

Independence Decomposition in Dynamic Bayesian Networks

Ildikó Flesch and Peter Lucas

Department of Information and Knowledge Systems
Institute for Computing and Information Science
Radboud University Nijmegen, the Netherlands
Email: {ildiko,peter}@cs.ru.nl

Abstract. Dynamic Bayesian networks are a special type of Bayesian network that explicitly incorporate the dimension of time. They can be distinguished into repetitive and non-repetitive networks. Repetitiveness implies that the set of random variables of the network and their independence relations are the same at each time step. Due to their structural symmetry, repetitive networks are easier to use and are, therefore, often taken as the standard. However, repetitiveness is a very strong assumption, which normally does not hold, as particular dependences and independences may only hold at certain time steps.

In this paper, we propose a new framework for independence modularisation in dynamic Bayesian networks. Our theory provides a method for separating atemporal and temporal independence relations, and offers a practical approach to building dynamic Bayesian networks that are possibly non-repetitive. A composition operator for temporal and atemporal independence relations is proposed and its properties are studied. Experimental results obtained by learning dynamic Bayesian networks from real data show that this framework offers a more accurate way for knowledge representation in dynamic Bayesian networks.

1 Introduction

Probabilistic graphical models are increasingly adopted as tools for the modelling of domains involving uncertainty. For the development of practical applications especially Bayesian networks have gained much popularity. When considering these application domains, it appears that so far only limited attention has been given to the modelling of uncertain time-related phenomena, which occur in many of these domains. Bayesian networks in which some notion of time is explicitly dealt with are usually called *dynamic* Bayesian networks (DBNs) [3]. In some domains involving time, such as speech recognition, the use of special DBNs has been extensively explored (e.g. [1]), and technical issues such as concerning reasoning (e.g. [7]) and learning (e.g. [2]) in DBNs have been investigated.

DBNs are distinguished into two main classes: repetitive and non-repetitive networks. Repetitive networks have the same set of random variables and independence relations at each time step, whereas in non-repetitive networks the set

of random variables and also the independence relations between these random variables may vary in time. The simpler structure of repetitive networks provides significant advantages in terms of ease of modelling and computational complexity. Therefore, they are often seen as the standard DBN model (see [6] for an overview). However, repetitiveness is a very strong assumption that normally will not hold.

Recently, scientific evidence has become available that non-repetitive DBNs may also be practically useful [8]. We think that separating temporal and atemporal information in DBNs can be valuable, as it: (i) helps experts gain more insight into the relations in the networks, (ii) provides an opportunity for learning procedures to obtain more accurate models, and (iii) may help overcome computational limitations. However, so far no research has been carried out to characterise temporal and atemporal independence relations.

In this paper, a new framework for independence modularisation in DBNs is proposed, based on a theoretically grounded separation of temporal and atemporal independences. This distinction allows us to investigate isolated parts of the independence relations. Having given these individual parts of the network, we can construct both repetitive and non-repetitive DBNs. In this paper, we analyse the necessary properties to correctly join independence relations from these individual parts and define a join operator to carry out the composition of these independence relations. Finally, we provide experimental evidence of the usefulness of non-repetitive DBNs.

2 Motivating Example: the Disease Course of VAP

A real-world non-repetitive DBN of the disease course of a form of pneumonia is used as motivating example in this paper. As we will see, at each time step we have different independence relations offering an accurate model of the evolution of the disease.

We briefly describe the clinical features of pneumonia and then discuss the construction of a DBN for this disease. Pneumonia develops frequently in ICU patients, as these patients are critically ill and often they need respiratory support by a mechanical ventilator. After admission to a hospital, all patients become colonised by bacteria. In particular, mechanically ventilated patients run the risk of subsequently developing pneumonia caused by these bacteria; this type of pneumonia is known as *ventilator-associated pneumonia*, or VAP for short. Typical signs and symptoms of VAP include: high body temperature, decreased lung function (measured by the $\text{PaO}_2/\text{FiO}_2$ ratio) and evidence of pneumonia on the chest X-ray. By carrying out a dependency analysis on a retrospective, temporal dataset, with data of ICU patients collected during a period of three years, we were able to study how independence information changed in the course of time. Taking the duration of mechanical ventilation as the parameter defining the time steps, we have focused on modelling the course of the development of VAP at 3, 4 and 5 days after admission. The resulting DBN is shown in Fig. 1.

