A Differential Approach for Staged Trees

Christiane Görgen, Manuele Leonelli and James Q. Smith

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

Abstract Symbolic inference algorithms in Bayesian networks have now been applied in a variety of domains. These often require the computation of the derivatives of polynomials representing probabilities in such graphical models. In this paper we formalise a symbolic approach for staged trees, a model class making it possible to visualise asymmetric model constraints. We are able to show that the probability parametrisation associated to trees has several advantages over the one associated to Bayesian networks. We then continue to compute certain derivatives of staged trees' polynomials and show their probabilistic interpretation. We are able to determine that these polynomials can be straightforwardly deduced by compiling a tree into an arithmetic circuit.

1 Introduction

The notion of probabilistic graphical models has been successfully established [10]. In particular, Bayesian networks (BNs) [13] have proved to provide an intuitive qualitative framework, based on various conditional independence constraints [8], as well as a computationally efficient inferential tool [11].

Probabilistic inference in BNs has been characterised in the literature not only using numerical approaches but also symbolic methods, where probabilities are treated as unknown quantities [5,7]. Symbolic approaches like these provide a natural framework around which to perform various sensitivity analyses. It has only recently been recognised that a variety of such probabilistic queries can be answered by computing derivatives of polynomials representing the model's probabilities [7]. In [7] it is further shown that the computational burden of calculating these polynomials can be reduced through an arithmetic circuit (AC) representation.

Symbolic methods have proved useful in BNs (e.g. [5]), although these techniques do come with a considerable computational cost. In this paper we study a different class of models called *staged trees* [18,19] where such difficulties are eased. We demonstrate that the *interpolating* polynomial [7,15] associated to a staged tree can be straightforwardly deduced by simply looking at the structure of the underlying graph. This is because the parametrisation associated to these models is more intuitive than the one of BNs.

It has been shown that in fact discrete BNs are a special case of the class of staged tree models [2,18,19]. The latter have the advantage over BNs of being able to explicitly represent asymmetric (conditional independence) constraints

Paper No. 15-06, www.warwick.ac.uk/go/crism

and relations between functions of random variables, explicitly modelling information which is only present in the probability structure of a BN model. Importantly, polynomials arising from this more general class of models have an interesting algebraic structure which is not necessarily homogeneous and multilinear as in the BN case. We are able to demonstrate that a probabilistic semantic can be attributed to the partial derivatives of interpolating polynomials. In addition, these can also be used to represent various causal assumptions under the Pearlean causal paradigm [14]. Typically, because of the wide variety of possible hypotheses they embody, staged trees are necessarily models over much smaller state spaces than BNs. Since this is the main computational issue for symbolic approaches associated with BNs, it follows that trees can be very practical for investigating inferential queries.

The polynomials of staged trees can be computed by compiling them into ACs just as for BNs. As noted in [12], the presence of asymmetries simply entails setting equal to zero some terms in the polynomial associated with a model with no such asymmetries. Therefore, the AC of a staged tree has often a substantially smaller number of leaves. Together with the point above this means that, when using a symbolic approach for our model class, computations and inferential challenges are therefore eased.

2 Staged Tree Models

In this paper, as in [17,19], we focus on graphical models represented by trees. We examine *event trees* $\mathcal{T} = (V, E)$, directed rooted trees where each inner vertex $v \in V$ has at least two children. In this context, the sample space of the model corresponds to the *set of root-to-leaf paths* in the graph and each directed path, which is a sequence of edges $r = (e \mid e \in E(r))$, for $E(r) \subset E$, has a meaning in the modelling context. To every edge $e \in E$ we associate a *primitive probability* $\theta(e) \in (0, 1)$ such that on each *floret* $\mathcal{F}(v) = (v, E(v))$, where $E(v) \subseteq E$ is the set of edges emanating from $v \in V$, the primitive probabilities sum to unity. The probability of an atom is then simply the product of primitive probabilities along the edges of its path: $\pi_\theta(r) = \prod_{e \in E(r)} \theta(e)$. After [6,19] we define:

Definition 1. *Let $\theta_v = (\theta(e) \mid e \in E(v))$ be the vector of primitive probabilities associated to the floret $\mathcal{F}(v)$, $v \in V$, in a tree $\mathcal{T} = (V, E)$. A* staged tree *is an event tree as above where, for some $v, w \in V$, the floret probabilities are identified $\theta_v = \theta_w$. Then, $w, v \in V$ are in the same* stage.

