A Differential Approach to Causality in Staged Trees

Christiane Görgen and Jim Q. Smith

PGM 2016

Warwick Statistics

This research is funded by EPSRC grant EP/L505110/1.

Causal manipulation on probability trees

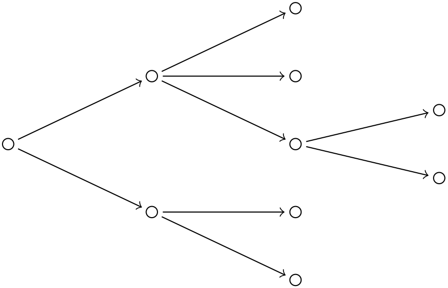
- more general than interventions on DAGs
- easily done in symbolic framework[1]

[1] Adnan Darwiche. A differential approach to inference in Bayesian networks. J. ACM, 50(3):280–305 (electronic), 2003.

Probability trees

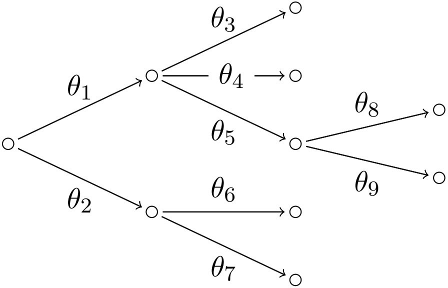
Probability trees

- *event tree* graphs



Probability trees

- *event tree* graphs
- edge labels (probabilities)



Staged probability trees

probability tree + conditional independence assumptions

Staged probability trees

probability tree + conditional independence assumptions

highly useful in asymmetric problems

Staged probability trees

probability tree + conditional independence assumptions

highly useful in asymmetric problems

- model selection techniques

Staged probability trees

probability tree + conditional independence assumptions

highly useful in asymmetric problems

- model selection techniques
- propagation algorithms

Staged probability trees

probability tree + conditional independence assumptions

highly useful in asymmetric problems

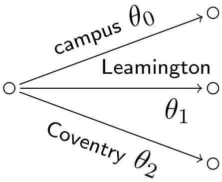
- model selection techniques
- propagation algorithms
- statistical equivalence classes

Example

Students at Warwick university...

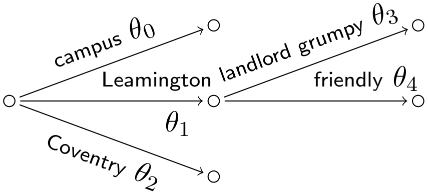
Example

Students at Warwick university...



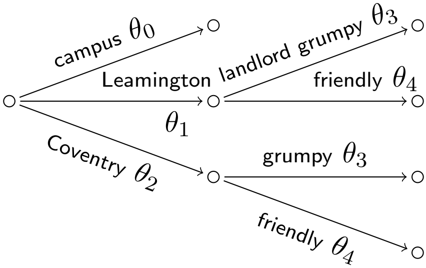
Example

Students at Warwick university. . .



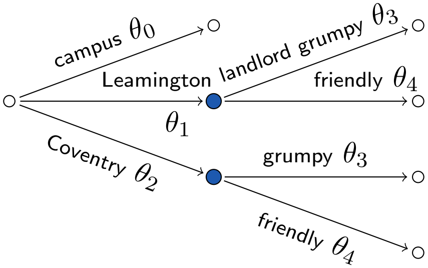
Example

Students at Warwick university. . .



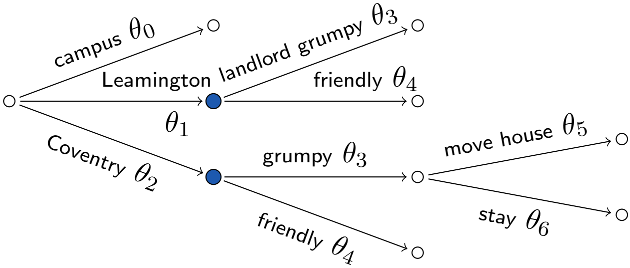
Example

Students at Warwick university...