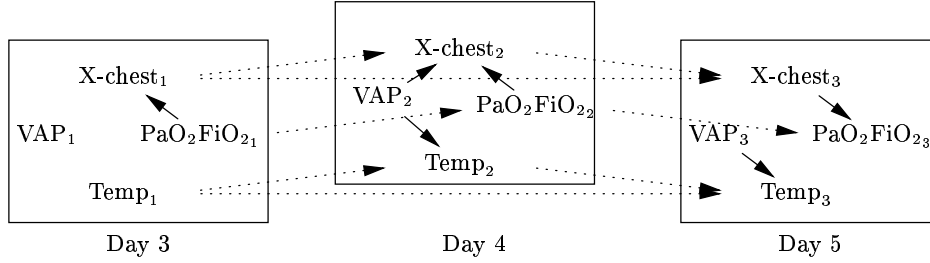


Fig. 1. The non-repetitive DBN for VAP; temporal arcs are depicted by dotted arrows.

This small network already shows the need for employing non-repetitiveness in the structure of a DBN, and, therefore, has motivated us to develop a theory that distinguishes between temporal and atemporal independence information and allows building a non-repetitive DBN in a correct and seamless fashion.

3 Basic Notions

We will be concerned in this paper with acyclic directed graphs (ADGs), denoted as a pair $G = (V, A)$, where V is a set of vertices and $A \subseteq V \times V$ is a set of *arcs*. A *directed path* is a sequence of vertices v_1, v_2, \dots, v_m , with $(v_k, v_{k+1}) \in A$ for each k , also denoted by $v_k \rightarrow v_{k+1}$, where v_1, v_2, \dots, v_{m-1} are required to be distinct. A *directed cycle* is a directed path with $v_1 = v_m$. A *trail* θ in a graph is a sequence of unique vertices v_1, v_2, \dots, v_m , where we have for each k that $v_k \rightarrow v_{k+1}$ or $v_{k+1} \rightarrow v_k$; each arc occurs only once. A *subtrail* of a trail v_1, v_2, \dots, v_m is a sequence v_i, v_{i+1}, \dots, v_j , $i < j$. A trail θ connecting vertices u and v is also written as $u \sim v$. The set of all trails of an ADG G is denoted by Θ . A graph $G'_{|\Theta'} = (V', A')$ is said to be a *reduced subgraph* of graph $G = (V, A)$ with associated set of trails Θ' if $V' \subseteq V$, A' consists of all arcs of the set of trails Θ' with $\Theta' \subseteq \Theta$, and Θ' is based on the set of vertices V' .

Let X be a set of *discrete* random variables and let V act as its index set, i.e., X_v with $v \in V$ denotes a random variable and X_W with $W \subseteq V$ denotes a set of random variables. Furthermore, let P denote a joint probability distribution (JPD) of X_V . The set X_U is said to be *conditionally independent* of X_W given X_Z , with $U, W, Z \subseteq V$, if

$$P(X_U | X_W, X_Z) = P(X_U | X_Z). \quad (1)$$

These independence relations in P can also be represented by means of an ADG G , in which the entire set of independence relations is denoted by $\perp\!\!\!\perp_G$. In the graph, arcs represent dependences, and absence of arcs represents (conditional) independences.

Independences can be read off from an ADG by the *d-separation criterion*, defined as follows [5]: a trail θ in an ADG G is said to be *blocked* by a set Z if one of the following conditions is satisfied: (i) $v \in Z$ and v appears on the trail

θ , and either *no* or *only one* of the arcs of θ meeting at v is directed to v ; (ii) $v \notin Z$, $\delta(v) \cap Z = \emptyset$, where $\delta(v)$ are the descendants of v , and *both* arcs meeting at v on θ are directed to v (*convergent connection*). It is said that the sets U and W are *d-separated* by Z if any trail between a vertex in U and a vertex in W is blocked by the set Z ; formally: $U \perp\!\!\!\perp_G W \mid Z$. Otherwise, U and W are *d-connected* by Z , denoted by $U \not\perp\!\!\!\perp_G W \mid Z$.

A *Bayesian network* is defined as a pair $\mathcal{B} = (G, P)$, where $G = (V, A)$ is an acyclic directed graph representing relations of random variables X_V , P is the JPD on X_V , and each independence represented in G is also a valid independence in the JPD P .

4 Dynamic Bayesian Networks

In this section, the foundation for independence modularisation in DBNs is developed, based on the separation of temporal and atemporal independences.