Setting floret probabilities equal can be thought of as representing conditional independence information. If vertices are linked to random variables [19,20] their edges are associated with a projection of the model's sample space. Two vertices are thus in the same stage if they have the same (conditional) distribution over their edges. When drawing a tree, vertices in the same stage are assigned the same colour in order to have a visual counterpart for that information.

Staged trees are flexible representations for many discrete models. They are capable of representing all conditional independence hypotheses within discrete

BNs, whilst at the same time being more flexible in expressing modifications of these, as we will see below. In particular, the graphical complexity is made up for by the extra expressiveness of these models [19]. In this paper, although the associated *Chain Event Graph (CEG)* is more convenient for displaying information in a staged tree model, we will stick to the latter graphs when representing their algebraic features.

Example 1. For the purposes of this short paper, we consider the following simplification of a real system described in [19]. A binary model is designed to explain a possible unfolding of the following events in a cell culture: a cell finds itself in a benign or hostile environment, the level of activity within this might be high or low, and if the environment is hostile then a cell might either survive or die.

We can model this narrative using a BN on three variables: the state of the environment is represented by Y_0 taking values in $\mathbb{Y}_0 = \{\text{hostile}, \text{benign}\}$, cell activity is measured by Y_1 as $\mathbb{Y}_1 = \{\text{high}, \text{low}\}$ and viability via Y_2 with $\mathbb{Y}_2 = \{\text{die}, \text{survive}\}$. Then $\mathbb{Y} = (\mathbb{Y}_0, \mathbb{Y}_1, \mathbb{Y}_2)$ is the model space.

If we argue that a high or low level of activity is independent of the environment being hostile or benign and that whether or not a cell dies does not depend on its activity, then our model corresponds to the collider BN in (1), stating that $Y_0 \perp\!\!\!\perp Y_1$ and $Y_0 \not\perp\!\!\!\perp Y_1 \mid Y_2$.

$$Y_0 \longrightarrow Y_2 \longleftarrow Y_1 \tag{1}$$

Observe that this graphical representation, though storing all conditional independence constraints between the Y_i variables, does not inform us about all of the assumptions above. It forces us to retain information which is meaningless in our context, as for instance the atom $\omega = (\text{benign}, \text{high}, \text{die}) \in \mathbb{Y}$ which has probability zero. The representation of (1) in terms of a staged tree $\mathcal{T}_{\text{BN}}$ in Fig. 1, where each root-to-leaf path represents one $\omega \in \mathbb{Y}$, is therefore large. As the number of variables gets larger, the percentage of information not described through the graph can increase dramatically.

The apparent symmetries in this representation are typical for event trees induced by BNs: all paths are of the same length and the stage structure (colouring) depends on the distance of a vertex from the root. Keeping in mind the assumptions made in our model, for example that there is no cell damage in a benign environment, we notice that the lower part of the tree in Fig. 1 does not contain any valuable information. There is even more redundancy if we add an extra level of complexity to the model, for instance by assessing the constitution of a surviving cell—which is meaningless if a unit has died. Thus, the model at hand is a context specific BN rather than a BN (see e.g. [19]), and there is a strong case for using a staged tree model.

We call the state space of our improved graphical model

$$\mathbb{X}_{\mathcal{T}} = \{\omega_1 = (\text{hostile, high, die}), \ldots, \omega_8 = (\text{benign, low})\},$$

which is the set of all meaningful unfoldings of events. It is canonically identified with the set of root-to-leaf paths of an event tree $\mathcal{T} = (V, E)$,

$$R_{\mathcal{T}} = \{r_1 = (e_{01}, e_{11}, e_{31}), r_2 = (e_{01}, e_{11}, e_{32}), \ldots, r_8 = (e_{02}, e_{22})\},$$

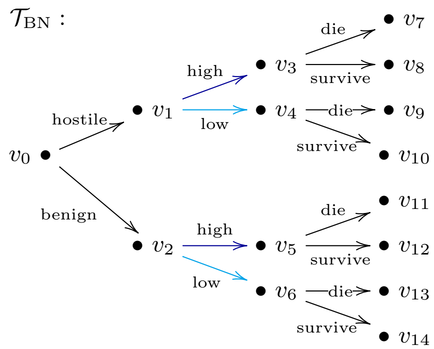

Figure 1. A staged tree $\mathcal{T}_{\text{BN}}$ representation of the BN in (1) of Ex. 1.

