Causal inference in staged tree models

Christiane Görgen and Jim Q. Smith

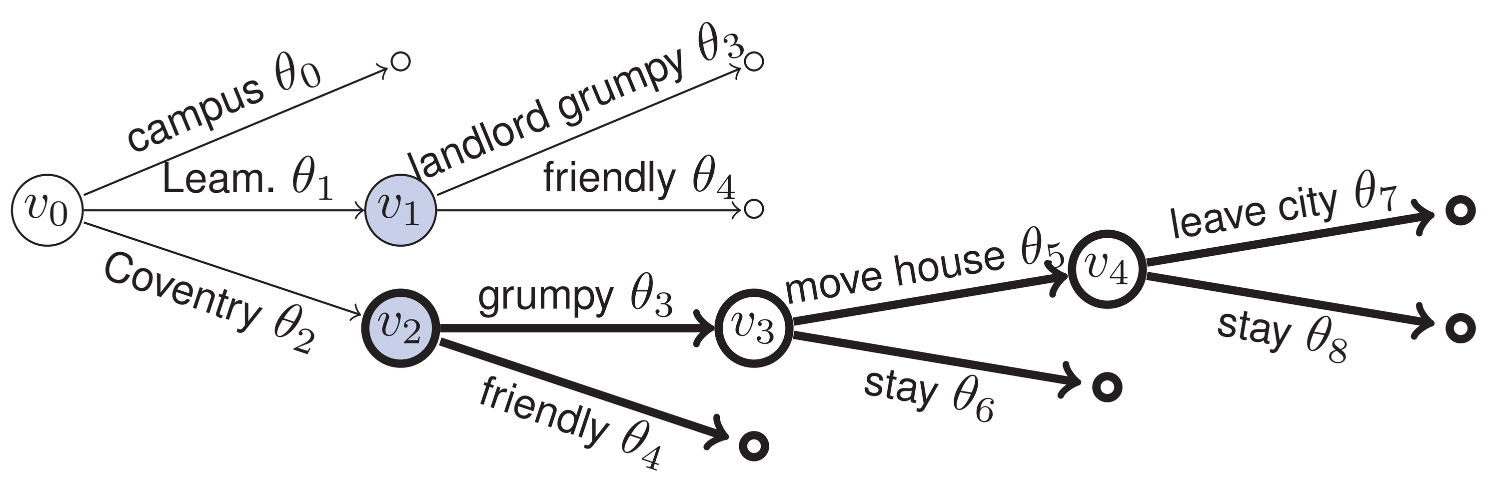
Department of Statistics, University of Warwick, Coventry CV4 7AL, UK. c.gorgen@warwick.ac.uk, j.q.smith@warwick.ac.uk

Abstract Staged tree models generalise discrete Bayesian networks (BNs). They are particularly useful when modelling asymmetric situations where BNs would unnecessarily assume an underlying state space to have a product structure. Centrally, staged trees do not rely on an a priori set of problem variables and are particularly strong when a model is specified in terms of relationships between a collection of events. We now apply a recent algebraic characterisation of these models to infer putative causal hypotheses and to translate these into easily calculable statements which do not rely on a graphical representation

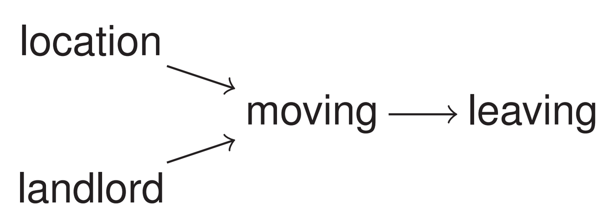
Staged tree models

A *probability tree* is an event tree $\mathcal{T} = (V, E)$ with labelled edges $\theta(e)$, $e \in E$. The product $\prod_{e \in E(\lambda)} \theta(e)$ of transition probabilities along root-to-leaf paths λ defines a strictly positive probability mass function on $\mathcal{T}$. Every probability tree represents a discrete parametric statistical model.

A *staged tree* is a probability tree where the emanating labels of certain vertices are identified:



A BN representation:



- Staged tree models include discrete (and context-specific) BNs as a special case.
- They are always *faithful* and all atomic probabilities are *strictly positive*.
- Tree graphs are efficient in **modelling *asymmetric problems***.

Causal manipulations in staged trees

We can **manipulate vertices** to force all units to follow a certain development. E.g. manipulating v_2 in $\mathcal{T}$ below forces students to live in one of the two cities.

Graphically, this is a *projection onto a subtree* with inherited labels. Here, the thick depicted $\mathcal{T}(v_2) \subseteq \mathcal{T}$.

- Situations upstream of the manipulation and counterfactuals are not affected.
- Vertex manipulation is **more general than Pearl's *atomic cause***, it is more like a context-specific intervention.
- **Causes and effects are *events***, not random variables.
- Effects are depicted downstream of a causal manipulation, so cause and effect are **ordered chronologically**.

An algebraic characterisation

We write the probability $P_{\boldsymbol{\theta}}(A)$ of an event A as a *formal polynomial*

$$c_{A,\mathcal{T}}(\boldsymbol{\theta}) = \sum_{\lambda \in A} \prod_{e \in E(\lambda)} \mathbb{1}_e(\lambda)\theta(e)$$

which is a sum of atomic probabilities together with an indicator function of a unit passing through the edge e.

