Article:
Barclay, LM, Collazo, RA, Smith, JQ et al. (2 more authors) (2015) The Dynamic Chain Event Graph. Electronic Journal of Statistics, 9 (2). 2130 - 2169. ISSN 1935-7524

https://doi.org/10.1214/15-EJS1068

Electronic Journal of Statistics
Vol. 9 (2015) 2130–2169
ISSN: 1935-7524
DOI: 10.1214/15-EJS1068

The dynamic chain event graph

Lorna M. Barclay, Rodrigo A. Collazo, Jim Q. Smith

Department of Statistics
University of Warwick
Coventry, CV4 7AL
United Kingdom
e-mail: L.M.Barclay@warwick.ac.uk; R.A.Collazo@warwick.ac.uk; J.Q.Smith@warwick.ac.uk

Peter A. Thwaites

School of Mathematics
University of Leeds
Leeds, LS2 9JT
United Kingdom
e-mail: P.A.Thwaites@leeds.ac.uk

and

Ann E. Nicholson

Faculty of Information Technology
Building 72, Clayton campus
Monash University
VIC 3800, Australia
e-mail: ann.nicholson@monash.edu

Abstract: In this paper we develop a formal dynamic version of Chain Event Graphs (CEGs), a particularly expressive family of discrete graphical models. We demonstrate how this class links to semi-Markov models and provides a convenient generalization of the Dynamic Bayesian Network (DBN). In particular we develop a repeating time-slice Dynamic CEG providing a useful and simpler model in this family. We demonstrate how the Dynamic CEG's graphical formulation exhibits asymmetric conditional independence statements and also how each model can be estimated in a closed form enabling fast model search over the class. The expressive power of this model class together with its estimation is illustrated throughout by a variety of examples that include the risk of childhood hospitalization and the efficacy of a flu vaccine.

Keywords and phrases: Chain Event Graphs, Markov processes, probabilistic graphical models, dynamic Bayesian networks.

Received September 2014.

1. Introduction

In this paper we propose a novel class of graphical models called Dynamic Chain Event Graph (DCEG) to model longitudinal discrete processes that exist in

many diverse domains such as medicine, biology and sociology. These processes often evolve over long periods of time allowing studies to collect repeated multivariate observations at different time points. In many cases they describe highly asymmetric unfoldings and context-specific structures, where the paths taken by different units are quite different. Our objective is to develop a graphical framework that facilitates reasoning about conditional independence statements and simplifies statistical inference for those dynamic processes. The last issue includes finding a well-fitted graph from data and quantifying the parameters of a given graphical model.

In the literature there are various dynamic graphical models to model longitudinal data. The most widely used is the Dynamic Bayesian Network (DBN) (Dean and Kanazawa (1989); Nicholson (1992); Kjærulff (1992)), where the process in each time-slice is modelled using a Bayesian Network (BN) and the temporal dynamic is embedded in the model by temporal edges connecting these different BNs. To allow for irregular time-steps, Nodelman et al. (2002) suggested the development of a Continuous-Time BN (CTBN) whose variables evolve continuously over time.

However a DBN (and also a BN) or a CTBN do not allow us to model context-specific conditional independencies directly in their graphs and thus in the statistical models. A DBN or a CTBN do this analytically but in a hidden way by absorbing these context-specific statements into the implicit structures within their conditional probability tables. This graphical limitation is illustrated in Example 1 using a BN to model a very simple process.

Example 1. *Suppose that we would like to analyse the impact of weather on traffic in a medium-size city. For this purpose, take the explanatory variable "weather" with categories dry, drizzle, rain and a binary response variable "traffic" which represents the risk of traffic jam with categories low and high. Assume now that dry or drizzle have the same effect on traffic but that rain increases the risk of a traffic jam substantially. Figure 1 depicts a standard BN for this process. Note that without defing new random variables it is not possible to represent the context-specific statement graphically (Dry weather and drizzle have the same impact on traffic).*

$$Weather \longrightarrow Traffic$$

FIG 1. *BN of traffic example.*

Another interesting class of dynamic graphical models is the local independence graph (Didelez (2008)) or the graphical duration model (Gottard (2007)). They have been developed to model event history data, which describe the relationship between a particular set of events that happen over time. These graphical models explore the local independence structures that may be presented in their corresponding processes. They, however, assume that the conditional independencies do not change with time, and relationships can be naturally expressed in terms of marked point processes.

Here we propose a different graphical framework based on a tree to model longitudinal data, which are observed at not necessarily regular time-steps. We can incorporate many potential context-specific conditional independencies that may vary over time within this class. This enables us to estimate each model in a tractable and transparent way. In spite of their power and flexibility to model diverse domains, previous graphical models are not able to enjoy all these advantages.

Recently tree-based graphical models have been successfully used to describe various phenomena. This tree provides a flexible graphical support through which time sequences can be easily incorporated. Each path in the tree describes the various possible sequences of events a unit can experience. One such alternative tree based model is the Chain Event Graph (CEG) (Smith and Anderson (2008); Freeman and Smith (2011a); Thwaites (2013); Barclay et al. (2014)). In a CEG not only conditional independencies, but also context-specific symmetries, are directly depicted in the *topology* of the graph, see Example 1 below. Furthermore structural zero probabilities in the conditional probability tables are directly depicted by the absence of edges in its graph. See, for example, Smith and Anderson (2008); Barclay et al. (2013); Cowell and Smith (2014).

Example 1 (continued)**.** *Observe that even without being formally familiar with the CEG semantic we can read from Figure 2 that the risk of traffic is identical for weather classed as dry or drizzle but differs in the case of rain.*

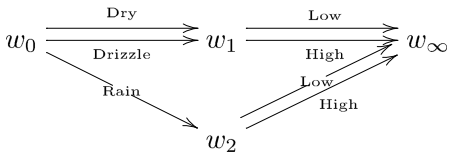


FIG 2. *CEG of traffic example.*

It has been recently discovered that a CEG also retains most of the useful properties of a BN like closure to learning under complete sampling (Freeman and Smith (2011a)) and causal expressiveness (Thwaites (2013); Thwaites et al. (2010); Riccomagno and Smith (2009); Thwaites and Smith (2006a)). It also supports efficient propagation of new information (Thwaites et al. (2008); Thwaites and Smith (2006b)). Hence CEGs provide an expressive framework for various tasks associated with graphical representation and statistical inference, especially when the tree of the underlying sample space is asymmetric (French and Insua (2010)).

The class of CEGs contains all discrete BNs (see Smith and Anderson (2008), section 3.2, p. 56 for the proof), as well as all the extension of the BN to context-specific BNs (Boutilier et al. (1996); Friedman and Goldszmidt (1998)) and Bayesian multinets (Geiger and Heckerman (1996); Bilmes (2000)). This fact guarantees that *all* the conditional independencies entailed by these model classes are embodied in the topology of the graph of a single CEG (see for

example Smith and Anderson (2008); Thwaites and Smith (2011) as well as many others). The topology of the CEG has hence been exploited to fully represent and generalize models such as context-specific BNs.

It has became increasingly apparent that in many contexts modelling variable changes explicitly over time provide better results. For a comparison between BNs and DBNs, see e.g. Rubio et al. (2014). Currently there is no such dynamic CEG defined in the literature. Freeman and Smith (2011b) developed one dynamic extension of CEGs where the underlying probability tree is finite but the stage structure of the possible CEGs is allowed to change across discrete time-steps. This model, however, develops an entirely distinct class of models to the one considered here. It looks at different cohorts of units entering the tree at discrete time-points rather than assuming that repeated measurements are taken over time.

In this paper we develop the DCEG model class that extends the CEG model class so that it contains all dynamic BNs as a special case. In this sense we discover an exactly parallel extension to the original CEG extension of the BN class. We show that any infinite tree can be rewritten as a DCEG, which represents the originally elicited tree in a much more compact and easily interpretable form. A DCEG actually provides an evocative representation of its corresponding process. It allows us to define many useful DCEG model classes, such as the Repeating Time-Slice DCEG (RT-DCEG), that have a finite model parameter space.

A DCEG also supports conjugate Bayesian learning where the prior distribution chosen in a family of probability distributions $\mathcal{A}$ together with the available likelihood function yield a posterior distribution in the same family $\mathcal{A}$. This is a necessary requirement to guarantee analytical tractability and hence to design clever model search algorithms that are able to explore the large numbers of collections of hypotheses encoded within the DCEG model space. We further demonstrate that we can extend this framework by attaching holding time distributions to the nodes in the graph, so that we can model processes observed at irregular time intervals. Our learning framework is closely related to the one developed by Nodelman et al. (2003) for CTBNs.

In Section 2 we present some important graph concepts and the definitions of a BN, a DBN and a CEG. In Section 3 we formally define the infinite staged tree and the DCEG. We further introduce the Extended DCEG which attaches conditional holding times to each edge within the graph. We also define a special class of DCEGs called the RT-DCEG which imposes certain restrictions on the more general class of DCEGs. In Section 4 we show how to perform fast conjugate Bayesian estimation of these model classes and demonstrate how a typical model can be scored. In Section 5, we demonstrate that any general DBN lies in the class of DCEGs and so show that DCEGs are a formal extension of DBN models. We then present some connections between the (Extended) DCEG model class and some (Semi-) Markov processes. We conclude the paper with a short discussion.

2. Background

In this section we revisit some graph notions that will be useful to discuss graphical models. Next we explain briefly the BN and DBN models and then define a CEG model. See Korb and Nicholson (2004); Neapolitan (2004); Cowell et al. (2007); Murphy (2012) for more detail on BNs and DBNs. The CEG concepts presented here are a natural extension of those in Smith and Anderson (2008); Thwaites et al. (2010); Freeman and Smith (2011a). These conceptual adaptations will allow us to directly use these concepts to define a DCEG model.

2.1. Graph Theory and conditional independence

Definition 1. *Graph Let a* graph $\mathcal{G}$ *have vertex set* $V(\mathcal{G})$ *and a (directed) edge set* $E(\mathcal{G})$*, where for each edge* $e(v_i, v_j) \in E(\mathcal{G})$*, there exists a directed edge* $v_i \to v_j$*,* $v_i, v_j \in V(\mathcal{G})$*. Call the vertex* v_i *a* parent *of* v_j *if* $e(v_i, v_j) \in E(\mathcal{G})$ *and let* $pa(v_j)$ *be the set of all parents of a vertex* v_j*. Also, call* v_k *a* child *of* v_i *if* $e(v_i, v_k) \in E(\mathcal{G})$ *and let* $ch(v_i)$ *be the set of all children of a vertex* v_i*. We say the graph is* infinite *when either the set* $V(\mathcal{G})$ *or the set* $E(\mathcal{G})$ *is infinite.*

Definition 2. *Directed Acyclic Graph*

A directed acyclic graph (DAG) $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ *is a graph all of whose edges are directed with no directed cycles – i.e. if there is a directed path from vertex* v_i *to vertex* v_j *then a directed path from vertex* v_j *to vertex* v_i *does not exist.*

Definition 3. *Tree*

A tree $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ *is a connected graph with no undirected cycles. Here we only consider a directed rooted tree. In this case, it has one vertex, called the root vertex* v_0*, with no parents, while all other vertices have exactly one parent. A* leaf vertex *in* $V(\mathcal{T})$ *is a vertex with no children. A* level $\mathcal{L}$ *is the set of vertices that are equally distant from the root vertex. A tree is an infinite tree if it has at least one infinite path.*

Definition 4. *Floret*

A floret *is a subtree* $\mathcal{F}(s_i) = (V(\mathcal{F}(s_i)), E(\mathcal{F}(s_i)))$ *of* $\mathcal{T}$*,* $s_i \in S(\mathcal{T})$ *where:*

- *its vertex set* $V(\mathcal{F}(s_i))$ *consists of* $\{s_i\} \cup ch(s_i)$*, and*
- *its edge set* $E(\mathcal{F}(s_i))$ *consists of all the edges between* s_i *and its children in* $\mathcal{T}$*.*

There are various alternative ways of defining conditional independence. For the purposes of this paper it is more convenient to use the definition below.

Definition 5. *Conditional Independence*

A discrete random variable X_a *is conditionally independent of a discrete random variable* X_b *given a set of discrete random variables* $\mathcal{X} = \{X_1, \ldots, X_n\}$ *if for every triple* $(x_a, x_b, \boldsymbol{x})$*, where* $\boldsymbol{x} = (x_1, \ldots, x_n)$*, we have that*

$$P(X_a = x_a | X_b = x_b, \mathcal{X} = \boldsymbol{x}) = P(X_a = x_a | \mathcal{X} = \boldsymbol{x}). \tag{1}$$

We write this conditional independence statement as $X_a \perp\!\!\!\perp X_b | \mathcal{X}$*.*

2.2. A Bayesian Network and a Dynamic Bayesian Network

A Bayesian Network is a probabilistic graphical model whose support graph is a DAG $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$. Each vertex $v_i \in V(G)$ represents a variable Z_i and the edge set $E(\mathcal{G})$ denotes the collection of conditional dependencies that are assumed in the variable set. Here we assume that the vertex set $V(\mathcal{G}))$ is well-ordered and that there can be a directed edge $e(v_i, v_j)$ if and only if $i < j$. Thus the variable Z_j is conditionally independent of variable Z_i given the variable set $\mathcal{Z}_j^i = \{Z_1, \ldots, Z_{j-1}\} \setminus \{Z_i\}$ whenever an edge $e(v_i, v_j)$ does not exist in $E(\mathcal{G})$. Formally,

$$Z_j \perp\!\!\!\perp Z_i | \mathcal{Z}_j^i \Leftrightarrow e(v_i, v_j) \notin E(\mathcal{G}). \tag{2}$$

Recall that the concept of conditional independence plays a central role in this model class (Dawid (1998); Pearl (2009), Chapter 1, p. 1–40). Below we give a simple example that is more complex than the first since it entails not only context-specific independencies but also various asymmetric developments.

