

# Modelling with non-stratified Chain Event Graphs

Aditi Shenvi, Jim Q. Smith, Robert Walton, Sandra Eldridge

**Abstract** Chain Event Graphs (CEGs) are a recent probabilistic graphical modelling tool that have proved successful in modelling scenarios with context-specific independencies. Although the theory underlying CEGs supports appropriate representation of structural zeroes, the literature so far does not provide an adaptation of the vanilla CEG methods for a real-world application presenting structural zeroes also known as the non-stratified CEG class. To illustrate these methods, we present a non-stratified CEG representing a public health intervention designed to reduce the risk and rate of falling in the elderly. We then compare the CEG model to the more conventional Bayesian Network model when applied to this setting.

**Keywords:** Chain Event Graphs; Event Tree; Public Health Intervention; Bayesian Networks; Bayesian Statistics

---

Aditi Shenvi  
Centre for Complexity Science, University of Warwick, Coventry, United Kingdom  
e-mail: a.shenvi@warwick.ac.uk

Jim Q. Smith  
Department of Statistics, University of Warwick, Coventry, United Kingdom  
e-mail: j.q.smith@warwick.ac.uk

Robert Walton  
Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom  
e-mail: rtwalton123@gmail.com

Sandra Eldridge  
Pragmatic Clinical Trials Unit, Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom  
e-mail: s.eldridge@qmul.ac.uk

## 1 Introduction

Bayesian Networks (BNs) are a widely used probabilistic graphical modelling tool in various domains. A BN  $X$  is represented by a directed acyclic graph (DAG)  $G = (V, E)$  where the lack of an edge between two nodes of  $G$  represents conditional independence between them. The directed edges encode information about the conditional dependencies between the random variables. A BN can be written as  $X = (X_v)_{v \in V}$  where  $X_v$  is a discrete random vector. A depth of resources are now available for learning and inferencing using BNs [11, 6]. A causal analysis can also be carried out wherever it is legitimate to assume a causal embedding of the model - a well developed technology where the topology of a fitted graph can indicate causal relationships [16]. Whilst the BN is a powerful modelling tool, it has several shortcomings. One is its inability to embed context-specific conditional independence relations directly into its graphical representation. In order to exploit such information from a BN, we need to make adjustments to the model using methods such as those proposed in [10, 2, 17]. However, a unified graphical model for encapsulating these context-specific details cannot be achieved by using these methods on the BN. The Chain Event Graph (CEG) was developed to deal with such asymmetries. More recently it has been shown that CEGs also share two properties enjoyed by BNs that the underlying framework can be elicited and explained using natural language [4, 20] and a causal algebra can be built around its topology [23, 22].

The event tree [18] for a real-world system may have some of its branches unpopulated. Zeroes present in the data could be one of two types: sampling zeroes or structural zeroes. While the former refers to unobserved values due to sampling limitations, the latter refers to a logical impossibility of observing a non-zero value. The inability to effectively represent this information graphically is another shortcoming of BNs. Event trees can represent such information succinctly by simply deleting the unpopulated edges where the absence of a non-zero value can be justified to be a logical constraint. The CEG, being a function of its underlying event tree, inherits the property of embedding such information directly in its structure.

Interventions in public health typically focus on some aspect(s) of the training of service providers such as doctors, nurses and pharmacists, or refinement of assessment methods, referral pathways, treatment plans or post-treatment care of patients. Asymmetries in models of such interventions manifest as context-specific conditional independence relations between the variables of the model. The lack of symmetry in such problems means that the approach of using a BN for modelling purposes is rather contrived. CEGs are much more sympathetic to the analysis of tree-like structures that directly express sequencing of events, contextual independence relations and structural zeroes in their topology as described above. In public health contexts like the one we describe here, where there is critical information about the underlying process concerning the order in which events unfold, we have found that non-stratified CEGs can provide a much more transparent framework for the analysis of various competing hypothesis and their causal extensions.

