Locally specified credal networks

Alessandro Antonucci and Marco Zaffalon
Istituto "Dalle Molle" Di Studi Sull'Intelligenza Artificiale
CH-6928 Manno (Lugano), Switzerland
{alessandro,zaffalon}@idsia.ch

Abstract

Credal networks are models that extend Bayesian nets to deal with imprecision in probability, and can actually be regarded as sets of Bayesian nets. Evidence suggests that credal nets are a powerful means to represent and deal with many important and challenging problems in uncertain reasoning. We give examples to show that some of these problems can only be modelled by credal nets called *non-separately specified*. These, however, are still missing a graphical representation language and solution algorithms. The situation is quite the opposite with separately specified credal nets, which have been the subject of much study and algorithmic development. This paper gives two major contributions. First, it delivers a new graphical language to formulate any type of credal network, both separately and non-separately specified. Second, it shows that *any* non-separately specified net represented with the new language can be easily transformed into an equivalent separately specified net, defined over a larger domain. This result opens up a number of new perspectives and concrete outcomes: first of all, it immediately enables the existing algorithms for separately specified credal nets to be applied to non-separately specified ones.

1 Introduction

We focus on credal networks (Section 3) (Cozman, 2005), which are a generalization of Bayesian nets. The generalization is achieved by relaxing the requirement that the conditional mass functions of the model be precise: with credal nets each of them is only required to belong to a closed convex set. Closed convex sets of mass functions are also known as *credal sets* after Levi (Levi, 1980). Using credal sets in the place of mass functions makes credal networks an *imprecise probability* model (Walley, 1991). It can be shown that a credal network is equivalent to a *set of Bayesian nets* with the same graph.

An important question is whether or not all credal networks can be represented in a way that emphasizes locality. The answer is positive if we restrict the attention to the most popular type of credal networks, those called *separately specified* (Section 4). In this case, each conditional mass function is allowed to vary in its credal set independently of the others. The representation is naturally local because there are no relationships between different credal sets. The question is more complicated with more general specifications of credal networks, which we call *non-separately specified*. The idea of non-separately specified credal nets is in fact to allow for relationships between conditional mass functions in different credal sets, which can also be far away in the net.

Although the idea of non-separately specified credal nets is relatively intuitive, it should be stressed that this kind of nets has been investigated very little: in fact, there has been no attempt so far to develop a general graphical language to describe them; and there is no algorithm to compute with them.[1] This appears

[1] An exception is the classification algorithm developed for the *naive credal classifier* (Zaffalon, 2001), but it is ad hoc for a very specific type of network. More generally speaking, it is not unlikely that some of the

to be an unfortunate gap in the literature as the non-separate specification seems to be the key to model many important problems, as illustrated in Section 6. Separately specified credal nets, on the other hand, have been the subject of much algorithmic development (Cozman, 2005).

In this paper we give two major contributions. First, we define a unified graphical language to locally specify credal networks in the general case (Section 5). The new representation is inspired, via the CCM transformation (Cano et al., 1994), by the formalism of influence diagrams, and more generally of decision graphs (Zhang et al., 1993). In this language the graph of a credal net is augmented with control nodes that express the relationships between different credal sets. We give examples to show that the new language provides one with a natural way to define non-separately specified nets; and we give a procedure to reformulate any separately specified net in the new language.

Second, we make a very simple observation (Section 7), which has surprisingly powerful implications: we show that for any credal network specified with the new language there is a separately specified credal network, defined over a larger domain, which is equivalent. The procedure to transform the former into the latter network is very simple, and takes only linear time. The key point is that this procedure can be used as a tool to "separate" the credal sets of non-separately specified nets. This makes it possible to model, by separately specified nets, problems formerly modelled by non-separately specified ones; and hence to use *any* (both exact and approximate) existing algorithm for separately specified nets to solve such problems.

Some comments on this result and perspectives for future developments are discussed in Section 8. The more technical parts of this paper are collected in Appendix A. ————-

existing algorithms for separately specified nets can be extended to special cases of non-separate specification, but we are not aware of any published work dealing with this issue.

