Artificial Intelligence 195 (2013) 291–315

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Causal identifiability via Chain Event Graphs

Peter Thwaites

School of Mathematics, University of Leeds, LS2 9JT, United Kingdom

ARTICLE INFO

Article history:
Received 15 April 2011
Received in revised form 6 September 2012
Accepted 12 September 2012
Available online 13 September 2012

Keywords:
Back Door theorem
Bayesian Network
Causal identifiability
Causal manipulation
Chain Event Graph
Conditional independence
Front Door theorem

ABSTRACT

We present the Chain Event Graph (CEG) as a complementary graphical model to the Causal Bayesian Network for the representation and analysis of causally manipulated asymmetric problems. Our focus is on causal identifiability — finding conditions for when the effects of a manipulation can be estimated from a subset of events observable in the unmanipulated system. CEG analogues of Pearl's Back Door and Front Door theorems are presented, applicable to the class of *singular* manipulations, which includes both Pearl's basic *Do* intervention and the class of functional manipulations possible on Bayesian Networks. These theorems are shown to be more flexible than their Bayesian Network counterparts, both in the types of manipulation to which they can be applied, and in the nature of the conditioning sets which can be used.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In this paper we consider cause and effect through the analysis of controlled models. The standard apparatus for such an approach is the Causal Bayesian Network (CBN) [8,14,15,24]. The CBN is a version of a Bayesian Network (BN) where the directionality of the edges of the graph is interpreted as *causal* and the BN as representing a causal model.

BNs are specifically designed to work with problems which have a natural product space structure, but many problems which we might wish to model do not have such a structure, and are *asymmetric* in that problem variables can have different sets of possible outcomes given different outcomes of their parental variables, or even no possible outcomes given some outcomes of their parents. The future development at any specific point depends on the particular history of the problem up to that point (i.e. on the outcomes of ancestral variables), and the values of a particular set of covariates at that point. It is these types of problem that we are primarily concerned with here.

So for instance, consider an infectious disease which is serious for people who have blood type O. Following a first treatment, patients with this blood type either die or need a second treatment; patients with other blood types either need a second treatment or make a full recovery. At the next stage of the process, patients who have died or fully recovered are not offered a second treatment, but all other patients are given one of three possible second treatments, the choice of which is dependent on factors such as hospital policy, consultant preference etc. Similar examples occur in many other areas (see for example [1,3,7,12]).

Context-specific variants of BNs have been developed for tackling asymmetric problems [2,13,18,20]. They can be used for the representation and analysis of problems whose future development at any specific point depends on the particular history of the problem up to that point, but their use is more circumscribed in problems where there may be no possible outcomes of some variables given certain histories or values of covariates. In both cases however the problems being analysed are no longer fully represented by the topology of the graph — context-specific BNs require supplementary information

E-mail address: P.A.Thwaites@leeds.ac.uk.

0004-3702/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.artint.2012.09.003

in the form of trees, or tree-like conditional probability tables attached to vertices to depict the asymmetries in the problem. The Chain Event Graph (CEG) introduced in [22,25] is specifically designed for the representation and analysis of such problems. It is a function of an event tree [21], whose topology expresses the full collection of independence properties associated with a problem. It is particularly useful when problems do not exhibit a product space structure, or when there is a lot of context-specific information present. All aspects of the model structure, including any context-specific dependencies are represented in the topology of the graph — these are not bolted on.

There have been many recent advances in CBN theory (see for example [5,6,9,16,27,28]), but little of this has made the causal analysis of such asymmetric problems simpler. In particular, techniques such as Pearl's Back and Front Door theorems [14,15] have conditions which are expressed in terms of the topology of the CBN — if the structure of the problem can no longer be expressed fully in terms of the topology of the graph, then this benefit is lost.

CBNs can be used for the basic *Do* intervention of Pearl [15], which sets a particular variable to a particular value; and their use has been extended to *functional* manipulations ($Do\, X = g(W)$ for some set of variables W), and *stochastic* manipulations which assign a new probability distribution over the outcomes of the manipulated variable. The ease with which necessary conditions can be checked on the unmanipulated graph however vanishes very rapidly as we move away from basic interventions.

It can be argued [4,21,26] that *causes* are more naturally expressed as *events* rather than the values of some random variable. The CEG provides an ideal graphical **representation** given this argument. It is also a sensible representation for the **analysis** of manipulations to events. By making additional assumptions concerning a CEG model we can give it a causal interpretation, and extend its use to causal analysis in an analogous manner to that in which CBNs extend the use of BNs. Unlike analysis using CBNs, the analysis of functional and stochastic manipulations using CEGs is no more complex than the analysis of the basic *Do* manipulation. In using the CEG for causal analysis we are building on the ideas of researchers who have attempted to use trees for this purpose [19,21,24].

Note also that in CBN analysis the standard methods for reducing the complexity of a manipulated probability expression (for example Pearl's Back and Front Door theorems) rely on the use of *blocking* sets consisting of problem variables. With CEG-based analysis our blocking sets are composed of events, allowing us more flexibility in their construction; so instead of conditioning on a set of variables $Z = \{Z_1, Z_2\}$ say, we might only need to condition on the events {*patient is male, patient is female and aged below* 40}.

A crude version of a Back Door theorem for CEGs was introduced in [26]. Here we present a much more general Back Door theorem as well as two alternative versions of a Front Door theorem. No knowledge of the content of [26] is assumed in this paper. The earlier paper touched briefly on some topics covered here, such as the use of CEGs for more sophisticated manipulations, but here we offer a comprehensive overview of causal analysis on CEGs, and look more carefully at causal identifiability — finding conditions for when the effects of a manipulation can be estimated from a subset of events observable in the idle system. Pearl's Back and Front Door theorems give sufficient conditions for causal identifiability in BNs, and their arrival prompted a search for a complete set of conditions, using which an analyst could gauge whether or not an expression could be estimated from a subset of observable variables [6,27,28]. This paper provides several sets of sufficient conditions for causal identifiability in CEGs. We anticipate that future work will allow us to find necessary and sufficient conditions for identifiability, expressed in terms of subsets of observable events as opposed to observable variables.

As the CEG is a comparatively new structure, there have been minor modifications since [22] and [26]. In particular we have removed the undirected edges from previous definitions so that the CEG is now a DAG. This has led to a less cluttered representation and made the CEG easier to read.

In Section 2 we define the CEG and manipulated CEG. Section 3 develops the Back Door theorem and the idea of *singular manipulations*. A Front Door theorem is then introduced in Section 4, and Section 5 provides a discussion of possible directions for future research.

2. Definitions and notation

In this section we define the CEG. We also provide some notation that will be used throughout the paper. We then turn our attention to what it means when we manipulate a CEG to an event, and present a definition of a manipulated CEG.

The CEG is a function of an event tree [21], retaining those features of the tree which allow for the transparent representation of asymmetric problems. They are a significant extension to trees since they express within their topology a more complete description of the conditional independence structure of a problem.

An event tree $\mathcal{T}$ is a directed tree with vertex set $V(\mathcal{T})$ and edge set $E(\mathcal{T})$. It has a single root vertex v_0, and a collection of leaf vertices (see for example Fig. 1, where there are 7 leaf vertices). The root-to-leaf paths $\{\lambda\}$ of $\mathcal{T}$ form the atoms of the event space. Events measurable with respect to this space are unions of these atoms.

Let $V^0(\mathcal{T})$ denote the set of non-leaf vertices of $\mathcal{T}$. Then each vertex $v \in V^0(\mathcal{T})$ labels a random variable $X(v)$ whose state space $\mathbb{X}(v)$ can be identified with the set of directed edges $e(v, v') \in E(\mathcal{T})$ emanating from v. For each $X(v)$ we let

$$\Pi(v) \equiv \{\pi_e(v' \mid v) \mid e(v, v') \in \mathbb{X}(v)\}$$

where $\pi_e(v' \mid v) \equiv P(X(v) = e(v, v'))$ are called the *primitive probabilities* of the tree; and

$$\Pi(\mathcal{T}) \equiv \{\Pi(v)\}_{v \in V(\mathcal{T})}$$

There are a number of modifications we can make to an event tree to enable it to portray conditional independence structure more transparently. The first of these is to highlight those vertices in $V(\mathcal{T})$ whose outgoing edges carry the same labels and the same probabilities. We do this through the concept of a *stage*, and through the colouring of edges.

Definition 1 (*Stages and colour*). For an event tree $\mathcal{T}$ with non-leaf vertex set $V^0(\mathcal{T})$ and edge set $E(\mathcal{T})$

1. The set $V^0(\mathcal{T})$ is partitioned into equivalence classes, called *stages*, as follows:
   Vertices $v^1,\ v^2 \in V^0(\mathcal{T})$ are members of the same equivalence class (stage) if (a) v^1, v^2 do not lie on the same root-to-leaf path, and (b) there is a bijection ψ between $\mathbb{X}(v^1)$ and $\mathbb{X}(v^2)$ such that if $\psi : e(v^1, v^{1\prime}) \mapsto e(v^2, v^{2\prime})$ then $\pi_e(v^{1\prime} \mid v^1) = \pi_e(v^{2\prime} \mid v^2)$.
2. The set $E(\mathcal{T})$ is partitioned into equivalence classes, whose members have the same colour, as follows:
   Edges $e(v^1, v^{1\prime}),\ e(v^2, v^{2\prime})$ have the same *colour* if and only if the vertices v^1 and v^2 are in the same stage, and $\pi_e(v^{1\prime} \mid v^1) = \pi_e(v^{2\prime} \mid v^2)$.

The set of stages of the tree $\mathcal{T}$ is labelled $L(\mathcal{T})$, and individual stages are labelled u.

Definition 2 (*Coloured tree*). An event tree $\mathcal{T}$ is a *coloured tree* if its edges are coloured in accordance with Definition 1.

The second modification is to highlight those vertices in $V(\mathcal{T})$ from which the complete future development of the process is essentially the same. We do this through the concept of a *position*.

For $v \in V^0(\mathcal{T})$, let $\mathcal{T}(v)$ denote the unique subtree of $\mathcal{T}$ whose root is v, and which contains all edges from $E(\mathcal{T})$ and vertices from $V(\mathcal{T})$ which lie on a v-to-leaf subpath of $\mathcal{T}$.

Definition 3 (*Positions*). For a coloured tree $\mathcal{T}$ with non-leaf vertex set $V^0(\mathcal{T})$ and edge set $E(\mathcal{T})$, the set $V^0(\mathcal{T})$ is partitioned into equivalence classes, called *positions*, as follows: Vertices $v^1,\ v^2 \in V^0(\mathcal{T})$ are members of the same equivalence class (position) if (a) the coloured subtrees $\mathcal{T}(v^1)$ and $\mathcal{T}(v^2)$ are topologically identical, and (b) there is a bijection between $\mathcal{T}(v^1)$ and $\mathcal{T}(v^2)$ such that edges in $\mathcal{T}(v^2)$ are coloured identically with their corresponding edges in $\mathcal{T}(v^1)$.

The set of positions is labelled $K(\mathcal{T})$, and the individual positions are labelled w. Note that the partition into positions is finer than the partition into stages, and that if two vertices are in the same position, they are necessarily in the same stage.

Note also that vertices are in the same stage when the sets of possible immediate outcomes at each vertex are the same and have the same (conditional) probability distribution. Vertices are in the same position when the sets of entire future evolutions from each vertex have the same probability distribution. Example 1 below puts the meanings of *stage* and *position* into a simple practical context.

Once we have the ideas of stages and positions we can make one last modification so that the conditional independence structure of the problem is expressed entirely through the topology of the graph. Example 1 demonstrates how a tree-to-CEG conversion is implemented.

Definition 4 (*Chain Event Graph*). The Chain Event Graph $\mathcal{C}(\mathcal{T})$ is the coloured graph with vertex set $V(\mathcal{C})$ and edge set $E(\mathcal{C})$ defined by:

1. $V(\mathcal{C}) \equiv K(\mathcal{T}) \cup \{w_\infty\}$, where w_∞ is called the *sink-vertex* of $\mathcal{C}(\mathcal{T})$.
2. (a) For $w,\ w' \in V(\mathcal{C}) \setminus \{w_\infty\}$, there exists a directed edge $e(w, w') \in E(\mathcal{C})$ iff there are vertices $v,\ v' \in V^0(\mathcal{T})$ such that $v \in w \in K(\mathcal{T}),\ v' \in w' \in K(\mathcal{T})$ and there is an edge $e(v, v') \in E(T)$.
   (b) For $w \in V(\mathcal{C}) \setminus \{w_\infty\}$, there exists a directed edge $e(w, w_\infty) \in E(\mathcal{C})$ iff there is a non-leaf vertex $v \in V^0(\mathcal{T})$ and a leaf vertex $v' \in V(\mathcal{T})$ such that $v \in w \in K(\mathcal{T})$ and there is an edge $e(v, v') \in E(\mathcal{T})$.

So the vertices of a CEG $\mathcal{C}(\mathcal{T})$ correspond to the **positions** of the underlying tree $\mathcal{T}$. We can therefore, without ambiguity, call the vertices of the CEG *positions*. Since the positions partition the vertices of the tree, and vertices in the same position are necessarily in the same stage, we can transfer the concepts of *stage* and *colour* from the tree to the CEG.

If positions w^1, w^2 in the CEG correspond to sets of vertices in the tree which are all members of the same stage u, we say that w^1 and w^2 are in the same stage u in the CEG. We label the set of stages of $\mathcal{C}(\mathcal{T})$ by $L(\mathcal{C})$. The colouring of the edges leaving any vertex v, a member of the stage u in $\mathcal{T}$, is inherited by the edges leaving any position w, a member of the stage u in $\mathcal{C}(\mathcal{T})$.

To aid in the reading of our CEGs we have in this paper given the same colour to positions in a CEG which are members of the same stage (as well as colouring the edges emanating from these positions).

There is a one-to-one correspondence between the root-to-leaf paths $\{\lambda\}$ of $\mathcal{T}$ and the root-to-sink ($w_0 \to w_\infty$) paths $\{\lambda\}$ of $\mathcal{C}(\mathcal{T})$, and these latter form the atoms of the event space of $\mathcal{C}(\mathcal{T})$. Events measurable with respect to this space are unions of these atoms.

