Conditional independence and chain event graphs

Jim Q. Smith† and Paul E. Anderson†,‡
†Department of Statistics and
‡Interdisciplinary Program for Cellular Regulation
University of Warwick
Coventry
United Kingdom
CV4 7AL
j.q.smith@warwick.ac.uk

Abstract

Graphs provide an excellent framework for interrogating symmetric models of measurement random variables and discovering their implied conditional independence structure. However, it is not unusual for a model to be specified from a description of how a process unfolds (i.e. via its event tree), rather than through relationships between a given set of measurements. Here we introduce a new mixed graphical structure called the chain event graph that is a function of this event tree and a set of elicited equivalence relationships. This graph is more expressive and flexible than either the Bayesian network — equivalent in the symmetric case — or the probability decision graph. Various separation theorems are proved for the chain event graph. These enable implied conditional independencies to be read from the graph's topology. We also show how the topology can be exploited to tease out the interesting conditional independence structure of functions of random variables associated with the underlying event tree.

1 Introduction

A Bayesian Network (BN) is an established framework for encoding and interrogating conditional independence statements. However, despite its advantages, many problems have been discovered whose underlying structure cannot be fully expressed by a single BN. Thus, for example, two well known Microsoft BN products incorporate special additional information [3]. Four of the common instances when BNs do not capture all of the problem's structure are listed in [20].

Such observations have prompted the development of so called context-specific networks, both to prove new analogues of Pearl's d-separation theorem, and to guide the search for efficient probabilistic representation, propagation, estimation and minimum cost variable assignment. Early models often supplemented BNs with additional structure, usually encoded via trees [3]. The majority of the most recent work has focused on propagation and estimation and has progressively become less graphical. For example, a powerful and ingenious method of propagation using context-specific tables as primitives (called confactors) has been devised [20].

Similar types of information can also be represented via collections of polynomial equations [21]. In a more inferential vein, other methods [6, 9] employ context-specific information for estimation in an undirected, graphical, log-linear framework. Further, very general methods based on the Case-Factor Diagram have been developed to solve a large class of problems [15], by employing directed (as opposed to mixed) graphs. Their methods, based on Boolean formulae, represent many different classes of probabilistic models and, in distinction to the objectives in this article, construct algorithms through minimising a given cost variable.

Another graphical framework, the *Probability Decision Graph* (PDG) [10], is also based on Boolean logic. The focus there is on fast propagation algorithms. Unlike our representation, this framework is not purely graphical and its semantics are not rich enough to contain all BNs as a special case. For example, Jaeger shows through exhaustive enumeration that the diamond shaped BN shown in figure 1 cannot be represented in his model class.

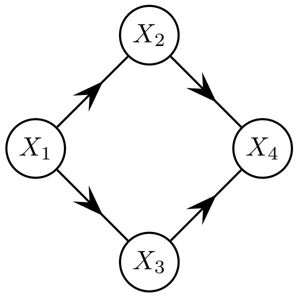

Figure 1: A Bayesian network for four variables that cannot be represented by a probability decision graph.

We do not start from a BN (as the context-specific models do) or a Boolean structure, but rather an event tree. In several different fields, for example Bayesian policy analysis [7], risk analysis [2], physics [14] and biological regulation [1, 5], models are often elicited as an event tree rather than a BN. In fact, one of the motivations for the earliest

BNs and influence diagrams was to efficiently depict, classify and store probability tables associated with problems whose event tree descriptions were highly symmetrical [23, 24]. (That is, the branches of the tree all have the same, or very similar, topologies).

An event tree represents how processes might unfold. The atoms of the resulting event space are its root-to-leaf paths. For illustration, consider the symmetric event tree $\mathcal{T}$ given in figure 2.

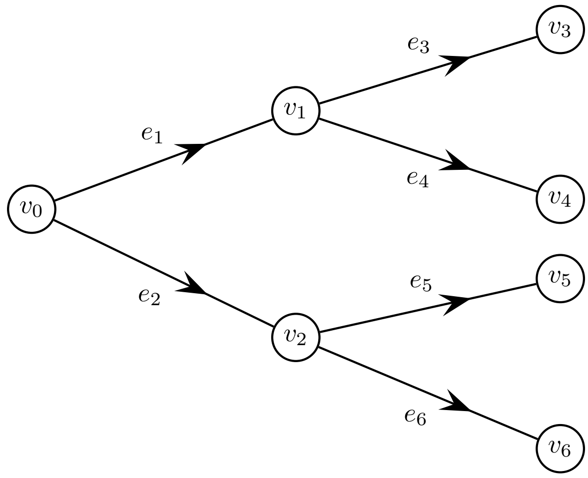

Figure 2: A simple symmetric event tree.

Its atoms are its four root-to-leaf paths $\{e_1, e_3\}, \{e_1, e_4\}, \{e_2, e_5\}, \{e_2, e_6\}$ which are labelled by the terminal vertices $\{v_3, v_4, v_5, v_6\}$ respectively. Two binary random variables X_1 and X_2 can be constructed (where X_2 does not happen before X_1) from this event space. Its atoms are thus:

$$(x_1, x_2) = \{(0,0) = v_3, (0,1) = v_4, (1,0) = v_5, (1,1) = v_6\}$$

Note that the topology of the tree can explicitly acknowledge which events are possible. Thus, if when $X_1 = 1$ it is a logical necessity that $X_2 = 1$, then the tree would have a different topology: the edge e_5 and the vertex v_5 would be missing from $\mathcal{T}$. The event tree therefore has a great advantage over the BN in that it can express this type of asymmetry explicitly.

Event trees have their own Boolean logic and so there are clear links with Jaeger [10] and McAllester *et al* [15] in this regard. However, unlike these authors, we see such trees (and not a construction from another framework, such as a Markov field, BN or junction tree) as the foundation of an elicited model.

In a seminal work [25], Shafer demonstrated that an elicited tree was often a much more powerful expression of an observer's beliefs about the process. He produced compelling arguments to show that this is particularly true when those beliefs are based on an underlying conjecture concerning a specific causal mechanism: a common occurrence in many disciplines.

There is an apparent redundancy in the event tree representation of the event space $\{v_3, v_4, v_5, v_6\}$ above: the interior vertices v_0, v_1 and v_2 (the *situations*) together with all the edges are an unnecessary embellishment. However, Shafer convincingly demonstrates that if situations are consistent with the order in which they unfold (in this case that X_2 does not occur before X_1) then the tree captures other useful "causal" structure. Hence, the edges e_1 and e_2 can be directly associated with the events $\{X_1 = 0\}$ and $\{X_1 = 1\}$ respectively. Furthermore the edges $\{e_3, e_4, e_5, e_6\}$ can be associated with the respective conditional events $\{X_2 = 0|X_1 = 0\}, \{X_2 = 1|X_1 = 0\}, \{X_2 = 0|X_1 = 1\}, \{X_2 = 1|X_1 = 1\}$ and the vertices v_1 and v_2 with the two different conditioning situations $\{X_1 = 0\}$ and $\{X_1 = 1\}$ under which the possible future evolution of the process is differentiated. The tree thus not only explicitly represents the joint event space but also certain conditional events and conditioning situations central to dependence relationships.

The topology of the tree does not represent conditional independence directly. However, we demonstrate in this paper that it is possible to construct a graph — the *Chain Event Graph (CEG)* — that does.

A CEG is a function of the tree and a collection of equations on certain conditional probabilities. Suppose it is asserted that $X_2 \coprod X_1$ (i.e. X_2 is independent of X_1) in the example above. Call the tree and this elicited assertion Model 1. The independence statement is equivalent to the two equations

$$
\begin{aligned}
P(X_2 = 0|X_1 = 0) &= P(X_2 = 0|X_1 = 1) \\
P(X_2 = 1|X_1 = 0) &= P(X_2 = 1|X_1 = 1)
\end{aligned}
$$

This implies that the set of all possible future unfoldings of the tree from situation v_1 are predictively equivalent to those from situation v_2. Furthermore in this predictive sense, the conditioned event e_3 is equivalent to e_5 and e_4 to e_6. The CEG defined formally in section 2 is able to express this type of elicited equivalence topologically by associating the predictively equivalent vertices and edges of $\mathcal{T}$ in the obvious way. Thus the CEG, $\mathcal{C}$, of Model 1 depicted in figure 3 has vertex set $V(\mathcal{C})$ and edge set $E(\mathcal{C})$ given by

$$
\begin{aligned}
V(\mathcal{C}) &= \{w_0 = \{v_0\}, w_1 = \{v_1, v_2\}, w_\infty = \{v_3, v_4, v_5, v_6\}\} \\
E(\mathcal{C}) &= \{e_1^*(w_0, w_1) = e_1, e_2^*(w_0, w_1) = e_2, \\
&\quad e_3^*(w_1, w_\infty) = \{e_3, e_5\}, e_4^*(w_1, w_\infty) = \{e_4, e_6\}\}
\end{aligned}
$$

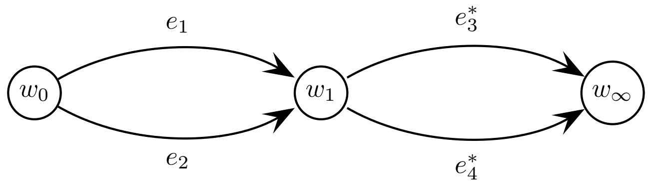

Figure 3: The CEG for Model 1.

Note that:

1. The root-to-sink paths, $\{e_1, e_3^*\}, \{e_1, e_4^*\}, \{e_2, e_3^*\}, \{e_2, e_4^*\}$ of $\mathcal{C}$ are in one-to-one correspondence with, respectively, the root-to-leaf paths $\{e_1, e_3\}, \{e_1, e_4\}, \{e_2, e_5\}, \{e_2, e_6\}$ of the original tree. So, as for the event tree, all atoms in the associated event space of $\mathcal{C}$ are explicitly represented as paths in its graph.

2. The topology of $\mathcal{C}$ is simpler than $\mathcal{T}$ in the sense that it has fewer vertices and edges.

3. Unlike $\mathcal{T}, \mathcal{C}$ represents the statement $X_2 \coprod X_1$ topologically. Hence we can read directly from the graph that, on reaching the vertex $w_1 = \{v_1, v_2\}$ the probabilities of the conditioned events $e_3^* = \{e_3, e_5\}$ and $e_4^* = \{e_4, e_6\}$ are the same. We show later that, with an appropriate definition, the set of conditional independence statements in a BN can be equivalently coded in the CEG.

4. Like the BN, the CEG expresses qualitative information such as whether or not certain sets of conditional probabilities emanating from different situations are the same and, unlike the BN, the explicit structure of the event space. However, we *don't* need the values of these conditional probabilities to actually draw a CEG.

One feature of a BN, sometimes not acknowledged in practice, is the critical role played by the underlying components $\{X_1, X_2, .., X_n\}$ of a random vector $\boldsymbol{X}$ labelling the vertices of the network. These components are given a preferred status over any other transformed random vector $\boldsymbol{g}(\boldsymbol{X}) = \{g_1(\boldsymbol{X}), g_2(\boldsymbol{X}), .., g_n(\boldsymbol{X})\}$, where $\boldsymbol{g}$ is invertible. This is despite the fact that the event space of $\boldsymbol{g}(\boldsymbol{X})$ is an equally good representation of the underlying sample space of the problem. This is fine in contexts when it is *only* reasonable to postulate model classes whose conditional independence relationships between subsets of the components $\{X_1, X_2, .., X_n\}$ are not functions of these variables. However, even in the simplest scenarios such model classes can appear very restrictive.

In the event tree above, suppose both X_1 and X_2 measure the presence of some attribute at an early and late time respectively. Instead of Model 1 ($X_2 \coprod X_1$), a reasonable alternative, Model 2, might assert that the probability that X_2 takes a different value to X_1 is independent of the value of X_1. This is equivalent to

$$P(X_2 = 0|X_1 = 0) = P(X_2 = 1|X_1 = 1)$$
$$P(X_2 = 1|X_1 = 0) = P(X_2 = 0|X_1 = 1)$$

Now, in contrast to Model 1, the conditioned event e_3 is equivalent to e_6 and e_4 to e_5. So the CEG $\mathcal{C}$ of Model 2 has vertex set $V(\mathcal{C})$ and edge set $E(\mathcal{C})$ given by

$$V(\mathcal{C}) = \{w_0 = \{v_0\}, w_1 = \{v_1, v_2\}, w_\infty = \{v_3, v_4, v_5, v_6\}\}$$
$$E(\mathcal{C}) = \{e_1, e_2, e_3'(w_1, w_\infty) = \{e_3, e_6\}, e_4'(w_1, w_\infty) = \{e_4, e_5\}\}$$

The new CEG is topologically the same but the edge equivalences are different: e_i' replaces e_i^*, $i = 3, 4$. Notice from the equations above that we automatically create a new indicator random variable Y that takes the value zero, say, when e_3' occurs (i.e. when $x_1 = x_2$) and one, say, when e_4' occurs (i.e. when $x_1 \neq x_2$). Analogous to Model 1, we prove later that the probability equations tell us that $Y \coprod X_1$. Therefore the tree and the collection of probability equivalences is embodied in the topology of the CEG and this allows a visual identification of a new pair of random variables (X_1, Y) that are independent of each other.

Note that Model 2 is not a BN on the variables (X_1, X_2). The only way to incorporate this information in a BN is to increase the sample space artificially to (X_1, X_2, Y). Then Model 1 would be a BN with directed edge set $\{(X_1, Y), (X_2, Y)\}$ and Model 2 with edge set $\{(X_1, X_2), (Y, X_2)\}$. When we need the flexibility to simultaneously consider these two types of model, both of which have been elicited from an explanation of how situations unfold, and want to examine the implicit conditional independence structure, we argue that the class of CEG models is a much more natural tool than the BN.

Once the CEG has been agreed with the expert observer, it can be used as a framework for further elaboration into a full probabilistic model in the same way as the other constructions discussed above. Furthermore, it gives a much more compact description of a problem than an event tree. For example, n k-state independent random variables $\{X_1, X_2, ...X_n\}$ are represented by a tree with k^n edges, whilst — like the directed acyclic graph of the related Case-Factor Diagrams [15] — the CEG has only nk edges. Unlike the PDG [10], we prove that all finite discrete BNs can be expressed as a CEG. In fact, this is also true for all context-specific BNs as defined in [3]. This is illustrated in the last example of this paper, see figure 15.

In section 2 we review the BN and the event tree and give a general definition of the CEG — a mixed graph with some of its directed edges coloured. We illustrate its construction and how it can be used to encode elicited qualitative information about a process. In section 3 we show how to construct useful random variables from the topology of a CEG and how to read off implied conditional independence relationships between these variables, even when the underlying process, unlike the one discussed in the introduction, is highly non-symmetric. We prove that all information in a BN can always be represented by a CEG, but not vice versa.

We give various analogues of the d-separation theorem for BNs for the CEG in section 4, and show how other dependence relationships, not encoded in the BN, can be read from the CEG when it is based on the tree of a context-specific BN. We also suggest a general algorithm for interrogating the dependencies of a given CEG. In the final section, we briefly discuss connections to other work and current developments in this field.