Example 2. *An individual is at risk of catching flu. Having caught flu he either decides to take antiviral treatment or not (Treatment variable, see Figure 3). If he takes the antiviral treatment we assume that he will always recover. On the other hand if he does not take the antiviral treatment he either manages to recover or he dies from the virus (Recovery variable, see Figure 3). Given a full recovery the individual can either decide to go back to his normal life or to receive an influenza vaccine to prevent him from being at risk again (Vaccine variable, see Figure 3). We further hypothesise that the decision of taking a vaccine is conditionally independent of the decision to take the antiviral treatment given, of course, that the individual is alive. Thus the Recovery and Vaccine variables depends, respectively, on the Treatment and Recovery variables, but the Vaccine variable is conditionally independent of the Treatment variable given the Recovery variable. Figure 3 shows a standard BN to model this process.*

$$Treatment \longrightarrow Recovery \longrightarrow Vaccine$$

FIG 3. *BN of flu example.*

Another class of graphical model we will discuss and compare in this paper is the Dynamic Bayesian Network (DBN) which models the temporal changes in the relationships among variables. It extends directly the BN conception. A DBN $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ can be interpreted as a collection of BNs $\{\mathcal{G}_t = (V(\mathcal{G}_t), E(\mathcal{G}_t)); t = 1, 2, \ldots\}$ where the variables in time-slices t can also be affected by variables in the previous time-slices but not by variables in the next ones. These dependencies between time-slices are represented graphically by edges called temporal edges. Formally, $V(\mathcal{G}) = \bigcup_t V(\mathcal{G}_t)$ and $E(\mathcal{G}) = \bigcup_t [E(\mathcal{G}_t) \cup E_t]$, where E_t is the set of temporal edges $\{e(v_{i,\tau}, v_{j,\tau}); \tau < t\}$ associated with a BN $\mathcal{G}_t$ and $v_{i,\tau}$ represents a variable Z_i in time-slice $\tau < t$.

In this paper we only consider discrete BNs and discrete DBNs where all variables have discrete state spaces.

2.3. A Chain Event Graph

A full description of this construction for finite processes can be found in Smith and Anderson (2008); Freeman and Smith (2011a); Thwaites et al. (2010). Here we summarise this development.

Definition 6. *Event Tree*
An event tree is a finite tree $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ where all vertices are chance nodes and the edges of the tree label the possible events that happen. A non-leaf vertex of a tree T is called a situation and $S(\mathcal{T}) \subseteq V(\mathcal{T})$ denotes the set of situations.

The path from the root vertex to a situation $s_i \in S(\mathcal{T})$ therefore represents a sequence of possible unfolding events. The situation denotes the state that is reached via those transitions. Here we assume that each situation $s_i \in S(\mathcal{T})$ has a finite number of edges, m_i, emanating from it. A leaf node symbolises a possible final situation of an unfolding process. An edge can be identified by two situations s_i and s_j (edge $e(s_i, s_j)$) or by a situation s_i and one of its corresponding unfolding events k (edge $e_{s_i k}$).

Definition 7. *Stage*
We say two situations s_i and s_k are in the same stage, *u, if and only if*

1. *there exists an isomorphism Φ_{ik} between the labels of $E(\mathcal{F}(s_i))$ and $E(\mathcal{F}(s_k))$, where $\Phi_{ik}(e_{s_i j}) = e_{s_k j}$, and*
2. *their corresponding conditional probabilities are identical.*

When there is only a single situation in a stage, then we call this stage and its corresponding situation trivial.

If two situations are in the same stage then we assign the same color to their corresponding vertices. In other publications, for example Smith and Anderson (2008), corresponding edges of situations in the same stage are also given the same color. For clarity here we only color vertices and edges corresponding to non-trivial situations. We can hence partition the situations of the tree $S(\mathcal{T})$ into stages, associated with a set of isomorphisms $\{\Phi_{ik} : s_i, s_k \in S(\mathcal{T})\}$, and embellish the event tree with colors to obtain the staged tree.

Definition 8. *Staged Tree*
A staged tree *version of $\mathcal{T}$ is one where*

1. *all non-trivial situations are assigned a color*
2. *situations in the same stage in $\mathcal{T}$ are assigned the same color, and*
3. *situations in different stages in $\mathcal{T}$ are assigned different colors.*

We illustrate these concepts through a simple example on influenza, which we later develop further using a DCEG model.

Example 2 (continued)**.** *After eliciting the event tree (see Figure 4) corresponding to the Example 2 we can hypothesize possible probabilistic symmetries in this process. For example, we might assume that recovering with or without*

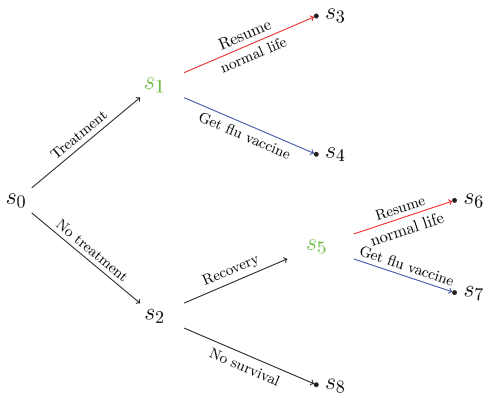


FIG 4. *Flu example.*

treatment will not affect the individual's probability to decide to get the vaccine. This demands that the probabilities on the edges emanating from s_1*, labeled "resume normal life" and "get vaccine", are identical to the probabilities on the edges emanating from* s_5 *with the same labels. This assumption can be visualized by coloring the vertices of the event tree and their corresponding edges.*

A finer partition of the vertices in a tree is given by the position partition. Let $\mathcal{T}(s_i)$ denote the full colored subtree with root vertex s_i.

Definition 9. *Position*

Two situations s_i, s_k *in the same stage, that is,* $s_i, s_k \in u \in U$*, are also in the same* position w *if there is a graph isomorphism* Ψ_{ik} *between the two colored subtrees* $\mathcal{T}(s_i) \to \mathcal{T}(s_k)$*. We denote the set of positions by* W*.*

The definition hence requires that for two situations to be in the same position there must not only be a map between the edge sets $E(\mathcal{T}(s_i)) \to E(\mathcal{T}(s_k))$ of the two colored subtrees but also the colors of any edges and vertices under this map must correspond. For example when all children of s_i, s_k are leaf nodes then $\mathcal{T}(s_i) = \mathcal{F}(s_i)$ and $\mathcal{T}(s_k) = \mathcal{F}(s_k)$. Therefore s_i and s_k will be in the same position if and only if they are in the same stage. But if two situations are further from a leaf, not only do they need to be in the same stage but also each child of s_i must correspond to a child of s_k and these must be in the same stage. This further applies to all children of each child of s_i and so on.

Definition 10. *Chain Event Graph (Smith and Anderson (2008))*

A CEG $\mathcal{C} = (V(\mathcal{C}), E(\mathcal{C}))$ *is a directed colored graph obtained from a staged tree by successive edge contraction operations. The situations in the staged tree are merged into the vertex set of positions and its leaf nodes are gathered into a single sink node* w_∞*.*

A CEG depicts not only the unfolding of events expressed in a tree but also the types of probabilistic symmetries.

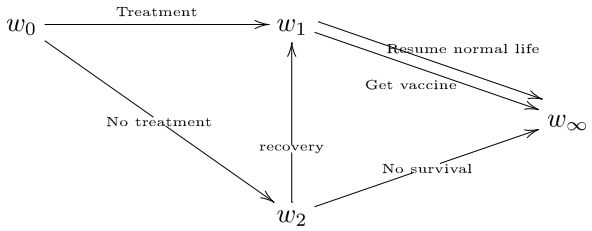


FIG 5. *CEG of flu example.*

Example 2 (continued)**.** *The hypothesis that recovering with or without treatment does not affect the probability of the individual taking the flu vaccine, places* s_1 *and* s_5 *in the same position. We then obtain the following CEG given in Figure 5 with stages and positions given by:*

$$w_0 = u_0 = \{s_0\}, w_1 = u_1 = \{s_1, s_5\}, w_2 = u_2 = \{s_2\}, w_\infty = \{s_3, s_4, s_6, s_7, s_8\}$$

Note that the corresponding BN (Figure 3) cannot depict graphically the asymmetric unfolding of this process and the context-specific conditional statements.

3. Infinite probability trees and DCEGs

In this section we extend the standard terminology used for finite trees and CEGs to infinite trees and DCEGs. In the first subsection we derive the infinite staged tree, followed by a formal definition of the DCEG. The next subsection we extend the DCEG to not only describe the transitions between the vertices of the graph but also the time spent at each vertex. Finally we define a useful class of DCEG models.

3.1. Infinite staged trees

Clearly an infinite event tree can be uniquely characterized by its florets, which retain the indexing of the vertices of $\mathcal{T}$. The edges of each floret can be labeled as $e_{s_i j} \in E(\mathcal{F}(s_i)), j = 1, \ldots, m_i$, where s_i has m_i children. As noted above, we can think of these edge labels as descriptions of the particular events or transitions that can occur after a unit reaches the root of the floret. In particular, we can also use the index $j = 1, \ldots, m_i$ to define a random variable taking values $\{x_1, \ldots, x_{m_i}\}$ associated with this floret.

Example 2 (continued)**.** *Assume that the individual is every month at risk of catching the flu. As before, given a full recovery from the virus, with or without treatment, the individual can either decide to go back to his normal life where he is at risk of catching flu again or decide to receive an influenza vaccine to prevent him from being at risk again. As the tree is infinite, only an informal depiction*

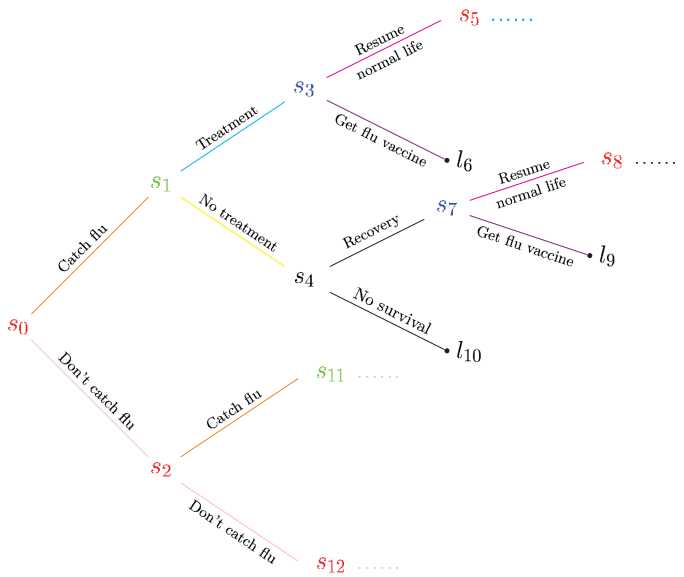


FIG 6. *Flu example: the beginning of the infinite staged tree,* $\mathcal{T}$.

of the corresponding tree can be given (Figure 6), where implicit continuations of the tree are given by the notation '...'.

In our example the edges $E(\mathcal{F}(s_0))$ *describe whether the individual catches the flu (edge* $e(s_0, s_1)$*) or not (edge* $e(s_0, s_2)$*), while the floret of vertex* s_1 *describes, having caught the flu, whether the individual takes the treatment (edge* $e(s_1, s_3)$*) or not (edge* $e(s_1, s_4)$*).*

From the above example we can observe that each path within an infinite event tree is a sequence through time. To embellish this event tree into a probability tree we need to elicit the conditional probability vectors (CPVs) associated with each floret $\mathcal{F}(s_i)$. This is given by

$$\pi_{s_i} = (\pi_{s_i 1}, \pi_{s_i 2}, \ldots, \pi_{s_i m_i}), \tag{3}$$

where $\pi_{s_i j} = P(e_{s_i j} | s_i)$ is the probability that the unit transitions from s_i along the j_{th} edge, and $\sum_{j=1}^{m_i} \pi_{s_i j} = 1$.

Collections of conditional independence (or Markovian) assumptions are intrinsic to most graphical models. For an event tree these ideas can be captured by coloring the vertices and edges of the tree as discussed for CEGs in Section 2.3. This idea immediately extends to this class of infinite trees.

Recall from Section 2.3 that a situation identifies a unique point in the development of a unit over a particular process. Situations that have identical conditional probabilities associated with their immediately subsequent events are colored the same and are said to be in the same stage. Situations whose unfolding subtrees are topologically and probabilistically equivalent constitute

a position. Observe that situations in the same position are always in the same stage but the inverse does not necessarily hold. This happens because stages impose probabilistic and topological equivalences for only one step ahead. In contrast position implies that these equivalences hold for the whole set of subsequent unfoldings along the event tree. Therefore if we have two distinct situations in the same stage but in different positions, they will be represented by two different vertices with the same color.

Now call U the stage partition of $\mathcal{T}$ and define the conditional probability vector (CPV) on stage u to be

$$\pi_u = (\pi_{u1}, \pi_{u2}, \ldots, \pi_{um_u}), \tag{4}$$

where u has m_u emanating edges. If U is the trivial partition, such that every situation is in a different stage, then the coloring contains no additional information about the process that is not contained in $\mathcal{T}$.

As above we can further define a CPV on each position:

$$\pi_w = (\pi_{w1}, \pi_{w2}, \ldots, \pi_{wm_w}). \tag{5}$$

Surprisingly, the positions of an infinite tree $\mathcal{T}$ are sometimes associated with a coarser partition of its situations than a finite subtree of $\mathcal{T}$ with the same root. This is because in an infinite tree two situations lying on the same directed path from the root can be in the same position. This is impossible for two situations s_i, s_k in a finite tree: the tree rooted at a vertex further up a path must necessarily have fewer vertices than the one closer to the root, so in particular no isomorphism between $\mathcal{T}(s_i)$ and $\mathcal{T}(s_k)$ can exist. We give examples below which explicate this phenomenon.

Note that, we would normally plan to elicit the *structural equivalences* of the model – here the topology of the tree and stage structure associated with its coloring – *before* we elicit the associated conditional probability tables. This would then allow the early interrogation and adoption of the qualitative features of an elicited model before enhancing it with supporting probabilities. These structural relationships can be evocatively and formally represented through the graph of the CEG and DCEG. In particular this graph can be used to explore and critique the logical consequences of the elicited qualitative structure of the underlying process before the often time consuming task of quantifying the structure with specific probability tables.