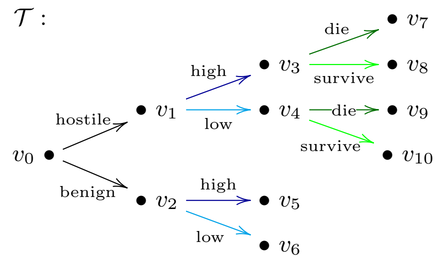

Figure 2. An asymmetric staged tree $\mathcal{T}$ representing the context specific information of the BN in (1) of Ex. 1.

where we reduce the vertex set of $\mathcal{T}_{\text{BN}}$ to $V = \{v_0, v_1, \dots, v_{10}\}$ and the edges are $E = \{e_{01}, e_{02}, \dots, e_{42}\}$, with e_{ij} corresponding to the jth edge emanating from v_i, for fitting i and j.

Following this approach, we obtain the staged tree $\mathcal{T}$ in Fig. 2. This new representation is far more expressive than the BN itself and less cluttered than the BN's tree $\mathcal{T}_{\text{BN}}$, whilst conveying the same information: the colouring expresses the given conditional independence assumptions that can also be read from the BN. For instance, by colouring the edges in $E(v_1)$ and $E(v_2)$ in the same manner, we visualise equality of the probability labels

$$\begin{aligned} &\theta(e_{11}) = \theta(e_{21}) \text{ or } P(Y_1 = \text{high} | Y_0 = \text{hostile}) = P(Y_1 = \text{high} | Y_0 = \text{benign}), \\ &\theta(e_{12}) = \theta(e_{22}) \text{ or } P(Y_1 = \text{low} | Y_0 = \text{hostile}) = P(Y_1 = \text{low} | Y_0 = \text{benign}). \end{aligned} \tag{2}$$

The same procedure is applied on the edges of v_7, v_9 and v_8, v_{10}. □

Having understood the advantages of a staged tree over a BN, we now present a symbolic approach to calculate probabilities in this type of models. Following concepts introduced in [15] in the context of designed experiments, we define:

Definition 2. *Let $\mathcal{T} = (V, E)$ be a staged tree with primitive probabilities $\theta(e)$, $e \in E$, and set of root-to-leaf paths $R_{\mathcal{T}}$. We call $\Lambda(e) = \{r \in R_{\mathcal{T}} \mid e \in E(r)\}$ an* edge-centred *event, and set $\lambda_e(r)$, for $e \in E$, to be an indicator of $r \in \Lambda(e)$. We call*

$$c_{\mathcal{T}}(\theta, \lambda) \;=\; \sum_{r \in R_{\mathcal{T}}} \pi_\theta(r) \prod_{e \in E(r)} \lambda_e(r) \;=\; \sum_{r \in R_{\mathcal{T}}} \prod_{e \in E(r)} \lambda_e(r) \theta(e)$$

the interpolating polynomial *of $\mathcal{T}$.*

The interpolating polynomial is a sum of atomic probabilities with indicators for certain conditional events happening or not happening. Even though all these unknowns sum to one, in our symbolic approach we treat them just like indeterminates. We will report in [9] some recent results that use interpolating polynomials to characterise when two staged trees are statistically equivalent.

We now look at this model class from an algebraic point of view. As seen in Ex. 1, the sample space of a BN with vertex set $\{Y_1, \ldots, Y_n\}$, $Y_i \in \mathbb{Y}_i$, $i = 1, \ldots, n$, gives rise to an event tree where each root-to-leaf path $r \in R_{\mathcal{T}}$ is associated to an atom $\omega \in \mathbb{Y}_1 \times \ldots \times \mathbb{Y}_n$ and is hence of length n. By definition, $P(\omega) = \pi_\theta(r) = \prod_{r \in E(r)} \theta(e)$ and therefore the interpolating polynomial of a BN is a sum of monomials each of which is of degree $2n$ and so *homogeneous*. Moreover, the stage structure of a BN tree as in Fig. 1 is such that no two vertices along the same directed path are in the same stage, in fact stages exist only along *orthogonal cuts* [20]. Thus in particular, the interpolating polynomial of a BN is also *multilinear*, that is linear in all components. Note that, although in this paper we consider Bayesian subjective probabilities only, other representations of uncertainty in directed graphical models entertain similar multilinear structures (see e.g. [1]).