2 Some Background on Graphs and the CEG

2.1 Bayesian networks: a review

Let $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$, where X_i are discrete random vectors which take one of the r_i values in the sample space $\mathbb{X}_i$, $1 \leq i \leq n$. Write $\mathbb{X}^{(i)} = \prod_{j=1}^{i} \mathbb{X}_j$, $\mathbb{X} = \mathbb{X}^{(n)}$, $r^{(i)} = \prod_{j=1}^{i} r_j$, $2 \leq j \leq n$ and $r = r^{(n)}$. There are many equivalent ways of defining a BN. For this paper it is most convenient to use the total order of the components in $\mathbf{X}$ and express the $n-1$ conditional independence statements

$$X_i \coprod \{X_1, X_2, \ldots, X_{i-1}\} | \mathbf{Q}_i$$

where $\mathbf{Q}_i \subseteq \{X_1, X_2, \ldots, X_{i-1}\}$, $2 \leq i \leq n$. As a notational convention, let $\mathbf{Q}_1$ be the empty set and call the set of random vectors $\mathbf{Q}_i$ the *parent set* of X_i, $1 \leq i \leq n$. The BN $\mathcal{D}$ is then the directed graph whose vertex set $V(\mathcal{D})$ is labelled by the set of n random variables and has edge set $E(\mathcal{D})$, where $e = (X_j, X_i) \in E(\mathcal{D})$ if and only if $X_j \in \mathbf{Q}_i$ [27]. The d-separation theorem ([17] and later re-expressed in, for example, [13] using constructions based on [12]) allows one to answer arbitrary conditional independence queries about relationships between disjoint subsets of the variables.

Let $X[1], X[2], X[3] \subseteq \{X_1, X_2, \ldots, X_n\}$ be disjoint subsets of components of $\mathbf{X}$. Let the set $A(B)$ of a set of vertices, B, consists of all vertices in $V(\mathcal{D})$ that are in B or that lie on a directed path in $\mathcal{D}$ which leads to a vertex in B. The *moralised graph* $\mathcal{D}_M$ of $\mathcal{D}$ is the mixed graph with vertex set $V(\mathcal{D})$ and directed edges $E(\mathcal{D})$, but with an undirected edge between any two vertices $v[1], v[2] \in V(\mathcal{D})$ such that whenever neither $(v[1], v[2])$ nor $(v[2], v[1])$ is in $E(\mathcal{D})$, there exists a vertex $v[3] \in V(\mathcal{D})$ where both $(v[1], v[3])$ and $(v[2], v[3])$ are in $E(\mathcal{D})$.

Let $\mathcal{D}_u$ denote the undirected graph obtained from $\mathcal{D}_M$ by replacing all directed edges in $E(\mathcal{D}_M)$ by undirected edges. For any $C \subseteq \{X_1, X_2, \ldots, X_n\}$, let $\mathcal{D}[C]$ have vertex set $V(\mathcal{D}[C]) = V(\mathcal{D}) \cap C$ and an edge between $v[1], v[2] \in V(\mathcal{D}[C])$ if and only if there is an edge between $v[1], v[2] \in V(\mathcal{D})$. The d-separation theorem [13] now states that

$$X[3] \coprod X[2] | X[1]$$

is a valid deduction if $X[1]$ separates $X[2]$ and $X[3]$ in $\mathcal{D}_U[A(X[1] \cup X[2] \cup X[3])]$. That is, all undirected paths in $\mathcal{D}_U[A(X[1] \cup X[2] \cup X[3])]$ from a vertex in $X[2]$ to a vertex in $X[3]$ must pass through a vertex in $X[1]$. Note that this theorem concerns only deductions about the relationships between subsets of $\{X_1, X_2, \ldots, X_n\}$ and not general functions of these variables.

A joint mass function $\pi(\mathbf{x})$ on the random variables $\{X_1, \ldots, X_n\}$ can be factored in the form

$$\pi(\mathbf{x}) = \prod_{i=1}^{n} \pi_i(x_i | \mathbf{x}^{(i-1)}) \tag{1}$$

where $\pi_i(x_i|\mathbf{x}^{(i-1)})$, $1 \leq i \leq n$, is a conditional mass function of x_i given $\mathbf{x}^{(i-1)} = (x_1, \ldots, x_{i-1}) \in \mathbb{X}^{(i-1)}$, for $2 \leq i \leq n$ (and $x^{(0)}$ denotes the empty set). These conditional mass functions have an important role in our subsequent discussion so call $\pi_i(x_i|\mathbf{x}^{(i-1)})$ with $x_i \in \mathbb{X}_i$, $\mathbf{x}^{(i-1)} \in \prod_{j=1}^{i-1} \mathbb{X}_i$ and $1 \leq i \leq n$, *primitive probabilities*. The factorisations in equation (1) can be seen as a set of r equations whose arguments are the primitive probabilities $\pi_i(x_i|\mathbf{x}^{(i-1)})$, $x_i \in \mathbb{X}_i$, having $(x_i|\mathbf{x}^{(i-1)}) \in \mathbb{X}^{(i)}$ as their indices.

The conditional probabilities obviously respect the simplex conditions for $1 \leq i \leq n$ and each $\mathbf{x}^{(i-1)} \in \mathbb{X}^{(i-1)}$

$$\sum_{x_i \in \mathbb{X}_i} \pi_i(x_i|\mathbf{x}^{(i-1)}) = 1$$

and

$$\pi_i(x_i|\mathbf{x}^{(i-1)}) \geq 0, \quad x_i \in \mathbb{X}_i$$

Using this representation, let $\mathcal{D}$ be the directed graph defined above and let $\mathbb{X}_{\mathbf{Q}_i}$ be the sample space for the random variables in $\mathbf{Q}_i$ in the parent set of X_i, $1 \leq i \leq n$. Consider two instantiations $\mathbf{x}^{(i-1)}$ and $\mathbf{x}'^{(i-1)} \in \mathbb{X}^{(i-1)}$ whose projection onto $\mathbb{X}_{\mathbf{Q}_i}$ coincide. In other words, for which

$$\mathbf{q}_i(\mathbf{x}^{(i-1)}) = \mathbf{q}_i(\mathbf{x}'^{(i-1)}) \tag{2}$$

where $\mathbf{q}_i(\mathbf{x}^{(i-1)})$ is the projection of $\mathbf{x}^{(i-1)}$ onto $\mathbb{X}_{\mathbf{Q}_i}$.

Let $r(\mathbf{q}_i) = \prod_{\{j:x_j \notin \mathbf{Q}_i\}} r_j$. The set of conditional independence statements above are then equivalent to the assertion that

$$\pi_i(x_i|\mathbf{x}^{(i-1)}) = \pi_i(x_i|\mathbf{x}'^{(i-1)}) \tag{3}$$

whenever equation (2) holds.

This in turn is equivalent to asserting that

$$\pi(\mathbf{x}) = \prod_{i=1}^{n} \pi_i(x_i|\mathbf{q}_i(\mathbf{x}^{(i-1)})) \tag{4}$$

which is the familiar factorisation of a joint probability mass function associated with a BN. However, implicitly specifying this factorisation through statements concerning the equality of the distributions of random variables with different conditioning sets, seamlessly transfers to classes of more heterogeneous models.

2.2 Factorisations from event trees

Here we will define and briefly review some properties of an event tree based on [25, 26] indicating when we diverge from their terminology. An event tree is a directed, rooted tree $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ where $V(\mathcal{T})$ denotes its vertex set, assumed finite, and $E(\mathcal{T})$ its edge set. Denote the *root vertex* (the only vertex of this tree with no edge into it) by v_0 and call any vertex with no edge out of it a *leaf vertex* v'. Throughout this paper, in distinction to Shafer, we call a non-leaf vertex v a *situation* and denote the set of situations by $S(\mathcal{T}) \subset V(\mathcal{T})$.

Henceforth, Λ will denote the set of root-to-leaf paths of $\mathcal{T}$. The paths $\lambda \in \Lambda$ which form the atoms of the event space (called the *path σ-algebra* of $\mathcal{T}$) label the different possible unfoldings of the described process. Each event $\{Y = y\}$ such that $y \in \mathbb{Y}$ (where $\mathbb{Y}$ denotes the sample space of a random variable Y measurable with respect to this event space) will label a subset $\Lambda(Y = y) \subseteq \Lambda$. Furthermore, the sets $\{\Lambda(Y = y) : y \in \mathbb{Y}\}$ will form a partition of Λ. We will demonstrate later how to identify topologically various interesting random variables associated with a process described by an event tree.

Unlike BNs, event trees can be used to describe highly non-symmetric processes. For example, consider the following fictitious but nevertheless typical model description of a biological regulatory system.

A culture is placed in an environment which: is benign ($B = 0$), can potentially disrupt gene interaction but is not physically damaging ($B = 1$), is physically damaging but does not disrupt gene interaction ($B = 2$) or can potentially disrupt gene interaction and is physically damaging ($B = 3$). Given that the environment damages the cell, it can repair itself: quickly ($R = 2$), slowly ($R = 1$) or be unable to repair ($R = 0$).

Assume the system hinges on two genes that can be under expressed ($G_i = -1$), normally expressed ($G_i = 0$) or highly expressed ($G_i = 1$), $i = 1, 2$. Suppose that we know from the gene pathways that if $G_1 = 1$ then $G_2 = 0$ or $G_2 = 1$ and if $G_1 = 0$ then $G_2 = 0$. Our interest is in whether the environment causes a cancerous increase in cells ($C = 1$) or not ($C = 0$). This increase can be affected either by enduring cell damage or disruption of the gene pathway in an otherwise undamaged cell.

When a process is described in this way, we note that the edge labels R, G_1 and G_2 are defined contingent on what has happened earlier in the unfolding. They can therefore be seen as labels of states of conditioned random variables defined, possibly only *conditionally*, on certain earlier developments. Thus it is meaningless to talk about the repair of a cell if it is not damaged, and expression of genes is only relevant to cancerous increase if the cell has been repaired but its interaction possibly disrupted. The full and direct expression of this description by a single BN is therefore not possible. However it is simple to express this process directly using an event tree, as shown in figure 4.

For example, the third path down $\lambda(B = 1, G_1 = -1, G_2 = -1, C = 0)$ expresses the unfolding that the environment only possibly disrupts cell interaction, the first gene becomes under expressed, the second gene becomes under expressed but there is no increase in cancerous cells. Note that each situation $v \in S(\mathcal{T})$ in this tree represents an attained state of the process that determines its subsequent development. Thus the first situation on this path v_0 defines the conditions under which the background environment is determined.

The edges label the possible states this background variable can take. The next situation $v(B = 1)$ tells us that we have an environment only possibly disruptive to gene interaction and edges from this situation determine the possible resulting expression states of the gene. The situation $v(B = 1, G_1 = -1)$ defines the state when $B = 1$ and the first gene has under expressed and finally $v(B = 1, G_1 = -1, G_2 = -1)$ the situation when the second gene has also under expressed. Note that the labels on the edges of a tree give the values a variable can take *conditional* on the circumstances defined by the situation from which those

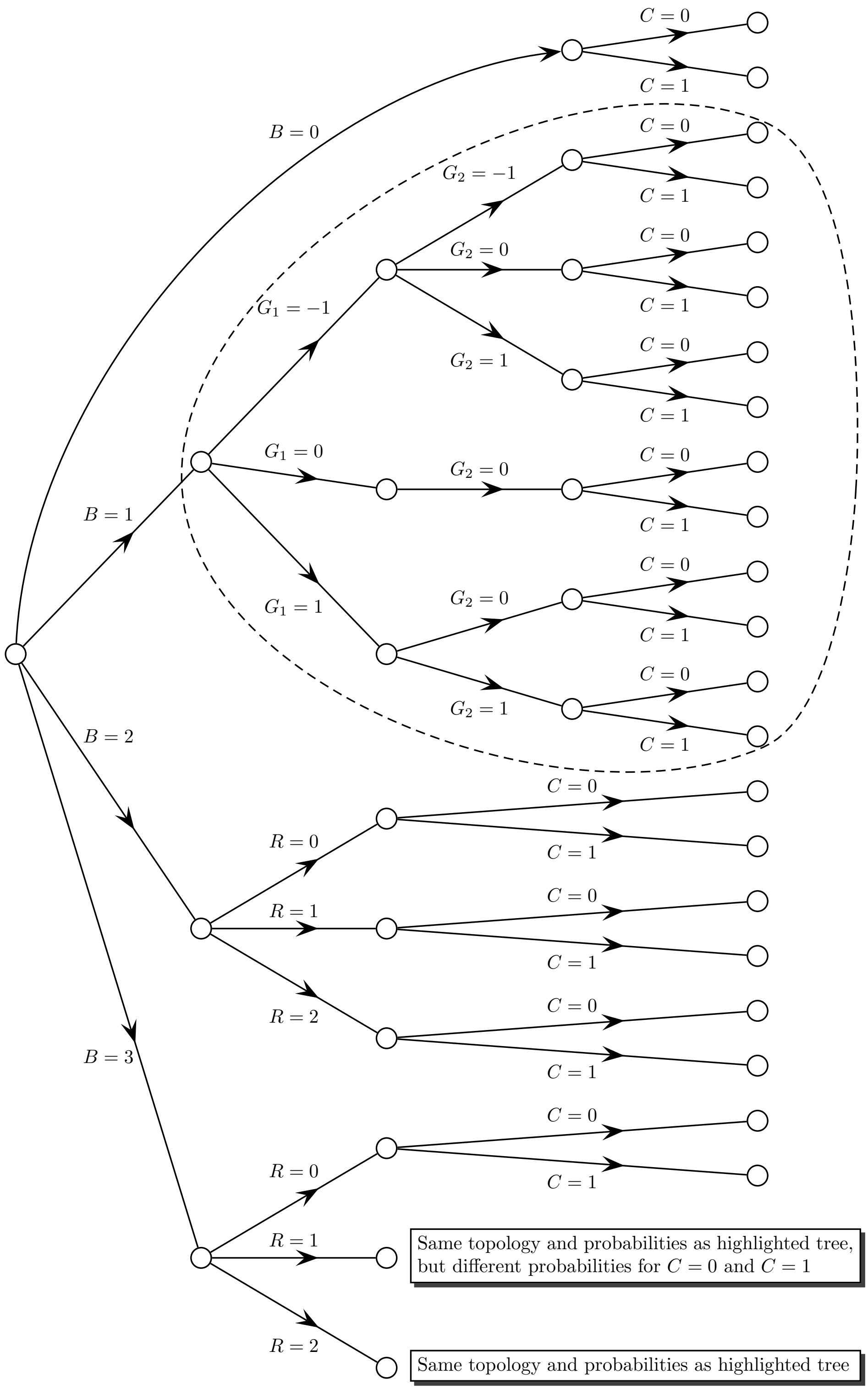

Figure 4: Event tree for the cell culture example. The dash-outlined tree is repeated (with minor differences) in other parts of the tree as indicated. See the text for an explanation of the variables.

edges emanate.

This means that each situation v in an event tree has a dual role: it expresses a state of a process and it also serves as an index of a random variable $X(v)$ whose values describe the next stage of possible developments of the unfolding process. The state space $\mathbb{X}(v)$ of $X(v)$ can be identified with the set of directed edges $(v, v') \in E(\mathcal{T})$ that emanate from v in $\mathcal{T}$.

For each $\{X(v) : v \in S(\mathcal{T})\}$, let

$$\Pi(v) = \{\pi(v'|v) : v' \in \mathbb{X}(v)\} \tag{5}$$

be the *primitive probabilities* associated with the random variable $X(v)$, where $\pi(v'|v) = P(X(v) = v'|v)$, and let $\Pi = \bigcup_{\{v \in S(\mathcal{T})\}} \Pi(v)$. Obviously these probabilities must satisfy, for all $v \in S$,

$$\sum_{v' \in \mathbb{X}(v)} \pi(v'|v) = 1$$

and for all $v' \in \mathbb{X}(v)$, $v \in S(\mathcal{T})$, $\pi(v'|v) \geq 0$.