Example 2 (continued)**.** *In the flu example we may have the staged tree as given in Figure 6. Hence we assume that the probability of catching flu does not change over the months and does not depend on whether flu has been caught before. This implies that* s_0*,* s_2*,* s_5*,* s_8 *and* s_{12} *are in the same stage, as well as all subsequent situations describing this event, which are not represented in Figure 6. Similarly,* s_1 *and* s_{11} *are in the same stage, such that whether the antiviral medication is taken or not is also independent of the number of months until the individual catches flu and independent of flu having been caught before. We further assume that the probability of the individual returning to his normal*

life after recovery is the same when he recovers after treatment as when he successfully recovers without treatment. This means that s_3 and s_7, as well as all other situations representing the probability of returning to a normal life after recovery, are in the same stage. It can be seen from the staged tree that, in this example, whenever two situations are in the same stage, they are also in the same position as their subtrees have the same topology and the same coloring of their situations. Note that in this example the stage partition and the position partition of the situations coincides. Hence our stage and position partition is as follows:

$$w_0 = u_0 = \{s_0, s_2, s_5, s_8, s_{12} \ldots\}, w_1 = u_1 = \{s_1, s_{11}, \ldots\},$$
$$w_2 = u_2 = \{s_3, s_7, \ldots\}, w_3 = u_3 = \{s_4, \ldots\}. \tag{6}$$

Not all paths in the tree are infinite and hence a set of leaf vertices, $\{l_6, l_9, l_{10}, \ldots\}$, exists.

3.2. Dynamic Chain Event Graphs

From the definition of a position, w, given a unit lies in w, any information about how that unit arrived at w is irrelevant for predictions about its future development. As for the CEG, the positions therefore become the vertices of the new graph, the DCEG, which we use as a framework to support inference. Further, colors represent probabilistic symmetries between positions in the same stage. Figure 7 depicts the DCEG corresponding to the staged tree shown in Figure 6 above.

We can now define the DCEG, which depicts a staged tree (see Definition 8 in Section 2.3) in a way analogous to the way the CEG represents structural equivalences.

Definition 11. *Dynamic Chain Event Graph*

A Dynamic Chain Event Graph *(DCEG) $\mathcal{D} = (V(\mathcal{D}), E(\mathcal{D}))$ of a staged tree $\mathcal{T}$ is a directed colored graph with vertex set $V(\mathcal{D}) = W$, the set of positions of the staged tree $\mathcal{T}$, together with a single sink vertex, w_∞, comprising the leaf nodes of T, if these exist. The edge set $E(\mathcal{D})$ is given as follows: Let $v \in w$ be a single representative vertex of the position w. Then there is an edge from w to a position $w' \in W$ for each child $v' \in ch(v), v' \in w'$ in the tree $\mathcal{T}$. When two positions are also in the same stage then they are colored in the same color as the corresponding vertices in the tree $\mathcal{T}$.*

We call the DCEG *simple* if the staged tree $\mathcal{T}$ is such that the set of positions equals the number of stages, $W = U$, and it is then uncolored.

A DCEG is actually obtained from the staged tree by edge contraction operations. Observe also that if two situations are in the same position w there is a bijection between their corresponding florets. Thus we can take any vertex in w to represent it.

Note that the DCEG class extends the CEG models since it could in principle have an infinite number of distinct vertices. When a tree is finite, a CEG is

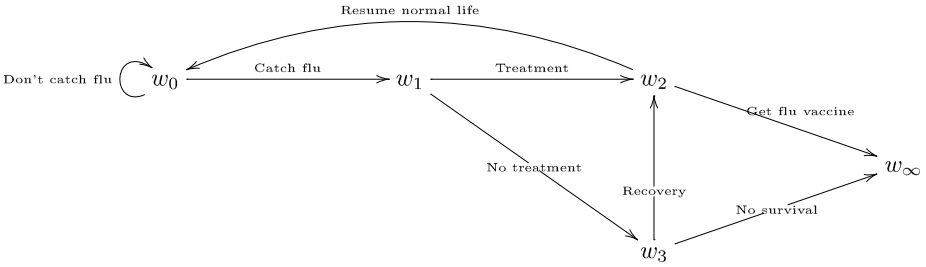


FIG 7. *Flu Example: DCEG of the infinite staged tree.*

actually a DCEG. However the CEG is always acyclic, whilst a DCEG can exhibit cycles – self-loops or loops across several vertices – when it has an infinite number of atoms but a finite graph. In this case a cycle represents a subprocess whose unfolding structure is unchangeable over time. We illustrate below that in many applications the number of positions of a staged tree is finite even though the tree's vertex set is infinite. When this is the case the DCEG is a finite graph and therefore provides a succinct picture of the structural relationships in the process.

Example 2 (continued)**.** *Figure 7 shows the corresponding DCEG of the staged tree given in Figure 6 with $V(\mathcal{D})$ given in Equation 6. The loop from w_0 into itself illustrates that every month the individual could remain well and not catch flu. Alternatively, the individual may move to w_1 at some point, meaning that he has caught flu. In this case he can recover either by getting treatment ($w_1 \to w_2$) or recover on his own ($w_1 \to w_3 \to w_2$). Having recovered the individual either decides to take a flu vaccine to avoid getting flu again ($w_2 \to w_\infty$) or to simply resume his normal life and risk getting flu again ($w_2 \to w_0$). Finally, when not taking treatment, the individual may not recover, and hence move from w_3 to w_∞. Here the position w_∞ can be interpreted as representing stopped subprocesses whose two different stopping reasons are indicated by the labels ("Get flu vaccine" or "No survival") of its incident edges.*

3.3. DCEGs with holding time distributions

Given the graph of a DCEG we can trace the possible paths a unit may take and the associated events that may occur across time. So far we have implicitly assumed that we have regular steps such as days or months. For instance in the DCEG of the flu example (Figure 7), every month the individual is at risk of catching flu: If he catches flu, he traverses through the rest of the DCEG ending up either at w_∞ or back at w_0; if not he loops back directly to w_0. In this case the time an individual stays in a particular position simply follows a geometric distribution, where the probability that an individual stays in position w for k time steps is equal to $P[e(w,w)|w]^k \times [1 - P\{e(w,w)|w\}]$. Further, it has been assumed that once an individual catches flu, only the events of taking

treatment, recovering, and receiving a vaccine are recorded and not the time until these events occur. These could, for example, be recorded retrospectively when measurements are taken a month later. Therefore, here the holding time distributions on a position without a loop into itself lose information.

However, in many cases our process is unlikely to be governed by regular time steps and it is much more natural to think of the time steps to be event driven. A process like this is naturally represented within a tree and hence a DCEG: when moving from one position to another the unit transitions away from a particular state into a different state associated with a new probability distribution of what will happen next. For example, the individual may not record whether he catches flu or not every month but instead monitor the time spent at w_0 not catching flu, until one day he falls ill. Similarly, the time until seeing the doctor for treatment or the time until recovery may be of different lengths and so he spends different amounts of time at each position in the DCEG. Motivated by this irregularity of events, we look at processes in which a unit stays a particular time at one vertex of the infinite tree and then moves along an edge to another vertex. We hence define in this section a generalization of the DCEG, called the Extended DCEG, which attaches a conditional holding time distribution to each edge in the DCEG.

We call the time a unit stays in a situation s_i the holding time H_{s_i} associated with this situation. We can further also define the conditional holding times associated with each edge $e_{s_i j}, j = 1, \ldots, m_i$ in the tree, denoted by $H_{s_i j}$. This describes the time a unit stays at a situation s_i given that he moves along the edge $e_{s_i j}$ next. Analogously to this we can further define holding times on the positions in the associated DCEG: We let H_w be the random variable describing the holding time on position $w \in W$ in the DCEG and $H_{wj}, j = 1, \ldots, m_w$ the random variable describing the conditional time on w given the unit moves along the edge e_{wj} next.

In this paper we assume that all DCEGs are *time-homogeneous*. This means that the conditional holding time distributions for two situations are the same whenever they are in the same stage u. Hence, given the identity of the stage reached, the holding times are independent of the path taken. We denote the random variable of the conditional holding time associated with each stage by $H_{uj}, j = 1, \ldots, m_u$. Time-homogeneity then implies that when two situations are in the same stage u then their conditional holding time distributions are also the same. We note that a unit may spend a certain amount of time in position $w \in u$ before moving along the j_{th} edge to a position $w^{'}$ which is in the same stage. So a unit may make a transition into a different position but arrive at the same stage.

We further assume throughout that the conditional probabilities of going along a particular edge after reaching a stage, do not vary with previous holding times. In the flu example this would mean that the time until catching flu does not effect the probability of taking treatment and the probability of recovery without treatment. Similarly, the holding times are assumed to be independent of previous holding times. So, for example, the time until recovery is independent of the time to catching flu. Contexts where the holding time

distribution may affect the transition probabilities and future holding times can provide an interesting extension to the DCEG and will be discussed in a later paper. Under these assumptions an *Extended DCEG* is defined below.

Definition 12. *Extended DCEG*

An Extended DCEG $\mathcal{D} = (V(\mathcal{D}), E(\mathcal{D}))$ *is a DCEG with no loops from a position into itself and with conditional holding time distributions conditioned on the current stage, u, and the next edge, e_{uj}, to be passed through:*

$$F_{uj}(h) = P(H_{uj} \leq h|u, e_{uj}), h \geq 0, \forall u \in U, j = 1, \ldots m_u. \tag{7}$$

Hence $F_{uj}(h)$ describes the time a unit stays in any position w merged into stage u before moving along the next edge e_{wj}.

Consequently, given a position $w \in W(\mathcal{D})$ is reached, the joint probability of staying at this position for a time less than or equal to h and then moving along the jth edge is

$$P(H_{wj} \leq h, e_{wj}|w) = P(H_{wj} \leq h|w, e_{wj})P(e_{wj}|w) = F_{uj}(h)\pi_{uj}, w \in u. \tag{8}$$

Finally, the joint density of e_{wj} and h is

$$p(e_{wj}, h|w) = \pi_{uj} f_{uj}(h),$$

where f_{uj} is the pdf or pmf of the holding time at stage u going along edge $e_{wj}, w \in u$ next.

An Extended DCEG with stage partition U is hence fully specified by its set of conditional holding time distributions $\{F_{uj}(.) : u \in U\}$ and its collection of CPVs $\{\pi_u : u \in U\}$. Note that it is simple to embed holding times into the staged tree and into the DCEG. Example 2 below discusses this issue in terms of qualitative and graphical modelling without using any specific holding time distribution. Observe also that an Extended DCEG differs from a DCEG in that the transition time between two positions depends on the initial and terminal stages. This fact links Extended DCEGs to semi-Markov processes.

Example 2 (continued)**.** *Return again to the flu example from Section 3.1 with a slightly different infinite tree given in Figure 8. Instead of measuring every month whether the individual catches flu, the individual will spend a certain amount of time at s_0 before moving along the tree. Hence the second edge emanating from s_0 in Figure 6 and its entire subtree have been removed. As before, it is assumed that the probability of catching flu and the decision to take treatment does not depend on whether the flu has been caught before. Also, recovery with or without treatment is assumed not to affect the probability of receiving a vaccine. The corresponding Extended DCEG is given in Figure 9 with positions given by*

$$w_0 = \{s_0, s_4, s_7, \ldots\}, w_1 = \{s_1, s_{10}, s_{11}, \ldots\},$$
$$w_2 = \{s_2, s_6, \ldots\}, w_3 = \{s_3, \ldots\}, w_\infty = \{l_5, l_8, l_9, \ldots\}.$$

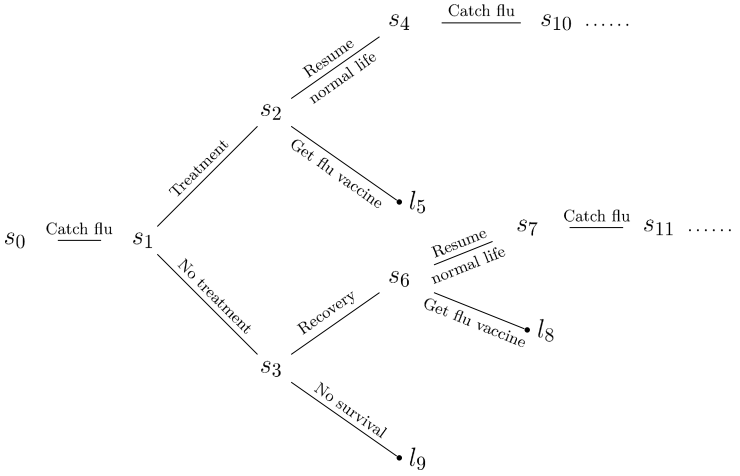


FIG 8. *Variant of flu example: infinite tree* $\mathcal{T}^*$.

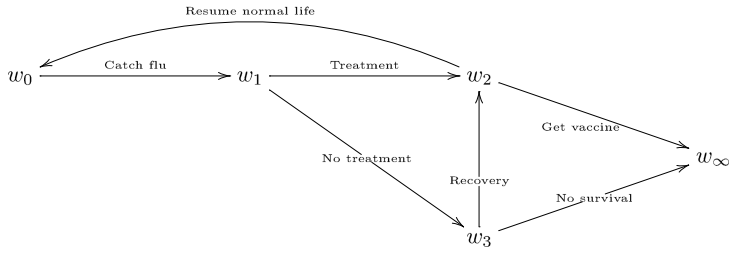


FIG 9. *Variant of the flu example: Extended DCEG of the infinite staged tree.*

In comparison to Figure 7 the loop from w_0 *into itself has been removed. Instead the time spent at* w_0 *is described by the holding time at position* w_0*. Similarly, the time until treatment is taken or not, the time until recovery or death and the time to receiving the flu vaccine or not are of interest and holding time distributions can be defined on these.*

3.4. The repeating time-slice DCEG

Now we can define a useful DCEG class, called the repeating time-slice DCEG (RT-DCEG), whose graph is composed by two different time-slice finite subgraphs.