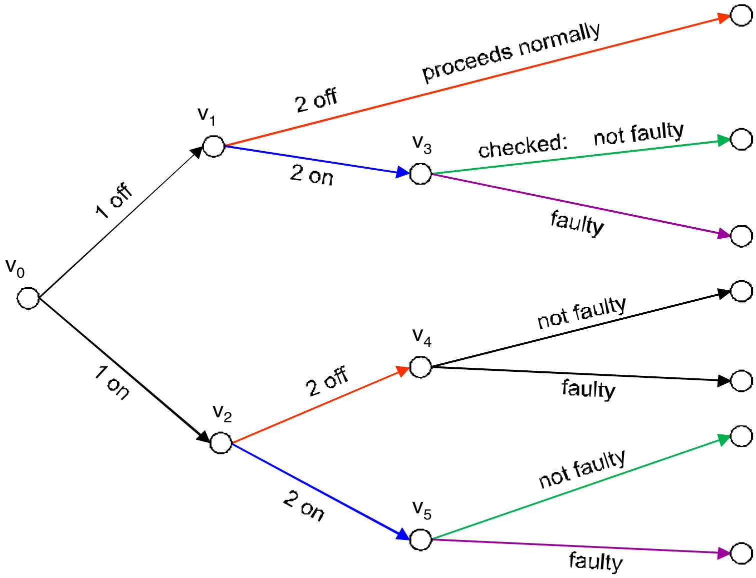

Fig. 1. Coloured tree for Example 1.

Stages: $\{v_0\}, \{v_1, v_2\}, \{v_3, v_5\}, \{v_4\}$
Positions: $\{v_0\}, \{v_1\}, \{v_2\}, \{v_3, v_5\}, \{v_4\}$

Each stage $u \in L(\mathcal{C})$ labels a random variable $X(u)$ whose state space $\mathbb{X}(u)$ can be identified with the set of directed edges $e(w, w') \in E(\mathcal{C})$ emanating from any position $w \in u$.

Note that Definition 1 (1)(a) implies that if the positions $w^1, w^2 \in V(\mathcal{C})$ are members of the same stage $u \in L(\mathcal{C})$, then w^1, w^2 do not lie on the same root-to-sink path in $\mathcal{C}(\mathcal{T})$.

For the remainder of this paper we abbreviate $\mathcal{C}(\mathcal{T})$ to $\mathcal{C}$, and assume that $\mathcal{C}$ has associated edge and vertex sets $E(\mathcal{C})$ and $V(\mathcal{C})$.

Example 1 (*CEG construction*). We illustrate the construction of a CEG through a fault diagnosis example, which for illustrative convenience uses only binary variables.

- A machine has two warning lights which indicate possible faults in two components.
- If either light is on, the machine is checked and judged to be either *faulty* or *not*.
- If both lights are off, operation proceeds as normal.

Operational evidence indicates that

(1) the warning lights come on independently of each other,
(2) whether the machine is judged faulty is independent of whether or not light 1 is on, provided that light 2 is on.

This information is represented in the coloured tree in Fig. 1. For ease of interpretation, only edges which share a probability have been coloured.

The edges leaving v_1 and v_2 are coloured to indicate that the probabilities of 2 *on* (blue) and 2 *off* (red) do not depend on whether light 1 is on or off. The vertices v_1 and v_2 are in the same stage.

The edges leaving v_3 and v_5 are coloured to indicate that the probabilities of *faulty* (mauve) and *not faulty* (green) do not depend on whether light 1 is on or off, provided that light 2 is on. The vertices v_3 and v_5 are in the same stage. As v_3 and v_5 are the roots of identically coloured subtrees, v_3 and v_5 are also in the same position.

Definition 2 tells us that the vertex set of a CEG consists of the positions of the coloured tree and a sink-vertex which groups all the leaf-nodes of the tree together. So v_3 and v_5 are merged, and the (blue) edges $e(v_1, v_3)$ and $e(v_2, v_5)$ in the tree become edges $e(w_1, w_4)$ and $e(w_2, w_4)$ in the CEG (Fig. 2).

It is possible to read both positions and stages in a CEG, in a manner similar to the reading of BNs. The position w_4 can be read to give the context-specific conditional independence property that *whether judged faulty or not is independent of whether light* 1 *is on or off, given that light* 2 *is on*.

Reading stages we only look into the immediate future. So reading $\{w_1, w_2\}$ gives the conditional independence property that *whether light* 2 *is on or off is independent of whether light* 1 *is on or off*.

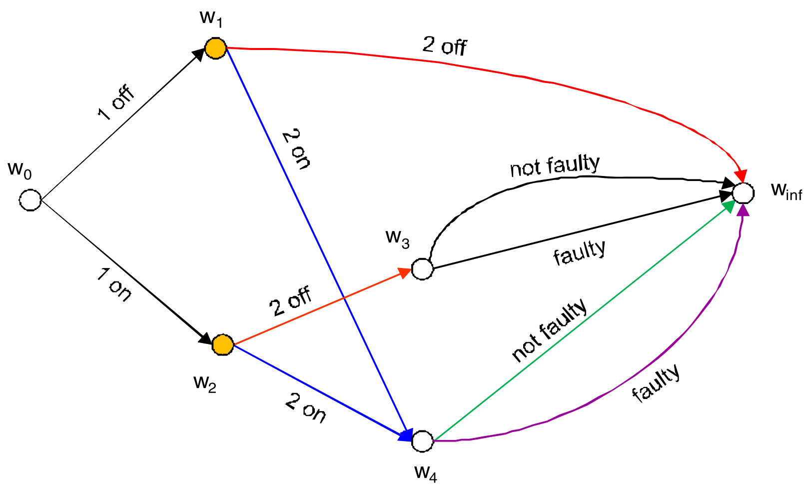

Fig. 2. Chain Event Graph for Example 1.

Stages: $\{w_0\}, \{w_1, w_2\}, \{w_3\}, \{w_4\}$

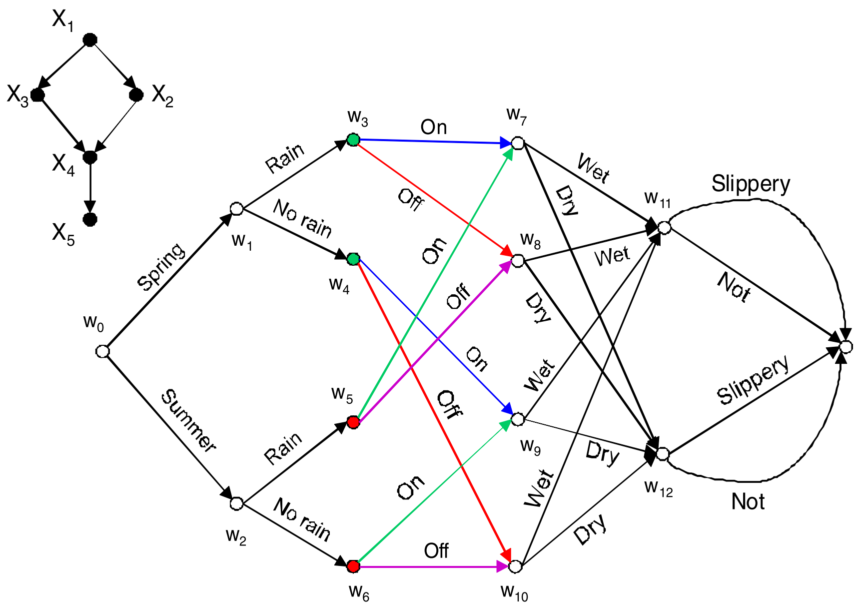

Fig. 3. BN and CEG for Example 2.

A BN-representation of this problem would consist of three variables (X_1 denoting *light* 1 *on/off*, X_2 denoting *light* 2 *on/off*, X_3 denoting *faulty/not faulty*), with edges from each of X_1 and X_2 into X_3. The additional knowledge that lights 1 & 2 both off implies *not faulty*, and that X_3 is independent of X_1 provided that light 2 is on, would not be represented in the topology of the graph, and would have to appear as supplementary information.

Example 2 (*Reading CEGs*). To show how CEGs relate to BNs we here provide a CEG-version of Pearl's *Sprinkler* example from [15].

Here there are five variables $X_1, \ldots, X_5$, the causal relationships between which are illustrated in the BN in Fig. 3.

X_1: *season*

X_2: *rain* {*yes, no* }

X_3: *(water) sprinkler* {*on, off* }

X_4: *(pavement) wet* {*wet, dry* }

X_5: *(pavement) slippery* {*slippery, not slippery* }

From the BN we can read that $X_3 \amalg X_2 \mid X_1$ (*knowing whether it is raining or not does not help us in estimating a probability for the event that the sprinkler is on, given that we know the season*), but $X_3 \not\amalg X_2 \mid X_4$ (*knowing whether it is raining or not* **does**

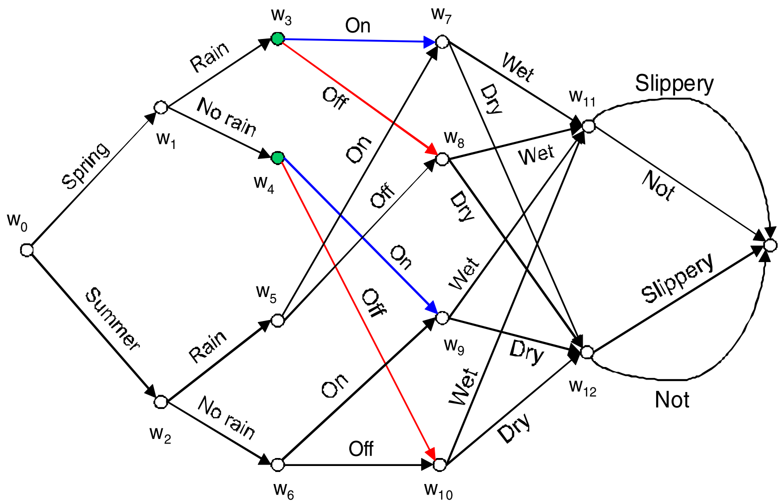

Fig. 4. Second CEG for Example 2.

help us in estimating a probability for the event that the sprinkler is on, given that we know whether the pavement is wet or dry). We can also read that $(X_4, X_5) \amalg X_1 \mid (X_2, X_3)$ and that $X_5 \amalg (X_1, X_2, X_3) \mid X_4$.

The CEG for this example is also given in Fig. 3, where for illustrative convenience I have shown only two seasons — Spring and Summer.

Here we have 13 positions $\{w_0, w_1, \ldots, w_{12}\}$, each of which can be read for a unique context-specific independence property. So for example, reading w_7 gives us that *the state of the pavement (wetness, slipperiness) is independent of the season given that it is raining and the sprinkler is on*. Combining the readings for w_7, w_8, w_9 and w_{10} we get (as in the BN) that $(X_4, X_5) \amalg X_1 \mid (X_2, X_3)$.

Similarly, reading w_{11} gives us that *the slipperiness of the pavement is independent of the season, whether it is raining or not, and whether the sprinkler is on or off, given that the pavement is wet*. Combining with the reading of w_{12} gives us that $X_5 \amalg (X_1, X_2, X_3) \mid X_4$.

There are also two stages which contain more than a single position. These are $\{w_3, w_4\}$ and $\{w_5, w_6\}$. When we read these we only look into the immediate future. So reading $\{w_3, w_4\}$ gives us that *whether the sprinkler is on or off is independent of whether or not it is raining, given that it is Spring*. Combining with the reading of $\{w_5, w_6\}$ gives us that $X_3 \amalg X_2 \mid X_1$.

The example as it stands is clearly symmetric, and the CEG gives us no more information than the BN. Indeed, given the number of nodes and edges, the CEG here is a considerably less efficient representation than the BN. However, as in Example 1, we can build context-specific information into the model much more readily than with a BN. So suppose that during the Summer a gardener is on hand to check the weather before turning on the sprinkler. Then X_3 is not independent of X_2 when X_1 takes the value corresponding to *Summer*, but **is** independent of X_2 when X_1 takes the value corresponding to *Spring*. In the BN-representation we would need to add an edge between X_2 and X_3, and supplement the graph with some additional information. In contrast, this information can be represented directly in the topology of the CEG, as shown in Fig. 4, where w_5 and w_6 are no longer in the same stage.

The following notation will be used throughout the remainder of the paper. Analogously with atoms in a tree, an atom λ is a $w_0 \to w_\infty$ path in $\mathcal{C}$. The set of atoms is denoted Ω. We write $w \prec w'$ when the position w precedes the position w' on a $w_0 \to w_\infty$ path (i.e. there exists a set of directed edges joining w to w' — in Fig. 3 for example $w_5 \prec w_{12}$, but $w_5 \not\prec w_9$). We call w a *parent* of w' if there exists an edge $e(w, w') \in E(\mathcal{C})$.

Events are denoted Λ. $\Lambda(w)$ is the event which is the union of all $w_0 \to w_\infty$ paths passing through the position w, and $\Lambda(e(w, w'))$ is the union of all paths passing through the edge $e(w, w')$.

We can now define what we call the *primitive probabilities* of the CEG: We use the notation $\pi_e(w' \mid w)$ to denote the probability of passing along the edge $e(w, w')$ having arrived at the position w. These probabilities have the same relationship to the CEG as the sets of conditional probability tables associated with a BN do to the corresponding DAG, so for example in Fig. 3, $\pi_e(w_{11} \mid w_7)$ is the probability that the pavement is wet given that it is raining and that the sprinkler is on. For each $u \in L(\mathcal{C})$ and random variable $X(u)$ we let

$$\Pi(u) \equiv \{\pi_e(w' \mid w) \mid w \in u\}$$

and

$$\Pi(\mathcal{C}) \equiv \{\Pi(u)\}_{u \in L(\mathcal{C})}$$

Note that if we label the probability of the event Λ by $\pi(\Lambda)$ then $\pi_e(w' \mid w) \equiv \pi(\Lambda(e(w,w')) \mid \Lambda(w))$. Each $w_0 \to w_\infty$ path λ corresponds to a sequence of outcomes at a set of positions, or equivalently a vector of values of a set of $X(u)$ variables. So the probability of an atom λ can be expressed as a product of edge-probabilities, each of the form $\pi(\Lambda(e(w,w')) \mid \Lambda(w))$, where $\Lambda(e(w,w'))$ corresponds to some value of $X(u)$ and $\Lambda(w)$ describes the *parental* configuration for the position w. But this is a product of probabilities of values of variables conditioned on the values of their parents, so $\Pi(\mathcal{C})$ satisfies the Directed Markov condition [15,10] with respect to the CEG $\mathcal{C}$.

A subpath of a root-to-sink path is denoted $\mu(w,w'')$, where w and w'' indicate the start and end positions of the subpath. $\Lambda(\mu(w,w''))$ is the event which is the union of all paths utilising the subpath $\mu(w,w'')$. $\pi_\mu(w'' \mid w) \equiv \pi(\Lambda(\mu(w,w'')) \mid \Lambda(w))$ is the probability of passing along the subpath $\mu(w,w'')$ having arrived at the position w.

Before moving on to manipulated CEGs we present a very useful lemma, a proof of which appears in Appendix A.