2 Basic notation and Bayesian nets

Let us first define some notation and the fundamental notion of Bayesian network. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a collection of random variables, which take values in finite sets, and $\mathcal{G}$ a directed acyclic graph (DAG), whose nodes are associated to the variables in $\mathbf{X}$. For each $X_i \in \mathbf{X}$, Ω_{X_i} is the possibility space of X_i, x_i a generic element of Ω_{X_i}, $P(X_i)$ a mass function for X_i and $P(x_i)$ the probability that $X_i = x_i$. The parents of X_i, according to $\mathcal{G}$, are denoted by the joint variable Π_i, whose possibility space is Ω_{Π_i}. For each $\pi_i \in \Omega_{\Pi_i}$, $P(X_i|\pi_i)$ is the conditional mass function for X_i given the joint value π_i of the parents of X_i. This formalism is sufficient to properly introduce the following:

Definition 1. A Bayesian network (BN) over $\mathbf{X}$ is a pair $\langle \mathcal{G}, \mathbb{P} \rangle$ such that $\mathbb{P}$ is a set of conditional mass functions $P(X_i|\pi_i)$, one for each $X_i \in \mathbf{X}$ and $\pi_i \in \Omega_{\Pi_i}$.

We assume the *Markov condition* to make $\mathcal{G}$ represent probabilistic independence relations between the variables in $\mathbf{X}$: every variable is independent of its non-descendant non-parents conditional on its parents. Thus, a BN determines a joint mass function $P(\mathbf{X})$ according to the following factorization:

$$P(\mathbf{x}) = \prod_{i=1}^{n} P(x_i|\pi_i), \tag{1}$$

for each $\mathbf{x} \in \Omega_{\mathbf{X}}$, where for each $i = 1, \ldots, n$ the values (x_i, π_i) are consistent with $\mathbf{x}$.

3 Credal sets and credal networks

Credal networks extend Bayesian nets to deal with imprecision in probability. This is obtained by means of closed convex sets of probability mass functions, which are called *credal sets* (Levi, 1980). We follow (Cozman, 2000) in considering only finitely generated credal sets, i.e., obtained as the convex hull of a finite number of mass functions. Geometrically, a credal set is a *polytope*. A credal set contains an infinite number of mass functions, but only a finite number of *extreme mass functions*: those corresponding to the *vertices* of the polytope, which are, in general, a subset of the generating mass functions. A credal set over X is denoted as $K(X)$. We similarly denote as $K(X|y)$ a *conditional credal set* over X given a value y of another random variable Y, i.e., a credal set of conditional mass functions $P(X|y)$. Given a *joint credal set* $K(X,Y)$, the *marginal credal set* for X is the credal set $K(X)$ obtained by the point-wise marginalization of Y from all the joint mass function $P(X,Y) \in K(X,Y)$. Finally, given a subset $\Omega'_X \subseteq \Omega_X$, a particularly important credal set for our purposes is the *vacuous credal set* for Ω'_X, i.e., the set of all mass functions over X assigning probability one to Ω'_X. In the following we will use the well known fact that the vertices of such a credal set are the $|\Omega'_X|$ degenerate mass functions assigning probability one to the single elements of Ω'_X.

Definition 2. A credal network (CN) over $\mathbf{X}$ is a pair $\langle \mathcal{G}, \{\mathbb{P}_1, \ldots, \mathbb{P}_m\}\rangle$ such that $\langle \mathcal{G}, \mathbb{P}_j\rangle$ is a Bayesian network over $\mathbf{X}$ for each $j = 1, \ldots, m$.

The BNs $\{\langle \mathcal{G}, \mathbb{P}_j\rangle\}_{j=1}^m$ are said to be the *compatible* BNs of the CN considered in Definition 2.

The CN $\langle \mathcal{G}, \{\mathbb{P}_1, \ldots, \mathbb{P}_m\}\rangle$ can be used to determine the following credal set:

$$K(\mathbf{X}) := \mathrm{CH}\{P_1(\mathbf{X}), \ldots, P_m(\mathbf{X})\}, \qquad (2)$$

where CH denotes the convex hull of a set of functions, and the joint mass functions $\{P_j(\mathbf{X})\}_{j=1}^m$ are those determined by the compatible BNs of the CN. With an abuse of terminology, we call the credal set in Equation (2) the *strong extension* of the CN, by analogy with the notion provided in the special case of *separately specified* CNs (see Section 4). Inference over a CN is intended as the computation of upper and lower expectations for a given function of $\mathbf{X}$ over the credal set $K(\mathbf{X})$, or equivalently over its vertices (Walley, 1991).

