SEPARATION THEOREMS FOR CHAIN EVENT GRAPHS

BY PETER A. THWAITES* AND JIM Q. SMITH*

University of Warwick

A separation theorem on a graphical model allows an analyst to identify the conditional independence statements it logically entails using only the topology of the graph. In this paper we prove separation theorems associated with a new coloured graphical model called a Chain Event Graph (CEG). The class of CEG models generalises the class of finite discrete Bayesian Network models. Here we formally define this model class, and consider the set of permissible conditional independence queries on this graph. We provide necessary and sufficient conditions for these conditional independence statements to hold on a subclass of uncoloured CEGs called simple CEGs. We then prove sufficient conditions for such statements to hold on a much larger subclass called regular CEGs. The paper is illustrated with a running example demonstrating the application of these theorems.

1. Introduction. If the DAG (directed acyclic graph) $\mathcal{G}$ of a Bayesian Network (BN) has a vertex set $\{X_1, X_2, \dots, X_n\}$, then there are n conditional independence assertions which can simply be read off the graph. These are the properties that state that a vertex-variable is independent of its non-descendants given its parents (the directed local Markov property [14]). Answering most conditional independence queries however, is not so straightforward. The d-separation theorem for BNs was first proved by Verma and Pearl [31], and an alternative version considered in [15, 14, 5]. The theorem addresses whether the conditional independence query $A \amalg B \mid C$? can be answered from the topology of the DAG of a BN, where A, B, C are disjoint subsets of the set of vertex-variables of the DAG. It allows the BN to be interrogated and irrelevances checked before any quantitative embellishments of distribution on its conditional probability tables are added. This provides a valuable tool in the process of discovering requisite models [21], as well as a logical framework for propagation algorithms and learning (see for example [5] and the TETRAD software of Scheines et al).

However for many problems the available quantitative dependence information cannot all be embodied in the DAG of a BN. Separation theorems

*This research is supported by EPSRC grant no. EP/F036752/1

AMS 2000 subject classifications: Primary 68T30, 68T37; secondary 62F30, 68R10

Keywords and phrases: Bayesian Network, Chain Event Graph, conditional independence, directed acyclic graph, graphical model, separation theorem

have been proved for more general classes of graphical model including chain graphs [3], alternative chain graphs [2], and ancestral graphs [23]. In this paper we prove separation theorems for a particularly expressive graphical model – the Chain Event Graph (CEG).

Our motivation for the development of this class is that CEGs are probably the most natural graphical models for discrete processes when elicitation involves questions about how situations might unfold. Although the topology of these graphs is more complicated than that of the BN, they are much more expressive, as they allow us to represent all structural quantitative information within the graph itself. Context-specific symmetries which are not intrinsic to the structure of the BN [4, 16, 22, 24] are fully expressed in the topology of the CEG, which also recognises logical zeros in probability tables, and the numbers of levels taken by problem-variables. This last has been found to be essential to understanding the geometry of BN models with hidden variables [1, 18].

The CEG has already been demonstrated to be a useful inferential framework for applications as diverse as forensic science [26], biological regulatory models [27], and education [8]. The graphs provide a framework for representation [27], probability propagation [29], learning and model selection [8], and for causal analysis [30].

These papers concentrate on the application of CEG-based techniques. Whilst they use the conditional independence properties of the graph, they do not provide a full formal development for the class of CEG models. This paper rectifies this lack. In doing so we identify the form of the types of conditional independence statements it is natural to query, and also prove a number of separation theorems which allow us to answer each query as always true or not, solely on the basis of the topology of the graph.

We note that, even more so than is the case with BNs, there are a number of conditional independence properties which can simply be *read* off the CEG. These are described in Sections 2.4 and 5.2, and given the tree-based nature of the CEG these properties are naturally context-specific. That is to say they are properties of the form $A \amalg B \mid \Lambda$ for some event Λ. An analogous statement for a discrete BN would be of the form

$$p(A = \boldsymbol{a} \mid B = \boldsymbol{b}, C = \boldsymbol{c}) = p(A = \boldsymbol{a} \mid C = \boldsymbol{c})$$

for some subsets of variables A, B, C, some **specific** vector value $\boldsymbol{c}$ of C and all vector values $\boldsymbol{a}$ of A and $\boldsymbol{b}$ of B. The class of conditioning events we can tackle with a CEG is however much richer than that generally considered when using BN-based analysis.

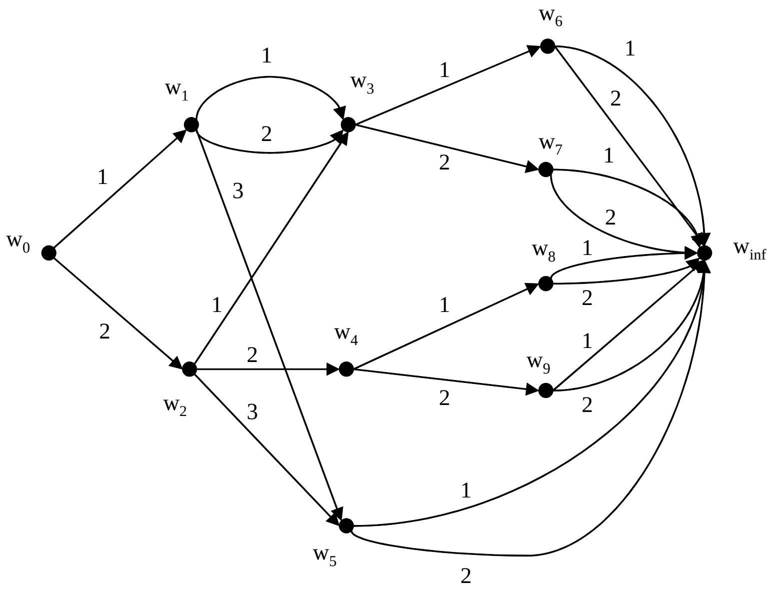

FIG 1. *An SCEG $\mathcal{C}$*

2. The Simple Chain Event Graph.

2.1. *The basic definition of an SCEG.* The *Chain Event Graph* $\mathcal{C}(V, E)$ is a directed acyclic graph (DAG), which is connected with a unique *root vertex* (with no incoming edges) and a unique *sink vertex* (with no outgoing edges). Unlike the BN more than one edge can exist between two vertices of a CEG. The *regular Chain Event Graph* (RCEG) discussed in section 4 also has its vertices and edges coloured.

We first consider a subclass of the class of CEGs called a *simple Chain Event Graph* (SCEG). Neither the vertices (called *positions*) $w \in V(\mathcal{C})$, nor the edges $e(w, w') \in E(\mathcal{C})$ of an SCEG are coloured. An example of an SCEG is given in Figure 1.

The root and sink vertices of a CEG are labelled w_0 and w_∞. Each position $w \in V(\mathcal{C}) \backslash \{w_\infty\}$ has a set $E(w)$ of $k(w)$ outgoing edges, which when we wish to emphasise their connection with the position w, may be labelled $\{e_x(w) : x = 1, 2, \ldots, k(w)\}$.

A directed $w_0 \to w_\infty$ path λ in $\mathcal{C}$ is called a *route*. The set of routes of $\mathcal{C}$ is labelled $\Lambda(\mathcal{C})$ (and corresponds to the set of atoms of the finite discrete

TABLE 1
Context for Figure 1

Descriptor	Edges
male	$e_1(w_0)$
female	$e_2(w_0)$
displayed symptom S before puberty	$e_1(w_1), e_1(w_2)$
displayed symptom S after puberty	$e_2(w_1), e_2(w_2)$
never displayed symptom S	$e_3(w_1), e_3(w_2)$
developed condition	$e_1(w_3), e_1(w_4)$
did not develop condition	$e_2(w_3), e_2(w_4)$
died before the age of 50	$e_1(w_5), e_1(w_6), e_1(w_7),$ $e_1(w_8), e_1(w_9)$
died at the age of 50 or older	$e_2(w_5), e_2(w_6), e_2(w_7),$ $e_2(w_8), e_2(w_9)$

probability space represented by $\mathcal{C}$ – see below). Note that each route is uniquely determined by a sequence of edges. Thus in the CEG in Figure 1, one such route is $\lambda_1 \equiv \{e_1(w_0), e_1(w_1), e_1(w_3), e_1(w_6)\}$. It is easy to check that $\mathcal{C}$ here has 20 such routes. We write $w \prec w'$ when the position w precedes the position w' on a route.

When our CEG is applied to a population, each route corresponds to a possible set of attributes that a member of the population could take. For example, if the CEG in Figure 1 is applied to a population of people whose parents sufferered from an inherited medical condition, and the edges of the CEG carry the descriptors given in Table 1, then the route λ_1 described above corresponds to *male, displayed symptom S before puberty, developed condition, died before the age of 50.*

An SCEG is *route compatible* for a population of units Ψ if each possible history of a unit in the population (or atom of the event space) corresponds to the unit passing along one of the routes $\lambda \in \Lambda(\mathcal{C})$. We use $\mathbb{F}(\mathcal{C})$ to denote the sigma field of events formed by these atoms. $\mathbb{F}(\mathcal{C})$ corresponds to the power set of $\Lambda(\mathcal{C})$. Since each atom of this event space codes what might happen to a unit in Ψ, the SCEG encodes an additional longitudinal development depicting the possible ways the future might unfold, not encoded by the sigma field $\mathbb{F}(\mathcal{C})$ alone (see [25]).

We label an event in $\mathbb{F}(\mathcal{C})$ by Λ, and note that because the CEG's atoms have this implicit longitudinal development associated with them, certain events in $\mathbb{F}(\mathcal{C})$ are particularly important. Let $\Lambda(w)$ denote the event that a unit takes a route that passes through the position $w \in V(\mathcal{C})$. $\Lambda(w, w')$ is then the union of all routes passing through the positions w and w', $\Lambda(e(w, w'))$ is the union of all routes passing through the edge $e(w, w')$, and $\Lambda(\mu(w, w'))$ is the union of all routes utilising the subpath $\mu(w, w')$.

Certain subsets of the set of positions also have an important status in this context. In this paper we will call a set $R \subset V(\mathcal{C})$ a *regular subset* if the events $\{\Lambda(w) : w \in R\}$ are disjoint. Note that R is regular if and only if there is no route $\lambda \in \Lambda(\mathcal{C})$ containing more than one position $w \in R$. Call R a *position-cut* if $\{\Lambda(w) : w \in R\}$ forms a partition of $\Lambda(\mathcal{C})$. A *position-cut* can be associated with a random variable that labels which of a class of developments a unit might take (see section 3).

2.2. *Probabilities on an SCEG.* Underlying the SCEG there is a probability space which is specified by assigning probabilities to the atoms. We do this as follows: For each position $w \in V(\mathcal{C})\backslash\{w_\infty\}$ and edge $e(w, w')$ emanating from w, we call $\pi_e(w' \mid w)$ a *primitive probability* if $\pi_e(w' \mid w) \geq 0$ and $\sum_{w'} \pi_e(w' \mid w) = 1$.

Definition 1. A probability mass function $p(\lambda)$, $\lambda \in \Lambda(\mathcal{C})$ is said to have the *monomial property* for a population Ψ if there exists a set of primitive probabilities $\Pi = \{\pi_e(w' \mid w) : e(w, w') \in E(w), w \in V(\mathcal{C})\backslash\{w_\infty\}\}$ on the edges of $\mathcal{C}$ such that for all routes $\lambda \in \Lambda(\mathcal{C})$

$$p(\lambda) = \prod_{e(w,w') \in \lambda} \pi_e(w' \mid w) \tag{2.1}$$

where $e(w, w') \in \lambda$ means that the edge $e(w, w')$ lies on the route λ.

Note that (2.1) fully defines a probability measure over $\mathbb{F}(\mathcal{C})$ by specifying each atomic probability as a function of its primitive probabilities.

The assignment of probabilities (2.1), determined by Π implicitly demands a Markov property over the flow of the units through the graph. Thus, in the context of our medical example, the probablility of an individual with attributes (*male, displayed symptom S before puberty*), (*male, displayed symptom S after puberty*) or (*female, displayed symptom S before puberty*) developing the condition depends only on the fact that the subpaths corresponding to these pairs of attributes terminate at the position w_3, and not on the particular subpath leading to w_3. The probability this individual develops the condition is then $\pi_e(w_6 \mid w_3) \equiv p(\Lambda(e(w_3, w_6)) \mid \Lambda(w_3))$. So we only need to know the position a unit has reached in order to predict as well as is possible what the next unfolding of its development will be.

This Markov hypothesis looks strong but in fact holds for many families of statistical model. For example all event tree descriptions of a problem satisfy this property, all finite state space context specific Bayesian Networks as well as many other structures [27].

We can go further and state that the sets of possible future developments (whether or not they developed the condition **and** whether or not they died before the age of 50) for individuals taking any of these three subpaths must be the same. Moreover the conditional probability of any particular subsequent development must be the same for individuals taking any of these three subpaths. This accounts for the term *position* for a (non-sink) vertex.

In this paper we discuss *minimal* CEGs where if positions w_α and w_β are such that the sets of possible future developments from w_α and w_β are identical, and the conditional probability distributions over these sets are identical, then w_α and w_β are the **same** position. Any reference to a CEG, SCEG or RCEG should therefore be taken to mean a minimal CEG, SCEG or RCEG.

DEFINITION 2. An SCEG $\mathcal{C}$ is said to be *valid* for a population Ψ if it is route compatible and has the monomial property for Ψ.

Note that like the BN, the SCEG can be valid without its associated primitive probabilities being known. We just need to believe that some set Π exists so that the associated Markov hypothesis holds. We are free to assign *any* set of probabilities Π to the edges of a valid SCEG within the simplex conditions above. So in particular the probability model space of a valid $\mathcal{C}$ can be defined as the product space of these $|V(\mathcal{C})|-1$ different simplices where the simplex associated with $w \in V(\mathcal{C})\backslash\{w_\infty\}$ has Euclidean dimension $k(w)-1$. The probability of any event Λ in $\mathbb{F}(\mathcal{C})$ is then of the form

$$p(\Lambda) = \sum_{\lambda\in\Lambda} p(\lambda) = \sum_{\lambda\in\Lambda} \prod_{e(w,w')\in\lambda} \pi_e(w' \mid w)$$

where $\lambda \in \Lambda$ means that λ is one of the component atoms of the event Λ. In this paper we will also use the following further notation:
$\pi_\mu(w' \mid w) \equiv p(\Lambda(\mu(w,w')) \mid \Lambda(w))$ denotes the probability of utilising the subpath $\mu(w,w')$ (conditional on passing through w),
$\pi(w' \mid w) \equiv p(\Lambda(w,w') \mid \Lambda(w)) = \sum_\mu \pi_\mu(w' \mid w)$ denotes the probability of arriving at w' conditional on passing through w.