DBNs are an extension of ordinary Bayesian networks and allow modelling the uncertainty involved in time-oriented processes. As a start, a representation of *time* is required, which in this paper is denoted by T and is assumed to be a subset of the set of the natural numbers; a *time point* t is then a member of T . The graphical representation of a DBN consists of two parts: (i) an atemporal part, and (ii) a temporal part. We subsequently define these parts.

An acyclic directed graph $G_t = (V_t, A_t^a)$, with set of vertices V_t and set of arcs $A_t^a \subseteq V_t \times V_t$, $t \in T$, is called a *timeslice* at time t , and its set of arcs A_t^a is called the set of *atemporal arcs*. Timeslices will be depicted by rectangles. The set of all timeslices G of a DBN is taken as:

$$G = \{G_t \mid t \in T\} = (V_T, A^a) . \quad (2)$$

Let $G_t = (V_t, A_t^a)$ and $G_{t'} = (V_{t'}, A_{t'}^a)$, $t, t' \in T$, $t \neq t'$, be two distinct timeslices. Then, an arc $(u, v) \in V_t \times V_{t'}$ with $t < t'$ is called a *temporal arc*. The set of temporal arcs of the set of all time slices G is denoted by A^t . Thus, temporal arcs connect vertices in different timeslices; they direct *always* from the past to the future and are drawn as dotted arrows.

Example 1. Consider Fig. 2; here the set of timeslices is equal to $G = \{G_1, G_2, G_3\}$. Timeslice G_2 is defined as $G_2 = (V_2, A_2^a)$, where the set of its vertices is equal to $V_2 = \{q_2, s_2, v_2, z_2\}$ and the set of its atemporal arcs $A_2^a = \{(q_2, s_2), (v_2, z_2)\}$. Moreover, the temporal arcs are equal to $A^t = \{(u_1, v_2), (w_1, v_2), (v_2, r_3)\}$.

Temporal arcs connect timeslices; this allows to construct temporal networks.

Definition 1. (*temporal network*) Let $G = (V_T, A^a)$ be a set of timeslices. Then, a temporal network N is defined as a pair $N = (V_T, A)$, where $A = A^a \cup A^t$.

Clearly, as each timeslice $G_t \in G$ is an acyclic directed graph, and timeslices are connected by temporal arcs pointing from the past to the future, N is also an

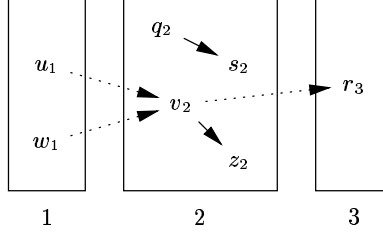


Fig. 2. An example of a dynamic Bayesian network.

acyclic directed graph. A DBN is defined as a pair $DBN = (N, P)$, where P is the JPD on the entire set of random variables.

Both temporal and atemporal relations in the network can be represented by means of trails. An *atemporal trail* contains no temporal arcs and is denoted by θ^a . A *temporal trail* consists of *at least one* temporal arc and is denoted by θ^t . The sets of all atemporal and temporal trails are denoted by Θ^a and Θ^t , respectively.

Example 2. Fig. 2 includes temporal trails $\theta_1^t = u_1 \rightarrow v_2 \leftarrow w_1$ and $\theta_2^t = z_2 \leftarrow v_2 \rightarrow r_3$, and the only two atemporal trails are $\theta_1^a = q_2 \rightarrow s_2$ and $\theta_2^a = v_2 \rightarrow z_2$.

With regards to the temporal relationships we only need to consider temporal trails that result into a reduced temporal network.

Definition 2. (*reduced temporal network*) Let $N = (V_T, A)$ be a temporal network. Then, $N_{|\Theta^t} = (V_T, A_{\Theta^t})$ is called a reduced temporal network if its set of arcs $A_{\Theta^t} \subseteq A$ consists of all the arcs included on the temporal trails in Θ^t .

Observe that the reduced temporal network is based on the set of temporal *trails*, which may consist of both atemporal and temporal trails. A further partitioning of the reduced temporal network is based on its set of *arcs*. This partitioning is obtained by decomposing a reduced temporal network into two parts, where one part consists of only atemporal and the another part of only temporal arcs. The *atemporal part of the reduced temporal network* is denoted by $N_{|\Theta^t}^a = (V_T, A_{\Theta^t}^a)$, where V_T is the set of vertices and $A_{\Theta^t}^a \subseteq A_{\Theta^t}^a$ consists of all atemporal arcs in the reduced temporal network. The *temporal part of the reduced temporal network* is denoted by $N_{|\Theta^t}^t = (V_T, A_{\Theta^t}^t)$, where $A_{\Theta^t}^t \subseteq A_{\Theta^t}^t$ consists of all temporal arcs in the reduced temporal network.