4 Separately specified credal nets

The main feature of probabilistic graphical models, which is the specification of a global model through a collection of sub-models local to the nodes of the graph, contrasts with Definition 2, which represents a CN as an explicit enumeration of BNs.

Nevertheless, there are specific subclasses of CNs that define a set of BNs as in Definition 2 through local specifications. This is for example the case of CNs with separately specified credal sets,[2] which are simply called *separately specified CNs* in the following. This specification requires each conditional mass function to belong to a (conditional) credal set, according to the following:

Definition 3. A separately specified CN over $\mathbf{X}$ is a pair $\langle \mathcal{G}, \mathbb{K}\rangle$, where $\mathbb{K}$ is a set of conditional credal sets $K(X_i|\pi_i)$, one for each $X_i \in \mathbf{X}$ and $\pi_i \in \Omega_{\Pi_i}$.

The *strong extension* $K(\mathbf{X})$ of a separately specified CN is defined as the convex hull of the joint mass functions $P(\mathbf{X})$, with, for each $\mathbf{x} \in \Omega_{\mathbf{X}}$:

$$P(\mathbf{x}) = \prod_{i=1}^n P(x_i|\pi_i), \qquad \begin{array}{l} P(X_i|\pi_i) \in K(X_i|\pi_i), \\ \text{for each } X_i \in \mathbf{X}, \pi_i \in \Pi_i. \end{array} \qquad (3)$$

Here $K(X_i|\pi_i)$ can be replaced by the set of its vertices (see Proposition 1 in the appendix). Separately specified CNs are the most popular type of CN.

As a more general case, some authors considered so-called *extensive* specifications of CNs (Ferreira da Rocha and Cozman, 2002), where instead of a separate specification for each conditional mass function as in Definition 3, the generic probability table $P(X_i|\Pi_i)$, i.e., a function of both X_i and Π_i, is defined to belong to a finite set of tables. The strong extension of an extensive CN is obtained as in Equation (3), by simply replacing the separate requirements for each single conditional mass function, with extensive requirements about the tables which take values in the corresponding finite sets.

In the next section, we provide an alternative definition of CN, with the same generality of Definition 2, but obtained through local specification as in Definition 3.

[2] Some authors use also the expression *locally defined* CNs (Cozman, 2000).

5 Local specification of credal nets

In this section we provide an alternative and yet equivalent definition for CNs with respect to Definition 2, which is inspired by the formalism of *decision networks* (Zhang et al., 1993) via the CCM transformation (Cano et al., 1994).

Definition 4. A locally specified credal network over $\mathbf{X}'$ is a triplet $\langle \mathcal{G}, (\mathbf{X}_D, \mathbf{X}'), (\mathbb{O}, \mathbb{P}) \rangle$ such that: (i) $\mathcal{G}$ is a DAG over $\mathbf{X} = \mathbf{X}_D \cup \mathbf{X}'$; (ii) $\mathbb{O}$ is a collection of sets $\Omega_{X_i}^{\pi_i} \subseteq \Omega_{X_i}$, one for each $X_i \in \mathbf{X}_D$ and[3] $\pi_i \in \Pi_i$; (iii) $\mathbb{P}$ is a set of conditional mass functions $P(X_i|\pi_i)$, one for each $X_i \in \mathbf{X}'$ and $\pi_i \in \Omega_{\Pi_i}$.

We intend to show that Definition 4 specifies a CN over the variables in $\mathbf{X}'$; the nodes corresponding to $\mathbf{X}'$ are therefore called *uncertain* and will be depicted by circles, while those corresponding to $\mathbf{X}_D$ are said *decision nodes* and will be depicted by squares. Let us associate each decision node $X_i \in \mathbf{X}_D$ with a so-called *decision function* $f_{X_i} : \Omega_{\Pi_i} \to \Omega_{X_i}$ returning an element of $\Omega_{X_i}^{\pi_i}$ for each $\pi_i \in \Omega_{\Pi_i}$. Call *strategy* $\mathbf{s}$ an array of decision functions, one for each $X_i \in \mathbf{X}_D$. We denote as $\Omega_{\mathbf{S}}$ the set of all the possible strategies.