Definition 13. *Repeating Time-Slice DCEG*

Consider a discrete-time process on $I = \{t_0, t_1, t_2, \ldots\}$ *characterised by a finite collection of variables* $\{Z_p, p = 1, 2, \ldots, P\}$ *where the index* p *defines the same unfolding variable order for each time slice* $t \in I$*. Denote by* $Z_{p,t}$ *the variable* Z_p *in the time slice* t *and assume that all situations corresponding*

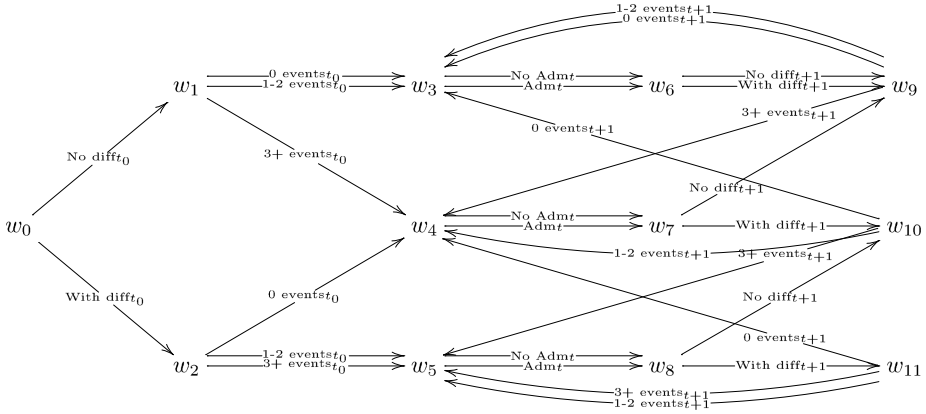


FIG 10. *Repeating Time-Slice DCEG.*

to a variable $Z_{p,t}$ define the level $\mathcal{L}_{p,t}$ in the corresponding event tree. Denote also by $\{s_{p_l,t_0}\}$ and $\{s_{p_l,t_1}\}$ the set of situations associated to the last variables of time-slices t_0 and t_1, respectively. We have a Repeating Time-Slice DCEG *(RT-DCEG) when all previous conditions are valid and there is a surjection map $\Upsilon : \{s_{p_l,t_0}\} \to \{s_{p_l,t_1}\}$ such that $\Upsilon(s_{p_l,t_0})$ is in the same position as s_{p_l,t_0} for all s_{p_l,t_0}.*

The main characteristic of the RT-DCEG topology is that in the end of the second time slice the edges loop back to the end of the first time slice (see Figure 10). Note that the levels $\mathcal{L}_{p,t}, t = 1, 2, \ldots$, correspond to the same variable Z_p. We will now illustrate a RT-DCEG modelling with a real-world example.

Example 3. *We here consider a small subset of the Christchurch Health and Development Study, previously analysed in Fergusson et al. (1986); Barclay et al. (2013). This study followed around 1000 children and collected yearly information about their family history over the first five years of the children's life. We here consider only the relationships of the following variables given below.*

- *Financial difficulty – a binary variable, describing whether the family is likely to have financial difficulties or not,*
- *Number of life events – a categorical variables distinguishing between 0, $1-2$ and ≥ 3 life events (e.g. moving house, husband changing job, death of a relative) that a family may experience in one year,*
- *Hospital admission – a binary variable, describing whether the child is admitted to hospital or not.*

In this setting each time slice corresponds to a year of a child's life starting from when the child is one year old, $t_0 = 1$. A plausible RT-DCEG could be the one given in Figure 10. Note that this RT-DCEG assumes that whether the individual is admitted to hospital or not does not affect the subsequent variables. This is evident from the double arrows from w_3 to w_6, w_4 to w_7 and w_5

to w_8. Also, observe that the variable describing the hospital admission is not included at time $t = 0$, as it does not provide additional information under this assumption.

We start at w_0 in order to follow the path an individual might take through the DCEG across time. The first part of the graph describes the initial CPVs at time t_0. It is first resolved whether or not the family has financial difficulties ($w_0 \to w_1$, $w_0 \to w_2$) and whether the individual experiences 0, $1-2$ or ≥ 3 life events during this year ($w_1 \to w_3$,$w_1 \to w_4$, $w_2 \to w_4$, $w_2 \to w_5$). She then reaches one of the three positions w_3, w_4 and w_5 describing a 'health state' the individual is in before a hospital admission may occur. Independent of whether an admission has occurred or not (w_6, w_7, w_8) she then moves to positions that describe the same three health states. Then, given the individual is in one of the three health states (w_3, w_4, w_5) at time t, for $t \geq t_1$, she traverses through the graph in the following year according to the financial difficulty and number of life events in year $t+1$ and ends up in one of the three previous health states again.

Note that the positions of the RT-DCEG encode the entire history of a unit and we can trace back the full path a unit has taken through the graph. This is a property inherited from the event tree that supports the RT-DCEG graph. For instance, in Example 3 the probability of an individual having a hospital admission at time t is given by $P(Adm = 1|w_i) = \pi_{w_i}$, $i = 3, 4, 5$. It therefore depends on the position where the individual is located at time t. These positions are reached depending on the number of life events and the financial difficulty in that year and the health state of the previous year, which is again determined by the financial difficulty and the number of life events of the year before.

4. Bayesian learning of the parameters of an extended DCEG

In this section we present the learning process of a finite Extended DCEG which extends those for the CEG and is closely related to the learning framework for CTBNs proposed by Nodelman et al. (2003). Conjugate learning in CEGs is now well documented (Smith (2010); Freeman and Smith (2011a)), where the developed methods resemble also the ones used for discrete BN learning – see Korb and Nicholson (2004); Neapolitan (2004); Cowell et al. (2007); Heckerman (2008).

Here we consider only conditional holding time distributions F_{uj} parametrised by a one-dimensional parameter λ_{uj}. Assuming random sampling and prior independence of the vector $\boldsymbol{\pi}$ of all stage parameters and the vector $\boldsymbol{\lambda}$ of different holding time parameters, we can show that the posterior joint density of $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ is given by:

$$p(\boldsymbol{\pi}, \boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D}) = p_1(\boldsymbol{\pi}|\mathbf{N}, \mathcal{D}) p_2(\boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D}), \tag{9}$$

where $\mathbf{h}$ and $\mathbf{N}$ are, respectively, the vector of holding times associated with each stage and the vector of the number of times each edge is taken in the sample;

and $p_1(\boldsymbol{\pi}|\mathbf{N}, \mathcal{D})$ and $p_2(\boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D})$ are the posterior distributions of parameters $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$, respectively. See Appendix A for more details.

Equation 9 makes sure that the parameters $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ can be updated independently. The Extended DCEG learning can then be divided in two distinct steps:

1. learning the stage parameters $\boldsymbol{\pi}$; and
2. learning the holding time parameters $\boldsymbol{\lambda}$.

Learning the posterior $p_1(\boldsymbol{\pi}|\mathbf{N}, \mathcal{D})$ therefore proceeds exactly analogously to learning within the standard CEG. Thus assuming local and global independence of stage parameters $\boldsymbol{\pi}$ and random sampling – these conditions are also assumed for conjugate learning of BNs – Freeman and Smith (2011a) show that with an appropriate characterisation each stage must have a Dirichlet distribution a priori and a posteriori. Here we also assume these conditions to update the stage parameters $\boldsymbol{\pi}$ in a DCEG model. It then follows that

$$\boldsymbol{\pi}_u \sim Dir(\alpha_{u1}, \ldots, \alpha_{um_u}) \tag{10}$$

and

$$\boldsymbol{\pi}_u|\mathbf{N}, \mathcal{D} \sim Dir(\alpha_{u1} + N_{u1}, \ldots, \alpha_{um_u} + N_{um_u}), \tag{11}$$

where α_{uj} is the hyperparameter of the prior distribution associated with an edge j of stage u and N_{uj} is the number of times an edge j of stage u is taken.

As with all Bayesian learning some care needs to be taken in the setting of the hyperparameter values $\boldsymbol{\alpha}_u$. In the simplest case we assume that the paths taken on the associated infinite tree are a priori equally likely. We then specify the hyperparameters associated with each floret accordingly. Given that the Extended DCEG has an absorbing position w_∞ we can find, under the above assumptions, the $\boldsymbol{\alpha}_u, u \in U$ of the Extended DCEG structure $\mathcal{D}$ derived from the infinite tree by simply summing the hyperparameters of the situations merged. This direct analogue to Freeman and Smith (2011a) does not however work when no absorbing position exists, for then these sums diverge. Hence we need to take a slightly different approach. There are many possible solutions. Here we will adapt the concept of 'equilibrium' in Markov chains and thus make the simplest assumption that our prior beliefs with respect to the dynamic stages are 'in equilibrium' (see the discussion of Example 3 below. In other words, our prior beliefs are assumed to be the stationary distribution of the stage transition matrix that assigns the same probability to any path in the corresponding DCEG.

Note that when the holding time distributions are identical across the model space an Extended DCEG is indeed a DCEG and thus it is only necessary to learn the stage parameters $\boldsymbol{\pi}$. To compare two different models we can then use the log Bayes Factor. We illustrate below how we can update the CPVs in a DCEG using the Christchurch example (Section 3.4).

Example 3 (continued)**.** *Take the RT-DCEG depicted in Figure 10 (Section 3.4). Note that again the stages and positions of the graph coincide and hence learning the stage parameters is equivalent to learning the position parameters of the graph. To specify the stage priors, we determine the hyperparameters* α_u

TABLE 1
Prior CPVs and data associated with each position

Position	Prior	Data
w_0	$Dir(\frac{5}{2}, \frac{1}{2})$	$(873, 189)$
w_1	$Dir(\frac{5}{6}, \frac{5}{6}, \frac{5}{6})$	$(135, 436, 302)$
w_2	$Dir(\frac{1}{6}, \frac{1}{6}, \frac{1}{6})$	$(9, 56, 124)$
w_3	$Dir(\frac{5}{6}, \frac{5}{6})$	$(1735, 122)$
w_4	$Dir(\frac{1}{2}, \frac{1}{2})$	$(766, 98)$
w_5	$Dir(\frac{1}{6}, \frac{1}{6})$	$(406, 59)$
w_6	$Dir(\frac{5}{6}, \frac{5}{6})$	$(1679, 178)$
w_7	$Dir(\frac{1}{2}, \frac{1}{2})$	$(700, 164)$
w_8	$Dir(\frac{1}{6}, \frac{1}{6})$	$(227, 238)$
w_9	$Dir(\frac{13}{18}, \frac{13}{18}, \frac{13}{18})$	$(616, 1323, 618)$
w_{10}	$Dir(\frac{2}{9}, \frac{2}{9}, \frac{2}{9})$	$(28, 180, 183)$
w_{11}	$Dir(\frac{1}{18}, \frac{1}{18}, \frac{1}{18})$	$(15, 74, 149)$

of the Dirichlet distribution associated with each stage u as suggested above as follows: We first find the limiting distribution of the Markov process with state space W=$\{w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}\}$ and with the following transition probability matrix that assumes all paths in the graph are equally likely:

$$P = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

So for example, the transition probability from position w_9 is 2/3 to position w_3, 1/3 to position w_4 and 0 to any other position. Solving the general balance equations we then deduce that $P(W = w_3) = P(W = w_6) = \frac{5}{27}$, $P(W = w_4) = P(W = w_7) = \frac{1}{9}$, $P(W = w_5) = P(W = w_8) = \frac{1}{27}$, $P(W = w_9) = \frac{13}{54}$, $P(W = w_{10}) = \frac{2}{27}$, $P(W = w_{11}) = \frac{1}{54}$. This limiting distribution together with an equivalent sample size of 3 (equal to the largest number of categories a variable of the problem takes Neapolitan (2004)) determines the strength of the prior on each stage. Therefore the strength of stages in the same level has to sum up 3. Here this implies that we need to multiply the limiting distribution associated with each stage by 9 to obtain its corresponding strength. Further, assuming that the probabilities on the edges emanating from each position are uniform we can deduce the stage priors to be as given in Table 1.

We can now update these priors separately and in closed form for each stage using the data. The data set has 1062 children born in Christchurch, New Zealand, for the first $2 - 5$ years of their lives. We use the data from

TABLE 2
Posterior CPVs and 95% credible intervals

Position	Posterior	Mean (95% credible interval)		
w_0	$Dir(875\frac{1}{2}, 189\frac{1}{2})$	$0.82(0.80, 0.84)$	$0.18(0.16, 0.20)$	
w_1	$Dir(135\frac{5}{6}, 436\frac{5}{6}, 302\frac{5}{6})$	$0.15(0.13, 0.18)$	$0.50(0.47, 0.53)$	$0.35(0.31, 0.38)$
w_2	$Dir(9\frac{1}{6}, 56\frac{1}{6}, 124\frac{1}{6})$	$0.05(0.02, 0.08)$	$0.30(0.23, 0.36)$	$0.65(0.58, 0.72)$
w_3	$Dir(1735\frac{5}{6}, 122\frac{5}{6})$	$0.93(0.92, 0.94)$	$0.07(0.06, 0.08)$	
w_4	$Dir(766\frac{1}{2}, 98\frac{1}{2})$	$0.89(0.86, 0.91)$	$0.11(0.09, 0.14)$	
w_5	$Dir(406\frac{1}{6}, 59\frac{1}{6})$	$0.87(0.84, 0.90)$	$0.13(0.10, 0.16)$	
w_6	$Dir(1679\frac{5}{6}, 178\frac{5}{6})$	$0.90(0.89, 0.92)$	$0.10(0.08, 0.11)$	
w_7	$Dir(700\frac{1}{2}, 164\frac{1}{2})$	$0.81(0.78, 0.84)$	$0.19(0.16, 0.22)$	
w_8	$Dir(227\frac{1}{6}, 238\frac{1}{6})$	$0.49(0.44, 0.53)$	$0.51(0.47, 0.56)$	
w_9	$Dir(616\frac{13}{18}, 1323\frac{13}{18}, 618\frac{13}{18})$	$0.24(0.22, 0.26)$	$0.52(0.50, 0.54)$	$0.24(0.23, 0.26)$
w_{10}	$Dir(28\frac{2}{9}, 180\frac{2}{9}, 183\frac{2}{9})$	$0.07(0.05, 0.10)$	$0.46(0.41, 0.51)$	$0.47(0.42, 0.52)$
w_{11}	$Dir(15\frac{1}{18}, 74\frac{1}{18}, 149\frac{1}{18})$	$0.06(0.04, 0.10)$	$0.31(0.25, 0.37)$	$0.63(0.56, 0.69)$

year 2 *to update the initial positions* w_0, w_1 *and* w_2 *and then use the hospital admissions variable of year* 2, *as well as years* $3-5$, *to update the remaining CPVs. The data available for each position is presented in Table 1.*

Doing so we obtain the posterior distributions associated with each stage given in Table 2. We also present their corresponding means and 95% *credible intervals. Observe that each stage has a Dirichlet posterior distribution whose parameter is obtained by summing up the parameter of its Dirichlet prior distribution and corresponding sample vector.*

Thus, for example, the expected probability of a child being admitted to hospital is 0.07 *given she has reached position* w_3. *This represents three possible developments: i) she was previously in position* w_3 *and had fewer than* 3 *life events in the current year; ii) she was previously in state* w_4 *and then had no financial difficulties and less than* 3 *events in the current year; or iii) she was previously in state* w_4 *and had financial difficulties but no life events in the current year. Similarly, we have that the probabilities of an admission when reaching* w_4 *and* w_5 *are* 0.11 *and* 0.13, *respectively.*

Next we consider the updating of the prior holding time distribution $p(\boldsymbol{\lambda}|\mathcal{D})$ to its posterior distribution $p(\boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D})$ using the holding time component of the likelihood. Here we restrict ourselves to discussing some examples for conjugate learning. An option is to assume that each holding time parameter λ_{uj} has a Weibull distribution $W(\lambda_{uj}, \kappa_{uj})$ with a known κ_{uj}. If we set $\kappa_{uj} = 1$, the parameter λ_{uj} corresponds to the average rate that transitions from a stage u using edge j occur. It corresponds to assuming an exponential distribution for λ_{uj}. We noted that Nodelman et al. (2003) also used an exponential distribution to model the holding time distribution in a CTBN. In this case, it is implicitly hypothesised that these transition events from a stage u using edge j happen at a constant expected rate over time and are mutually exchangeable given a DCEG model. Observe that this implies that the holding time distributions are stationary and conditionally independent given their corresponding

stages. However using a Weibull distribution we are also able to allow for the possibility that the transition average rate varies over time by adjusting the hyperparameter κ_{uj}; for $\kappa_{uj} < 1$ this rate decreases over time and for $\kappa_{uj} > 1$ this rate increases over time. For more detail about the Weibull distribution and its use for Bayesian learning see Johnson et al. (1995).