Note that the indicators $\lambda_e(r)$ on the edges $e \in E(r)$ are associated to the (conditional) event represented by e, having probability $\theta(e)$. This notation is apparently redundant, but will turn out to be useful in Sect. 3. We observe that this redundancy is one of the great advantages of a staged tree: whilst [7] needs to compute conditional probabilities of all *compatible parent structures* of an event, which is a rather obscure concept in a symbolic framework, and [5] computes the product space of any indeterminates' combination regardless of their meaning, a tree visualisation of our model gives us the necessary structure immediately: events can be simply read from the paths in the graph. Recently, [12] developed an algorithm which automatically computes only the required monomials in BN models. Although this makes computations more efficient the parametrisation in [12] is still not as transparent as the one associated to trees.

Example 2. Recall the model analysed in Ex. 1. Ignoring the equalities implied by the stage structure, the interpolating polynomial of a model represented by the BN in (1) or the tree in Fig. 1 equals

$$\begin{aligned} c_{\mathrm{BN}}(\theta) \;=\; & \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{11}\theta_{32} + \theta_{01}\theta_{12}\theta_{41} + \theta_{01}\theta_{12}\theta_{42} \\ & + \theta_{02}\theta_{21}\theta_{51} + \theta_{02}\theta_{22}\theta_{52} + \theta_{02}\theta_{22}\theta_{61} + \theta_{02}\theta_{22}\theta_{62}, \end{aligned} \tag{3}$$

where we simplified our notation to $\theta_{ij} = \theta(e_{ij})$ for each i, j. We also omitted for ease of notation the indicator functions on all terms. This polynomial has been simply read from the event tree by first multiplying over all primitive probabilities along one root-to-leaf path, and then summing over all of these paths. This is a lot easier done using Fig. 1 than in (1), where we would have had to sum over compatible parent configurations, which could have not been read directly from a DAG. Observe that here, as outlined above, c_{BN} is homogeneous of degree 3. The number of terms equals the number of paths in the tree representation.

Conversely, the more adequate improved model without meaningless terms or terms with probability zero has the interpolating polynomial

$$c(\theta) \;=\; \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{11}\theta_{32} + \theta_{01}\theta_{12}\theta_{41} + \theta_{01}\theta_{12}\theta_{42} + \theta_{02}\theta_{21} + \theta_{02}\theta_{22}. \tag{4}$$

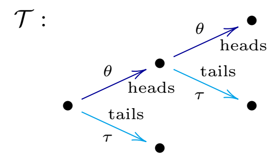

Figure 3. The staged tree of a repeated coin toss with interpolating polynomial (7).

This is a lot easier to handle than c_{BN} but still conveys exactly the same information. When plugging in the conditional independence constraints as in (2), we obtain the interpolating polynomial of the staged tree in Fig. 2 as:

$$\begin{aligned} c_{\mathcal{T}}(\theta) &= \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{11}\theta_{32} + \theta_{01}\theta_{12}\theta_{31} + \theta_{01}\theta_{12}\theta_{32} + \theta_{02}\theta_{11} + \theta_{02}\theta_{12}, \quad (5) \\ &= \theta_{01}(\theta_{11}(\theta_{31} + \theta_{32}) + \theta_{12}(\theta_{31} + \theta_{32})) + \theta_{02}(\theta_{11} + \theta_{12}), \quad (6) \end{aligned}$$

where we substituted $\theta_{1j} = \theta_{2j}$ and $\theta_{3j} = \theta_{4j}$, for $j = 1, 2$. This is now inhomogeneous but still multilinear, and it has total degree 3 with individual monomial terms having degree 2 or 3. Notice that $c_{\mathcal{T}}$ can be easily factorised in (6) by simply following the structure of the underlying graph [9]. In [4], polynomials of this type are called *factored*. This representation entails great computational advantages since the compilation into an AC is almost instantaneous. Whilst for BNs the factored representation might be difficult to obtain, it comes almost for free in tree models.