The probabilities $\mathbf{Q} = \{\pi(\lambda) : \lambda \in \Lambda\}$ of the elementary events $\lambda \in \Lambda$ can now be given as products of these primitive probabilities Π, [25, 26]. Assume that each root-to-leaf path $\lambda = (v_{0,\lambda}, v_{1,\lambda}, \ldots, v_{n[\lambda],\lambda}) \in \Lambda$ with $v_{0,\lambda} = v_0$, is $n[\lambda] \geq 0$ edges from the root vertex. Then the probabilities $\pi(\lambda)$ for every $\lambda \in \Lambda$ must satisfy the equations

$$\pi(\lambda) = \prod_{j=0}^{n[\lambda]-1} \pi(v_{j+1,\lambda}|v_{j,\lambda}) \tag{6}$$

Like the BN, the probabilities of elementary events can be expressed as a set of monomials in the primitive probabilities. However, unlike the BN these monomials can be of different degrees. This is the case in figure 4. Note that a necessary and sufficient condition for these equations to hold is that $\{X(v), v \in A\}$ are mutually independent whenever all $v \in A$ lie on a single path in $\mathcal{T}$. Henceforth we shall assume this is true for consistency with other work such as [25].

Clearly, a full specification of the probability model is given by $(\mathcal{T}, \Pi(\mathcal{T}))$: the tree and its set of primitive probabilities. It is common, having elicited an event tree, to learn that one of a set $\Lambda^* \subset \Lambda$ of root-to-leaf paths in $(\mathcal{T}, \Pi(\mathcal{T}))$ has occurred and it is necessary to condition on this event in the event space associated with the paths of $\mathcal{T}$. Within the event tree framework it is in fact simple to construct a tree that reflects this change.

Notation 1 *The* Λ^*-tree $\mathcal{T}_{\Lambda^*} = (V(\mathcal{T}_{\Lambda^*}), E(\mathcal{T}_{\Lambda^*}))$ *has vertex set* $V(\mathcal{T}_{\Lambda^*})$, *edge set* $E(\mathcal{T}_{\Lambda^*})$ *and situations* $S(\mathcal{T}_{\Lambda^*})$ *defined by*

$$\begin{aligned}
V(\mathcal{T}_{\Lambda^*}) &= \{v \in V(\mathcal{T}) : v \textit{ is on a root-to-leaf path } \lambda \in \Lambda^*\} \\
E(\mathcal{T}_{\Lambda^*}) &= \{e \in E(\mathcal{T}) : e \textit{ is on a root-to-leaf path } \lambda \in \Lambda^*\} \\
S(\mathcal{T}_{\Lambda^*}) &= \{v \in S(\mathcal{T}) : v \textit{ is on a root-to-leaf path } \lambda \in \Lambda^*\}
\end{aligned}$$

Using an obvious extension of notation, for each $v \in S(\mathcal{T}_{\Lambda^*})$, let each $X_{\Lambda^*}(v)$ have sample space $\mathbb{X}_{\Lambda^*}(v) \subseteq \mathbb{X}(v)$. Directly from Bayes' rule, it is simple to find the associated primitive probabilities as functions of the primitives in $\Pi(\mathcal{T})$:

$$\Pi_\Lambda(v) = \{\pi_\Lambda(v'|v) : v' \in \mathbb{X}_{\Lambda^*}(v)\}$$

These constitute the new set of primitive probabilities $\Pi_{\Lambda^*}(\mathcal{T}_{\Lambda^*})$ associated with $\mathcal{T}_{\Lambda^*}$ after conditioning. Thus whenever v is a terminal situation in $S(\mathcal{T}_{\Lambda^*})$, then

$$\pi_{\Lambda^*}(v'|v) = \mu_{\Lambda^*}^{-1}[v]\pi(v'|v)$$

where $v' \in \mathbb{X}_{\Lambda^*}(v)$ and $\mu_{\Lambda^*}[v] = \sum_{v' \in V(\mathcal{T}_{\Lambda^*})} \pi(v'|v)$.

The remaining primitives associated with a non-terminal situation can now be recursively calculated backwards along the tree as a function of the revised primitives associated with its children and the original primitives associated with $X(v)$. Thus

$$\pi_{\Lambda^*}(v'|v) = \mu_{\Lambda^*}^{-1}[v]\mu_{\Lambda^*}[v']\pi(v'|v)$$

where $v' \in \mathbb{X}_{\Lambda^*}(v)$ and $\mu_{\Lambda^*}[v] = \sum_{v' \in V(\mathcal{T}_{\Lambda^*})} \mu_{\Lambda^*}[v']\pi(v'|v)$. The set of primitive probabilities associated with edges of the new conditional tree is now simply $\Pi_{\Lambda^*} = \bigcup_{\{v \in S(\mathcal{T}) \cap V(\mathcal{T}^*_\Lambda)\}} \Pi_{\Lambda^*}(v)$.

The formulae above give a local propagation of the information that $\lambda \in \Lambda^*$ through $(\mathcal{T}, \Pi(\mathcal{T}))$ analogous to junction tree algorithms for BNs [11]. Just as clique probability tables are sequentially revised to admit new information so, in the case of the tree, the distributions of $\{X(v) : v \in S(\mathcal{T})\}$ are revised to $\{X_\Lambda(v) : v \in S(\mathcal{T}_{\Lambda^*})\}$ using the algorithm above. In general, when conditioning on the observation of a general function measurable with respect to the path σ-algebra associated with the tree, the updating algorithm given above will not necessarily be quick: this should not however be surprising. Updating probability tables after observing a *general* function of variables in a BN can also be very time consuming, often requiring a new customised triangulation step.

Useful fast junction tree algorithms assume observations need to be of subsets of the variables depicted by the vertices of the BN. The speed of algorithms is therefore linked to conditioning on a compatible type of observation as well as utilising conditional independence structure. Jaeger [10] has now established several fast algorithms based on important classes of these models, see also [15]. Note that conditioning can destroy symmetries in a tree. In particular, it is common for the distributions of $X(v[1])$ and $X(v[2])$ to be the same, but for $X_{\Lambda^*}(v[1])$ and $X_{\Lambda^*}(v[2])$ to differ.

2.3 Probability graphs and chain event graphs

Define the *floret* of v in $\mathcal{T}$ as the subtree

$$\mathcal{F}(v, \mathcal{T}) = (V(\mathcal{F}(v, \mathcal{T})), E(\mathcal{F}(v, \mathcal{T})))$$

of an event tree $\mathcal{T}$ with $v \in S(\mathcal{T})$, where the vertex set $V(\mathcal{F}(v, \mathcal{T}))$ and edge set $E(\mathcal{F}(v, \mathcal{T}))$ are given by

$$\begin{aligned} V(\mathcal{F}(v, \mathcal{T})) &= \{v\} \cup \{v^* \in V(\mathcal{T}) : (v, v^*) \in E(\mathcal{T})\} \\ E(\mathcal{F}(v, \mathcal{T})) &= \{e \in E(\mathcal{T}) : e = (v, v^*) \text{ for some } v^* \in V(\mathcal{T})\} \end{aligned}$$

We noted above that the random variable $X(v)$ has sample space $\mathbb{X}(v) = \{x_1(v), \ldots, x_{n(v)}(v)\}$ where $x_i(v)$ can be used to label an edge in $E(\mathcal{F}(v, \mathcal{T}))$, $1 \leq i \leq n(v)$. It is often possible to elicit information that two situations v and v' are equivalent in the sense that the distribution of their associated random variables $X(v)$ and $X(v')$ are the same. We now set out two key definitions.

Definition 1 *We say that the situations v, v' are in the same stage u if and only if the random variables $X(v)$ and $X(v')$ have the same distribution under a bijection $\psi_u(v, v')$, $v, v' \in u$, where*

$$\begin{aligned} \psi_u(v, v') : \mathbb{X}(v) &= E(\mathcal{F}(v, \mathcal{T})) \mapsto E(\mathcal{F}(v', \mathcal{T})) = \mathbb{X}(v') \\ : x_i(v) &= e(v, v^*) \mapsto e(v', v'^*) = x_{i'}(v') \end{aligned}$$

Note that the set of stages $L(\mathcal{T})$ of a tree $\mathcal{T}$ form a partition of the set of situations $S(\mathcal{T})$. We call $\mathcal{L}(\mathcal{T}) = \{\Psi_u(v, v') : v, v' \in u, u \in L(\mathcal{T})\}$ a *staging* of T.

Definition 2 *A* staged tree *$\mathcal{G}(\mathcal{T}, L(\mathcal{T}), \mathcal{L}(\mathcal{T}))$ is a tree with vertex set $V(\mathcal{G}) = V(\mathcal{T})$, edge set $E(\mathcal{G}) = E(\mathcal{T})$, stage set $L(\mathcal{T})$ and staging $\mathcal{L}(\mathcal{T})$. Its edges are coloured as follows.*
When $v \in u$ and u contains a single vertex, then all edges emanating from v in $E(\mathcal{G})$ are uncoloured.
When $v \in u$ and u contains more than one vertex, then all edges emanating from v in $E(\mathcal{G})$ are coloured.

Two edges $e(v, v^*), e(v', v'^*) \in E(\mathcal{G})$ emanating from v and v' respectively have the same colour if and only if $e(v, v^*) \mapsto e(v', v'^*)$ under $\psi_u(v, v') \in \mathcal{L}(\mathcal{T})$.

Note that a staged tree contains only qualitative information. In particular, the stages of the tree specify the collections of situations the expert believes are equivalent in the sense that they share the same distribution over the next stage of their development. We show later that all information in a BN can equally well be represented in a staged tree. For example, in figure 2, we can identify $X(v_1)$ with $X_2|X_1 = 0$ and $X(v_2)$ with $X_2|X_1 = 1$. The statement that $X_2 \coprod X_1$ is equivalent to the assertion that, under the obvious map of edges, v_1 and v_2 are in the same stage.

The types of stage partitions that are expressible through a BN are highly restricted. For instance, two situations v and v' can *only* lie in the same stage if those situations are the same distance from the root vertex. This type of symmetry is not exhibited through the information which we might elicit to supplement the tree of the biological regulation experiment given in this section. Nevertheless, it can be expressed with a staged tree. For example, suppose we are given the following qualitative information about the regulatory network.

The expression level G_1 of the first gene has the same distribution whenever the disrupted environment does not cause irreparable cell damage (u_1). Also, the distribution of the expression of the second gene G_2 given that the first is highly or lowly expressed has the same distribution in the same circumstance (u_4, u_5). Further, the probability of cancerous increase when both genes are lowly expressed is the same whether they are in a gene disruptive environment where cell damage (if it occurs) is quickly repaired or, similarly, when the genes are both highly expressed (u_7 and u_8). The distribution of cancerous increase when there is irreparable cell damage and neither gene is normally expressed is always the same.

This type of information allows us to identify distributions associated with the random variable $X(v)$ over certain v, giving us a staged tree. Labelling the situations of $\mathcal{T}$ by the numbering of their incoming edges and the root vertex as v_0, gives the stages: $u_0 = \{v_0\}$,
$u_1 = \{v(1), v(3,1), v(3,2)\}, u_2 = \{v(2)\}, u_3 = \{v(3)\}$,
$u_4 = \{v(1,-1), v(3,1,-1), v(3,2,-1)\}, u_5 = \{v(1,1), v(3,1,1), v(3,2,1)\}$,
$u_6 = \{v(0)\}, u_7 = \{v(1,-1,-1), v(3,2,-1,-1)\}$,
$u_8 = \{v(1,1,1), v(3,2,1,1)\}$,
$u_9 = \{v(3,1,-1,-1), v(3,1,1,-1), v(3,1,-1,1), v(3,1,1,1)\}$,
$u_{10} = \{v(3,2)\}$.

A second useful partition $K(\mathcal{T}) = \{w(v) : v \in S(\mathcal{T})\}$ can be defined from a staged tree $\mathcal{G}(\mathcal{T}, L(\mathcal{T}), \mathcal{L}(\mathcal{T}))$. For each situation $v \in S(\mathcal{T})$, let $\Lambda(v, \mathcal{T})$ denote the set of paths in $\mathcal{T}$ from v to a leaf vertex of $\mathcal{T}$. Two situations v, v' are defined to be in the same *position* $w \in K(\mathcal{T})$ if there is a bijective map

$$\begin{aligned}\phi_w(v, v') &: \Lambda(v, \mathcal{T}) \to \Lambda(v', \mathcal{T}) \\ &: \lambda(v) \mapsto \lambda(v')\end{aligned}$$

such that

1. all edges in all the paths in $\Lambda(v, \mathcal{T})$ and $\Lambda(v', \mathcal{T})$ are coloured in $\mathcal{G}(\mathcal{T}, L(\mathcal{T}), \mathcal{L}(\mathcal{T}))$
2. for all paths $\lambda(v) \in \Lambda(v, \mathcal{T})$, the ordered sequence of colours in $\lambda(v)$ equals the ordered sequence of colours in $\lambda(v') = \phi_w(v, v')[\lambda(v)] \in \Lambda(v', \mathcal{T})$.

Two situations v and v' are therefore in the same position when (under the map $\phi_w(v, v')$) the future evolution from both v and v' is governed by the same probability law.

In the cell culture example, we can group the 13 positions w_i as follows: $u_0 = w_0, u_1 = \{w_1, w_6\}, u_2 = w_2, u_3 = w_3, u_4 = \{w_4, w_7\}, u_5 = \{w_5, w_8\}, u_6 = w_9, u_7 = w_{10}, u_8 = w_{11}, u_9 = w_{12}$ and $u_{10} = w_{13}$. Consider, for example, u_1. We choose to distinguish the two cases $w_1 \sim$ value of G_1 given $\{B = 1\}$ or $\{B = 3, R = 2\}$ from $w_6 \sim$ value of G_1 given $\{B = 3, R = 1\}$. We do this because the distribution of C corresponding to these two scenarios may be different later on: the slow repair of the cell during gene interaction may influence cancer cell growth. The other positions distinct from stages are:

- $w_4 \sim$ value of G_2 given $\{B = 1, G_1 = -1\}$ or $\{B = 3, R = 2, G_1 = -1\}$
- $w_5 \sim$ value of G_2 given $\{B = 1, G_1 = 1\}$ or $\{B = 3, R = 2, G_1 = 1\}$
- $w_7 \sim$ value of G_2 given $\{B = 3, R = 1, G_1 = -1\}$
- $w_8 \sim$ value of G_2 given $\{B = 3, R = 1, G_1 = 1\}$

Positions are a very obvious way of equating situations, because two situations in the same position will be impossible to differentiate through subsequent events. For example, the stage u_1 is partitioned into two positions $\{w_1, w_2\}$ because the value of B has a bearing on the distribution of a future but not immediate unfolding. Note that, by an abuse of notation, the stages $\{u : u \in L(\mathcal{T})\}$ partition the set of positions $\{w : w \in K(\mathcal{T})\}$.

A new graph — the *Chain Event Graph* (CEG) — which is useful for deducing implied conditional independencies from a staged tree can now be constructed. Unlike an event tree, the vertices and edges of a CEG play different roles. Its non-leaf vertices will define circumstances in which a unit may find itself. The directed edges emanating from that vertex position label the different possible outcomes that might subsequently be experienced. Finally, undirected edges join positions whose next stage of evolution is governed by the same probability law. The construction of a CEG is based on the probability graph of this event tree model [4, 16, 25].

