

# Constructing a Chain Event Graph from a Staged Tree

Aditi Shenvi<sup>1,2</sup>

A.SHENVI@WARWICK.AC.UK

Jim Q. Smith<sup>1,2</sup>

J.Q.SMITH@WARWICK.AC.UK

<sup>1</sup> *University of Warwick, Coventry, United Kingdom*

<sup>2</sup> *The Alan Turing Institute, London, United Kingdom*

## Abstract

Chain Event Graphs (CEGs) are a recent family of probabilistic graphical models - a generalisation of Bayesian Networks - providing an explicit representation of structural zeros, structural missing values and context-specific conditional independences within their graph topology. A CEG is constructed from an event tree through a sequence of transformations beginning with the colouring of the vertices of the event tree to identify one-step transition symmetries. This coloured event tree, also known as a staged tree, is the output of the learning algorithms used for this family. Surprisingly, no general algorithm has yet been devised that automatically transforms any staged tree into a CEG representation. In this paper we provide a simple iterative backward algorithm for this transformation. Additionally, we show that no information is lost from transforming a staged tree into a CEG. Finally, we demonstrate that with an optimal stopping criterion, our algorithm is more efficient than the generalisation of a special case presented in Silander and Leong (2013). We also provide Python code using this algorithm to obtain a CEG from any staged tree along with the functionality to add edges with sampling zeros.

**Keywords:** Chain event graphs; event trees; context-specific independence; structural zeroes; structural missing values; directed graphical models; conditional independence.

## 1. Introduction

Many real-world processes contain non-symmetric sample space structures. Examples of such processes can be frequently found in public health, medicine, risk analysis and policing (see Collazo et al. (2018)). Such asymmetries may arise due to the existence of structural zeros and structural missing values (collectively referred here as structural asymmetries) in the sample space of a variable conditional on the realisation of other variable(s). A structural zero refers to observing zero frequencies for a count variable or a category of a categorical variable when a non-zero observation is a logical impossibility rather than a sampling limitation (e.g. days or amount as low, medium, high of alcohol consumption by teetotallers). Structural missing value are observations which are missing as they are not defined for a subset of the individuals/units (e.g. variables relating to post-operative health of individuals who had the illness but weren't operated). It is easy to see how such asymmetries may give rise to context-specific conditional independences which are independence relationships of the form  $X \perp\!\!\!\perp Y|Z = z_1$  but  $X \not\perp\!\!\!\perp Y|Z = z_2$  where  $\perp\!\!\!\perp$  stands for probabilistic independence and the vertical bar shows conditioning variables on the right. In fact, context-specific independences regularly arise naturally in many applications (Zhang and Poole, 1999).

Graphical models such as Bayesian Networks (BNs) are unable to fully describe asymmetric processes. They are primarily stymied in this respect as they force the process description on a set of variables that are defined *a priori*. Indeed, in order to scale up BN methodologies to large problems, good BN software contain functions that copy parts of one conditional probability table

to another. Thus BNs implicitly embed context-specific independences through probability assignments within their conditional probability tables. However, this structural information is never explicitly represented in their topologies. Uncovering these independences requires serious modifications (typically involving trees in some form) to their standard representation and/or inferential process (Boutilier et al., 1996; Zhang and Poole, 1999; Jabbari et al., 2018). Additionally, structural zeros too are hidden away in their conditional probability tables.

Chain Event Graphs (CEGs) are a family of probabilistic graphical models whose graphical representation make structural asymmetries and context-specific conditional independences explicit (Collazo et al., 2018). CEGs contain the class of finite discrete BNs as a special case (Smith and Anderson, 2008). They are constructed from *event trees* which provide a natural and intuitive framework for describing the unfolding of a process through a sequence of *events*. Although the size of an event tree increases linearly with the number of events involved in the evolution of the process which may become unwieldy for large complex processes, they are nonetheless easy for the statistician to transparently elicit from the natural language descriptions of a domain expert. Embedding structural asymmetries within an event tree is a matter of simply not drawing the corresponding branch in the tree (Shenvi et al., 2018). However, a more compact representation of an event tree while retaining its properties and transparency is desirable. A CEG provides such a compact representation. Hence it is a powerful modelling tool for processes exhibiting significant asymmetries, particularly in domains such as medicine (Barclay et al., 2013), public health (Shenvi et al., 2018), forensic science (Collazo et al., 2018) where experts often offer event based descriptions of processes.

To obtain a CEG, an event tree is first transformed into a *staged tree* by colouring its vertices to represent symmetries within its structure. The vertices of the staged tree are then merged to provide a more concise representation of these symmetries in the form of the graph of a CEG. Such a transformation results in a much simpler graph often with an order of magnitude fewer vertices and edges than the generating tree. Like an event tree, a CEG also describes a process through a sequence of events and thus inherits the ability to graphically represent structural asymmetries. A CEG representation is especially useful because various implicit conditional independences, including of the context-specific nature, hidden within the patterns of colouring of the tree can be read directly from its topology using sets of events called *cuts* and *fine cuts* (Smith and Anderson, 2008).

Several fast learning algorithms now exist for the CEG (Freeman and Smith, 2011; Silander and Leong, 2013; Cowell and Smith, 2014). The output of these algorithms is a staged tree. A staged tree typically must go through a sequence of non-trivial transformations before it represents the graph of a CEG. In fact, a CEG is uniquely defined by its staged tree, and we show that the staged tree can be recovered from the graph of the CEG alone.

Silander and Leong (2013) present an algorithm to transform a *stratified* staged tree into a *stratified* CEG (SCEG). A stratified staged tree/SCEG is one in which events broadly corresponding to the same variable are at the same distance from a leaf/ the sink. Intuitively this corresponds to there being no events which become redundant conditional on the past events that have occurred. SCEGs have been studied extensively as any process that can be represented by a finite discrete BN can also be represented within this wider class. In particular, the advantages of the CEG over a BN can be demonstrated (Barclay et al., 2013). However, we are increasingly finding many applications where the CEG representation is not stratified (Shenvi et al., 2018; Shenvi and Smith, 2019). So it is timely that automatic algorithms are available to make this transformation for *any* staged tree.