## 2 Chain Event Graphs

We briefly review the necessary concepts for understanding the construction of the CEG. For a detailed description, see [20].

**Event Tree:** An event tree  $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$  is an acyclic, connected, directed graph. The vertices represent events that may occur to a unit during the unfolding of the process being modelled. There exists only one vertex with no parents, this vertex is the root vertex  $s_0$ . All other vertices have exactly one parent. The vertices with no children are called *leaves*. Let  $L(\mathcal{T})$  be the set of leaves. The non-leaf vertices are called *situations*. Let  $S(\mathcal{T})$  form the set of situations where  $S(\mathcal{T}) = V(\mathcal{T}) \setminus L(\mathcal{T})$ . Let  $\mathbb{X}(s_i)$  be the set of children of situation  $s_i$ .

**Floret:** A floret  $\mathcal{F}$  of situation  $s_i$  is defined as  $\mathcal{F}(s_i) = (V(\mathcal{F}(s_i)), E(\mathcal{F}(s_i)))$  where  $V(\mathcal{F}(s_i)) = s_i \cup \mathbb{X}(s_i)$  and  $E(\mathcal{F}(s_i))$  is the set of edges induced by  $V(\mathcal{F}(s_i))$  in  $\mathcal{T}$ .

**Stage:** Two situations  $s_i$  and  $s_j$  in  $\mathcal{T}$  are said to be in the same stage  $u$  if and only if there exists a bijection  $\psi_u$  under which  $\mathbb{X}(s_i)$  and  $\mathbb{X}(s_j)$  have the same distribution given by  $\psi_u(s_i, s_j) : \mathbb{X}(s_i) \rightarrow \mathbb{X}(s_j)$ , where  $\psi_u$  leads to a mapping of the edge labels which is meaningful for the real-world application.

**Staged Tree:** In event tree  $\mathcal{T}$ , situations in the same stage are given the same colour. A coloured event tree  $\mathcal{T}$  is called a staged tree  $\mathcal{ST}$ .

**Position:** Two situations  $s_i$  and  $s_j$  in  $\mathcal{T}$  are said to be in the same position  $w$  if and only if the staged trees  $\mathcal{ST}_{s_i}$  and  $\mathcal{ST}_{s_j}$  rooted at  $s_i$  and  $s_j$  respectively are isomorphic in the sense of isomorphism between coloured graphs.

**Chain Event Graph:** The Chain Event Graph is obtained from its underlying staged tree by collapsing situations in the same position into one vertex and adding an additional vertex  $w_\infty$  into which all the leaf vertices are collected.

## 3 Intervention

Falls-related injuries and fatalities are a serious problem among the elderly. According to NICE guidelines [8], 30% of people older than 65 and 50% of people older than 80 fall at least once a year. Eldridge *et al* [7] modelled an intervention to reduce falls-related injuries in older people using a probability tree for short-term analysis and a Markov model for long-term effectiveness. The intervention was designed to enhance assessment, referral pathways and treatment for high-risk individuals aged over 65 years who have a substantial risk of falling. After assessment, individuals are classified as high-risk or low-risk of falling as per the recommendations in FRAT (Falls Risk Assessment Tool) [13]. In our model, we have additionally classified in-

