parents $\pi(X)$ in G, formally:

$$P(V(G)) = \prod_{X \in V(G)} P(X \mid \pi(X))$$

Example 3.1. Let the cardiac disease example in Fig. 1 be a Bayesian network $\mathcal{B}$ with $V(G) = \{\text{DT}, \text{CM}, \text{LHT}, \text{TROP}, \text{HF}, \text{HA}, \text{BW}\}$. We obtain the joint probability P by multiplying conditional distributions over the variables $V(G)$ according to the graph:

$$\begin{aligned} P(V(G)) = {} & P(\text{CM} \mid \text{DT}, \text{LHT})\, P(\text{HF} \mid \text{CM})\, P(\text{BW} \mid \text{HF})\, P(\text{LHT} \mid \text{HA}) \\ & P(\text{TROP} \mid \text{LHT})\, P(\text{HA})\, P(\text{DT}) \end{aligned}$$

3.2. Dynamic Bayesian Networks (DBNs)

A DBN is defined as a pair $(\mathcal{B}_0, \mathcal{B}_{\rightarrow})$ over discrete-time variables $\mathbf{D}$, where $\mathcal{B}_0$ is taken as an initial Bayesian network model, and $\mathcal{B}_{\rightarrow}$ is defined as a conditional distribution for a 2-time-slice Bayesian network (2-TBN) given by:

$$P(\mathbf{D}_{s(\alpha)} \mid \mathbf{D}_{\alpha}) = \prod_{D \in \mathbf{D}} P(D_{s(\alpha)} \mid \pi(D_{s(\alpha)}))$$

for some time-point α.

Given a set of discrete time points of interest $\mathbf{A}$, an initial segment of $\mathbb{N}_0$, $\mathbf{A} = \{0, 1, 2, \ldots, n\}$, the joint distribution for a DBN with $|\mathbf{A}|$ slices is defined by a product of the conditional probability distributions in the initial model and in the 2-TBN:

$$P(\mathbf{D}_{\mathbf{A}}) = \prod_{D \in \mathbf{D}} P_{\mathcal{B}_0}(D_0 \mid \pi(D_0)) \prod_{D \in \mathbf{D}} \prod_{\alpha \in \mathbf{A} \setminus \{\max \mathbf{A}\}} P_{\mathcal{B}_{\rightarrow}}(D_{s(\alpha)} \mid \pi(D_{s(\alpha)}))$$

where $D_{s(\alpha)}$ is the random variable D at time $s(\alpha)$. The parent set $\pi(D_{s(\alpha)})$ includes variables from the same or the previous time slice. We can obtain a standard Bayesian network by unrolling the DBN over the time points of interest $\mathbf{A}$. In the remainder it is assumed that the intra-slice arcs of this BN are the same for every α.

Example 3.2. Consider a dynamic Bayesian network that has two random variables, HF and BW (see Fig. 2), with an initial model and a transition model as shown in Fig. 2a and 2b, respectively. Then the joint distribution for the DBN over time points of interest $\mathbf{A}$ with the corresponding Bayesian network as shown in Fig. 2c is: $P(\text{HF}_{\mathbf{A}}, \text{BW}_{\mathbf{A}}) = P(\text{HF}_0) P(\text{BW}_0 | \text{HF}_0) \prod_{\alpha=0}^{|\mathbf{A}|-2} P(\text{BW}_{s(\alpha)} | \text{BW}_{\alpha}, \text{HF}_{s(\alpha)}) P(\text{HF}_{s(\alpha)} | \text{HF}_{\alpha})$.

3.3. Continuous Time Bayesian Networks (CTBNs)

CTBNs [5] represent dynamic systems with continuous-time variables as a factorized homogeneous Markov process parameterized by *intensity matrices*. A CTBN is defined as a tuple $\mathcal{N} = (\mathcal{B}, G_{\rightarrow}, \Lambda)$, where $\mathcal{B} = (G_0, P)$ denotes the Bayesian network that specifies the initial model; $G_{\rightarrow}$ denotes the graph of the transition model and Λ is a set of intensity matrices.

An entry (i, j) with $i \neq j$ in an intensity matrix gives the intensity of transitioning from state i to state j. Furthermore, the main diagonal makes each row sum to zero.

Example 3.3. Suppose we want to model the random process of body weight as the variable BW, which describes a patient's weight. Variable BW has three possible states, i.e., $\text{BW} = \{\text{low}, \text{normal}, \text{high}\}$, with an intensity matrix as follows:

$$Q_{\text{BW}} = \begin{pmatrix} -0.13 & 0.09 & 0.04 \\ 0.13 & -0.23 & 0.1 \\ 0.07 & 0.16 & -0.23 \end{pmatrix}$$

For example, the entry $(3, 2)$ means that the process will transition from high at time β to normal at time $\beta + \epsilon$ with a probability of $0.16/0.23 = 0.696$ if a transition occurs at $\beta + \epsilon$, where $\epsilon \downarrow 0$.

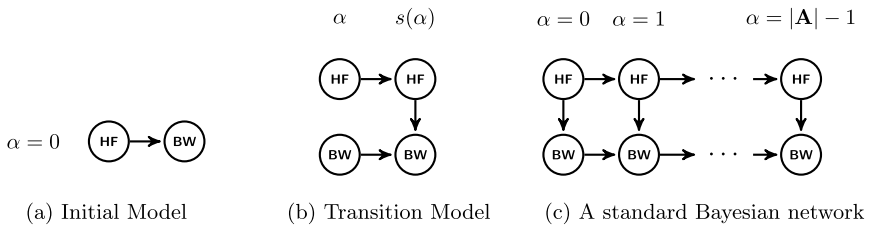


Fig. 2. A DBN and its corresponding Bayesian network.

The notion of a *conditional intensity matrix* (CIM) describes the dependence of a variable C on the current values of its parents $\pi(C)$. A *full amalgamation* product operator $*$ [5] is defined over a set of CIMs to compute the joint intensity matrix, resulting in a single continuous-time Markov process for the entire system.

Example 3.4. Suppose we have a CTBN with graph HF $\rightarrow$ BW with CIMs:

$$Q_{\text{HF}} = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}$$

$$Q_{\text{BW}|\text{HF}} = \begin{pmatrix} -0.13 & 0.09 & 0.04 \\ 0.13 & -0.23 & 0.1 \\ 0.07 & 0.16 & -0.23 \end{pmatrix} \quad Q_{\text{BW}|\overline{\text{HF}}} = \begin{pmatrix} -0.04 & 0.01 & 0.03 \\ 0.05 & -0.3 & 0.25 \\ 0.1 & 0.2 & -0.3 \end{pmatrix}$$

Let the states in the joint intensity matrix be ordered as (low, T), (normal, T), (high, T), (low, F), (normal, F), (high, F), then the joint intensity matrix is computed by the amalgamation operation, which looks up the intensity for the changing variable in the corresponding CIM. Therefore, the joint intensity matrix is:

$$Q_{\text{BWHF}} = Q_{\text{BW}|\text{HF}} * Q_{\text{HF}} = \begin{pmatrix} -1.13 & 0.09 & 0.04 & 1 & 0 & 0 \\ 0.13 & -1.23 & 0.1 & 0 & 1 & 0 \\ 0.07 & 0.16 & -1.23 & 0 & 0 & 1 \\ 2 & 0 & 0 & -2.04 & 0.01 & 0.03 \\ 0 & 2 & 0 & 0.05 & -2.3 & 0.25 \\ 0 & 0 & 2 & 0.1 & 0.2 & -2.3 \end{pmatrix}$$

For a homogeneous Markov process over variables $\mathbf{C}$ with an intensity matrix $Q_{\mathbf{C}}$ and an initial distribution $P(\mathbf{C}_0)$, we can compute the distribution over the values of $\mathbf{C}$ at a particular time point or the joint distribution at different time points. The distribution at time point β is given by:

$$P(\mathbf{C}_\beta) = P(\mathbf{C}_0)\exp(Q_{\mathbf{C}}\beta)$$

The distribution over a finite set of time points of interest $\mathbf{B}$ is given by:

$$P(\mathbf{C}_\mathbf{B}) = P(\mathbf{C}_0) \prod_{\beta \in \mathbf{B} \setminus \{\max \mathbf{B}\}} \exp(Q_{\mathbf{C}}(s(\beta) - \beta))$$

Below we will sometimes use only the graphical structure of the CTBN, i.e., the *CTBN graph*, which is the tuple $(G_0, G_\rightarrow)$, where G_0 is the initial Bayesian network graph and $G_\rightarrow$ the transition graph. Furthermore, we will call $\mathbf{V}_\mathbf{B}$ the associated *time-indexed variables* if $\mathbf{V}$ is the set of variables in the CTBN, i.e., $\mathbf{V} = V(G_0)$ and $\mathbf{B}$ is a set of time points of interest.

4. Hybrid time Bayesian networks

In this section, we define hybrid-time Bayesian networks and give the semantics of these models in terms of their factorization.