The contribution of our paper is threefold. First we provide an algorithm that can transform any staged tree into a CEG and provide an optimal stopping criterion for this algorithm. Secondly, we

prove that the transformation of a staged tree into a CEG does not lead to the loss of any information. Lastly, we provide Python code (<https://github.com/ashenvi10/Chain-Event-Graphs>) that obtains a staged tree using an Agglomerative Hierarchical Clustering (AHC) algorithm and then transforms it into a CEG using our algorithm. Unlike the existing ‘ceg’ (Collazo and Taranti, 2017) and ‘stagedtrees’ (Carli et al., 2020) R packages, our code is not restricted to SCEGs and it also allows manual addition of edges with sampling zeros.

In Section 2 we review the notation and the preliminary concepts. In Section 3 we present a simple recursive backward algorithm - coded within supporting software - that can construct a CEG from any staged tree. Here we also prove some properties of the algorithm and of the transformation itself. In Section 4 we compare an adapted version of the algorithm presented in Silander and Leong (2013) to our algorithm. We conclude the paper with a short discussion in Section 5.

## 2. Notation and Preliminaries

A CEG construction begins by eliciting an event tree of a process either from a domain expert or from existing literature. Alternatively, it can be constructed directly from data. Below we outline the transformations an event tree goes through to obtain the graph of a CEG:

- Vertices in the event tree whose one step ahead evolutions, i.e. conditional transition probabilities, are equivalent are assigned the same colour to indicate this symmetry;
- Vertices whose rooted subtrees (the subtree formed by considering that vertex as the root) are isomorphic - in the structure and colour preserving sense - are merged into a single vertex which retains the colouring of its merged vertices;
- All the leaves of the tree are merged into a single vertex called the sink.

**Example 1** *Here we consider a simplified topical example. The staged tree in Figure 1 shows a hypothesised example of testing for a certain disease available to individuals exhibiting symptoms in three different settings: hospitals, care homes and in the general community. For simplicity, we assume here that the test is 100% sensitive and specific, and that we are only interested in the outcomes related to the disease. By “recovery\*” we collectively refer to those who recover and those who never had the disease. We further assume that death can only be caused by the disease in the time period considered. The coloured vertices represent equivalence of their conditional transition probabilities. For instance, the probability of dying is the same for individuals in hospitals and care homes who exhibit symptoms but do not get a test. The CEG for this staged tree is shown in Figure 2. It is not hard to see how this tree can be refined to be more realistic.*

Let  $\mathcal{T}$  denote an event tree with a finite vertex set  $V(\mathcal{T})$  and an edge set  $E(\mathcal{T})$ . An edge  $e \in E(\mathcal{T})$  from vertex  $v$  to vertex  $v'$  with edge label  $l$  is an ordered triple given by  $(v, v', l)$ . Denote by  $L(\mathcal{T})$  the set of leaves in  $\mathcal{T}$ . The non-leaf vertices in  $\mathcal{T}$  are called *situations* and their set is denoted by  $S(\mathcal{T}) = V(\mathcal{T}) \setminus L(\mathcal{T})$ . The set of children of a vertex  $v$  are denoted by  $\text{ch}(v)$ . Let  $\Phi_{\mathcal{T}} = \{\theta_v | v \in V(\mathcal{T})\}$  where  $\theta_v = (\theta(e) | e = (v, v', l) \in E(\mathcal{T}), v' \in V(\mathcal{T}))$  denotes the parameters for each vertex  $v \in V(\mathcal{T})$ .

Two situations  $v$  and  $v'$  are said to be in the same *stage* whenever  $\theta_v = \theta_{v'}$  and if  $\theta(e) = \theta(e')$  then  $e = (v, \cdot, l)$  and  $e' = (v', \cdot, l)$  for edge  $e$  emanating from  $v$  and  $e'$  emanating from  $v'$ . The latter condition states that the edges emanating from situations in the same stage which have the same estimated conditional transition probability must also share the same edge label. Note that

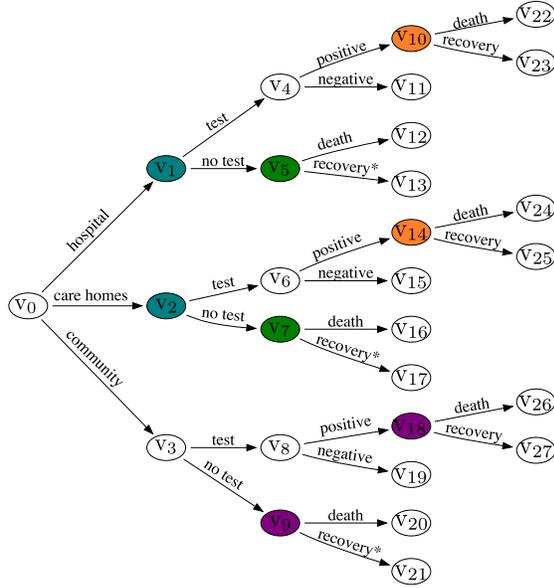


Figure 1: Staged tree for Example 1

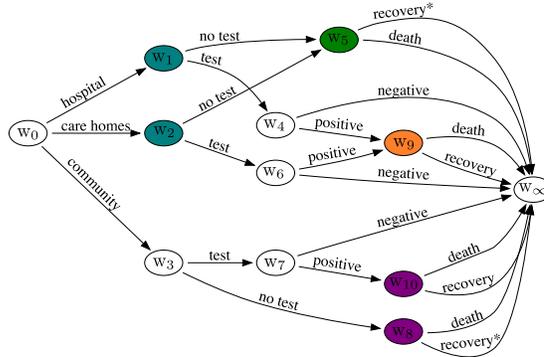


Figure 2: CEG for Example 1

when edge labels are not fixed, this condition is relaxed. In this case, edges of vertices in the same stage are coloured to represent which edges share the same conditional transition probabilities. This allows the statistician and domain expert to retrospectively assign labels to events which have the same meaning but which could have initially been assigned different labels.

**Example 1** (Continued) The domain expert may decide that the edge labels “recovery” and “recovery\*” can be treated as equivalent. Then situations  $v_9$  and  $v_{18}$  would be in the same position.