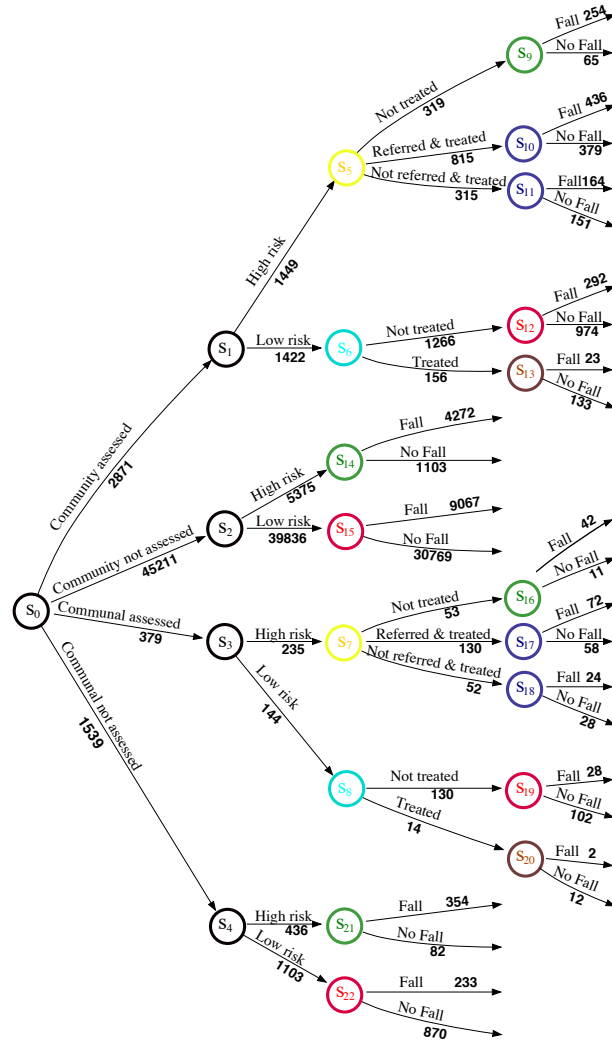
dividuals by their type of residence as we learned from domain literature that the fall rates are higher for individuals living in institutionalised care [14]. We incorporate this information by setting the proportion of high-risk individuals in communal establishments higher than in the community. The variables used to study this intervention are  $X = \{X_A, X_R, X_T, X_F\}$ . Here  $X_A$  indicates whether the individual aged over 65 resides in the community or in a communal establishment (such as nursing homes, care homes, hospitals) and whether they have been assessed or not;  $X_R$  indicates the risk level of the individual as high or low;  $X_T$  indicates whether the individual has been referred & treated, not referred & treated or not treated; and  $X_F$  indicates whether the individual suffered from a fall or not. As per the design of this intervention, all referred persons are treated and all those who fall under the not treated category have not been referred. Low-risk individuals do not receive referrals for treatment and hence we only classify them as treated or not treated.

## 4 Methods

For the intervention described above, we first constructed a staged tree as shown in figure 1. Observe that this tree is non-stratified as the treatment variable  $X_T$  has no logical interpretation for individuals who haven't been assessed as by intervention design they cannot receive any treatment. This staged tree is our data generating model. We simulated data for 50,000 individuals passing through this system by forward sampling. The numbers along the edges represent the observations along each branch. Observe that several of the branches are sparsely populated. For instance, there are only two observations along the edge indicating falls for assessed low-risk individuals who live in a communal establishment and have received treatment. Sparsely populated branches may pose a problem for model selection. We discuss this further in section 6.

We use the R package **bnlearn** for learning the network structure from the data using the Hill-Climbing algorithm. We use a weakly informative equivalent sample size or alpha of 4 and set the prior on the edges of the tree using the mass conservation property as described in [4]. All CEG structures are assumed a priori equally likely. We use an adapted form of the Agglomerative Hierarchical Clustering (AHC) algorithm developed in [9] to fit a CEG to the data. We compared the two graphical models based on their BDeu score as well as their ability to express context-specific information and structural zeroes which are essential in assessing the effectiveness of the intervention.