2.3. *Conditioning on intrinsic events.* In this paper we are interested in conditioning sets which give rise to conditional independence queries that can be answered purely by inspecting the topology of an SCEG $\mathcal{C}$. An important subclass of these are events in $\mathbb{F}(\mathcal{C})$ which are called intrinsic.

DEFINITION 3. An *intrinsic event* Λ in $\mathbb{F}(\mathcal{C})$ is a set of routes of $\mathcal{C}$ which are also routes of $\mathcal{C}_\Lambda$ where $\mathcal{C}_\Lambda$ is a subgraph of $\mathcal{C}$ that contains the root vertex

w_0 and the sink vertex w_∞ of $\mathcal{C}$ in its vertex set, and where w_0 is the only vertex in $V(\mathcal{C}_\Lambda)$ with no parent, and w_∞ is the only vertex in $V(\mathcal{C}_\Lambda)$ with no child. Call such a subgraph $\mathcal{C}_\Lambda$ a *sub SCEG*.

Note that the sub SCEG $\mathcal{C}_\Lambda$ is itself an SCEG. All atoms of $\mathbb{F}(\mathcal{C})$ are intrinsic, as are $\Lambda(w)$ and $\Lambda(w, w')$ (provided this is non-empty) for all $w, w' \in V(\mathcal{C})$, and as is the exhaustive set $\Lambda(w_0)$. If we include the empty set in the set of intrinsic events then we note that intrinsic sets are closed under intersection and so technically form a π-system (see for example [12]) we can associate with the SCEG $\mathcal{C}$.

Not all events in $\mathbb{F}(\mathcal{C})$ are necessarily intrinsic because the class of intrinsic events is not closed under union. For example, for the CEG in Figure 1, the event Λ consisting of the union of the two atoms $(e_1(w_0), e_1(w_1), e_1(w_3), e_1(w_6))$ and $(e_1(w_0), e_2(w_1), e_1(w_3), e_2(w_6))$ produces a subgraph $\mathcal{C}_\Lambda$ which has four distinct routes, so Λ is not intrinsic. However the class of intrinsic events is rich enough to encompass virtually all of the conditioning events in the conditional independence statements we would like to query. In particular, if our model can be expressed as a BN then any set of observations expressible in the form $O(\boldsymbol{A}) = \{X_j \in A_j\}$ (for subsets $\{A_j\}$ of the sample spaces of $\{X_j\}$, the vertex-variables of the BN) is a proper subset of the set of intrinsic events defined on the CEG of our model [29].

The first important property of the class of valid SCEG models is that they are closed under conditioning by an intrinsic event:

THEOREM 1. *If an SCEG $\mathcal{C}$ is valid on a population Ψ then the probability model on $\mathbb{F}(\mathcal{C}|\Lambda)$ of any of its sub SCEGs $\mathcal{C}_\Lambda$ is a probability model on $\mathbb{F}(\mathcal{C}_\Lambda)$ which is also valid.*

The obvious set of primitive probabilities for the sub-SCEG $\mathcal{C}_\Lambda$ is given by

$$\Pi^* = \{\pi_e^*(w' \mid w) : e(w, w') \in E(w), w \in V(\mathcal{C}) \backslash \{w_\infty\}\}$$

where

$$\pi_e^*(w' \mid w) = \frac{p(\Lambda \mid \Lambda(e(w, w')))}{p(\Lambda \mid \Lambda(w))} \, \pi_e(w' \mid w)$$

providing this is well-defined. A proof of this theorem can be found in the appendix. We note that this property has now been successfuly used to develop fast propagation algorithms for CEGs (see [29]).

Note that the probability of an atom λ in $\mathcal{C}$ conditioned on the intrinsic event Λ is the probability of that atom in the SCEG $\mathcal{C}_\Lambda$. We denote this probability $p_\Lambda(\lambda)$. It is then trivially the case that the probability of an event in $\mathcal{C}$ conditioned on the event Λ is the probability of that event in the SCEG $\mathcal{C}_\Lambda$.

2.4. *Random variables on an SCEG..* Random variables measurable with respect to $\mathbb{F}(\mathcal{C})$ partition the set of atoms into events. So for example, we can define variables X, Y, measurable with respect to $\mathbb{F}(\mathcal{C})$, which partition the set of atoms into events $\{\Lambda_X\}, \{\Lambda_Y\}$. Moreover for an event Λ (with $p(\Lambda) \neq 0$) we can write $X \amalg Y \mid \Lambda$ if $p(X = x \mid Y = y, \Lambda) = p(X = x \mid \Lambda)$ for all values x of X and y of Y (see for example [7]).

LEMMA 1. *For a CEG $\mathcal{C}$, variables X, Y measurable with respect to $\mathbb{F}(\mathcal{C})$, and intrinsic conditioning event Λ, the statement $X \amalg Y \mid \Lambda$ is true if and only if $X \amalg Y$ is true in the CEG $\mathcal{C}_\Lambda$.*

The proof of this lemma is in the appendix. This is a particularly useful property because it allows us to check any context-specific conditional independence property by checking a non-conditional independence property on a sub-SCEG.

We now turn our attention to two types of elementary random variables, measurable with respect to $\mathbb{F}(\mathcal{C})$, that can be identified with each position $w \in V(\mathcal{C})\backslash\{w_\infty\}$. These are the variables $\{I(w) : w \in V(\mathcal{C})\backslash\{w_\infty\}\}$ defined by

$$I(w) = \begin{cases} 1 & \text{if } \lambda \text{ passes through } w \\ 0 & \text{otherwise} \end{cases}$$

and the variables $\{X(w) : w \in V(\mathcal{C})\backslash\{w_\infty\}\}$ defined by

$$X(w) = \begin{cases} x & \text{if } \lambda \text{ passes along edge } e_x(w) \in E(w) \\ 0 & \text{if the position } w \text{ does not lie on } \lambda \end{cases}$$

where $x = 1, 2, \dots k(w)$ index the edges emanating from w. Notice that since $I(w)$ is clearly a function of $X(w)$, to specify a full joint distribution over $\{(I(w), X(w)) : w \in V(\mathcal{C})\backslash\{w_\infty\}\}$ it is sufficient to specify the joint distribution of $\{X(w) : w \in V(\mathcal{C})\backslash\{w_\infty\}\}$. Note that all atomic events λ can be expressed as the intersection of events

$$\lambda = \bigcap_{w \in \lambda} \{X(w) = x_\lambda\}$$

and events in $\mathbb{F}(\mathcal{C})$ as the union of these atomic events

$$\Lambda = \bigcup_{\lambda \in \Lambda} \left\{ \bigcap_{w \in \lambda} \{X(w) = x_\lambda\} \right\}$$

where $w \in \lambda$ denotes that the position w lies on the route λ, and $x_\lambda \neq 0$ is the unique value of $X(w)$ of the edge in the route λ.

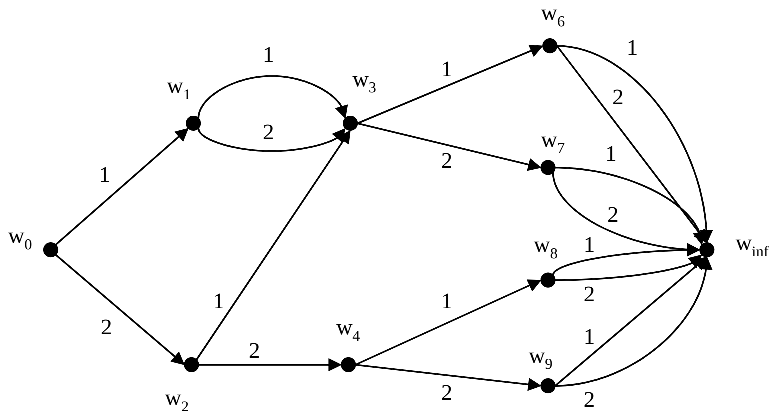

FIG 2. $\mathcal{C}$ *conditioned on the event that* displayed symptom S

Figure 2 shows the SCEG $\mathcal{C}$ from Figure 1 conditioned on the intrinsic event $\Lambda = (X(w_1) = 1) \cup (X(w_1) = 2) \cup (X(w_2) = 1) \cup (X(w_2) = 2)$ or in the context of our medical example, *displayed symptom S*.

For any set $A \subset V(\mathcal{C})$, let $\boldsymbol{X}_A$ denote the set of random variables $\{X(w) : w \in A\}$ and $\boldsymbol{I}_A$ the set $\{I(w) : w \in A\}$. Also, for any $w \in V(\mathcal{C})$, let $U(w)$ be the set of positions in $V(\mathcal{C})$ which lie upstream of the position w, $D(w)$ the set of positions which lie downstream of w, $U^c(w)$ the set of positions which do not lie upstream of w, and $D^c(w)$ the set of positions which do not lie downstream of w.

LEMMA 2. *For any SCEG $\mathcal{C}$ and position $w \in V(\mathcal{C}) \backslash \{w_\infty\}$, the variables $I(w), X(w)$ exhibit the* position independence property *that*

$$X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid I(w)$$

The result given in this lemma is analogous to that which Pearl [20] uses to define BNs, which states that a BN vertex-variable is independent of its non-descendants given its parents. It provides a set of conditional independence statements that can simply be read from the graph, one for each position in $V(\mathcal{C})$. The proof of the lemma is in the appendix.

The statement that $X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid (I(w) = 1)$ can be read as: *Given a unit reaches a position $w \in V(\mathcal{C})$, whatever happens immediately after w is independent of not only all developments through which that position was* *reached, but also of all positions that logically have not happened or could not now happen because the unit has passed through w.* Thus, in the sense above, the position of a valid SCEG $\mathcal{C}$ is *sufficient* to describe the future development of units passing through it.

As already noted, the product space defined by $\{(I(w), X(w)) : w \in V(\mathcal{C})\}$ is over specified. This is so firstly because $I(w) = 0 \Leftrightarrow X(w) = 0$ and $I(w) = 1 \Rightarrow X(w') = 0$ for $w' \in D^c(w) \cap U^c(w)$. Probability distributions exist which satisfy the set of statements of the form $X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid I(w)$ which do not obey these implications, but such distributions cannot be represented on an SCEG.

More significantly, if the SCEG is used for the purpose for which it was intended, as a representation of an asymmetric process or problem, then there will be many probabilities in the joint probability tables over the space defined by $\{(I(w), X(w)) : w \in V(\mathcal{C})\}$ which are identically zero. The joint mass function is then extremely sparse. These zeros correspond to *impossible* events which nonetheless are given equal significance with *possible* events in a BN-representation of the problem. In many cases these events are not just impossible but meaningless. For example if $X(w_a) = 1$ corresponds to *patient dies*, $X(w_b) = 1$ corresponds to *patient is given treatment 2*, and $w_a \prec w_b$, then the event $(X(w_a) = 1, X(w_b) = 1)$ has no logical meaning.

As the set of statements of the form $\{(I(w), X(w)) : w \in V(\mathcal{C})\}$ do not define the SCEG, these additional *counterfactual* statements produced by the product space representation are not an integral part of the CEG-framework. The product space defined by the full set of statements is nevertheless a useful construct because it allows us to encode sets of conditional independence statements into a valid SCEG and so allows us to quickly prove separation theorems for such graphs.

The structure of the CEG illustrates a further aspect of the graphical modelling process which is not transparent in the topology of the BN. The CEG depicts all possible histories of a unit in a population, and gives a probability distribution over these histories. However, when a single unit traverses one of the routes in the CEG, values are assigned to $I(w), X(w)$ for all positions $w \in V(\mathcal{C})$. Those conditional independence statements encoded by the positions and edges through which our unit has **not** passed are now truly counterfactual [6] in that they answer queries of the form *If X had not been the case, what would be the chance of Y happening?* So the CEG simultaneously depicts both the "reality" and the counterfactual aspects of the problem once we start to observe the actual behaviour of units in the population. It also makes it a powerful framework for expressing rich varieties of causal hypotheses [30].

3. A separation theorem for Simple CEGs. We call a position $w \in V(\mathcal{C})$ a *stalk* if the removal of w from $V(\mathcal{C})$ would result in a graph with two disconnected components. In (non-probabilistic) graph theory such a vertex is called a *cut vertex* (see for example [11]).

THEOREM 2. *In an SCEG $\mathcal{C}$ with $w_1, w_2 \in V(C)$ and $w_2 \not\prec w_1$, $X(w_1) \amalg X(w_2)$ if and only if either w_2 is a stalk, or there exists a stalk downstream of w_1 and upstream of w_2, for $w_0 \preceq w_1 \prec w_\infty$, $w_0 \prec w_2 \prec w_\infty$.*

The proof of this theorem is in the appendix. Theorem 2 has a number of powerful corollaries, which we give after introducing two new variables. Call $J(R)$ the *incidence variable* of a regular subset R if

$$J(R) \equiv \sum_{w \in R} I(w) \equiv \sup_{w \in R} I(w)$$

and call $Y(R)$ the *criterion variable* of a regular subset R if

$$Y(R) \equiv \sum_{w \in R} X(w) \equiv \sup_{w \in R} X(w)$$

LEMMA 3. *For an SCEG $\mathcal{C}$ with position cuts $R_a = \{w_a\}$, $R_b = \{w_b\}$: If $X(w_a) \amalg X(w_b)$ for* **any** *$w_a \in R_a$, $w_b \in R_b$, then $Y(R_a) \amalg Y(R_b)$ in every distribution compatible with $\mathcal{C}$.*
Conversely, if $Y(R_a) \amalg Y(R_b)$ holds for all distributions compatible with $\mathcal{C}$, then $X(w_a) \amalg X(w_b)$ for all $w_a \in R_a$, $w_b \in R_b$.

This lemma and Corollary 5 in Section 7 formalise and generalise the result given in [27] Theorem 2. The proof of the lemma is in the appendix. The converse result is somewhat surprising, but is a consequence of the particular structure of the sigma field associated with an SCEG.