Lemma 1 *(Limited Memory). For a CEG $\mathcal{C}$ and positions $w_1, w_2, w_3 \in V(\mathcal{C})$ such that $w_1 \prec w_2 \prec w_3$*

$$\pi\big(\Lambda(w_3) \mid \Lambda(w_1), \Lambda(w_2)\big) = \pi\big(\Lambda(w_3) \mid \Lambda(w_2)\big)$$

Here w_1 precedes w_2 precedes w_3 in the sense described above. The lemma states that the probability of passing through the position w_3 given that we have previously passed through both w_1 and w_2 is equal to the probability of passing through the position w_3 given only the information that we have previously passed through w_2.

This result can be extended so that the positions w_1 and w_2 can each be replaced by edges, and the position w_3 can be replaced by a union of positions and/or edges.

Essentially this tells us that being at a position (w_3) or edge (or collection of positions or edges), given that we have been at an earlier position (w_2) or edge, is independent of the path taken to that earlier position or edge. This result is used in the proof of Theorem 1.

2.1. Manipulated CEGs

Anything that we observe about a system or do to a system will change the topology of a graphical representation of that system. In [25] we considered how the topology of a CEG is altered when we observe an event Λ. Here we investigate how the topology of a CEG is altered when we manipulate to an event Λ. As the following definitions suggest, the process of updating our beliefs following a manipulation is very similar to that which happens following the observation of an event.

How does this relate to the manipulation of a BN? When we talk about manipulating a BN or enacting an intervention on a BN, we are in general setting a variable or a collection of variables to some specified values. In the case of a *stochastic* manipulation we are altering the probability distribution associated with a variable or set of variables. In the former case we can see that this can be described as manipulation to an event (e.g. the event that $X = x_1$ for some variable X and value x_1). In the latter case the equivalent action on a CEG would be a reassignment of a subset of the edge-probabilities of the CEG; and we allow for this in Definition 5 below.

The type of events we consider in this paper are *intrinsic* events (called *$\mathcal{C}$-compatible* events in [25]). An intrinsic event Λ is one which defines a subgraph of $\mathcal{C}$, so every atom of Λ is a $w_0 \to w_\infty$ path of a subgraph of $\mathcal{C}$, and every $w_0 \to w_\infty$ path in this subgraph is an atom of Λ. The set of intrinsic events is large, and if the CEG is expressible as a BN then it contains as a proper subset all sets of the form $\{X_j \in A_j\}$ for subsets $\{A_j\}$ of the sample spaces of $\{X_j\}$, the vertex variables of the BN. It excludes a few events of a more convoluted structure, but if we wish to manipulate to such an event we can convert it to intrinsic by relaxing some of the conditional independence structure represented by the CEG and expanding one or more merged vertices.

A manipulation to an intrinsic event Λ is hence a reassignment of edge-probabilities so that $\hat{\pi}(\Lambda) = 1$, where $\hat{\pi}$ denotes a probability following manipulation.

Clearly such a reassignment might be done in several ways, so the description of any specified manipulation to Λ should include details of how edge-probabilities are to be assigned. Also, if Λ is a proper subset of Ω, a manipulation to Λ must assign zero-probabilities to a proper subset of $E(\mathcal{C})$. We normally *prune* our manipulated CEG by removing edges and positions which only lie on paths which are not elements of Λ.

Example 3. Consider again Pearl's Sprinkler example (Example 2) and the CEG for this in Fig. 3. Suppose we were to enact the intervention *Put sprinkler on*. This is setting the variable X_3 to the value associated with *sprinkler on*. If we let this value be 1 then Pearl would call this manipulation *Do* $X_3 = 1$. The manipulated BN for this intervention is in Fig. 5.

In the CEG-representation we are manipulating to the event Λ which is the union of all paths passing through an edge labelled *on*. As we are not controlling any other aspect of the system, this is done by assigning an edge-probability of 1 to each *on* edge, an edge-probability of 0 to each *off* edge, and leaving all other edge-probabilities unchanged. As suggested above we prune edges and positions which only lie on paths which are not elements of Λ — here the positions $\{w_8, w_{10}\}$ and the edges entering or leaving these positions. The resultant manipulated CEG is given in Fig. 5.

This is a symmetric manipulation of a symmetric system, and the CEG is a less efficient representation than a CBN. But we can also consider *functional* manipulations of the system such as: *If it is Summer put the sprinkler on; if it is Spring and it is raining put the sprinkler off.* In the BN-representation of the problem, instead of removing the edge from X_1 to X_3 as in

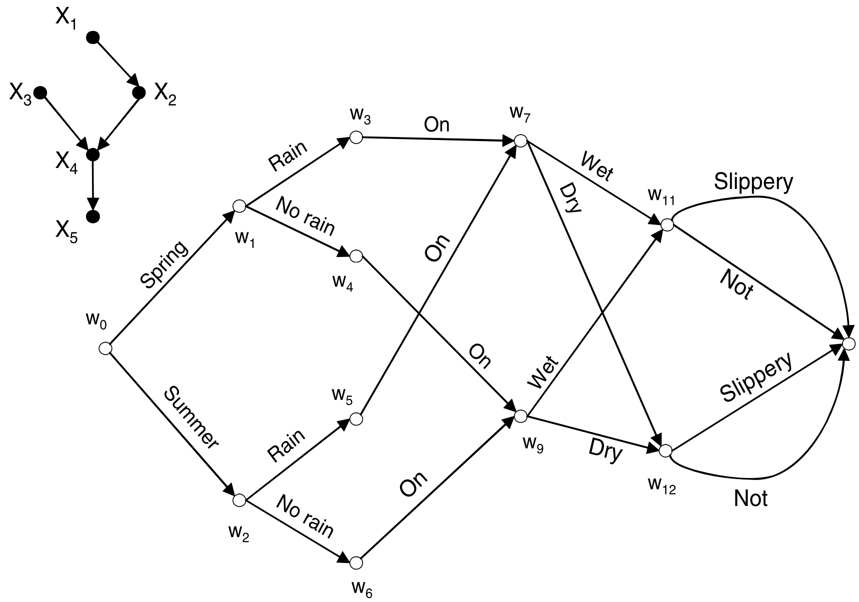

Fig. 5. Manipulated BN and CEG for Example 3.

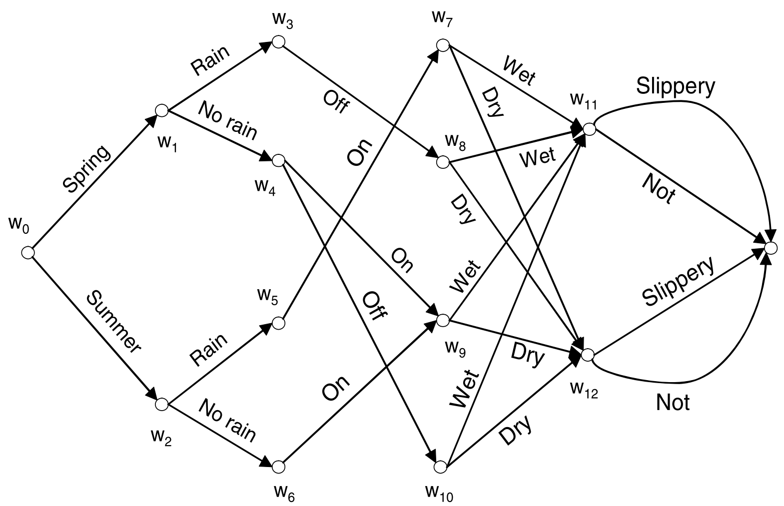

Fig. 6. Second manipulated CEG for Example 3.

Fig. 5, there is now an additional edge from X_2 to X_3 since whether the sprinkler is on or off depends on both the season and whether it is raining. The BN would also have to be supplemented with the extra information given above. So any advantages gained through the simplicity of the BN topology are lost.

Using a CEG, representing this manipulation is as straightforward as representing the simple *Put sprinkler on* intervention. The resultant CEG is shown in Fig. 6. Here, as again we are not controlling any other aspect of the system, the only edge-probabilities which might be modified are those leaving the positions $\{w_3, w_4, w_5, w_6\}$. The edges $e(w_3, w_8)$, $e(w_5, w_7)$ and $e(w_6, w_9)$ are given probability 1; the edges $e(w_3, w_7)$, $e(w_5, w_8)$ and $e(w_6, w_{10})$ are given probability 0 and are pruned. As the description of our manipulation does not specify what happens when it is Spring but doesn't rain, we leave the edge-probabilities for $e(w_4, w_9)$ and $e(w_4, w_{10})$ unchanged.

Definition 5 (*Manipulated CEG*). For a CEG $\mathcal{C}$ and intrinsic event Λ, let $\hat{\mathcal{C}}^{\Lambda}$ (the CEG manipulated to the event Λ) be the subgraph of $\mathcal{C}$ with

(a) $V(\hat{\mathcal{C}}^{\Lambda}) \subset V(\mathcal{C})$ contains precisely those positions which lie on a $w_0 \to w_\infty$ path $\lambda \in \Lambda$;
(b) $E(\hat{\mathcal{C}}^{\Lambda}) \subset E(\mathcal{C})$ contains precisely those edges which lie on a $w_0 \to w_\infty$ path $\lambda \in \Lambda$;

(c) For $w_1, w_2 \in V(\hat{\mathcal{C}}^\Lambda)$, and $e(w_1, w_2) \in E(\hat{\mathcal{C}}^\Lambda)$, the edge $e(w_1, w_2)$ has primitive probability $\pi_e(w_2 \mid w_1)$ replaced by $\hat{\pi}_e^\Lambda(w_2 \mid w_1)$.

Clearly there are many probability distributions over $\hat{\mathcal{C}}^\Lambda$ that could be assigned via Definition 5(c) which do not correspond to any practical or theoretical manipulation. As with CBNs we make causal assumptions about the system. So in the Sprinkler example, we know that we cannot control the season; we might be able to control whether it rains or not (or at least alter the probability distributions over these outcomes) by seeding clouds; we can definitely control whether the sprinkler is on or off; and can also alter the probability distributions over whether the path is wet or not. The allowable edge-probabilities in our manipulated CEG are governed by our causal assumptions. In Example 3 we specified that the only aspect of the system being controlled was the sprinkler, so these are the only edge-probabilities that might alter, and their new values are determined by our causal model. The probability following manipulation of any atom λ is equal to the probability of the corresponding $w_0 \to w_\infty$ path in the manipulated CEG $\hat{\mathcal{C}}^\Lambda$. This is, as before, expressible as the product of probabilities of values of (stage) variables conditioned on the values of their parents, so $\Pi(\hat{\mathcal{C}}^\Lambda)$ satisfies the Directed Markov condition with respect to $\hat{\mathcal{C}}^\Lambda$.

For completeness we also define a conditioned CEG (first defined in [25]).

Definition 6 (*Conditioned CEG*). For a CEG $\mathcal{C}$ and intrinsic event Λ, let $\mathcal{C}^\Lambda$ (the CEG conditioned on the event Λ) be the subgraph of $\mathcal{C}$ with $V(\mathcal{C}^\Lambda), E(\mathcal{C}^\Lambda)$ defined and coloured analogously with $V(\hat{\mathcal{C}}^\Lambda)$, $E(\hat{\mathcal{C}}^\Lambda)$ in Definition 5, and

(c) For $w_1, w_2 \in V(\mathcal{C}^\Lambda)$, and $e(w_1, w_2) \in E(\mathcal{C}^\Lambda)$, the edge $e(w_1, w_2)$ has primitive probability $\pi_e(w_2 \mid w_1)$ replaced by

$$\pi_e^\Lambda(w_2 \mid w_1) = \frac{\sum_{\lambda \in \Lambda} \pi(\lambda, \Lambda(e(w_1, w_2)))}{\sum_{\lambda \in \Lambda} \pi(\lambda, \Lambda(w_1))}$$

Probabilities in $\mathcal{C}$ are denoted π, in $\mathcal{C}^\Lambda$ are denoted π^Λ, and in $\hat{\mathcal{C}}^\Lambda$ are denoted $\hat{\pi}^\Lambda$.

If instead of manipulating to Λ in Example 3, we had simply observed that the sprinkler was on, then in the conditioned CEG $\mathcal{C}^\Lambda$ all edges associated with the variables *season* or *rain* would retain their original probabilities, but edges associated with the variables *wet* and *slippery* would have new probabilities as detailed in Definition 6(c).

Causal analysis on CEGs was first discussed in [26], but the emphasis in the earlier paper was on manipulations which were exact analogues of BN interventions, and on manipulations to sets of positions within the CEG (a rather small subset of the set of intrinsic events). This was a very narrow focus, and the manipulations considered here are more generic — the manipulation *Put sprinkler on* in Example 3 could be thought of as a manipulation to a set of positions, but the second manipulation in this example cannot be characterised in this way — it is a functional manipulation which, as noted above, can only be expressed as a BN by sacrificing some of the simplicity of the DAG-representation.

3. The Back Door theorem

Since 1995 there has been considerable effort put in to finding conditions for causal identifiability on BNs [6,16,17,27,28] — that is conditions for when the effects of a manipulation can be estimated from a subset of variables observed in the idle system. The initial spur for this activity was the publication of Pearl's Back Door theorem in [14], which provided sufficient conditions for such an analysis. The advantages in using Pearl's formulation are threefold: the factors needing to be considered in the analysis are clearly identified, the probability of observing an effect following a manipulation is expressed in terms of the idle or unmanipulated system, and the conditions for the analysis to be valid can be checked on the unmanipulated BN. In particular, when a manipulation is impossible or unethical in practice, or its effects difficult or impossible to observe, an analyst may still be able to estimate the probabilities of the theoretically possible effects of this manipulation.

Pearl's Back Door theorem states that under certain conditions on sets of variables X, Y, Z, we can write down the probability expression

$$p(y \parallel x) = \sum_z p(y \mid x, z)\, p(z)$$

where $p(y \parallel x)$ denotes the probability of observing that an effect or response variable Y takes the value y, following a manipulation of the variable X to a specified value x. The manipulation here is sometimes described as the *control* or *setting* of the variable X to the value x, and the notation is that of Lauritzen [11]. By careful choice of the set Z we may be able to calculate or estimate $p(y \parallel x)$ without conditioning on the full set of measurement variables.

As the Back Door theorem for BNs is the tool most widely used by causal analysts, a CEG-analogue is a principal focus of this paper. We have already noted that the CEG is a very good means of depicting asymmetric problems (e.g. Example 1), and that it also allows for transparent representation of what might be termed asymmetric manipulations (e.g. Example 3). The Back Door theorem for CEGs presented here also allows the analyst a fair amount of flexibility in the choice of the

CEG-analogue of Z. The topology of the CEG can be utilised to find functions of the data which can be observed in the idle system and which fulfil this role — these do not need to be vectors of values of the measurement variables of the problem.