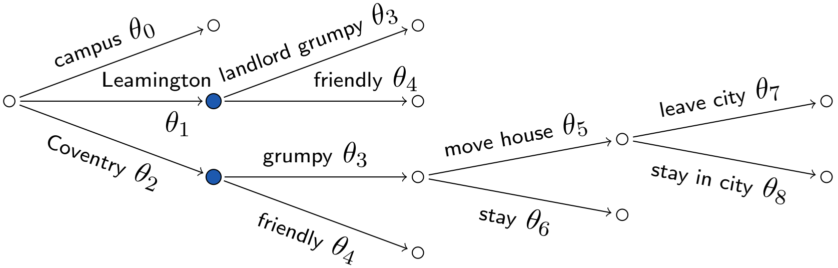
Example

Students at Warwick university...

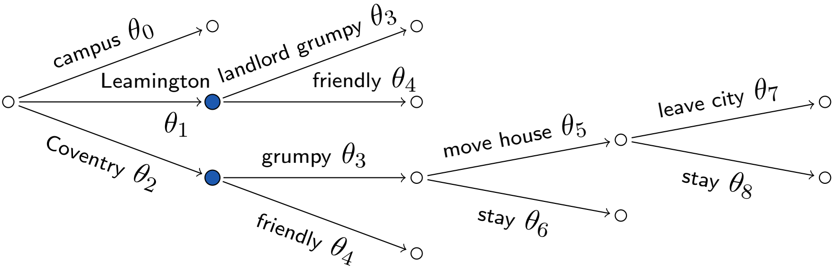


Example

Students at Warwick university...



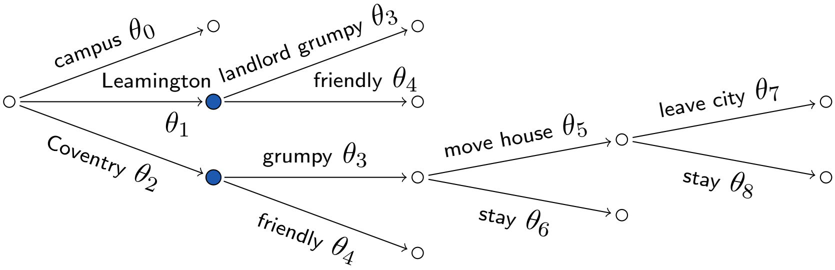
Interventions (graphically)



[2]Peter Thwaites. Causal Identifiability via Chain Event Graphs. Artificial Intelligence, 195:291–315, 2013.

Interventions (graphically)

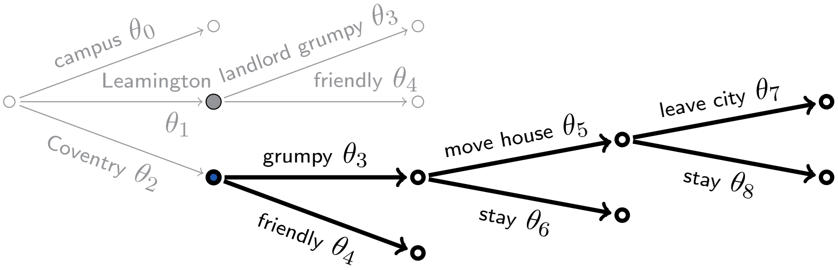
Impose policy forcing students to live in Coventry:



[2] Peter Thwaites. Causal Identifiability via Chain Event Graphs. Artificial Intelligence, 195:291–315, 2013.

Interventions (graphically)

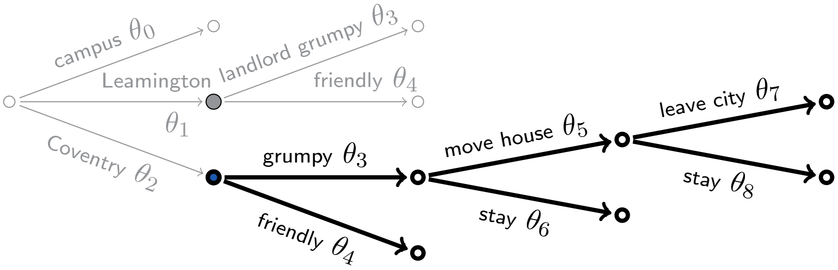
Impose policy forcing students to live in Coventry:



[2] Peter Thwaites. Causal Identifiability via Chain Event Graphs. Artificial Intelligence, 195:291–315, 2013.