COROLLARY 1. *Let $\mathcal{C}$ be an SCEG, Λ an intrinsic event, $R_a = \{w_a\}$, $R_b = \{w_b\}$ be position cuts of $\mathcal{C}$.*
If in the sub-CEG $\mathcal{C}_\Lambda$, w_a and w_b are separated by a stalk, for **any** *$w_a \in R_a$, $w_b \in R_b$, $w_a, w_b \in V(\mathcal{C}_\Lambda)$, then $Y(R_a) \amalg Y(R_b) \mid \Lambda$.*

The proof of this corollary is in the appendix. This has major consequences for models which admit a product space structure, where othogonal cuts of the CEG have a natural meaning corresponding to measurement variables of the problem. Models of this sort can be represented as BNs, with possible annotation of context-specific conditional independence properties.

Corollary 2. *If an SCEG $\mathcal{C}$ is of a model which admits a product space structure, A, B are measurement variables of the model, and R_a, R_b are the position cuts of $\mathcal{C}$ correponding to these variables, then:*
If $X(w_a) \amalg X(w_b)$ *for* **any** $w_a \in R_a$, $w_b \in R_b$, *then* $A \amalg B$ *in every distribution compatible with* $\mathcal{C}$.
Conversely, if $A \amalg B$ *holds for all distributions compatible with* $\mathcal{C}$, *then* $X(w_a) \amalg X(w_b)$ *for all* $w_a \in R_a$, $w_b \in R_b$.

The proof of this follows immediately from Lemma 3.

Corollary 3. *Let $\mathcal{C}$ be an SCEG of a model which admits a product space structure, A, B be measurement variables of the model, Λ an intrinsic event, R_a, R_b be position cuts of $\mathcal{C}$ correponding to the variables A and B. If in the sub-CEG $\mathcal{C}_\Lambda$, w_a and w_b are separated by a stalk, for* **any** $w_a \in R_a$, $w_b \in R_b$, $w_a, w_b \in V(\mathcal{C}_\Lambda)$, *then* $A \amalg B \mid \Lambda$.

The proof of this follows directly from Corollaries 1 and 2. In the case where our model has a natural product space structure, the topology of the SCEG allows us to replace conditional independence queries such as $A \amalg B \mid C$? by sets of context-specific queries such as $\{A \amalg B \mid (C = \mathbf{c})\ ?\}$, allowing us to interrogate the graph using Corollary 3. If in addition our model admits no context-specific conditional independence properties, then the symmetries in the SCEG mean that we need only check the answer to a **single** query, for instance $A \amalg B \mid (C = 1)$?

Example 3.1. Figure 3 shows the SCEG $\mathcal{C}$ from Figure 1 conditioned on the intrinsic event $\Lambda = (X(w_1) = 1) \cup (X(w_2) = 1)$ or *displayed symptom S before puberty*. This graph has a stalk at w_3, and by Theorem 2 we have that $X(w_0) \amalg \{X(w_3), X(w_6), X(w_7)\}$ in this graph.

Consider the position cuts $R_0 = \{w_0\}, R_1 = \{w_1, w_2\}, R_2 = \{w_3, w_4, w_5\}$, $R_3 = \{w_5, w_6, w_7, w_8, w_9\}$ of $\mathcal{C}$. Then as Figure 3 depicts a conditioned CEG $\mathcal{C}_\Lambda$ for the intrinsic event Λ, Corollary 1 gives us that

$$Y(R_0) \amalg (Y(R_2), Y(R_3)) \mid \Lambda$$

Now the CEG $\mathcal{C}$ from Figure 1 does not have a **natural** product space structure, but this is no obstacle to our using Corollary 3 here. As $\mathcal{C}_\Lambda$ does admit a product space structure we can impose this onto $\mathcal{C}$ by for example defining $A \equiv Y(R_0), B \equiv Y(R_1)$, $D \equiv Y(R_3)$ and

$$C = \begin{cases} 1 & \text{if } \sup(X(w_3), X(w_4)) = 1 \\ 2 & \text{otherwise} \end{cases}$$

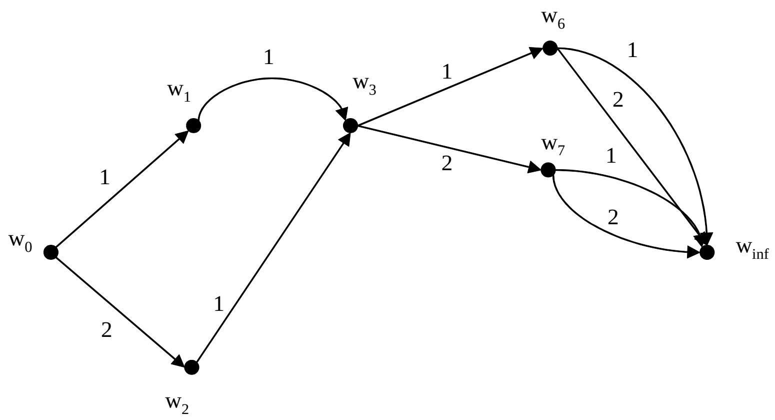

FIG 3. $\mathcal{C}$ *conditioned on* displayed symptom S before puberty

This allows us to use Corollary 3 and gives us that

$$A \amalg (C, D) \mid (B = 1)$$

which in our medical context reads as *whether an individual develops the condition and whether they die before 50 are independent of their gender given that they displayed symptom S before puberty.*

4. Regular CEGs. Although SCEGs form an important class of graphical model, by adding extra structure to them we can make them even more expressive. We do this by colouring positions and edges. The resultant graph is called a *regular Chain Event Graph* (RCEG). We note that coloured graphs have recently been found to provide a valuable embellishment to other graphical models (see for example [9]).

An RCEG is a coloured SCEG $\mathcal{C}$ where the set $V(\mathcal{C})$ has an associated partition $U(\mathcal{C}) = \{u_1, u_2, \ldots u_t\}$ for which each set $u \subset V(\mathcal{C})$ is regular. The set u is called a *stage* and is such that for each $w \in u$ the distribution function of $X(w) \mid (I(w) = 1)$ is dependent only on u and not on the particular $w \in u$.

DEFINITION 4. $w_1, w_2 \in V(\mathcal{C}) \backslash \{w_\infty\}$ are in the same *stage* u if there exists a bijection $\psi(w_1, w_2)$ between $E(w_1)$ and $E(w_2)$ such that if $\psi : e_x(w_1) \mapsto e_x(w_2)$ then $p(\Lambda(e_x(w_1)) \mid \Lambda(w_1)) = p(\Lambda(e_x(w_2)) \mid \Lambda(w_2))$.
The positions w_1, w_2 have the same *colour* if they are in the same stage,

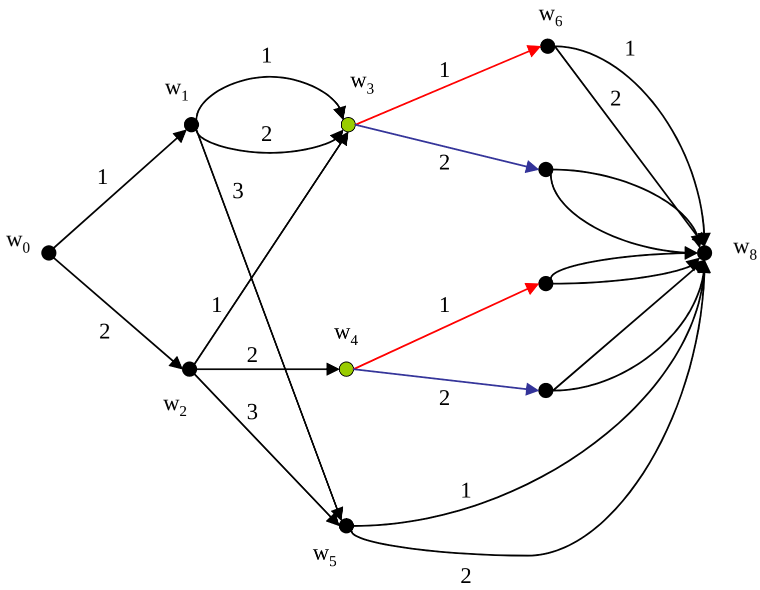

FIG 4. *The RCEG for Example 4.1*

and the edges $e_x(w_1), e_x(w_2)$ have the same *colour* if w_1, w_2 are in the same stage and $e_x(w_1)$ maps to $e_x(w_2)$ under this bijection.

The existence or otherwise of a bijection between two edge sets is normally apparent from the context of the problem. Note that if $e_x(w_1)$ maps to $e_x(w_2)$ under a bijection ψ, then these edges must correspond to the same outcome (for example *patient dies*) given the two histories $\Lambda(w_1)$ and $\Lambda(w_2)$. We call the colouring of the RCEG the *stage-structure* of the graph.

EXAMPLE 4.1. Producing an RCEG from the SCEG in Figure 1 we can add the extra information that the positions w_3 and w_4 are in the same stage – that is the probability of developing the condition (or not) is the same whether a member of the population has attributes corresponding to the sub-paths $(e_1(w_0), e_1(w_1)), (e_1(w_0), e_2(w_1)), (e_2(w_0), e_1(w_2))$ or $(e_2(w_0), e_2(w_2))$. The RCEG $\mathcal{C}$ is given in Figure 4.

This additional structure allows us to express a richer set of context-specific properties and sample space information than we can with the SCEG. The class of models expressible as an RCEG includes as a proper subset the class of models expressible as faithful regular or context-specific BNs on finite variables. Unlike the BN, the RCEG embodies the structure of the model state space and any context-specific information in its topology and colouring.

RCEGs are route-compatible and have the monomial property for a population Ψ if their underlying SCEG does, and hence are valid for a population Ψ if their underlying SCEG is. The subgraph $\mathcal{C}_\Lambda$ of an RCEG $\mathcal{C}$ conditioned on an intrinsic event Λ is an RCEG. Theorem 1 holds for RCEGs. Note however that $\mathcal{C}_\Lambda$ may not have the same stage-stucture as $\mathcal{C}$ in that positions or edges which have the same colour in $\mathcal{C}$ may have different colours in $\mathcal{C}_\Lambda$. Lemma 1 and the *position independence property* hold for RCEGs.

The conditions stipulated in Corollaries 1 and 3 can now be relaxed. It is sufficient that $\mathcal{C}_\Lambda$ should be simple (rather than $\mathcal{C}$) for these results to hold.

The subgraph of a CEG which consists of a position w, the sink-node w_∞, and all edges and positions which lie on a $w \to w_\infty$ subpath is called the subgraph rooted in w. When the CEG is used as a practical tool it is important to maximise its representational efficiency. So if in the subgraph $\mathcal{C}_\Lambda$, the subgraphs rooted in the positions w_α and w_β have identical topologies and colouring we can combine the positions w_α and w_β into a single position [30]. Note that if we do this then $\mathcal{C}_\Lambda$ although now minimal, is no longer a subgraph of $\mathcal{C}$ (see Definition 3).

Following the ideas of section 3, we let

$$J(u) = \sup_{w \in u} I(w) \qquad \text{and} \qquad Y(u) = \sup_{w \in u} X(w)$$

The RCEG is also a powerful tool for interrogation purposes, but to maximise its potential in this area we use the *Augmented Chain Event Graph* (ACEG) described in the next section.

5. Augmented CEGs.

5.1. *Definition of an Augmented CEG.* Analogously to the definition of $\boldsymbol{X}_A$, let $\boldsymbol{Y}_A = \{Y(u) : u \in A\}$ and $\boldsymbol{J}_A = \{J(u) : u \in A\}$. Since the CEG $\mathcal{C}$ is a DAG, there exists a partial order of the stages in the set $U(\mathcal{C})$. Let $P(u)$ be the set of all u' stages that precede u in this partial order. Let $\boldsymbol{Y}_{Q(u)}$ be a minimal subset of $\boldsymbol{Y}_{P(u)}$ such that

$$J(u) \amalg \boldsymbol{Y}_{P(u)} \mid \boldsymbol{Y}_{Q(u)}$$

Definition 5. An *augmented CEG (ACEG)* $\mathcal{A}(\mathcal{C})$ is a function of the CEG $\mathcal{C}$ with vertex set $V(\mathcal{A}(\mathcal{C})) = \{J(u) : u \in U(\mathcal{C})\} \cup \{Y(u) : u \in U(\mathcal{C})\}$.

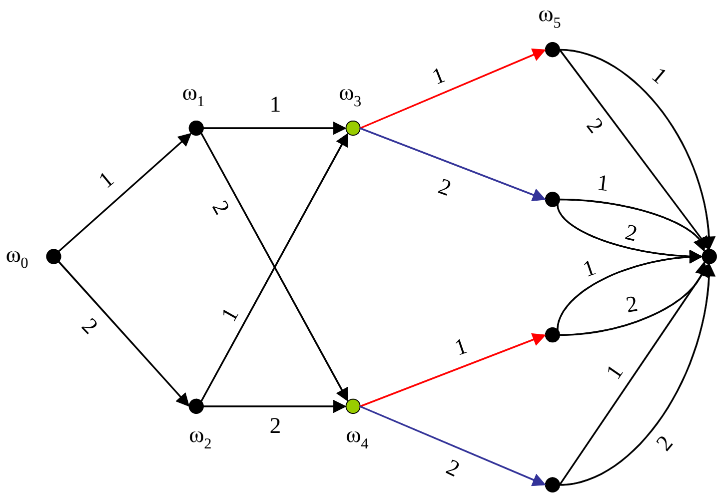

FIG 5. *The RCEG for Example 5.1*

The edge set $E(\mathcal{A}(\mathcal{C}))$ consists of directed edges connecting the parents of any vertex in $V(\mathcal{A}(\mathcal{C}))$ to that vertex. Each vertex $Y(u)$ has a single parent $J(u)$, and the parents of $J(u)$ are precisely those $Y(u')$ vertices that are members of $\boldsymbol{Y}_{Q(u)}$.

EXAMPLE 5.1. A research group has taken a sample from the population described in Section 2.1 which contains only people who displayed symptom S. Analysis of this sample suggests that *whether an individual develops the condition* **and** *whether they die before 50 are independent of their gender given when they displayed symptom S*. The RCEG for this is given in Figure 5. An ACEG for this graph is given in Figure 6, where for illustrative convenience the edges emanating from $Y(u)$ nodes have been labelled with values of A $(= Y(R_0)$ for $R_0 = \{w_0\})$, B $(= Y(R_1)$ for $R_1 = \{w_1, w_2\})$, and C $(= Y(R_2)$ for $R_2 = \{w_3, w_4\})$.