Our Back Door theorem also refers explicitly to manipulation **to** events (such as *Put sprinkler on*) rather than manipulation **of** variables (set the variable *sprinkler* to the value corresponding to *on*), which reflects the difference in topology between the CEG and the BN. A primitive version of a Back Door theorem was given in [26], but this was restricted to manipulations to sets of positions. The conditions required for this theorem were also very complex, and not easily checkable on the topology of the CEG. The Back Door theorem presented here works for a far larger class of manipulations, the conditions are simpler, and they can be checked on the topology of the CEG.

The use of the description *Back Door* is not perhaps an obvious choice for our CEG-analogue, but we retain it to emphasise its affinity with the Back Door theorem for BNs.

So consider a manipulation to an event Λ_x (a specified union of root-to-sink paths). Suppose we wish to find the probability of (observing) an event Λ_y given that the manipulation to Λ_x has been enacted — that is we wish to produce an expression for $\pi(\Lambda_y \parallel \Lambda_x)$. This is equal to the probability of the event Λ_y on the CEG $\hat{\mathcal{C}}^{\Lambda_x}$, which is the sum of the probabilities of the $w_0 \to w_\infty$ paths in $\hat{\mathcal{C}}^{\Lambda_x}$ which are consistent with the event Λ_y:

$$\pi(\Lambda_y \parallel \Lambda_x) = \hat{\pi}^{\Lambda_x}(\Lambda_y)$$

Consider a partition of the atomic events ($w_0 \to w_\infty$ paths in $\mathcal{C}$) $\{\Lambda_z\}$. Then

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \hat{\pi}^{\Lambda_x}\left(\bigcup_z \Lambda_z, \Lambda_y\right) = \sum_z \hat{\pi}^{\Lambda_x}(\Lambda_z, \Lambda_y)$$

since the events $\{\Lambda_z\}$ form a partition of Ω

$$= \sum_z \hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda_z)\, \hat{\pi}^{\Lambda_x}(\Lambda_z)$$

Definition 7 (*Back Door partition*). The partition $\{\Lambda_z\}$ forms a *Back Door partition* of Ω if

$$\pi(\Lambda_y \parallel \Lambda_x) = \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z)\pi(\Lambda_z)$$

Note that this expression holds if

(A) $\hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda_z) = \pi(\Lambda_y \mid \Lambda_x, \Lambda_z)$;
(B) $\hat{\pi}^{\Lambda_x}(\Lambda_z) = \pi(\Lambda_z)$

for all $\Lambda_z \in \{\Lambda_z\}$.

The sets of variables Z in the BN-based Back and Front Door theorems are called *blocking* sets because they *block* certain paths between X and Y in the BN. Z also blocks the effect on Y of other problem variables so that they can be ignored when calculating the manipulated probability expression $p(y \parallel x)$. In our CEG-analogue, the blocking set becomes a partition of the $w_0 \to w_\infty$ paths of the CEG into sets $\{\Lambda_z\}$. This is not so very different from the BN version, where the events $Z = z$ (over the outcome space of Z) partition the set of all possible vectors of problem variable values. As with the BN version, if we choose $\{\Lambda_z\}$ carefully, we can calculate or estimate $\pi(\Lambda_y \parallel \Lambda_x)$ from a partially observed idle system.

3.1. Singular manipulations

The commonest form of manipulation on a BN is Pearl's $Do\, X = x$ intervention, where a variable X is set to a specific value x. There are also *functional* manipulations of the form *If* $Z = z_1$, *do* $X = x$, or *If* $Z = z_1$, *do* $X = x_1$; *if* $Z \neq z_1$, *do* $X = x_2$ etc. In each of these cases a variable X is set to a specified value. In a CEG this is equivalent to manipulating the system so as to pass through a set of edges carrying a specified label, so for instance in Example 3, the intervention $Do\, X_3 = 1$ (*Put Sprinkler on*) imposes a probability of 1 on all the edges labelled *on*.

Pearl's *Do* interventions and functional manipulations when enacted on a CEG are examples of *singular* manipulations. These are manipulations where every $w_0 \to w_\infty$ path passes through *one* of a collection of positions, and the manipulation imposes a probability of 1 on *one* edge emanating from each of these positions.

The class of singular manipulations includes more subtle interventions than these, so for instance in Example 3 the intervention *If it is raining, put the sprinkler off*; *otherwise make pavement wet* (*If* $X_2 = 1$ (*say*), *do* $X_3 = 0$; *otherwise do* $X_4 = 1$ (*say*)) is a singular manipulation. Whereas this would involve quite complicated analysis on a BN, with a CEG it is no more difficult than the simple $Do\, X_3 = 1$. *Stochastic* manipulations are not singular. These are interventions which impose a new probability distribution on the values of the variable X, or in the case of the CEG on the outcome spaces for a selection of positions. Seeding the clouds in Example 3 would be an example of this type of intervention.

Definition 8 (*Singular manipulation*). A manipulation of a CEG $\mathcal{C}$ to an event Λ is called *singular* if there exist sets $W \subset V(\mathcal{C})$, $E_\Lambda \subset E(\mathcal{C})$ such that

(i) the elements of W partition Ω (i.e. every $w_0 \to w_\infty$ path in $\mathcal{C}$ passes through precisely one $w \in W$),
(ii) for each $w \in W$, there exists precisely one emanating edge $e(w, w')$ which is an element of E_Λ,
(iii) Λ is the union of precisely those $w_0 \to w_\infty$ paths that pass through some $e(w, w') \in E_\Lambda$,
(iv) all edge probabilities in $\hat{\mathcal{C}}^\Lambda$ are equal to the corresponding edge probabilities in $\mathcal{C}$, except that $\hat{\pi}_e^\Lambda(w' \mid w) = 1$ for $w \in W$, $e(w, w') \in E_\Lambda$.

So in Example 3 our *Put Sprinkler on* manipulation is singular, with $W = \{w_3, w_4, w_5, w_6\}$, $E_\Lambda = \{e(w_3, w_7), e(w_4, w_9), e(w_5, w_7), e(w_6, w_9)\}$ is the set of edges labelled *on*. Λ is then the set of $w_0 \to w_\infty$ paths defined by the CEG in Fig. 5, where all edge-probabilities are as in Fig. 3 except those on edges labelled *on*, which have a probability of 1.

Similarly, for the intervention *If it is raining, put the sprinkler off; otherwise make pavement wet*, $W = \{w_3, w_5, w_9, w_{10}\}$ and $E_\Lambda = \{e(w_3, w_8), e(w_5, w_8), e(w_9, w_{11}), e(w_{10}, w_{11})\}$. Λ is the set of $w_0 \to w_\infty$ paths defined by the manipulated CEG produced from that in Fig. 3 by pruning the position w_7 and all edges entering or leaving w_7, as well as the edges $e(w_9, w_{12})$ and $e(w_{10}, w_{12})$. Edge-probabilities are as in Fig. 3 except for edges in E_Λ, which have a probability of 1.

3.2. A Back Door theorem for singular manipulations

As we also consider effect events (Λ_y) and conditioning sets (Λ_z), we distinguish our manipulation event Λ by adding a suffix to give Λ_x. We also relabel the set W as W_X, the positions within W_X as w_X, and the edges of Definition 8(ii) as $e(w_X, w'_X)$.

As the set of positions in W_X partitions Ω, we can consider a random variable X, defined on Ω, which takes values labelled by the emanating edges of w_X (for each w_X) with probabilities dependent on the history of the problem up to that position w_X (in Example 2, the histories up to the positions w_7 and w_8 are respectively *rain, on* and *rain, off*, and the probability of the pavement being *wet* is different for these two histories).

The singular manipulation to Λ_x assigns a probability of 1 to one of the values of X at each w_X, dependent on the history of the problem up to that position w_X. So Λ_x is of the form

$$\Lambda_x \equiv \bigcup_{w_X \in W_X} \Lambda\bigl(e\bigl(w_X, w'_X\bigr)\bigr)$$

We are also interested in an *effect event*, the probability of which may be altered by our manipulation. We can define this in exactly the same way as we defined Λ_x. So we define an effect variable Y, and a set of positions W_Y (such that $w_Y \not< w_X$ for all $w_X \in W_X$, $w_Y \in W_Y$) which partitions Ω (i.e. every $w_0 \to w_\infty$ path in $\mathcal{C}$ passes through one of the positions in W_Y). Then Λ_y consists of all paths that passing through some $w_Y \in W_Y$, utilise some prespecified edge emanating from that w_Y. So Λ_y is of the form

$$\Lambda_y \equiv \bigcup_{w_Y \in W_Y} \Lambda\bigl(e\bigl(w_Y, w'_Y\bigr)\bigr)$$

Pearl's Back Door Theorem depends on the assumption that the CBN expresses a causal model, and that its edges represent causal dependencies. Given this assumption the conditions for the theorem are then expressed in terms of the topology of the unmanipulated CBN. Since then Dawid [5] and Lauritzen [11] have produced versions of the Back Door theorem which incorporate similar causal assumptions but express the conditions in terms of conditional independence statements.

Dawid and Lauritzen express their conditions with reference to *augmented* DAGs, but it is also possible to produce conditions in the form of conditional independence statements on unaugmented DAGs. So, assuming that our CBN is a valid expression of our causal model, and that Pearl's two Back Door conditions hold, then the following conditional independence properties can be read from the DAG of the BN:

$$Z \perp\!\!\!\perp X \mid Q(X) \quad \text{and} \quad Y \perp\!\!\!\perp Q(X) \mid (X, Z)$$

where Z is the Back Door blocking set, and $Q(X)$ are the variable parents of X. Note that we are here ignoring the possibility that $Z \equiv Q(X)$. We return to this case in Section 3.4.

As noted in Section 2.1, and in direct analogy with CBNs, causal modelling with CEGs requires acceptance of the assumption that the CEG expresses a causal model. If this is the case then these conditional independence statements can be used to produce *Back Door* conditions on the topology of the CEG.

The first of our properties tells us that

$$p(z \mid q_1) = p(z \mid q_1, x)$$

where x is any value of X, z is any vector of values of the variables Z, and q_1 is any specified vector of values of the variables $Q(X)$.

The **events** $X = x$, $Z = z$, $Q(X) = q_1$ have direct analogues for singular manipulations of CEGs: $X = x$ corresponds to the event Λ_x (the union of all paths passing through an edge to which the manipulation will assign a probability of 1); $Z = z$ corresponds to Λ_z (an element in our Back Door partition); and noting that positions store the history of a process up to that point (and so in particular the values taken by any *parent variables* at that position), $Q(X) = q_1$ corresponds to $\Lambda(w_{X(1)})$, where $w_{X(1)}$ is a specific $w_X \in W_X$. So $p(z \mid q_1)$ becomes $\pi(\Lambda_z \mid \Lambda(w_{X(1)}))$, the probability of observing that the event Λ_z has occurred, given that the event $\Lambda(w_{X(1)})$ has occurred. The term $p(z \mid q_1, x)$ is translated similarly, and the expression $p(z \mid q_1) = p(z \mid q_1, x)$ becomes

$$\begin{aligned}\pi\left(\Lambda_z \mid \Lambda(w_{X(1)})\right) &= \pi\left(\Lambda_z \mid \Lambda(w_{X(1)}), \Lambda_x\right) \\ &= \pi\left(\Lambda_z \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda\big(e(w_X, w'_X)\big)\right) \\ &= \pi\left(\Lambda_z \mid \Lambda\big(e(w_{X(1)}, w'_{X(1)})\big)\right) \qquad (3.1)\end{aligned}$$

for any position $w_{X(1)} \in W_X$.

Condition (3.1) translates to the topology of the CEG as: The probability (in the idle CEG) of the event Λ_z conditioned on passing through the position w_X (in W_X) must equal the probability (in the idle CEG) of Λ_z conditioned on passing through the edge $e(w_X, w'_X)$, where this is the edge leaving w_X assigned a probability of 1 by the manipulation. This condition must hold for all $w_X \in W_X$.

The second of our properties tells us that

$$p(y \mid x, z) = p(y \mid q_1, x, z)$$

where y is any value of Y, and the event $Y = y$ corresponds to the event Λ_y. Substituting into this expression gives

$$\begin{aligned}\pi(\Lambda_y \mid \Lambda_x, \Lambda_z) &= \pi\left(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x, \Lambda_z\right) \\ &= \pi\left(\Lambda_y \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda\big(e(w_X, w'_X)\big), \Lambda_z\right) \\ &= \pi\left(\Lambda_y \mid \Lambda\big(e(w_{X(1)}, w'_{X(1)})\big), \Lambda_z\right) \qquad (3.2)\end{aligned}$$

and

$$\pi\left(\Lambda_y \mid \Lambda\big(e(w_{X(1)}, w'_{X(1)})\big), \Lambda_z\right) = \pi\left(\Lambda_y \mid \Lambda\big(e(w_{X(2)}, w'_{X(2)})\big), \Lambda_z\right)$$

for any positions $w_{X(1)}, w_{X(2)} \in W_X$.

Condition (3.2) translates as: The probability (in the idle CEG) of the event Λ_y ($Y = y$) conditioned on Λ_z and on passing through an edge $e(w_X, w'_X)$ is independent of which $w_X \in W_X$ this edge emanates from.

Theorem 1 *(Back Door theorem). With W_X, W_Y, Λ_x, Λ_y detailed as above, and $\{\Lambda_z\}$ a partition of Ω, then $\{\Lambda_z\}$ is a Back Door partition if conditions* (3.1) *and* (3.2) *hold for all elements of* $\{\Lambda_z\}$, $w_X \in W_X$.

A proof of this theorem appears in Appendix A.

Recall that Pearl's Back Door conditions for BNs refer explicitly to sets of problem (or measurement) variables, and that our conditions (3.1) and (3.2) are expressed in terms of events. An event here may be as simple as a measurement variable taking a particular value, but may just as easily correspond to some collection of variables taking some combination of values given one set of possible prior developments (histories), and another (perhaps overlapping) collection of variables taking a different combination of values given a different set of possible prior developments.

So our CEG-based Back Door theorem does not restrict analysis to manipulations which apply to one variable or to a specific set of variables, nor to blocking sets which consist only of problem variables. In Example 4 (below) there is no Back Door blocking set composed of problem variables which we can use, but we can still produce an identifiable probability expression using our event-based Back Door theorem and conditions (3.1) and (3.2).

3.3. Checking the conditions for the Back Door theorem

One of the strengths of Pearl's Back Door theorem is that his conditions can be checked directly on the topology of the idle BN. The conditions for our Back Door theorem for singular manipulations can also be checked on the topology of the unmanipulated CEG. This is best illustrated by example, and this is done in Example 4 below. Before doing this we look at condition (3.2) in a little more detail.