Definition 3 *The* probability graph $\mathcal{H}(\mathcal{G}(\mathcal{T})) = \mathcal{H}(\mathcal{T}) = (V(\mathcal{H}), E(\mathcal{H}))$ *of a staged tree $\mathcal{G}(\mathcal{T})$ of an event tree $\mathcal{T}$ is a directed graph with, possibly, some coloured edges. Its vertex set is given by $V(\mathcal{H}) = K(\mathcal{T}) \cup \{w_\infty\}$. Its edge set $E(\mathcal{H})$ is constructed as follows.*

For each position $w \in \mathcal{K}(\mathcal{T})$ choose a single representative situation $v(w) \in \mathcal{S}(\mathcal{T})$. For each edge from $v(w)$ to $v'(w) \in E(\mathcal{T})$, denoted by $e(v(w), v'(w))$, construct a single *edge $e(w, w') \in E(\mathcal{H})$ where $w'(w) = w_\infty$ if $v'(w)$ is a leaf vertex of $\mathcal{T}$ and $w'(w) = w(v'(w))$ otherwise, where $w(v'(w)) \in K(\mathcal{T})$ is the position containing $v'(w)$.*

The colour of the edge $(w, w') \in E(\mathcal{H})$ is the colour of the edge $(v(w), v'(w)) \in S(\mathcal{T})$ if $u(w) \neq \{w\}$, where $u(w)$ is the stage containing $v(w)$ and is otherwise uncoloured.

Because $\mathcal{T}$ is finite, all its paths are of finite length so, by definition, $\mathcal{H}(\mathcal{T})$ is directed and acyclic, having a single root vertex $w_0 = v_0$ (the root vertex of its tree) and a single sink vertex w_∞. There is a one-to-one correspondence between all root-to-leaf paths in $\mathcal{T}$ and all root-to-leaf paths in $\mathcal{H}(\mathcal{T})$. Thus each elementary event generated by the root-to-leaf paths in $\mathcal{T}$ appear as w_0 to w_∞ paths $\lambda(w_0, w_\infty)$ in $\mathcal{H}(\mathcal{T})$. Unlike the BN, the probability graph is always rooted but not usually simple, i.e. there can be several directed edges from a node w to another w'.

As for situations on an event tree, for two positions w, w' write $w \prec w'$ when there is a directed path λ in $\mathcal{H}(\mathcal{T})$ from w to w'. Note that each edge in $E(\mathcal{H}(\mathcal{T}))$ can be associated with a primitive probability $\pi \in \Pi$ (although not necessarily uniquely), but that in general $\mathcal{H}(\mathcal{T})$ has far fewer vertices and edges than $\mathcal{T}$.

It is useful to supplement the topology of the probability graph so that the stages are represented explicitly. Thus we have:

Definition 4 *Call the* chain event graph (CEG) $\mathcal{C}(\mathcal{T})$ *the mixed graph with vertex set* $V(\mathcal{C}(\mathcal{T})) = V(\mathcal{H}(\mathcal{T}))$*, directed edges* $E_d(\mathcal{C}(\mathcal{T})) = E(\mathcal{H}(\mathcal{T}))$ *and undirected edges* $E_u(\mathcal{C}(\mathcal{T})) = \{(w, w') : u(w) = u(w'),$ $w, w' \in V(\mathcal{C}(\mathcal{T}))\}$*. The colours of the directed edges of* $\mathcal{C}(\mathcal{T})$ *are inherited from the corresponding probability graph* $\mathcal{H}(\mathcal{T})$*.*

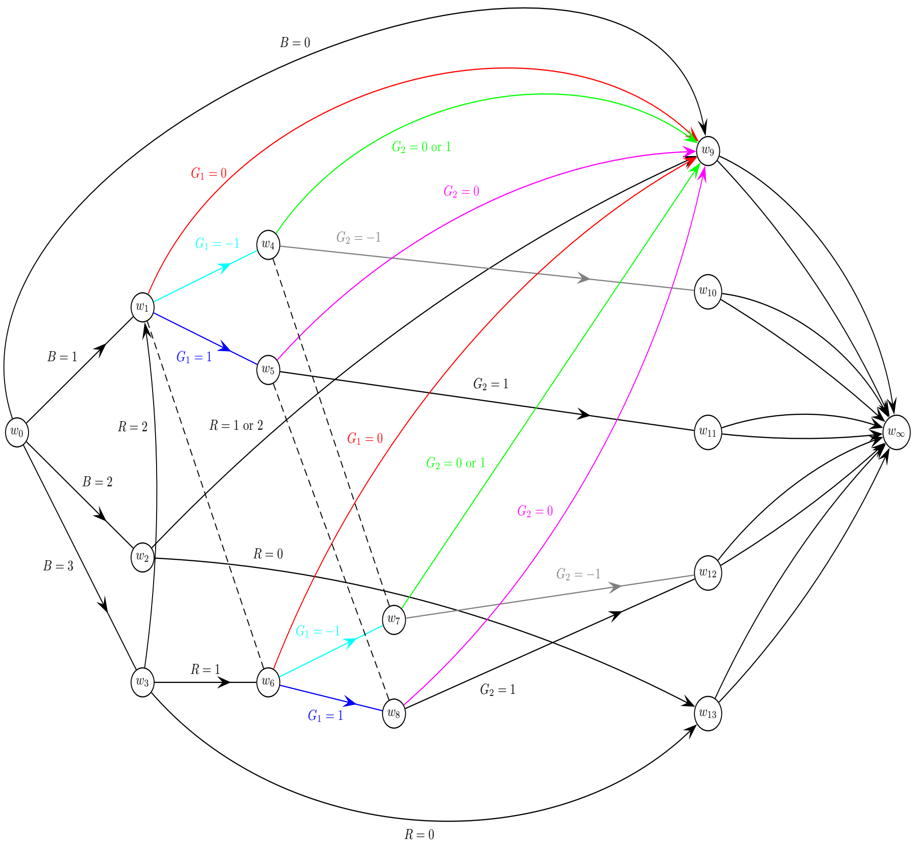

Figure 5: Chain event graph for the cell culture example. Notice the colour identification of edges from positions in the same stage. The dashed lines are the undirected edges which join situations in the same stage. The double edges from w_9, w_{10}, w_{11} and w_{12} represent $C = 0$ and $C = 1$. The positions, w_i, are used to label the nodes. See the text for an explanation of the variables.

Note that, by definition, positions connected to w_∞ in $\mathcal{C}(\mathcal{T})$ are never connected by an undirected edge. When the set of stages $L(\mathcal{T})$ equals the set of positions $K(\mathcal{T})$ of a staged tree $\mathcal{G}(\mathcal{T})$, we call $\mathcal{C}(\mathcal{T})$ *simple*. By definition, simple CEGs have no undirected edges and no coloured edges and so are acyclic, directed graphs. An example of a simple CEG can be found in the introduction. Later in the paper we will show how to read conditional independence relationships from the topology of a general CEG.

For a staged tree $\mathcal{G}(\mathcal{T})$, the pair of primitive probabilities $(\mathcal{T}, \Pi(\mathcal{T}))$ (where $\Pi(\mathcal{T}) = \{\Pi(u) : u \in L(\mathcal{T})\}$) associated with the distributions $\{\mathbf{X}(u) : u \in L(\mathcal{T})\}$ give a complete description of an event space and its associated probability model. It follows that $(\mathcal{C}(\mathcal{T}), \Pi(\mathcal{C}))$ also gives a complete specification of a probability model, where $\Pi(\mathcal{C}) = \Pi(\mathcal{T})$. So, like the BN, the CEG can be seen as a graph whose topology embodies sets of conditional independence statements and, when supplemented by a set of conditional probability distributions, can be elaborated into a full probability model. But unlike the BN, because there is an explicit invertible map between the set of directed root-to-sink paths of $\mathcal{C}(\mathcal{T})$ and the root-to-leaf paths of $\mathcal{T}$, the topology of the CEG expresses the structure of the sample space of $\mathcal{T}$ and, in particular, impossible events.

The CEG of the staged tree of the cell culture example is given in figure 5. Note that the labelling and colouring of the edges is consistent with the set of maps $\mathcal{L}(\mathcal{T})$ and that all information in the staged tree is expressed within the topology of this graph.

3 Conditional Independence in CEGs

3.1 Cuts and CEGs

As with a faithful BN, it is possible to read the various implied conditional independence statements of a staged tree directly from the topology of a CEG. We demonstrated in the introduction that because the CEG is constructed from an explanation of how situations happen (unlike the BN) there is no intrinsic set of measurement random variables over which conditional independence is defined. The random variables that explain the underlying symmetries can, however, be deduced from the topology of a CEG and its associated maps $\mathcal{L}(\mathcal{T})$. Two important constructions of these intrinsic random variables, linked to the underlying filtration represented in the event tree, are the cut and fine cut, as illustrated in figure 6.

Definition 5 *Call a collection W of positions $w \in K(\mathcal{T})$ a* fine cut *of $\mathcal{H}(\mathcal{T})$ (or $\mathcal{C}(\mathcal{T})$) if all root-to-leaf paths in $\mathcal{H}(\mathcal{T})$ pass through exactly one $w \in W$. For any fine cut W of $\mathcal{H}(\mathcal{T})$, let $\mathcal{T}^{*W}$ denote the subtree of $\mathcal{T}$ whose paths are those paths of $\mathcal{T}$ which end in a $v \in w$, for some $w \in W$. Let $\mathcal{H}(\mathcal{T}^{*W})$ and $\mathcal{C}(\mathcal{T}^{*W})$ represent the probability graph and the chain event graph of $\mathcal{T}^{*W}$, respectively.*

Definition 6 *Call a collection U of stages $u \in L(\mathcal{T})$ a* cut *of $\mathcal{H}(\mathcal{T})$ if all root-to-leaf paths in $\mathcal{H}(\mathcal{T})$ pass through exactly one $w \in u$ for some*

$u \in U$. For any cut U of $\mathcal{H}(\mathcal{T})$, let $\mathcal{T}^U$ denote the subtree of $\mathcal{T}$ whose paths are those paths of $\mathcal{T}$ which end in a $v \in u$, for some $u \in U$. Let $\mathcal{H}(\mathcal{T}^U)$ and $\mathcal{C}(\mathcal{T}^U)$ represent, respectively, the probability graph and the chain event graph of $\mathcal{T}^U$. Let $\mathbb{P}^U(\mathcal{C}(\mathcal{T}))$ denote the set of probability distributions associated with the stages of positions of $\mathcal{C}(\mathcal{T}^U)$.

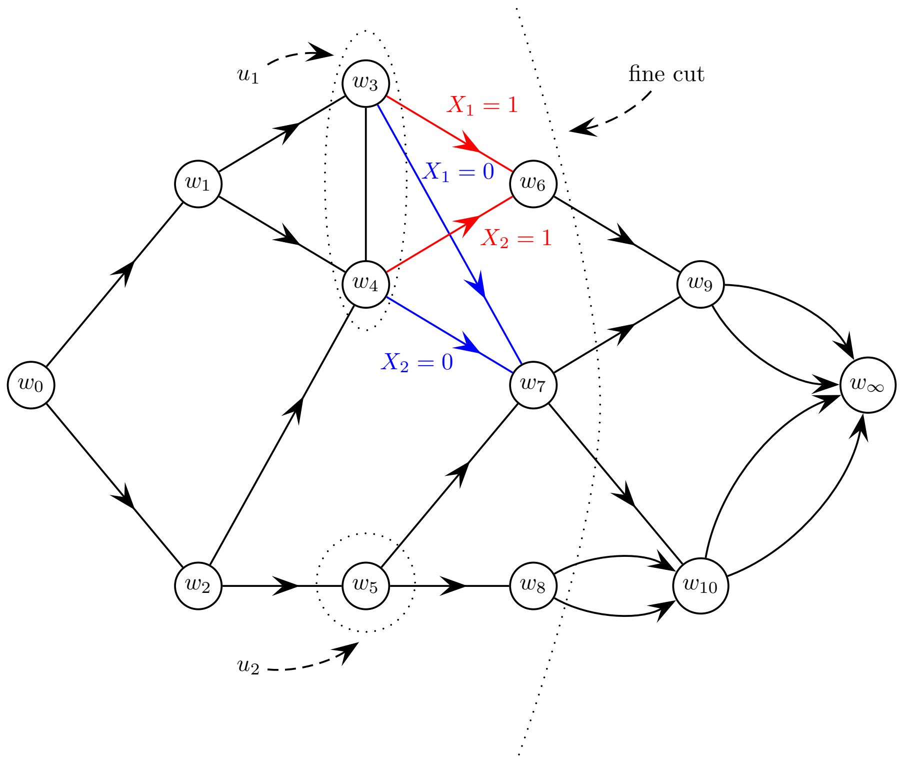

Figure 6: An example of a CEG displaying various constructions: a cut $\{u_1\} \cup \{u_2\} = \{w_3, w_4\} \cup \{w_5\}$ and a fine cut $\{w_6, w_7, w_8\}$. A fine cutting sequence is given by $\{w_0\}\cup\{w_1, w_2\}\cup\{w_2, w_3, w_4\}\cup\{w_5, w_6, w_7\}\cup \{w_6, w_7, w_8\} \cup \{w_9, w_{10}\}$. Some edges have been labelled for illustration.

For convenience, let the set consisting solely of the root vertex $\{v_0\}$ be both a cut and a fine cut. By definition, since $(\mathcal{C}(\mathcal{T}), \Pi(\mathcal{C}(\mathcal{T}))$ provides a full description of a probability model on $\mathcal{T}$, $(\mathcal{C}(\mathcal{T}^U), \Pi^U(\mathcal{C}(\mathcal{T}^U))$ provides a full description of a probability model on $\mathcal{T}^U$. In particular, the probabilities associated with each of its paths sum to unity and are expressible as a monomial in primitive probabilities in $\Pi^U(\mathcal{C}(\mathcal{T}^U))$. To explore the relationship between the graphical depiction of conditional independence in the BN and the analogous depiction in the CEG, it is first necessary to introduce further definitions.

Definition 7 *Call a sequence of fine cuts* $(W_0, W_1, W_2, \dots, W_N)$ *of* $\mathcal{H}(\mathcal{T})$, *a* fine cutting sequence *of* $\mathcal{H}(\mathcal{T})$ *if:*

1. $W_0 = \{w_0\}$, *where* w_0 *is the root vertex.*

2. *For each* $w_i \in W_i$ *there is a* $w_{i-1}(w_i) \in W_{i-1}$ *such that either* $w_{i-1}(w_i) = w_i$, *so that* w_i *lies in both* W_i *or* W_{i-1}, *or* $(w_{i-1}, w_i) \in E(\mathcal{H}(\mathcal{T}))$, *so that there is an edge from a vertex* w_{i-1} *to* w_i, $1 \leq i \leq n$.

3. *All* $v_i \in w_i$, $w_i \in W_i$, *lie on a path* $\lambda(v_0, v_{i+1})$ *in* $\mathcal{T}$ *from its root to some vertex* $v_{i+1} \in w_{i+1}, w_{i+1} \in W_{i+1}$ *or* $v_i \in w_{i+1}$, *for some* $w_{i+1} \in W_{i+1}$, $1 \leq i \leq N-1$.

4. $S(\mathcal{T}) = \bigcup_{1 \leq i \leq N} \{v \in W_i\} \cup \{w_0\}$.