As a DBN includes temporal and atemporal elements, the question is how to distinguish between these relations. Vertices U and W are said to be *atemporally d-separated (temporally d-separated)* by Z , denoted by $\perp\!\!\!\perp_G$ ($\perp\!\!\!\perp_{N_{|\Theta^t}}$), if all the atemporal (temporal) trails connecting U and W are d-separated given Z . Atemporal d-separation of vertices belonging to only one timeslice G_t is denoted by $\perp\!\!\!\perp_{G_t}$. Finally, $\perp\!\!\!\perp_N$ denotes the set of independences in the temporal network N .

Fig. 3 shows an example of a temporal network and includes the various temporal and atemporal parts defined above.

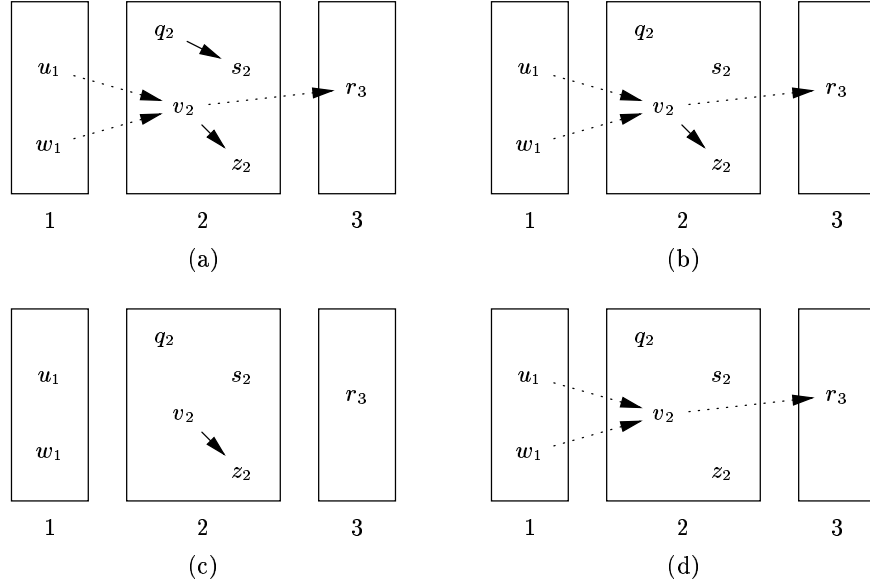


Fig. 3. Temporal and atemporal parts of a temporal network: (a) temporal network, (b) reduced temporal network, (c) atemporal part of the reduced temporal network, and (d) temporal part of the reduced temporal network.

5 The Join Operator

In this section, we study the correct composition of the temporal and atemporal independence relations in DBNs.

Suppose we want to join two independence relations $\perp\!\!\!\perp$ and $\perp\!\!\!\perp'$, both defined on the same set of variables. Then, the following three situations need to be considered: (i) joining one dependence and one independence statement; (ii) joining two dependence statements; (iii) joining two independence statements.

To correctly join the statements from situations (i) and (ii) a dependence preservation property is proposed below. Similarly, in order to deal with situation (iii), an independence concatenation property is defined below.

5.1 Dependence Preservation

The reason that dependence preservation is required for joining dependence and independence or dependence and dependence statements can be explained in terms of the concepts of consistency and dominance; these terms are interpreted as follows.

Let the independence relations $\perp\!\!\!\perp$ and $\perp\!\!\!\perp'$ be defined on the same vertex set V . Then, if there exist statements $U \perp\!\!\!\perp W \mid Z$ and $U \perp\!\!\!\perp' W \mid Z$ for arbitrary, mutually disjoint sets of vertices $U, W, Z \subseteq V$, then these independence statements and, therefore, independence relations $\perp\!\!\!\perp$ and $\perp\!\!\!\perp'$ are said to