Each strategy $\mathbf{s} \in \Omega_{\mathbf{S}}$ determines a BN over $\mathbf{X}$ via Definition 4. A conditional mass function $P(X_i|\pi_i)$ for each uncertain node $X_i \in \mathbf{X}'$ and $\pi_i \in \Omega_{\Pi_i}$ is already specified by $\mathbb{P}$. To determine a BN we have then to simply represent decisions functions by mass functions: for each decision node $X_i \in \mathbf{X}_D$ and $\pi_i \in \Omega_{\Pi_i}$, we consider the conditional mass functions $P_{\mathbf{s}}(X_i|\pi_i)$ assigning all the mass to the value $f_{X_i}(\pi_i)$, where f_{X_i} is the decision function corresponding to $\mathbf{s}$. The BN obtained in this way will be denoted as $\langle \mathcal{G}, \mathbb{P}_{\mathbf{s}} \rangle$, while for the corresponding joint mass function, we clearly have, for each $\mathbf{x} = (\mathbf{x}_D, \mathbf{x}') \in \Omega_{\mathbf{X}}$:

$$P_{\mathbf{s}}(\mathbf{x}_D, \mathbf{x}') = \prod_{X_l \in \mathbf{X}_D} P_{\mathbf{s}}(x_l|\pi_l) \cdot \prod_{X_i \in \mathbf{X}'} P(x_i|\pi_i). \tag{4}$$

The next step is then obvious: we want to define a CN by means of the set of BNs determined by all the possible strategies. But the point is whether or not all these networks have the same DAG, as required by Definition 2. To show this we need the following:

Transformation 1. *Given a locally specified CN $\langle \mathcal{G}, (\mathbf{X}_D, \mathbf{X}'), (\mathbb{P}, \mathbb{O}) \rangle$, obtain a DAG $\mathcal{G}'$ associated to the variables $\mathbf{X}'$ iterating, for each $X_d \in \mathbf{X}_D$, the following operations over $\mathcal{G}$: (i) connect with an arc all the parents of X_d with all the children of X_d; (ii) remove the node corresponding to X_d.*

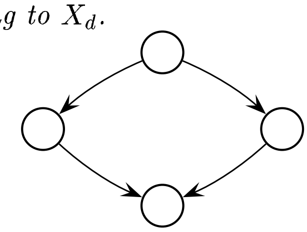

Figure 1: The DAG $\mathcal{G}'$ returned by Transformation 1 given a locally specified CN whose DAG is that in Figure 2 (or also Figure 3 or Figure 4).

Figure 1, reports an example of Transformation 1. The DAG $\mathcal{G}'$ returned by Transformation 1 is considered by the following:

Theorem 1. *The marginal for $\mathbf{X}'$ relative to $\langle \mathcal{G}, \mathbb{P}_{\mathbf{s}} \rangle$, i.e. the mass function $P_{\mathbf{s}}(\mathbf{X}')$ such that*

$$P'_{\mathbf{s}}(\mathbf{x}') = \sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} P_{\mathbf{s}}(\mathbf{x}_D, \mathbf{x}'), \tag{5}$$

for each $\mathbf{x}' \in \Omega_{\mathbf{X}'}$, factorizes as the joint mass function of a BN $\langle \mathcal{G}', \mathbb{P}'_{\mathbf{s}} \rangle$ over $\mathbf{X}'$, where $\mathcal{G}'$ is the DAG obtained from $\mathcal{G}$ by Transformation 1.

From which, considering the BNs $\langle \mathcal{G}', \mathbb{P}'_{\mathbf{s}} \rangle$ for each strategy $\mathbf{s} \in \Omega_{\mathbf{S}}$, it is possible to conclude:

Corollary 1. *A locally specified CN as in Definition 4 properly defines a CN over $\mathbf{X}'$, based on the DAG $\mathcal{G}'$ returned by Transformation 1.*

It is worthy to note that any CN defined as in Definition 2 can be reformulated as in Definition 4, by simply adding a single decision node, which is parent of all the other nodes (see Figure 2).

The conditional mass functions corresponding to different values of the decision node are assumed to be those specified by the compatible BNs. This means that, if D denotes the decision

[3] If X_i corresponds to a parentless node of $\mathcal{G}$, a single set equal to the whole Ω_{X_i} is considered.

