Embellishing a Bayesian Network using a Chain Event Graph

L. M. Barclay J. L. Hutton J. Q. Smith

University of Warwick

4th Annual Conference of the Australasian Bayesian Network Modelling Society

Introduction

- Chain Event Graphs (CEGs) (Smith and Anderson, 2008) are derived from probability trees by merging the nodes in a tree whose associated conditional probabilities are the same
- The CEG generalises the discrete BN by allowing for asymmetric dependence structures between the variables

We demonstrate on a real-world health study that

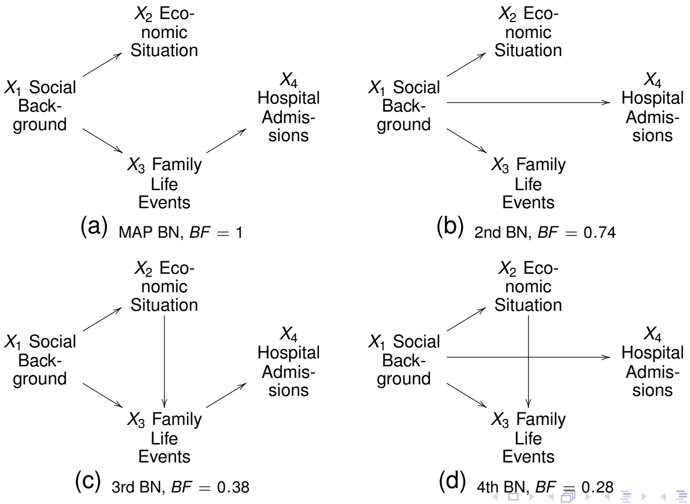
- CEGs can embellish an initial Bayesian Network model description
- They lead to promising higher scoring models
- CEGs allows us to make more refined conclusions about the given problem

Example: 'Effect of Social and Family Factors on Childhood Hospital Admissions'
(Fergusson et al., 1986)

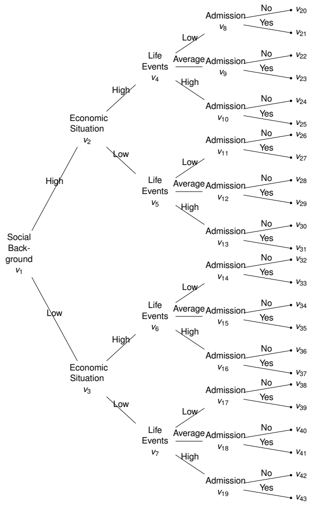
Looks at the effect of the social background, the economic situation and the family life events on the child's physical health.

- $X_1 =$ Social background: binary, high or low social background
- $X_2 =$ Economic situation: binary, high or low eonomic situation
- $X_3 =$ Number of family life events: three categories, low, average or high number of events
- $X_4 =$ Hospital admission: binary, no admission, at least one admission

Highest scoring BN structures



CEG: Stages and Positions



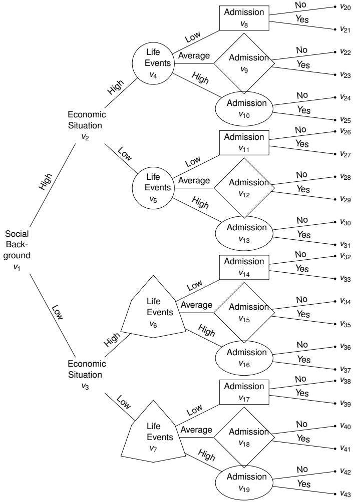
Two situations $v, v' \in S(T)$ are in the same **stage** u if and only if

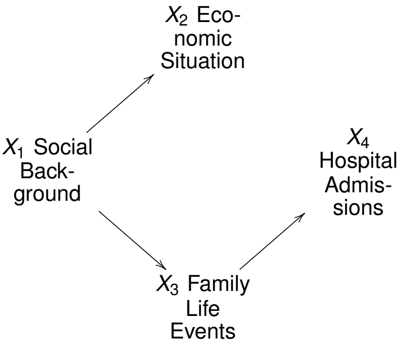
- The topology of the florets $F(v)$ and $F(v')$ are the same
- There is a bijection between the florets such that the probabilities on corresponding edges are the same