The AHC algorithm in [9] sequentially merges situations into stages by combining the two situations at every step which give the highest improvement to the BDeu score of the graph. This was further refined by the introduction of the concept of hyperstages in [3]. A hyperstage consists of sets of situations such that two situations can be merged into a single stage if and only if they belong to the same set in the

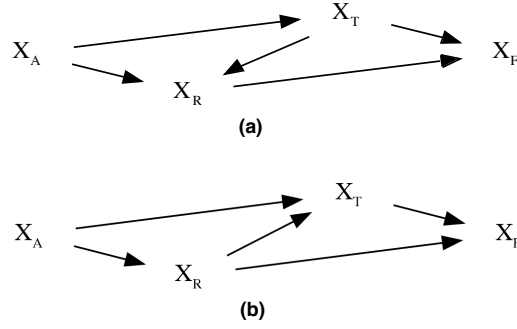


**Fig. 1** Staged tree representing the data generating model.

hyperstage. This allows further information about the domain to be embedded into the model and can also greatly reduce the dimension of the model search space. Under this adaptation, we used the domain knowledge to set the hyperstage structure and used this structure to run the AHC algorithm to output the best fitting CEG to the data.

## 5 Results

The Hill-Climbing algorithm outputs the BN in figure 2(a). As the intervention gives rise to a total order of  $X_A \leq X_R \leq X_T \leq X_F$ , certain edges must be suppressed in order for the BN to be representative of the real-world application. For instance, the directed edge from *Treatment* to *Risk* is not permissible given the total order. This gives rise to the BN in figure 2(b). The BDeu score of this BN structure is -68709.99.



**Fig. 2** (a) Original BN returned using the Hill-Climbing algorithm; (b) Best-fitting BN which admits the total order of  $X_A \leq X_R \leq X_T \leq X_F$ .

We set up the hyperstage structure for the AHC algorithm as  $[\{s_1, s_2, s_3, s_4\}, \{s_5, s_7\}, \{s_6, s_8\}, \{s_9, s_{10}, s_{11}, s_{14}, s_{16}, s_{17}, s_{18}, s_{21}\}, \{s_{12}, s_{13}, s_{15}, s_{19}, s_{20}, s_{22}\}]$ . With an alpha of size 4, the best-fitting CEG returned by the AHC is given in figure 4 with position structure given by  $w_0 = \{s_0\}$ ,  $w_1 = \{s_1\}$ ,  $w_2 = \{s_2\}$ ,  $w_3 = \{s_3\}$ ,  $w_4 = \{s_4\}$ ,  $w_5 = \{s_5, s_7\}$ ,  $w_6 = \{s_6, s_8\}$ ,  $w_7 = \{s_9, s_{14}, s_{16}, s_{21}\}$ ,  $w_8 = \{s_{12}, s_{15}, s_{19}, s_{22}\}$ ,  $w_9 = \{s_{10}, s_{11}, s_{17}, s_{18}\}$ ,  $w_{10} = \{s_{13}, s_{20}\}$ ,  $s_\infty = L(\mathcal{T})$ . Note that this is exactly the position structure of the data generating tree given in figure 1. The BDeu score of this CEG is -68671.59, thus giving an extremely high Bayes Factor of  $4.7523 \times 10^{16}$  in favour of the CEG model. By Kass and Raftery's interpretation [12], this indicates *very strong evidence* in favour of the CEG model being a better fit to the data.

The following context-specific conditional independence statements can be read from the CEG in figure 4 using *cuts* as defined in [20]:

$$X_T \perp\!\!\!\perp X_R \mid \{X_A = \text{Not Assessed}\}, \quad (1)$$

$$X_T \not\perp\!\!\!\perp X_R \mid \{X_A = \text{Assessed}\}, \quad (2)$$

whereas figure 2(b) fails to show these contextual independencies.

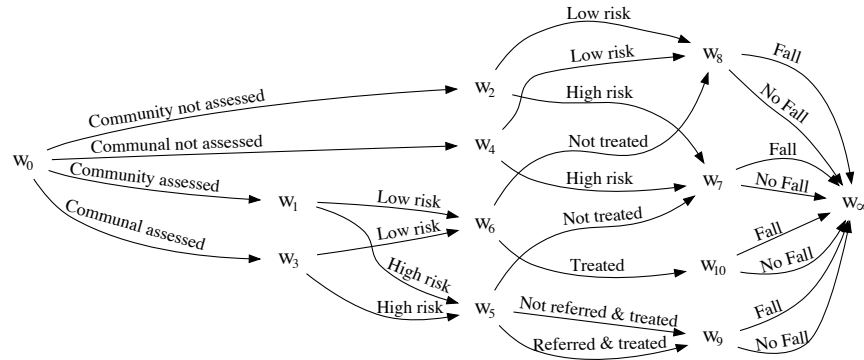


Fig. 3 CEG returned by the AHC algorithm.