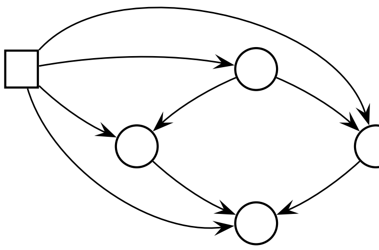

Figure 2: Local specification of a non-separately specified CN over the DAG in Figure 1. Remember that circles denote uncertain nodes, while the square is used for the decision node.

node, the states of D index the compatible BNs, and $P(X_i|\pi_i, d) := P_d(X_i|\pi_i)$, where $P_d(X_i|\pi_i)$ are the conditional mass functions specified by the d-th compatible BN for each $X_i \in \mathbf{X}'$ and $\pi_i \in \Omega_{\Pi_i}$ and $d \in \Omega_D$. This formulation, which is an example of the CCM transformation (Cano et al., 1994), is only seemingly local, because of the arcs connecting the decision node with all the uncertain nodes. In the remaining part of this section, we show how different CNs specifications can be reformulated as required by Definition 4.

For example, we can represent extensive CNs by introducing a decision parent for each node of the original CN (Figure 3).

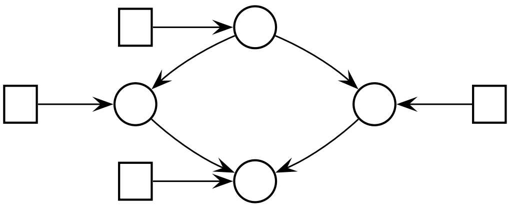

Figure 3: Local specification of an extensive CN over the DAG in Figure 1.

The conditional mass functions of the uncertain nodes corresponding to different values of the related decision nodes are assumed to be those specified by the different tables in the extensive specification of the uncertain node. This means that, if X_i in an uncertain node and D_i the corresponding decision node, the states $d_i \in \Omega_{D_i}$ index the tables $P_{d_i}(X_i|\Pi_i)$ of the extensive specification for X_i, and $P(X_i|d_i, \pi_i)$ is the joint mass function $P_{d_i}(X_i|\pi_i)$ associated to the d_i-th table of the extensive specification.

More generally, constraints for the specifications of conditional mass functions relative to different nodes are similarly represented by decision nodes which are the parents of these nodes.

Finally, to locally specify, as required by Definition 4, a separately specified CN, it would suffice to reformulate the separately specified CN as an extensive CN whose tables are obtained considering all the combinations of the vertices of the separately specified conditional credal sets of the same variable. Yet, this approach suffers an obvious exponential explosion of the number of tables in the input size.

A more effective procedure consists in adding a decision node in between each node and its parents, regarding the nodes of the original model as uncertain nodes (Figure 4). To complete the local specification proceed as follows. For each uncertain node X_i, the states $d_i \in \Omega_{D_i}$ of the corresponding decision node D_i are assumed to index the vertices of *all* the conditional credal sets $K(X_i|\pi_i) \in \mathbb{K}$, with $\pi_i \in \Omega_{\Pi_i}$. In this way, for each uncertain node X_i, it is possible to set the conditional mass function $P(X_i|d_i)$ to be the vertex of the conditional credal set $K(X_i|\pi_i)$ associated to d_i, for each $d_i \in \Omega_{D_i}$. Regarding decision nodes, for each decision node D_i and the related value π_i of the parents, we simply set the subset $\Omega_{D_i}^{\pi_i} \subseteq \Omega_{D_i}$ to be such that $\{P(X_i|d_i)\}_{d_i \in \Omega_{D_i}^{\pi_i}}$ are the vertices of $K(X_i|\pi_i)$. This approach, which is clearly linear in the input size, takes inspiration from *probability trees* representations (Cano and Moral, 2002).

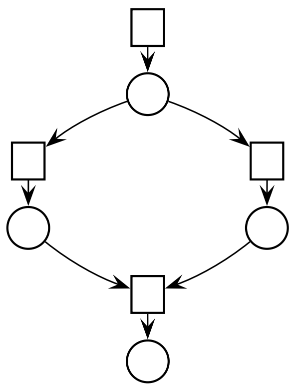

Figure 4: Local specification of a separately specified CN over the DAG in Figure 1.