4.1. General idea

Random variables model dynamical systems evolving in discrete time steps or smoothly over time. DBNs are a general framework for modeling dynamic probabilistic systems on a set of discrete time points. The dynamics are parameterized by a conditional probability distribution of variables at time $s(\alpha)$ conditioned on variables at time α. Building a DBN requires finding a single time granularity, though variables change at their own rate. When the modeled time granularity is coarser than the time spent between two consecutive observations, the conditional probabilities for some observations between discrete time slices are difficult to estimate, that is, it can give rise to information loss. In contrast, when time is continuous, intensity matrices in CTBNs are used to describe transient behavior of random variables. Transition probabilities

can be induced from intensity matrices, yet, these probabilities cannot be represented explicitly in CTBNs. This implies that probabilistic knowledge from domain experts has to be mapped to intensity matrices which are hard to interpret. As a result, the question arises whether it is possible to have a probabilistic representation modeling dynamical systems with multiple changing rates. One answer is to define a probabilistic representation for modeling both discrete-time and continuous-time variables.

Dependences of random variables are represented by arcs in probabilistic graphical models, such as DBNs and CTBNs. In discrete-time models one can distinguish arcs which are inter-time-slice (also known as temporal arcs), going between time slices, or intra-time-slice (called atemporal arcs), connecting variables within the same time slice. The decision on how to relate two variables is dependent on how tight the coupling between them is in time. If the dependence of one variable with another is immediate, that is, much shorter than the time granularity in the model, an atemporal arc is appropriate. If the dependence manifests after some time but longer than modeled time granularity, this can be modeled as an arc from one slice to the next. In contrast, arcs in CTBNs are always temporal: the dependence between variables manifests itself in the probability distribution of states mediated by an intensity matrix. The time it takes for a variable to influence another can be arbitrarily small, yet non-zero, which we denote by ϵ.

In these two modeling approaches, the common factor is that the dependences manifest a *delay*. In DBNs dependences have delay zero represented as atemporal arcs or some non-zero delay for temporal arcs, whereas in CTBNs all arcs have the same delay ϵ. For our definition of hybrid-time Bayesian networks, we can therefore incorporate continuous-time and discrete-time BNs in a single graph by assigning a delay to each arc, capturing the dependence behavior over time in both DBNs and CTBNs.

4.2. Model definition

To define hybrid-time models, we first formalize the concept of arcs that represent delayed dependence. To this end, we define an arc with an associated attribute *delay*, denoted as d, explicitly indicating when the influence between variables manifests. The value of the delay can take either a natural number n, $n \in \mathbb{N}_0$, or ϵ, $\epsilon \downarrow 0$. Thus, the dependences in CTBNs can be represented by arcs with a delay having a value ϵ. Whereas atemporal and temporal dependences in DBNs are represented by arcs having a delay 0 and 1, respectively (if we restrict ourselves to first-order Markov models).

Now we can give a formal definition for arcs in a hybrid-time graph. Given a graph $G = (V(G), E(G))$, the arcs are defined as $E(G) \subseteq V(G) \times V(G) \times (\mathbb{N}_0 \cup \{\epsilon\})$, each of which represents direct dependence of one variable on another.

The formal definition of hybrid time Bayesian networks is as follows.

Definition 4.1 *(Hybrid Time Bayesian Networks (HTBNs)).* A hybrid time Bayesian network is a tuple $\mathcal{H} = (\mathcal{B}, G_{\mathcal{H}}, \Phi, \Lambda)$, where $\mathcal{B} = (G_0, P)$ is a Bayesian network specifying an initial distribution, $G_{\mathcal{H}} = (V(G_{\mathcal{H}}), E(G_{\mathcal{H}}))$ is a directed graph specifying a transition model with each vertex in $V(G_{\mathcal{H}})$ either a continuous-time variable, collectively denoted by **C**, or a discrete-time variable, collectively denoted by **D**, Φ is a set of conditional probability distributions for variables **D**, and Λ is a set of conditional intensity matrices for variables **C**. Furthermore, graph $G_{\mathcal{H}}$ has the following properties:

1. An arc to a continuous-time variable has a delay ϵ;
2. An arc between discrete-time variables has a delay d, $d \in \mathbb{N}_0$;
3. An arc from a continuous-time variable to a discrete-time variable has delay 0;
4. $(V(G_{\mathcal{H}}), \{(x, y) \mid (x, y, 0) \in E(G_{\mathcal{H}})\})$ is acyclic.

A cycle of arcs with delay 0 is not permitted, analogous to the requirement for BN graphs to be acyclic. A cycle containing arcs with non-zero delay, however, is allowed as an inherited property from discrete-time and continuous-time Bayesian networks. Although in principle arbitrary delays are possible, we here restrict ourselves to temporal dependences between continuous time variables with delay ϵ.

Example 4.1. In the example discussed in Section 2, regular variables, i.e., BW, DT, HF and the hidden variable CM can be represented in a discrete-time manner. The irregular variables, i.e., LHT, TROP, HA are modeled as continuous-time variables. The example is then represented in a hybrid time Bayesian network $\mathcal{H}$ as shown in Fig. 3.

4.3. Factorization

The joint probability distribution for hybrid time Bayesian networks is defined by the multiplication of the conditional joint probabilities for the continuous-time and discrete-time Bayesian network parts. To this end, we first need to introduce some graph theoretical notions.

The *skeleton* $G^{\sim}$ of a directed graph G is obtained by changing the arcs in G to (undirected) edges. Every directed graph can be defined as the union of *connected components* by an equivalence relation $X - Y$, meaning that vertex Y can be reached by an undirected path from vertex X in its skeleton. Vertices X and Y are then members of the same equivalence class $[X]$ and the corresponding graph is a connected component. A graph $G' = (V(G'), E(G'))$ is said to be a *continuous-time*

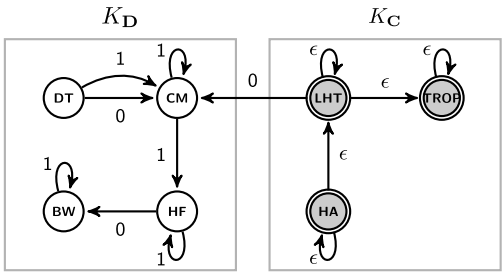


Fig. 3. A transition model of an HTBN for the heart failure problem. Continuous-time variables are graphically represented by double-edged shaded circles.

induced subgraph of G, denoted as $G_{\mathbf{C}}$, if $V(G') = \mathbf{C}$ and $E(G') = (V(G') \times V(G') \times \{\epsilon\}) \cap (V(G) \times V(G) \times \{\epsilon\})$. Similarly, G' is a *discrete-time induced subgraph* of G, denoted as $G_{\mathbf{D}}$, if $V(G') = \mathbf{D}$ and $E(G') = (V(G') \times V(G') \times \mathbb{N}_0) \cap (V(G) \times V(G) \times \mathbb{N}_0)$.

Both $G_{\mathbf{C}}$ and $G_{\mathbf{D}}$ can be decomposed into connected components; each individual connected component is indicated by $K_{\mathbf{C}}$ and $K_{\mathbf{D}}$, respectively. Clearly connected components are disjoint as they represent equivalence classes and together the connected components form partitions of the continuous-time and discrete-time subgraphs, respectively. A subset $\mathbf{X} \subseteq V(G_{\mathbf{D}})$ is said to constitute the parents of $V(K_{\mathbf{C}})$, denoted as $\pi(V(K_{\mathbf{C}}))$, if and only if there exists at least an arc (D, C) in G, $C \in V(K_{\mathbf{C}})$, for every $D \in \mathbf{X}$. Parents $\pi(V(K_{\mathbf{D}}))$ are defined analogously.

Example 4.2. In the graph shown in Fig. 3, there is only one continuous-time connected component with $V(K_{\mathbf{C}}) = \{\text{LHT}, \text{TROP}, \text{HA}\}$ and one discrete-time connected component with $V(K_{\mathbf{D}}) = \{\text{DT}, \text{CM}, \text{HF}, \text{BW}\}$.

We are now in the position to define a conditional distribution of connected components given their parents.