We observe that the graphical simplicity of a staged tree model in comparison to an uncoloured tree or a BN is also reflected algebraically: the polynomial in (5) has fewer indeterminates than the one in (4) and a lot fewer than the polynomial associated to a tree which is derived from a BN in (3). This is because in the BN the redundancy of atoms gives rise to redundant terms. □

Observe that, although the interpolating polynomial of the staged tree in Ex. 2 is multilinear, the concept of stages allows for enough flexibility to construct models where this is not the case. Suppose we are interested in a situation where we flip a coin and repeat this experiment only if the first outcome is heads. This is depicted graphically by the coloured tree in Fig. 3. The interpolating polynomial of this model is non-homogeneous and not multilinear:

$$c_{\mathcal{T}}(\theta, \tau, \lambda, \lambda') = \lambda^2\theta^2 + \lambda\lambda'\theta\tau + \lambda'\tau, \quad (7)$$

for $\theta + \tau = 1$ and indicators λ of 'heads' and λ' of 'tails'. Again, this algebraic structure and model type cannot, without significant obscuration, be expressed in terms of a BN. If the polynomial is multilinear, we call our model *square-free*. The focus of this paper lies on these.

By construction, Theorem 1 of [7] holds for a staged tree interpolating polynomial:

Lemma 1. *For any event A represented by a set of root-to-leaf paths R_A in a staged tree $\mathcal{T}$, we know that*

$$P(A) \;=\; \sum_{r\in R_A}\pi_\theta(r) \;=\; \sum_{r\in R_A}\prod_{e\in E(r)}\lambda_e(r)\theta(e) = c_{\mathcal{T}}(\theta,\lambda|_{R_A}),$$

where $\lambda|_{R_A}$ indicates that $\lambda_e(r) = 1$ for all $e \in E(r)$ with $r \in R_A$, and else zero.

We are therefore able to symbolically compute the probability of any event associated to a tree.

Example 3. In the notation of Examples 1 and 2, suppose we are interested in calculating the probability of death of a cell. This is captured by the event $A = \{x \in \mathbb{X}_{\mathcal{T}} \mid x_3 = \text{die}\}$. Thus $R_A = \Lambda(e_{31}) \cup \Lambda(e_{41}) = \{r_1, r_2\}$ corresponds to all root-to-leaf paths going through an edge labelled 'die' which translates in summing all terms in (5) which include the label θ_{31}. Therefore, again omitting the λ indicators, $P(A) = \sum_{r\in R_A}\pi_\theta(r) = \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{12}\theta_{31}$. □

3 The Differential Approach

We are now able to provide a probabilistic semantic, just as [7] for BNs, to the derivatives of polynomials associated to staged trees. For ease of notation we let in this section $\lambda_e = \lambda_e(r)$.

Proposition 1. *For equally coloured edges $e \in E$ and an event A represented by the root-to-leaf paths R_A, the following results hold:*

$$P(\Lambda(e)|A) = \frac{1}{c_{\mathcal{T}}(\theta,\lambda|_{R_A})}\frac{\partial c_{\mathcal{T}}(\theta,\lambda|_{R_A})}{\partial \lambda_e},\;\; P(\Lambda(e), A) = \theta(e)\frac{\partial c_{\mathcal{T}}(\theta,\lambda|_{R_A})}{\partial \theta(e)}, \quad (8)$$

where $\Lambda(e)$ is an edge-centred event.

All the probabilities in (8) are equal to zero whenever $e \notin E(r)$ for all $r \in R_A$. Notice that the derivatives of tree polynomials have the exact same interpretation of the ones of BNs as in [7]. Here we restricted our attention to square-free staged trees but analogous results hold in the generic case: each monomial with indeterminate λ_e and $\theta(e)$ of degree higher than one would need to be differentiated a number of times equal to the degree of that indeterminate.

Proposition 2. *In the notation of Prop. 1, we have that for $e, e_1, e_2 \in E$:*

$$P(\Lambda(e_1), \Lambda(e_2) \mid A) \;=\; \frac{1}{c_{\mathcal{T}}(\theta,\lambda|_{R_A})}\frac{\partial^2 c_{\mathcal{T}}(\theta,\lambda|_{R_A})}{\partial \lambda_{e_1}\partial \lambda_{e_2}}, \quad (9)$$

$$P(\Lambda(e_1), \Lambda(e_2), A) \;=\; \theta(e_1)\theta(e_2)\frac{\partial c_{\mathcal{T}}(\theta,\lambda|_{R_A})}{\partial \theta(e_1)\partial \theta(e_2)}, \quad (10)$$

$$P(A \mid \Lambda(e)) \;=\; \frac{\partial^2 c_{\mathcal{T}}(\theta,\lambda|_{R_A})}{\partial \theta(e)\partial \lambda_e}. \quad (11)$$

It is an easy exercise to deduce from Prop. 2 the probabilistic meaning of higher order derivatives.