Interventions (graphically)

Impose policy forcing students to live in Coventry:



causal interventions[2] $\rightsquigarrow$ projections onto a subtree

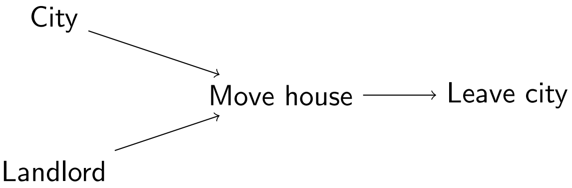
[2]Peter Thwaites. Causal Identifiability via Chain Event Graphs. Artificial Intelligence, 195:291–315, 2013.

More flexible than DAGs

Staged trees contain discrete Bayesian networks as a special case

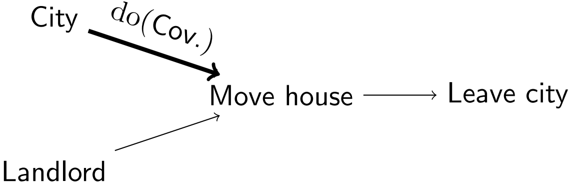
More flexible than DAGs

Staged trees contain discrete Bayesian networks as a special case



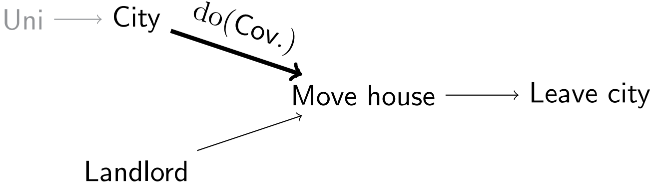
More flexible than DAGs

Staged trees contain discrete Bayesian networks as a special case



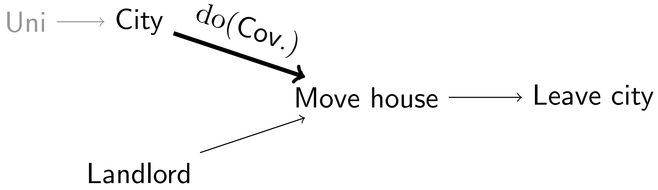
More flexible than DAGs

Staged trees contain discrete Bayesian networks as a special case



More flexible than DAGs

Staged trees contain discrete Bayesian networks as a special case



but are more general and more expressive!

Two questions

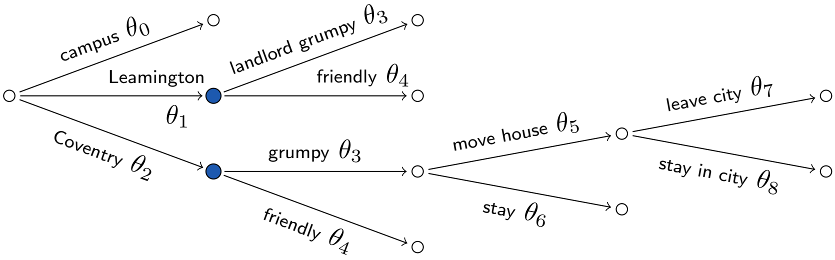
Two questions

- What if we want to do causal manipulations in staged trees without referring to a graph?

Two questions

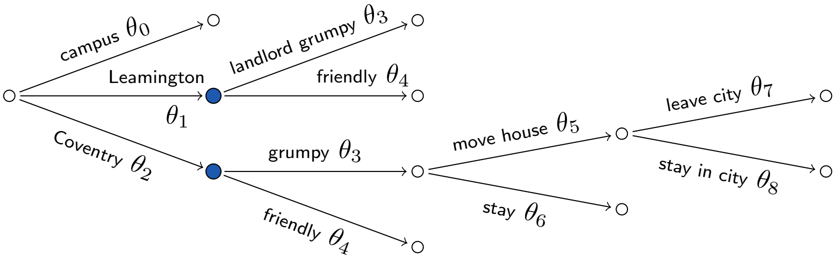
- What if we want to do causal manipulations in staged trees without referring to a graph?
- What if there is a sequence of manipulations we want to perform consecutively?