An advantage of letting $\{\Lambda_z\}$ be a collection of events is that we have considerable flexibility in what type of events we choose. Each Λ_z could be of the form $\Lambda(w)$ (the union of all root-to-sink paths passing through the position w), $\Lambda(e)$ (the

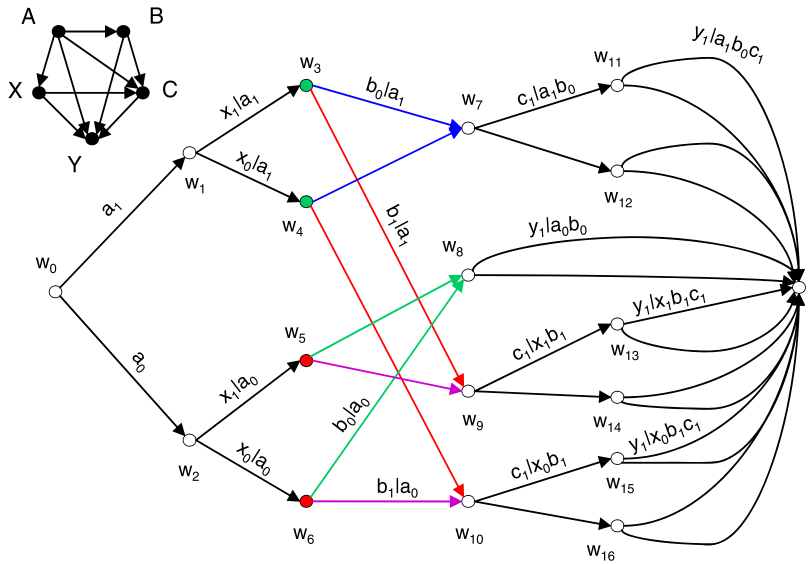

Fig. 7. BN and CEG $\mathcal{C}$ for Example 4.

union of all root-to-sink paths passing through the edge e), some union of such events, or a union of root-to-sink paths which fits none of these descriptions.

For the remainder of Section 3 we use partitions where each Λ_z is an event associated with a collection of positions, and in Section 3.3 in particular, concentrate on sets of positions *downstream* of the manipulation (in that none of these positions precedes any $w_X \in W_X$ on any root-to-sink path). Translating this work to events associated with edges, and to events associated with positions *upstream* of the manipulation (no positions succeed any $w_X \in W_X$ on any root-to-sink path) is straightforward. Some results on *upstream* positions are provided in Section 3.4 (and this was also the focus of the basic Back Door theorem presented in [26]).

We show here how we can choose each member of Λ_z (associated with a collection of positions) so that condition (3.2) is automatically satisfied.

Let each element Λ_z in our Back Door partition be a union of events of the form $\Lambda(w)$ for some (small) collection of positions. Formally, let $V_z \subset V(\mathcal{C})$ be a set of positions which partitions Ω, such that V_z lies *downstream* of W_X and *upstream* of W_Y (*downstream* and *upstream* used as above). Now, whereas all elements of V_z exist in $\mathcal{C}$, not all will exist in $\hat{\mathcal{C}}^{\Lambda_x}$. So consider a partition of V_z into a collection $\{V_z^1, \dots, V_z^N\}$, where N is the number of elements of V_z which exist in $\hat{\mathcal{C}}^{\Lambda_x}$, and each V_z^i contains precisely one element of V_z which exists in $\hat{\mathcal{C}}^{\Lambda_x}$ (plus some or none of the elements of V_z which do not exist in $\hat{\mathcal{C}}^{\Lambda_x}$).

Now let $\{\Lambda_z\}$ consist of N elements Λ_z^i $(i = 1, \dots, N)$, where

$$\Lambda_z^i = \bigcup_{w_z \in V_z^i} \Lambda(w_z)$$

If we choose our Back Door partition $\{\Lambda_z\}$ in this way then condition (3.2) is satisfied. A demonstration of this is provided in Appendix A.

Example 4 (*Using the Back Door theorem*). We illustrate the use of our Back Door theorem through a medical example. As with earlier examples we use binary variables for illustrative convenience.

Our interest is in a condition which can manifest itself in one of two forms ($C = 1$ or 2). Individuals who will as adults develop the condition (in either of its forms) display either symptom S_A before the age of ten, or S_B in their late teens, or both. Whether or not an individual displays S_A is labelled by a variable A, and whether or not they display S_B by a variable B. In both cases the variable takes the value 1 if the symptom is displayed, and the value 0 if it is not. There is a treatment T available which has some efficacy if given in an individual's early teens. Being treated is labelled $X = 1$, and not treated $X = 0$. We have a life expectancy indicator Y, and dying before the age of fifty is labelled $Y = 1$, dying at fifty or older $Y = 2$.

The relationships between the variables A, X, B, C and Y are described below, and are portrayed by the CEG in Fig. 7, where for convenience edges are labelled a_0 for $A = 0$ etc.

Symptom S_A is often missed by doctors, but if it is detected an individual is more likely to be given treatment T. We therefore do not know the distributions of $A, X \mid A = 0$ or $X \mid A = 1$. We do know however that $X \not\perp\!\!\!\perp A$.

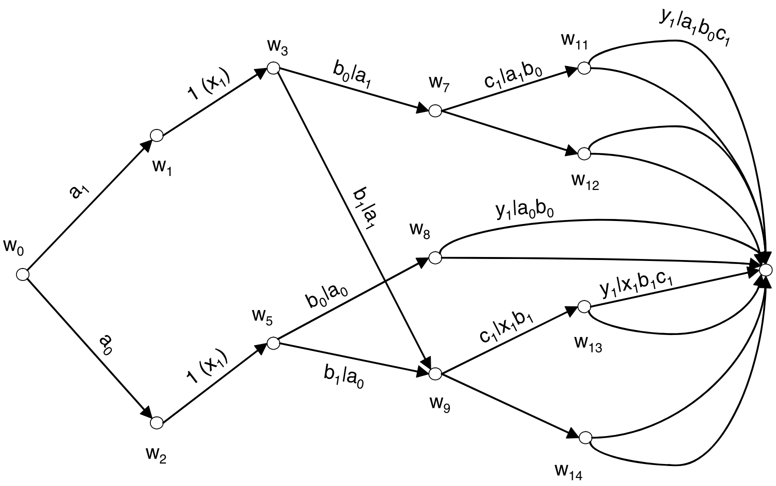

Fig. 8. Manipulated CEG $\hat{\mathcal{C}}^{\Lambda_x}$ for Example 4.

Evidence from previous studies indicates that

- whether or not an individual displays symptom S_B depends only on whether or not they displayed symptom S_A ($B \amalg X \mid A$),
- displaying either symptom means that an individual **will** develop the condition in one of its two forms,
- for individuals displaying S_A but not S_B, developing the condition in form 1 does not depend on whether or not they had treatment T ($C \amalg X \mid A = 1, B = 0$). Also, how long they live depends only on which form of the condition they develop ($Y \amalg X \mid A = 1, B = 0, C$),
- for individuals displaying S_B, developing the condition in form 1 does not depend on whether or not they displayed S_A, irrespective of whether they were treated or not ($C \amalg A \mid X, B = 1$). Also, how long they live depends on whether or not they were treated and on which form of the condition they develop ($Y \amalg A \mid X, B = 1, C$).

If we were to attempt to portray the problem via a BN it would look like the one in Fig. 7. Without considerable annotation the BN cannot express the context-specific conditional independence structure illustrated by the CEG.

We are interested in the effects on life expectancy if we were to treat everybody in the population in their early teens. So we consider the singular manipulation to Λ_x equivalent to $Do\, X = 1$, and calculate the probability $\pi(\Lambda_y \parallel \Lambda_x) \equiv P(Y = 1 \parallel X = 1)$. The CEG satisfies the conditions that every path passes through a position from $W_X = \{w_1, w_2\}$ and a position from $W_Y = \{w_8, w_{11}, w_{12}, \ldots, w_{16}\}$. Also, every position in W_X has an outgoing edge labelled x_1 ($X = 1$), and every position in W_Y has an outgoing edge labelled y_1 ($Y = 1$).

Clearly A is a required variable in any Back Door blocking set Z based on the BN representation of the problem. But from above we do not know the distribution of A or of any joint distribution involving A. Can we use our Back Door theorem for CEGs to find an identifiable expression not involving A?

In these situations we generally have a lot of flexibility in determining our Back Door partition/blocking set, and some experimentation may be needed before we find the ideal allocation. Here we consider Λ_z of the form $\bigcup \Lambda(w)$. The choice of positions will depend on what we can observe, and may be heavily influenced by observation costs. Note that the connection between these constraints and our choice of positions can be very subtle — in this example we clearly cannot estimate $P(A = 1, B = 0, C = 1)$, but we can still include the position w_{11} in our blocking set. Here we simply imagine that these constraints and our experimentation have produced a blocking set of positions V_z, lying between W_X and W_Y, comprising $\{w_8, w_9, w_{11}, w_{12}, w_{15}, w_{16}\}$. The CEG $\hat{\mathcal{C}}^{\Lambda_x}$ is given in Fig. 8.

Checking condition (3.2) on the graph. There are 6 positions in V_z, but only 4 of these appear in $\hat{\mathcal{C}}^{\Lambda_x}$. We can combine the original 6 positions to produce 4 events of the form $\Lambda_z = \bigcup \Lambda(w)$ in 65 different ways, and our choice of which way will depend on a number of factors including whether a particular combination satisfies condition (3.1).

Here we have chosen $\Lambda_z^1 = \Lambda(w_8)$, $\Lambda_z^2 = \Lambda(w_{11})$, $\Lambda_z^3 = \Lambda(w_{12})$, $\Lambda_z^4 = \Lambda(w_9) \cup \Lambda(w_{15}) \cup \Lambda(w_{16})$. The positions w_8, w_9, w_{11}, w_{12} are the members of V_z that exist in $\hat{\mathcal{C}}^{\Lambda_x}$, and w_{15}, w_{16} are those that don't. So, using the result from the paragraphs immediately preceding Example 4, this choice of $\{\Lambda_z\}$ satisfies condition (3.2).

Checking condition (3.1) on the graph. In Fig. 7 we have, as in earlier figures, labelled edges with probabilistic descriptions such as $c_1|a_1b_0$. This is not actually necessary — the forms of these probabilities are completely specified by the topology and colouring of the CEG. In checking condition (3.1) we do not need to refer to the labels on the edges, or to probability tables or any separate lists of conditional independence properties.

For (3.1) we need to show that $\pi(\Lambda_z \mid \Lambda(w_X)) = \pi(\Lambda_z \mid \Lambda(e(w_X, w'_X)))$ for each w_X and each Λ_z. This can be done in a purely graphical manner.

- Does $\pi(\Lambda(w_{11}) \mid \Lambda(w_1)) = \pi(\Lambda(w_{11}) \mid \Lambda(e(w_1, w_3)))$?
  $\pi(\Lambda(w_{11}) \mid \Lambda(e(w_1, w_3))) = \pi(\Lambda(w_{11}) \mid \Lambda(w_3))$ by Lemma 1. To get from w_1 to w_{11} we go through either w_3 or w_4 (with probability 1), and then follow a *blue* edge, and then the edge $e(w_7, w_{11})$.
  To get from w_3 to w_{11} we follow a *blue* edge, and then the edge $e(w_7, w_{11})$. So these probabilities are equal.
- Similarly for w_{12}.
- $w_1 \not< w_8$.
- Does $\pi(\Lambda(w_9) \cup \Lambda(w_{15}) \cup \Lambda(w_{16}) \mid \Lambda(w_1)) = \pi(\Lambda(w_9) \cup \Lambda(w_{15}) \cup \Lambda(w_{16}) \mid \Lambda(e(w_1, w_3)))$? From Fig. 7 we see that $\Lambda(w_{15}) \cup \Lambda(w_{16}) = \Lambda(w_{10})$. To get from w_1 to w_9 or w_{10} we go through either w_3 or w_4 (with probability 1), and then follow a *red* edge.
  $w_3 \not< w_{10}$, and to get from w_3 to w_9 we follow a *red* edge. So these probabilities are equal.

Similar quick checks for $w_X = w_2$ confirm that $\{\Lambda_z\}$ satisfies condition (3.1). Our manipulated probability expression

$$p(y_1 \parallel x_1) = \pi(\Lambda_y \parallel \Lambda_x) = \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z)\pi(\Lambda_z)$$

is evaluated on $\mathcal{C}$, and simplifies to

$$p(b_0)\ p(y_1 \mid b_0) + p(b_1)\ p(y_1 \mid x_1 b_1)$$

So we need only know the distribution of B (the incidence of symptom S_B), and the conditional distributions of Y (life expectancy) on the events $B = 0$ (S_B not displayed) and $X = 1, B = 1$ (treated and S_B displayed). This expression does not involve A (the incidence of S_A), and interestingly neither does it involve C (which form the condition takes). It does however involve B, which would be impossible if we used the BN from Fig. 7 for this model, as B does not block all Back Door paths from X to Y.

Note that this example gives an insight into how to choose the component Λ_z of our partition. If we can find w_z such that $\Lambda(w_z)$ satisfies

$$\pi\big(\Lambda(w_z) \mid \Lambda\big(e\big(w_X, w'_X\big)\big)\big) = \pi\big(\Lambda(w_z) \mid \Lambda(w_X)\big) \quad \forall\, w_X \in W_X$$

then we can make $\Lambda(w_z)$ a Λ_z.

Other Λ_z are produced by combining **one** position w_z that exists in $\hat{\mathcal{C}}^{\Lambda_x}$ with other positions $\{w_z\}$ that disappear when we create $\hat{\mathcal{C}}^{\Lambda_x}$, in such a way that the union of their associated events satisfies condition (3.1) for all $w_X \in W_X$.

3.4. Using W_X to create a Back Door partition

If we apply Pearl's Back Door theorem to the BN in Fig. 3 we could use X_1 (*Season*) as the blocking set for the manipulation $Do\, X_3 = 1$ described in Example 3. Here $X_1 = Q(X_3)$, the set of parents of X_3, and using a blocking set of this type gives us a revised Back Door expression

$$p(y \parallel x) = \sum_{q(x)} p\big(y \mid x, q(x)\big)\ p\big(q(x)\big)$$

where $q(x)$ runs over the possible vectors of values of $Q(X)$.