The collection of stages  $\mathbb{U}$  partitions  $V(\mathcal{T})$ . Each stage  $u \in \mathbb{U}$  is a set of situations in  $V(\mathcal{T})$  that belong to the stage  $u$ . Stage memberships are represented by colouring the situations of  $\mathcal{T}$  such that each stage  $u \in \mathbb{U}$  is represented by a unique colour. An event tree whose situations are coloured according to their stage memberships is called a *staged tree* and is denoted by  $\mathcal{S}^1$ . Situations in the staged tree whose rooted subtrees are isomorphic<sup>2</sup> have equivalent sets of parameters. That is,

1. Note that for simplicity, like in Figures 1 and 2, the colouring of the trivial stages may be suppressed.  
 2. In this paper isomorphism is in a structure and colouring preserving sense.

for two isomorphic subtrees  $\mathcal{S}'$  and  $\mathcal{S}''$  rooted at  $v$  and  $v'$ ,  $\Theta_{\mathcal{S}'} = \Theta_{\mathcal{S}''}$ . In a non-technical sense, this implies that  $v$  and  $v'$  have identical future evolutions. Situations whose rooted subtrees are isomorphic belong to the same *position*. The collection of positions  $\mathbb{W}$  is a finer partition of  $V(\mathcal{T})$  and each position  $w \in \mathbb{W}$  is a set of situations of  $V(\mathcal{T})$  that belong to the position  $w$ . Merging the situations in  $\mathcal{S}$  which are in the same position and collecting all the leaves in  $L(\mathcal{S})$  into a sink node denoted by  $w_\infty$  result in the graph of a CEG  $C$  for the process being modelled. Thus a CEG is uniquely defined by its staged tree, or in other words, it is uniquely defined by the pair  $(\mathcal{T}, \mathbb{U})$  where  $\mathcal{T}$  is its underlying event tree and  $\mathbb{U}$  is the set of stages.

**Definition 1 (Chain Event Graph)** A Chain Event Graph (CEG)  $C = (V(C), E(C))$  of a process represented by a staged tree  $\mathcal{S} = (V(\mathcal{S}), E(\mathcal{S}))$  with set of parameters  $\Theta_{\mathcal{S}}$  is a directed acyclic graph with  $V(C) = W \cup w_\infty$  where  $W$  is a set constructed by choosing a representative situation from each set in the collection<sup>3</sup>  $\mathbb{W}$ . The edges in  $C$  are constructed as follows: For a  $w \in W$ , create an edge  $(w, w', l) \in E(C)$  for every edge  $(w, s', l) \in E(\mathcal{S})$ , with  $\theta_w^C = \theta_w^{\mathcal{S}}$  where  $s'$  belongs to a set in  $\mathbb{W}$  which is represented by  $w'$  in  $W$ . Additionally,  $w \in V(C)$  retains the colouring of  $w \in V(\mathcal{S})$ .

A *floret* of a vertex  $v$  in a directed graph is denoted by  $F(v) = (V(F(v)), E(F(v)))$  where  $V(F(v)) = \{v \cup \text{ch}(v)\}$  and  $E(F(v))$  is the set of edges induced by  $V(F(v))$  in the graph. Denote the set of root-to-sink (root-to-leaf) paths in a CEG  $C$  (event tree  $\mathcal{T}$  /staged tree  $\mathcal{S}$ ) by  $C_\Lambda$  ( $\mathcal{T}_\Lambda$  /  $\mathcal{S}_\Lambda$ ) where a path is a sequence of tuples of the form ('vertex colour', 'edge label') from the root vertex to the sink following the directed edges. Say that an event tree, staged tree or a CEG is *stratified* whenever the vertices representing the same type of event (e.g. severity of illness) have the same number of edges between them and the root vertex along any path connecting them, and otherwise say it is *non-stratified*. Non-stratified CEGs provide a more realistic representation of a wide range of processes containing structural asymmetries (see e.g. Shenvi et al. (2018); Shenvi and Smith (2019)).

## 2.1 Why not just Staged Trees?

Staged trees are a graphical representation of a parametric statistical model and encapsulate within their colouring conditional independence information about the events describing a process (Görge and Smith, 2016, 2018). So why do we need CEGs when staged trees are themselves powerful tools?

While we show that staged tree and CEG representations are equivalent, the graph of a CEG is simpler and more compact. Typically, a CEG contains far fewer vertices and edges than its corresponding staged tree. Let  $V_k \subseteq V(\mathcal{T})$  denote the vertices of an event tree  $\mathcal{T}$  with  $k$  outgoing edges and let  $n_k = |V_k|$ . Then  $\mathcal{T}$  has  $n(\mathcal{T}) = \sum_{k=0}^d n_k$  vertices and  $k(\mathcal{T}) = \sum_{k=1}^d k n_k$  edges where  $d = \max\{k : n_k \geq 0\}$ . When a CEG  $C$  partitions  $V_k$  into  $1 \leq m_k \leq n_k$  positions, it is trivial to check that it has  $n(C) = \sum_{k=1}^d m_k + 1$  vertices (including the sink) and  $k(C) = \sum_{k=1}^d k m_k$  edges. So we have

$$\begin{aligned} n - 1 &\leq n(\mathcal{T}) - n(C) = \sum_{k=1}^d (n_k - m_k) + n - 1 \leq \sum_{k=0}^d (n_k - 1), \\ 0 &\leq k(\mathcal{T}) - k(C) = \sum_{k=1}^d k (n_k - m_k) \leq \sum_{k=1}^d k (n_k - 1), \end{aligned}$$

where  $n = |L(\mathcal{T})|$ . Let  $m = \max\{|\lambda| : \lambda \in \mathcal{T}_\Lambda\}$  be the length (i.e. number of tuples) of the longest root-to-leaf path of  $\mathcal{T}$ . It is easy to check that  $n(\mathcal{T})$  and  $k(\mathcal{T})$  typically increase as a power of  $m$ , whilst when its CEG expresses many symmetries  $n(C)$  and  $k(C)$  increase linearly in  $m$ . In fact, for dynamic processes, the staged tree is infinite but the corresponding CEG might be finite (Shenvi

3. Notice that  $\mathbb{U}$  and  $\mathbb{W}$  are sets of sets and to disambiguate, we refer to sets of sets as collections in this paper.

and Smith, 2019). Crucially, while there now exists a d-separation theorem (to be reported soon) for CEGs, such methodologies are yet to be developed for staged trees. Note that there is an interesting framework called conditional independence trees (Su and Zhang, 2005; Zhang and Su, 2004) which decompose decision trees into subtrees by exploiting the conditional independence relationships (including those of the context-specific nature) exhibited by the process. However, these are not yet fully developed and have been primarily used for improving prediction on classification problems.