5.2. *ACEGs are Bayesian Networks.* We extend the notation of section 3 to let $\boldsymbol{X}_{D^c(u)}$ be the vector of random variables of the form $X(w)$ associated with positions in $\mathcal{C}$ which do not lie downstream of the stage u. Let

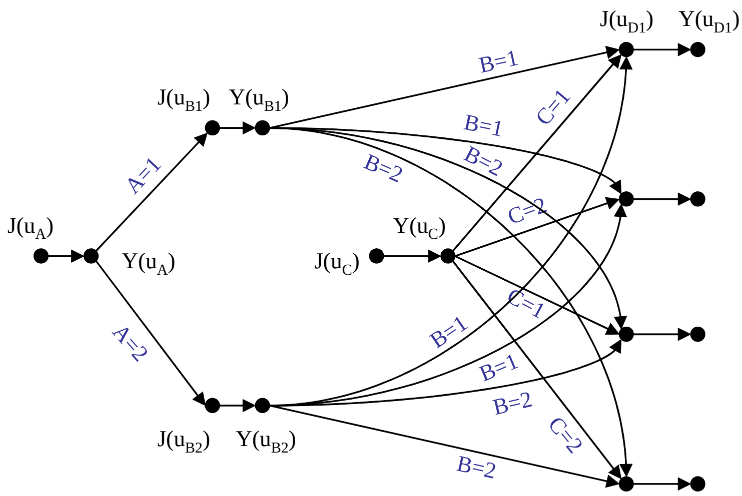

FIG 6. *An ACEG for the RCEG in Figure 5*

$\boldsymbol{Y}_{D^c(u)}, \boldsymbol{J}_{D^c(u)}$ be the vectors of random variables of the form $Y(u'), J(u')$ associated with stages in $\mathcal{C}$ which do not lie downstream of the stage u.

LEMMA 4. *For CEG $\mathcal{C}$, and stage $u \in U(\mathcal{C})$*

$$Y(u) \amalg \boldsymbol{Y}_{D^c(u)} \mid J(u)$$

The result given in this lemma is analogous to that given in Lemma 2 for positions, and so also to the result quoted there for BNs. It provides a set of conditional independence statements that can simply be read from the graph, one for each stage in $U(\mathcal{C})$. A partial reading of the lemma gives us that the immediate future for a unit at a stage u is independent of how the unit reached that stage. The proof of the lemma is in the appendix.

By construction, if a stage u' is not downstream of u in $\mathcal{C}$, then $J(u'), Y(u')$ are not downstream of $J(u), Y(u)$ in $\mathcal{A}(\mathcal{C})$. Since for every stage u', $J(u')$ is a function of $Y(u')$, it follows that

$$Y(u) \amalg (\boldsymbol{J}_{D^c(u)}, \boldsymbol{Y}_{D^c(u)}) \mid J(u)$$

and hence that

$$Y(u) \amalg (\boldsymbol{J}_{P(u)}, \boldsymbol{Y}_{P(u)}) \mid J(u)$$

in any partial order of $U(\mathcal{C})$. Clearly we also have that

$$J(u) \amalg (\boldsymbol{J}_{P(u)}, \boldsymbol{Y}_{P(u)}) \mid \boldsymbol{Y}_{Q(u)}$$

and hence that **all** vertices in an ACEG $\mathcal{A}(\mathcal{C})$ are independent of their predecessor vertices given their parental vertices in any partial order of $U(\mathcal{C})$.

In [19] it is shown that a probability distribution P is Markov relative to a DAG $\mathcal{G}$ if and only if each variable in $\mathcal{G}$ is independent of all its predecessors conditional on its parents, in some ordering of the variables that agrees with the arrows of $\mathcal{G}$. Clearly our ACEG is a DAG, and from the above reasoning each variable in $\mathcal{A}(\mathcal{C})$ is independent of all its predecessors conditional on its parents for all P defined on the CEG $\mathcal{C}$. So our ACEG obeys what Pearl [20] calls the *ordered* Markov condition, and hence also obeys the *local* Markov condition [13]. Results in [10] allow us therefore to deduce that the ACEG is itself a BN.

This deduction means that any result available for use with BNs can also be used with ACEGs. In particular we can use d-separation to allow us to interrogate ACEGs for conditional independence properties. The advantage that the ACEG has here over the BN is that in the former context-specific conditional independence properties are depicted **explicitly** in the topology of the graph, and so it can be interrogated directly for such properties. We begin however by looking at models which can be represented by BNs.

6. Models depictable by Bayesian Networks and others. If a model has a natural product space structure and admits no context-specific conditional independence properties then it can be depicted by a BN without any further annotation. In this section we show that if our CEG is of such a model then any separation-based conditional independence property readable from the BN can also be read from its associated ACEG.

If our CEG is of a model which has a natural product space structure then for each variable X_i in the BN there exists a collection of vertices $\{J(u_i)\}$ in the ACEG whose members correspond to the possible configurations of $Q(X_i)$ (the parent variables of X_i), and a collection of vertices $\{Y(u_i)\}$ whose members correspond to X_i given those configurations.

THEOREM 3. *If a model with a natural product space structure admitting no context-specific conditional independence properties, has a BN representation $\mathcal{G}$, and a CEG representation $\mathcal{C}$, then:*

If $\{Y(u_i)\}$ *is d-separated from* $\{Y(u_j)\}$ *by* $\{Y(u_k)\}$ *in* $\mathcal{A}(\mathcal{C})$, *then* $X_i \amalg X_j \mid X_k$ *in every distribution compatible with* $\mathcal{C}$ *and* $\mathcal{G}$.
Conversely, if $X_i \amalg X_j \mid X_k$ *holds for all distributions compatible with* $\mathcal{C}$ *and* $\mathcal{G}$, *then* $\{Y(u_i)\}$ *is d-separated from* $\{Y(u_j)\}$ *by* $\{Y(u_k)\}$ *in* $\mathcal{A}(\mathcal{C})$.

The proof of this theorem is given in the appendix. This result can be explained as follows: The CEG $\mathcal{C}$ is of a model which has a natural product space structure admitting no context-specific conditional independence properties, and can be depicted by a BN $\mathcal{G}$. Therefore there exist (in $\mathcal{A}(\mathcal{C})$) edges from vertices in $\{Y(u_i)\}$ to vertices in $\{J(u_j)\}$ if and only if there exists an edge from X_i to X_j in $\mathcal{G}$, and so there is a 1:1 correspondence between the *parental* conditional independence statements in $\mathcal{G}$ and the *parental* conditional independence statements in $\mathcal{A}(\mathcal{C})$. By [31] Corollary 1, the conditional independence statements in a DAG can be derived from d-separation if and only if they can be derived from the list of *parental* conditional independence statements using the semi-graphoid axioms [28]. As both $\mathcal{G}$ and $\mathcal{A}(\mathcal{C})$ are DAGs, we can infer that there is a 1:1 correspondence between the conditional independence statements derived from d-separation in $\mathcal{G}$ and the conditional independence statements derived from d-separation in $\mathcal{A}(\mathcal{C})$.

Essentially, Theorem 3 allows us to use the collections $\{Y(u_i)\}$ in the ACEG as *surrogates* for X_i in the BN, when answering conditional independence queries.

Example 6.1. For the RCEG in Figure 5, let A, B, C be as in Example 5.1, $R_3 = \{w_5, w_6, w_7, w_8\}$ and $D = Y(R_3)$. Then using Theorem 3 on the ACEG for this RCEG (given in Figure 6), we see that

$$\begin{aligned}
Y(u_A) \text{ is d-separated from } \{Y(u_C), Y(u_D)\} \text{ by } \{Y(u_B)\} &\Rightarrow A \amalg (C, D) \mid B \\
\{Y(u_A), Y(u_B)\} \text{ are d-separated from } \{Y(u_C)\} &\Rightarrow (A, B) \amalg C \\
Y(u_A) \text{ is not d-separated from } Y(u_C) \text{ by } \{Y(u_D)\} &\Rightarrow A \not\amalg C \mid D \\
Y(u_B) \text{ is not d-separated from } Y(u_C) \text{ by } \{Y(u_D)\} &\Rightarrow B \not\amalg C \mid D
\end{aligned}$$

for the RCEG in Figure 5. In our medical context *whether an individual develops the condition and whether they die before the age of 50 are independent of their gender given when they displayed symptom S; whether they develop the condition is independent of their gender and when they displayed symptom S;* but *whether they develop the condition is* **not** *independent of either their gender or when they displayed symptom S given whether or not they die before the age of 50*, for the sample considered in Example 5.1.

COROLLARY 4. *If X_i, X_j, X_k are distinct subsets of the vertex-variables of $\mathcal{G}$, and $\{Y(u_i)\}, \{Y(u_j)\}, \{Y(u_k)\}$ are the corresponding collections in $\mathcal{A}(\mathcal{C})$, then the results of Theorem 3 still hold.*

This follows from the proof of Theorem 3 (which does not depend on X_i, X_j, X_k being single variables).

We have called the collections $\{Y(u_i)\}$ in $\mathcal{A}(\mathcal{C})$ surrogates for X_i in $\mathcal{G}$, but when we replace the statement "$X_i \amalg X_j \mid X_k$ *in* $\mathcal{G}$" by "$\{Y(u_i)\}$ *is d-separated from* $\{Y(u_j)\}$ *by* $\{Y(u_k)\}$ *in* $\mathcal{A}(\mathcal{C})$", only $\{Y(u_k)\}$ is actually a surrogate. This is because the latter statement implies that $\{Y(u_i)\} \amalg \{Y(u_j)\} \mid \{Y(u_k)\}$ (since $\mathcal{A}(\mathcal{C})$ is a BN), and if this statement is true then $X_i \amalg X_j \mid \{Y(u_k)\}$ since X_i ($\equiv \sup Y(u_i)$) is a function of $\{Y(u_i)\}$. By construction only one $Y(u_i)$ within the set $\{Y(u_i)\}$ can take a non-zero value, and the value this variable takes is equal to the value taken by X_i.

Note that the ACEG has $J(u)$ and $Y(u)$ nodes for each stage $u \in U(\mathcal{C})$, and each stage u in a CEG is associated with a particular collection of parents – the set of $u' \in U(\mathcal{C})$ corresponding to $\boldsymbol{Y}_{Q(u)}$ in the ACEG. Indeed, if our CEG has sufficient symmetry to be embedded into a family of models with a product space structure, then the positions constituting each stage u are members of a specific orthogonal cut R (section 2.1), and u encodes a particular configuration of the parental variables of $Y(R)$.

For this reason an ACEG has many more nodes than a standard BN, and so admits a far larger collection of conditional independence statements. This collection includes many context-specific properties which can only be represented in BNs by modifying their structure [4, 16, 22, 24]. It also includes many counterfactual statements of the type described in Section 2.5 on SCEGs. So for example, if we consider the CEG in Figure 2, but combine the positions w_6, w_7, w_8 and w_9 into a sink-node w_∞, we get the ACEG depicted in Figure 7, where for convenience we have let $A = Y(R_0)$, $B = Y(R_1)$, $C = Y(R_2)$ with R_0, R_1, R_2 defined as in Example 5.1 above.

Using the ACEG in Figure 7 we can deduce that $C \amalg B \mid (A = 1)$ – *whether an individual develops the condition is independent of when they displayed symptom S given that their gender is male (and symptom S was displayed)*, and that $C \amalg A \mid (B = 1)$ – *whether they develop the condition is independent of their gender given that they displayed symptom S before puberty.* But we also have statements such as $Y(u_{C2}) \amalg Y(u_{B1}) \mid Y(u_A)$, which has no obvious meaning in the context of the problem.

7. More on context-specific conditional independence. One of the distinct advantages of the CEG when representing and analysing asymmetric problems is that we can examine the effects of conditioning on a

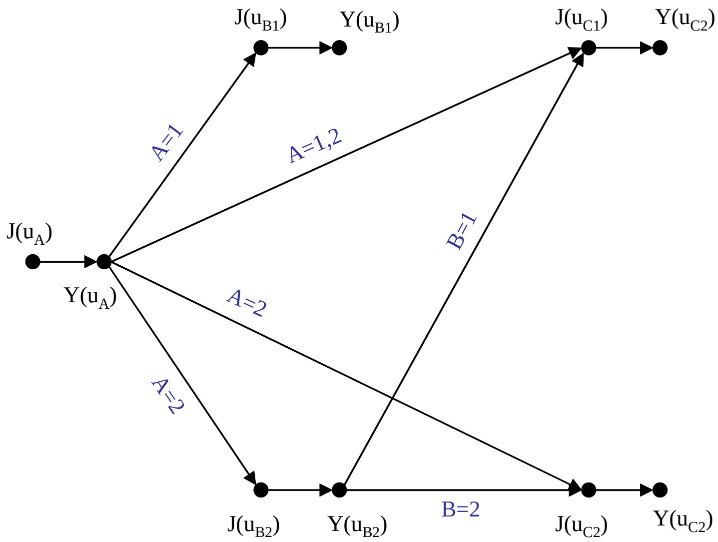

FIG 7. *An ACEG for the adapted CEG from Figure 2*

specific event, perhaps a specific value of a variable, and use the CEG's topology to discover conditional independencies which would not exist if we were to condition on a related event, such as a different value of our variable. If we are interested in these context-specific conditional independence properties then in this discrete context we need to concentrate our attention on statements where the conditioning element is an event. In most cases this event will be expressible as a value of a single $Y(u)$ (or $J(u)$) variable, and so queries can be checked directly on an ACEG without the need of the *surrogate* argument of the last section. What happens in cases where our conditioning event cannot be expressed as a value of a $Y(u)$ (or $J(u)$) variable? An example of this is the event $\Lambda = (B = 1)$ for the model depicted in Figure 4. We could draw an ACEG for the full CEG $\mathcal{C}$ here, but the ACEG for $\mathcal{C}_\Lambda$ is much more useful. The sub-CEG $\mathcal{C}_\Lambda$ for this event is given in Figure 3. Note that the edge-probabilities on this graph are now $A = 1 \mid B = 1$, $A = 2 \mid B = 1$ for the edges leaving w_0; 1 for the edges leaving w_1 & w_2 (the positions w_1 & w_2 could be combined into a single position as suggested in Section 4); $C = 1,\ C = 2$ for the edges leaving w_3; $D = 1 \mid B = 1, C = 1, \quad D = 2 \mid B = 1, C = 1, \quad D = 1 \mid B = 1, C = 2$

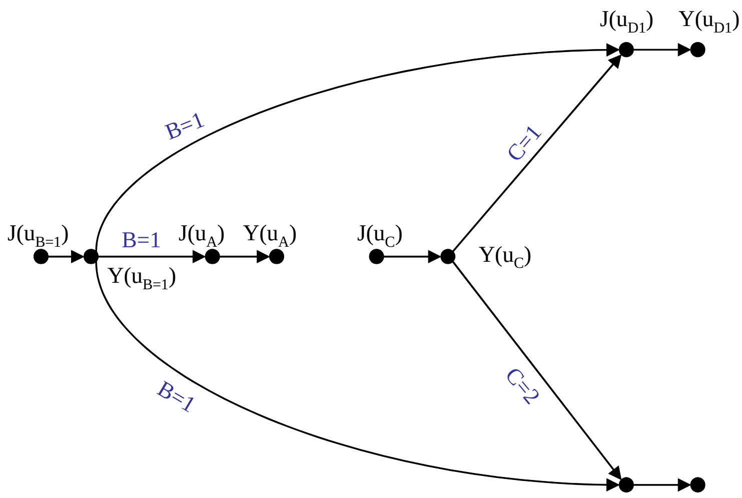

FIG 8. *An ACEG for the sub-CEG from Figure 3*

and $D = 2 \mid B = 1, C = 2$ for the edges leaving w_5 & w_6 (see Section 2.3 and [29]).