In our example we would get

$$\begin{aligned} p(\mathit{slippery} \parallel \mathit{sprinkler\ on}) = {} & p(\mathit{slippery} \mid \mathit{sprinkler\ on}, \mathit{Spring})\ p(\mathit{Spring}) \\ & + p(\mathit{slippery} \mid \mathit{sprinkler\ on}, \mathit{Summer})\ p(\mathit{Summer}) \end{aligned}$$

The CEG-analogue of this is to use W_X to create our Back Door partition. So, as in Section 3.3 we let each element Λ_z of our partition be a union of events of the form $\Lambda(w)$, but these positions are now members of W_X.

In Section 3.2 we suggested an analogy between $Q(X) = q(x)$ for BNs and $\Lambda(w_X)$ for CEGs. In fact this analogy is not perfect; a better analogy for parents in a BN is a set of *stages* rather than positions. Recall that two (or more) positions are in the same stage if the immediate future developments from these positions have the same probability distribution (i.e. their sets of emanating edges have the same colour scheme).

Let each element of our Back Door partition have the form

$$\Lambda_z = \bigcup_{w_X \in u_X} \Lambda(w_X) = \Lambda(u_X)$$

for some u_X (the event $\Lambda(u_X)$ is the union of all root-to-sink paths passing through some position w_X, a member of the stage u_X).

Note that (i) each $w_X \in W_X$ needs to be a member of some stage u_X, such that $\Lambda(u_X) \in \{\Lambda_z\}$, and (ii) each $w_X \in u_X$ (where $\Lambda(u_X) \in \{\Lambda_z\}$) needs to be an element of W_X. Also, for each $w_X \in u_X$, the manipulated edges $e(w_X, w'_X)$ must carry the same label (and hence *colour*). These labels can differ for different stages.

This is not actually particularly restrictive, as the set of manipulations we can consider still contains all basic *Do* interventions on BNs and all functional manipulations where the argument of the function is (a subset of) the parent set of the manipulated variable. In fact we can argue that this set contains **all** functional manipulations of a BN: If a manipulation is functional in that the value we manipulate X to depends on the value taken by another variable W, then essentially we have a *decision* problem and the BN representation of the system becomes an Influence Diagram (ID) representation with X as a decision node. Clearly the value of W must be known before X is manipulated, so in this ID representation there must be an edge from W to X (see for example [23]) and so W is a parent of X. Hence we argue that for all functional manipulations of BNs the argument of the function is (a subset of) the parent set of the manipulated variable.

Given these conditions on Λ_z, our new CEG-based Back Door probability expression is

$$\pi(\Lambda_y \parallel \Lambda_x) = \sum_{u_X} \pi\big(\Lambda_y \mid \Lambda(u_X), \Lambda_x\big)\pi\big(\Lambda(u_X)\big)$$

A demonstration of this result appears in Appendix A.

For the intervention $Do\, X_3 = 1$ (*Put sprinkler on*), our set $W_X = \{w_3, w_4, w_5, w_6\}$ (see Fig. 3), which separates into two stages $u_3 = \{w_3, w_4\}$ and $u_4 = \{w_5, w_6\}$. Now $\Lambda(u_3) = \Lambda(w_3) \cup \Lambda(w_4) = (\textit{Spring},\ \textit{rain}) \cup (\textit{Spring},\ \textit{no rain}) = (\textit{Spring})$, so our CEG-based expression is identical to the BN-based one above.

4. A Front Door theorem for CEGs

Pearl's Front Door theorem [14,15] for BNs can be used in cases where the Back Door theorem conditions do not hold or where the events needing to be observed for the Back Door theorem have too large an observational cost. Like the Back Door theorem, the Front Door theorem provides conditions for when the effects of a manipulation can be estimated from a subset of variables observed in the unmanipulated system.

Pearl's Front Door theorem states that under certain conditions on sets of variables X, Y, Z, we can write

$$p(y \parallel x) = \sum_z p(z \mid x) \sum_{x'} p\big(y \mid x', z\big)\, p\big(x'\big)$$

an expression whose value can be estimated from a partially observed idle system.

This expression is more complex than that for the Back Door theorem, and in our CEG-analogue this imposes greater restrictions on the types of manipulation we can consider. We concentrate here on singular manipulations.

The expression also suggests that we will need to sum over some variable corresponding to the variable X. Hence we need to produce a partition of Ω, of which Λ_x (the event to which we manipulate) is **one** element. So, for instance in Example 3 we would partition Ω into two events (Λ_x^1 and Λ_x^2) — the union of all root-to-sink paths passing through an edge labelled *on*, and the union of all root-to-sink paths passing through an edge labelled *off*.

For simplicity we look at manipulations of the form $Do\, X = x$, and consider positions $w_X \in W_X$ which each have the same number of emanating edges and these edges carry the same labels for each w_X (e.g. in Example 3 the positions w_3, w_4, w_5, w_6 each have emanating edges labelled *on* and *off*).

Note that even for fairly regular problems depictable by BNs there may be histories or parental configurations of a variable X for which the probability of a particular outcome is zero. Although normally we do not draw zero-probability edges in a CEG, in this case it is advisable to do so, although only for the edges emanating from those positions associated with the variable X.

Suppose we have a CBN which is a valid expression of our causal model, and that Pearl's Front Door conditions hold. Then, just as with the Back Door Theorem, there are two conditional independence properties that can be read from the DAG of the BN. These are

$$Y \amalg X \mid \big(Z, Q(X)\big) \quad \text{and} \quad Z \amalg Q(X) \mid X$$

Note that Z is normally a descendant of X and an ancestor of Y.

If we have a CEG which expresses our causal model then these conditional independence statements can be used to produce *Front Door* conditions on the topology of the CEG. The first of our properties tells us that

$$p(y \mid q_1, z) = p(y \mid q_1, x_i, z)$$

where as before y is any value of Y, q_1 is any specified vector of values of $Q(X)$, and z is any vector of values of Z. Here x_i is any value of X, and the event $X = x_i$ corresponds to an event on our CEG of the form $\Lambda_x^i = \bigcup_{w_X \in W_X} \Lambda(e(w_X, w_X^i))$, where for each $w_X \in W_X$, the edge $e(w_X, w_X^i)$ is the edge leaving w_X labelled x_i. The set of root-to-sink paths is partitioned by the $\{\Lambda_x^i\}$, which collection contains as many members as there are edges leaving any $w_X \in W_X$. The event Λ_x to which we manipulate is one element of this partition.

Substituting the CEG-analogues for these events from Section 3.2 into $p(y \mid q_1, z) = p(y \mid q_1, x_i, z)$ gives

$$\begin{aligned}\pi\big(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_z\big) &= \pi\big(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^i, \Lambda_z\big)\\ &= \pi\Bigg(\Lambda_y \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda\big(e\big(w_X, w_X^i\big)\big), \Lambda_z\Bigg)\\ &= \pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(1)}, w_{X(1)}^i\big)\big), \Lambda_z\big)\end{aligned} \tag{4.1}$$

and

$$\pi\big(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^i, \Lambda_z\big) = \pi\big(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^j, \Lambda_z\big)$$

for any position $w_{X(1)} \in W_X$ and any elements Λ_x^i, Λ_x^j of our partition (i.e. any edges leaving $w_{X(1)}$).

Note that $\{\Lambda_z\}$ is a (*Front Door*) partition of Ω, and at present we have imposed no further restrictions on the form of each Λ_z (as for example is done in Section 3.3).

Condition (4.1) translates to the topology of the CEG as: The probability (in the idle CEG) of the event Λ_y $(Y = y)$ conditioned on Λ_z and on passing along an edge emanating from $w_X \in W_X$ is not dependent on which edge is utilised. This condition must hold for all $w_X \in W_X$, and any Λ_z in our Front Door partition.

The second of our properties tells us that

$$p(z \mid x_i) = p(z \mid q_1, x_i)$$

Substituting into this expression gives

$$\begin{aligned}\pi\big(\Lambda_z \mid \Lambda_x^i\big) &= \pi\big(\Lambda_z \mid \Lambda(w_{X(1)}), \Lambda_x^i\big)\\ &= \pi\Bigg(\Lambda_z \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda\big(e\big(w_X, w_X^i\big)\big)\Bigg)\\ &= \pi\big(\Lambda_z \mid \Lambda\big(e\big(w_{X(1)}, w_{X(1)}^i\big)\big)\big)\end{aligned} \tag{4.2}$$

and

$$\pi\big(\Lambda_z \mid \Lambda(w_{X(1)}), \Lambda_x^i\big) = \pi\big(\Lambda_z \mid \Lambda(w_{X(2)}), \Lambda_x^i\big)$$

for any positions $w_{X(1)}, w_{X(2)} \in W_X$, and any element Λ_x^i of our partition (i.e. any edge leaving a $w_X \in W_X$).

Condition (4.2) translates as: The probability (in the idle CEG) of the event Λ_z conditioned on leaving a position in W_X via the edge labelled x_i is independent of which position $w_X \in W_X$ is passed through. This condition must hold for all edges x_i (i.e. values of X).

4.1. A Front Door theorem for singular manipulations

Theorem 2 *(Front Door theorem). If W_X, W_Y, Λ_y are detailed as in Section 3, $\{\Lambda_x^i\}$ is a partition of Ω with each individual Λ_x^i of the form detailed above, and $\{\Lambda_z\}$ is a partition of Ω which satisfies conditions (4.1) and (4.2) above, then*

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_i \pi\big(\Lambda_y \mid \Lambda_x^i, \Lambda_z\big)\pi\big(\Lambda_x^i\big)$$

We call such a $\{\Lambda_z\}$ a Front Door *partition.*

A proof of this theorem is in Appendix A.

Example 5 (*Using the Front Door theorem*). We here consider the example from [15] Section 3.3.3, but without reference to Pearl's hypothetical data. This example relates to the debate concerning the relationship between smoking and lung cancer summarised in [24].

In Pearl's example the vertices of the BN in Fig. 9 correspond to binary variables as follows:

$X = 1$: *smoker*, $X = 0$: *non-smoker*,
$Y = 1$: *lung cancer*, $Y = 0$: *no lung cancer*,
$B = 1$: *tar in lungs*, $B = 0$: *no tar in lungs*.

The variable A is associated with an unobservable genetic tendency, the presence of which $(A = 1)$ in an individual affects both the probability that the individual smokes and that they get lung cancer. The variable B by contrast is observable. Pearl

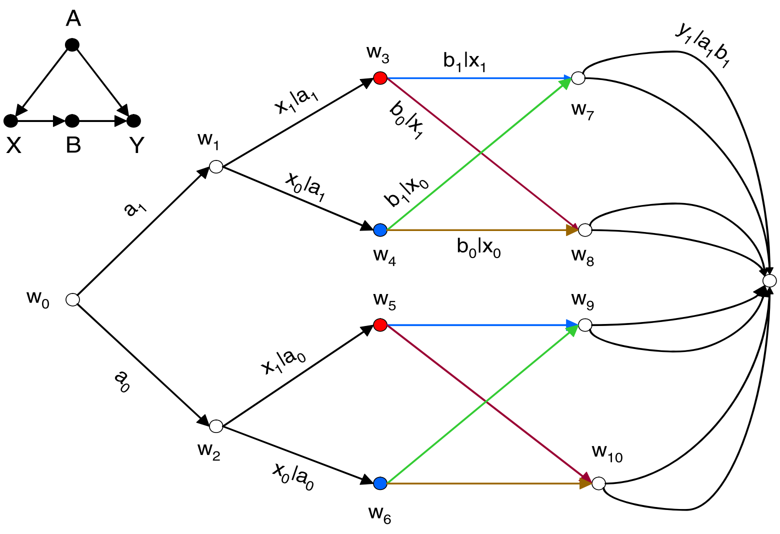

Fig. 9. BN and CEG $\mathcal{C}$ for Example 5.

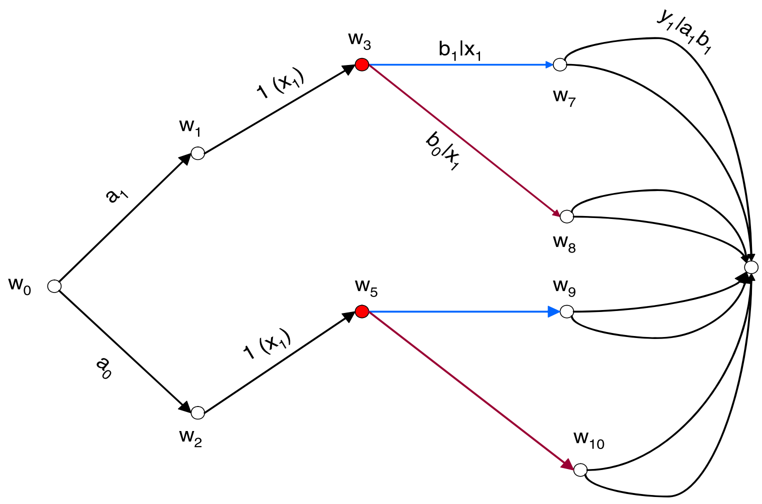

Fig. 10. Manipulated CEG $\hat{\mathcal{C}}^{\Lambda_x}$ for Example 5.

uses the BN to show that it is possible to estimate $p(\mathit{lung\ cancer} \parallel \mathit{smoker})$ from joint or conditional distributions of the variables X, B and Y even if there were to exist such an unobservable genetic tendency.

We demonstrate the use of the Front Door theorem for CEGs using this example. The unmanipulated CEG is given in Fig. 9, where as before edges are labelled a_0 for $A = 0$ etc. We consider the manipulation to Λ_x equivalent to $Do\, X = 1$, and use Theorem 2 to find an expression for $\pi(\Lambda_y \parallel \Lambda_x) \equiv P(Y = 1 \parallel X = 1)$. The manipulated CEG $\hat{\mathcal{C}}^{\Lambda_x}$ is given in Fig. 10.

Note that if A was observable we could use the Back Door theorem for CEGs here with $W_X = \{w_1, w_2\}$ doubling up as the blocking set (as in Section 3.4), which would be possible since each element of W_X is a distinct stage.

For our Front Door theorem we again have $W_X = \{w_1, w_2\}$, and our partition of Ω corresponding to the values of X is given by

Λ_x^1 is the union of all paths passing through an edge labelled x_1,

Λ_x^2 is the union of all paths passing through an edge labelled x_0.

The event Λ_y is the union of all paths passing through an edge labelled y_1.

We use the flexibility of CEG analysis to give each Λ_z a different form from that used in Section 3.3 — instead of being a union of events of the form $\Lambda(w)$, we make them a union of events of the form $\Lambda(e)$. So let

Λ_z^1 be the union of all paths passing through an edge labelled b_1,

Λ_z^2 be the union of all paths passing through an edge labelled b_0.