Definition 4.2 *(Conditional joint distribution for component $K_{\mathbf{D}}$).* Given a discrete-time component $K_{\mathbf{D}}$, the conditional joint distribution for $K_{\mathbf{D}}$ over time points of interest $\mathbf{A}$ is defined as:

$$P(V(K_{\mathbf{D}})_{\mathbf{A}} \mid \pi(V(K_{\mathbf{D}}))_{\mathbf{A}}) = \prod_{D \in V(K_{\mathbf{D}})} (P(D_0 \mid \pi(D)_0) \prod_{\alpha \in \mathbf{A} \setminus \{0\}} P(D_\alpha \mid \pi(D)_{\alpha - d}))$$

Definition 4.3 *(Conditional joint distribution for component $K_{\mathbf{C}}$).* Given a continuous-time component $K_{\mathbf{C}}$ over variables $V(K_{\mathbf{C}})$ with an initial distribution $P(V(K_{\mathbf{C}})_0)$ and corresponding parents $\pi(V(K_{\mathbf{C}}))$ over time points of interest $\mathbf{A}$. The conditional joint distribution for $K_{\mathbf{C}}$ over a finite set of time points of interest $\mathbf{B}$, $\{0\} \subset A \subseteq \mathbf{B} \subset \mathbb{R}^+$, is defined as:

$$P(V(K_{\mathbf{C}})_{\mathbf{B}} \mid \pi(V(K_{\mathbf{C}}))_{\mathbf{A}}) = P(V(K_{\mathbf{C}})_0) \prod_{\beta \in \mathbf{B} \setminus \{\max \mathbf{B}\}} \exp(Q_{V(K_{\mathbf{C}}) \mid \pi(V(K_{\mathbf{C}}))_a}(s(\beta) - \beta))$$
$$a = \max\{\alpha \mid \alpha < \beta, \alpha \in \mathbf{A}\}$$

where $Q_{V(K_{\mathbf{C}}) \mid \pi(V(K_{\mathbf{C}}))_a}$ is the conditional intensity matrix for variables $V(K_{\mathbf{C}})$ given the values of parents $\pi(V(K_{\mathbf{C}}))$ at time a.

Now we can define the full joint probability distribution of a hybrid-time BN given sets of time points of interest.

Definition 4.4 *(Joint probability distribution).* Given a hybrid time Bayesian network $\mathcal{H}$ and sets of components $\mathbf{K_D}$, $\mathbf{K_C}$ with associated time points of interest $\mathbf{A}$, $\mathbf{B}$. The joint distribution for $\mathcal{H}$ over $\mathbf{B}$ is defined as:

$$P(V(G)_{\mathbf{B}}) = \prod_{K_{\mathbf{C}} \in \mathbf{K_C}} P(V(K_{\mathbf{C}})_{\mathbf{B}} \mid \pi(V(K_{\mathbf{C}}))_{\mathbf{A}}) \prod_{K_{\mathbf{D}} \in \mathbf{K_D}} P(V(K_{\mathbf{D}})_{\mathbf{A}} \mid \pi(V(K_{\mathbf{D}}))_{\mathbf{A}})$$

Example 4.3. Consider the example in Fig. 3 with time points of interest $\mathbf{A}$ and $\mathbf{B}$ and joint intensity matrix Q for continuous-time variables LHT, HA and TROP. As the continuous component has no parents, the joint distribution of the transition model is then given by:

$$P = \prod_{\alpha \in \mathbf{A} \setminus \{0\}} P(\text{DT}_{s(\alpha)}) P(\text{BW}_{s(\alpha)} \mid \text{BW}_\alpha, \text{HF}_{s(\alpha)}) P(\text{HF}_{s(\alpha)} \mid \text{HF}_\alpha, \text{CM}_\alpha)$$
$$P(\text{CM}_{s(\alpha)} \mid \text{CM}_\alpha, \text{DT}_\alpha, \text{DT}_{s(\alpha)}, \text{LHT}_{s(\alpha)}) \prod_{\beta \in \mathbf{B} \setminus \{\max \mathbf{B}\}} P(\exp(Q(s(\beta) - \beta)))$$

The following propositions establish that HTBNs are proper generalizations of both DBNs and CTBNs. For DBNs, the transition graph can be converted to an HTBN graph, where the intra-temporal arcs are replaced by arcs with delay 0 and inter-temporal arcs are replaced by arcs with delay 1.

Proposition 4.1. *A DBN* $(\mathcal{B}_0, \mathcal{B}_\rightarrow)$ *and an HTBN* $(\mathcal{B}, G_\mathcal{H}, \Phi, \varnothing)$ *define the same joint probability distribution for any set of time points of interest* **A**, *if* $\mathcal{B}_0 = \mathcal{B}$, $G_\mathcal{H}$ *corresponds to the graph of* $\mathcal{B}_\rightarrow$, *and* Φ *are the parameters of the DBN.*

Similarly, a CTBN can be interpreted as an HTBN by replacing its arcs by arcs with ϵ delay.

Proposition 4.2. *A CTBN* $(\mathcal{B}, G_\rightarrow, \Lambda)$ *and an HTBN* $(\mathcal{B}, G_\mathcal{H}, \varnothing, \Lambda)$ *define the same probability distribution for any set of time points of interest* **B** *if* $G_\mathcal{H}$ *is* $G_\rightarrow$ *such that each edge is replaced by an edge with delay* ϵ.

For both propositions, the results follow directly from the factorization of CTBNs, DBNs, and HTBNs.

5. Discrete-time characterization

A natural question is whether the joint distribution defined on an HTBN, given the fixed time points of interest, can also be graphically represented as a regular (discrete-time) Bayesian network. The benefit is that the parameters of the resulting Bayesian network are conditional probabilities, which are easier to understand for domain experts. Furthermore, this construction is convenient as this implies that we may (dynamically) generate discrete-time versions of the model given time points for which we have observations, and in which we would like to compute marginals. Given this translation, existing algorithms for probabilistic inference in BNs may be employed.

In Section 5.1 we describe the procedure to construct a discrete-time graph structure of a hybrid model. The construction for the complete representative BN is given in Section 5.2, employing an EM algorithm to obtain the parameters.

5.1. Structural discretization

The discretization of a hybrid-time model requires discretizing the continuous components $K_\mathbf{C}$ given time points of interest **B**. Since each continuous component is itself a CTBN, we focus in this subsection mainly on the discretization procedure in terms of CTBNs. The resulting CTBNs are subsequently combined with the discrete components $K_\mathbf{D}$ to obtain the discrete version of the HTBN.

The remainder of this subsection is structured as follows. In Definition 5.1 and Lemma 1 we characterize the independence structure of a CTBN in terms of an infinite set of Bayesian networks which follow the causal structure of the graph. Then, in Definition 5.2 and Theorem 2, we show the existence of a single Bayesian network, called a *representative Bayesian network*, that includes all possible dependences of this set of causal networks, and therefore of the CTBN. In Theorem 3, it is shown how to directly construct the graph of this representative Bayesian network from a given CTBN. Finally, in Definition 5.3, this is tied together with the discrete components of an HTBN.

Definition 5.1. Consider a CTBN with graph $(G_0, G_\rightarrow)$ and time-indexed variables $\mathbf{V_B}$. An *associated causal graph* $G_\mathbf{B}$ is the graph with nodes $\mathbf{V_B}$ such that for all $X, Y \in \mathbf{V_B}$:

- If $X \rightarrow Y \in G_0$, then $X_0 \rightarrow Y_0 \in G_\mathbf{B}$
- If $X \rightarrow Y \in G_\rightarrow$ and $\{\beta, s(\beta)\} \subseteq \mathbf{B}$, then $X_\beta \rightarrow Y_{s(\beta)} \in G_\mathbf{B}$

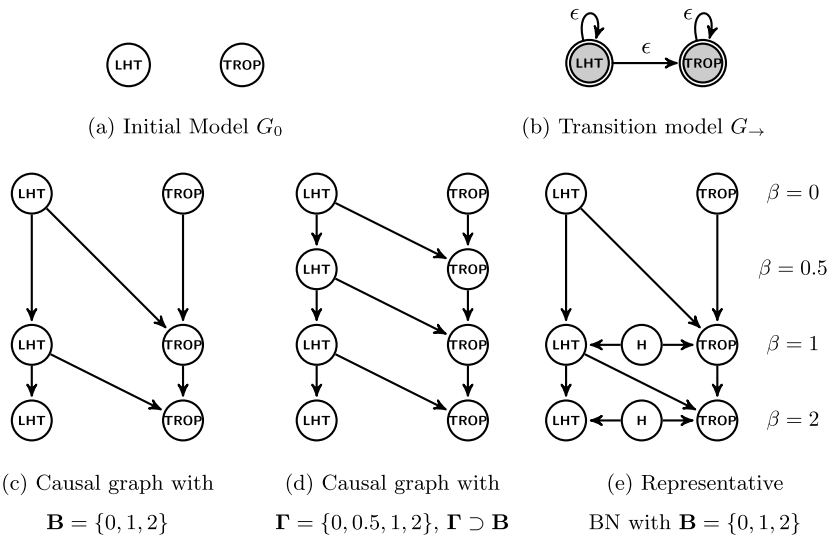
and $G_\mathbf{B}$ does not contain any other arcs.

Example 5.1. Consider a CTBN with the initial model and transition model as shown in Fig. 4a and 4b. The causal graphs of the continuous component with time points of interest **B** and $\mathbf{\Gamma}$ are illustrated in Fig. 4c and 4d.

In the following lemma, we give a precise characterization of the relationship between CTBNs and causal graphs in terms of independence structures. Note that proofs of the lemmas and theorems can be found in Appendix A.