An ACEG for $\mathcal{C}_\Lambda$ is given in Figure 8. Notice that unlike in Figure 6, $J(u_A)$ is not a root-vertex as A is now dependent on B. Also, because we have conditioned on the event $(B = 1)$, the set of $Y(u_B)$ vertices has become a **single** vertex $Y(u_{B=1})$ with no ancestors except $J(u_{B=1})$. We now use d-separation to read that

$$\begin{aligned} & \{Y(u_C), Y(u_D)\} \text{ are d-separated from } Y(u_A) \text{ by } Y(u_{B=1}) \\ \Rightarrow\ & (Y(u_C), \{Y(u_D)\}) \amalg Y(u_A) \mid Y(u_{B=1}) \\ \Rightarrow\ & (C, D) \amalg A \mid Y(u_{B=1}) \end{aligned}$$

since the ACEG is a BN, and A, C, D are functions of $Y(u_A), Y(u_C), \{Y(u_D)\}$. Also $Y(u_{B=1}) = 1 \Leftrightarrow B = 1$, so this in turn implies that $(C, D) \amalg A \mid (B = 1)$ without the use of the *surrogate* argument.

This method will always work for cases like this since conditioning on an event such as $B = 1$ always produces a single vertex of the form $Y(u_{B=1})$, which takes the value 1 if and only if $B = 1$.

As with SCEGs (see Section 3), standard conditional independence queries on an RCEG can generally be answered by looking at context-specific conditional independence queries on subgraphs of the CEG which are often simple. Such conditioning can only remove colouring from the graph and not add it. Because of the way CEGs are constructed, the conditioning event in a context-specific conditional independence query can very often be written as $\Lambda(w)$ for some $w \in V(\mathcal{C})$. But if we condition on an event $\Lambda = \Lambda(w)$, we can read conditional independence properties off the graph $\mathcal{C}_\Lambda$ even if $\mathcal{C}_\Lambda$ is not simple.

COROLLARY 5. *Let C be an RCEG with position cuts $R_a = \{w_a\}$, $R_b = \{w_b\}$. If w_a, w_b are separated by a stalk, for* **any** *$w_a \in R_a, w_b \in R_b$, then $Y(R_a) \amalg Y(R_b)$.*

The proof of this corollary is in the appendix. This result can obviously be extended to give sufficient conditions for $Y(R_a) \amalg Y(R_b) \mid \Lambda$ just as Corollary 1 extends the result for SCEGs.

So if $\Lambda = \Lambda(w)$ for some $w \in V(\mathcal{C})$, then w is a stalk in $\mathcal{C}_\Lambda$, and in this graph, $Y(R_a) \amalg Y(R_b)$ for any position cuts R_a upstream of w, and R_b downstream of w. Hence $Y(R_a) \amalg Y(R_b) \mid \Lambda$ providing we have defined these variables consistently on the CEGs $\mathcal{C}$ and $\mathcal{C}_\Lambda$ (see proof of Corollary 1).

EXAMPLE 7.1. Consider the RCEG from Figure 4 conditioned on the event $\Lambda = (X(w_1) = 1) \cup (X(w_1) = 2) \cup (X(w_2) = 1) \cup (X(w_2) = 2)$. The RCEG for this is given in Figure 9.

Ignoring the medical context here, we note that in this graph the event $\Lambda = \Lambda(w_3)$ can be characterised as $(\min(A, B) = 1)$. If we condition on this event we get $\mathcal{C}_\Lambda$ as in Figure 10, which as already noted **must** have a stalk. For illustrative convenience edges in Figure 10 have been given probability labels.

Using Corollary 5 we get $(C, D) \amalg (A, B) \mid (\min(A, B) = 1)$, and conditioning on $\Lambda = \Lambda(w_4)$ we get a CEG from which we can trivially read that $(C, D) \amalg (A, B) \mid (\min(A, B) = 2)$. Combining these we get $(C, D) \amalg (A, B) \mid \min(A, B)$.

8. Conclusion. In summary, the results of section 3 give us conditions for the truth of $A \amalg B$ statements on SCEGs which are directly analogous to those given in (for example Pearl's [20] or Lauritzen's [14] versions of) the d-separation theorem for BNs. Corollary 1 also gives us sufficient conditions for $A \amalg B \mid \Lambda$ statements to hold. Subsequent sections give us sufficient

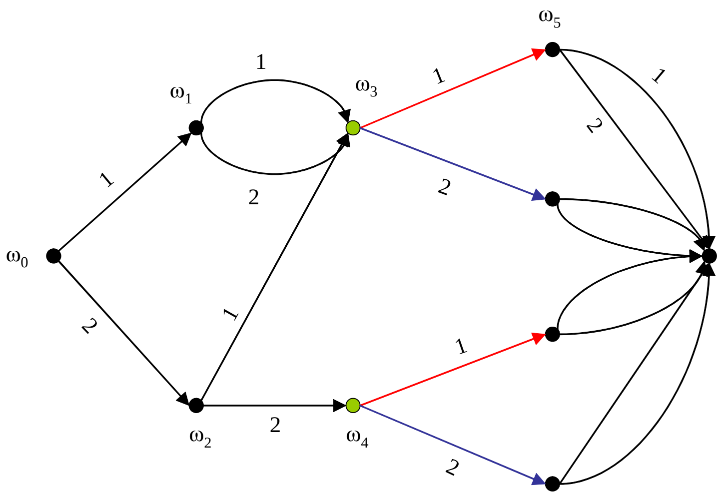

FIG 9. *The RCEG for Example 7.1*

conditions for $A \amalg B \mid \Lambda$ statements to hold on RCEGs. Queries such as $A \amalg B \mid C$? where the conditioning element is also a variable (or collection of variables) can generally be answered by considering sets of queries of the form $A \amalg B \mid \Lambda$? Methods for doing this are suggested at various points in the text, but in the special case where the RCEG describes a model which can be depicted by a BN, Theorem 3 gives conditions for $A \amalg B \mid C$ directly analogous to those given in the d-separation theorem for BNs. The ACEG from Section 5 is very useful for all types of conditional independence query, but is particularly useful for queries of the form $A \amalg B \mid \Lambda$? in situations where using other techniques is not straightforward. The fact that the ACEG is itself a BN opens up an exciting range of possibilities still to be explored.

Analysts working with BNs have found that attempts to feed back to a user **all** the implicit conditional independencies associated with a given graph can be rather overwhelming unless the BN is very simple. Clearly this would also be the case with CEG-based models. However, within any given context the types of independencies that it is natural for the user to be able to understand, examine and verify are small in number. Since the identification of such *natural* relationships is dependent on the domain of

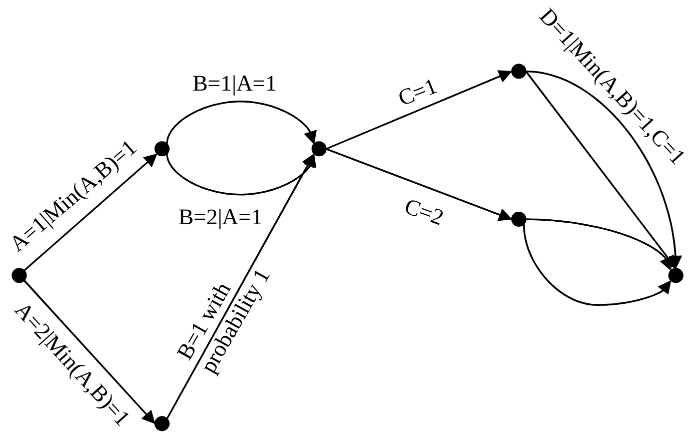

FIG 10. *The RCEG from Figure 9 conditioned on $\Lambda = \Lambda(w_3)$*

application of the CEG we defer this discussion to a future paper.

APPENDIX 1: PROOFS AND ONE ADDITIONAL LEMMA

LIMITED MEMORY LEMMA. *For any CEG $\mathcal{C}$, $w_1, w_2, w_3 \in V(\mathcal{C})$ with $w_1 \prec w_2 \prec w_3$,*

$$I(w_3) \amalg I(w_1) \mid (I(w_2) = 1)$$

PROOF. It is sufficient to prove that

$$p(I(w_3) = 1 \mid I(w_1) = 1, I(w_2) = 1) = p(I(w_3) = 1 \mid I(w_2) = 1)$$

So consider a single route λ passing through w_1, w_2, w_3. This route consists of a set of edges and by construction the probability $p(\lambda)$ of the route is equal to the product of the probabilities labelling each of these edges. Moreover, the probability of any subpath of λ is equal to the product of the probabilities labelling each of its edges. So $p(\lambda)$ 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

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

Consider now the event $(I(w_1) = 1, I(w_2) = 1, I(w_3) = 1)$ or $\Lambda(w_1, w_2, w_3)$, which is the union of all $w_0 \to w_\infty$ routes passing through w_1, w_2, w_3. Then since $\Lambda(w_1, w_2, w_3)$ is an intrinsic event we can write

$$\begin{aligned} p(\Lambda(w_1, w_2, w_3)) = \Big(\sum_{\mu_0 \in M_0} \pi_{\mu_0}(w_1 \mid w_0) \Big) \Big(\sum_{\mu_1 \in M_1} \pi_{\mu_1}(w_2 \mid w_1) \Big) \\ \times \Big(\sum_{\mu_2 \in M_2} \pi_{\mu_2}(w_3 \mid w_2) \Big) \Big(\sum_{\mu_3 \in M_3} \pi_{\mu_3}(w_\infty \mid w_3) \Big) \end{aligned}$$

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 $\sum_{\mu_0 \in M_0} \pi_{\mu_0}(w_1 \mid w_0)$ is simply the probability of reaching w_1 from w_0, or $\pi(w_1 \mid w_0)$, so

$$\begin{aligned} p(\Lambda(w_1, w_2, w_3)) &= \pi(w_1 \mid w_0)\ \pi(w_2 \mid w_1)\ \pi(w_3 \mid w_2)\ \pi(w_\infty \mid w_3) \\ &= \pi(w_1 \mid w_0)\ \pi(w_2 \mid w_1)\ \pi(w_3 \mid w_2) \times 1 \end{aligned}$$

since all paths passing through w_3 terminate in w_∞. Therefore

$$\begin{aligned} p(I(w_3) = 1 \mid I(w_1) = 1, I(w_2) = 1) &= \frac{p(\Lambda(w_1, w_2, w_3))}{p(\Lambda(w_1, w_2))} \\ &= \frac{\pi(w_1 \mid w_0)\ \pi(w_2 \mid w_1)\ \pi(w_3 \mid w_2) \times 1}{\pi(w_1 \mid w_0)\ \pi(w_2 \mid w_1)\ \times 1} \\ &= \pi(w_3 \mid w_2) \\ &= p(I(w_3) = 1 \mid I(w_2) = 1) \qquad \square \end{aligned}$$

If we replace $I(w_1) = 1$ by $\Lambda(\mu(w_0, w_2))$ for any subpath $\mu(w_0, w_2)$, and $I(w_3) = 1$ by $\Lambda(e(w_2, w_2'))$ for some edge $e(w_2, w_2')$ then we obtain

COROLLARY A. *For any CEG $\mathcal{C}$ with $w \in V(\mathcal{C})$*

$$p(\Lambda(e(w, w')) \mid \Lambda(\mu(w_0, w)), \Lambda(w)) = p(\Lambda(e(w, w')) \mid \Lambda(w))$$

Similarly, if $w_1' \prec w_2$ and we replace $I(w_1) = 1$ by $\Lambda(e(w_1, w_1'))$ $(X(w_1) = x_1$ for some $x_1 \in 1, \ldots k(w_1))$, and $I(w_3) = 1$ by $\Lambda(e(w_3, w_3'))$ $(X(w_3) = x_3$ for some $x_3 \in 1, \ldots k(w_3))$ then we obtain

COROLLARY B. *For any CEG $\mathcal{C}$, $w_1, w_2, w_3 \in V(\mathcal{C})$, with $w_1' \prec w_2 \prec w_3$*

$$p(\Lambda(e(w_3, w_3')) \mid \Lambda(e(w_1, w_1')), \Lambda(w_2)) = p(\Lambda(e(w_3, w_3')) \mid \Lambda(w_2))$$
$$\Big(p(X(w_3) = x_3 \mid X(w_1) = x_1, I(w_2) = 1) = p(X(w_3) = x_3 \mid I(w_2) = 1) \Big)$$

PROOF OF THEOREM 1. Since the atoms of $\mathbb{F}(\mathcal{C}|\Lambda)$ are routes in $\mathcal{C}_\Lambda$, $\mathbb{F}(\mathcal{C}|\Lambda) = \mathbb{F}(\mathcal{C}_\Lambda)$, and so the conditioned model is route compatible (section 2.1). For each atom $\lambda \in \Lambda$, the probability mass function of this atom in $\mathcal{C}_\Lambda$ is given by $p_\Lambda(\lambda) = p(\lambda|\Lambda)$ which equals

$$\begin{aligned}
&p(\Lambda(w_0), \Lambda(e(w_0, w_1)), \Lambda(e(w_1, w_2)), \ldots \Lambda(e(w_m, w_\infty)) \mid \Lambda) \\
&= p_\Lambda(\Lambda(w_0), \Lambda(e(w_0, w_1)), \Lambda(e(w_1, w_2)), \ldots \Lambda(e(w_m, w_\infty))) \\
&= p_\Lambda(\Lambda(w_0))\; p_\Lambda(\Lambda(e(w_0, w_1) \mid \Lambda(w_0)) \\
&\quad \times p_\Lambda(\Lambda(e(w_1, w_2) \mid \Lambda(w_0), \Lambda(e(w_0, w_1)) \\
&\quad \times \ldots\; p_\Lambda(\Lambda(e(w_m, w_\infty)) \mid \Lambda(w_0), \ldots\; \Lambda(e(w_{m-1}, w_m)))
\end{aligned}$$

which by Corollary A of the Limited Memory Lemma equals

$$\begin{aligned}
&1 \times p_\Lambda(\Lambda(e(w_0, w_1) \mid \Lambda(w_0))\; p_\Lambda(\Lambda(e(w_1, w_2) \mid \Lambda(w_1)) \\
&\quad \times p_\Lambda(\Lambda(e(w_2, w_3) \mid \Lambda(w_2)) \times \ldots\; p_\Lambda(\Lambda(e(w_m, w_\infty)) \mid \Lambda(w_m)) \\
&= \prod_{e(w,w') \in \lambda} p_\Lambda(\Lambda(e(w, w')) \mid \Lambda(w))
\end{aligned}$$

So letting $\pi_e^*(w' \mid w) = p_\Lambda(\Lambda(e(w, w')) \mid \Lambda(w))$ we have

$$p(\lambda|\Lambda) = \prod_{e(w,w') \in \lambda} \pi_e^*(w' \mid w)$$

and the probability mass function $p(\lambda|\Lambda)$ has the monomial property. Hence $\mathcal{C}_\Lambda$ is a valid SCEG.