### 3. A Recursive Algorithm to Construct a CEG

We present a simple recursive backward algorithm for constructing the graph of a CEG  $C$  from any staged tree  $\mathcal{S}$  irrespective of whether it is stratified. While a variety of model selection techniques exist for the CEG family (Freeman and Smith, 2011; Silander and Leong, 2013; Cowell and Smith, 2014), we do not discuss these in this paper. The outcome of any model selection algorithm for a CEG is a collection of stages  $\mathbb{U}$  for its underlying event tree. The vertices of the event tree can be coloured according to  $\mathbb{U}$ ; giving us the associated staged tree. Here we assume that we are only given the staged tree - obtained either as an output of a model selection algorithm or elicited by domain experts - from which we can deduce the collection of stages  $\mathbb{U}$ . The collection  $\mathbb{U}$  and the topology of the staged tree are then used to iteratively identify the collection of positions.

The recursion progressively melds situations together according to the position structure incrementally more distant from the leaves of the staged tree  $\mathcal{S}$ . This produces a sequence of coloured graphs  $\mathcal{G}_0 = \mathcal{S}, \mathcal{G}_1, \dots, \mathcal{G}_m = C$  where  $m$  is the depth of  $\mathcal{S}$ . Each graph in the sequence has the same root-to-leaf/sink paths, that is  $\mathcal{G}_{0\Lambda} = \mathcal{G}_{1\Lambda} = \dots = \mathcal{G}_{m\Lambda}$ , and the following relationship holds

$$|V(\mathcal{G}_i)| \geq |V(\mathcal{G}_{i+1})|, \quad |E(\mathcal{G}_i)| \geq |E(\mathcal{G}_{i+1})|; \quad i = 0, 1, \dots, m-1.$$

We specify our construction by writing the vertex and edge sets of each graph  $\mathcal{G}_i$  as a function of the vertex and edge sets of the graph  $\mathcal{G}_{i-1}$ . Note that the vertices in  $\mathcal{G}_i$  retain their colouring from the graph  $\mathcal{G}_{i-1}$ . Henceforth, we will say  $\mathcal{G}_i = \mathcal{G}_j$ ,  $i \neq j$ , whenever the two graphs  $\mathcal{G}_i$  and  $\mathcal{G}_j$  are isomorphic. Say that a vertex  $v$  is at a distance  $k$  from the sink vertex  $w_\infty$  (or equivalently, a leaf in a tree) if the shortest directed path from  $v$  to the sink (or a leaf) contains  $k$  tuples. Let  $V^{-k}$  be the set of vertices in a given graph such that every  $v \in V^{-k}$  is at a distance of  $k$  from the sink vertex  $w_\infty$  (or a leaf) of the graph. We describe our iterative algorithm below.

**Step 1: Initialisation.** From  $\mathcal{G}_0 = \mathcal{S}$  where  $\mathcal{S}$  is the staged tree, define the following:

$$\begin{aligned} v_1^- &\triangleq L(\mathcal{G}_0), & v_1^+ &\triangleq \{w_\infty\}, \\ \epsilon_1^- &\triangleq \{e \in E(\mathcal{G}_0) : e = (v, v', l) \text{ where } v \in S(\mathcal{G}_0), v' \in L(\mathcal{G}_0)\}, \\ \epsilon_1^+ &\triangleq \{\sigma_1(e) : e \in \epsilon_1^-\}, \end{aligned}$$

where  $\sigma_1(e) = \sigma_1(v, v', l) \triangleq (v, w_\infty, l)$ . Graph  $\mathcal{G}_1 = (V(\mathcal{G}_1), E(\mathcal{G}_1))$  where

$$V(\mathcal{G}_1) \triangleq V(\mathcal{G}_0) \setminus v_1^- \cup v_1^+, \quad E(\mathcal{G}_1) \triangleq E(\mathcal{G}_0) \setminus \epsilon_1^- \cup \epsilon_1^+.$$

**Step 2: Generalisation.** To construct graph  $\mathcal{G}_i$  from  $\mathcal{G}_{i-1}$ ,  $i \leq m$ , proceed as follows:

1. Create a sub-collection  $U_i = \{u_{1i}, \dots, u_{m_i}\}$  informed by the collection of stages  $\mathbb{U}$  such that each situation  $v \in V^{-(i-1)}$  belongs to only one set  $u_{ji} \in U_i$  for some  $j = 1, \dots, m_i$ , and two situations  $v, v' \in V^{-(i-1)}$  belong to the same set  $u_{ji}$  if and only if there exists a stage  $u \in \mathbb{U}$  such that  $v, v' \in u$ . Thus, the collection  $U_i$  gives us the stage structure for the vertices in  $V^{-(i-1)}$ .

2. Construct a collection  $U_i^*$  such that each  $u_{ji} \in U_i$  is replaced in  $U_i^*$  by the sets  $u_{ji}^1, \dots, u_{ji}^{n_{ji}}, n_{ji} \geq 1$ . Each situation  $v \in u_{ji}$  belongs to only one set  $u_{ji}^k$  for some  $k = 1, \dots, n_{ji}$ , and two situations  $v, v' \in u_{ji}$  belong to the same set  $u_{ji}^k$  if and only if there exists an edge  $(v', v'', l) \in E(\mathcal{G}_{i-1})$  for every edge  $(v, v'', l) \in E(\mathcal{G}_{i-1})$ . Thus, we have that  $u_{ji}^k \cap u_{ji}^l = \emptyset, k \neq l, \cup_k u_{ji}^k = u_{ji}$ , and  $U_i^* = \cup_j \cup_k u_{ji}^k$ . The collection  $U_i^*$  partitions the situations in  $V^{-(i-1)}$  into positions.
3. Define the following terms for each  $u_{ji}^k, j = 1, \dots, m_i, k = 1, \dots, n_{ji}$ ,

$$v^-(u_{ji}^k) \triangleq u_{ji}^k, \quad v^+(u_{ji}^k) \triangleq \{v\} \text{ for some } v \in v^-(u_{ji}^k).$$