Lemma 1. *Let* $P(\mathbf{V})$ *be the distribution defined in a CTBN with graph* $(G_0, G_\rightarrow)$ *and time-indexed variables* $\mathbf{V_B}$, *and let* $X_\beta, Y_{\beta'} \in \mathbf{V_B}$. *Then if for all* $\mathbf{V_\Gamma}$ *with* $\mathbf{\Gamma} \supseteq \mathbf{B}$, *in the* associated causal graph $G_\mathbf{\Gamma}$ *over* $\mathbf{V_\Gamma}$ *it holds that* $X_\beta \perp\!\!\!\perp_{G_\mathbf{\Gamma}} Y_{\beta'} \mid \mathbf{Z}$, *then* $X_\beta \perp\!\!\!\perp_{P(\mathbf{V})} Y_{\beta'} \mid \mathbf{Z}$, $\mathbf{Z} \subseteq \mathbf{V_B} \setminus \{X_\beta, Y_{\beta'}\}$.

Next, we define a Bayesian network that represents independences of the CTBN, using the causal graphs from Definition 5.1. Theorem 2 proves that this BN is indeed representative.



(a) Initial Model G_0 (b) Transition model $G_{\rightarrow}$

(c) Causal graph with $\mathbf{B} = \{0, 1, 2\}$ (d) Causal graph with $\mathbf{\Gamma} = \{0, 0.5, 1, 2\}$, $\mathbf{\Gamma} \supset \mathbf{B}$ (e) Representative BN with $\mathbf{B} = \{0, 1, 2\}$

Fig. 4. (a) CTBN initial model; (b) CTBN transition model; (c) causal graph for $\mathbf{B}$; (d) causal graph for $\mathbf{\Gamma}$; (e) the corresponding representative BN. See Examples 5.1 and 5.2 for details.

Definition 5.2. Consider a CTBN with graph $(G_0, G_{\rightarrow})$ and time-indexed variables $\mathbf{V_B}$. A *representative discrete-time Bayesian network* is a Bayesian network with graph $G_{\mathbf{B}}$ that has at least nodes $\mathbf{V_B}$ and given any $X_\beta, Y_{\beta'} \in \mathbf{V_B}$, if:

$$X_\beta \not\perp\!\!\!\perp_{G_{\mathbf{\Gamma}}} Y_{\beta'} \mid \mathbf{Z}$$

for all $\mathbf{Z} \subseteq \mathbf{V_B} \setminus \{X_\beta, Y_{\beta'}\}$ in some associated causal graph $G_{\mathbf{\Gamma}}$, $\mathbf{\Gamma} \supseteq \mathbf{B}$, then:

- $\beta' = s(\beta)$ implies $X_\beta \rightarrow Y_{\beta'} \in G_{\mathbf{B}}$
- $\beta = s(\beta')$ implies $X_\beta \leftarrow Y_{\beta'} \in G_{\mathbf{B}}$
- $\beta = \beta' \neq 0$ implies $H^{XY}_\beta \rightarrow X_\beta \in G_{\mathbf{B}}$, $H^{XY}_\beta \rightarrow Y_\beta \in G_{\mathbf{B}}$
- $\beta = \beta' = 0$ and $X_0 \rightarrow Y_0$ in all causal graphs implies $X_0 \rightarrow Y_0 \in G_{\mathbf{B}}$

and $G_{\mathbf{B}}$ does not contain any other nodes or arcs.

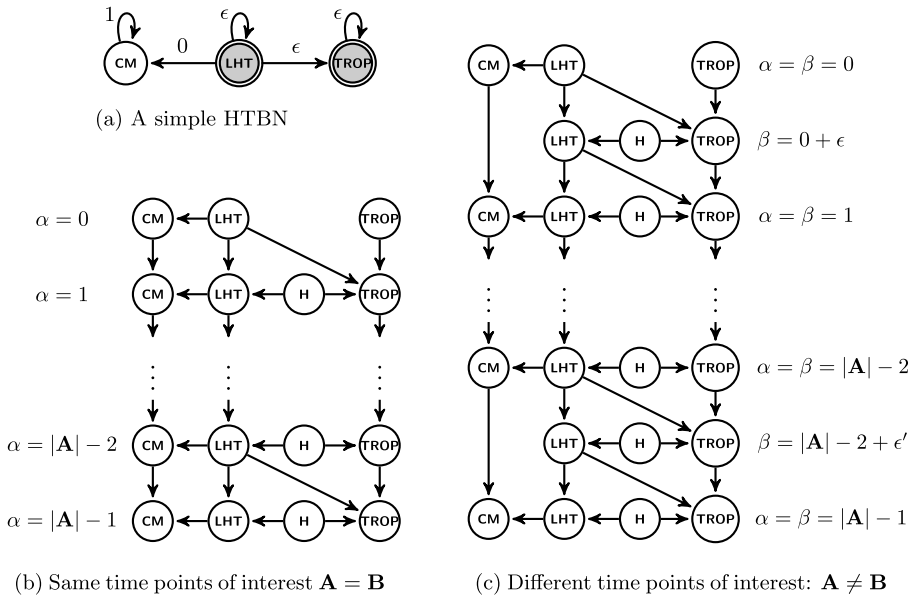
Example 5.2. Recall the CTBN example in Fig. 4. In causal graph $G_{\mathbf{B}}$, it holds that $\text{LHT}_1 \perp\!\!\!\perp \text{TROP}_1 \mid \{\text{LHT}_{\mathbf{B}}, \text{TROP}_{\mathbf{B}}\} \setminus \{\text{LHT}_1, \text{TROP}_1\}$, but this independence does not hold in causal graph $G_{\mathbf{\Gamma}}$, $\mathbf{\Gamma} \supset \mathbf{B}$, which implies that there exists a possible dependence in the CTBN distribution which cannot be represented by $G_{\mathbf{B}}$. Since every causal BN contains such independence assumptions that may not hold in the CTBN, which is revealed by increasing the set of time points, it implies there may not exist a causal BN that models all the CTBN's dependences. Therefore, additional variables are introduced in Definition 5.2 which create additional dependences, e.g. a hidden variable with arcs to LHT_1 and TROP_1 in the representative BN as shown in Fig. 4e.

Theorem 2. *Let $P(\mathbf{V})$ be the distribution defined in a CTBN with graph $(G_0, G_{\rightarrow})$ and time-indexed variables $\mathbf{V_B}$, and let $X_\beta, Y_{\beta'} \in \mathbf{V_B}$. Then if for the graph of the representative discrete-time Bayesian network $G_{\mathbf{B}}$ it holds that $X_\beta \perp\!\!\!\perp_{G_{\mathbf{B}}} Y_{\beta'} \mid \mathbf{Z}$, then $X_\beta \perp\!\!\!\perp_{P(\mathbf{V})} Y_{\beta'} \mid \mathbf{Z}$, with $\mathbf{Z} \subseteq \mathbf{V_B} \setminus \{X_\beta, Y_{\beta'}\}$.*

According to Definition 5.2, additional hidden variables $\mathbf{H}$ are introduced for a continuous-time variable C in the associated representative Bayesian network graph when C is d-connected with other continuous-time variables. The introduction of hidden variables makes sure that the states of C at any time point are correlated with the other continuous-time variables when their temporal evolutions are taken into consideration. This models entanglement, a concept that occurs in CTBNs and DBNs [5].

The graphs of representative Bayesian network are defined above by the existence of some causal graph of the CTBN. Clearly, it is not feasible to go over all causal graphs and their independences. Therefore, we provide in the following theorem a simple procedure to directly construct this Bayesian network graph from a given CTBN.

Theorem 3. *Consider a CTBN, with graph $(G_0, G_{\rightarrow})$ and time-indexed variables $\mathbf{V_B}$, and let $G_{\mathbf{B}}$ be the graph of the representative discrete-time Bayesian network. Then if $X \rightarrow Y \in G_0$, then $X_0 \rightarrow Y_0 \in G_{\mathbf{B}}$. Furthermore, for all $X, Y \in \mathbf{V}$, if X and Y are d-connected in $G_{\rightarrow}$ given $\varnothing$, and for all $\beta \in \mathbf{B} \setminus \{0\}$:*



(a) A simple HTBN

(b) Same time points of interest $\mathbf{A} = \mathbf{B}$

(c) Different time points of interest: $\mathbf{A} \neq \mathbf{B}$

Fig. 5. Discretization of an HTBN.

- $H_\beta^{XY} \to X_\beta, H_\beta^{XY} \to Y_\beta \in G_{\mathbf{B}}$
- *if there is a directed path from X to Y in* $G_{\to}$*, then* $X_\beta \to Y_{s(\beta)} \in G_{\mathbf{B}}$

and $G_{\mathbf{B}}$ *does not contain any other nodes or arcs.*

Once we have a discrete version of the continuous components, we are able to discretize a whole HTBN by assembling the components. We use $G_{K_{\mathbf{D}}}$ to denote the unrolled graph for a discrete-time component $K_{\mathbf{D}}$ (given time points **A**). We denote by $G_{K_{\mathbf{C}}}$ the representative Bayesian network graph of the continuous component $K_{\mathbf{C}}$ (given time points **B**). In the following definition, we tie these two discretizations together to discretize the full HTBN. This includes the Bayesian networks associated to each component, together with a number of edges between these components.