Replacing the graph by a polynomial



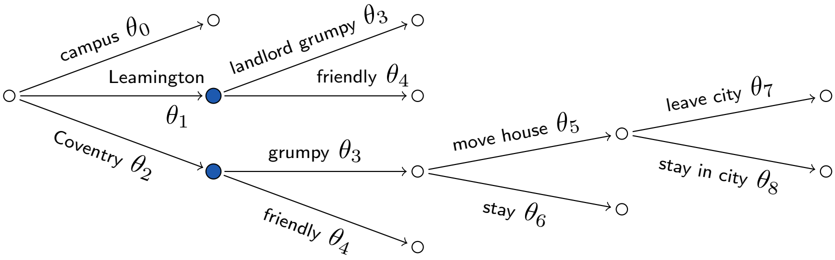
Replacing the graph by a polynomial

Every staged tree is in one-to-one correspondence with a nested polynomial:



Replacing the graph by a polynomial

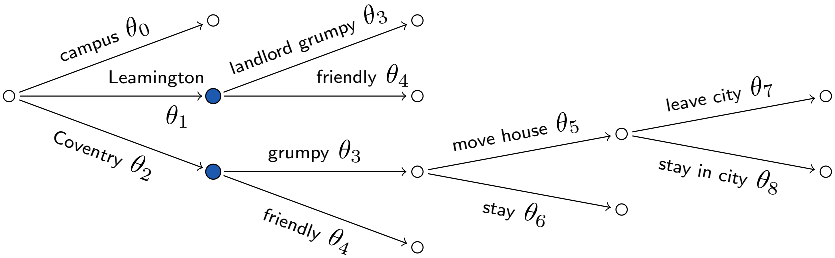
Every staged tree is in one-to-one correspondence with a nested polynomial:



$$c_{\mathcal{T}}(\boldsymbol{\theta}) = \theta_0 + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_3\theta_5\theta_7 + \theta_2\theta_3\theta_5\theta_8 + \theta_2\theta_3\theta_6 + \theta_2\theta_4$$

Replacing the graph by a polynomial

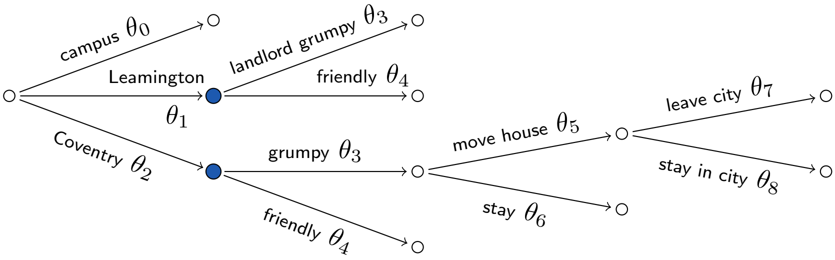
Every staged tree is in one-to-one correspondence with a nested polynomial:



$$
\begin{aligned}
c_{\mathcal{T}}(\boldsymbol{\theta}) &= \theta_0 + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_3\theta_5\theta_7 + \theta_2\theta_3\theta_5\theta_8 + \theta_2\theta_3\theta_6 + \theta_2\theta_4 \\
&= \theta_0 + \theta_1(\theta_3 + \theta_4) + \theta_2(\theta_3(\theta_5(\theta_7 + \theta_8) + \theta_6) + \theta_4)
\end{aligned}
$$

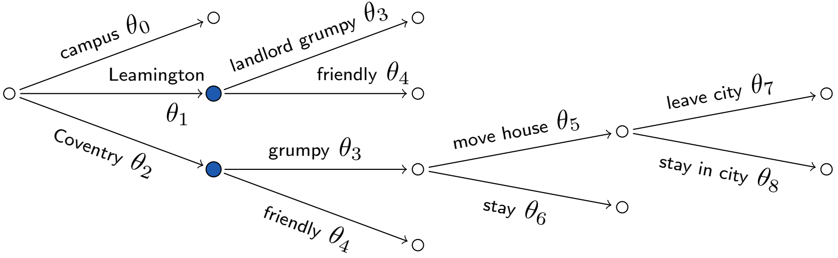
Interventions (symbolically)

Manipulate a tree using a **differentiation operation** on this polynomial:



Interventions (symbolically)