We now define the following terms to enable us to construct the vertex and edge sets of  $\mathcal{G}_i$ ,

$$\begin{aligned} v_i^- &\triangleq \cup_j \cup_k v^-(u_{ji}^k), & v_i^+ &\triangleq \cup_j \cup_k v^+(u_{ji}^k), \\ \epsilon_i^f &\triangleq \{e \in E(\mathcal{G}_{i-1}) : e = (v, v', l) \text{ where } v \in v_i^- \setminus v_i^+\}, \\ \epsilon_i^b &\triangleq \{e \in E(\mathcal{G}_{i-1}) : e = (v, v', l) \text{ where } v' \in v_i^- \setminus v_i^+\}, \\ \epsilon_i^- &\triangleq \epsilon_i^f \cup \epsilon_i^b, & \epsilon_i^+ &\triangleq \{\sigma_i(e) : e \in \epsilon_i^b\}, \end{aligned}$$

where  $\sigma_i(e) = \sigma_i(v, v', l) \triangleq (v, v'', l)$  in which  $v'' \in v^+(u_{ji}^k)$  for  $v' \in v^-(u_{ji}^k), k = 1, \dots, n_{ji}$ . Setting  $V(\mathcal{G}_i) \triangleq V(\mathcal{G}_{i-1}) \setminus v_i^- \cup v_i^+$  and  $E(\mathcal{G}_i) \triangleq E(\mathcal{G}_{i-1}) \setminus \epsilon_i^- \cup \epsilon_i^+$  gives us the graph of  $\mathcal{G}_i$ .

We now prove that the above construction of  $U_i^*$  does in fact result in a collection of positions of the vertices in  $V^{-(i-1)}$ . The associated theorem is stated below with a proof in Appendix A.1.

**Theorem 1** *Given graph  $\mathcal{G}_{i-1}, i \leq m$  in the sequence of graphs transforming a staged tree  $\mathcal{G}_0 = \mathcal{S}$  to a CEG  $\mathcal{G}_m = \mathcal{C}$ , two situations  $v_1, v_2 \in V^{-(i-1)}$  are in the same position if and only if they belong to the same stage and for every  $(v_1, v', l)$  there exists a  $(v_2, v', l)$  in  $\mathcal{G}_{i-1}$ .*

We now show that the recursion may in fact be stopped for some  $0 < r < m$ . This optimal stopping point for the recursion is given in Theorem 2 with proof in Appendix A.2.

**Theorem 2 (Optimal stopping)** *In the sequence of graphs transforming a staged tree  $\mathcal{G}_0 = \mathcal{S}$  to a CEG  $\mathcal{G}_m = \mathcal{C}$  and  $m \geq 2$  where  $m$  is the depth of  $\mathcal{S}$ , the earliest stopping time in this transformation that guarantees the required CEG  $\mathcal{C}$  is the recursion step  $r$  such that  $\mathcal{G}_r = \mathcal{G}_{r-1} \neq \mathcal{G}_{r-2}, 0 < r < m$ .*

Theorem 3, with proof in Appendix A.3 implies that for every staged tree there is a unique CEG and also that the staged tree can be recovered given this CEG. This is equivalent to saying that no information is lost in transforming a staged tree into a CEG.

**Theorem 3 (Preservation of information)** *The mapping from a staged tree to a CEG is bijective.*

### 3.1 Related Work

Silander and Leong (2013) presented an algorithm to learn a stratified staged tree and to transform it into an SCEG (although the stratified terminology was not used). Their algorithm is a special case - albeit with no early stopping criterion - of the general algorithm we presented in Section 3.

They define the structure of a CEG for  $n$ -dimensional data as a “layered directed acyclic graph with  $n + 1$  layers”. They assumed that the vertices in layer  $k$  correspond to the same variable, say

$X_k$ . They also assume that from each vertex in layer  $k$ , there are exactly  $r_k$  emanating edges, all entering vertices in layer  $k + 1$ . The stratified staged tree to SCEG transformation algorithm states a weaker form of Theorem 1 without a proof and carries out a backward iteration from one layer to the previous one, all the way to the root, by merging situations which satisfy Theorem 1. However, it is easy to see that using their definition of layers, this algorithm fails for non-stratified CEGs where events don't necessarily satisfy a symmetric product space structure.

We adapt their algorithm so that layer  $k$  in their algorithm corresponds to what we defined as set  $V^{-k}$  in Section 3. The main differences between the adapted version of their algorithm and ours is that (1) we provide an optimal stopping criterion which saves on computational effort of searching the entire staged tree, (2) we provide all the necessary proofs for our algorithm. For convenience, call their adapted algorithm the *baseline algorithm* and ours the *optimal time algorithm*.

## 4. Experiments

Here we compare the performance of the baseline and optimal algorithms on 7 datasets. The first four datasets are from the UCI repository (Dua and Graff, 2019). The missing values were removed and sampling zeros were treated as structural. The fifth dataset is from the Christchurch Health and Development Study (CHDS) conducted at the University of Otago, New Zealand (see Fergusson et al. (1986)). The penultimate dataset is from Shenvi et al. (2018) and its asymmetric nature can be seen from the CEG in Figure 3 in that paper. The final dataset is an extension of this dataset and has been used in Shenvi and Smith (2019). The last two datasets have structural zeroes and so, they are not stratified and do not have symmetric product space structures. The remaining datasets are stratified. It has also been shown that the last three datasets exhibit context-specific conditional independences (Collazo et al., 2018; Shenvi et al., 2018; Shenvi and Smith, 2019).

These experiments were carried out using our Python code <sup>4</sup> on a 2.9 GHz MacBook Pro with 32GB memory. Our code can handle datasets with structural asymmetries (stored as NaNs or null values) and also provides the capability to manually add sampling zero paths to the tree. It is currently set up to learn the staged tree from the event tree of the dataset using the AHC algorithm.

Dataset	$ S(S) $	Depth $m$	$T_{\text{Baseline}}$	$ V(C_{\text{Baseline}}) $	$T_{\text{Optimal}}$	$ V(C_{\text{Optimal}}) $
Iris	52	5	1.635	42	1.414	42
Hayes-Roth	124	5	12.118	58	12.085	58
Balance scale	327	5	145.052	90	143.321	90
Glass	636	10	389.272	308	376.689	308
CHDS	19	4	0.586	10	0.556	10
Falls	39	6	1.564	27	1.453	27
Falls dynamic	346	5	585.789	242	550.990	242

Table 1: Comparison of the baseline algorithm and the optimal time algorithm.