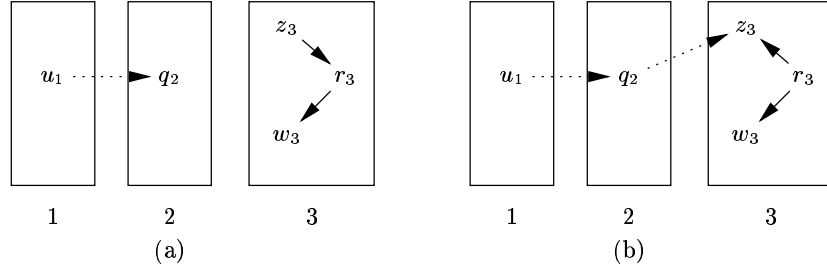


Fig. 4. Temporal networks (a) and (b).

be *inconsistent*. Otherwise, the statements are *consistent*. If we wish to join independence relations together, and two independence relations are inconsistent, a choice has to be made between the independence and dependence. In other words, one statement has to *dominate* the other one. If the relations $\perp\!\!\!\perp$ and $\perp\!\!\!\perp'$ are inconsistent due to the statements $U \perp\!\!\!\perp W \mid Z$ and $U \not\perp\!\!\!\perp' W \mid Z$, then $U \not\perp\!\!\!\perp' W \mid Z$ is said to *dominate* $U \perp\!\!\!\perp W \mid Z$. Furthermore, the dominance of dependences also indicates that two dependence statements, mentioned for situation (ii), also have to be joined into a dependence statement. Since dominance has to be taken into account when joining independence relations, the following property is defined.

Definition 3. (dependence preservation) Let $\perp\!\!\!\perp$, $\perp\!\!\!\perp'$ and $\perp\!\!\!\perp''$ be independence relations all defined on V . Suppose that $U \not\perp\!\!\!\perp W \mid Z$ or $U \not\perp\!\!\!\perp' W \mid Z$, or both, then it holds that $U \not\perp\!\!\!\perp'' W \mid Z$ for all $U, W, Z \subseteq V$. It is said that $\perp\!\!\!\perp''$ satisfies the dependence preservation property with regard to $\perp\!\!\!\perp$ and $\perp\!\!\!\perp'$.

5.2 Independence Concatenation

The independence concatenation property takes into account how independence relations are combined.

Definition 4. (independence concatenation) Let $\perp\!\!\!\perp$, $\perp\!\!\!\perp'$ and $\perp\!\!\!\perp''$ be independence relations all defined on V . Suppose that for some $U, W, Z \subseteq V$ independences $U \perp\!\!\!\perp W \mid Z$ and $U \perp\!\!\!\perp' W \mid Z$ hold and then it also holds that $U \perp\!\!\!\perp'' W \mid Z$. Then, it is said that $\perp\!\!\!\perp''$ satisfies the independence concatenation property with regard to the set of vertices U , W and Z in relations $\perp\!\!\!\perp$ and $\perp\!\!\!\perp'$.

In the remaining part of this section, we investigate how the independence concatenation property can be applied to ADGs. The following example indicates that joining independences is *not* straightforward in these graphs, since an independence may change into a dependence.

Example 3. Consider Fig. 4. In both the atemporal and temporal parts of (a) and (b) vertex u_1 is conditionally independent of w_3 given z_3 ; formally, we have that $u_1 \perp\!\!\!\perp_{N_{\emptyset}^a} w_3 \mid z_3$ and $u_1 \perp\!\!\!\perp_{N_{\emptyset}^t} w_3 \mid z_3$. In the reduced temporal network of

(a) u_1 is still conditionally independent of w_3 given z_3 ; however, in the reduced temporal network of (b), u_1 is conditionally dependent of w_3 given z_3 ; formally: $u_1 \perp\!\!\!\perp_{N_{|\Theta^t}} w_3 \mid z_3$ and $u_1 \not\perp\!\!\!\perp_{N_{|\Theta^t}} w_3 \mid z_3$.

Thus, it is necessary to investigate how to join two independence statements. As mentioned above, in Bayesian networks d-separation criterion is used to read-off independence statements, based on the study of the arc directions in the trails in the graphical representations. Therefore, when we want to join two independence relations $\perp\!\!\!\perp_G$ and $\perp\!\!\!\perp_{G'}$ into $\perp\!\!\!\perp_{G''}$, we need to investigate the trails in graph G'' obtained from the set of trails of the graphs G and G' . Then, any trail in G'' can be partitioned into a set of subtrails, called *special subtrails*, where each subtrail consists of arcs obtained from only *one* of the graphs G or G' .