□

Note that as stated in section 2.3

$$\begin{aligned}
p_\Lambda(\Lambda(e(w, w')) \mid \Lambda(w)) &= p(\Lambda(e(w, w')) \mid \Lambda, \Lambda(w)) \\
&= \frac{p(\Lambda \mid \Lambda(w), \Lambda(e(w, w')))}{p(\Lambda \mid \Lambda(w))}\; p(\Lambda(e(w, w')) \mid \Lambda(w)) \\
&= \frac{p(\Lambda \mid \Lambda(w), \Lambda(e(w, w')))}{p(\Lambda \mid \Lambda(w))}\; \pi_e(w' \mid w)
\end{aligned}$$

PROOF OF LEMMA 1. X, Y partition the set of atoms of $\mathcal{C}$, and since $\Lambda \subset \Lambda(C)$, X, Y also partition the set of atoms of $\mathcal{C}_\Lambda$. Consider arbitrary events Λ_X and Λ_Y from the sets $\{\Lambda_X\}$ and $\{\Lambda_Y\}$ (partitions of the set of atoms of $\mathcal{C}$), and the event $\Lambda_A = \Lambda_X \cap \Lambda_Y$. Then $p(\Lambda_X \mid \Lambda) = p_\Lambda(\Lambda_X)$, $p(\Lambda_Y \mid \Lambda) = p_\Lambda(\Lambda_Y)$ and $p(\Lambda_X, \Lambda_Y \mid \Lambda) = p(\Lambda_A \mid \Lambda) = p_\Lambda(\Lambda_A) = p_\Lambda(\Lambda_X, \Lambda_Y)$. The statement

$$p(\Lambda_X, \Lambda_Y \mid \Lambda) = p(\Lambda_X \mid \Lambda)\; p(\Lambda_Y \mid \Lambda)$$

is then true if and only if the statement

$$p_\Lambda(\Lambda_X, \Lambda_Y) = p_\Lambda(\Lambda_X)\; p_\Lambda(\Lambda_Y)$$

is true; and this holds for all $\Lambda_X \in \{\Lambda_X\}$ $\Lambda_Y \in \{\Lambda_Y\}$. $\square$

PROOF OF LEMMA 2. By definition $I(w) = 0 \Rightarrow X(w) = 0$, so if $I(w) = 0$, $X(w)$ is known and so in particular is independent of all other variables. If $I(w) = 1$ then $X(w') = 0$ for all $w' \in D^c(w) \cap U^c(w)$.
So consider $I(w) = 1$ and $w' \in U(w)$. We now use the monomial property to show that $X(w) \amalg \boldsymbol{X}_{U(w)} \mid (I(w) = 1)$.

The primitive probability $\pi_e(w^+ \mid w)$ is a factor of the probability $p(\lambda)$ for a number of routes. Consider one of these routes and denote the subpath of this route between w_0 and w by $\mu(w_0, w)$. Then by Corollary A of the Limited Memory Lemma, we can write

$$\begin{aligned}\pi_e(w^+ \mid w) &= p(\Lambda(e(w, w^+)) \mid \Lambda(w)) \\ &= p(\Lambda(e(w, w^+)) \mid \Lambda(\mu(w_0, w)), \Lambda(w))\end{aligned}$$

and this is clearly true for all subpaths between w_0 and w. But the set of these subpaths is in 1:1 correspondence with the set of vectors of values of $\boldsymbol{X}_{U(w)}$ which are consistent with the topology of the SCEG and with the event $I(w) = 1$. The event $\Lambda(w)$ can be written as $I(w) = 1$, and the event $\Lambda(e(w, w'))$ as $X(w) = x$ for some $x > 0$. Hence

$$p(X(w) = x \mid I(w) = 1) = p(X(w) = x \mid \boldsymbol{X}_{U(w)}, I(w) = 1)$$

and $X(w) \amalg \boldsymbol{X}_{U(w)} \mid (I(w) = 1)$.
Combining this with the two previous results we therefore have

$$X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid I(w) \qquad \square$$

PROOF OF THEOREM 2. (1) SUFFICIENT CONDITIONS FOR INDEPENDENCE
Consider an SCEG $\mathcal{C}$, and two positions $w_1, w_2 \in V(\mathcal{C})$, where $w_2 \not\prec w_1$, and by construction $I(w_1) \not\equiv 0$, $I(w_2) \not\equiv 0$.

Suppose there exists a stalk downstream of w_1 and upstream of w_2. Label this position w. Then all paths passing through w_1 pass through w, all paths passing through w_2 pass through w, and necessarily $w_1 \prec w \prec w_2$. Also

$$\begin{aligned}p(X(w_2) = x_2 \mid X(w_1) = x_1) &= p(X(w_2) = x_2 \mid X(w_1) = x_1, I(w) = 1) \\ &= p(X(w_2) = x_2 \mid I(w) = 1)\end{aligned}$$

since $w_1 \prec w \prec w_2$, and using Corollary B of the Limited Memory Lemma

$$= p(X(w_2) = x_2 \mid X(w_1) = x_1', I(w) = 1)$$
$$= p(X(w_2) = x_2 \mid X(w_1) = x_1')$$

for all values x_1, x_1' of $X(w_1)$ and all values x_2 of $X(w_2)$

$$\Rightarrow X(w_1) \amalg X(w_2)$$

If w_2 is itself a stalk, then we replace $I(w) = 1$ by $I(w_2) = 1$ in the above argument with the same result.

So a **sufficient** condition for $X(w_1) \amalg X(w_2)$ is that either w_2 is itself a stalk, or there exists a stalk downstream of w_1 and upstream of w_2.

(2) NECESSARY CONDITIONS FOR INDEPENDENCE

Let $X(w_1) \amalg X(w_2)$ (and since $I(w)$ is a function of $X(w)$, $X(w_1) \amalg I(w_2)$ and $I(w_1) \amalg I(w_2)$). Let the set of routes of $\mathcal{C}$ be partitioned into four subsets. Call a route Type A if it passes through w_2, but not through w_1, Type B if it passes through neither w_1 nor w_2, Type C if it passes through both w_1 and w_2, and Type D if it passes through w_1, but not through w_2. Our proof proceeds as follows:
(a) We show that we must have $w_1 \prec w_2$ (ie. the set of Type C routes is non-empty.
(b) We show that every route intersects with every other route at some point downstream of w_0 and upstream of w_∞. If two $w_0 \to w_\infty$ routes share no vertices except w_0 and w_∞, we call them *internally disjoint* (see for example [11]). So we can say that there cannot be two internally disjoint directed routes in $\mathcal{C}$
(c) We show that there cannot be two internally disjoint routes in the **undirected** version of the CEG (the CEG with its edge arrows removed), and that therefore there must be a stalk between w_0 and w_∞.
(d) We show that either w_1 is a stalk or w_2 is a stalk, or there exists a stalk downstream of w_1 and upstream of w_2.
(e) Finally we show that if w_1 is a stalk then there must also either be a stalk at w_2 or a stalk downstream of w_1 and upstream of w_2.

(a) Suppose that $w_1 \not\prec w_2$ (and recall that $w_2 \not\prec w_1$). Then $p(I(w_2) = 1 \mid I(w_1) = 1) \equiv 0$. $I(w_1) \amalg I(w_2) \Rightarrow p(I(w_2) = 1) \equiv 0 \Rightarrow I(w_2) \equiv 0$. This is impossible by construction. Therefore $w_1 \prec w_2$.

(b) We first show that each Type C route intersects with every other route at w_1 or at w_2 or at some point between these positions.

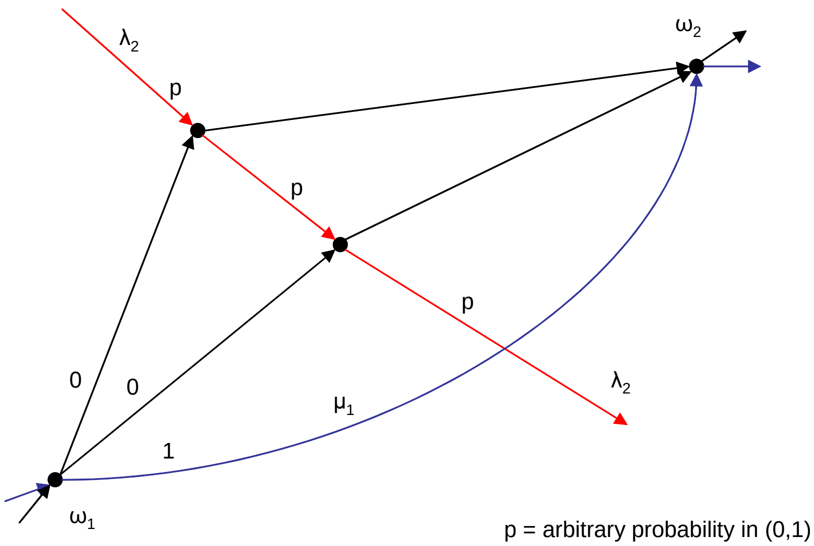

FIG 11. *Illustration for Type C and Type B routes*

Let λ_1 be a Type C route, and $\mu_1(w_1, w_2)$ the subpath coincident with λ_1 between w_1 and w_2. If the set of Type B routes is non-empty then let λ_2 be a Type B route which does not intersect with μ_1 (ie. λ_2 and μ_1 have no positions in common).

Consider a distribution P which (1) assigns a probability of 1 to every edge of the subpath $\mu_1(w_1, w_2)$, and (2) an arbitrary probability greater than 0 and less than 1 to each edge of the route λ_2 (Figure 11). If our SCEG is minimal and λ_2 does not intersect with μ_1 then this is always possible. Under P, assignment (1) gives us that

$$p(I(w_2) = 1 \mid I(w_1) = 1) = 1$$

and $I(w_1) \amalg I(w_2)$ implies that under this P

$$p(I(w_2) = 1 \mid I(w_1) = 0) = 1 \;\Rightarrow\; p(I(w_2) = 0 \mid I(w_1) = 0) = 0$$

But assignment (2) gives us that $p(I(w_2) = 0 \mid I(w_1) = 0) > 0$ ※

The assumption $I(w_1) \amalg I(w_2)$ is incompatible with the assignments of (1) and (2). But these assignments are always possible if λ_2 does not intersect with μ_1. Hence λ_2 must intersect with μ_1.

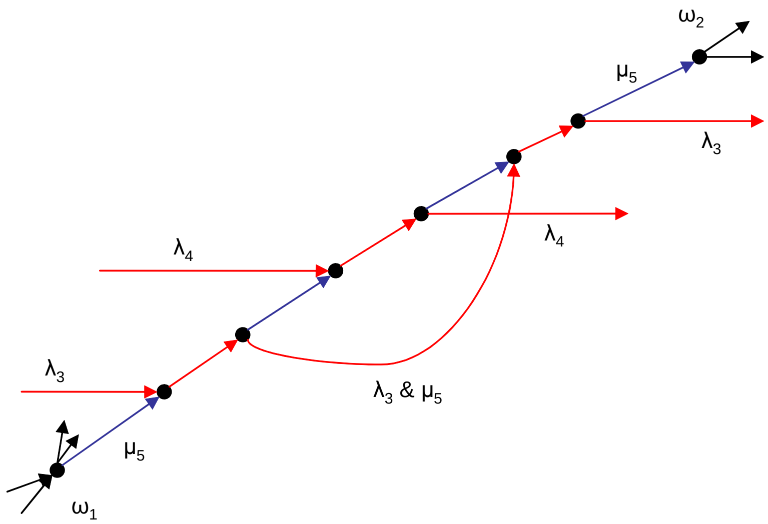

FIG 12. *Illustration for non-Type C routes: $w_{4n} \prec w_{3m}$*

Hence each Type C route intersects with every Type B route at some point downstream of w_1 and upstream of w_2. Also each Type C route intersects with every Type A route (at w_2), with every Type D route (at w_1) and with every other Type C route (at both w_1 and w_2).