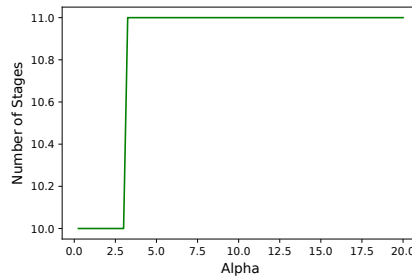


Fig. 4 The CEG model is robust given varying values of alpha.

To assess the robustness of the CEG model, we fit the data to varying values of alpha. Figure 4 shows the number of stages for each value of alpha between 0.25 and 20 with increments of 0.25. The number of stages remains steadily at 11 for alpha greater than three and at 10 for alpha less than three. Moreover, we observe that the stage structure also remains the same for the alpha values returning 10 stages and similarly for those returning 11 stages. In conclusion, we demonstrated that in scenarios with context-specific independencies and structural zeroes, a CEG is much more expressive and flexible while also being a better fit to the data as evidenced by the high Bayes Factor score.

## 6 Discussion

We observed in section 5 that the CEG is robust against varying values of alpha. For alpha greater than three, the resultant CEG was the data generating model. The

smaller values of alpha were unable to return this due to sparsity along some of the branches of the event tree. In the falls intervention scenario we had domain literature to support the veracity of the staging produced by the AHC. In the absence of sufficient domain information, situations whose emanating edges contain sparse data-points should be maintained as singletons in the hyperstage structure. Failing this, spurious stages may be merged by the AHC resulting in an unreliable CEG.

In this paper, we compared two types of graphical models - Bayesian Networks and Chain Event Graphs - demonstrating the superiority of CEGs in modelling structural zeroes and context-specific independencies by modelling a falls intervention. In [7], which provided the falls intervention design, the effectiveness of the intervention was assessed using a probability tree for the short-term and a Markov model for the long-term. We briefly outline the advantages of CEGs over these modelling techniques when presented with a scenario such as the falls intervention.

A probability tree is the same as an uncoloured event tree. Unlike the BN, probability trees and Markov models can satisfactorily express asymmetric information introduced by structural zeroes. However, an essential property which these models lack is the ability to read conditional independence statements from their topology. They also do not admit causal manipulations. The colouring of the staged tree, followed by the collapsing of vertices in the CEG not only provide a succinct and complete description of the various paths that an individual may traverse as they pass through the system but also allow us to read context-specific conditional independence relationships and under a legitimate causal setting, can be subject to manipulations and a causal analysis.

It is important to note that the Markov model in [7] was for assessing long-term effectiveness. The CEG described in this paper caters to short-term analysis. Dynamic variants of the CEG have been developed in [1, 5]. For applications such as the falls intervention, we observe that individuals take varying amounts of time to move from certain states. For instance, individuals living in the community who have been assessed and have been referred and treated will not all fall. Also, those who do suffer from a fall would do so after varying amounts of time since they received their treatment. In such settings, it is also typical to record observations when a transition occurs rather than when an individual remains in the same state. For instance, we would note that an individual has suffered a fall and the time that has elapsed since their treatment but we are unlikely to record every day or every month that an individual has not suffered a fall. This type of setting corresponds more closely to a semi-Markov process rather than a Markov process. However, like a Markov model, a semi-Markov model will not allow reading of conditional independence statements. For this purpose, we have developed a dynamic variant of the CEG that has an underlying semi-Markov structure. This is called the Reduced Dynamic Chain Event Graph (RDCEG). An early application of this class of models can be found in [21] and a paper describing the technical details is in preparation [19].