Definition 5.3. Let $\mathcal{H}$ be an HTBN with continuous components $\mathbf{K_C}$, discrete components $\mathbf{K_D}$, an initial Bayesian network $\mathcal{B} = (G_0, P)$ and a transition model with graph $G_{\mathcal{H}}$. Given sets of time points **B** for continuous-time variables and **A** for discrete-time variables, then a *representative Bayesian network structure* is a graph $G = (V(G), E(G))$ where $V(G) = \bigcup_{K \in \mathbf{K_D} \cup \mathbf{K_C}} V(G_K)$, $E(G)$ includes $\bigcup_{K \in \mathbf{K_D} \cup \mathbf{K_C}} E(G_K)$, and for any continuous-time variable C and discrete-time variable D with $D \in \mathbf{D}, C \in \mathbf{C}$:

- $C \xrightarrow{0} D \in G_{\mathcal{H}}$ implies $C_\alpha \to D_\alpha \in G$, for all $\alpha \in \mathbf{A}$;
- $D \xrightarrow{\epsilon} C \in G_{\mathcal{H}}$ implies $D_a \to C_\beta \in G$, $a = \max\{\alpha \mid \alpha < \beta\}$ for all $\beta \in \mathbf{B}$;
- $D_0 \xrightarrow{0} C_0 \in G_0$ implies $D_0 \to C_0 \in G$;
- $C_0 \xrightarrow{0} D_0 \in G_0$ implies $C_0 \to D_0 \in G$;

and G does not contain any other nodes or arcs.

Example 5.3. The graph of the representative Bayesian network for an HTBN is illustrated in Fig. 5. Two cases are shown, the graph for $\mathbf{A} = \mathbf{B}$ and $\mathbf{A} \neq \mathbf{B}$.

5.2. Constructing representative BNs

After obtaining the structure of the discrete version of an HTBN, the second step in the procedure of discretization is to estimate the parameters for the representative BN given an HTBN and sets of time points of interest **A**, **B**. The full procedure, using the results above, is given in Algorithm 1 and explained below.

The algorithm is based on an Expectation Maximization (EM) procedure [6]. This well-known procedure is traditionally used to learn distributions from data with missing values. In our case, we exploit EM for learning parameters of the representative Bayesian network, as this network includes a number of hidden variables for the continuous components. In order

to fit the original joint distribution of a continuous component to a distribution that includes hidden variables, we look upon the hidden variables as if they are missing values. In an ordinary EM procedure parameters of the model are maximized based on expected sufficient statistics, which are determined based on the distribution of observed data and expectations computed at each iteration. In this case, we replace the distribution of observed data by the known distribution given by an HTBN. This is analogous to learning from data since the HTBN parameters are equivalent to the expected sufficient statistics of data generated from the HTBN. As a consequence, this parameter fitting procedure has the same properties as any other EM algorithm, in particular, the procedure results in a local minimum of the Kullback–Leibler divergence between the distributions associated to the HTBN and the representative BN (see e.g. Section 11.4.7 of [7]).

In the following, the parameters in the BN for continuous-time variables and hidden variables are denoted as $\boldsymbol{\theta}$. We use $\boldsymbol{\theta}^H_{s(\beta)}$ for the prior probability of $H_{s(\beta)}$. Furthermore, we use $\boldsymbol{\theta}^C_{s(\beta)}$ to represent the conditional probability of $C_{s(\beta)}$ conditioned on its parents in the representative Bayesian network. Finally, to distinguish between the distributions of the hybrid model and representative BN, we use $P_{\mathcal{H}}$ for the distribution of the HTBN and P for the distribution in the BN.

Algorithm 1: Construct representative BN.

Input: A HTBN $\mathcal{H}$ and sets of time points $\mathbf{A}$, $\mathbf{B}$
Result: Representative BN for $\mathcal{H}$
1 Construct the representative BN structure according to Definition 5.3
2 Inherit parameters for discrete-time variables $\mathbf{D}$ from $\mathcal{H}$
3 Initialize parameters $\boldsymbol{\theta}$ for continuous-time and hidden variables in BN
4 **while** $\boldsymbol{\theta}$ *is not converged* **do**
5 **foreach** $s(\beta) \in \mathbf{B}$ **do**
6 Compute expectation for $\mathbf{H}_{s(\beta)}$ in BN ▷ E-Step
7 $P(\mathbf{H}_{s(\beta)} \mid \mathbf{C}_\beta, \mathbf{C}_{s(\beta)}, \mathbf{D}_a)$
8 where $a = \max\{\alpha \mid \alpha < s(\beta), \alpha \in \mathbf{A}\}$
9 Let $\Delta = \mathbf{H}_{s(\beta)} \cup \mathbf{C}_\beta \cup \mathbf{C}_{s(\beta)} \cup \mathbf{D}_a$
10 **foreach** $C_{s(\beta)} \in \mathbf{C}_{s(\beta)}$ **do**
11 Update parameters for $C_{s(\beta)}$ in BN ▷ M-Step
12 $\boldsymbol{\theta}^C_{s(\beta)} \propto \sum_{\Delta \setminus \{\pi(C_{s(\beta)}) \cup \{C_{s(\beta)}\}\}} P(\mathbf{H}_{s(\beta)}, \mathbf{C}_\beta, \mathbf{C}_{s(\beta)}, \mathbf{D}_a)$
13 where $P(\mathbf{H}_{s(\beta)}, \mathbf{C}_\beta, \mathbf{C}_{s(\beta)}, \mathbf{D}_a) = P(\mathbf{H}_{s(\beta)} \mid \mathbf{C}_\beta, \mathbf{C}_{s(\beta)}, \mathbf{D}_a) P_{\mathcal{H}}(\mathbf{C}_\beta, \mathbf{C}_{s(\beta)} \mid \mathbf{D}_a) P_{\mathcal{H}}(\mathbf{D}_a)$
14 **end**
15 **foreach** $H_{s(\beta)} \in \mathbf{H}_{s(\beta)}$ **do**
16 Update prior probability for $H_{s(\beta)}$ in BN ▷ M-Step
17 $\boldsymbol{\theta}^H_{s(\beta)} = \sum_{\Delta \setminus H_{s(\beta)}} P(\mathbf{H}_{s(\beta)}, \mathbf{C}_\beta, \mathbf{C}_{s(\beta)}, \mathbf{D}_a)$
18 **end**
19 **end**
20 **end**
21 **return** BN

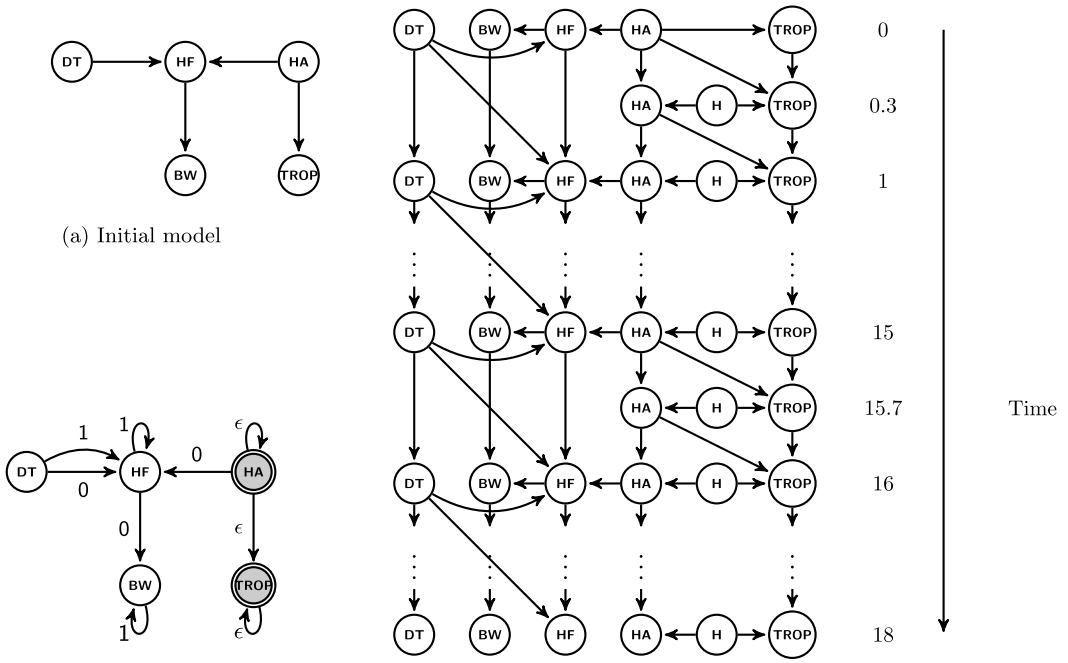
Algorithm 1 outlines the procedure of constructing a representative BN given an HTBN and time points. First, it constructs the structure of the representative BN according to Definition 5.3. In the remainder, it estimates parameters for continuous-time variables given their parents, including the hidden variables. The estimation of parameters for discrete-time variables is straightforward as they can be directly copied from the parameters defined in the HTBN $\mathcal{H}$.