Checking condition (4.1) on the graph. $\pi(\Lambda_y \mid \Lambda(w_1), \Lambda_x^1, \Lambda_z^1)$ is the probability of taking an edge labelled y_1, given that we have passed through the position w_1, along an edge labelled x_1, and then an edge labelled b_1. Note that we do not need to know anything about the probabilities on these edges; nor do we need to refer to any separate list of conditional independence properties. Using Lemma 1 this probability is simply that of taking the upper edge leaving w_7 given that we have passed through w_7.

But this is clearly also equal to the probability $\pi(\Lambda_y \mid \Lambda(w_1), \Lambda_x^2, \Lambda_z^1)$.

Using the symmetry of the problem, condition (4.1) holds.

Checking condition (4.2) on the graph. $\pi(\Lambda_z \mid \Lambda(w_1), \Lambda_x^1)$ is the probability of taking an edge labelled b_1, given that we have passed through the position w_1, and along an edge labelled x_1. Using Lemma 1 this is the probability of a *blue* edge.

But clearly $\pi(\Lambda_z \mid \Lambda(w_2), \Lambda_x^1)$ is also the probability of a *blue* edge, so these probabilities are equal.

Using the symmetry of the problem, condition (4.2) holds.

Our manipulated probability expression from Theorem 2 is evaluated on $\mathcal{C}$, and is equal to

$$p(y_1 \parallel x_1) = \sum_b p(b \mid x_1) \sum_x p(y_1 \mid x, b)\ p(x)$$

So as Pearl found, the expression p(*lung cancer* $\parallel$ *smoker*) can be estimated from joint or conditional distributions of the variables X (*smoker*), B (*tar in lungs*) and Y (*lung cancer*) only.

4.2. An alternative form of the Front Door theorem

At the start of Section 4 we produced a partition of Ω of which Λ_x was one element, and confined ourselves to manipulations of the form $Do\,X = x$. This required us to consider positions $\{w_X\}$ which had the same number of emanating edges and where these edges carried the same label for each w_X. This restriction dilutes one of the powerful reasons for using CEGs for causal analysis — the possibility of analysing more complicated *functional* manipulations of the form $Do\,X = g(W)$ for some set of variables W. Corollary 1 gives an alternative Front Door expression which is appropriate for the full range of singular manipulations. It does however require knowledge of the probability distribution over the events $\{\Lambda(w_X)\},\ w_X \in W_X$, and joint or conditional distributions including these events. As these events do not always correspond to the values of some problem measurement variable, these distributions might not be easy to quantify. The Theorem 2 version of the Front Door expression requires knowledge of the probability distribution over the events $\{\Lambda_x^i\}$, which are often just the values of a problem variable X, and as such are more likely to be known.

Corollary 1. *If W_X, Λ_x, Λ_y are detailed as in Section* 3*, and $\{\Lambda_z\}$ is a partition of Ω which satisfies conditions* (4.1) *and* (4.2)*, then $\{\Lambda_z\}$ is a Front Door partition, and*

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_z\big)\pi\big(\Lambda(w_X)\big)$$

The proof of this corollary follows the proof of Theorem 2 until line (A.4).

Note that the partitions $\{\Lambda(w_X)\}$ and $\{\Lambda_x^i\}$ can differ considerably in size, with either partition being the larger. If we can estimate all the relevant probabilities, we can choose between the two versions of the Front Door theorem dependent on the relative size of these sets.

5. Discussion

As noted in the Introduction, there have been a number of recent advances in BN theory which concentrate on the representation and analysis of asymmetric problems, and on the analysis of controlled models. The CEG is presented here as a complementary graphical model, appropriate for analysis in both these areas.

In this section we consider when it is appropriate to use CEGs rather than BNs, and in particular when it is appropriate to use them for causal analysis rather than CBNs. We also consider how CEG-based causal analysis might develop in the future.

We use CEGs for problems which do not admit a natural product space structure, and where problem variables have different sets of possible outcomes (or no possible outcomes) given different vectors of values of ancestral variables. In Example 1, if neither warning light shows then the machine is not checked; and in Example 4, if an individual displays neither symptom they do not develop the condition in either of its forms. We also use the CEG when the degree of problem asymmetry is such that the topology of the associated BN yields little information of interest. In Example 1 the property that *whether judged faulty or not is independent of whether light* 1 *is on or off, given that light* 2 *is on* cannot be read from the DAG. In Example 4 the only conditional independence property readable from the BN is that $B \amalg X|A$. The four significant context-specific conditional independence properties are obscured.

We note here that every BN model is expressible as a CEG, but there are many problems where it is more sensible to use the former graph. This is certainly the case when problems are essentially symmetric and the benefits of a concise graph outweigh those of having an explicit representation of the outcome spaces of the problem variables.

We use CEGs for causal analysis when the idle system is itself better modelled via a CEG, so for example they are ideal for the analysis of asymmetric controlled models such as treatment regimes. We also use them to analyse the effects of asymmetric manipulations (e.g. functional and stochastic interventions). These manipulations tend to have a fairly simple representation on a CEG, but such analysis is not necessarily straightforward on a BN, particularly if both the manipulated variable and the value this variable takes are dependent on the values of other variables. These types of intervention often require the addition of edges to the DAG-representation of the problem, which can cause difficulties for an analyst trying to find suitable blocking sets. We also use CEGs for causal analysis when, using BN-based techniques, we cannot find identifiable expressions for manipulated probabilities.

Any manipulated probability identifiable as a result of using Pearl's Back Door theorem for a simple or functional manipulation of a BN is also identifiable from a CEG representation. However, as is the case for non-causal analysis, there are many problems for which CBNs are the most appropriate graph. The obvious example here is when the idle system is essentially symmetric, and the manipulations of interest are simple $Do\, X = x$ interventions; although even here there are occasions when a CEG-based analysis might be useful (as we indicate below).

Obviously we would use the CEG-based Back Door theorem if we were already using the CEG for our causal analysis. However there are other reasons for doing so. Pearl's Back Door conditions are easily checked on the topology of the CBN, but this asset diminishes if either there is a lot of context-specific information not encoded in the DAG of the CBN, or the manipulations of interest are more complex than $Do\, X = x$ interventions. Furthermore, using the Back Door theorem for BNs requires the analyst to be able to calculate or estimate $p(z)$ and $p(y \mid x, z)$ for all values z of the blocking set of variables Z. With the CEG our partition does not need to correspond to any fixed subset of the measurement variables that define a BN. This flexibility is very useful when some of the events in the system (X, Y and Z taking certain vectors of values) are unobservable or have large observational costs. It is also very useful if measurements of the system have not yet been taken, in that the identification of some usable **function** over the measurement variables can prevent the possibly expensive collection of unnecessary information. We have also seen in Example 4 that we can use the CEG version of the theorem in cases where it would be impossible to use the BN version, as the model does not obey the conditions specified by Pearl.

The reasons outlined here for using the CEG-based Back Door theorem apply equally to the CEG-based Front Door theorem.

As indicated in Example 4, where there were 65 possible partitions $\{\Lambda_z\}$ which satisfied condition (3.1) (even after restricting Λ_z to the form $\bigcup \Lambda(w_z)$ and specifying the positions $\{w_z\}$), our ability to choose our Back Door partition so that its elements do not correspond to vectors of values of problem variables gives us a lot of flexibility. Now our partition $\{\Lambda_z\}$ is *fixed* in the sense that its membership is constant. Suppose we let the membership depend in some way on whichever $w_X \in W_X$ our $w_0 \to w_\infty$ path passes through. If we could do this then our choice of possible partitions would be enormous.

It would also be useful to adapt our Back and Front Door theorems to produce workable versions for some of the non-singular manipulations of the type described in [26] Section 3.2; and to automate the search over the event space of the CEG for partitions $\{\Lambda_z\}$ which satisfy the conditions for these theorems.

Longer term, we aim to produce necessary conditions for causal identifiability, expressed as functions of the topology of the idle CEG. In this we will be mirroring the work of [6,17,27,28] on causal identifiability on BNs.

Acknowledgements

Much of the initial work leading to this paper was done whilst the author was supported by the EPSRC through grant no. EP/F036752/1. Thanks are due to the referees for their very useful comments which have helped to markedly improve the clarity of the paper.

Appendix A

Proof of Lemma 1. Consider a single root-to-sink path λ passing through w_1, w_2, w_3. This path consists of a set of edges, and we can assign a probability to it, equal to the product of the primitive probabilities labelling each of these edges. Call this probability $\pi(\lambda)$. Moreover, the probability of any subpath of λ is equal to the product of the primitive probabilities labelling each of its edges. So $\pi(\lambda)$ can be written as the product of the probabilities of four subpaths: $\mu_0(w_0, w_1)$, $\mu_1(w_1, w_2)$, $\mu_2(w_2, w_3)$, and $\mu_3(w_3, w_\infty)$. Thus

$$\pi(\lambda) = \pi_{\mu_0}(w_1 \mid w_0)\pi_{\mu_1}(w_2 \mid w_1)\pi_{\mu_2}(w_3 \mid w_2)\pi_{\mu_3}(w_\infty \mid w_3)$$

Consider now the event $\Lambda(w_1, w_2, w_3)$, which is the union of all root-to-sink paths passing through w_1, w_2, w_3. Since $\Lambda(w_1, w_2, w_3)$ is an intrinsic event (see Section 2.1) we have

$$\pi\big(\Lambda(w_1, w_2, w_3)\big) = \Bigg(\sum_{\mu_0 \in M_0} \pi_{\mu_0}(w_1 \mid w_0)\Bigg)\Bigg(\sum_{\mu_1 \in M_1} \pi_{\mu_1}(w_2 \mid w_1)\Bigg)$$
$$\times \Bigg(\sum_{\mu_2 \in M_2} \pi_{\mu_2}(w_3 \mid w_2)\Bigg)\Bigg(\sum_{\mu_3 \in M_3} \pi_{\mu_3}(w_\infty \mid w_3)\Bigg)$$

where M_i ($i = 0, 1, 2$) is the set of all subpaths from w_i to w_{i+1}, and M_3 is the set of all subpaths from w_3 to w_∞. But $\sum_{\mu_3 \in M_3} \pi_{\mu_3}(w_\infty \mid w_3)$ is simply the probability of reaching w_∞ from w_3, which equals 1. And $\sum_{\mu_0 \in M_0} \pi_{\mu_0}(w_1 \mid w_0)$ is the probability of reaching w_1 from w_0, which is $\pi(\Lambda(w_1) \mid \Lambda(w_0))$ etc. So

$$\pi\big(\Lambda(w_1, w_2, w_3)\big) = \pi\big(\Lambda(w_1) \mid \Lambda(w_0)\big)\pi\big(\Lambda(w_2) \mid \Lambda(w_1)\big)\pi\big(\Lambda(w_3) \mid \Lambda(w_2)\big) \times 1$$

Similarly

$$\pi\big(\Lambda(w_1, w_2)\big) = \pi\big(\Lambda(w_1) \mid \Lambda(w_0)\big)\pi\big(\Lambda(w_2) \mid \Lambda(w_1)\big) \times 1$$

and

$$\pi\big(\Lambda(w_3) \mid \Lambda(w_1), \Lambda(w_2)\big) = \frac{\pi(\Lambda(w_1, w_2, w_3))}{\pi(\Lambda(w_1, w_2))}$$

($\Lambda(w_1) \cap \Lambda(w_2) = \Lambda(w_1, w_2)$ etc. by construction)

$$= \pi\big(\Lambda(w_3) \mid \Lambda(w_2)\big) \qquad \Box$$

Proof of Theorem 1.

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_{w_X \in W_X} \hat{\pi}^{\Lambda_x}\big(\Lambda(w_X), \Lambda_y\big) = \sum_{w_X \in W_X} \hat{\pi}^{\Lambda_x}\big(\Lambda(w_X)\big)\, \hat{\pi}^{\Lambda_x}\big(\Lambda_y \mid \Lambda(w_X)\big)$$

since $\{\Lambda(w_X)\}$ form a partition of the atomic events

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\, \hat{\pi}^{\Lambda_x}\big(\Lambda_y \mid \Lambda(w_X)\big)$$

since every w_X lies *upstream* of our manipulation (Definition 8(iv))

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\, \hat{\pi}^{\Lambda_x}\big(\Lambda_y \mid \Lambda(w_X), \Lambda\big(w'_X\big)\big)$$

since $\Lambda(w_X) = \Lambda(e(w_X, w'_X)) \subset \Lambda(w'_X)$ in $\hat{\mathcal{C}}^{\Lambda_x}$

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\, \hat{\pi}^{\Lambda_x}\big(\Lambda_y \mid \Lambda\big(w'_X\big)\big)$$

using the form specified for Λ_y, the fact that $w_X \prec w'_X \prec w_Y$ for some $w_Y \in W_Y$ in $\hat{\mathcal{C}}^{\Lambda_x}$, and the result of Lemma 1.

From the definition of our manipulation, any edge lying on a $w'_X \to w_\infty$ path in $\mathcal{C}$ remains in $\hat{\mathcal{C}}^{\Lambda_x}$, and retains its original probability. Hence any set of path-segments starting at w'_X in $\hat{\mathcal{C}}^{\Lambda_x}$ corresponds to a set of path-segments in $\mathcal{C}$, and has the same probability as this set. Given the form specified for Λ_y, $\hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w'_X))$ is the probability of a set of path-segments starting at w'_X in $\hat{\mathcal{C}}^{\Lambda_x}$. Hence

$$\hat{\pi}^{\Lambda_x}\big(\Lambda_y \mid \Lambda\big(w'_X\big)\big) = \pi\big(\Lambda_y \mid \Lambda\big(w'_X\big)\big)$$

and

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\pi\big(\Lambda_y \mid \Lambda\big(w'_X\big)\big)$$
$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_X, w'_X\big)\big), \Lambda\big(w'_X\big)\big)$$

using the form specified for Λ_y, the fact that $e(w_X, w'_X) \prec w'_X \prec w_Y$ for some $w_Y \in W_Y$ in $\mathcal{C}$, and the result of Lemma 1

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_X, w'_X\big)\big)\big)$$

since $\Lambda(e(w_X, w'_X)) \subset \Lambda(w'_X)$ in $\mathcal{C}$

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big) \sum_z \pi\big(\Lambda_z, \Lambda_y \mid \Lambda\big(e\big(w_X, w'_X\big)\big)\big)$$

since $\{\Lambda_z\}$ form a partition of the atomic events

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big) \sum_z \pi\big(\Lambda_y \mid \Lambda\big(e\big(w_X, w'_X\big)\big), \Lambda_z\big)\pi\big(\Lambda_z \mid \Lambda\big(e\big(w_X, w'_X\big)\big)\big)$$

$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big) \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z)\pi\big(\Lambda_z \mid \Lambda(w_X)\big) \tag{A.1}$$

substituting from (3.1) and (3.2)

$$= \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z)\pi(\Lambda_z) \qquad \square$$

Satisfying condition (3.2). Let $\{\Lambda_z\}$ consist of N elements of the form $\Lambda_z^i = \bigcup_{w_z \in V_z^i} \Lambda(w_z)$ (for $i = 1, \ldots, N$), where for each V_z^i, only one position w_z exists in $\hat{\mathcal{C}}^{\Lambda_x}$. Call this position w_z^{i1}.