To learn the parameters of the conditional holding time distributions, the priors on λ_{uj} are assumed to be mutually independent and have inverse-Gamma distributions $IG(\alpha_{uj}, \beta_{uj})$. This enables us to perform conjugate analyses which are analytically tractable. It also allows us to incorporate some background prior domain information. The hyperparameters α_{uj} and β_{uj} have a strict link with the expected mean and variance of the transitions events about which domain experts can provide prior knowledge. Of course in certain contexts the parameter independence assumption a priori may not be appropriate because the transition times are mutually correlated. In these situations it is likely that conjugacy would be lost, requiring other methods such as MCMC to find the corresponding posterior distribution.

Under the assumptions discussed above to obtain a conjugate learning, the posterior of the rate under this model is given by

$$\lambda_{uj}|\mathbf{h}_{uj}, N_{uj}, \mathcal{D} \sim IG(\alpha_{uj} + N_{uj}, \beta_{uj} + \sum_{l=1}^{N_{uj}} (h_{ujl})^{k_{uj}}), \tag{12}$$

where $h_{ujl}, l = 1, \ldots, N_{uj}$ are the conditional holding times for each unit l that emanates from a stage u through an edge j. Example 2 below exemplifies how we can apply this framework to learn the parameters of an Extended DCEG that has a loop and also a sink node w_∞.

Example 2 (continued)**.** *Recall again the Extended DCEG of the flu example given in Figure 9. To first set up the Dirichlet priors on* $\boldsymbol{\pi}_u$ *and the Inverse-Gamma priors on* λ_{uj} *we again assume an uninformative prior on the paths of the associated tree. Since the equivalent sample size has to be greater than* 2.5 *to ensure that the prior Inverse-Gamma distributions have a mean, we chose an equivalent sample size of* 3 *to be only weakly informative. To determine the hyperparameters* $\boldsymbol{\alpha}_u$ *of the Dirichlet priors we can here use the standard approach of summing the hyperparameters of the situations in each stage, as, due to the sink node* w_∞*, the sum will not diverge as in the previous example.*

Recall from Equation 9 that, for example, $u_1 = \{s_1, s_{10}, s_{11}, \ldots\}$*. Then under the above assumptions and the tree structure in Figure 8 the situations in* u_1 *have the distributions:* $v_1 \sim Dir(1.5, 1.5), v_{11} \sim Dir(1.5\rho_1, 1.5\rho_1), v_{12} \sim Dir(1.5\rho_2, 1.5\rho_2)$*, where* $\rho_1 = 0.25$ *and* $\rho_2 = 0.125$*. Similarly, the next situations of* u_1 *will have the distributions* $Dir(1.5\rho_1^2, 1.5\rho_1^2)$*,* $Dir(1.5\rho_1\rho_2, 1.5\rho_1\rho_2)$*,* $Dir(1.5\rho_2\rho_1, 1.5\rho_2\rho_1)$*,* $Dir(1.5\rho_2^2, 1.5\rho_2^2)$*, ... The infinite sum of the hyperparameters of these distributions is hence a geometric serie whose initial term is* 3 *and whose rate is equal to* $\rho_1 + \rho_2 = 0.375$*. So we can obtain the hyperparameters of the prior on* u_1 *as the limit of these two series, such that we have* $\pi_{u_1} \sim Dir(4.8, 4.8)$*. The hyperparameters of the remaining priors on* u_2 *and* u_3

TABLE 3
Influenza Example: Prior distributions on CPVs and conditional holding times

Description	Holding time distribution	Prior
Time until catching flu	$H_{u_0 1} \sim Exp(\lambda_{u_0 1})$	$\lambda_{u_0 1} \sim IG(4.8, 3.8)$
Take treatment	$N_{u_1} \sim Mult(\pi_{u_1})$	$\pi_{u_1} \sim Dir(4.8, 4.8)$
Time until recovery with treatment	$H_{u_1 1} \sim Weibull(\lambda_{u_1 1}, K_1)$	$\lambda_{u_1 1}^{K_1} \sim IG(4.8, 3.8)$
Time until decide against treatment	$H_{u_1 2} \sim Exp(\lambda_{u_1 2})$	$\lambda_{u_1 2} \sim IG(4.8, 3.8)$
Recovery	$N_{u_3} \sim Mult(\pi_{u_3})$	$\pi_{u_3} \sim Dir(1.2, 1.2)$
Time until recovery	$H_{u_3 1} \sim Weibull(\lambda_{u_3 1}, K_2)$	$\lambda_{u_3 1}^{K_2} \sim IG(1.2, 0.2)$
Time until death	$H_{u_3 2} \sim Weibull(\lambda_{u_3 2}, K_3)$	$\lambda_{u_3 2}^{K_3} \sim IG(1.2, 0.2)$
Get vaccine	$N_{u_2} \sim Mult(\pi_{u_2})$	$\pi_{u_2} \sim Dir(3.8, 3.8)$
Time until resume normal life	$H_{u_2 1} \sim Exp(\lambda_{u_2 1})$	$\lambda_{u_2 1} \sim IG(3.8, 2.8)$
Time until vaccine taken	$H_{u_2 2} \sim Exp(\lambda_{u_2 2})$	$\lambda_{u_2 2} \sim IG(3.8, 2.8)$

can be found in a similar way. They are given together with the priors of the conditional holding times in Table 3.

In this example, it may further be plausible to assume an exponential distribution on $H_{u_0 1}$, which describes the time until catching flu, with scale parameter $\lambda_{u_0 1}$, the average time until the individual gets ill. Further it could be assumed that $H_{u_1 1}$ has the more general Weibull distribution, with scale parameter $\lambda_{u_1 1}$ and with known shape parameter $k_1 > 1$, describing the time until taking treatment and recovering. As $k_1 > 1$ it is assumed that the recovery rate increases with time. The time until the individual decides not to take the treatment could again be exponentially distributed with scale parameter $\lambda_{u_1 2}$, i.e. it is assumed to occur at a constant rate. Similarly to $H_{u_1 1}$, $H_{u_3 1}$ could also have a Weibull distribution with known shape parameter $k_2 > 1$. In contrast to this, $H_{u_3 2}$ could have a Weibull distribution with scale parameter $\lambda_{u_3 2}$ and known shape parameter $k_3 < 1$ indicating that the death rate decreases with time. The holding times $H_{u_2 1}$ and $H_{u_2 2}$ could again have exponential distributions with parameters $\lambda_{u_2 1}$ and $\lambda_{u_2 2}$ respectively. Here the time until getting the vaccine or resuming a normal life is measured.

If Inverse-Gamma priors on $\lambda_{u_0 1}$, $\lambda_{u_1 1}^{k_1}$, $\lambda_{u_1 2}$, $\lambda_{u_2 1}$, $\lambda_{u_2 2}$, $\lambda_{u_3 1}^{k_2}$ and $\lambda_{u_3 2}^{k_3}$ are assumed, a conjugate analysis as described above can be carried out. The priors can be specified by assuming two conditions: i) a prior mean equal to 1 for all prior holding times; and ii) an equivalent sample size corresponding to the strength of the prior belief on the edge associated with each conditional holding time distribution (see Table 3). Then, given a complete random sample of individuals going through the Extended DCEG for a certain length of time, the number of times, N_{uj}, each edge, e_{uj}, is used can be recorded, as well as the time spent at each position before moving along a particular edge. The prior distributions on $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ could then be updated in closed form by Equations 12 and 11, respectively. The CPVs and expected time spent at each position, before moving along a certain edge, can thus be calculated.

Because the estimation above is in closed form, the corresponding marginal likelihood can easily be computed. Thus, note that the marginal likelihood of

an Extended DCEG structure given a complete random sample $L(\mathcal{D}|\mathbf{h},\mathbf{N})$ separates into two parts – one associated with the stages and another with the holding times:

$$L(\mathcal{D}|\mathbf{h},\mathbf{N}) = L_1(\mathcal{D}|\mathbf{N})L_2(\mathcal{D}|\mathbf{h},\mathbf{N}). \tag{13}$$

Then, the marginal likelihood of an Extended DCEG takes the form:

$$L_1(\mathcal{D}|\mathbf{N}) = \prod_{u\in U} \frac{\Gamma(\sum_{j=1}^{m_u}\alpha_{uj})}{\Gamma(\sum_{j=1}^{m_u}\alpha_u + N_u)} \prod_{j=1}^{m_u} \frac{\Gamma(\alpha_{uj}+N_{uj})}{\Gamma(\alpha_{uj})}. \tag{14}$$

After a little algebra the second component of the marginal likelihood associated with, for example, exponential holding times distributions can be written as:

$$L_2(\mathcal{D}|\mathbf{h},\mathbf{N}) = \prod_{u\in U}\prod_{j=1}^{m_u} \frac{\beta_{uj}^{\alpha_{uj}}}{\Gamma(\alpha_{uj})} \frac{\Gamma(\alpha_{uj}+N_{uj})}{\beta_{uj}+\sum_{l=1}^{N_{uj}} h_{ujl}^{\alpha_{uj}+N_{uj}}}. \tag{15}$$

When the prior distributions on $\boldsymbol{\lambda}$ are the same for all Extended DCEG structures the log marginal likelihood, $log\ L(\mathcal{D}|\mathbf{h},\mathbf{N})$, can be written as a linear function of scores associated with different components of the models. The overall linearity of the score is an important property to be explored to devise clever techniques for traversing the Extended DCEG model space since the size of this space is vast without further constraints.

5. Discussion

In this section we discuss the association between DCEGs and three other dynamic models: DBNs, Markov chains and semi-Markov processes.

5.1. The relationship between a DBN and a DCEG

Here we demonstrate that discrete DBNs (see Section 2.2) constitute a special DCEG class. We then discuss some pros and cons in using one or other model. Smith and Anderson (2008) and Barclay et al. (2013) have shown how a BN can be written as a staged tree and hence as a CEG. This can be simply extended to a dynamic setting and we explain below how a DBN can be represented as an infinite staged tree and therefore as a DCEG. It is also easy to check that many other processes such as dynamic context-specific BNs (Boutilier et al. (1996); Friedman and Goldszmidt (1998)) or dynamic Bayesian multinets (Geiger and Heckerman (1996); Bilmes (2000)) are amenable to this representation. Here to match our methods against the usual formulation of the DBN we are focusing only on the DCEG (Section 3.2), where one-step transitions are known and holding times do not need to be explicitly considered.

Let $\{\mathbf{Z}_t : t \in I\}$ where $I = \{t_0, t_1, t_2, \ldots\}$ be a vector stochastic process. Assume that at each time point t, we have a vector of n_t variables $\mathbf{Z}_t = (Z_{1,t}, \ldots, Z_{n_t,t})$, and that the components $Z_{p,t}, p = 1, \ldots, n_t$ all take a finite

number of values. The variables $\mathbf{Z}_t$ then form a time-slice of the DBN for each time point t. In the most general case, the DBN on $\mathbf{Z}_t$ has an associated infinite acyclic directed graph $\mathcal{G}$ where the component $Z_{p,t}$ of $\mathbf{Z}_t$ has parents

$$pa(Z_{p,t}) = \{Z_{q,s} : s < t, q \in \{1, \ldots, n_s\}\} \cup \{Z_{q,s} : s = t, q \in \{1, \ldots, p-1\}\}.$$

Next we claim that any general DBN can be written as an infinite staged tree. To demonstrate this, we first show how to write the variables of the DBN as an infinite tree. We then define the conditional independence statements of the DBN by coloring the florets in the tree to form a stage partition of the situations.

Reindex the variables as $Z_k = Z_{p,t}, k = 1, 2, 3, \ldots$ so that, whenever $Z_i = Z_{q,s} \in pa(Z_{p,t})$, then the index $i < k$. This will ensure that parent variables come before children variables and time-slices come before each other. There is clearly always such an indexing because of the acyclicity and time element of $\mathcal{G}$. This gives a potential total ordering of the variables in $\{\mathbf{Z}_t : t \in I\}$ from which we choose one. Let $a(Z_k) = \{Z_i : i < k\}$ be the set of antecedents of Z_k for the chosen variable order. Note that $pa(Z_{k,t}) \subseteq a(Z_k)$.