Example 4. Reconsider Fig. 3 and suppose we want to join the atemporal and temporal parts (c) and (d) of the reduced temporal network obtaining the entire independence relation of the reduced temporal network (b). Suppose we want to determine the relation between vertices u_1 and z_2 . Then, we need to study trail $u_1 \rightarrow v_2 \rightarrow z_2$ in graph (b), which consists of two special subtrails, atemporal trail $v_2 \rightarrow z_2$ in (c) and temporal trail $u_1 \rightarrow v_2$ in (d).

Considering the issues mentioned above, when we need to join two independence statements, we only need to analyse trails in G'' that consist of more than one special subtrail; otherwise the relation is just an independence statement, since its arcs belong only to one of the graphs G or G' .

After having singled out the set of special subtrails of a trail, we can determine the independence relations from the trail by looking at these subtrails. This is done by examining the independence and dependence statements, derived from $\perp\!\!\!\perp_G$ and $\perp\!\!\!\perp_{G'}$, between the initial and end vertex of each special subtrail, while conditioning on the same set of vertices as from the entire trail. Then, if at least one of these two statements is a dependence, according to the dependence preservation property, the initial and end vertex of the special subtrail will be conditionally dependent in G'' . However, in the case of two independence statements we need to apply d-separation. Furthermore, we also need to consider the set of so-called *shared vertices* of a trail, which are the vertices that connect the special subtrails of this trail. These considerations give rise to an easier way of the composition of two independence statements:

Proposition 1. *Let the independence relations $\perp\!\!\!\perp_G$, $\perp\!\!\!\perp_{G'}$ and $\perp\!\!\!\perp_{G''}$ both defined on V and let $U \perp\!\!\!\perp_G W \mid Z$, $U \perp\!\!\!\perp_{G'} W \mid Z$ for $U, W, Z \subseteq V$ hold in graphs $G = (V, A)$ and $G' = (V, A')$, respectively. Let $A'' \subseteq A \cup A'$. Let trail θ connect the two vertices $u \in U$ and $w \in W$ with each other, and let $\theta_1, \dots, \theta_n$ be the special subtrails of θ . Then, θ is blocked by Z if one of the following condition holds:*

- the trail θ consists of only one special subtrail;
- for one of the special subtrails $\theta_i = v_1, v_2, \dots, v_m$ we have $v_1 \perp\!\!\!\perp_G v_m \mid Z$ and $v_1 \perp\!\!\!\perp_{G'} v_m \mid Z$ and θ_i is blocked by Z in G'' according to d-separation;
- one of the shared vertices blocks θ according to d-separation.

The independence $U \perp_{G''} W \mid Z$ holds if each trail connecting any vertex in U and W is blocked by Z satisfying the independence concatenation property.

5.3 The Join Operator

Next, the join operator is defined and a significant property of this operator is considered.

Definition 5. (join operator) Let \perp and \perp' be two independence relations defined on the same vertex set V . The join of these two relations, denoted by $\perp \circ \perp' = \perp''$, is then again an independence relation, \perp'' , defined on V , that satisfies the dependence preservation and the independence concatenation properties.

Proposition 2. Let $G = (V, A)$, $G' = (V, A')$ and $G'' = (V, A'')$ be three ADGs, where $A \cup A' \subseteq A''$. Then, it holds that $\perp_{G''} \subseteq \perp_G \circ \perp_{G'}$.

Proof. The graph G'' contains at least as many arcs as the union of the graphs G and G' . Thus, $\perp_{G''}$ contains possibly extra dependences in addition to those in \perp_G and $\perp_{G'}$. As the join operator satisfies the independence concatenation property, any independence that results from applying d-separation to G'' is preserved by joining \perp_G and $\perp_{G'}$. Hence, it follows that $\perp_{G''} \subseteq \perp_G \circ \perp_{G'}$. \square

6 Temporal and Atemporal Interaction

Based on the results above, in this section we investigate how to employ the join operator for combining the temporal and atemporal relations underlying temporal networks to support the modelling of *non-repetitive* DBNs.

6.1 Joining Atemporal and Reduced Temporal Networks

In this section, we start by considering the relations \perp_G and \perp_{G_t} . The following proposition establishes that the join operator \circ can be interpreted as the intersection of the independence relations in G_t .

Proposition 3. Let $DBN = (N, P)$ with temporal network $N = (V_T, A)$, set of timeslices $G = (V_T, A^a)$, $A = A^a \cup A^t$, and the joint probability distribution P . Then, it holds that:

- (i) $\perp_G = \bigcap_{t \in T} \perp_{G_t}$;
- (ii) $\perp_G = \bigcup_{t \in T} \perp_{G_t}$.