The above propositions demonstrate that the results of [7] are transferable to the class of staged trees. In addition we are able to derive that in the staged tree model class derivatives can be associated to causal propositions in the sense of the Perlean concept of causal intervention on trees, as formalised in [21]. Note that such a result does not hold in general for the polynomials describing BN probabilities.

Proposition 3. *Suppose the staged tree is believed to be causal as in [18]. Then under the notation of Prop. 2,*

$$P(A \mid\mid \Lambda(e)) = \frac{\partial^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\theta_e \partial\lambda_e} \tag{12}$$

is the probability of the event A when the system is forced to go through edge e.

Note that all the quantities in (8)–(12) can be used in sensitivity analysis, for instance by investigating the changes in probability estimates when the system is set to be in a certain scenario of interest.

Example 4. We now compute a set of derivatives on the interpolating polynomial $c_{\mathcal{T}}$ in (5) with respect to λ_{31} and θ_{31} to perform probabilistic inference over the event A that a cell dies, as in Ex. 3. Thus, we consider the edge $e = (v_3, v_7)$ and

$$\frac{1}{c_{\mathcal{T}}(\theta, \lambda|_{R_A})} \frac{\partial c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\lambda_e} = \frac{\theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{12}\theta_{31}}{\theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{12}\theta_{31}} = 1, \tag{13}$$

$$\theta(e) \frac{\partial c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\theta(e)} = \theta_{13}(\theta_{01}\theta_{11} + \theta_{01}\theta_{12}) = P(A), \tag{14}$$

$$\frac{\partial^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\theta(e)\partial\lambda_e} = \theta_{01}\theta_{11} + \theta_{01}\theta_{12}. \tag{15}$$

Observe that (13) is equal to unity since every path associated to the event A must go through e. From the same argument follows that (14) is equal to $P(A)$. Eq. (15) is a simple consequence of Bayes' theorem, which can be checked algebraically. □

4 Trees as Circuits

The previous sections have introduced a comprehensive symbolic inferential toolbox for trees, based on the computation of the interpolating polynomial and its derivatives. In [7] it is shown that an efficient method to compute such polynomials is by representing them as an *AC*. This is a DAG whose leaves are the indeterminates and the inner nodes are labelled by multiplication and summation operations. The *size* of the circuit equals its number of edges.

ACs of staged tree polynomials are smaller in size than the ones associated to BNs for two reasons: first, a tree might have fewer root-to-leaf paths (as in Ex. 1);

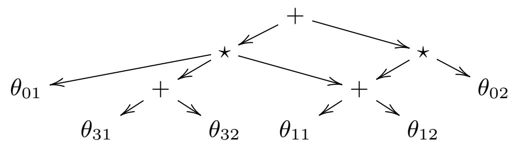

Figure 4. The arithmetic circuit of the model represented in Ex. 2, yielding (5).

second, there can be less indeterminates because unfoldings with probability zero are not included in the model and coloured labels further decrease the number of indeterminates. Therefore, in asymmetric settings we can expect computations to be much faster for trees than for BNs.

A major problem in the compilation of BN polynomials consists in the identification of the AC of smallest size. This usually entails the computation of the BN's jointree and the application of more complex algorithms [7]. We note here that in tree models this is straightforward since the interpolating polynomial is naturally factored.

Example 5. Recall the interpolating polynomial of the staged tree from Ex. 2. We notice that (6) can be rewritten as $c_{\mathcal{T}}(\theta) = \theta_{01}(\theta_{11}+\theta_{12})(\theta_{31}+\theta_{31})+\theta_{02}(\theta_{11}+\theta_{12})$.