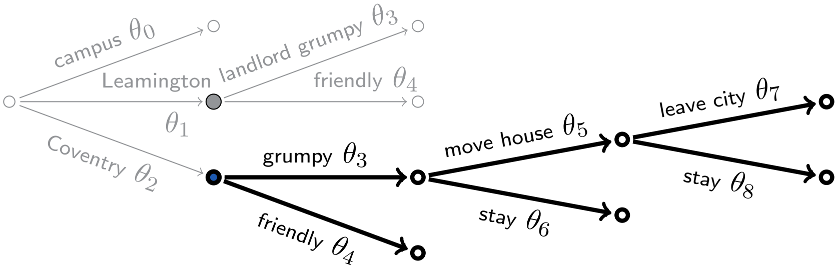
Manipulate a tree using a **differentiation operation** on this polynomial:



$$c_{\mathcal{T}}(\boldsymbol{\theta}) = \theta_0 + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_3\theta_5\theta_7 + \theta_2\theta_3\theta_5\theta_8 + \theta_2\theta_3\theta_6 + \theta_2\theta_4$$

Interventions (symbolically)

Manipulate a tree using a **differentiation operation** on this polynomial:

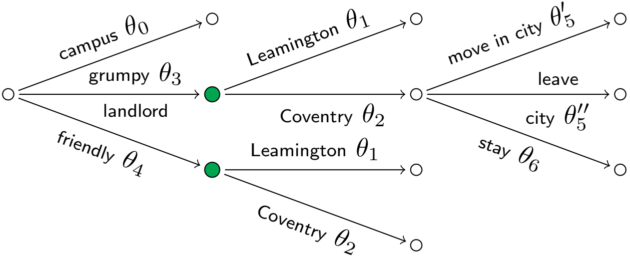


$$c_{\mathcal{T}}(\boldsymbol{\theta}) = \theta_0 + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_3\theta_5\theta_7 + \theta_2\theta_3\theta_5\theta_8 + \theta_2\theta_3\theta_6 + \theta_2\theta_4$$

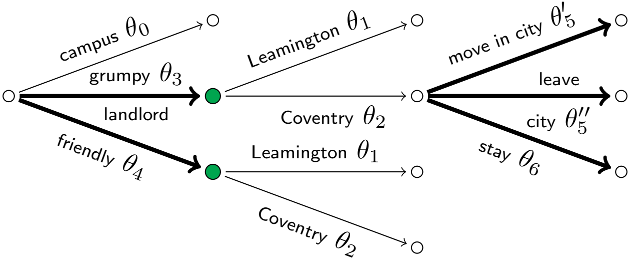
$$\frac{\partial}{\partial\theta_2} c_{\mathcal{T}}(\boldsymbol{\theta}) = \theta_3\theta_5\theta_7 + \theta_3\theta_5\theta_8 + \theta_3\theta_6 + \theta_4$$

Advantages of the symbolic approach

Advantages of the symbolic approach



Advantages of the symbolic approach



Interventions on the polynomial are more general than vertex manipulations!

Local differentiation operation

Replace a staged tree by a polynomial

$$c_{\mathcal{T}}(\boldsymbol{\theta}) = \sum_{(v_0,v_1)\in E(v_0)} \theta(v_0,v_1) \left(\sum_{(v_1,v_2)\in E(v_1)} \theta(v_1,v_2) \left(\cdots \left(\sum_{(v_{k-1},v_k)\in E(v_{k-1})} \theta(v_{k-1},v_k) \right) \right) \right)$$

Local differentiation operation

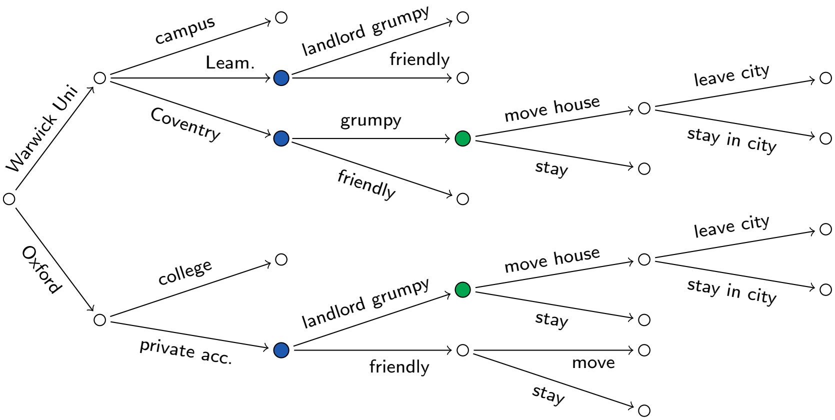
Replace a staged tree by a polynomial

$$c_{\mathcal{T}}(\boldsymbol{\theta}) = \sum_{(v_0,v_1)\in E(v_0)} \theta(v_0, v_1) \left(\sum_{(v_1,v_2)\in E(v_1)} \theta(v_1, v_2) \left(\cdots \left(\sum_{(v_{k-1},v_k)\in E(v_{k-1})} \theta(v_{k-1}, v_k) \right) \right) \right)$$

and perform a local differentiation on that:

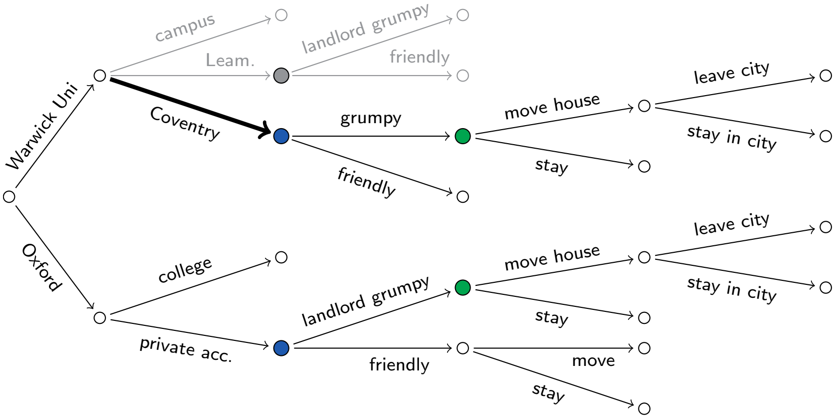
$$\sum_{(v_0,v_1)\in E(v_0)} \theta(v_0, v_1) \left(\cdots \left(\frac{\partial}{\partial \theta_j^\star} \sum_{(v_{j-1},v_j)\in E(v_j)} \theta(v_{j-1}, v_j) \left(\cdots \left(\sum_{(v_{k-1},v_k)\in E(v_{k-1})} \theta(v_{k-1}, v_k) \right) \right) \right) \right)$$

Example: two interventions



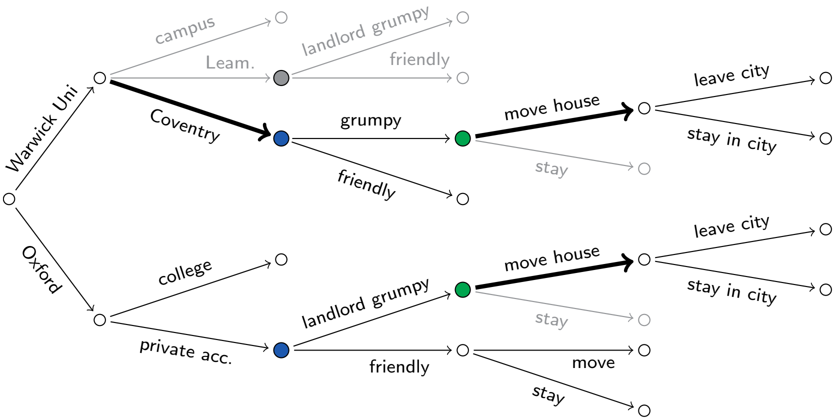
Example: two interventions

Differentiate locally on florets:



Example: two interventions

Differentiate locally on florets:



Advantages of this new differentiation

- can do sequence of interventions

Advantages of this new differentiation

- can do sequence of interventions
- does not rely on graphical representation $\rightsquigarrow$ flexible and very general

Advantages of this new differentiation

- can do sequence of interventions
- does not rely on graphical representation $\rightsquigarrow$ flexible and very general
- used only algebraic description of a parametric model: method can be used in models far more general than staged trees

Thank you very much for your attention!

JMLR: Workshop and Conference Proceedings vol 52, 207-215, 2016 PGM 2016

A Differential Approach to Causality in Staged Trees

Christiane Görgen C.GORGEN@WARWICK.AC.UK
Jim Q. Smith J.Q.SMITH@WARWICK.AC.UK
Department of Statistics
University of Warwick
Coventry CV4 7AL, United Kingdom

Abstract

In this paper, we apply a recently developed differential approach to inference in staged tree models to causal inference. Staged trees generalise modelling techniques established for Bayesian networks (BN). They have the advantage that they can depict highly nuanced structure impossible to express in a BN and also enable us to perform causal manipulations associated with very general types of interventions on the system. Conveniently, what we call the interpolating polynomial of