Two situations $v, v' \in S(T)$ are in the same **position** w if and only if

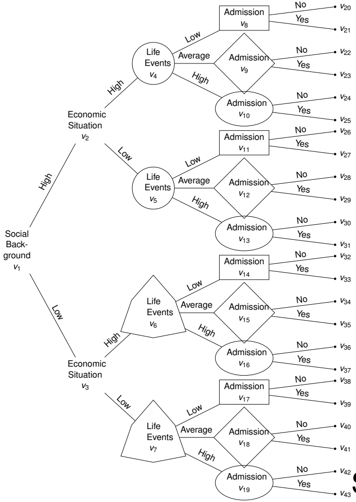
- The topology of the subtrees $T(v)$ and $T(v')$ are the same
- There is a bijection between the subtrees such that the probabilities on corresponding edges are the same

Stages and Positions





Definition of a Chain Event Graph

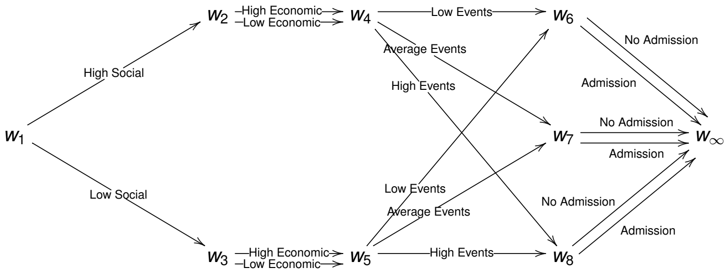


- The set of vertices is the set of all positions of the tree T and the position of all leaf nodes.
- For each position w choose a single representative situation $v(w) \in S(T)$. We have an edge from w to w' for each edge from $v(w)$ to a vertex $v' \in w'$.
- If $u(w) \neq \{w\}$, there is more than one position in the stage, then we connect two positions by an undirected dotted line.

Smith and Anderson (2008)

Writing the BN as a CEG

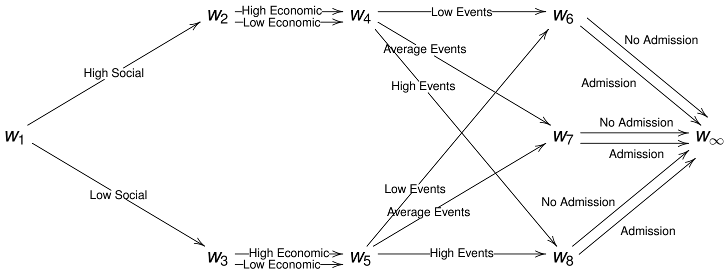
$$w_1 = \{v_1\}, w_2 = \{v_2\}, w_3 = \{v_3\}, w_4 = \{v_4, v_5\}, w_5 = \{v_6, v_7\},$$
$$w_6 = \{v_8, v_{11}, v_{14}, v_{17}\}, w_7 = \{v_9, v_{12}, v_{15}, v_{18}\}, w_8 = \{v_{10}, v_{13}, v_{16}, v_{19}\},$$
$$w_\infty = \{v_{20}, \ldots, v_{43}\}$$



$\Rightarrow$ Strong symmetry due to BN

Writing the BN as a CEG

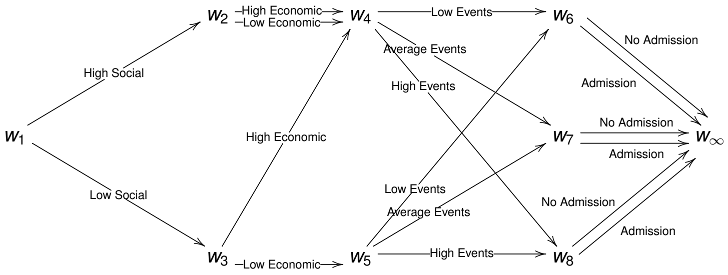
$$w_1 = \{v_1\}, w_2 = \{v_2\}, w_3 = \{v_3\}, w_4 = \{v_4, v_5\}, w_5 = \{v_6, v_7\},$$
$$w_6 = \{v_8, v_{11}, v_{14}, v_{17}\}, w_7 = \{v_9, v_{12}, v_{15}, v_{18}\}, w_8 = \{v_{10}, v_{13}, v_{16}, v_{19}\},$$
$$w_\infty = \{v_{20}, ..., v_{43}\}$$