By the assumptions of the ordering the components up to index k can be represented by a finite event tree denoted by $\mathcal{T}_k = (V_k, E_k)$. Recall from Section 3.1 that each floret in the tree can be associated with a random variable Z_i and the edges $e_{ij}, j = 1, \ldots, m_i$ describe the m_i values in the sample space that this random variable can take. Hence the paths in the tree $\mathcal{T}_k$ correspond to the set of all combinations of values that variables Z_k can take. Then a sequential construction of the stochastic process allows us to define a set of trees $\{\mathcal{T}_k\}_{k \geq 1}$, such that $\mathcal{T}_k$ is a subtree of $\mathcal{T}_{k+1}$, recursively as follows:

Let $L_k = V_k \setminus V_{k-1}$ be the set of leaf vertices of $\mathcal{T}_k$. Let also $l_{ki} \in L_k$, $i = 1, 2, \ldots, N_k$ be a single leaf vertex i of $\mathcal{T}_k$ which has N_k leaf vertices.

1. For $k = 1$, let $\mathcal{T}_1$ be the floret, $\mathcal{F}(s_0)$, associated with Z_1 which can take m_1 values. Therefore $V_1 = \{s_0, l_{11}, l_{12}, \ldots, l_{1m_1}\}$ and $E_1 = \{e_{s_0 j} : j = 1, \ldots, m_1\}$. Given $\mathcal{T}_k = (V_k, E_k)$, define the edge set of the tree $\mathcal{T}_{k+1}$ as follows:

$$E_{k+1} = E_k \cup E^+_{k+1},$$

where

$$E^+_{k+1} = \{e_{l_{ki} j} : l_{ki} \in L_k, j = 1, 2, \ldots, m_{k+1}\} \tag{16}$$

is a set of $N_k \times m_{k+1}$ new edges such as m_{k+1} edges emanate from each vertex $l_{ki}, i = 1, 2, \ldots, N_k$. Each of these edge $e_{l_{ki} j}$ describes a specific value that the random variable Z_{k+1} can take. To define the vertex set of $\mathcal{T}_{k+1}$, attach now a new leaf vertex to each of the edges in E^+_{k+1} and let

$$V^+_{k+1} = \{ch(l_{ki}) : l_{ki} \in L_k\} \tag{17}$$

and $V_{k+1} = V_k \cup V^+_{k+1}$.

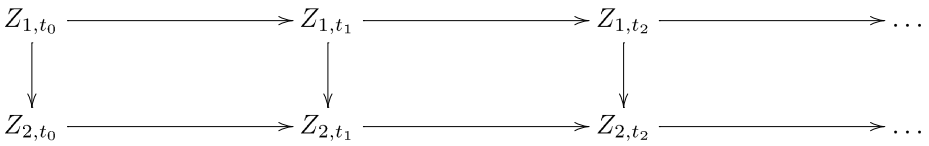


FIG 11. *A simple DBN for Christchurch Example.*

2. The infinite tree $\mathcal{T}$ of this DBN is now simply defined as $\mathcal{T} = (V, E)$, where the vertex and edges sets are, respectively, given by

$$V = \lim_{k\to\infty} V_k \quad \text{and} \quad E = \lim_{k\to\infty} E_k.$$

Note that the infinite length directed paths starting from the root of this tree correspond to the atoms of the sample space of the process.

We demonstrate this recursive construction of the infinite tree below.

Example 4. *Here we remodel the Christchurch example (Section 3.4) where we take only the binary variables Financial Difficulty and Hospital Admission into account. Let $Z_{1,t}$ and $Z_{2,t}$ denote, respectively, the variables Financial Difficulty and Hospital Admission in time-slice t. Suppose now that the financial life enjoyed by a family in time t depends only on its previous financial situation in time $t-1$. Assume also that the probability of a child been admitted in the hospital in time t depends if she visited the hospital in time $t-1$ as well as on the current financial difficulty faced by her family. The DBN given in Figure 11 represents this process over time. Note that this is a 1-Markov BN: a variable is only affected by variables of the previous and current time-slices.*

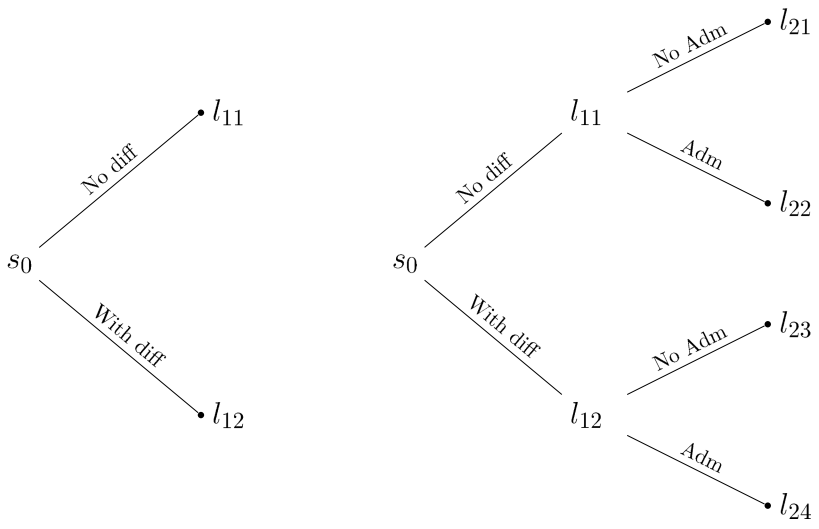
We can now reindex the variables of the DBN as follows:

$$Z_1 = Z_{1,t_0},\ Z_2 = Z_{2,t_0},\ Z_3 = Z_{1,t_1},\ Z_4 = Z_{2,t_1},\ Z_5 = Z_{1,t_2},\ Z_6 = Z_{2,t_2},\ \ldots$$

where Z_i represents a variable Financial Difficulty, if the index i is an odd number, and a variable Hospital Admission, otherwise. Thus for example in the event tree $a(Z_6) = \{Z_1, Z_2, Z_3, Z_4, Z_5\}$ will be the antecedents of the variable hospital admission associated with the third time-slice (Z_{2,t_2}).

Because we have defined $Z_1 = Z_{1,t_0}$, $\mathcal{T}_1$ hence corresponds to the tree given in Figure 12 (a) with root vertex s_0 and two emanating edges labeled No difficulty *and* With difficulty*. To obtain $\mathcal{T}_2$ (Figure 12 (b)) from $\mathcal{T}_1$ attach $m_2 = 2$ edges to each leaf vertex of $\mathcal{T}_1$ as defined by Equation 16 and attach a child to each new edge as defined in Equation 17. Similarly, to obtain $\mathcal{T}_3$ from $\mathcal{T}_2$ attach $m_3 = 2$ edges describing $Z_3 = 0$ and $Z_3 = 1$ to each leaf of $\mathcal{T}_2$ and attach a new leaf to each new edge. Continuing in this way a representation of the infinite tree is provided in Figure 17 (Appendix B), where again the notation '...' describes the continuation of the process.*

We next represent the conditional independencies of the DBN by coloring the vertices and associated edges that are in the same stage as described in



(a) Illustration of $\mathcal{T}_1$ (b) Illustration of $\mathcal{T}_2$

FIG 12. *Illustration of $\mathcal{T}_1$ and $\mathcal{T}_2$ of a DBN.*

Section 3.1. The resulting staged tree then encodes the same conditional independencies as the DBN.

Notice that the vertex $l_{ki} \in V_k \subseteq V$ labels the conditioning history of the variable Z_{k+1} based on the values of its antecedent variables. By the definition of a DBN

$$Z_{k+1} \perp\!\!\!\perp a(Z_{k+1}) | pa(Z_{k+1}), \tag{18}$$

which means that a variable Z_{k+1} is independent of its antecedents given its parents. So by the DCEG definition the leaf nodes l_{ki_1} and l_{ki_2} associated with the event tree $\mathcal{T}_k$ are in the same stage whenever their edge probabilities are the same. More formally

$$P(e_{l_{ki_1}j} | l_{ki_1}) = P(e_{l_{ki_2}j} | l_{ki_2}) \tag{19}$$

for all edges $e_{l_{ki_1}j}$ and $e_{l_{ki_2}j}$, $j = 1, \ldots, m_{k+1}$ or alternatively,

$$P(Z_{k+1} = z_{k+1} | l_{ki_1}) = P(Z_{k+1} = z_{k+1} | l_{ki_2}), \tag{20}$$

where z_{k+1} is a value the variable Z_{k+1} can take. If this is true then we assign the same color to l_{ki_1} as to l_{ki_2}. Thus the corresponding DCEG follows directly from the staged tree by performing edge contraction operations according to the position partition (see Section 3.2).

Example 4 (continued)**.** *Recall the previous Example 4. Assume now that the conditional probability tables remain the same across the time-slices t, $t \geq t_1$, for the Financial Difficulty variable set and the Hospital Admission variable set. Consider also that the probability of hospital admission in a specific time-slice*

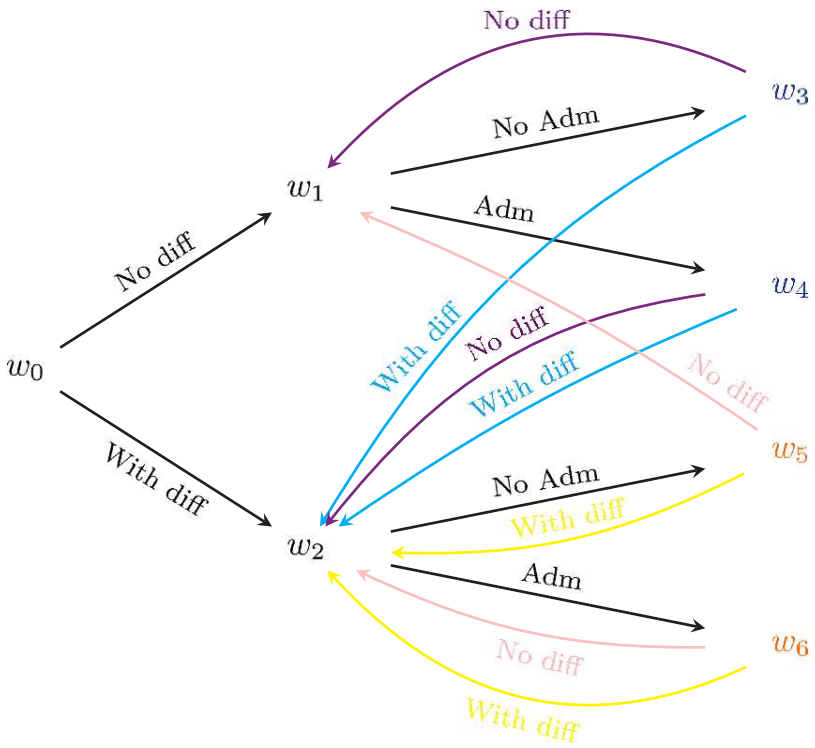


FIG 13. *Illustration of the RT-DCEG of a DBN.*

t, $t \geq t_1$, only changes – in this case positively – if a family currently enjoys a good financial situation and their child was not admitted to the hospital in the previous time-slice. This probability is hypothesised to be equal to the one assigned to a child that lives in a financially stable family in the first time-slice. Appendix B shows the staged tree corresponding to these hypotheses. Note that the colors alternate between odd and even levels because of the invariance of the conditional probability tables over time. Observe that it is not possible to represent these context-specific conditional statements graphically using a DBN model on these variables although they are encoded in the DBN's conditional probabilistic tables. In contrast, these additional conditions are not only directly depicted in a DCEG – which is actually a RT-DCEG – but also the corresponding graph is quite compact and easily interpreted (Figure 13).

Note that the re-expression of the DBN as a staged tree emphasizes how the usual classes of DBNs only represent graphically certain specific families of symmetric conditional independencies. In contrast, the DCEG can allow us to depict asymmetric dependence structures between the variables of a time-slice and also across time-slices. When the dependence structure is defined through symmetric conditional independencies then the DBN is topologically much simpler than

the corresponding DCEG. But when, as is often the case, many combinations of values of states are logically impossible and the number of non-zero probability transitions between states is small then the DCEG depicts these zeros explicitly and can sometimes be topologically simpler than the DBN.

Consider the staged tree of Example 4 (Figure 17, Appendix B). If the conditional probability tables of the BN state that $P(Z_{2,t_0} = \text{Adm}|Z_{1,t_0} = \text{No diff}) = 0$ then the edge describing this probability can be omitted from the tree and the tree is hence reduced to three quarters of its size. Hence unlike the BN and its dynamic analogue, as well as depicting independence relationships the DCEG also allow us to read zeros in the corresponding transition matrix, represented by missing edges in the tree. This is particularly helpful when representing processes which have many logical constraints. However, these gains imply that the DCEG model space scales up super-exponentially with the number of variables. Here the main challenge is to devise clever algorithms to search the DCEG model space.

5.2. DCEG and Markov chain

In this section we use some examples to illustrate some topological links between DCEG graphs and state-transition diagrams of Markov Chains. These connections constitute a promising start pointing to extend many of the well-developed results on Markov processes to the DCEG domain – see, for example, the use of limiting distribution to initialise the DCEG learning process (Section 4). In its turn, the DCEG framework can be used to verify if there is statistical evidence that supports modelling a real-world process as a Markov Chain, and (if there is) to infer its corresponding transition matrix.

Note that the topology of the DCEG graph resembles the familiar state-transition diagram of a Markov process, where the positions of the DCEG can be reinterpreted as states of the Markov process. However, as mentioned at the end of Section 3.1 the DCEG is usually constructed from a *description* of a process as a *staged tree* rather than from a prespecified Markov chain. Thus there are also some differences between the DCEG graph and standard state-transition diagrams such as the one-to-one relationship between the atoms of the space of the DCEG and its paths and its coloring as will be illustrated in the simple examples below.