Table 1 gives for each dataset the number of situations in the staged tree output by the AHC algorithm ( $|S(S)|$ ), the maximum depth of the staged tree ( $m$ ) and the time taken (in milliseconds) by the two compacting algorithms ( $T_{\text{Baseline}}$  and  $T_{\text{Optimal}}$ ) as well as the number of positions in the resulting CEG found by the two algorithms ( $|V(C_{\text{Baseline}})|$  and  $|V(C_{\text{Optimal}})|$ ). From this table we

4. <https://github.com/ashenvi10/Chain-Event-Graphs>

can see that the optimal time algorithm takes less time than the baseline algorithm while arriving at the same CEG as it stops as soon as Theorem 2 is satisfied. However, the gain in efficiencies are inversely proportional to the number of symmetries exhibited by the process (see Section 2.1). Thus, if there are more symmetries (more situations in non-trivial stages) across the tree, we need to search across more  $V^{-k}$  sets before we arrive at the CEG.

## 5. Discussion

We have provided a simple iterative backward algorithm along with supporting Python code to transform any staged tree into a CEG. Research in CEGs and their applications has been an increasingly active field in recent years. However, such a general algorithm and proofs of the validity of the staged tree to CEG transformation have been missing in the literature so far. We know through personal correspondence that, a soon to be published, d-separation theorem for CEGs has been developed. Construction of the minimal ancestral CEGs in this theorem follows the same procedure as our algorithm. Hence, automating this process, as we have done, is a very timely development.

## Acknowledgments

We would like to thank John Horwood and the CHDS research group for the CHDS dataset. We would also like to thank the reviewers whose insightful comments greatly improved the original version. AS was supported by the University of Warwick Chancellor's International Scholarship and the Alan Turing Institute. JQS was supported by the Alan Turing Institute and funded by the EPSRC [grant number EP/K03 9628/1].

## Appendix A. Proofs

### A.1 Proof for Theorem 1

We have a graph  $\mathcal{G}_{i-1}$  belonging to the sequence of graphs converting a staged tree  $\mathcal{S}$  into a CEG  $\mathcal{C}$ . This implies that all the vertices in  $V^{-j}$ ,  $j = 1, \dots, i-2$  in  $\mathcal{G}_{i-1}$  represent positions.

$\Rightarrow$  Given that two situations  $v_1, v_2 \in V^{-(i-1)}$  are in the same position. We show that (1)  $v_1$  and  $v_2$  belong to the same stage; (2) for every  $(v_1, v', l)$  there exists a  $(v_2, v', l)$  in  $\mathcal{G}_{i-1}$ .

If  $v_1$  and  $v_2$  are in the same position, it is trivially true that they are also in the same stage. Additionally, by the definition of a position, the subtrees rooted at  $v_1$  and  $v_2$ , call them  $\mathcal{S}_{v_1}$  and  $\mathcal{S}_{v_2}$  in the staged tree  $\mathcal{S}$  are isomorphic. Thus also, for every subtree rooted at a child of  $v_1$  in  $\mathcal{S}_{v_1}$ , there exists an isomorphic subtree rooted at a child of  $v_2$  in  $\mathcal{S}_{v_2}$ . In fact, stages by definition require that edges with the same estimated conditional transition probability must also have the same edge label. Therefore, there necessarily exists a situation  $v_2^{ch}$  along edge  $(v_2, v_2^{ch}, l)$  such that the subtree rooted at  $v_2^{ch}$  is isomorphic to the subtree rooted at situation  $v_1^{ch}$  which is along the edge  $(v_1, v_1^{ch}, l)$ . Notice that  $v_1^{ch}$  and  $v_2^{ch}$  belong to the set  $V^{-(i-2)}$  in  $\mathcal{G}_{i-2}$ . Since their rooted subtrees in  $\mathcal{S}$  are isomorphic, they belong to the same position and are represented by a single vertex, say  $v_{1,2}^{ch}$  in  $\mathcal{G}_{i-1}$ . The edges  $(v_1, v_1^{ch}, l)$  in  $\mathcal{S}_{v_1}$  and  $(v_2, v_2^{ch}, l)$  in  $\mathcal{S}_{v_2}$  are represented by edges  $(v_1, v_{1,2}^{ch}, l)$  and  $(v_2, v_{1,2}^{ch}, l)$  in  $\mathcal{G}_{i-1}$ . This result extends to every  $(v_1, v', l)$  in  $\mathcal{G}_{i-1}$ .

$\Leftarrow$  Given that  $v_1, v_2 \in V^{-(i-1)}$  in  $\mathcal{G}_{i-1}$  belong to the same stage and for every  $(v_1, v', l)$  there exists a  $(v_2, v', l)$  in  $\mathcal{G}_{i-1}$ . We need to show that  $v_1$  and  $v_2$  are in the same position.

Recall that two situations are in the same position when the subtrees rooted at these vertices in  $\mathcal{S}$  are isomorphic. Since  $v_1$  and  $v_2$  are in the same stage, they have the same number of emanating edges and also, the edges from  $v_1$  and  $v_2$  which share the same edge label have the same estimated conditional transition probability. Consider edges  $(v_1, v_{1,2}^{ch}, l)$  and  $(v_2, v_{1,2}^{ch}, l)$  emanating from situations  $v_1$  and  $v_2$  in  $\mathcal{G}_{i-1}$  respectively where  $v_{1,2}^{ch}$  is the common situation along these two edges. In a tree each vertex has at most one parent. So in the staged tree  $\mathcal{S}$ , the position  $v_{1,2}^{ch}$  would be represented by two separate vertices, call them  $v_1^{ch}$  and  $v_2^{ch}$  in the subtrees rooted at  $v_1$  and  $v_2$  respectively. Thus, the edge  $(v_1, v_{1,2}^{ch}, l)$  would be replaced by an edge  $(v_1, v_1^{ch}, l)$  in the subtree rooted at  $v_1$ , call this  $\mathcal{S}_{v_1}$  in  $\mathcal{S}$ . Similarly, the edge  $(v_2, v_{1,2}^{ch}, l)$  would be replaced by an edge  $(v_2, v_2^{ch}, l)$  in  $\mathcal{S}_{v_2}$  which is the subtree rooted at  $v_2$  in  $\mathcal{S}$ . Since  $v_1^{ch}$  and  $v_2^{ch}$  are in the same position in  $\mathcal{G}_{i-1}$ , they have isomorphic subtrees in  $\mathcal{S}_{v_1}$  and  $\mathcal{S}_{v_2}$ . Similarly, the subtrees rooted at the children of  $v_1$  and  $v_2$  in  $\mathcal{S}_{v_1}$  and  $\mathcal{S}_{v_2}$  respectively are isomorphic whenever the edges from  $v_1$  and  $v_2$  to their respective children share the same edge label. Since  $v_1$  and  $v_2$  are in the same stage, the florets  $F(v_1)$  in  $\mathcal{S}_{v_1}$  and  $F(v_2)$  in  $\mathcal{S}_{v_2}$  are also isomorphic. Thus  $\mathcal{S}_{v_1}$  and  $\mathcal{S}_{v_2}$  are isomorphic and hence, they belong to the same position.