Call this sequence an orthogonal fine cut *if all positions in* W_i *lie the same distance from the root position for* $1 \leq i \leq N$, *and no* $W_i = W_j$ *for* $1 \leq i, j \leq N$.

Definition 8 *Call a sequence of cuts* $(U_0, U_1, U_2, \dots, U_N)$ *of* $\mathcal{H}(\mathcal{T})$, *a* cutting sequence *of* $\mathcal{H}(\mathcal{T})$ *if:*

1. $U_0 = \{w_0\}$ *where* w_0 *is the root vertex.*

2. *For each* $w_i \in U_i$ *there is a* $w_{i-1}(w_i) \in U_{i-1}$ *such that either* $w_{i-1}(w_i) = w_i$ *or* $(w_{i-1}, w_i) \in E(\mathcal{H}(\mathcal{T}))$, $1 \leq i \leq n$.

3. *All* $v_i \in u_i, u_i \in U_i$, *either lie on a directed path* $\lambda(v_0, v_{i+1})$ *in* $\mathcal{T}$ *from its root* v_0 *to some vertex* $v_{i+1} \in u_{i+1}, u_i \in U_{i+1}$ *or are such that* $v_i \in u_{i+1}$, *for some* $u_{i+1} \in U_{i+1}$, $1 \leq i \leq N-1$.

4. $S(\mathcal{T}) = \bigcup_{1 \leq i \leq N} \{v \in u_i : u_i \in U_i\}$.

Call this sequence an orthogonal cut *if all positions in* U_i *lie the same distance from the root position* $1 \leq i \leq N$, *and no* $U_i = U_j$, $1 \leq i, j \leq N$.

Note that an orthogonal fine cut partitions the set of positions $K(\mathcal{T})$. There are three useful random variables which can be defined using the concept of a cut.

Notation 2 *Let* $\mathcal{C}(\mathcal{T})$ *be a CEG and* $\Pi(\mathcal{C})$ *be the set of probability distributions on its positions. For any cut* U, *let* $\boldsymbol{X}(U) = (X[u] : u \in U)$ *where* $X[u]$ *is the random variable associated with the stage* u. *Let* $\mathbf{Q}(U)$ *be a random vector of parents of* $\boldsymbol{X}(U)$, *whose state space is the set of* *stages $u \in U$ where the probability $\pi_{\mathbf{Q}(U)}(u)$ is the sum of all the monomials in primitives associated with paths $\lambda_u \in \Lambda_u$ from the root vertex of $\mathcal{H}(\mathcal{T})$ to an element $w \in u$. Explicitly,*

$$\pi_{\mathbf{Q}(U)}(u) = \sum_{\lambda_u \in \Lambda_u} \prod_{w \in \lambda_u, w \notin u} \pi(w'|w)$$

Let $\mathbf{Z}(U)$ denote a random variable whose state space Λ_U consists of all paths $\lambda_u \in \Lambda_u$ in $\mathcal{H}(\mathcal{T})$ from its root vertex to a vertex $w \in u$, for some $u \in U$: the upstream variable. This has an associated probability mass function $\pi_{\mathbf{Z}(U)}$ given by

$$\pi_{\mathbf{Z}(U)} = \prod_{w \in \lambda_u, w \notin u} \pi(w'|w)$$

These constructions give answers to conditional independence statements, like those embedded in BNs, that are valid for all values of the conditioning variables. By definition, once the stage u is given, or equivalently once the value of $\mathbf{Q}(U)$ is observed, the inputs to any random variable associated with a stage $u \in U$ are known. So, in particular, none of the positions in $\mathcal{H}(\mathcal{T}^U)$ can have any bearing on the realisation of $\boldsymbol{X}(U)$. Thus we have that, given a set of primitives, by construction

$$\boldsymbol{X}(U) \coprod \mathbf{Z}(U)|\mathbf{Q}(U)$$

Conversely,

Theorem 1 *If a function $B(\mathbf{Z}(U))$, where U is a cut, satisfies*

$$\boldsymbol{X}(U) \coprod \mathbf{Z}(U)|B(\mathbf{Z}(U))$$

then $\mathbf{Q}(U)$ is a function of $B(\mathbf{Z}(U))$ with probability one.
Proof. *Suppose $\mathbf{Q}(U)$ is not a function of $B(\mathbf{Z}(U))$ with probability one. Then there exist two positions $w[1]$ and $w[2]$ in different stages ($u[1]$ and $u[2]$ respectively) which each have non-zero probability and for which $X(w[1]) = X(w[2])$. But this would imply that $w[1]$ and $w[2]$ were at the same stage, giving a contradiction.* ■

The theorem above also tells us that these are the *only* independencies between upstream and downstream random variables defined on the path event space that can be deduced from the CEG of a staged tree $\mathcal{T}$.

3.2 Homogeneous staged trees and the BN

In this paper we focus much of our attention on event trees that are *n-homogeneous*: that is, all their root-to-leaf paths are of length n edges. One important n-homogeneous event tree is compatible with finite discrete random variables $X_1, X_2, \ldots, X_n$ taking values on a subset of the product event space $\{\mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_n\}$ where each root-to-leaf path of $\lambda \in \Lambda$ corresponds to an event of the form $\cap_{i=1}^{n}\{X_i = x_i\}$. An example of such a tree on two binary variables is given in the introduction.

Suppose an observer's beliefs are fully and accurately given by a BN $\mathcal{D}$. Suppose this is unknown to the analyst who constructs the client's event tree $\mathcal{T}$ and then its CEG $(\mathcal{C}(\mathcal{T}), \Pi)$. It is shown below that the underlying BN is identified from the CEG $\mathcal{C}(\mathcal{T})$ alone: the primitive probabilities of $(\mathcal{C}(\mathcal{T}), \Pi)$ project directly on to the primitive probabilities of $\mathcal{D}$.

When a staged tree represents all the conditional independence statements depicted in a BN, its stages $L(\mathcal{C}(\mathcal{T}))$ must be in one-to-one correspondence with the different possible configurations $\mathbf{q}_i$ of the $n-1$ parent sets $\mathbf{Q}_i$ of the random variables X_i, $2 \leq i \leq n$, with the root vertex of $\mathcal{C}(\mathcal{T})$ being associated with X_1. Let the set of stages U_{i-1} label the different possible configurations $\mathbf{q}_i$ of parents of X_i, $2 \leq i \leq n$. Clearly, each of the sets U_{i-1} , forms a cut in $\mathcal{C}(\mathcal{T})$ for $1 \leq i \leq n-1$, and furthermore $(U_0, U_1, \ldots, U_{n-1})$ is a cutting sequence of $\mathcal{C}(\mathcal{T})$.

One of many equivalent definitions [17, 27] of a valid Bayesian network $\mathcal{D} = (V(\mathcal{D}), E(\mathcal{D}))$, with $V(\mathcal{D}) = \{X_1, X_2, \ldots, X_n\}$ is that $(X_j, X_i) \in E(\mathcal{D}) \Leftrightarrow X_j \in \mathbf{Q}_i$ where $1 \leq j < i$ and, for all configurations $\mathbf{q}_i$ of the parents $\mathbf{Q}_i$ of X_i,

$$X_i \coprod \{X_1, X_2, .., X_{i-1}\} | \mathbf{Q}_i = \mathbf{q}_i$$

It has already been noted that the associated tree $(\mathcal{T}(\mathcal{D}), \Pi(\mathcal{T}(\mathcal{D}))$ and hence its CEG $(\mathcal{C}(\mathcal{D}), \Pi(\mathcal{C}(\mathcal{D}))$ gives a full description of this probability model. So $(\mathcal{T}^{U_i}(\mathcal{D}), \Pi(\mathcal{T}^{U_i}(\mathcal{D}))$ (or equivalently $(\mathcal{C}(\mathcal{T}^{U_i}(\mathcal{D}), \Pi(\mathcal{C}^{U_i}(\mathcal{D})))$ has its root-to-leaf paths — the sample space of $\mathbf{Z}(U_i)$ — labelling the set of events $\{X_1 = x_1, X_2 = x_2, .., X_i = x_i\}$.

By definition, whenever a situation labelled by the path $\{X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i\}$ corresponds to the same configuration of parents of X_{i+1}, it will be placed in the same stage in U_i. Different configurations of parents of X_i will correspond to different events $\{\mathbf{Q}_i = \mathbf{q}_i\}$ where $\mathbf{Q}_i = \mathbf{Q}(U_{i+1}), 1 \leq i \leq n-2$. It follows that for each $u_i \in U_i$, $0 \leq i \leq n-1$,

$$\mathbf{X}(U_i) \coprod \mathbf{Z}(U_i) | \mathbf{Q}(U_i)$$

Note here that $X(U_i)$ can be identified with $X_{i+1}|\mathbf{Q}_{i+1}$, $1 \leq i \leq n$, $\mathbf{Q}(U_i)$ with $\mathbf{Q}_{i+1}$ and $\mathbf{Z}(U_i)$ with the set of random variables $\{X_1, \ldots, X_i\} \setminus \mathbf{Q}_{i+1}$. Henceforth we will identify edges emanating from the positions of v_i of the CEG on an n-homogeneous tree $\mathcal{T}$ a distance $i-1$ from the root of $\mathcal{T}$ by $X_i|\mathbf{Q}_i$, $2 \leq i \leq n$, the collection of such edges by X_{i+1} and $\mathbf{X}(W_0)$ by X_1.

Since the stages of a CEG can be identified visually, it follows that all the conditional independencies that are needed to build the associated faithful BN $\mathcal{D}$ can be read directly from the chain event graph $\mathcal{C}(\mathcal{T}(\mathcal{D}))$. Note here that each of the cuts consists of stages, all of whose positions are the same number of edges from the root vertex. They are therefore trivial to identify from $\mathcal{C}(\mathcal{T}(\mathcal{D}))$. Note also that the sequence of cuts defined by the the parent configurations $(U_1, U_2, \ldots, U_{n-1})$ provide an orthogonal cutting sequence of $\mathcal{C}(\mathcal{T}(\mathcal{D}))$.

3.3 Event conditioned independence in CEGs

A position w^* of $\mathcal{C}$ is called a *stalk* if every root-to-sink path $\lambda \in \Lambda(\mathcal{C})$ passes through w^*. A stalk is a fine cut that is also a singleton and has a particularly important role in the interpretation of a CEG. Because $\mathcal{H}(\mathcal{T})$ is acyclic, all paths in $\mathcal{C}$ pass through its stalks in the same order. Label the m stalks $\{w_0, w_1^*, w_2^*, \ldots, w_m^*\}$ consistently with this order.

Definition 9 *A* shelling *of a CEG* $(\mathcal{C}, \Pi(\mathcal{C}))$ *into* peas $\{(\mathcal{C}_i, \Pi^{(i)}(\mathcal{C}_i)) : 1 \leq i \leq m\}$ *is a map*

$$(\mathcal{C}, \Pi(\mathcal{C})) \;\rightarrow\; \{(\mathcal{C}_1, \Pi^{(1)}(\mathcal{C}_1)), (\mathcal{C}_2, \Pi^{(2)}(\mathcal{C}_2)), \ldots, (\mathcal{C}_m, \Pi^{(m)}(\mathcal{C}_m)\}$$

where

1. *The vertex set* $V(\mathcal{C}_1)$ *of the mixed graph* $\mathcal{C}_1$ *is the set of positions* $\{w \in K(\mathcal{C}) : w \prec w_1^*\}$ *together with its sink vertex* w_1^*. *The vertex set* $V(\mathcal{C}_i)$ *of the mixed graph* $\mathcal{C}_i$ *is the set of positions* $\{w \in K(\mathcal{C}) : w_{i-1}^* \preceq w \prec w_i^*\}$ *together with its sink vertex* w_i^* *where we use the convention that* $w_m^* = w_\infty$. *For* $1 \leq i \leq m$, *if* $E(\mathcal{C})$ *denotes the edge set of* $\mathcal{C}$ *then the edge set* $E(\mathcal{C}_i)$ *is defined by*

$$e \in E(\mathcal{C}_i) \Leftrightarrow e \in E(\mathcal{C})$$

2. *The primitive probabilities* $\pi_i(v_i'|u_i) \in \Pi^{(i)}(\mathcal{C}_i)$ *of* $X_i(u_i)$ *are such that* $\pi_i(v_i'|u_i) = \pi(v_i'|u_i) \in \Pi(\mathcal{C})$.

Theorem 2 *Suppose a CEG has at least two peas. The random vector* $(Y_i, Y_{i+1}, \ldots, Y_m)$ *whose sample space can be identified with a set of paths* *through vertices in* $\cup_{k=i}^{m} V(\mathcal{C}_k)$ *is then independent of the random vector* $(Y_1, Y_2, \ldots, Y_{i-1})$ *whose sample space is defined by paths through vertices in* $\cup_{k=1}^{i-1} V(\mathcal{C}_k)$*,* $2 \leq i \leq m$*.*

Proof. *Each atom of the* σ*-field associated with* $(Y_1, Y_2, \ldots, Y_{i-1}) \times (Y_i, Y_{i+1}, \ldots, Y_m) = (Y_1, Y_2, \ldots, Y_m)$ *corresponds to a root-to-leaf path* λ *in* $\mathcal{C}$*. By definition,* $\lambda = (w_0, w_{1,\lambda}, \ldots, w_{t(\lambda),\lambda} = w_{i-1}^*, \ldots, w_{n(\lambda)})$ *must pass through the position* w_{i-1}^**. Let* $\Lambda_1(\lambda)$ *denote the set of all paths that agree with* λ *until the position* w_{i-1}^* *and are otherwise arbitrary. Let* $\Lambda_2(\lambda)$ *denote the set of all paths that are arbitrary until reaching* w_{i-1}^* *but that agree after* w_{i-1}^**. By definition*

$$\pi(\lambda) = \prod_{i=1}^{n(\lambda)} \pi(w_{i,\lambda} | w_{i-1,\lambda})$$

Clearly, by definition and the fact that all paths pass through w_{i-1}^*

$$P(\lambda \in \Lambda_1(\lambda)) = \prod_{i=1}^{t(\lambda)} \pi(w_{i,\lambda} | w_{i-1,\lambda})$$

whilst, by the same argument,

$$P(\lambda \in \Lambda_2(\lambda)) = \prod_{i=t(\lambda)+1}^{n(\lambda)} \pi(w_{i,\lambda} | w_{i-1,\lambda})$$

It follows that

$$\pi(\lambda) = P(\lambda \in \Lambda_1(\lambda)) P(\lambda \in \Lambda_2(\lambda))$$

Since this is true for all atoms in the space, the theorem is proved. ∎

This result is important, since it is now possible to immediately identify a CEGs independent components (its peas) visually from its stalks. Thus from the topology of the CEG of figure 7, we can immediately identify three mutually independent random variables (Y_1, Y_2 and Y_3) on the event space of its 18 root-to-sink paths. We have that $Y_1 = y_1(1)$ when a path contains e_1 and e_3, $Y_1 = y_1(2)$ when a path contains e_1 and e_4 and $Y_1 = y_1(3)$ when a path contains e_2. $Y_2 = y_2(1)$ when a path passes through e_6, $Y_2 = y_2(2)$ when it passes through e_7 and $Y_2 = y_2(3)$ when it passes through e_8. Finally, $Y_3 = y_3(1)$ when a path passes through e_9 and $Y_3 = y_3(3)$ when it passes through e_9. In fact, the theorem also allows us to identify certain conditional independence statements,