We now consider routes that are not of Type C. If the set of non-Type C routes is non-empty let λ_3, λ_4 be members of this set which do not intersect except at w_0 and w_∞. Let $\mu(w_1, w_2)$ be a subpath between w_1 and w_2.
From above both λ_3 and λ_4 must intersect with μ. Let λ_3 intersect with μ only at the positions $w_{31}, \ldots w_{3m}$, where $w_{31} \prec \cdots \prec w_{3m}$; and let λ_4 intersect with μ only at the positions $w_{41}, \ldots w_{4n}$, where $w_{41} \prec \cdots \prec w_{4n}$. Without loss of generality let $w_1 \preceq w_{31} \prec w_{41} \preceq w_2$, so that λ_3 could be a route of Type B or Type D, and λ_4 could be a route of Type A or Type B.
Suppose that $w_{4n} \prec w_{3m}$ (Figure 12). Consider the subpath $\mu_5(w_1, w_2)$ which coincides with μ from w_1 to w_{31} (if $w_{31} \neq w_1$), coincides with λ_3 from w_{31} to w_{3m}, and coincides with μ from w_{3m} to w_2. This subpath μ_5 does not intersect with the route λ_4. This is impossible since every route in $\mathcal{C}$ intersects with every $\mu(w_1, w_2)$ subpath.
Suppose therefore that $w_{3m} \prec w_{4n}$ (Figure 13). Consider the subpath

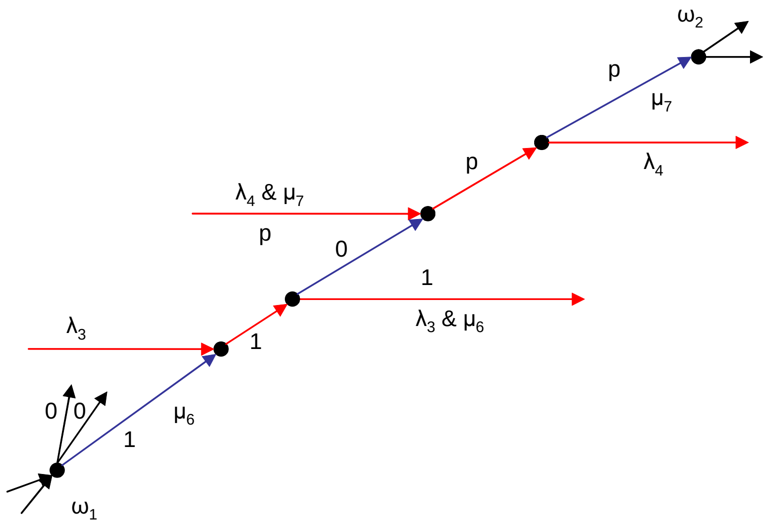

FIG 13. *Illustration for non-Type C routes: simplest case of* $w_{3m} \prec w_{4n}$

$\mu_6(w_1, w_\infty)$ which coincides with μ from w_1 to w_{31} (if $w_{31} \neq w_1$) and coincides with λ_3 from w_{31} to w_∞; and the subpath $\mu_7(w_0, w_2)$ which coincides with λ_4 from w_0 to w_{4n} and coincides with μ from w_{4n} to w_2 (if $w_{4n} \neq w_2$). Consider also a distribution P which (1) assigns a probability of 1 to every edge of μ_6, and (2) an arbitrary probability in $(0,1)$ to each edge of μ_7. If our SCEG is minimal and λ_3 and λ_4 do not intersect then this is always possible. Under P, assignment (1) gives us that

$$p(I(w_2) = 0 \mid I(w_1) = 1) = 1$$

and $I(w_1) \amalg I(w_2)$ implies that under this P

$$p(I(w_2) = 0 \mid I(w_1) = 0) = 1 \;\Rightarrow\; p(I(w_2) = 1 \mid I(w_1) = 0) = 0$$

But assignment (2) gives us that $p(I(w_2) = 1 \mid I(w_1) = 0) > 0$ ※

The assumption $I(w_1) \amalg I(w_2)$ is incompatible with the assignments of (1) and (2). But these assignments are always possible if λ_3 and λ_4 do not intersect. Hence λ_3 and λ_4 must intersect.

Hence each Type B route intersects with every Type A, Type B or Type D route, and each Type A route intersects with every Type D route. Also, each Type A route intersects with every other Type A route (at w_2), and each Type D route intersects with every other Type D route (at w_1). So each route in $\mathcal{C}$ intersects with every other route downstream of w_0 and upstream of w_∞.

Hence there cannot be two internally disjoint **directed** routes from w_0 to w_∞.

(c) Suppose that in the **undirected** version of the CEG $\mathcal{C}$ there are two internally disjoint paths between w_0 and w_∞. One of these necessarily corresponds to a representative directed $w_0 \to w_\infty$ route λ in $\mathcal{C}$. The other must correspond to a path (not a route) in $\mathcal{C}$ consisting of edges some of which meet head to head. In the simplest possible case this latter path will consist of a directed $w_0 \to w_A$ subpath $(\mu_\alpha(w_0, w_A))$, a directed $w_B \to w_\infty$ subpath $(\mu_\alpha(w_B, w_\infty))$, and a subpath joining w_A to w_B but directed $w_B \to w_A$, for some positions w_A and w_B.
In a CEG all positions lie on a directed $w_0 \to w_\infty$ route. So there must exist a directed subpath from w_0 to w_B $(\mu_\beta(w_0, w_B))$ and a directed subpath from w_A to w_∞ $(\mu_\beta(w_A, w_\infty))$.
Suppose these subpaths $\mu_\beta(w_0, w_B)$ and $\mu_\beta(w_A, w_\infty)$ intersect at a position w ($w_0 \prec w \prec w_B$, $w_A \prec w \prec w_\infty$). Then there exists a cycle in $\mathcal{C}$: $w \to w_B \to w_A \to w$. This is impossible since a CEG is a directed acyclic graph.
Suppose therefore that $\mu_\beta(w_0, w_B)$ and $\mu_\beta(w_A, w_\infty)$ do not intersect.
If the subpath $\mu_\beta(w_0, w_B)$ intersects with our original directed route λ but $\mu_\beta(w_A, w_\infty)$ does not, or if neither of these subpaths intersects with λ, then the directed route $(\mu_\alpha(w_0, w_A), \mu_\beta(w_A, w_\infty))$ is internally disjoint from λ.
If the subpath $\mu_\beta(w_A, w_\infty)$ intersects with λ but $\mu_\beta(w_0, w_B)$ does not, then the directed route $(\mu_\beta(w_0, w_B), \mu_\alpha(w_B, w_\infty))$ is internally disjoint from λ.
If both the subpaths $\mu_\beta(w_0, w_B)$ and $\mu_\beta(w_A, w_\infty)$ intersect with λ then the two routes $(\mu_\alpha(w_0, w_A), \mu_\beta(w_A, w_\infty))$ and $(\mu_\beta(w_0, w_B), \mu_\alpha(w_B, w_\infty))$ are internally disjoint.

So in this simplest possible case, if there exist two internally disjoint undirected paths between w_0 and w_∞ then there exist two internally disjoint directed routes. Clearly if we assume that there are more than two internally disjoint undirected paths between w_0 and w_∞ then this argument still holds. If we allow our second path to have more than one *reversed* section, we simply let w_A be the first position w on the path where edges meet head to head, and w_B the last position w on the path where both edges are directed away from w. Doing this our argument is then identical to that given above.

Hence there cannot be two internally disjoint routes in the undirected version of the CEG.

LEMMA. *An undirected graph $\mathcal{G}$ has no stalk between the vertices v and w if and only if there exist at least two internally disjoint paths between v and w.*

This is a corollary of Whitney's [32] Theorem 7, which is sometimes described as the *2nd variation of Menger's Theorem* [17]. A proof can be found in [11] where it appears as Theorem 7.4.

Hence there is a stalk lying downstream of w_0 and upstream of w_∞.

(d) Suppose there exists a stalk upstream of w_1. Then relabel this stalk as w_0 and repeat the argument of (b)(c) to show that there exists a stalk between this new w_0 and w_∞. Since the number of positions in $\mathcal{C}$ is finite, repeated use of this argument shows us that either w_1 is a stalk or there exists a stalk downstream of w_1. A complementary argument shows that there exists a stalk at w_2 or upstream of w_2.

(e) Suppose w_1 is a stalk. We know that $w_1 \prec w_2$ so there must exist a position exactly one edge downstream of w_1 which lies on a $w_1 \to w_2$ subpath. Call this position w_1^1. Then $w_1^1 \prec w_2$.
Now $I(w_1^1)$ is a function of $X(w_1)$ (because w_1 is a stalk): If $X(w_1)$ takes a value corresponding to an edge from w_1 to w_1^1 then $I(w_1^1) = 1$; otherwise $I(w_1^1) = 0$. So $X(w_1) \amalg X(w_2) \Rightarrow X(w_1) \amalg I(w_2) \Rightarrow I(w_1^1) \amalg I(w_2)$, and using the argument of (b)(c)(d) above there must be a stalk at w_1^1 or at w_2 or between them.

Therefore there exists a stalk downstream of w_1, either at or upstream of w_2.

□

PROOF OF LEMMA 3. Let $X(w_a) \amalg X(w_b)$ for some $w_a \in R_a, w_b \in R_b$. Then w_a is separated from w_b by a stalk (from Theorem 2). So since R_a, R_b are position cuts, each element of R_a is separated from each element of R_b by a stalk, and $X(w_a) \amalg X(w_b)$ for any pair of positions $w_a \in R_a, w_b \in R_b$. Hence $\boldsymbol{X}_{R_a} \amalg \boldsymbol{X}_{R_b}$. But $Y(R_a)(= \sup_{w \in R_a} X(w))$ is a function of $\boldsymbol{X}_{R_a}$, and hence $Y(R_a) \amalg Y(R_b)$.

Since our SCEG is minimal we can let the distribution P impose the probabilities

$$\begin{aligned}
p(X(w_a) = x_a) &= \alpha \quad \forall\, w_a \in R_a \\
p(X(w_b) = x_b) &= \beta \quad \forall\, w_b \in R_b \\
p(X(w_a) = x_a, X(w_b) = x_b) &= \gamma \quad \forall\, w_a \in R_a, w_b \in R_b.
\end{aligned}$$

for some **specified** $x_a, x_b > 0$ $(\alpha, \beta, \gamma > 0)$.

Now suppose that $X(w_a) \not\perp\!\!\!\perp X(w_b)$ for some $w_a \in R_a, w_b \in R_b$. Then there is **no** stalk between w_a and w_b, and hence **no** stalk between R_a and R_b. So by Theorem 2 $p(X(w_a) = x_a, X(w_b) = x_b)$ cannot equal $p(X(w_a) = x_a) \times p(X(w_b) = x_b)$ for all $w_a \in R_a, w_b \in R_b$. Hence $\gamma \neq \alpha\beta$.
Now for any $x_a, x_b > 0$ (greater than zero since R_a, R_b are position cuts)

$$\begin{aligned}
p(Y(R_a) = x_a, Y(R_b) = x_b) &= p(\sup_{w_a \in R_a} X(w_a) = x_a, \sup_{w_b \in R_b} X(w_b) = x_b) \\
&= p((X(w_{a1}) = x_a, X(w_{b1}) = x_b) \text{ or } (X(w_{a1}) = x_a, X(w_{b2}) = x_b) \\
&\quad \dots \text{ or } (X(w_{a2}) = x_a, X(w_{b1}) = x_b) \\
&\quad \dots \text{ or } X(w_{a|R_a|}) = x_a, X(w_{a|R_b|}) = x_b))
\end{aligned}$$

(noting that $X(w_{a1}) = x_a \Leftrightarrow X(w_{a1}) = x_a, X(w_{aj}) = 0$ for any $j \neq 1$)

$$\begin{aligned}
&= p(X(w_{a1}) = x_a, X(w_{b1}) = x_b) + p(X(w_{a1}) = x_a, X(w_{b2}) = x_b) \\
&\quad \dots + p(X(w_{a2}) = x_a, X(w_{b1}) = x_b) \\
&\quad \dots + p(X(w_{a|R_a|}) = x_a, X(w_{a|R_b|}) = x_b) \\
&= |R_a||R_b| \; \gamma
\end{aligned}$$

Similarly $p(Y(R_a) = x_a)\; p(Y(R_b) = x_b) = |R_a| \; \alpha \; |R_b| \; \beta$.
Now $\gamma \neq \alpha\beta \Rightarrow p(Y(R_a) = x_a)\; p(Y(R_b) = x_b) \neq p(Y(R_a) = x_a, Y(R_b) = x_b)$. So under this P, $X(w_a) \not\perp\!\!\!\perp X(w_b)$ (for some $w_a \in R_a, w_b \in R_b$) necessitates that $Y(R_a) \not\perp\!\!\!\perp Y(R_b)$. Hence if $X(w_a) \not\perp\!\!\!\perp X(w_b)$ then $Y(R_a) \not\perp\!\!\!\perp Y(R_b)$ in at least one distribution compatible with $\mathcal{C}$.

So if $Y(R_a) \perp\!\!\!\perp Y(R_b)$ holds for all distributions compatible with $\mathcal{C}$ then $X(w_a) \perp\!\!\!\perp X(w_b)$.

□

PROOF OF COROLLARY 1. Since the event Λ is intrinsic, $\mathcal{C}_\Lambda$ is a subgraph of $\mathcal{C}$ and $V(\mathcal{C}_\Lambda) \subset V(\mathcal{C})$. Let R_a in $\mathcal{C}_\Lambda$ be the set of $w_a \in V(\mathcal{C}_\Lambda)$ that are members of R_a in $\mathcal{C}$. Then $X(w_a)$ and R_a are well-defined on $\mathcal{C}_\Lambda$.
$Y(R_a)$ is measurable with respect to $\mathbb{F}(\mathcal{C})$ so it partitions the set of atoms of $\mathcal{C}$. Since $\Lambda \subset \Lambda(\mathcal{C})$ it also partitions the set of atoms of $\mathcal{C}_\Lambda$, and is well-defined on $\mathcal{C}_\Lambda$ as

$$Y(R_a) = \sup_{\substack{w_a \in R_a \\ w_a \in V(\mathcal{C}_\Lambda)}} X(w_a)$$

Hence $p_\Lambda(Y(R_a) = x_a) = p(Y(R_a) = x_a \mid \Lambda)$, and all necessary terms are defined on $\mathcal{C}_\Lambda$ consistently with their definitions on $\mathcal{C}$.

In $\mathcal{C}_\Lambda$, w_a and w_b are separated by a stalk, so by Theorem 2, $X(w_a) \amalg X(w_b)$ in $\mathcal{C}_\Lambda$, and by Lemma 3, $Y(R_a) \amalg Y(R_b)$ in $\mathcal{C}_\Lambda$, and by Lemma 1, $Y(R_a) \amalg Y(R_b) \mid \Lambda$ in $\mathcal{C}$.