This gives us the AC in Fig. 4 where leaves with the same parent are labelled by primitive probabilities from the same floret, and labels belonging to leaves in the tree are first summed in the AC. It is easy to deduce that the AC associated to the BN's polynomial in (3) would be much larger than the one in Fig. 4. We note also that, whilst all the ACs deriving from BNs in [7] are trees, ours is more generally a DAG. This is a consequence of the more flexible stage structure of generic staged trees than the one of trees depicting BNs. □

5 Discussion

Staged tree models, whilst representing a much larger model class than discrete BNs, have proven to have a much more intuitive symbolic representation. We have been able to show that in this framework polynomial derivatives have a probabilistic semantic which is of use in sensitivity analysis. Our parametrisation further led to computational advantages because of the almost automatic compilation into an AC.

Importantly, this paper relates the symbolic definition of discrete BNs to the one of generic trees via the notion of an interpolating polynomial introduced in Def. 2. We can therefore now start investigating classes of models that are defined only symbolically, since the interpolating polynomial is able to capture all the probabilistic information of the model. This can then lead to the definition of new models that in general cannot be depicted by a graph.

In addition, the recognition that the probabilities associated to certain statistical models have a polynomial form started a whole new area of research called *algebraic statistics* [16]. We are now developing results which apply new exciting methodologies from this subject to staged tree models. We are also starting to develop computer algebra methods to work with trees that exploit the symbolic definition of the model we provided here and that will facilitate the use of such models in practical applications. The examples we work with are of course larger than those presented here (see [2,3]) and provide the framework for sensitivity analyses in important areas of application.

References

1. Antonucci, A., De Campos, C.P., Huber, D., Zaffalon, M.: Approximating credal network inferences by linear programming. In: ECSQARU, pp. 13–24. Springer (2013)
2. Barclay, L.M., Hutton, J.L., Smith, J.Q.: Refining a Bayesian Network using a Chain Event Graph. Int. J. Approx. Reason. 54, 1300–1309 (2013)
3. Barclay, L.M., Hutton, J.L., Smith, J.Q.: Chain event graphs for informed missingness. Bayesian Analysis 9(1), 53 – 76 (2014)
4. Brandherm, B., Jameson, A.: An extension of the differential approach for Bayesian network inference to dynamic Bayesian networks. Int. J. Intell. Syst. 19(8), 727–748 (2004)
5. Castillo, E., Gutiérrez, J.M., Hadi, A.S.: A New Method for Efficient Symbolic Propagation in Discrete Bayesian Networks. Networks 28(1), 31–43 (1996)
6. Cowell, R.G., Smith, J.Q.: Causal discovery through MAP selection of stratified Chain Event Graphs. Electron. J. Stat. 8, 965–997 (2014)
7. Darwiche, A.: A differential approach to inference in Bayesian networks. Journal of ACM 50(3) (2003)
8. Dawid, A.P.: Conditional independence in statistical theory. J. Roy. Stat. Soc. B 41(1), 1–31 (1979)
9. Görgen, C., Smith, J.Q.: Equivalence Classes of Chain Event Graph Models. In preparation.
10. Jordan, M.I.: Graphical models. Stat. Sci. pp. 140–155 (2004)
11. Lauritzen, S.L., Spiegelhalter, D.J.: Local computations with probabilities on graphical structures and their application to expert systems. J. Roy. Stat. Soc. B pp. 157–224 (1988)
12. Leonelli, M., Smith, J.Q., Riccomagno, E.: Using computer algebra for the symbolic evaluation of discrete influence diagrams. Tech. rep., CRISM (2015)
13. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann (1988)
14. Pearl, J.: Causality: models, reasoning and inference. CUP (2000)
15. Pistone, G., Riccomagno, E., Wynn, E.P.: Gröbner bases and factorisation in discrete probability and Bayes. Stat. Comput. 11, 37–46 (2001)
16. Riccomagno, E.: A short history of algebraic statistics. Metrika 69(2-3), 397–418 (2009)
17. Shafer, G.: The Art of causal Conjecture. MIT Press, Cambridge (1996)
18. Smith, J.Q.: Bayesian decision analysis: principles and practice. CUP (2010)
19. Smith, J.Q., Anderson, P.E.: Conditional independence and Chain Event Graphs. Artif. Intell. 172, 42–68 (2008)
20. Thwaites, P.A., Smith, J.Q.: Separation theorems for Chain Event Graphs. CRiSM 11-09 (2011)
21. Thwaites, P.A., Smith, J.Q., Riccomagno, E.: Causal Analysis with Chain Event Graphs. Artif. Intell. 174, 889–909 (2010)