Let $\Lambda[\boldsymbol{w}] \subset \Lambda$ denote the event in the path σ-algebra Λ of $\mathcal{C}$ consisting of the set of all paths $\lambda \in \Lambda$ passing through $\boldsymbol{w}^* = (w_1^*, w_2^*, \ldots, w_{m-1}^*)$,

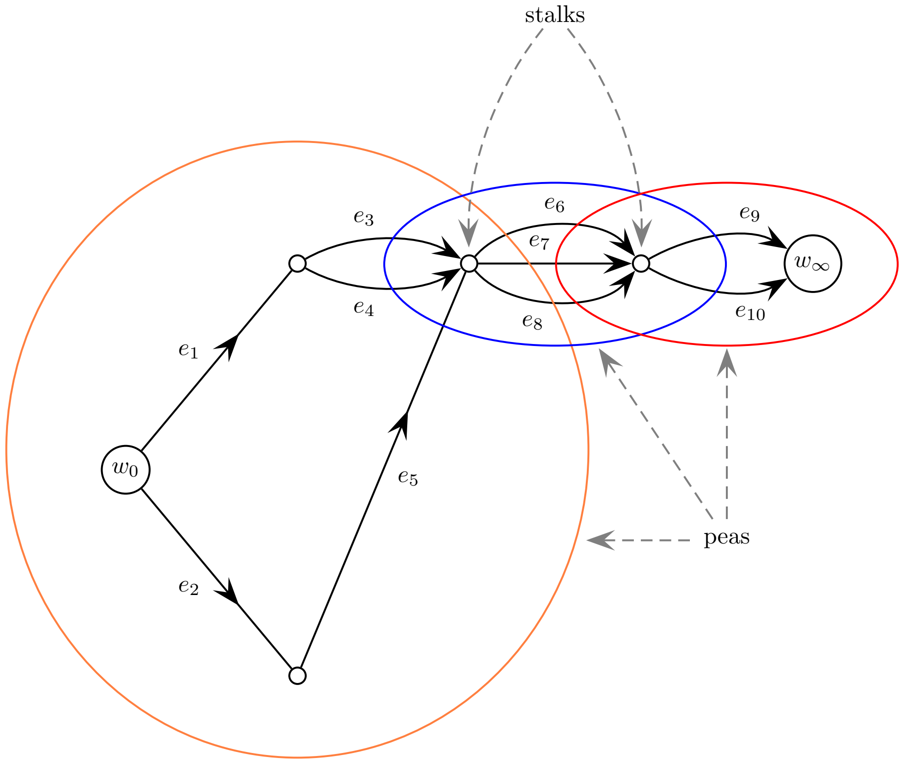

Figure 7: A CEG with three peas.

where $w_i^* \in K(\mathcal{C}(\mathcal{G}))\backslash\{w_0\}$, $1 \leq i \leq m-1$ and, for $k > 1$, $w_i^* \prec w_{i+1}^*$, $1 \leq i \leq k-1$. Let $\mathcal{G}_{\Lambda[\mathbf{w}^*]}(\mathcal{T}_{\Lambda[\mathbf{w}^*]}, L(\mathcal{T}_{\Lambda[\mathbf{w}^*]}), \mathcal{L}(\mathcal{T}_{\Lambda[\mathbf{w}^*]}))$ denote the subtree of a staged tree $\mathcal{G}$ whose tree is $\mathcal{T}_{\Lambda[\mathbf{w}^*]}$ — the subtree of $\mathcal{T}$ whose paths are $\Lambda[\boldsymbol{w}^*]$ — and which inherits its stages and staging bijections from $\mathcal{G}(\mathcal{T}, L(\mathcal{T}), \mathcal{L}(\mathcal{T}))$.

Corollary 1 *If $\mathcal{C}(\mathcal{T}_{\Lambda[\mathbf{w}^*]})$ has stalks $\{w_0, w_1^*, w_2^* \ldots w_{m-1}^*\}$ with $\{Y_i(\boldsymbol{w}^*) : 1 \leq i \leq m\}$, as defined in theorem 2 but with $(\mathcal{C}(\mathcal{T}_{\Lambda[\mathbf{w}^*]}), \Pi(\mathcal{C}(\mathcal{T}_{\Lambda[\mathbf{w}^*]}))$ replacing $(\mathcal{C}, \Pi(\mathcal{C}))$, then*

$$\coprod_{i=1}^{m} Y_i | \Lambda[\boldsymbol{w}^*]$$

Proof. *This is immediate from the observations at the end of section 2.2 that $\mathcal{C}(\mathcal{T}_{\Lambda[\mathbf{w}^*]})$ is in fact the CEG of the tree $\mathcal{T}_{\Lambda[\mathbf{w}^*]}$ conditioned on the event $\Lambda[\boldsymbol{w}^*]$ and that, under our formula, the necessary separation of probabilities in $\Pi(\mathcal{C}(\mathcal{T}_{\Lambda[\mathbf{w}^*]})$ still holds.* $\blacksquare$

This allows many conditional independence statements to be read from the CEG $\mathcal{C}(\mathcal{T})$ of the unconditioned tree $\mathcal{T}$, when the conditioning is on an event rather than a variable.

For example, consider the cell culture CEG of figure 5. Suppose we take a measurement that tells us that there is possible disruption of epistatic interaction but if cell damage has occurred it is quickly repaired (position w_1) and that these circumstances preclude a larger than usual probability of an increase in cancer cells (position w_8). Then $\Lambda[w_1, w_8]$ is the set of 12 paths in Λ passing through w_1 and w_8. The corollary allows us to construct three random variables Y_1, Y_2 and Y_3 which are mutually independent *given* the event $\Lambda[\boldsymbol{w}^*]$. Thus Y_1 is determined given $\Lambda[\boldsymbol{w}^*]$, whether or not cell damage has occurred (i.e. which path from w_0 to w_1 is taken: $\{B = 3, R = 1\}$ or $\{B = 1\}$). Variable Y_2 labels which of the three passive switching pathways $\{G_1 = 0\}, \{G_1 = -1, G_2 \neq -1\}, \{G_1 = 1, G_2 \neq 1\}$ is taken: the legal paths between w_1 and w_8 given what we have learned. Finally, Y_3 is an indicator of whether or not there has been a cancerous increase: represented by the two paths from w_8 to w_∞.

Obviously many other examples of event conditioned conditional independence can be read from this CEG. They allow us to address interesting implications of this staged tree without demanding the construction of a conditioning random variable: a construction that is needed in the interrogation of a BN. In contrast, all we need is a conditioning *event*: here $\Lambda[w^*]$. It is argued in [23] that this is intrinsic to implications associated with causal models.

For the remainder of the paper we return to more familiar conditional independence relationships and investigate how cuts and fine cuts of the CEG of a a staged tree can be used to identify pairs of variables that are independent of each other given a third.

3.4 Constructing variables from CEGs: a simple example

In the last example, through naming the positions and edges of a CEG we created a semantics within which we could construct variables that exhibited conditional independence over an event. Here we demonstrate a similar simple method for finding conditional independencies over variables using cuts.

An explorer in a forest may die tomorrow. There is a possibility that she is bitten by a venomous snake. If she carries an antidote and uses it, such a bite will certainly not be deadly and will have no effect on her health. But without the antidote, the probability that she will die tomorrow will increase.

Naively constructing a BN of this scenario encourages us to represent the dependence structure in terms of the three indicators we could measure: X_1 (whether she is bitten), X_2 (whether she carries the antidote) and X_3 (whether she dies tomorrow). Unfortunately, the story as relayed above would simply give a degenerate (complete) BN. Indeed, the only possibly plausible additional conditional independence that might be added to the story is that $X_1 \coprod X_2$, but this looks suspicious since if she is more likely to be bitten then, unless she is very ignorant, she is more likely to decide to carry the antidote.

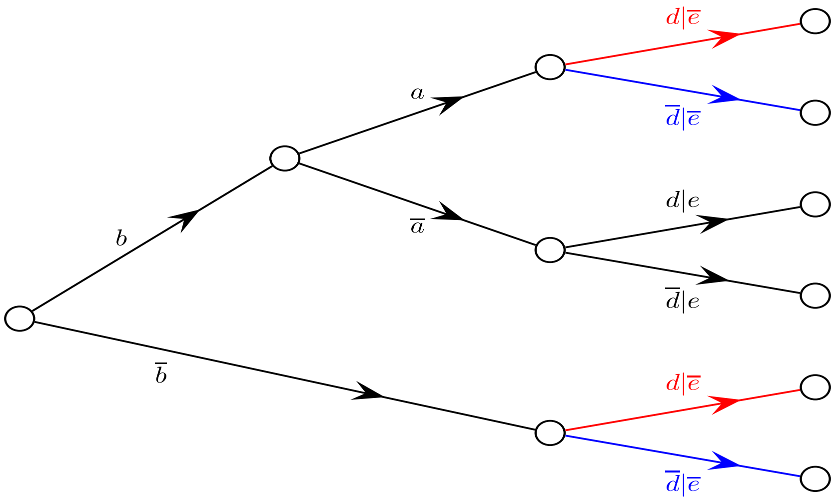

Figure 8: The event tree for the snake bite example. b is the event that she gets bitten, a that she is carrying the antidote, d that she dies tomorrow and $e \equiv b \cap \overline{a}$ that she is endangered. An over-line denotes the complement.

However, once the event tree and then the CEG of this scenario is drawn, see figures 8 and 9 respectively, random variables that might exhibit conditional independencies can be automatically derived from the cuts and fine cuts of the CEG.

Having drawn the CEG and acknowledged that two situations in the tree can be combined into a single position w_3, it is easy to see that the interior positions (w_1, w_2, w_3) represent (bitten, endangered by bite, not endangered). The edges (w_0, w_1) and (w_0, w_3) hold the primitive probability of being bitten, $\pi[b]$, or not, $\pi[\overline{b}]$; (w_1, w_2) that she does not carry the antidote when bitten, $\pi[\overline{a}]$; (w_1, w_3) that she carries the antidote when bitten, $\pi[a]$; the two edges from w_2 to w_∞ whether she dies, $\pi[d|e]$, or not, $\pi[\overline{d}|e]$, if endangered; and the two edges from w_3 to w_∞ whether she dies, $\pi[d|\overline{e}]$, or not, $\pi[\overline{d}|\overline{e}]$, when there is no effect from the bite. Clearly she is endangered only if b and $\overline{a}$ have occurred, and not endangered if b and a, or just $\overline{b}$ have taken place.

Note that because $\mathcal{C}(\mathcal{T}) = \mathcal{H}(\mathcal{T})$, positions and stages are identical and consequently the primitive probabilities of the problem can be as-

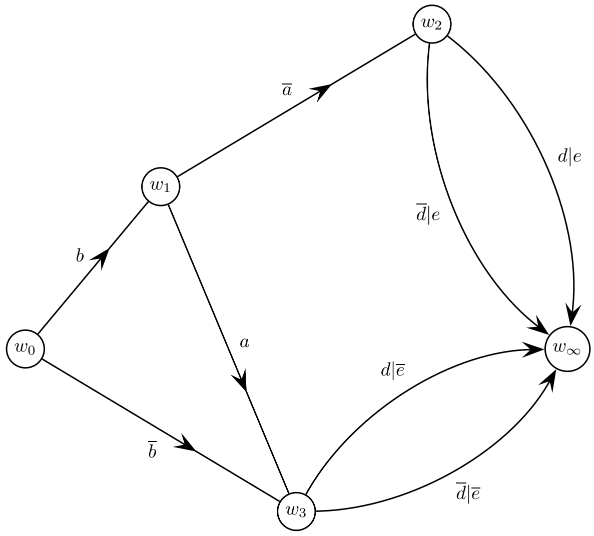

Figure 9: The CEG for the event tree of the snake bite example shown in figure 8.

signed uniquely to the edges of this graph. Our sum to one conditions reduce these eight probabilities to four functionally independent values. The only non-degenerate cut here is that provided by $U = \{w_2, w_3\}$. From the results above it is possible to read directly from figure 9 that

$$\mathbf{X}(U) \coprod \mathbf{Z}(U)|\mathbf{Q}(U)$$

where $\mathbf{Q}(U)$ is the indicator of whether or not the individual is endangered by the snake bite, $\mathbf{X}(U) = X_3$ is the indicator of whether or not the individual dies, and $\mathbf{Z}(U)$ is an indicator differentiating between the event that the person was bitten and then took the antidote, and the event that she was not bitten. Thus, the conditional independence embedded in the CEG is over these three variables, all of which can be constructed directly from the graph.

The substantive statement being made by the observer and encapsulated by this conditional independence statement is that the fact that the person was not bitten, or bitten and then given the antidote, is irrelevant to predictions about her probability of death tomorrow.

Following the common practice of simply searching over dependence structures between these three variables, either from an elicitation process or a search over a sample space, will fail to detect this structure. But tracing how events might happen leads us to appropriate random variables which do express any exchangeability. In general, any non-degenerate cut corresponds to a substantive conditional independence statement associated with the description of the problem as captured by the CEG. Furthermore, an associated parent variable $\mathbf{Q}$ and residual variable $\mathbf{Z}$ can be visually identified, and subsequently interpreted, via the original description from the client.

Note that the BN demands that all cuts can be expressed as invertible functions of a subset of the measurement vector whose sample space defines $\mathbf{Z}(U)$. This fierce invertibility condition is completely unnecessary in the CEG: $\mathbf{Q}(U)$ can be *any* function of the space determined by the previously listed variables. The only implicit condition on the examined functions $\mathbf{Q}(U)$ is that they must be consistent with a natural order of an associated tree, i.e., consistent with the partial order in which the client believes situations will take place. This covers a large class of models. Indeed Shafer, [25] would appear to assert that these are the *only* conditional independence statements that one can reasonably expect to elicit from a client by direct questioning. Certainly when the whole of the client's beliefs are captured by a single "causally" faithful event tree, Shafer's assertion appears a compelling one.

4 Dependence Enquiries Using CEGs

4.1 Fine cuts and conditional independence concerning the past given the future

Identifying cuts allows us to define independence structures associated with subsequent unfoldings of situations on a tree. However, fine cuts address global independence statements associated with a graph. In particular, they allow deductions to be made about conditional independencies of causes given effects in the same way as the constructions associated with d-separation in BNs. This is important since it is common to observe effects but not causes. For example, a doctor sees a patient's symptoms but she is interested in observing the disease itself.

To address this type of enquiry, three random variables associated with a fine cut on a CEG must be defined.