## A.2 Proof for Theorem 2

Suppose that  $0 < r < m$  recursions have taken place and  $\mathcal{G}_r = \mathcal{G}_{r-1} \neq \mathcal{G}_{r-2}$ . We show that  $\mathcal{G}_r = C$ . As the graph of a CEG is the most parsimonious representation of the event tree describing a process, this is equivalent to showing that  $|V(\mathcal{G}_r)| = |\mathbb{W}| + 1$  where  $\mathbb{W}$  is the collection of positions. Graph  $\mathcal{G}_{r-1}$  contains the positions for all situations in  $V^{-k}$ ,  $0 \leq k < r - 1$ . Since  $\mathcal{G}_r = \mathcal{G}_{r-1}$ , the problem can be framed as showing that if there are no non-trivial positions in  $V^{-(r-1)}$  then there are no non-trivial positions in any of  $V^{-k}$ ,  $r \leq k \leq m$ . We prove this by contradiction.

Let there be no non-trivial positions in  $V^{-(r-1)}$ . Suppose that two situations  $v_1, v_2 \in V^{-r}$  are in the same position and hence, the same stage. This implies that the subtrees of  $\mathcal{S}$  rooted at  $v_1$  and  $v_2$ , say  $\mathcal{S}_{v_1}$  and  $\mathcal{S}_{v_2}$  respectively are isomorphic. Let  $v_1^{ch}$  be a child of  $v_1$  along the edge  $(v_1, v_1^{ch}, l)$  and let  $\mathcal{S}_{v_1^{ch}}$  be the subtree rooted at  $v_1^{ch}$ . By the definition of a stage, there exists an edge  $(v_2, v_2^{ch}, l)$  in  $\mathcal{S}_{v_2}$  with rooted subtree  $\mathcal{S}_{v_2^{ch}}$ . The subtrees  $\mathcal{S}_{v_1^{ch}}$  and  $\mathcal{S}_{v_2^{ch}}$  are isomorphic as  $\mathcal{S}_{v_1}$  and  $\mathcal{S}_{v_2}$  are isomorphic. By the definition of a position,  $v_1^{ch}$  and  $v_2^{ch}$  are in the same position. As  $v_1, v_2 \in V^{-r}$ , we have that  $v_1^{ch}, v_2^{ch} \in V^{-(r-1)}$ . This contradicts that there are no non-trivial positions in  $V^{-(r-1)}$ . A similar argument can be made for any  $v_1, v_2 \in V^{-k}$ ,  $r \leq k \leq m$ . Since  $\mathcal{G}_r = \mathcal{G}_{r-1}$ ,  $V^{-(r-1)}$  has no non-trivial positions and all the positions in  $V^{-k}$ ,  $0 \leq k < r$  have been identified. By the above result,  $V^{-k}$ ,  $r \leq k \leq m$  also do not contain any non-trivial positions. Thus  $\mathcal{G}_r = C$ .

We have that  $\mathcal{G}_{r-2} \neq \mathcal{G}_{r-1} = \mathcal{G}_r = \dots = \mathcal{G}_m = C$ . While stopping at graph  $\mathcal{G}_{r-1}$  gives us the required graph of the CEG, this recursive step is indistinguishable from any of the other  $k < r - 1$  steps. Hence, the isomorphism of  $\mathcal{G}_{r-1}$  and  $\mathcal{G}_r$  is needed to stop the recursions with certainty. Thus the earliest stopping point for the recursion is step  $r$  such that  $\mathcal{G}_r = \mathcal{G}_{r-1} \neq \mathcal{G}_{r-2}$ ,  $0 < r < m$ .

## A.3 Proof for Theorem 3

We prove bijection by proving injection and surjection.

**Injection:** We prove the injective contrapositive; that is, given staged trees  $\mathcal{S}_1 \neq \mathcal{S}_2$ , we show that their corresponding CEGs  $C_1$  and  $C_2$  are not isomorphic. It is straightforward to show that if  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are structurally not isomorphic, then  $C_1 \neq C_2$ . Suppose that  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are structurally isomorphic and that they differ only in the colouring of one of their vertices. Let these vertices be

$v_1$  with colour  $c_1$  in  $\mathcal{S}_1$  and  $v_2$  with colour  $c_2 \neq c_1$  in  $\mathcal{S}_2$ . Since vertices retain their colouring in the CEG, the positions representing  $v_1$  and  $v_2$  in  $C_1$  and  $C_2$  will be coloured by  $c_1$  and  $c_2$  respectively. Hence,  $C_1$  and  $C_2$  will not have colour preserving isomorphism. Additionally,  $C_1$  and  $C_2$  will not be structurally isomorphic if either or both of  $v_1$  and  $v_2$  create non-trivial positions in their respective staged trees as the collection of positions in  $\mathcal{S}_1$  and  $\mathcal{S}_2$  will not be equivalent.

**Surjection:** From a given CEG  $C$ , construct a staged tree  $\mathcal{S}$  as follows:

1. Sort the paths in  $C_\Lambda$  in ascending order of the length (number of tuples) of the paths.
2. For each path  $\{(c, l)\}$  of length 1 where  $c$  is a colour and  $l$  is an edge label, construct an edge from  $v_0$ , the root of  $\mathcal{S}$  to a new vertex (labelled as  $v_i$  where  $i$  is an integer index which hasn't been assigned thus far in the construction) and label it  $l$ . Assign colour  $c$  to  $v_0$ .
3. In general, for any path of length  $k$  given by  $\{(c_1, l_1), \dots, (c_{k-1}, l_{k-1}), (c_k, l_k)\}$ , there necessarily exists a path  $\{(c_1, l_1), \dots, (c_{k-1}, l_{k-1})\}$  ending in a vertex, say  $v_{k-1}$  in the staged tree constructed so far. To add the  $k$ th tuple  $(c_k, l_k)$  to this path, colour  $v_{k-1}$  by  $c_k$ , add a vertex  $v_k$  and construct a directed edge from  $v_{k-1}$  to  $v_k$  with edge label  $l_k$ .