The **effect of the causal manipulation** on an event A can be easily **calculated as a *differentiation* on this polynomial**

$$P_{\boldsymbol{\theta}}(A \mid\mid v) = \frac{\partial^2 c_{\mathcal{T},A}(\boldsymbol{\theta})}{\partial\theta(e)\partial\mathbb{1}_e} = c_{\mathcal{T}(v),A}(\boldsymbol{\theta})$$

and coincides with an analogous formal polynomial defined on the subtree rooted at the intervention vertex $v \in V$.

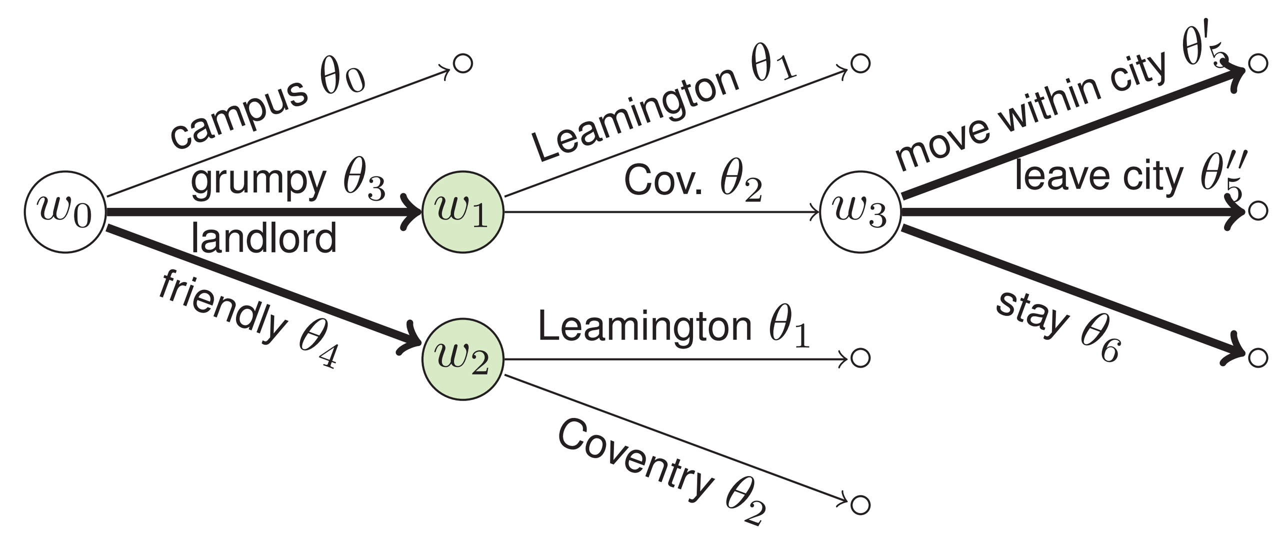
Example: Advantages of an algebraic approach to causality

The intervention on v_2 in $\mathcal{T}$ transforms the polynomial $c_{\mathcal{T}}(\boldsymbol{\theta}) = \theta_0 + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_3\theta_5\theta_7 + \theta_2\theta_3\theta_5\theta_8 + \theta_2\theta_3\theta_6 + \theta_2\theta_4$ to the polynomial of the subtree $\mathcal{T}(v_2)$, so $c_{\mathcal{T}(v_2)}(\boldsymbol{\theta}) = \theta_3\theta_5\theta_7 + \theta_3\theta_5\theta_8 + \theta_3\theta_6 + \theta_4$.

The probability of a student leaving the city, assuming she was initially forced to live there, is $P_{\boldsymbol{\theta}}(v_4 \mid\mid v_2) = \frac{\partial c_{\mathcal{T},v_2}(\boldsymbol{\theta})}{\partial\theta_7\partial\mathbb{1}_{e_7}} = \theta_3\theta_5$. This is simply the probability of the subpath $v_2 \to v_4$ in $\mathcal{T}$.

There is **not always a straight forward graphical way** of representing a manipulation operation, e.g.:

- Forcing students to live in one of the two cities is not a vertex manipulation in the tree on the right.
- Forcing students who are renting with a grumpy landlord to move can be represented by $\mathcal{T}(v_4) \subseteq \mathcal{T}$ but on the right we would force a unit to go through two mutually exclusive edges simultaneously, following two different unfoldings from w_3.



Conclusions

Algebraic and differential methods allow us to perform causal manipulations on staged tree models without referring to the graph structure. They can be greatly generalised to models with a monomial parametrisation that do not rely on a priori problem variables.

References Jim Q. Smith and Paul E. Anderson. *Conditional independence and Chain Event Graphs*, Journal of Artifical Intelligence 172, 42–68, 2008.

Robert G. Cowell and Jim Q. Smith. *Causal discovery through MAP selection of stratified Chain Event Graphs*, Electronic Journal of Statistics 8, 965–997, 2014.

Peter A. Thwaites, Jim Q. Smith, and Eva Riccomagno. *Causal analysis with Chain Event Graphs*. Artificial Intelligence, 174(12-13):889–909, 2010.

Christiane Görgen, Manuele Leonelli and Jim Q. Smith. *A Differential Approach for Staged Trees*, Conference Proceedings of ECSQARU 2015.

UK Causal Inference Meeting 2016, April 12–15, London School of Hygiene and Tropical Medicine.

This research (with grant number EP/L505110/1) is funded by EPSRC Engineering and Physical Sciences Research Council