Notation 3 *Let $(\mathcal{C}(\mathcal{T}), \Pi)$ be a CEG. Let $\mathcal{H}(\mathcal{T}(w))$, $w \in K(\mathcal{T})$, denote the subgraph $\mathcal{H}(\mathcal{T})$ whose root-to-leaf paths are exactly those paths in $\mathcal{H}(\mathcal{T})$ beginning at w and ending at w_∞. Let $\boldsymbol{X}(\mathcal{H}(\mathcal{T}(w)))$ be the random vector with event space atoms consisting of all these paths from w to w_∞ and write*

$$\boldsymbol{X}[W] = (X(\mathcal{H}(\mathcal{T}(w)) : w \in W)$$

for the vector of such variables associated with a fine cut W. Let $\mathbf{Z}(W)$ denote the random variable whose state space Λ_W consists of all paths

$\lambda(w_0, w')$ *in* $\mathcal{H}(\mathcal{T})$ *from its root vertex to the vertex* $w' \in W$*. The associated probability* $\pi_{\mathbf{Z}(W)}$ *is given by*

$$\pi_{\mathbf{Z}(W)}(\lambda(w_0, w')) = \prod_{w \in \lambda_u, w \neq w'} \pi(w'|w)$$

Let $\mathbf{Q}(W)$ *be the random variable whose state space is the set of positions* $w' \in W$ *and the probability* $\pi_{\mathbf{Q}(W)}(w')$ *is the sum of monomials in the primitives associated with all paths* $\lambda(w_0, w')$ *from the root vertex of* $\mathcal{H}(\mathcal{T})$ *to* w'*. We then call* $\mathbf{Q}(W)$ *(a function of* $\mathbf{Z}(W)$*), the* separator *of* $\boldsymbol{X}(W)$ *from* $\mathbf{Z}(W)$*. Explicitly,*

$$\pi_{\mathbf{Q}(W)}(w') = \sum_{\lambda(w_0,w')} \prod_{w \in \lambda_u, w \neq w'} \pi(w'|w)$$

These constructions allow us to move directly from the geometry of $\mathcal{H}(\mathcal{T})$, or $\mathcal{C}(\mathcal{T})$, to large collections of conditional independence relationships between vectors of functions of random variables which, as for the BN, can be read from the separation properties of the graph $\mathcal{H}(\mathcal{T})$. Thus, for any fine cut W we can assert immediately from the construction above that

$$\boldsymbol{X}[W] \coprod \mathbf{Z}(W)|\mathbf{Q}(W) \tag{7}$$

From the usual conditional independence algebra, we can deduce from equation (7) that, for a vector function of the downstream variables, $\boldsymbol{g}(\boldsymbol{X}[W])$, given each possible position $w \in W$ learned through observing $\mathbf{Q}(W)$:

$$(\boldsymbol{X}[W], \boldsymbol{g}(\boldsymbol{X}[W]) \coprod \mathbf{Z}(W)|\mathbf{Q}(W)$$

and therefore through symmetry and weak union [18]:

$$\boldsymbol{X}[W] \coprod \mathbf{Z}(W)|\mathbf{Q}(W), g(\boldsymbol{X}[W])$$

Thus, *even after we observe* a function $\boldsymbol{g}(\boldsymbol{X}[W])$, any function $T = \boldsymbol{h}(\mathbf{Z}(W))$ of $\mathbf{Z}(W)$ (a random vector measurable with respect to Λ_W) remains uninformative about any function $Y = \boldsymbol{f}(\boldsymbol{X}[W])$ downstream of $\mathbf{Q}[W]$ if we know the value of $\mathbf{Q}[W]$. So, if we need to learn about Y there is no point in learning or remembering the value of T given $\mathbf{Q}$: the value of $\mathbf{Q}$ will suffice.

This fact is useful both for designing efficient sampling schemes for Y and for developing efficient propagation algorithms which we are currently developing. Note that the corresponding observation is *not* necessarily true for the separator variable of cuts $\mathbf{Q}(U)$. Hence refining cuts into fine cuts to obtain general separation criteria for a CEG is analogous to coarsening a BN by adding edges in the moralisation step of the d-separation theorem.

In fact, there is also an important converse to this observation. Let $\mathbf{Q}(W)$ be the separator associated with an arbitrary cut W. Let $\mathbf{Q}^*(W)$ be a function of $\mathbf{Q}(W)$ which is not a cut. By definition, this implies that it is possible to find two positions $w_1, w_2 \in W$, such that $\mathbf{Q}(w_i) = \mathbf{q}_i$ $(i = 1, 2)$ are distinct values for which the joint distributions of $(\boldsymbol{X}[W]|\mathbf{Q}(W) = \mathbf{q}_1, \mathbf{Q}^*(W))$ and $(\boldsymbol{X}[W]|\mathbf{Q}(W) = \mathbf{q}_2, \mathbf{Q}^*(W))$ are not identical. It immediately follows that

$$\boldsymbol{X}[W] \coprod \mathbf{Q}(W)|\mathbf{Q}^*(W)$$

cannot hold, and hence in particular that

$$\boldsymbol{X}[W] \coprod \mathbf{Z}(W)|\mathbf{Q}^*(W)$$

is also false. As a consequence, *all* functions $\mathbf{Q}$ of upstream variables which on conditioning make all downstream variables independent of upstream variables must define fine cuts.

4.2 Dependence enquiries on BNs using CEGs

If a CEG can be expressed as a faithful BN and a dependence enquiry only concerns the relationship between subsets of the variables of the BN, then the simplest and recommended method for answering the enquiry is to use the d-separation theorem. However, it is instructive to see how an n-homogeneous CEG (see section 3.2) and its cuts can be used as an alternative way of answering these queries. Consider the BN given in figure 10.

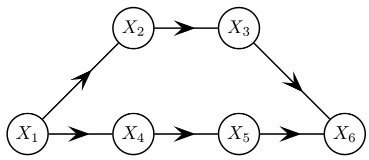

Figure 10: An example of a BN on which we wish to make an enquiry. X_6 can be observed, but our variable of interest, X_2, cannot.

The variable X_6 can be observed, but not X_2, which is our variable of interest. The values of which remaining variables can be ignored without loss? To answer this question using d-separation, the undirected version of the moralised ancestral graph must be constructed, see figure

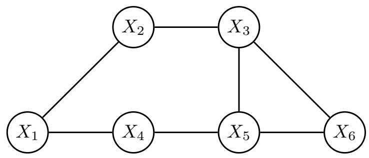

Figure 11: The undirected moralised graph of figure 10. d-separation can be used to deduce conditional independencies.

11. Clearly, $\{X_1, X_3, X_6\}$ separate X_2 from the other variables $\{X_4, X_5\}$ and so $\{X_4, X_5\}$ give no useful additional information about X_2 over that given by $\{X_1, X_3, X_6\}$. Furthermore, discarding a variable from the subsets $\{X_1, X_6\}$ and $\{X_3, X_6\}$ will inevitably lose information.

Now construct a CEG of the BN. The five fine cuts, defined by positions associated with a tree that introduces situations compatible with variables in the order $\{X_1, X_2, X_3, X_4, X_5, X_6\}$, are defined by functions of $(\{X_1\}, \{X_1, X_2\}, \{X_1, X_3\}, \{X_3, X_4\}, \{X_4, X_5\})$. This is illustrated in figure 12. The third fine cut gives us that

$$\{X_4, X_5, X_6\} \coprod \{X_1, X_2, X_3\} | \{X_1, X_3\} \tag{8}$$

implying

$$\{X_4, X_5, X_6\} \coprod X_2 | \{X_1, X_3\}$$

and thus

$$\{X_4, X_5\} \coprod X_2 | \{X_1, X_3, X_6\}$$

This is the same irrelevance statement obtained from the d-separation theorem above.

The only other substantive conditional independencies that can be read from the fine cut and are also readable from d-separation are

$$\{X_5, X_6\} \coprod \{X_1, X_2, X_3, X_4\} | \{X_3, X_4\} \tag{9}$$

$$\{X_5, X_6\} \coprod \{X_1, X_2\} | \{X_3, X_4\} \tag{10}$$

Equation (10) is derived from the property of decomposition [18] and implies, by symmetry and weak union, that

$$X_5 \coprod \{X_1, X_2\} | \{X_3, X_4, X_6\}$$

Because a CEG focuses on relationships between upstream and downstream variables, and the definition of upstream and downstream is partly a function of the underlying tree, it is not always possible to

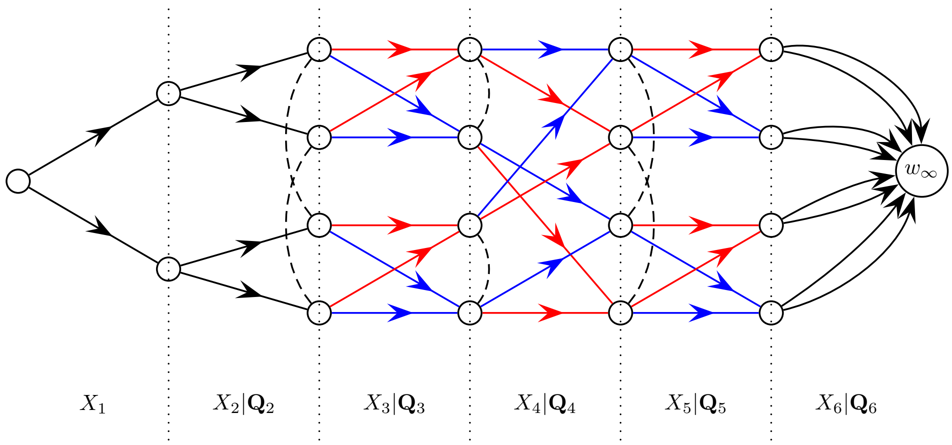

Figure 12: The CEG with cuts for the Bayesian network shown in figure 10. The dashed lines are the undirected edges. The variables X_i and $\mathbf{Q}_i$, $1 \leq i \leq 6$, are defined in section 3.2. For example, for the two vertices and four edges labelled $X_2|\boldsymbol{Q}_2$, the vertices correspond to the configurations of the parent X_1, and the edges correspond to the possible values X_2 can take given these two possible configurations.

read all implied conditional independencies from the cuts and fine cuts of a single CEG of a BN. It is in fact sometimes necessary to repeat the procedure above on a subset of different trees, all compatible with the BN, before deriving a complete list using intersection and conditioning.

For example, the tree taking variables into the CEG in total order $\{X_1, X_4, X_5, X_2, X_3, X_6\}$, gives the analogous statements

$$\{X_2, X_3\} \coprod X_4 | \{X_1, X_5, X_6\}$$
$$X_3 \coprod \{X_1, X_4\} | \{X_3, X_4, X_6\}$$

We conjecture that, in general, all substantive statements implicit in a BN can be generated by searching through a small subset of all compatible trees and evoking the properties of symmetry, decomposition and weak union. Alternatively, we can use somewhat more complex topological structure and edge colouring: see [22] for some analogues of such constructions.

4.3 Representing BNs with compact CEGs

An event tree can define rich classes of conditional independencies via the sets of random vectors $\{\boldsymbol{X}(U_i), \mathbf{Z}(U_i), \mathbf{Q}(U_i) : 2 \leq i \leq n\}$ associated with the cuts/fine cuts of a CEG $\mathcal{C}(\mathcal{T}(\mathcal{D}))$. In particular, the cuts and fine cuts define random variables $\mathbf{Q}(U_i)$ intrinsic in separating downstream variables that might be observed from upstream variables whose distributions are of interest. Because the event tree can be more expressive than the BN and is often not unique, different CEGs describing the same problem can highlight different sets of conditional independencies.

Due to the fine cut property discussed above, the most useful CEGs for interrogation purposes are ones where the sizes of stages have as small a number of positions in them as possible. It is therefore of some interest to find out when a given BN has a compact CEG representation.

Suppose four binary variables respect the BN in figure 13. A tree compatible with the total order (X_1, X_2, X_3, X_4) gives the CEG in figure 14.

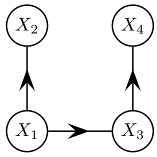

Figure 13: A Bayesian network that can be represented differently by event trees and CEGs according to the order in which you take the four variables. See figure 14.

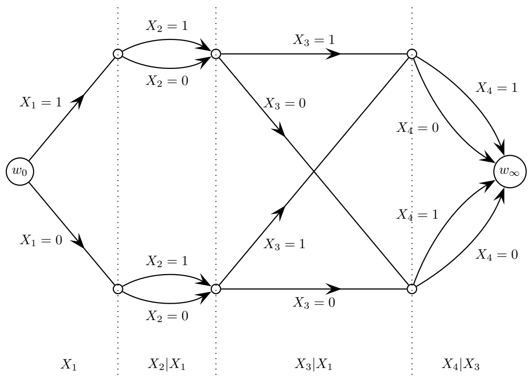

Figure 14: The CEG corresponding to the Bayesian network in figure 13, taking the variables in the order (X_1, X_2, X_3, X_4).

It is straightforward to characterise those BNs with simple representations: that is, those whose CEGs need no undirected edges. Hence, the set of fine cuts and the set of cuts are identical. Let the vertices $(X_1, X_2, \ldots, X_n)$ of a BN, $\mathcal{D}$, be such that $\mathbf{Q}_i \subseteq \{X_1, X_2, \ldots, X_{i-1}\}$ are the parents of X_i for $2 \leq i \leq n$, and write $\mathbf{R}_i = \{X_1, X_2, \ldots, X_{i-1}\} \backslash \mathbf{Q}_i$ for $2 \leq i \leq n$.

Definition 10 *A BN, $\mathcal{D}$, is said to be* moral *if all its parent sets $\mathbf{Q}_i$, $1 \leq i \leq n$, are complete.*

Theorem 3 *A BN, $\mathcal{D}$, with variables $(X_1', X_2', \ldots, X_n')$ can be represented as a simple CEG if and only if there is a permutation of the components $(X_1', X_2', \ldots, X_n') \mapsto (X_1, X_2, \ldots, X_n)$ such that*

$$\mathbf{R}_i \subseteq \mathbf{R}_{i+1} \text{ for } 2 \leq i \leq n-1$$

Proof. *Let $u(\mathbf{Q}_j = \mathbf{q}_j)$ denote the stage associated with each configuration of the parents of each random variable X_j, $1 \leq j \leq n$. These label the stages of the tree $\mathcal{T}$ compatible with the total order of this particular indexing of variables. Note that this equivalence class of situations is precisely*

$$u(\mathbf{Q}_j = \mathbf{q}_j) = \{v(\mathbf{Q}_j = \mathbf{q}_j, \mathbf{R}_j = \mathbf{r}_j) : \mathbf{Q}_j = \mathbf{q}_j\}$$

Because $\mathbf{R}_i \subseteq \mathbf{R}_j$, $j \geq i$ implies that the index $\mathbf{Q}_j = \mathbf{q}_j$ does not depend on the situation, with the choice of element of $u(\mathbf{Q}_i = \mathbf{q}_i)$ labelled by $\mathbf{R}_i = \mathbf{r}_i$. It follows that the positions of $\mathcal{T}$ are exactly its stages. On the other hand, again by definition, if the condition above is violated, then for any compatible total ordering of the variables, there exist values $1 \leq i < j \leq n$ such that the index $\mathbf{Q}_j = \mathbf{q}_j$ depends on the value of $\mathbf{R}_i$. It then follows that $u(\mathbf{Q}_i = \mathbf{q}_i)$ must contain at least two stages, implying that $\#[w] - \#[u] \geq 1$. Thus any such CEG cannot be simple. ■

Corollary 2 *If a CEG is simple and fully represents a BN, $\mathcal{D}$, then $\mathcal{D}$ must be moral.*
Proof. *If a CEG is not representable as a faithful decomposable BN, then there must exist random variables X_i, X_j, X_k, $1 \leq i < j < k \leq n$ in the vertex set of the BN such that $X_i \in \mathbf{Q}_k$ (so that $X_i \notin \mathbf{R}_k$). So there are at least two configurations of $\{X_i, X_j\}$ which define different positions, but for which X_i is not connected by an edge to X_j. Thus $X_i \in \mathbf{R}_j$. But this violates the condition of the theorem.* ■

In the example above, the two fine cuts $\{w(0, X_3), w(1, X_3)\}$ and $\{w(0, X_4), w(1, X_4)\}$ on the original random variables give the respective statements

$$\{X_4, X_3\} \coprod X_2 | X_1$$

$$X_4 \coprod \{X_2, X_1\} | X_3$$

Notice that all the conditional independence statements implied by this BN can be derived from the fine cuts.