Example 5. *Example of a Markov Chain I*

Let $\{X_n : n \in \mathbb{N}\}$ *be a discrete-time Markov process on the state space* $\{a, b, c\}$ *with transition matrix* P *given by*

$$P = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.5 & 0.3 & 0.2 \end{pmatrix},$$

and with initial distribution $\alpha = (0.4, 0.4, 0.2)$. *Note that the transition probabilities from states* b *and* c *are the same. The state-transition diagram of the associated Markov process is given in Figure 14.*

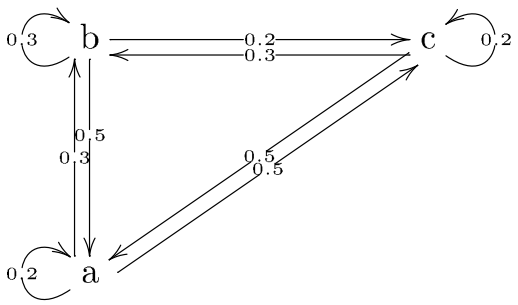


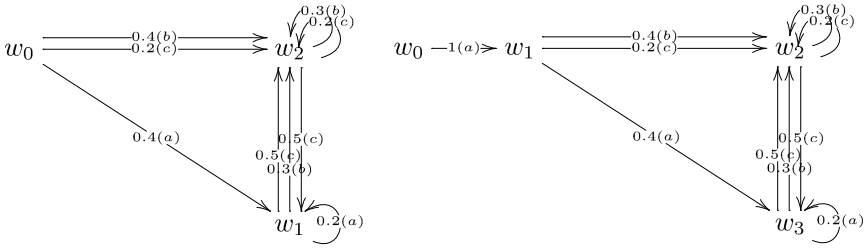
FIG 14. *State-transition diagram of the Markov process in Example 5.*

However, the DCEG representation gives a different structure, which becomes apparent when looking first at the tree representation of the problem. As the process is infinite, the number of situations of the tree is also infinite. The initial situation s_0*, the root of the tree, has emanating edges which represent the choice of initial state with associated CPV* $\pi_{s_0} = (0.4, 0.4, 0.2)$*. The other situations could be indexed as* $\{s_{i,n}, i = a, b, c, n \in \mathbb{N}\}$ *with CPVs* $\pi_{s_a,n} = (0.2, 0.3, 0.5)$ *and* $\pi_{s_b,n} = \pi_{s_c,n} = (0.5, 0.3, 0.2)$*. It is then immediate that the corresponding DCEG only has three stages and positions with the stage and position partition given by*

$$u_0 = w_0 = \{s_0\}, u_1 = w_1 = \{s_{a,n}, n \in \mathbb{N}\}, u_2 = w_2 = \{s_{b,n}, s_{c,n}, n \in \mathbb{N}\}.$$

There is no w_∞ *as all paths are infinite and hence no leaf vertices exist in the tree. The DCEG can then be drawn as given in Figure 15a and the associated CPVs are* $\pi_{w_0} = (0.4, 0.4, 0.2)$*,* $\pi_{w_1} = (0.2, 0.3, 0.5)$ *and* $\pi_{w_2} = (0.5, 0.3, 0.2)$*. For a better comparison the CPVs have here also been attached to the edges of the DCEG. Figure 15b depicts the same process when it has a degenerate initial distribution* $\pi_{s_0} = (1, 0, 0)$*.*

Even here, where the process is initially defined through a transition matrix, the graph of the DCEG automatically identifies states which have equivalent roles: here state b being identified with state c, and illustrates the identical



(a) $\pi_{s_0} = (0.4, 0.4, 0.2)$ (b) $\pi_{s_0} = (1, 0, 0)$

FIG 15. *DCEG representation of the Markov process in Example 4.*

conditional probabilities associated with the two states by putting $s_{b,n}$ and $s_{c,n}$, for $n \in \mathbb{N}$ in the same position w_1 in Figure 15.

The DCEG also depicts explicitly the initial distribution of the process given by the edges emanating from w_0 and acknowledges the initially elicited distinctions of the states b and c through the double edge from w_0 to w_2. Observe also that if a process has a degenerate initial distribution – for example, the one depicted in Figure 15b – the DCEG will show this phenomenon transparently and it only implies minor changes in the DCEG topology.

These topological properties often have important interpretive value, as the DCEG can discover a different partition of the states of a variable or even help to construct new informative variables to represent a problem. Further, any residual coloring, inherited from the staged tree allows us to elaborate the structure of the transitions in a natural and consistent way, highlighting some possible common underlying structures between the states of a Markov process. This can bring new questions and motivate a deeper understanding of the process under analysis. For example, the state-transition diagram and DCEG graph (Figure 16) corresponding to the simple Example 6 are identical except for the colors. By coloring the positions red, the DCEG model stresses that their transition processes are probabilistically identified with each other (i.e. they are in the same stage). In real-world problems if these coloring properties appear in the best scoring model based on the given data set, then these features are explicitly depicted and fed back to domain experts who can then speculate about the possible reasons for their presence.

Example 6. *Example of a Markov Chain II*

A coin is tossed independently, with probability $P(H) = \lambda$ of throwing heads and probability $P(T) = 1 - \lambda = \bar{\lambda}$ of throwing tails. The coin is tossed until N heads have appeared when the game terminates. Its DCEG has thus $N+1$ positions describing whether $0, 1, 2, \ldots, N-1$ or N heads have been tossed and is given in Figure 16.

Notice here that because each toss has the same probability λ of heads the positions $w_0, w_1, w_2, \ldots, w_{N-1}$ are all in the same stage and so its vertices $w_0, w_1, w_2, \ldots, w_{N-1}$ are colored red. If this was a model discovered from observations an expert could immediately deduce that a coin with the same probability of heads was being used at each time.

Being able to embed the state-transition diagrams and to register the entire unfolding process in its coloring and topology, a DCEG model provides a very expressive graphical representation of a stochastic process. After searching a well-fitting DCEG model and learning it, we would then be able to identify

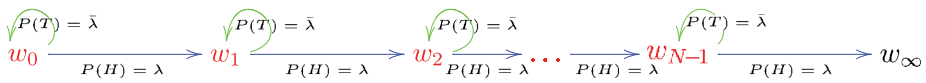


Fig 16. *DCEG representation of coin tossing example.*

if there is at least a subgraph in this DCEG model that represents a Markov process. If this is possible, this will enable us to further explore some asymptotic behaviours of this subprocess using the well-established Markov theory.

5.3. Extended DCEG and semi-Markov process

As there is a connection between a DCEG and Markov processes, an Extended DCEG is closely linked to semi-Markov processes (Barbu and Limnios (2008); Medhi (1994)). These are a generalization of Markov processes that allow for the holding times to have any distribution instead of restricting them to have a geometric distribution (discrete-time Markov processes) or an exponential distribution (continuous-time Markov processes). We recall the definition of a semi-Markov process below:

Definition 14. *Semi-Markov Process (Medhi (1994))*

Let $\{Y_t, t \geq 0\}$ *be a stochastic process with discrete state space and with transitions occurring at times* $t_0, t_1, t_2, \ldots$. *Also, let* $\{X_n, n \in \mathbb{N}\}$ *describe the state of the process at time* t_n *and let* H_n *be the holding time before transition to* X_n. *Hence* $Y_t = X_n$ *on* $t_n \leq t < t_{n+1}$. *If*

$$\begin{aligned} &P(X_{n+1} = j, H_{n+1} \leq t | X_0, X_1, \ldots, X_n, H_1, \ldots, H_n) \\ &\quad = P(X_{n+1} = j, H_{n+1} \leq t | X_n), \end{aligned} \tag{21}$$

then $\{X_n, H_n\}$ *is called a Markov Renewal process and* $\{Y_t, t \geq 0\}$ *a semi-Markov process. Also,* $\{X_n, n \in \mathbb{N}\}$ *is the embedded Markov chain with transition probability matrix* $P = (p_{ij})$, *where* $p_{ij} = P(X_{n+1} = j | X_n = i)$.

A semi-Markov process is usually specified by an initial distribution α and by its semi-Markov kernel Q whose ij^{th} entry is given by

$$Q_{ij}(t) = P(X_{n+1} = j, H_{n+1} \leq t | X_n = i). \tag{22}$$

We assume here that all Markov processes considered are time-homogeneous and hence the above equations do not depend on the index n. In order to illustrate a link between the Extended DCEG and semi-Markov processes we write the semi-Markov kernel as

$$Q_{ij}(t) = p_{ij} F_{ij}(t), \tag{23}$$

where

$$F_{ij}(t) = P(H_{n+1} \leq t | X_{n+1} = j, X_n = i) \tag{24}$$

is the conditional holding time distribution, i.e. the holding time at $X_n = i$ assuming that we move to $X_{n+1} = j$ next and p_{ij} is given in Definition 14. We can then show that a particular subclass of the time-homogeneous Extended DCEG corresponds to a semi-Markov Process.

Theorem 1. *Let an Extended DCEG* $\mathcal{D}$ *with holding times be simple and let no two edges lead from the same parent into the same child. Then this Extended*

DCEG is a semi-Markov process with state space $\mathcal{S} = \{V(\mathcal{D})\backslash w_0\}$ *and with the entries of its transition matrix given by*

$$p_{ij} = \begin{cases} \pi_{w_i j} & : \text{ if } e_{w_i j} = e(w_i, w_j) \text{ exists} \\ 1 & : \text{ if } w_i = w_j = w_\infty \\ 0 & : \text{ otherwise,} \end{cases}$$

and with conditional holding time distributions

$$F_{ij}(t) = \begin{cases} P(H_{w_i j} \leq t | e_{w_i j}, w_i) & : \text{ if } e_{w_i j} = e(w_i, w_j) \text{ exists} \\ 1 & : \text{ if } w_i = w_j = w_\infty \\ 0 & : \text{ otherwise.} \end{cases}$$

If the position w_0 *is a source vertex then the initial distribution is given by* $\alpha = \pi_{w_0}$. *Otherwise the initial distribution assigns probability* 1 *to* w_0 *and* w_0 *is included in the state space.*

Proof. See Appendix C. □

Results such as the one in Theorem 1 allow us to identify particular Extended DCEG subclasses whose models have a strong connection with semi-Markov processes. This can indeed be very useful as many of the well-developed results on Markov processes could be extended to the DCEG. For instance, from Equation 8 the probability of staying at a position w for a time $\leq h$ and then moving along the edge e_{wk} can be calculated. This equation corresponds to the entries of the semi-Markov kernel (Equation 22) of a semi-Markov process. Then, for example, Barbu and Limnios (2008) or Kulkarni (1995) have shown how to derive the transition matrix of the semi-Markov process from the semi-Markov kernel, in order to calculate the probability of being in state j at time t given that we are initially in state i. These types of calculations could be directly extended to the Extended DCEG. This would further enable the Extended DCEG to be applicable to the wide-ranging domain of semi-Markov processes, which includes reliability theory, finance and insurance or traffic modelling.

6. Conclusion

We have demonstrated here that a dynamic version of the CEG is straightforward to develop and that this class enjoys most of the convenient properties of the CEG. It further usefully generalizes the discrete DBN when the context demands it.

Although we do not envisage the DCEG taking over from the DBN as a representational device and framework for learning we nevertheless believe that it provides a valuable alternative tool. It is particularly suited to domains where the levels of state vectors are numerous but the associated transitions are sparse, or when context-specific symmetries abound. The fact that the DCEGs express

DBNs as a special case means that the DCEG and DBN representations are particularly complementary. The first focuses on the micro structure of the transitions between states of the process whilst the other focuses on the macro elements of the relationships between relevant variables within the study domain.

Despite the closed form of their score functions, the major challenge that exists is to develop effective model search algorithms to discover potential causal mechanisms within the DCEG class. Faster and more efficient algorithms are now becoming available for CEG model search (Collazo and Smith (2015)) and the technology is now being transferred to address DCEG model selection. Early results on this topic are promising and will be reported in a later paper.

Appendix A: Parameter independence

Here we show that under some mild conditions the Extended DCEG parameters associated with stages and holding times can be learnt independently. Recall that the Extended DCEG has a set of stages $u \in U$ and that each position in u has m_u edges emanating from it. As defined in Equation 4 we have associated with each stage u a CPV $\boldsymbol{\pi}_u = (\pi_{u1}, \pi_{u2}, \ldots, \pi_{m_u})$ and we denote the concatenation of these vectors of stage parameters by a vector $\boldsymbol{\pi}$. As in Section 3.3, we can further attach a vector of conditional holding time distributions $(F_{u1}, F_{u2}, \ldots, F_{um_u})$ to each stage u with parameters $\boldsymbol{\lambda}_u = (\lambda_1, \lambda_2, \ldots, \lambda_{m_u})$. This could be, for example, a set of exponential holding time distributions with scale parameters λ_u or more general distributions such as th Weibull distribution with scale parameter λ_u and known shape parameter. We call the concatenation of these different holding time parameters $\boldsymbol{\lambda}$.