$\Rightarrow$ Strong symmetry due to BN

Writing the BN as a CEG

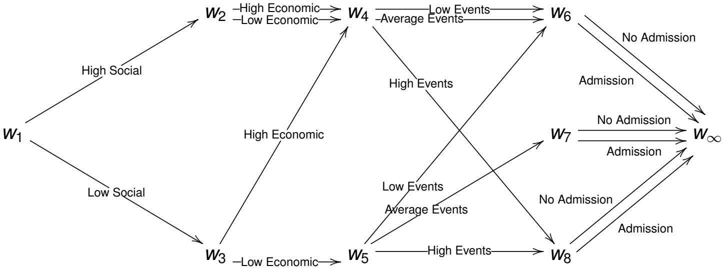
$$w_1 = \{v_1\}, w_2 = \{v_2\}, w_3 = \{v_3\}, w_4 = \{v_4, v_5\}, w_5 = \{v_6, v_7\},$$
$$w_6 = \{v_8, v_{11}, v_{14}, v_{17}\}, w_7 = \{v_9, v_{12}, v_{15}, v_{18}\}, w_8 = \{v_{10}, v_{13}, v_{16}, v_{19}\},$$
$$w_\infty = \{v_{20}, \ldots, v_{43}\}$$



⇒ Strong symmetry due to BN

Writing the BN as a CEG

$$w_1 = \{v_1\}, w_2 = \{v_2\}, w_3 = \{v_3\}, w_4 = \{v_4, v_5\}, w_5 = \{v_6, v_7\},$$
$$w_6 = \{v_8, v_{11}, v_{14}, v_{17}\}, w_7 = \{v_9, v_{12}, v_{15}, v_{18}\}, w_8 = \{v_{10}, v_{13}, v_{16}, v_{19}\},$$
$$w_\infty = \{v_{20}, ..., v_{43}\}$$



$\Rightarrow$ Strong symmetry due to BN

Scoring CEGs

Freeman and Smith (2011) set up a scoring method for CEGs equivalent to the BDe-metric (Heckerman et al., 1995):

- Let $\Pi_u = \{\pi(e(w', w)|w) : w \in u\}$, the set of conditional probabilities associated with the floret $F(u)$
- Assumptions:
  - Stage priors are mutually independent
  - Equivalent stages in different CEGs have the same prior distribution
- $\Rightarrow \Pi_u \sim Dir(\boldsymbol{\alpha}_u)$, $\boldsymbol{\alpha}_u = (\alpha_{u1}, .., \alpha_{ur_u})$
- $\Rightarrow \Pi_u|D \sim Dir(\boldsymbol{\alpha}_u + \boldsymbol{N}_u)$, $\boldsymbol{N}_u = (N_{u1}, .., N_{ur_u})$
- Simplest case: uniform prior on the root-to-leaf paths of the associated tree

Scoring CEGs

- The joint probability $p(C, D)$ of a CEG structure C and a dataset of cases D is given by

$$p(C)p(D|C) = p(C) \prod_{u \in J(T)} \frac{\Gamma(\alpha_u)}{\Gamma(\alpha_u + N_u)} \prod_{k=1}^{r_u} \frac{\Gamma(\alpha_{uk} + N_{uk})}{\Gamma(\alpha_{uk})},$$

  where $\alpha_u = \sum_k \alpha_{uk}$ and $N_u = \sum_k N_{uk}$.
- Assuming that structures are a priori equally likely we compare two CEG structures using log Bayes factors:

$$log\ p(D|C_1) - log\ p(D|C_0).$$

- Note: Calculation depends only on the stages in which they differ.