## References

1. Barclay, L.M., Collazo, R.A., Smith, J.Q., Thwaites, P.A., Nicholson, A.E.: The dynamic chain event graph. *Electronic Journal of Statistics* **9(2)**, 2130–2169 (2015)
2. Boutilier, C., Friedman, N., Goldszmidt, M., Koller, D.: Context-specific independence in Bayesian networks. In: *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pp. 115–123. Morgan Kaufmann Publishers Inc. (1996)
3. Collazo R.A.: The dynamic chain event graph. PhD thesis, University of Warwick (2017)
4. Collazo R.A., Görgen C., Smith J.Q.: *Chain Event Graphs*. Chapman & Hall/CRC, Boca Raton, FL (2017)
5. Collazo, R.A., Smith, J.Q.: An N Time-Slice Dynamic Chain Event Graph. *International Journal of Approximate Reasoning* (to appear)
6. Darwiche, A.: *Modeling and reasoning with Bayesian networks*. Cambridge University Press (2009)
7. Eldridge, S., Spencer, A., Cryer, C., Parsons, S., Underwood, M., Feder, G.: Why modelling a complex intervention is an important precursor to trial design: lessons from studying an intervention to reduce falls-related injuries in older people. *Journal of Health Services Research & Policy* **10(3)**, 133–142 (2005)
8. Falls in older people: assessing risk and prevention. In: *Guidance and guidelines — NICE* (2013).  
<https://www.nice.org.uk/guidance/cg161>
9. Freeman, G., Smith, J.Q.: Bayesian MAP model selection of chain event graphs. *Journal of Multivariate Analysis* **102.7**, 1152–1165 (2011)
10. Friedman, N., Goldszmidt, M.: Learning Bayesian networks with local structure. In: *Learning in graphical models*, pp. 421–459. Springer (1998)
11. Heckerman, D., Geiger, D., Chickering, D.: Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning* **20(3)**, 197–243 (1995)
12. Kass, R.E., Raftery, A.E.: Bayes factors. *Journal of the american statistical association* **90(430)**, 773–795 (1995)
13. Nandy, S., Parsons, S., Cryer, C., Underwood, M., Rashbrook, E., Carter, Y., Eldridge, S., Close, J., Skelton, D., Taylor, S.: Development and preliminary examination of the predictive validity of the Falls Risk Assessment Tool (FRAT) for use in primary care. *Journal of Public Health* **26(2)**, 138–143 (2004)
14. Nurmi, I., Lühje, P.: Incidence and costs of falls and fall injuries among elderly in institutional care. *Scandinavian journal of primary health care* **20(2)**, 118–122 (2002)
15. Office for National Statistics: 2011 Census data. This information is licensed under the terms of the Open Government Licence [<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3>]  
<http://dx.doi.org/10.5257/census/aggregate-2011-1>
16. Pearl, J.: *Causality*. Cambridge university press (2009)
17. Poole, D., Zhang, N.L.: Exploiting contextual independence in probabilistic inference. *Journal of Artificial Intelligence Research* **18**, 263–313 (2003)
18. Shafer G.: *The art of causal conjecture*. MIT press (1996)
19. Shenvi, A., Smith, J.Q.: Elicitation, Estimation and Model Selection for the Reduced Dynamic Chain Event Graph. (in preparation)
20. Smith, J.Q., Anderson, P.E.: Conditional independence and chain event graphs. *Artificial Intelligence* **172(1)**, 42–68 (2008)
21. Smith, J.Q., Shenvi, A.: *Assault Crime Dynamic Chain Event Graphs*. University of Warwick repository (2018)  
<http://wrap.warwick.ac.uk/104824/>
22. Thwaites, P.: Causal identifiability via chain event graphs. *Artificial Intelligence* **195**, 291–315 (2013)
23. Thwaites, P., Smith, J.Q., Riccomagno, E.: Causal analysis with chain event graphs. *Artificial Intelligence* **174(12-13)**, 889–909 (2010)