The parameters $\boldsymbol{\theta}^C_{s(\beta)}$ and $\boldsymbol{\theta}^H_{s(\beta)}$ start with an arbitrary initialization in the BN. In the E-step, we compute the expectation for hidden variables $\mathbf{H}_{s(\beta)}$ given the Markov blanket of $\mathbf{H}_{s(\beta)}$, each of which is represented as a conditional probability as shown on line 7. In the M-step, i.e., from line 11 onwards, the parameters $\boldsymbol{\theta}^C_{s(\beta)}$ and $\boldsymbol{\theta}^H_{s(\beta)}$ are updated by marginalizing out variables from Δ by making use of the expectations computed in the E-step and information derived from the HTBN distribution $P_{\mathcal{H}}$. In particular, the distributions $P_{\mathcal{H}}(\mathbf{D}_a)$ and $P_{\mathcal{H}}(\mathbf{C}_\beta, \mathbf{C}_{s(\beta)} \mid \mathbf{D}_a)$ can be computed by marginalization from the joint distribution of the HTBN as given in Definition 4.4.

6. Experiments

The power of HTBNs is illustrated in the domain of myocardial contractility in relationship to heart attack, heart failure and its medical treatment, introduced in Section 2 and summarized in Fig. 3. Of particular interest is the question how the dynamics of the occurrence of heart failure are affected by heart attacks and the administration of digitalis. As discussed in Section 2, a single DBN and CTBN cannot provide a satisfactory representation of the evolution of variables with different rates: the administering of digitalis or measurement of body weight is regular, in contrast to the more sparse and irregular occurrence of heart attacks.

We now show a simplified HTBN of the heart failure problem given time points of interest $\mathbf{A}$, $\mathbf{B}$. As the data were unavailable for heart tissue and contractility, the model was simplified by leaving out variables CM and LHT. The graphical representation of the resulting hybrid-time model is shown in Fig. 6a and Fig. 6b. We parameterized the model partly using medical expert knowledge and partly from data. The unit of time for the transitions is weeks. There are three weekly changing variables in the model, i.e., digitalis (DT), heart failure (HT) and body weight (BW). The parameters of the continuous-time variables (heart attack (HA) and troponin (TROP)) were derived from real data. The transition rates for



(a) Initial model

(b) Transition model

(c) A representative Bayesian network for a simplified hybrid model of heart failure

Fig. 6. Simplified hybrid model of heart failure and its associated representative Bayesian network graph given time points of interest.

$P(\mathrm{DT}_0 = \mathrm{T})$	0.01
$P(\mathrm{HA}_0 = \mathrm{T})$	0.01
$P(\mathrm{TROP}_0 = \mathrm{N} \mid \mathrm{HA}_0 = \mathrm{T})$	0.15
$P(\mathrm{TROP}_0 = \mathrm{N} \mid \mathrm{HA}_0 = \mathrm{F})$	0.98
$P(\mathrm{HF}_0 = \mathrm{T} \mid \mathrm{DT}_0 = \cdot, \mathrm{HA}_0 = \cdot)$	0.01

DT_t	DT_{t+1}	HF_t	HA_{t+1}	HF_{t+1}
F	F	F	F	0.08
T	F	F	F	0.07
F	T	F	F	0.05
T	T	F	F	0.04
F	F	T	F	0.45
T	F	T	F	0.40
F	T	T	F	0.40
T	T	T	F	0.30
F	F	F	T	0.40
T	F	F	T	0.35
F	T	F	T	0.30
T	T	F	T	0.25
F	F	T	T	0.99
T	F	T	T	0.70
F	T	T	T	0.65
T	T	T	T	0.70

$$Q_{\mathrm{HA}} = \begin{array}{c} \\ F \\ T \end{array} \begin{array}{c} \begin{array}{cc} F & T \end{array} \\ \begin{pmatrix} -0.002 & 0.002 \\ 0.4 & -0.4 \end{pmatrix} \end{array}$$

$$Q_{\mathrm{TROP|HA=T}} = \begin{array}{c} \\ N \\ Abn \end{array} \begin{array}{c} \begin{array}{cc} N & Abn \end{array} \\ \begin{pmatrix} -0.05 & 0.05 \\ 0.02 & -0.02 \end{pmatrix} \end{array}$$

$$Q_{\mathrm{TROP|HA=F}} = \begin{array}{c} \\ N \\ Abn \end{array} \begin{array}{c} \begin{array}{cc} N & Abn \end{array} \\ \begin{pmatrix} -0.01 & 0.01 \\ 0.015 & -0.015 \end{pmatrix} \end{array}$$

Fig. 7. Parameters in the hybrid model. Prior probabilities (top) and transition parameters, i.e. $P(\mathrm{HF}_{t+1} = \mathrm{T} \mid \mathrm{DT}_t, \mathrm{DT}_{t+1}, \mathrm{HF}_t, \mathrm{HA}_{t+1})$ for all parent configurations (bottom left) and the intensity matrices for HA and for TROP given HA (bottom right).

troponin were parameterized using the MIMIC II Clinical Database [8,9] which contains thousands of Intensive Care Unit (ICU) patient records gathered from 2001 to 2008. Similarly, some parameters for troponin in the initial model were obtained from the same database. However, due to incompleteness of the clinical information from the MIMIC II Database, we derived the distribution of troponin in the absence of heart attack from the literature [10].

Now we show the procedure to determine the parameters of a representative BN given the HTBN and time points of interest. According to Definition 5.3, a representative BN was generated as shown in Fig. 6, given the sets of time points **A** and **B**, $\mathbf{A} \subset \mathbf{B}$, where **B** includes two additional irregular time points for the heart attack events. Taking the parameters

Table 1
Parameters in the representative BN learned by EM. Parameters for DT, HF are the same as Fig. 7. Prior probabilities for hidden variables H in (a), transition parameters for HA, TROP for all parent configuration, i.e. $P(\mathrm{HA}_{s(\beta)} = \mathrm{T} \mid \mathrm{HA}_\beta, \mathrm{H}_{s(\beta)})$ in (b), (c), $P(\mathrm{TROP}_{s(\beta)} = \mathrm{Abn} \mid \mathrm{HA}_\beta, \mathrm{TROP}_\beta, \mathrm{H}_{s(\beta)})$ in (d), (e), $s(\beta) \in \{0.3, 1\}$.

(a)

$\mathrm{H}_{0.3}$	0.51
H_1	0.51

(b)

HA_0	$\mathrm{H}_{0.3}$	$\mathrm{HA}_{0.3}$
F	F	$9 * 10^{-4}$
T	F	0.84
F	T	$2 * 10^{-4}$
T	T	0.94

(c)

$\mathrm{HA}_{0.3}$	H_1	HA_1
F	F	$2 * 10^{-3}$
T	F	0.70
F	T	$5 * 10^{-4}$
T	T	0.81

(d)

HA_0	TROP_0	$\mathrm{H}_{0.3}$	$\mathrm{TROP}_{0.3}$
F	Abn	F	0.99
T	Abn	F	0.99
F	N	F	$5 * 10^{-3}$
T	N	F	0.02
F	Abn	T	0.99
T	Abn	T	0.99
F	N	T	$1 * 10^{-3}$
T	N	T	$8 * 10^{-3}$

(e)

$\mathrm{HA}_{0.3}$	$\mathrm{TROP}_{0.3}$	H_1	TROP_1
F	Abn	F	0.98
T	Abn	F	0.98
F	N	F	0.01
T	N	F	0.04
F	Abn	T	0.99
T	Abn	T	0.99
F	N	T	$3 * 10^{-3}$
T	N	T	0.02

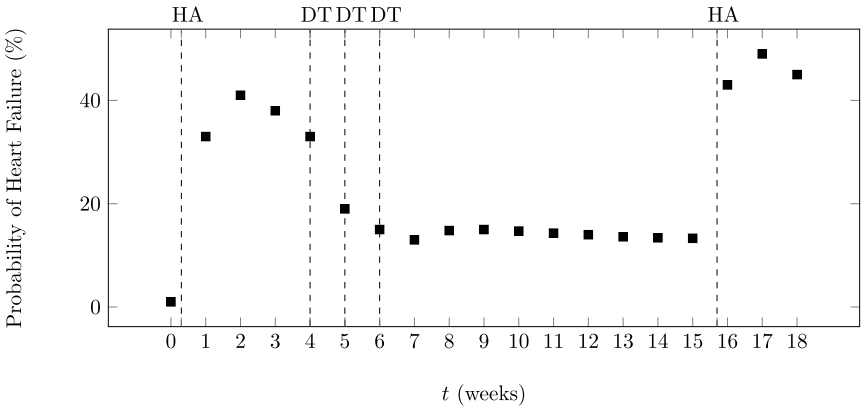


Fig. 8. Effects of heart attack and digitalis on heart failure. 'DT' indicates that digitalis was administered at that moment in time. 'HA' indicates that a heart attack was observed. Note that observations for HA are continuous-time, so observed at an arbitrary point in time; digitalis is observed once a week at most.

in the hybrid model shown in Fig. 7 and the constructed BN as input for the EM procedure described in Algorithm 1, we computed the parameters of the representative BN. The EM algorithm is implemented in R using the CTBN-RLE reasoning engine [11] and its R interface [12]. Some of the learned parameters are shown in Table 1. The parameters for HA at time 0.3 and 1 are different because the time points are unevenly spaced. In the following, we illustrate the equivalence of the HTBN with the representative BN by computing one of the marginal probabilities, i.e. $P(\mathrm{HA}_{0.3})$.