The graph-theoretic interpretation of dependence preservation and independence concatenation for reduced temporal networks is investigated next.

Proposition 4. Let $N_{|\Theta^t}^a$ and $N_{|\Theta^t}^t$ be the atemporal and temporal parts of reduced temporal network $N_{|\Theta^t}$. Then, the dependence preservation and independence concatenation properties hold for the independence relations $\perp_{N_{|\Theta^t}^a}$, $\perp_{N_{|\Theta^t}^t}$ and $\perp_{N_{|\Theta^t}}$.

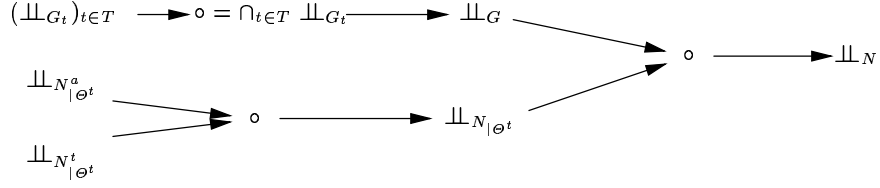


Fig. 5. Joining temporal and atemporal independence relations.

Next we will show that the join operator can be used to merge the two independence relations, proving its soundness and completeness. *Soundness* of the join operator means that all independence statements obtained by joining two independence relations can be read off from the union of the underlying graphs, whereas *completeness* means that none of the independence statements of the union of the graphs has been omitted in the resulting independence relation.

Theorem 1. *Let $\perp_{N_{|\Theta^t}^a}$, $\perp_{N_{|\Theta^t}^t}$, and $\perp_{N_{|\Theta^t}}$ be constructed as defined above. Then, it holds that $\perp_{N_{|\Theta^t}} = \perp_{N_{|\Theta^t}^a} \circ \perp_{N_{|\Theta^t}^t}$, i.e. the join operator is sound and complete.*

Proof: *Soundness* follows from the independence concatenation property. By this property only independence statements that hold can be derived, because it is based on Proposition 1 and, therefore, on temporal d-separation. *Completeness* follows from Proposition 2 by substituting $N_{|\Theta^t}^a$, $N_{|\Theta^t}^t$, and $N_{|\Theta^t}$ for acyclic directed graphs G , G' and G'' , respectively. \square

6.2 Joining It All Together

In this subsection, the temporal and atemporal independence relations are joined together, yielding the relation \perp_N . To start, the following proposition and theorem show that these relations can be linked to each other by means of the join operator.

Proposition 5. *Let G and $N_{|\Theta^t}$ be the atemporal and temporal parts of temporal network N . Then, the dependence preservation and independence concatenation properties hold for the independence relations \perp_G , $\perp_{N_{|\Theta^t}}$ and \perp_N .*

Theorem 2. *Let \perp_G , $\perp_{N_{|\Theta^t}}$ and \perp_N be constructed as defined above. Then, it holds that $\perp_N = \perp_G \circ \perp_{N_{|\Theta^t}}$, i.e. the join operator is sound and complete.*

Proof: *Soundness* and *completeness* are similar to those of Theorem 1. \square

Fig. 5 provides a summary of propositions 3, 4 and 5. Finally, the various independence relations can be compared to each other.

Proposition 6. *The following properties hold:*

- $\perp_{N_{|\Theta^t}} \subseteq \perp_{N_{|\Theta^t}^a}$ and $\perp_{N_{|\Theta^t}} \subseteq \perp_{N_{|\Theta^t}^t}$;
- $\perp_N \subseteq \perp_G$ and $\perp_N \subseteq \perp_{N_{|\Theta^t}}$.

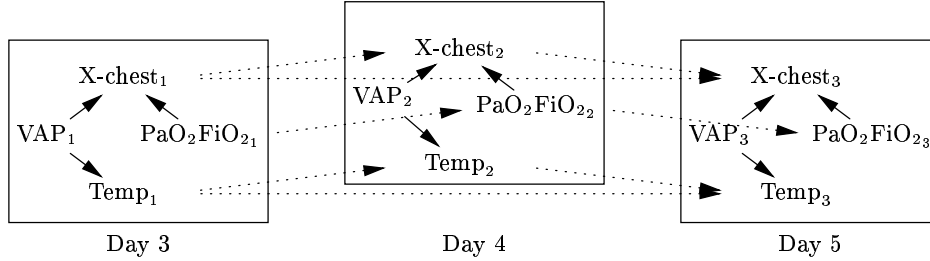


Fig. 6. The repetitive DBN for VAP.