Consider the expression $\pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z^i)$, where $w_{X(1)}$ is a specified element of W_X, and the edge $e(w_{X(1)}, w'_{X(1)})$ is the edge emanating from $w_{X(1)}$ which is assigned a probability of 1 under our manipulation. Then

$$\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(1)}, w'_{X(1)}\big)\big), \Lambda_z^i\big) = \pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(1)}, w'_{X(1)}\big)\big), \Lambda\big(w'_{X(1)}\big), \Lambda_z^i\big)$$

since passing through the edge $e(w_{X(1)}, w'_{X(1)})$ necessarily entails passing through the position $w'_{X(1)}$

$$= \frac{\pi(\Lambda_z^i, \Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda(w'_{X(1)}))}{\pi(\Lambda_z^i \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda(w'_{X(1)}))}$$

$$= \frac{\pi(\Lambda_z^i, \Lambda_y \mid \Lambda(w'_{X(1)}))}{\pi(\Lambda_z^i \mid \Lambda(w'_{X(1)}))}$$

using Lemma 1. Hence

$$\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(1)}, w'_{X(1)}\big)\big), \Lambda_z^i\big) = \pi\big(\Lambda_y \mid \Lambda\big(w'_{X(1)}\big), \Lambda_z^i\big) \tag{A.2}$$

But any path-segment in $\mathcal{C}$ starting at $w'_{X(1)}$ remains in $\hat{\mathcal{C}}^{\Lambda_x}$, and we know that $\{w_z^{ij}\}_{j \geqslant 2}$ do **not** exist in $\hat{\mathcal{C}}^{\Lambda_x}$, so there are no path-segments joining $w'_{X(1)}$ to w_z^{ij} (for $j \geqslant 2$) in $\hat{\mathcal{C}}^{\Lambda_x}$, and hence no path-segments joining $w'_{X(1)}$ to w_z^{ij} (for $j \geqslant 2$) in $\mathcal{C}$. Therefore

$$\Lambda\big(w'_{X(1)}\big) \cap \Lambda\big(w_z^{ij}\big) = \phi \quad \text{for } j \geqslant 2$$

and

$$\Lambda\big(w'_{X(1)}\big) \cap \Lambda_z^i = \Lambda\big(w'_{X(1)}\big) \cap \Lambda\big(w_z^{i1}\big)$$

so expression (A.2) becomes

$$\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(1)}, w'_{X(1)}\big)\big), \Lambda_z^i\big) = \pi\big(\Lambda_y \mid \Lambda\big(w'_{X(1)}\big), \Lambda\big(w_z^{i1}\big)\big)$$
$$= \pi\big(\Lambda_y \mid \Lambda\big(w_z^{i1}\big)\big)$$

using Lemma 1. So Λ_y is conditionally independent of which position $w_X \in W_X$ is considered, and

$$\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(1)}, w'_{X(1)}\big)\big), \Lambda_z^i\big) = \pi\big(\Lambda_y \mid \Lambda\big(e\big(w_{X(2)}, w'_{X(2)}\big)\big), \Lambda_z^i\big)$$

where $w_{X(1)}, w_{X(2)}$ are any two positions in W_X. Hence condition (3.2) is satisfied.

Lemma 2 and W_X-based Back Door probability expression. *For a CEG $\mathcal{C}$, $w_X \in W_X \subset V(\mathcal{C})$, $w_X \in u_X \in L(\mathcal{C})$, and $\Lambda_x = \bigcup_{w_X \in W_X} \Lambda(e(w_X, w'_X))$; if each edge $e(w_X, w'_X)$ for $w_X \in u_X$ carries the same label then*

$$\pi\big(\Lambda_x \mid \Lambda(w_X)\big) = \pi\big(\Lambda_x \mid \Lambda(u_X)\big)$$

This tells us that the probability of leaving a stage by an edge carrying a particular label is the same as that of leaving any of its component positions by an edge carrying this label.

The equality holds if the edges $e(w_X, w'_X)$ label the **same** value of X for each $w_X \in u_X$. This is the case for all basic *Do* interventions and all functional manipulations.

Proof.

$$\pi\big(\Lambda_x \mid \Lambda(u_X)\big) = \pi\left(\Lambda_x \mid \bigcup_{w_X \in u_X} \Lambda(w_X)\right)$$
$$= \frac{\pi(\Lambda_x, \bigcup_{w_X \in u_X} \Lambda(w_X))}{\pi(\bigcup_{w_X \in u_X} \Lambda(w_X))}$$

And the events $\{\Lambda(w_X)\}_{w_X \in u_X}$ are disjoint (Definitions 1 and 3) so this equals

$$= \frac{\sum_{w_X \in u_X} \pi(\Lambda_x, \Lambda(w_X))}{\sum_{w_X \in u_X} \pi(\Lambda(w_X))}$$

Consider a specific $w_X \in u_X$. Call this $w_{X(1)}$. Then

$$\Lambda_x \cap \Lambda(w_{X(1)}) = \left[\bigcup_{w_X \in u_X} \Lambda\big(e\big(w_X, w'_X\big)\big)\right] \cap \Lambda(w_{X(1)})$$
$$= \Lambda\big(e\big(w_{X(1)}, w'_{X(1)}\big)\big) \cap \Lambda(w_{X(1)}) \tag{A.3}$$

since $\Lambda(e(w_{X(2)}, w'_{X(2)})) \cap \Lambda(w_{X(1)}) = \emptyset$ for $w_{X(2)} \neq w_{X(1)}$ $(w_{X(2)} \in W_X)$.

So $\pi(\Lambda_x, \Lambda(w_X))$ can be written as

$$\pi\big(\Lambda(w_X), \Lambda\big(e\big(w_X, w'_X\big)\big)\big) = \pi\big(\Lambda\big(e\big(w_X, w'_X\big)\big) \mid \Lambda(w_X)\big)\pi\big(\Lambda(w_X)\big).$$

But $\pi(\Lambda(e(w_X, w'_X)) \mid \Lambda(w_X))$ is constant for all $w_X \in u_X$, since u_X is a stage, and the edges $\{e(w_X, w'_X)\}_{w_X \in u_X}$ carry the same label. So we can take this probability outside the summation to give

$$\pi\big(\Lambda_x \mid \Lambda(u_X)\big) = \frac{\pi(\Lambda(e(w_X, w'_X)) \mid \Lambda(w_X)) \sum_{w_X \in u_X} \pi(\Lambda(w_X))}{\sum_{w_X \in u_X} \pi(\Lambda(w_X))}$$
$$= \pi\big(\Lambda\big(e\big(w_X, w'_X\big)\big) \mid \Lambda(w_X)\big)$$
$$= \frac{\pi(\Lambda(e(w_X, w'_X)), \Lambda(w_X))}{\pi(\Lambda(w_X))} = \frac{\pi(\Lambda_x, \Lambda(w_X))}{\pi(\Lambda(w_X))}$$

from $(A.3)$ above

$$= \pi\big(\Lambda_x \mid \Lambda(w_X)\big) \qquad \square$$

***W*$_X$-based Back Door probability expression.** Using the proof of Theorem 1 above we can write

$$\pi(\Lambda_y \parallel \Lambda_x) = \hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\pi\big(\Lambda_y \mid \Lambda\big(e\big(w_X, w'_X\big)\big)\big)$$
$$= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big)\pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_x\big)$$
$$= \sum_{u_X} \sum_{w_X \in u_X} \pi\big(\Lambda(w_X)\big)\pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_x\big)$$
$$= \sum_{u_X} \sum_{w_X \in u_X} \left[\frac{\pi(\Lambda(w_X), \Lambda_x, \Lambda_y)}{\pi(\Lambda_x \mid \Lambda(w_X))}\right]$$
$$= \sum_{u_X} \left[\frac{\sum_{w_X \in u_X} \pi(\Lambda(w_X), \Lambda_x, \Lambda_y)}{\pi(\Lambda_x \mid \Lambda(u_X))}\right] = \sum_{u_X} \left[\frac{\pi(\Lambda(u_X), \Lambda_x, \Lambda_y)}{\pi(\Lambda_x \mid \Lambda(u_X))}\right]$$
$$= \sum_{u_X} \pi\big(\Lambda_y \mid \Lambda(u_X), \Lambda_x\big)\pi\big(\Lambda(u_X)\big) \qquad \square$$

Proof of Theorem 2. This follows the proof of Theorem 1 until line (A.1). We then invoke conditions (4.1) and (4.2) to give

$$\begin{aligned}\hat{\pi}^{\Lambda_x}(\Lambda_y) &= \sum_{w_X \in W_X} \pi\big(\Lambda(w_X)\big) \sum_z \pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_z\big)\pi(\Lambda_z \mid \Lambda_x) \\ &= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_z\big)\pi\big(\Lambda(w_X)\big) \\ &= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \sum_i \pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_z\big)\pi\big(\Lambda(w_X), \Lambda_x^i\big) \qquad \text{(A.4)}\end{aligned}$$

since $\{\Lambda_x^i\}$ forms a partition of Ω

$$= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \sum_i \pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_x^i, \Lambda_z\big)\pi\big(\Lambda(w_X), \Lambda_x^i\big)$$

using condition (4.1). But

$$\pi\big(\Lambda(w_X), \Lambda_x^i\big) = \frac{\pi(\Lambda(w_X), \Lambda_x^i, \Lambda_z)}{\pi(\Lambda_z \mid \Lambda(w_X), \Lambda_x^i)} = \frac{\pi(\Lambda(w_X), \Lambda_x^i, \Lambda_z)}{\pi(\Lambda_z \mid \Lambda_x^i)}$$

using condition (4.2)

$$= \pi\big(\Lambda(w_X) \mid \Lambda_x^i, \Lambda_z\big)\pi\big(\Lambda_x^i\big)$$

So

$$\begin{aligned}\hat{\pi}^{\Lambda_x}(\Lambda_y) &= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \sum_i \pi\big(\Lambda_y \mid \Lambda(w_X), \Lambda_x^i, \Lambda_z\big)\pi\big(\Lambda(w_X) \mid \Lambda_x^i, \Lambda_z\big)\pi\big(\Lambda_x^i\big) \\ &= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_i \pi\big(\Lambda_y \mid \Lambda_x^i, \Lambda_z\big)\pi\big(\Lambda_x^i\big) \qquad \square\end{aligned}$$

References

[1] T. Bedford, R. Cooke, Probabilistic risk analysis: Foundations and methods, in: Probabilistic Risk Analysis: Foundations and Methods, Cambridge, 2001, pp. 99–151.
[2] C. Boutilier, N. Friedman, M. Goldszmidt, D. Koller, Context-specific independence in Bayesian Networks, in: Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, 1996, pp. 115–123.
[3] G.A. Churchill, Accurate restoration of DNA sequences, in: C. Gatsaris, et al. (Eds.), Case Studies in Bayesian Statistics, vol. 2, Springer-Verlag, 1995, pp. 90–148.
[4] A.P. Dawid, Causal inference without counterfactuals, Journal of the American Statistical Association 95 (2000) 407–448.
[5] A.P. Dawid, Influence diagrams for causal modelling and inference, International Statistical Review 70 (2002) 161–189.
[6] A.P. Dawid, V. Didelez, Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview, Statistics Surveys 4 (2010) 184–231.
[7] S. French (Ed.), Readings in Decision Analysis, Chapman and Hall/CRC, 1989.
[8] D. Glymour, G.F. Cooper, Computation, Causation and Discovery, MIT Press, 1999.
[9] D. Heckerman, A Bayesian approach to Learning Causal Networks, in: W. Edwards, et al. (Eds.), Advances in Decision Analysis, CUP, 2007, pp. 202–220.
[10] S.L. Lauritzen, Graphical Models, Oxford, 1996.
[11] S.L. Lauritzen, Causal inference from graphical models, in: O.E. Barndorff-Nielsen, et al. (Eds.), Complex Stochastic Systems, Chapman and Hall, 2001.
[12] R. Lyons, Random walks and percolation on trees, Annals of Probability 18 (1990) 931–958.
[13] D. McAllester, M. Collins, F. Periera, Case factor diagrams for structured probabilistic modeling, in: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, 2004, pp. 382–391.
[14] J. Pearl, Causal diagrams for empirical research, Biometrika 82 (1995) 669–710.
[15] J. Pearl, Causality: Models, Reasoning and Inference, Cambridge, 2000.
[16] J. Pearl, Statistics and causal inference: A review, Sociedad de Estadistica e Investigacion Operativa. Test 12 (2003) 281–345.
[17] J. Pearl, J.M. Robins, Probabilistic evaluation of sequential plans from causal models with hidden variables, in: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, 1995, pp. 444–445.
[18] D. Poole, N.L. Zhang, Exploiting contextual independence in probabilistic inference, Journal of Artificial Intelligence Research 18 (2003) 263–313.
[19] J.M. Robins, A new approach to causal inference in mortality studies with sustained exposure period — application to control of the healthy worker survivor effect, Mathematical Modelling 7 (1986) 1393–1512.
[20] A. Salmeron, A. Cano, S. Moral, Importance sampling in Bayesian Networks using probability trees, Computational Statistics and Data Analysis 34 (2000) 387–413.
[21] G. Shafer, The Art of Causal Conjecture, MIT Press, 1996.
[22] J.Q. Smith, P.E. Anderson, Conditional independence and Chain Event Graphs, Artificial Intelligence 172 (2008) 42–68.
[23] J.Q. Smith, P.A. Thwaites, Influence diagrams, in: E.L. Melnick, B.S. Everitt (Eds.), Encyclopedia of Quantitative Risk Analysis and Assessment, vol. 2, Wiley, 2008, pp. 897–910.
[24] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction and Search, Springer-Verlag, 1993.

[25] P.A. Thwaites, J.Q. Smith, R.G. Cowell, Propagation using Chain Event Graphs, in: Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, Helsinki, 2008, pp. 546–553.
[26] P.A. Thwaites, J.Q. Smith, E.M. Riccomagno, Causal analysis with Chain Event Graphs, Artificial Intelligence 174 (2010) 889–909.
[27] J. Tian, Identifying dynamic sequential plans, in: Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, Helsinki, 2008, pp. 554–561.
[28] J. Tian, J. Pearl, A general identification condition for causal effects, in: Proceedings of the 18th National Conference on Artificial Intelligence, AAAI Press, 2002, pp. 567–573.