In the BN, we can compute the marginal of $\mathrm{HA}_{0.3}$ by:

$$\begin{aligned} P(\mathrm{HA}_{0.3} = T) &= \sum_{\mathrm{HA}_0, \mathrm{H}_{0.3}} P(\mathrm{HA}_0, \mathrm{H}_{0.3}, \mathrm{HA}_{0.3} = T) \\ &= \sum_{\mathrm{HA}_0, \mathrm{H}_{0.3}} P(\mathrm{HA}_{0.3} = T \mid \mathrm{HA}_0, \mathrm{H}_{0.3}) P(\mathrm{HA}_0) P(\mathrm{H}_{0.3}) \\ &= 9 * 10^{-4} * 0.99 * 0.49 + 0.84 * 0.01 * 0.49 \\ &\quad + 2 * 10^{-4} * 0.99 * 0.51 + 0.94 * 0.01 * 0.51 = 9 * 10^{-3} \end{aligned}$$

As HA has no parent in the HTBN model, we can directly compute its distribution at time 0.3 using the intensity matrix and initial distribution of HA described in Fig. 7:

$$P(\mathrm{HA}_{0.3}) = P(\mathrm{HA}_0) \exp(Q_{\mathrm{HA}} * 0.3) = 9 * 10^{-3}$$

Finally, to show the behavior of the model we computed the probability distribution of heart failure for a period of 19 weeks given the observed (regular or irregular) evidence. Results of this experiment are plotted in Fig. 8. The plot shows the

negative effects of a heart attack (see the jumps at time $t = 1$, $t = 2$ and $t = 16$) and the positive effect of digitalis on heart failure (see the rapid fall at time $t = 5$). The model also implies that the condition of the heart stabilizes after administering the drug through an increase in the contractility. However, a damaged heart does not fully recover, not even with the help of digitalis.

7. Discussion

We have described hybrid time Bayesian networks as a means to model dynamic systems. In particular, HTBNs are suitable when a combination of regular and irregularly changing variables best describe the domain or when multiple time granularities need to be modeled. The proposed approach generalizes both continuous-time and discrete-time Bayesian networks. HTBNs allow reasoning over irregularly spaced evidence, a property inherited from CTBNs, as well as regular time sliced evidence, derived from DBNs. This is a significant contribution in terms of ease of modeling complex dynamic systems. Each variable can be modeled using the appropriate mechanism, which is particularly helpful when constructing models from expert knowledge. Furthermore, we established a mapping of hybrid-time networks into a standard BN given a set of time points of interest. The inference problem in HTBNs is therefore reduced to a problem for which efficient solutions exist.

DBNs were proposed by Dean and Kanazawa as early as 1989 [13] and have been extensively studied and applied since then, notably by Murphy [4]. CTBNs [5] are of a more recent date, but have seen various applications over the last decade. Applications range from detecting attacks in computer networks [14], analyzing social network dynamics [15] and modeling heart failure [16]. Note that in this article we do not aim to improve upon the medical analysis provided in the work by Gatti et al. [16], but instead studied the heart failure domain to show an application of HTBNs in a relevant use case.

There is also some relation between non-stationary dynamic Bayesian networks (nsDBN) and the formalism proposed here. In nsDBNs the structure and parameters change at particular time points [17,18]. The approaches are related in the sense that non-stationary Bayesian networks allow for different time granularities of the (complete) temporal process. The key difference here is that we consider the case where different random variables evolve at different rates.

A limitation of HTBNs is that so far the granularities of discrete-time variables are assumed to be fixed, as the focus of this paper has been on the combination of continuous and discrete-time models. A future extension of the work presented here would be combining different discrete-time granularities within the hybrid-time framework. This is related to the work on irregular-time Bayesian networks (ITBNs) [19] and to the work by van der Heijden and Lucas [20].

There is also some possible future work on inference. Currently, inference in HTBNs is limited by the exponential increase of dependencies when discretizing large continuous components, as currently all variables are connected. It appears possible to optimize the BN construction to only take into account dependencies that have significant influence on the temporal process. A more efficient, approximate, structure would alleviate the complexity problem. Note that CTBNs also suffer from this problem, preventing exact inference in large networks.

Since the BN construction presented here may become infeasible for large networks with many time points of interest, another direction for further study would be inference methods that operate directly on the HTBN instead of constructing a BN first. Possibilities include a sampling approach to inference as well as expectation propagation, which has been proposed for inference in CTBNs [21].

Finally, it remains of interest to study learning HTBNs from data. We expect that a general approach to parameter learning can be constructed based on existing work on parameter learning in BNs and CTBNs. Yet, it will require some extensions to take the interactions between discrete and continuous components into account. Structure learning from data is for now an open problem. Nevertheless, hand-crafted HTBN models can be applied already using the inference approach developed in this paper.

Appendix A. Proofs

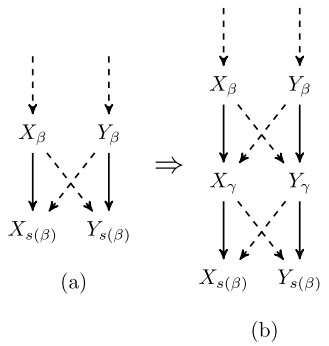
Proof of Lemma 1. The data-generating process of CTBNs generates a sequence of events, i.e., observations of random variables indexed by time. Since there is no feedback, i.e., it is a causal process, the joint distribution of these sequences can be described by a causal DAG as given by Definition 5.1 for a set of time points $\mathbf{\Gamma}$ that occur in these sequences. As a result, each dependency $X_\beta \not\perp\!\!\!\perp_{P(\mathbf{V})} Y_{\beta'} \mid \mathbf{Z}$ will be represented by some DAG that includes the variables $\mathbf{V_B}$. □

Proof of Theorem 2. Take some dependence $X_\beta \not\perp\!\!\!\perp_{P(\mathbf{V})} Y_{\beta'} \mid \mathbf{Z}$, $\mathbf{Z} \subseteq \mathbf{V_B} \setminus \{X_\beta, Y_{\beta'}\}$. Because of Lemma 1, we know that this dependence is included in some causal graph associated to the CTBN. Let $\mathbf{\Gamma}_i \supset \mathbf{B}$ be a set of time points such that the associated causal graph $G_{\mathbf{\Gamma}_i}$ represents the ith dependence statement over the set of variables $\mathbf{V_B}$ in the CTBN. Observe that for any $\mathbf{\Gamma}' \supset \mathbf{\Gamma} \supseteq \mathbf{B}$, if $X_\beta \not\perp\!\!\!\perp_{G_{\mathbf{\Gamma}}} Y_{\beta'} \mid \mathbf{Z}$ then $X_\beta \not\perp\!\!\!\perp_{G_{\mathbf{\Gamma}'}} Y_{\beta'} \mid \mathbf{Z}$. Thus, all the dependences of the CTBN over $V_{\mathbf{B}}$ will be represented by a causal graph $G_{\mathbf{\Gamma}}$ such that $\mathbf{\Gamma} = \bigcup_i \mathbf{\Gamma}_i$.