7 Experimental Results

Using available temporal data of 876 patients with and without VAP, we experimentally compared repetitive and non-repetitive DBNs with the aim of demonstrating the usefulness of the framework. The experiments were done using Murphy’s BNT toolbox [6].

First, based on the non-repetitive DBN of progression of VAP, shown in Fig. 1, a repetitive DBN was constructed; it is shown in Fig. 6. The graphical structure of the non-repetitive network was obtained by using the join operator described in Section 6. Evaluation of the quality of the two DBNs was done using stratified tenfold cross-validation, i.e., the VAP dataset was divided randomly into 10 parts, where 9 parts acted as training sets, whereas the remaining part acted as a test set. The entire cross-validation process was repeated 10 times. In the process, the probability distributions of the DBNs were estimated using the EM algorithm [4].

As a measure of the quality of the two DBNs use was made of the log-likelihood function l [4]:

$$l(\mathcal{B}; D) = \sum_{d \in D} \log P_{\mathcal{B}}(d),$$

where \mathcal{B} denotes a DBN, D denotes the test set of data of patients from the ICU with and without VAP, and $d \in D$ denotes a tuple with patient data. The higher the value of the log-likelihood is, the better the distribution fits to the data. For each iteration of cross-validation, the average log-likelihood was computed.

The experimental results obtained are shown in Fig. 7. On the abscis are the iteration numbers of individual tenfold cross-validation processes. Although the log-likelihood varied for each of the DBNs between the different iterations, the log-likelihood for the non-repetitive DBN is *always* better than for the repetitive DBN. Thus, there exist datasets for which non-repetitive DBNs outperform repetitive DBNs, at the same time offering a more natural and precise form of knowledge representation.

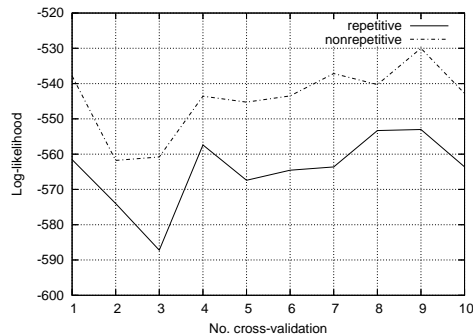


Fig. 7. The log-likelihoods of the repetitive and non-repetitive DBNs.

8 Conclusions

The aim of the research described in this paper was to develop a framework for modelling non-repetitive and repetitive DBNs, based on a modularisation of independence relations. It appeared that by distinguishing between temporal and atemporal independence relations, modelling DBNs can be greatly facilitated.

How to build a DBN from its atemporal and temporal parts is not obvious. This problem was tackled by the introduction of a join operator with special semantics. Using the join operator allows one to build DBNs in a modular fashion, which in particular is important when designing non-repetitive DBNs. As far as we know, our paper offers the first systematic method for building non-repetitive DBNs.

The practical usefulness of the method in the context of learning DBNs from data was also explored, using a real-world problem, and evidence of the validity of the method was obtained.

References

1. M. Deviren and K. Daoudi. Continuous speech recognition using dynamic Bayesian networks: a fast decoding algorithm. In: Proc PGM'02, Spain, 2002, pp. 54–60.
2. N. Friedman, K. Murphy and S. Russell. Learning the structure of dynamic probabilistic networks. In: Proc 14th UAI, 1998, pp. 139–147.
3. F.V. Jensen. Bayesian Networks and Decision Graphs. Springer, New York, 2001.
4. T. Hastie, R. Tibshirani and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, 2001.
5. R.G. Cowell, A. Philip Dawid, S. L. Lauritzen and D.J. Spiegelhalter. Probabilistic Networks and Expert Systems Springer-Verlag, New York, 1999.
6. K.P. Murphy. Dynamic Bayesian Networks: Representation, Inference and Learning. PhD Thesis, UC Berkeley, 2002.
7. U. Kjaerulff. A computational scheme for reasoning in dynamic probabilistic networks. In: Proc UAI'92, 1992, pp. 121–129.
8. A. Tucker and X. Liu. Learning Dynamic Bayesian Networks from Multivariate Time Series with Changing Dependencies. IDA, 2003.