Given an Extended DCEG $\mathcal{D}$, for each unit that traverses the DCEG, the edges he passes along can be recorded as well as the holding times at each position. Assume the unit ι takes the path $\epsilon^\iota = (e_{w_{i_0}j_0}, e_{w_{i_1}j_1}, \ldots, e_{w_{i_{n_\iota}}j_{n_\iota}})$ along $n_\iota + 1$ edges starting at $w_{i_0} = w_0$. Then, let $w^\iota_{i_a}$ describe the a^{th} position reached by unit ι, $h^\iota_{i_a}$ the holding time at position $w^\iota_{i_a}$ and $e^\iota_{i_a j_a}$ the ath edge passed along, where $a = 0, 1, \ldots, n_\iota$. Then, by the definition of a DCEG (see Definition 12) the likelihood, given a unit ι, with path ϵ^ι and vector of holding times $\mathbf{h}^\iota = (h^\iota_{i_0}, h^\iota_{i_1}, \ldots, h^\iota_{i_n})$, is given by

$$L(\boldsymbol{\pi}, \boldsymbol{\lambda} | \epsilon^\iota, \mathbf{h}^\iota, \mathcal{D}) = \prod_{a=0}^{n_\iota} p(e^\iota_{i_a j_a}, h^\iota_{i_a} | w^\iota_{i_a}) = \prod_{a=0}^{n_\iota} \pi_{w^\iota_{i_a} j_a} f_{w^\iota_{i_a} j_a}(h^\iota_{i_a}). \tag{25}$$

This can now be generalized to a complete random sample $\mathcal{S}$ of n units going through the tree to obtain the likelihood

$$L(\boldsymbol{\pi}, \boldsymbol{\lambda} | \mathcal{S}, \mathcal{D}) = \prod_{\iota=1}^{n} L(\boldsymbol{\pi}, \boldsymbol{\lambda} | \epsilon^\iota, \mathbf{h}^\iota, \mathcal{D}) = \prod_{\iota=1}^{n} \prod_{a=0}^{n_\iota} \pi_{w^\iota_{i_a} j_a} f_{w^\iota_{i_a} j_a}(h^\iota_{i_a}). \tag{26}$$

This likelihood can then be rewritten by counting the number of times the units pass through a position $w \in u$ and go along the jth edge, $j = 1, \ldots, m_u$, which is denoted by N_{uj}. Let $\mathbf{h}_{uj}$ be the vector of conditional holding times for the units who arrive at stage u and move along the jth edge next and let h_{ujl} be the holding time of the lth pass along this edge. Denote the vector of holding times by $\mathbf{h} = \{\mathbf{h}_{uj}, u \in U, j = 1, \ldots, m_u\}$ and the vector of the number of times each edge is taken by $\mathbf{N} = \{N_{uj}, u \in U, j = 1, \ldots, m_u\}$. The likelihood of $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ given a complete random sample and an Extended DCEG $\mathcal{D}$ is therefore given by

$$L(\boldsymbol{\pi}, \boldsymbol{\lambda}|\mathbf{N}, \mathbf{h}, \mathcal{D}) = \prod_{u \in U} \prod_{j=1}^{m_u} \pi_{uj}^{N_{uj}} \prod_{l=1}^{N_{uj}} f_{uj}(h_{ujl}), \tag{27}$$

where the units go N_{uj} times along edges $e_{wj}, w \in u$ each time staying for a time h_{ujl} at the previous position. Then, immediately from Equation 27 the likelihood $L(\boldsymbol{\pi}, \boldsymbol{\lambda}|\mathbf{N}, \mathbf{h}, \mathcal{D})$ of a complete random sample separates. Explicitly, we have that

$$L(\boldsymbol{\pi}, \boldsymbol{\lambda}|\mathbf{N}, \mathbf{h}, \mathcal{D}) = \overbrace{\prod_{u \in U} \prod_{j=1}^{m_u} \pi_{uj}^{N_{uj}}}^{L_1(\boldsymbol{\pi}|\mathbf{N}, \mathcal{D})} \times \overbrace{\prod_{u \in U} \prod_{j=1}^{m_u} \prod_{l=1}^{N_{uj}} f_{uj}(h_{ujl})}^{L_2(\boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D})}. \tag{28}$$

If $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}$ are believed to be a priori independent so that

$$p(\boldsymbol{\pi}, \boldsymbol{\lambda}|\mathcal{D}) = p_1(\boldsymbol{\pi}|\mathcal{D}) p_2(\boldsymbol{\lambda}|\mathcal{D}),$$

then $p_1(\boldsymbol{\pi}|\mathcal{D})$ and $p_2(\boldsymbol{\lambda}|\mathcal{D})$ can be updated independently using $L_1(\boldsymbol{\pi}|\mathbf{N}, \mathcal{D})$ and $L_2(\boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D})$ respectively, to obtain the posterior density

$$p(\boldsymbol{\pi}, \boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D}) = p_1(\boldsymbol{\pi}|\mathbf{N}, \mathcal{D}) p_2(\boldsymbol{\lambda}|\mathbf{h}, \mathbf{N}, \mathcal{D}), \tag{29}$$

which also separates.

B. The staged tree of a DBN – Example 4

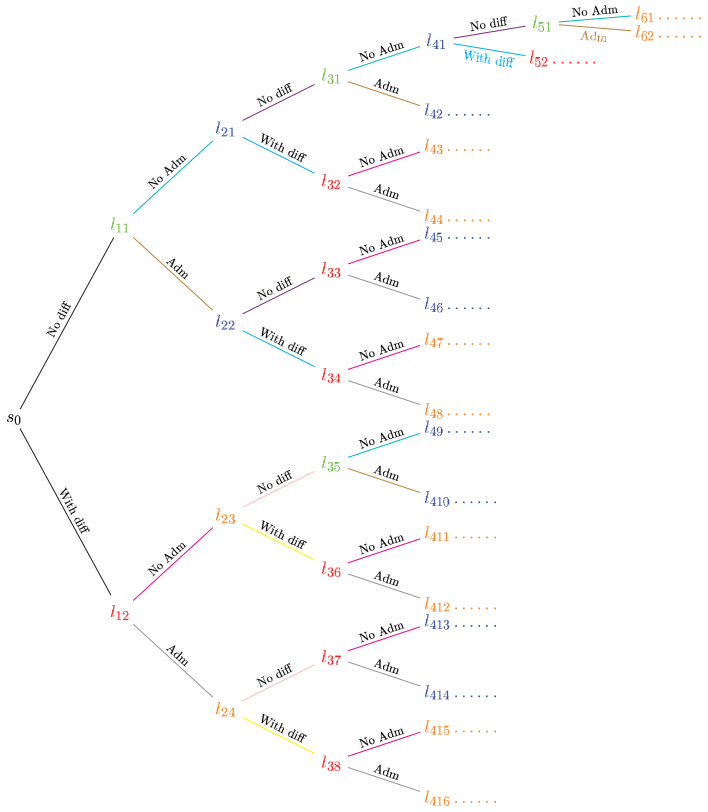


FIG 17. *Illustration of the staged tree of $\mathcal{T}$ of a DBN.*

C. Proof of Theorem 1

The proof of Theorem 1 is given below:

Proof. Assume we have a Extended DCEG $\mathcal{D}$ which is simple and which has no double edges from one vertex into another. To show that this can be written as a semi-Markov process the state space needs to be defined and the semi-Markov kernel and initial distribution need to be specified.

Define the state space of the semi-Markov process and its jump process to be $\mathcal{S} = \{V(\mathcal{D})\backslash w_0\}$, the set of positions not including w_0. As no two edges lead from the same parent into the same child each edge is uniquely determined by the two positions it connects. First consider the case where $w_i \neq w_\infty$ and then the case where $w_i = w_\infty$. Note that not every Extended DCEG will have a final position of leaf vertices, in which case the second case does not apply.

Case 1: $w_i \neq w_\infty$**:** If $e(w_i, w_j)$ exists, then the ij^{th} entry of the transition matrix P of the jump process is given by

$$p_{ij} = P(X_{n+1} = w_j | X_n = w_i) = P(e(w_i, w_j)|w_i).$$

Assuming without loss of generality that the j^{th} edge of w_i leads to w_j, then,

$$\begin{aligned} P(X_{n+1} = w_j | X_n = w_i) &= P(e_{w_i j}|w_i) \\ &= \pi_{w_i j} \\ &= \pi_{u_i j}, \end{aligned}$$

where $u_i = w_i$ as the Extended DCEG is simple. The conditional holding time distributions can be derived in a similar way. Assuming again that the j^{th} edge of w_i leads to w_j

$$\begin{aligned} F_{ij}(t) &= P(H_{n+1} \leq t | X_{n+1} = w_j, X_n = w_i) \\ &= P(H_{w_i j} \leq t | e_{w_i j}, w_i) \\ &= P(H_{u_i j} \leq t), \end{aligned}$$

where $u_i = w_i$ as the Extended DCEG is simple. By Equation 22 the ij^{th} entry of the semi-Markov kernel is then given by $Q_{ij}(t) = p_{ij} F_{ij}(t)$. If $e(w_i, w_j)$ does not exist then the ij^{th} entry of the semi-Markov kernel is zero as no transition from w_i to w_j occurs.

Case 2: $w_i = w_\infty$**:** When $w_i = w_\infty$, then the unit stays in w_∞ forever once reaching this state and hence $Q_{ij}(t) = 1$ when $w_j = w_\infty$ and 0 otherwise.

When w_0 in the Extended DCEG is a source node and no edges lead back to w_0, so that it solely serves as a starting point of the process, then the initial distribution of the corresponding semi-Markov process is given by $\alpha = \pi_{w_0} = \pi_{u_0}$. If w_0 can be reached again throughout, then w_0 is included in the state space and the initial distribution of the semi-Markov process assigns w_0 probability 1. □

Acknowledgements

The authors would like to thank John Horwood and the CHDS research group for providing the data set. The authors would like also to thank the reviewers and editors of the journal for their valuable comments which have helped improve this paper. M. Barclay was funded by the EPSRC. Rodrigo A. Collazo was supported by the Brazilian Navy and CNPq-Brazil [grant number 229058/2013-2].

References

Barbu, V. S. and Limnios, N. *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*, volume 191. Springer, 2008. MR2452304

Barclay, L. M., Hutton, J. L., and Smith, J. Q. Refining a Bayesian Network using a Chain Event Graph. *International Journal of Approximate Reasoning*, 54 (9): 1300–1309, 2013. MR3115418

Barclay, L. M., Hutton, J. L., and Smith, J. Q. Chain Event Graphs for Informed Missingness. *Bayesian Analysis*, 9 (1): 53–76, 2014. MR3188299

Bilmes, J. A. Dynamic Bayesian Multinets. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 38–45. Morgan Kaufmann Publishers Inc., 2000.

Boutilier, C., Friedman, N., Goldszmidt, M., and Koller, D. Context-specific independence in Bayesian Networks. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, 1996*, pages 115–123. Morgan Kaufmann Publishers Inc., 1996. MR1617129

Collazo, R. A. and Smith, J. Q. A new family of non-local priors for chain event graph model selection. CRiSM Research Report 15-02, 2015.

Cowell, R. G. and Smith, J. Q. Causal discovery through MAP selection of stratified chain event graphs. *Electronic Journal of Statistics*, 8 (1): 965–997, 2014. MR3263109

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. *Probabilistic Networks and Expert Systems*. Springer Verlag, New York, USA, 2007. MR1697175

Dawid, A. P. Conditional independence. In S. Kotz, C. B. Read, and D. L. Banks, editors, *Encyclopedia of Statistical Science*, volume 2, pages 146–153. Wiley-Interscience, update edition, 1998. MR1605063

Dean, T. and Kanazawa, K. A model for reasoning about persistence and causation. *Computational Intelligence*, 5 (3): 142–150, 1989.

Didelez, V. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70 (1): 245–264, 2008. MR2412641

Fergusson, D. M., Horwood, L. J., and Shannon, F. T. Social and family factors in childhood hospital admission. *Journal of Epidemiology and Community Health*, 40 (1): 50, 1986.

Freeman, G. and Smith, J. Q. Bayesian MAP model selection of Chain Event Graphs. *Journal of Multivariate Analysis*, 102 (7): 1152–1165, 2011a. MR2805655

Freeman, G. and Smith, J. Q. Dynamic staged trees for discrete multivariate time series: forecasting, model selection and causal analysis. *Bayesian Analysis*, 6 (2): 279–305, 2011b. MR2806245

French, S. and Insua, D. Rios. *Statistical Decision Theory: Kendall's Library of Statistics 9*. Wiley, 2010.

Friedman, N. and Goldszmidt, M.. Learning Bayesian Networks with local structure. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 421–460. MIT Press, 1998.

Geiger, D. and Heckerman, D. Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence*, 82 (1): 45–74, 1996. MR1391056

Gottard, A. On the inclusion of bivariate marked point processes in graphical models. *Metrika*, 66 (3): 269–287, 2007. MR2336480

Heckerman, D. A tutorial on learning with Bayesian Networks. *Innovations in Bayesian Networks*, pages 33–82, 2008.

Johnson, N. L., Kotz, S., and Balakrishnan, N. *Continuous Univariate Distributions*. Number v. 1 in Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley & Sons, 1995. MR1326603

Kjærulff, U. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the Eighth International Conference on Uncertainty in Artificial Intelligence*, UAI'92, pages 121–129, 1992.

Korb, K. B. and Nicholson, A. E. *Bayesian Artificial Intelligence*, volume 1. CRC Press, 2004. MR2130189

Kulkarni, V. G. *Modeling and analysis of stochastic systems*, volume 36. CRC Press, 1995. MR1357414

Medhi, J. *Stochastic Processes*. New Age International, 1994.

Murphy, K. P. *Machine Learning: a Probabilistic Perspective*. The MIT Press, 2012.

Neapolitan, R. E. *Learning Bayesian Networks*. Pearson Prentice Hall Upper Saddle River, 2004.

Nicholson, A. E. Monitoring Discrete Environments Using Dynamic Belief Networks. PhD thesis, Department of Engineering Sciences, Oxford, 1992.

Nodelman, U., Shelton, C. R., and Koller, D. Continuous time Bayesian networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 378–387, 2002.

Nodelman, U., Shelton, C. R., and Koller, D. Learning continuous time Bayesian networks. In *Proceedings of the Nineteenth International Conference on Uncertainty in Artificial Intelligence*, pages 451–458, 2003.

Pearl, J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, second edition, 2009. MR2548166

Riccomagno, E. and Smith, J. Q. The geometry of causal probability trees that are algebraically constrained. *Optimal Design and Related Areas in Optimization and Statistics*, pages 133–154, 2009. MR2513349

Rubio, F., Flores, M. J., Gómez, J. M., and Nicholson A. Dynamic Bayesian Networks for semantic localization in robotics. In *XV Workshop of Physical Agents: Book of Proceedings, WAF 2014, June 12th and 13th, 2014 León, Spain*, pages 144–155, 2014.

Smith, J. Q. *Decision Analysis – Principles and Practice*. Cambridge University Press, 2010. MR2828346

Smith, J. Q. and Anderson, P. E. Conditional independence and Chain Event Graphs. *Artificial Intelligence*, 172 (1): 42–68, 2008. MR2388535

Thwaites, P. A. Causal identifiability via Chain Event Graphs. *Artificial Intelligence*, 195: 291–315, 2013. MR3024205

Thwaites, P. A. and Smith, J. Q. Evaluating causal effects using Chain Event Graphs. In *Proceedings of PGM, 2006*, pages 293–300, 2006a.

Thwaites, P. A. and Smith, J. Q. Non-symmetric models, Chain Event Graphs and propagation. In *Proceedings of IPMU, 2006*, pages 2339–2347, 2006b.

Thwaites, P. A. and Smith, J. Q. Separation theorems for Chain Event Graphs. *CRiSM Research Report 11-09*, 2011.

Thwaites, P. A., Smith, J. Q., and Cowell, R. G. Propagation using Chain Event Graphs. In *Proceedings of the Twenty-Fourth Annual Conference on Uncertainty in Artificial Intelligence (UAI-08)*, pages 546–553. AUAI Press, 2008.

Thwaites, P. A., Smith, J. Q., and Riccomagno, E. Causal analysis with Chain Event Graphs. *Artificial Intelligence*, 174 (12): 889–909, 2010. MR2722255