$\square$

PROOF OF LEMMA 4.
(A) If $J(u) = 0$ then $Y(u) = 0$, so $Y(u) \amalg \boldsymbol{X}_{D^c(u)} \mid (J(u) = 0)$ (1)
(B) Let $J(u) = 1$. From section 3 we have $X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid I(w)$, so in particular, since $I(w) = 1$ implies both that $J(u) = 1$ for $w \in u$ and that $X(w') = 0$ for all $w' \in u,\ w' \neq w$

$$X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid (I(w) = 1) \Rightarrow\ X(w) \amalg \boldsymbol{X}_{D^c(w)} \mid (I(w) = 1, J(u) = 1)$$
$$\Rightarrow\ \boldsymbol{X}_u \amalg \boldsymbol{X}_{D^c(w)} \mid (I(w) = 1, J(u) = 1)$$

And since $Y(u)$ $(= \sup_{w' \in u} X(w'))$ is a function of $\boldsymbol{X}_u$, and $\boldsymbol{X}_{D^c(u)} \subset \boldsymbol{X}_{D^c(w)}$ for $w \in u$, this implies that

$$Y(u) \amalg \boldsymbol{X}_{D^c(w)} \mid (I(w) = 1, J(u) = 1)$$
$$\Rightarrow\ Y(u) \amalg \boldsymbol{X}_{D^c(u)} \mid (I(w) = 1, J(u) = 1)$$

Now suppose that $u = \{w_i\}_{i=1,\dots n}$. It follows that for $i = 1, \dots n$

$$Y(u) \amalg \boldsymbol{X}_{D^c(u)} \mid (I(w_i) = 1, J(u) = 1) \Rightarrow\ Y(u) \amalg \boldsymbol{X}_{D^c(u)} \mid (J(u) = 1) \quad (2)$$

Combining expressions (1) and (2) gives $Y(u) \amalg \boldsymbol{X}_{D^c(u)} \mid J(u)$.
But if we know $X(w')$ for all $w' \in u'$, then we know $Y(u')$; so $Y(u')$ is a function of the set $\{X(w')\}_{w' \in u'}$, and $\boldsymbol{Y}_{D^c(u)}$ is a function of $\boldsymbol{X}_{D^c(u)}$. Hence

$$Y(u) \amalg \boldsymbol{Y}_{D^c(u)} \mid J(u) \qquad \square$$

PROOF OF THEOREM 3. We first show that $\{Y(u_i)\}$ is d-separated from $\{Y(u_j)\}$ by $\{Y(u_k)\}$ in $\mathcal{A}(\mathcal{C})$ if and only if X_i is d-separated from X_j by X_k in $\mathcal{G}$.

Suppose X_i is d-separated from X_j by X_k in $\mathcal{G}$, but $\{Y(u_i)\}$ is **not** d-separated from $\{Y(u_j)\}$ by $\{Y(u_k)\}$ in $\mathcal{A}(\mathcal{C})$. Then there exists a path between $\{Y(u_i)\}$ and $\{Y(u_j)\}$ in the moralised ancestral version [14] of $\mathcal{A}(\mathcal{C})$ which does not pass through $\{Y(u_k)\}$.

Now in $\mathcal{A}(\mathcal{C})$ there exist edges from each $J(u)$ vertex to the corresponding $Y(u)$ vertex; and edges from $Y(u_a)$ vertices to $J(u_b)$ vertices only if there exists an edge from X_a to X_b in $\mathcal{G}$. When we produce the moralised ancestral version of $\mathcal{A}(\mathcal{C})$ we introduce two sorts of undirected edges – those between distinct vertices belonging to the **same** collection $\{Y(u_a)\}$, and those between $Y(u_a)$ and $Y(u_b)$ vertices belonging to different collections $\{Y(u_a)\}$, $\{Y(u_b)\}$, where $Y(u_a)$ and $Y(u_b)$ are both parents of a vertex $J(u_c)$. This latter only occurs when X_a and X_b are both parents of X_c in the moralised ancestral version of $\mathcal{G}$. Note that we introduce **no** undirected edges which connect $J(u)$ vertices, or connect $J(u)$ vertices to $Y(u)$ vertices.

So in the moralised ancestral version of $\mathcal{A}(\mathcal{C})$ there only exist undirected edges between different collections of vertices if there exist undirected edges between the corresponding variables in the moralised ancestral version of $\mathcal{G}$. Hence there cannot be a path between $\{Y(u_i)\}$ and $\{Y(u_j)\}$ in the moralised ancestral version of $\mathcal{A}(\mathcal{C})$ which does not pass through $\{Y(u_k)\}$.

Suppose instead that $\{Y(u_i)\}$ is d-separated from $\{Y(u_j)\}$ by $\{Y(u_k)\}$ in $\mathcal{A}(\mathcal{C})$, but that X_i is **not** d-separated from X_j by X_k in $\mathcal{G}$. Then there must exist **either** a moralising edge in the moralised ancestral version of $\mathcal{G}$ that has no corresponding edges in the moralised ancestral version of $\mathcal{A}(\mathcal{C})$ **or** a directed edge in $\mathcal{G}$ that has no corresponding edges in $\mathcal{A}(\mathcal{C})$.

The latter is impossible by construction – if there are no edges from $\{Y(u_a)\}$ to $\{J(u_b)\}$ then X_a is not a parent of X_b. In the former case this would mean that there existed variables X_a, X_b, both parents of X_c, such that there was no $J(u_c)$ vertex which was the child of both a $Y(u_a)$ vertex and a $Y(u_b)$ vertex.

But if the CEG is of a model which has a natural product space but which admits no context-specific conditional independence properties then each $J(u_c)$ vertex must have as parents both $Y(u_a)$ and $Y(u_b)$ vertices, since each $J(u_c)$ corresponds to a particular configuration of the parents of X_c, which include both X_a and X_b. So this also is impossible.

Using the above result and results from [31] the first statement in Theorem 3 holds, and also if $\{Y(u_i)\}$ is not d-separated from $\{Y(u_j)\}$ by $\{Y(u_k)\}$, then $X_i \not\perp\!\!\!\perp X_j \mid X_k$ in at least one distribution compatible with $\mathcal{C}$ and $\mathcal{G}$.

So if $X_i \perp\!\!\!\perp X_j \mid X_k$ holds for all distributions compatible with $\mathcal{C}$ and $\mathcal{G}$, then X_i is d-separated from X_j by X_k in $\mathcal{G}$, and from above $\{Y(u_i)\}$ is d-separated from $\{Y(u_j)\}$ by $\{Y(u_k)\}$ in $\mathcal{A}(\mathcal{C})$.

□

PROOF OF COROLLARY 5. If w_a, w_b are separated by a stalk then $X(w_a) \perp\!\!\!\perp X(w_b)$, simply by replacing *SCEG* by *RCEG* in part (1) of the proof of Theorem 2.

If w_a, w_b are separated by a stalk, then every $w_a \in R_a$ is separated from every $w_b \in R_b$ by a stalk, and hence $\boldsymbol{X}_{R_a} \perp\!\!\!\perp \boldsymbol{X}_{R_b}$. $Y(R_a)(= \sup_{w \in R_a} X(w))$ is a function of $\boldsymbol{X}_{R_a}$, and hence $Y(R_a) \perp\!\!\!\perp Y(R_b)$.

□

APPENDIX 2: A CAUTIONARY TALE

Suppose we have a CEG and an ACEG of a model which satisfies the conditions for Theorem 3. Suppose also that in the BN-representation of this model, A is a parent of both B and C, and B is a parent of C. Then $\{Y(u_C)\} \amalg \{J(u_B)\} \mid \{Y(u_B)\}$, since $J(u_B)$ is a function of $Y(u_B)$.

But in an ACEG of this model, $\{Y(u_C)\}$ is apparently not d-separated from $\{J(u_B)\}$ by $\{Y(u_B)\}$, since there are paths from $Y(u_C)$ vertices to $J(u_B)$ vertices which are not blocked by $\{Y(u_B)\}$ – see Figure 14.

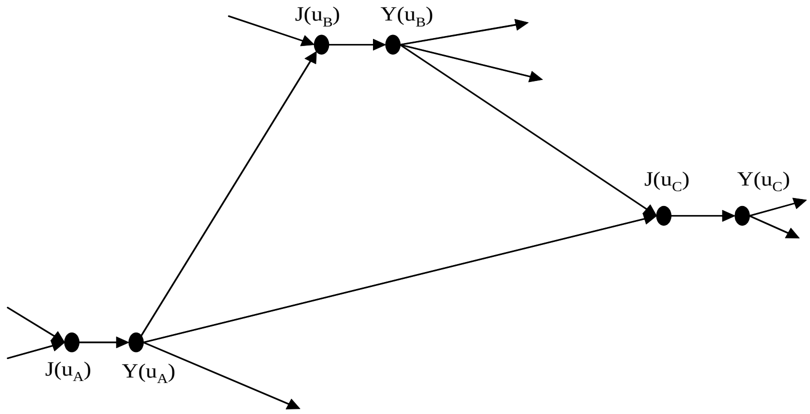

FIG 14. *ACEG for example in Appendix 2*

We use the word *apparently* here with justification. In [31] section 4, the authors briefly discuss D-separation (as opposed to d-separation) for graphs where there are functional (as opposed to stochastic) dependencies. An otherwise *active* path between two nodes is rendered *inactive* by a set of nodes Z under D-separation if a node on the path is *determined* by Z. Here each $J(u_B)$ is a function of its child $Y(u_B)$, so $\{Y(u_C)\}$ **is** D-separated from $\{J(u_B)\}$ in this example.

Note that in this paper we have, with one exception, just discussed d-separation expressions which involve only $Y(u)$-type vertices; between which there are **no** functional dependencies. The one exception is where we have considered expressions of the form $Y(u) \amalg \boldsymbol{Y}_{D^c(u)} \mid J(u)$. Here it is quite clear that $Y(u)$ is d-separated from the set of vertices associated with $\boldsymbol{Y}_{D^c(u)}$ by $J(u)$, since $J(u)$ is the sole parent of $Y(u)$, and $Y(u)$ must be d-separated from its non-descendants by its parents.

Acknowledgements. We would like to thank Jon Warren and Milan Studeny for their perceptive and helpful comments on the research presented in this paper.

REFERENCES

[1] E. S. Allman, C. Matias, and J. A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37:3099–3132, 2009.

[2] S. A. Andersson, D. Madigan, and M. D. Perlman. Alternative Markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28:33–85, 2001.

[3] R. R. Bouckaert and M. Studeny. Chain graphs: Semantics and expressiveness. In C. Froidevaux and J. Kohlas, editors, *Symbolic and Quantative approaches to Reasoning and Uncertainty*, number 946 in Lecture Notes in Artificial Intelligence, pages 67–76. Springer-Verlag, 1995.

[4] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in Bayesian Networks. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence*, pages 115–123, Portland, Oregon, 1996.

[5] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

[6] A. P. Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95:407–448, 2000.

[7] A. P. Dawid and M. Studeny. Conditional products: an alternative approach to conditional independence. In *Proceedings of the 7th Workshop on Artificial Intelligence and Statistics*, pages 32–40, 1999.

[8] G. Freeman and J. Q. Smith. Bayesian map model selection of Chain Event Graphs. Research Report 09–06, CRiSM, 2009.

[9] S. Hojsgaard and S. L. Lauritzen. Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society, Series B*, 70:1005–1027, 2008.

[10] R. A. Howard and J. E. Matheson. Influence diagrams. In R. A. Howard and J. E. Matheson, editors, *Readings in the Principles and Applications of Decision Analysis*. Strategic Decisions Group, 1984.

[11] L-H. Hsu and C-K. Lin. *Graph theory and Interconnection networks*. CRC Press, 2009.

[12] O. Kallenberg. *Foundations of Modern Probability*. Springer, 1997.

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

[14] S. L. Lauritzen. *Graphical Models*. Oxford, 1996.

[15] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H. G. Leimer. Independence properties of directed Markov fields. *Networks*, 20:491–505, 1990.

[16] D. McAllester, M. Collins, and F. Periera. Case factor diagrams for structured probabilistic modeling. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pages 382–391, 2004.

[17] K. Menger. Zur allgemeinen Kurventheorie. *Fundamenta Mathematicae*, 10:95–115, 1927.

[18] D. M. Q. Mond, J. Q. Smith, and D. Van Straten. Stochastic factorisations, sandwiched simplices and the topology of the space of explanations. *Proceedings of the Royal Society of London*, A 459:2821–2845, 2003.

[19] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[20] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge, 2000.

[21] L. D. Phillips. Requisite decision modelling. *Journal of the Operational Research Society*, 33:303–311, 1982.

[22] D. Poole and N. L. Zhang. Exploiting contextual independence in probabilistic inference. *Journal of Artificial Intelligence Research*, 18:263–313, 2003.

[23] T. S. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30:962–1030, 2002.

[24] A. Salmeron, A. Cano, and S. Moral. Importance sampling in Bayesian Networks using probability trees. *Computational Statistics and Data Analysis*, 34:387–413, 2000.

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

[26] J. Q. Smith. *Bayesian Decision analysis: Principles and Practice*. Cambrifge, 2010.

[27] J. Q. Smith and P. E. Anderson. Conditional independence and Chain Event Graphs. *Artificial Intelligence*, 172:42–68, 2008.

[28] M. Studeny. *Probabilistic Conditional Independence Structures*. Springer, 2005.

[29] P. A. Thwaites, J. Q. Smith, and R. G. Cowell. Propagation using Chain Event Graphs. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, pages 546–553, Helsinki, 2008.

[30] P. A. Thwaites, J. Q. Smith, and E. M. Riccomagno. Causal analysis with Chain Event Graphs. *Artificial Intelligence*, 174:889–909, 2010.

[31] T. Verma and J. Pearl. Causal networks: semantics and expressiveness. In *Proceedings of the 4th Conference on Uncertainty in Artificial Intelligence*, pages 352–359, 1988.

[32] H. Whitney. Congruent Graphs and the Connectivity of Graphs. *American Journal of Mathematics*, 54(1):150–168, 1932.

Department of Statistics
University of Warwick
Coventry CV4 7AL
United Kingdom
E-mail: Peter.Thwaites@warwick.ac.uk
J.Q.Smith@warwick.ac.uk