AHC algorithm for CEGs (Freeman and Smith, 2011)

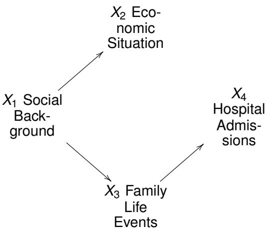
1. Start with CEG C_0, finest partition into stages
2. For each pair of situations with the same number of edges emanating from it calculate the log Bayes Factor $log\ p(D|C_1^*) - log\ p(D|C_0)$, where C_1^* is the CEG constructed by putting the two situations into the same stage.
3. Let $C_1 = \arg\max_{C_1^*}(log\ p(D|C_1^*) - log\ p(D|C_0))$.
4. Calculate C_2^* by merging two situations in C_1 and hence find C_2.
5. Continue until the coarsest partition, is reached.
6. Select the CEG of $\{C_0, C_1, C_2, ..., C_\infty\}$ which has the highest score.

Example: The AHC Algorithm

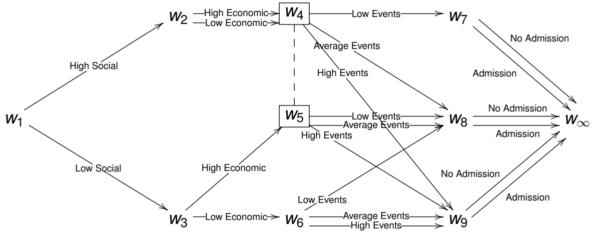
CEG	Stages merged	Log-Bayes Factor	CEG score
C_0			-2512.708
C_1	$\{v_5, v_6\}$	5.528	-2507.18
C_2	$\{v_{18}, v_{19}\}$	3.731	-2503.449
C_3	$\{v_9, v_{17}\}$	3.453	-2499.996
C_4	$\{v_{13}, v_{18}, v_{19}\}$	3.377	-2496.619
C_5	$\{v_8, v_{11}\}$	3.305	-2493.314
C_6	$\{v_9, v_{12}, v_{17}\}$	3.060	-2490.254
C_7	$\{v_{10}, v_{13}, v_{18}, v_{19}\}$	3.041	-2487.213
C_8	$\{v_{14}, v_{15}\}$	2.565	-2484.648
C_9	$\{v_{10}, v_{13}, v_{16}, v_{18}, v_{19}\}$	2.514	-2482.134
C_{10}	$\{v_9, v_{12}, v_{14}, v_{15}, v_{17}\}$	2.342	-2479.792
C_{11}	$\{v_4, v_5, v_6\}$	1.302	-2478.490
C_{12}	$\{v_9, v_{10}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}\}$	-0.812	-2479.302
C_{13}	$\{v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}\}$	-8.764	-2488.066
C_{14}	$\{v_4, v_5, v_6, v_7\}$	-36.638	-2524.704
C_∞	$\{v_2, v_3\}$	-62.440	-2587.144

Results

- Improvement in score by the CEG: 11.286
- Bayes Factor: 79 698



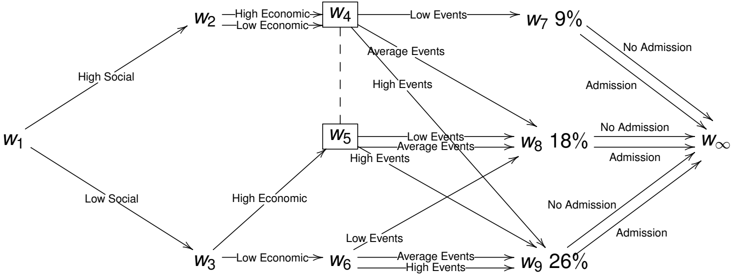
Score: -2489.776



Score: -2478.490

Results

- Improvement in score by the CEG: 11.286
- Bayes Factor: 79 698



Conclusions

CEGs

- provide a useful embellishment to the BN
- provide a significantly better score than the discrete BN
- retains the expressiveness to the client

Further work:

- Accommodation of missing data structures
- Development of a dynamic CEG, two time-slice CEG
- Development of a CEG software

References I

D.M. Fergusson, L.J. Horwood, and F.T. Shannon. Social and family factors in childhood hospital admission. *Journal of epidemiology and community health*, 40(1):50, 1986.

G. Freeman and J.Q. Smith. Bayesian map model selection of chain event graphs. *Journal of Multivariate Analysis*, 2011.

D. Heckerman, D. Geiger, and D.M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

J.Q. Smith and P.E. Anderson. Conditional independence and chain event graphs. *Artificial Intelligence*, 172(1):42–68, 2008.