4.4 An example of interrogating a CEG that cannot be fully represented by a BN

CEGs have real advantages over BNs when there is additional staging information that cannot be expressed directly by a BN. Such structures are very common. Perhaps the simplest of these lie in the category of context-specific BNs [3, 8, 20]. These also have n-homogeneous trees.

Suppose $(X_1, X_2, \ldots, X_6)$ are binary random variables and

$$X_3 \coprod X_1 | X_2$$
$$X_5 \coprod \{X_1, X_2, X_4\} | X_3$$
$$X_6 \coprod \{X_1, X_2, X_3\} | \{X_4, X_5\}$$

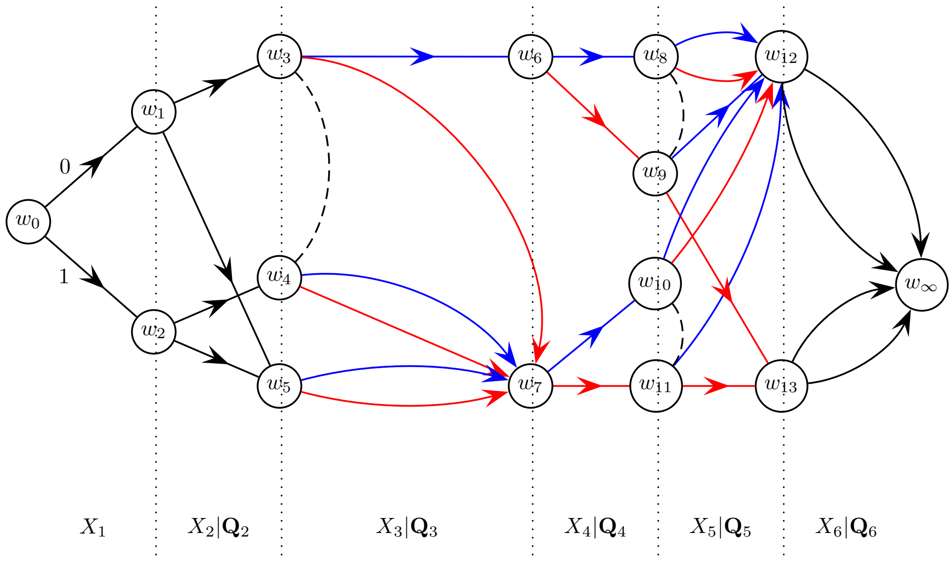

Figure 15: The CEG of binary variables with noisy OR and AND gates as described in the text. The dashed lines are undirected edges. The conditioning variables $(\mathbf{Q}_2, \mathbf{Q}_3, \mathbf{Q}_4, \mathbf{Q}_5, \mathbf{Q}_6)$ defined in section 3.2 no longer correspond to subvectors of $\mathbf{x}$.

It is also known that X_4 is a noisy OR gate on $\{X_1, X_2, X_3\}$: i.e. the distribution of X_4 depends only on whether or not at least one of $\{X_1, X_2, X_3\}$ takes the value 1. Also, X_6 is a noisy AND gate on $\{X_4, X_5\}$: the distribution of X_4 depends only on whether or not both $\{X_4 \text{ and } X_5\}$ take the value 1. The CEG of this situation, see figure 15, not only explicitly depicts the conditional independencies above but also, unlike the BN, the OR and AND gates. There are 32 stages but only 15 positions. These correspond to various configurations of the parents.

Let $(x_1, \ldots, x_k)$ denote the event $(X_1 = x_1, \ldots, X_k = x_k)$, $1 \leq k \leq 5$, where a missing entry corresponds to union over all those coordinates, and let $(\overline{x_1, x_2, \ldots, x_k})$ denote the complement of $(x_1, x_2, \ldots, x_k)$ in the path event space determined by the variables i to k. The positions can then be listed as w_0, the root vertex, $w_1 = (0)$, $w_2 = (1)$, $w_3 = (0, 0)$, $w_4 = (1, 0)$, $w_5 = (1, 1)$, $w_6 = (0, 0, 0)$, $w_7 = (\overline{0, 0, 0})$, $w_8 = (0, 0, 0, 0)$, $w_9 = (0, 0, 0, 1)$, $w_{10} = (\overline{0, 0, 0}, 0)$, $w_{11} = (\overline{0, 0, 0}, 1)$. $w_{13} = (x_4 = 1, x_5 = 1)$ and w_{12} is the complement of w_{13}. The stages that are not positions are $u_{3,5} = \{w_3, w_4\}$, $u_{4,6} = \{w_5\}$, $u_{9,10} = \{w_8, w_9\}$ and $u_{11,12} = \{w_{10}, w_{11}\}$.

The orthogonal fine cuts that can be read automatically from the graphs are the non-informative fine cuts $W_0 = \{w_0\}$ and $W_1 = \{w_1, w_2\}$, together with $W_2 = \{w_3, w_4, w_5\}$, $W_3 = \{w_6, w_7\}$,
$W_4 = \{w_8, w_9, w_{10}, w_{11}\}$ and $W_5 = \{w_{12}, w_{13}\}$, and the orthogonal cuts $U_0 = \{w_0\}$, $U_1 = \{w_1, w_2\}$, $U_2 = \{u_{3,5}, u_{4,6}\}$, $U_3 = \{w_6, w_7\}$, $U_4 = \{u_{9,10}, u_{11,12}\}$ and $U_5 = \{w_{12}, w_{13}\}$. Potentially informative separators can also be read directly from the CEG. Of course, these are defined only up to invertible transformations because they define conditioning sets. Thus, suitable representatives are $\mathbf{Q}(W_2) = (X_1, X_2)$, $\mathbf{Q}(W_3) = \max\{X_1, X_2, X_3\}$, $\mathbf{Q}(W_4) = (X_3, X_4)$, $\mathbf{Q}(W_5) = \min\{X_4, X_5\}$. Here, $\mathbf{Q}(W_2)$ is uninformative because $\mathbf{Z}(W_2)$ is the constant function, but all others convey conditional independence relationships concerning the whole space. For example, the fact that $\mathbf{Q}(W_3)$ is a separator tells us that

$$\{X_4, X_5, X_6\} \coprod \{X_1, X_2, X_3\} | \max\{X_1, X_2, X_3\}$$

This clearly cannot be read from the BN on $\{X_1, X_2, \ldots, X_6\}$, unless $\mathbf{Q}(W_3)$ is added to the variables listed in the BN. Furthermore, using corollary 1, it can be deduced that

$$\{X_5, X_6\} \coprod \{X_1, X_2, X_3, X_4\} | \min\{X_4, X_5\} = 1$$

because conditioning on the position w_{13} gives us that w_7 is a pea. Notice that since it is not true that

$$\{X_5, X_6\} \coprod \{X_1, X_2, X_3, X_4\} | \min\{X_4, X_5\} = 0$$

it is impossible to read this statement from any BN since the value of the conditioning variable must be the same at all levels to be representable as a BN.

4.5 A simple interrogation algorithm

Suppose interest centres on the value of the variable $\boldsymbol{Y}$ — the *queries* variable — measurable with respect to the path σ-field of an event tree, $\mathcal{T}_c$. You have observed a vector of measurements $\boldsymbol{X}$, also measurable with respect to $\mathcal{T}_c$. Your task is to determine which features of $\boldsymbol{X}$ you can discard with no loss of information about $\boldsymbol{Y}$. Equivalently, you want to determine which functions $f(\boldsymbol{X})$ you can keep and still be fully informed about $\boldsymbol{Y}$: i.e. which $f(\boldsymbol{X})$ satisfy $\boldsymbol{Y} \coprod \boldsymbol{X}|f(\boldsymbol{X})$. This type of question has a solution for BNs, when f is a subvector of $\boldsymbol{X}$, through the d-separation theorem.

We present a similar protocol for CEGs when f is allowed to be a general function of $\boldsymbol{X}$. Note that this construction is based on what Shafer calls a simplification, see chapter 13 in [25].

1. From a given enquiry to an elicited tree, $\mathcal{T}_e$, construct a CEG, $\mathcal{C}(\mathcal{T}_e)$. Find a cut U so that all the possible positions of interest, B, are upstream of the cut or in the cut. The set A must contain all positions that define the query and the possible positions that could be observed. Choose a cut to be minimal in the sense that it has the property described above but has the smallest number of positions upstream. A therefore contains all situations in $\mathcal{C}(\mathcal{T}_e)$ whose positions are upstream of all query or observed vertices in $\mathcal{C}(\mathcal{T}_e)$.

2. Beginning again with the situations in A, draw a tree $\mathcal{T}_A$ describing the unfolding of these situations. Construct the CEG of $\mathcal{T}_A$, $\mathcal{C}(\mathcal{T}_A)$. The most expressive CEGs tend to be those that introduce as many situations associated with the observed variables into $\mathcal{T}_A$ as early as possible, and introduce situations involving the query object as late as possible. In the case when $\mathcal{C}(\mathcal{T}_A)$ is cross-sectional, its net can be used to help construct a $\mathcal{T}_A$ sympathetic to the considerations above.

3. Any fine cuts in $\mathcal{C}(\mathcal{T}_A)$ with the property that all situations pertaining to the query on Y lie downstream will now define a conditional independence associated with the observation vector. In particular, if $B(\mathcal{T})$ can be expressed as a function of $\mathbf{Q}(U)$, where U is a fine cut of $\mathcal{C}(\mathcal{T}_A)$, then

$$Y \coprod \mathcal{T}|B(\mathcal{T})$$

Note that the first two steps are analogous to the construction of an appropriate moralised ancestor set under the d-separation procedure, whilst the third is a construction of paths blocking a set of variables from the others. These constructions provide lists of many valid conditional independence statements implicit in a CEG.

5 Discussion

Because CEGs encode conditional independence statements, they can be used as a framework for fast probability propagation both for general inference (using fine cuts) and abductive inference (using cuts). Such algorithms are under development and will be reported in a later paper. Although classes of models defined by how events unfold are not ubiquitous, experience suggests that they are very common. One important subclass are the so-called causal models.

Although the most common graphical method for expressing such hypotheses is an adapted Bayesian network — the causal BN [18, 19, 28] — it has recently been recognised that this representation is unnecessarily restrictive and other methodologies have been suggested. In particular, one author [25] has argued compellingly that causal hypotheses should be expressed through the framework of event trees rather than BNs. The representation of manipulative causal structures by CEGs is extended in [21] and [22]. Such causal modelling [18] is much better addressed within the event tree framework of the CEG than the BN [22]. Furthermore, generalised classes of discrete models can be developed. However, their richness often precludes the use of a graphical representation, see [21, 29].

Finally, within the framework described above, classes of prior probability distributions over the simplices of a CEG can be defined so that CEGs can be estimated. This methodology is discussed from a Bayesian perspective in [22] and it is shown that a conjugate product Dirichlet prior-to-posterior analysis is often possible.

6 Acknowledgements

Jim Q. Smith and Paul E. Anderson are members of the Department of Statistics and the Interdisciplinary Program for Cellular Regulation (IPCR) (soon to become the Centre for Systems Biology) at the University of Warwick. Paul E. Anderson thanks EPSRC, BBSRC and BioSim for financial support. Both authors thank the referees, whose insightful comments greatly improved the original version.

References

[1] Paul E. Anderson and Jim Q. Smith. A graphical framework for representing the semantics of asymmetric models. Technical Re-

port 05-12, CRiSM paper, Department of Statistics, University of Warwick, 2005.

[2] T. Bedford and R. Cooke. *Probabilistic Risk Analysis — Foundations and Methods*, pages 99–151. Cambridge University Press, 2001.

[3] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-Specific Independence in Bayesian Networks. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, pages 115–123, San Francisco, CA, 1996. Morgan Kaufmann Publishers.

[4] R. E. Bryant. Graphical algorithms for Boolean function manipulation. *IEEE Transactions of Computers*, C-35:677–691, 1986.

[5] G. A. Churchill. Accurate Restoration of DNA Sequences. In C. Gatsaris, J. S. Hodges, R. E. Kass, and N. D. Singpurwalls, editors, *Case Studies in Bayesian Statistics*, volume 2, pages 90–148. Springer-Verlag, 1995.

[6] Jukka Corander. Labelled graphical models. *Scandinavian Journal of Statistics*, 30(3):493–508, 2003.

[7] S. French, editor. *Readings in Decision Analysis.* Chapman Hall, 1989.

[8] D. Geiger, D. Heckerman, and C. Meek. Asymptotic model selection for directed networks with hidden variables. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence (UAI-96)*, pages 283–290, Portland, OR, 1996. Morgan Kaufmann Publishers.

[9] S. Hojsgaard. Statistical Inference in Context Specific Interaction Models for Contingency Tables. *Scandinavian Journal of Statistics*, 31(1):143–158, 2004.

[10] M. Jaeger. Probabilistic Decision Graphs — Combining Verification and AI Techniques for Probabilistic Inference. *Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems*, 12:19–42, 2004.

[11] F. V. Jensen. *Bayesian Networks and Decision Graphs.* Springer Verlag, 2001.

[12] H. Kiiveri, T. P. Speed, and J. B. Carlin. Recursive causal models. *Journal of Australian Mathematical Society A*, 36:30–52, 1984.

[13] S. L. Lauritzen. *Graphical models.* Oxford Science Press, 1st edition, 1996.

[14] R. Lyons. Random walks and percolation on trees. *Ann. Probab.*, 20:2043–2088, 1992.

[15] David McAllester, Michael Collins, and Fernando Pereira. Case-Factor Diagrams for Structured Probability Modelling. In *Proceedings of the 20th Annual Conference on Uncertainty in Artificial*

Intelligence (UAI-04), pages 382–391, 2004.

[16] S. M. Olmsted. *On representing and solving decision problems*. PhD thesis, Engineering-Economic Systems, Stanford University, 1983.

[17] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kauffman, 1988.

[18] J. Pearl. *Causality, models, reasoning and inference*. Cambridge University Press, 2000.

[19] J. Pearl. Statistics and Causal Inference: A Review (with discussion). *Sociedad de Estadistica e Investigacion Operativa Test*, 12(2):281–345, 2003.

[20] David Poole and Nevin Lianwen Zhang. Exploiting Contextual Independence in Probabilistic Inference. *Journal of Artificial Intelligence Research*, 18:263–313, 2003.

[21] E. Riccomagno and J. Q. Smith. Identifying a cause in models which are not simple Bayesian networks. In *Proceedings of the 10th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1315–1322, Perugia, 2004.

[22] E. Riccomagno and J. Q. Smith. The Causal Manipulation and Bayesian Estimation of Chain Event Graphs. Technical report, Univeristy of Warwick, 2005.

[23] R. D. Shachter. Evaluating Influence Diagrams. In A. P. Basu, editor, *Reliability and Control*, pages 321–344. Elsevier, 1986.

[24] R. D. Shachter. Intelligent Probabilistic Inference. In L. N. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 371–382. Amsterdam: North-Holland, 1986.

[25] G. Shafer. *The Art of Causal Conjecture*. Cambridge, MA, MIT Press, 1996.

[26] G. Shafer, P. R. Gillett, and R. B. Scherl. A new understanding of subjective probability and its generalization to lower and upper prevision. *International Journal of Approximate Reasoning*, 33(1):1–49, 2003.

[27] J. Q. Smith. Influence Diagrams for Statistical Modelling. *Annals of Statistics*, 17:654–672, 1989.

[28] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. Springer Verlag, 1993.

[29] M. Studeny. *Probabilistic Conditional Independence Structures*. Springer-Verlag, 2005.