This construction results in a tree as it is connected (no vertex - with the exception of the root - is added until it is connected by an edge to an existing vertex) and has no directed cycles (each edge is constructed from an existing vertex to a new vertex). Call this tree  $\mathcal{T}^*$ . We prove that  $\mathcal{T}^*$  is the unique staged tree whose transformation, as described in Section 3, results in our given CEG  $C$ .

Observe that a staged tree is uniquely and unambiguously defined by its underlying event tree and its collection of stages  $\mathbb{U}$ . The structure of any event tree can be recovered from its set of uncoloured root-to-leaf paths, which is equivalent to the uncoloured root-to-sink paths  $C_\Lambda$  of the CEG  $C$ . As  $\mathcal{T}^*$  is constructed from the set  $C_\Lambda$ , the uncoloured version of  $\mathcal{T}^*$  is the required underlying event tree for  $C$ . The vertices of  $\mathcal{T}^*$  inherit their colourings from the positions of  $C$ . Recall that colouring of positions in a CEG is indicative of stage memberships. Hence, two positions  $w$  and  $w'$  with the same colour, say  $c$  in  $C$  are in the same stage. By definition of a stage,  $\theta_w = \theta_{w'}$ , and for each edge  $e = (w, \cdot, l)$  there exists an  $e' = (w', \cdot, l)$  such that  $\theta(e) = \theta(e')$  in  $C$ . Two vertices  $v$  and  $v'$  with the colour  $c$  in  $\mathcal{T}^*$  either belong to the same position in  $C$  - without loss of generality assume this is  $w$  - or belong to two distinct positions in  $C$ , assume these are  $w$  and  $w'$ . If both  $v, v'$  belong to position  $w$ , then  $v$  and  $v'$  in  $\mathcal{T}^*$  are created from two separate root-to- $w$  subpaths, say  $p$  and  $p'$  in  $C$ . Floret  $F(v)$  is formed by creating  $k$  copies of subpath  $p$  and appending each with a distinct  $(c, l_i)$  where  $i = 1, \dots, k$  and  $l_i$  is the label of the  $i$ th edge emanating from  $w$  in  $C$ . Floret  $F(v')$  is constructed in a similar manner. Thus  $v$  and  $v'$  have the same number of emanating vertices in  $\mathcal{T}^*$  and share the same vertex colour as they satisfy the conditions of being in the same stage by belonging to the same position in  $C$ . This also holds when  $v$  and  $v'$  belong to positions  $w$  and  $w'$  respectively, where  $w$  and  $w'$  share the same colour in  $C$ , with the exception that  $p$  will be a root-to- $w$  subpath and  $p'$  a root-to- $w'$  subpath. Thus  $\mathcal{T}^*$  is the underlying staged tree of  $C$  as it has the structure of the event tree of  $C$  and a collection of stages equivalent to that of  $C$ .

## References

- L. M. Barclay, J. L. Hutton, and J. Q. Smith. Refining a Bayesian network using a chain event graph. *International Journal of Approximate Reasoning*, 54(9):1300–1309, 2013.
- 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 Intel-*

- ligence*, pages 115–123, 1996.
- F. Carli, M. Leonelli, E. Riccomagno, and G. Varando. The r package stagedtrees for structural learning of stratified staged trees. *arXiv preprint arXiv:2004.06459*, 2020.
- R. Collazo and P. Taranti. *ceg: Chain event graph*, 2017. URL <https://CRAN.R-project.org/package=ceg>. R package version 0.1.0.
- R. A. Collazo, C. Görden, and J. Q. Smith. *Chain event graphs*. CRC Press, 2018.
- R. G. Cowell and J. Q. Smith. Causal discovery through MAP selection of stratified chain event graphs. *Electronic Journal of Statistics*, 8(1):965–997, 2014.
- D. Dua and C. Graff. UCI machine learning repository, 2019. URL <http://archive.ics.uci.edu/ml>.
- D. Fergusson, L. Horwood, and F. Shannon. Social and family factors in childhood hospital admission. *Journal of Epidemiology & Community Health*, 40(1):50–58, 1986.
- G. Freeman and J. Q. Smith. Bayesian MAP model selection of chain event graphs. *Journal of Multivariate Analysis*, 102(7):1152–1165, 2011.
- C. Görden and J. Q. Smith. A differential approach to causality in staged trees. In *Conference on Probabilistic Graphical Models*, pages 207–215, 2016.
- C. Görden and J. Q. Smith. Equivalence classes of staged trees. *Bernoulli*, 24(4A):2676–2692, 2018.
- F. Jabbari, S. Visweswaran, and G. F. Cooper. Instance-specific Bayesian network structure learning. *Proceedings of machine learning research*, 72:169, 2018.
- A. Shenvi and J. Q. Smith. A Bayesian dynamic graphical model for recurrent events in public health. *arXiv preprint arXiv:1811.08872*, 2019.
- A. Shenvi, J. Q. Smith, R. Walton, and S. Eldridge. Modelling with non-stratified chain event graphs. In *International Conference on Bayesian Statistics in Action*, pages 155–163, 2018.
- T. Silander and T.-Y. Leong. A dynamic programming algorithm for learning chain event graphs. In *International Conference on Discovery Science*, pages 201–216. Springer, 2013.
- J. Q. Smith and P. E. Anderson. Conditional independence and chain event graphs. *Artificial Intelligence*, 172(1):42–68, 2008.
- J. Su and H. Zhang. Representing conditional independence using decision trees. In *AAAI*, pages 874–879, 2005.
- H. Zhang and J. Su. Conditional independence trees. In *European Conference on Machine Learning*, pages 513–524. Springer, 2004.
- N. L. Zhang and D. Poole. On the role of context-specific independence in probabilistic inference. In *Proceedings of the 16th International Joint Conference on Artificial intelligence*, volume 2, pages 1288–1293, 1999.