Now consider this causal graph $G_{\mathbf{\Gamma}}$. It is possible to marginalize out all variables $\mathbf{\Gamma} \setminus \mathbf{B}$ using the procedure described in [22, Definition 4.2.1], which coincides with Definition 5.2 instantiated for the structure of these causal graphs, except that we represent bidirectional arcs in ancestral graphs by means of additional hidden variables. Since $G_{\mathbf{\Gamma}}$ represents the same independence structure as $G_{\mathbf{B}}$ [22, Theorem 4.18], we obtain $X_\beta \not\perp\!\!\!\perp_{G_{\mathbf{B}}} Y_{\beta'} \mid \mathbf{Z}$. □

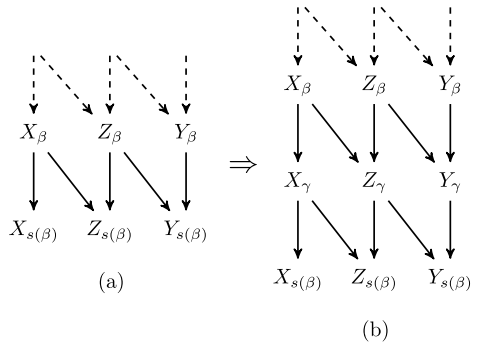
Proof of Theorem 3. The nodes and arcs for the initial time slices are obvious because these are exactly included in the causal graphs. Now take some $\beta \in \mathbf{B} \setminus \{0\}$. We will consider the possible relationships between X and Y in the CTBN.

(i) Suppose X and Y are d-connected. Then we need to show that there is some causal graph such that X_β is dependent of Y_β for any $\mathbf{Z}$. Consider the case that X and Y are d-connected through some path in the CTBN and consider the causal graph associated to this CTBN in the following two figures:



In (a), X_β and Y_β are conditionally independent, e.g., $X_{s(\beta)} \perp\!\!\!\perp Y_{s(\beta)} \mid X_\beta, Y_\beta$. Now observe that by adding time slices between β and $s(\beta)$, these independences disappear, see (b). Similarly, if the shortest connecting path between X and Y is of length n, any independence between $X_{s(\beta)}$ and $Y_{s(\beta)}$ in a set of variables $\mathbf{V_B}$ will not hold in a causal graph with n additional intermediate time slices between $s(\beta)$ and its predecessor in $\mathbf{B}$. Hence, there will always be a causal graph where there are no independences between $X_{s(\beta)}$ and $Y_{s(\beta)}$ for some conditioning set in $\mathbf{V_B}$.

(ii) Suppose that there is a directed path between X and Y. Then we again need to show that there is some causal graph such that X_β is dependent of $Y_{s(\beta)}$ for any $\mathbf{Z}$. If X and Y are directly connected, then there will clearly be a dependency of X_β and $Y_{s(\beta)}$, for example a direct dependency in $\mathbf{V_B}$. Now suppose that X and Y are not directly connected. Consider the following figures.

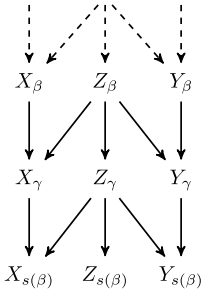


In the causal graph (a) there will be conditional independences between X_β and $Y_{s(\beta)}$, i.e., if we condition on parents of $Y_{s(\beta)}$. Again, by adding intermediate time slices between β and $s(\beta)$, dependences between X_β and $Y_{s(\beta)}$ appear that cannot be blocked by variables in (a), in particular the path $X_\beta \to Z_\gamma \to Y_{s(\beta)}$. Similar to the previous case, directed paths of length n lead to dependences if we add $n-1$ time-slices between β and $s(\beta)$ that cannot be blocked by $\mathbf{V_B}$.

(iii) Suppose X and Y are d-connected, yet, not through a fully directed path. Then we show that in all causal graphs there is a conditional independence between X_β and $Y_{s(\beta)}$. Let $\mathbf{W}$ be the set of variables which are also d-connected to X and Y, i.e., this set includes all variables on paths between X and Y. Then, the claim is that for all causal graph, it holds that:

$$X_\beta \perp\!\!\!\perp Y_{s(\beta)} \mid (\mathbf{W} \setminus \{X, Y\})_{\mathbf{B}}, Y_\beta$$

Consider some causal graph, for which we know that on all d-connected paths between X and Y there is some divergent node Z (otherwise this path would be fully directed, so case (ii) applies). Then consider the following graph to illustrate why this is true:



Consider that $\beta, s(\beta) \in \mathbf{B}$, but $\gamma \notin \mathbf{B}$. In this case $\mathbf{W} = \{Z\}$, so the conditioning set contains at least $\{Z_\beta, Z_{s(\beta)}, Y_\beta\}$. To d-connect X_β with $Y_{s(\beta)}$ on this path of the CTBN, there should therefore be a path through some $Z_{\gamma'}$, where $\gamma' \notin \mathbf{B}$. Suppose $\gamma' < \beta$, then all paths will be blocked at β, since all the variables at β are in the conditioning set, except for X_β. Now suppose $\beta < \gamma' < s(\beta)$. Then we have the case similar to the figure where $\gamma = \gamma'$. It is clear in this case that such a path would be blocked by X_β because of the v-structure. Finally suppose $\gamma' > s(\beta)$. Then the only way that this opens a path from Y to X is if there is a v-structure on Z, i.e., $X \to Z \leftarrow Y$. But then there is a directed path $X \to Z \to Y$ in the CTBN, so this contradicts the assumption.

(iv) Suppose X and Y are not d-connected. Again, we need show that in all causal graphs there will be at least a conditional independence between X_β and $Y_{s(\beta)}$. This is trivial, because in all causal graph $X_\beta \perp\!\!\!\perp Y_{s(\beta)} \mid \varnothing$ because either X and Y are not connected or on all paths between X and Y, there is some v-structure, which then also occurs in the causal graph. The case for X_β and Y_β is analogous.

(v) For any variables X and Y it holds that $X_\beta \perp\!\!\!\perp Y_{s(s(\beta))} \mid \mathbf{V}_{s(\beta)}$. That is, the Markov property holds in CTBNs.

Finally, note that (i)–(v) together imply the claim of the theorem. □

References

[1] M. Liu, A. Hommersom, M. van der Heijden, P. Lucas, Hybrid time Bayesian networks, in: S. Destercke, T. Denoeux (Eds.), ECSQARU '15: Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, in: Lect. Notes Artif. Intell., vol. 9161, 2015, pp. 376–386.
[2] C. Bettini, S. Jajodia, S. Wang, Time Granularities in Databases, Data Mining, and Temporal Reasoning, Springer, 2000.
[3] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, Palo Alto, 1988.
[4] K.P. Murphy, Dynamic Bayesian networks: representation, inference and learning, Ph.D. thesis, University of California, Berkeley, 2002.
[5] U. Nodelman, C.R. Shelton, D. Koller, Continuous time Bayesian networks, in: UAI '02: Proceedings of the Eighteenth conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc., 2002, pp. 378–387.
[6] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. B (1977) 1–38.
[7] K.P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.
[8] M. Saeed, M. Villarroel, A.T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T.H. Kyaw, B. Moody, R.G. Mark, Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database, Crit. Care Med. 39 (5) (2011) 952.
[9] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, H. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[10] J.A. de Lemos, M.H. Drazner, T. Omland, C.R. Ayers, A. Khera, A. Rohatgi, I. Hashim, J.D. Berry, S.R. Das, D.A. Morrow, et al., Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population, JAMA 304 (22) (2010) 2503–2512.
[11] C. Shelton, Continuous time Bayesian network reasoning and learning engine, http://rlair.cs.ucr.edu/ctbnrle/, 2013.
[12] S. Villa, CTBN R interface, http://rlair.cs.ucr.edu/ctbnrle/Rinterface, 2014.
[13] T. Dean, K. Kanazawa, A model for reasoning about persistence and causation, Comput. Intell. 5 (3) (1989) 142–150.
[14] J. Xu, C.R. Shelton, Continuous time Bayesian networks for host level network intrusion detection, in: Machine Learning and Knowledge Discovery in Databases (ECML/PKDD), in: Lect. Notes Artif. Intell., vol. 5212, 2008, pp. 613–627.
[15] Y. Fan, C.R. Shelton, Learning continuous-time social network dynamics, in: UAI '09: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, 2009, pp. 161–168.
[16] E. Gatti, D. Luciani, F. Stella, A continuous time Bayesian network model for cardiogenic heart failure, Flexible Serv. Manuf. J. 24 (4) (2012) 496–515, http://dx.doi.org/10.1007/s10696-011-9131-2.
[17] M. Grzegorczyk, D. Husmeier, Non-stationary continuous dynamic Bayesian networks, in: Advances in Neural Information Processing Systems, 2009, pp. 682–690.
[18] J.W. Robinson, A.J. Hartemink, Non-stationary dynamic Bayesian networks, in: Advances in Neural Information Processing Systems, 2009, pp. 1369–1376.
[19] M. Ramati, Y. Shahar, Irregular-time Bayesian networks, in: UAI '10: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, 2010, pp. 484–491.
[20] M. van der Heijden, P.J. Lucas, Probabilistic reasoning with temporal indeterminacy, in: PGM '12: Proceedings of the 6th European Workshop on Probabilistic Graphical Models, 2012.
[21] U. Nodelman, D. Koller, C. Shelton, Expectation propagation for continuous time Bayesian networks, in: UAI '05: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, 2005, pp. 431–440.
[22] T. Richardson, P. Spirtes, Ancestral graph Markov models, Ann. Stat. (2002) 962–1030.