Summarizing, we can obtain in efficient time a local specification of a CN, by simply considering one of the transformations described in this section.[4] It is worth noting that local specifications of CNs based on the DAGs respectively in Figure 4, Figure 3 and Figure 2 correspond to CNs based on the DAG in Figure 1 which are respectively separately specified, extensively specified and non-separately (and non-extensively) specified. Vice versa, it is easy to check that Transformation 1 can be regarded as the inverse of these three transformations.

6 Reasons for non-separately specified credal nets

Let us illustrate by a few examples that the necessity of non-separately specified credal nets arises naturally in a variety of problems.

Conservative Inference Rule The *conservative inference rule* (CIR) is a new rule for updating beliefs with incomplete observations (Zaffalon, 2005). CIR models the case of mixed knowledge about the incompleteness process, which is assumed to be nearly unknown for some variables and unselective for the others. This leads to an imprecise probability model, where all the possible completions of the incomplete observations of the first type of variables are considered. In a recent work, updating beliefs with CIR has been considered for BNs (Antonucci and Zaffalon, 2006). The problem is proved to be equivalent to a traditional belief updating problem in a CN: the equivalence is made possible by non-separately specified credal nets.

Qualitative networks Qualitative probabilistic networks (Wellman, 1990) are an abstraction of BNs, where the probabilistic assessments are replaced by qualitative relations of the influences or the synergies between the variables. If we regard qualitative nets as credal nets, we see that not all types of relations can be represented by separate specifications of the conditional credal sets. This is, for instance, the case of (positive) *qualitative influence*, which requires, for two boolean variables A and B, that

$$P(a|b) \geq P(a|\neg b). \tag{6}$$

The qualitative influence between A and B can therefore be modeled by requiring $P(A|b)$ and $P(A|\neg b)$ to belong to credal sets, which cannot be separately specified because of the constraint in Equation (6). An extensive specification for A should therefore be considered to model the positive influence of B (Cozman et al., 2004).

Equivalent graphs for CNs Remember that DAGs represent independencies between variables according to the Markov condition. Different DAGs describing the same independencies are said to be *equivalent* (Verma and Pearl, 1990). Thus, a BN can be reformulated using an equivalent DAG. The same holds with CNs, when (as implicitly done in this paper) *strong independence* replaces standard probabilistic independence in the Markov condition (Moral and Cano, 2002).

Consider, for example, $A \to B$ and $A \leftarrow B$, which are clearly equivalent DAGs. One problem with separately specified CNs is that they are not closed under this kind of (equivalent) structure changes: if we define a separately specified CN for $A \to B$, and then reverse the arc, the resulting net will not be separately specified in general. Consider the following separate specifications of the conditional credal sets for a CN over $A \to B$:

$$\begin{aligned} K(B|a) &:= [.9, .1] \quad K(B|\neg a) := [.8, .2] \\ K(A) &:= \mathrm{CH}\{[.7, .3], [.5, .5]\}, \end{aligned}$$

where two-dimensional horizontal arrays are used to denote mass functions for boolean variables. Such a separately specified CN has two compatible BNs, one for each vertex of $K(A)$. From the joint mass functions corresponding to

[4] As a side note, it is important to be aware that a credal set can have a very large number of vertices, and this can still be a source of computational problems for algorithms (such as those based on the CCM transformation) that explicitly enumerate the vertices of a net's credal sets. This is a well-know issue, which in the present setup is related to the possibly large number of states for the decision nodes in the locally specified representation of a credal net.

these BNs, say $P_1(A,B)$ and $P_2(A,B)$, we obtain the conditional mass functions for the corresponding BNs over $B \to A$:

$$\begin{array}{ll} P_1(B) = [.87,.13] & P_2(B) = [.85,.15] \\ P_1(A|b) = [.72,.28] & P_2(A|b) = [.52,.47] \\ P_1(A|\neg b) = [.54,.46] & P_2(A|\neg b) = [.33,.67]. \end{array}$$

According to Definition 2, these two distinct specifications define a CN over $B \to A$, which cannot be separately specified as in Definition 3. To see this, note for example that the specification $P(a|b) = .52$, $P(a|\neg b) = .33$ and $P(b) = .85$, which is clearly a possible specification if the conditional credal sets were separately specified, would lead to the unacceptable mass function $P(A) = [.49,.51] \notin K(A)$.

It is useful to observe that general, non-separately specified, CNs do not suffer for these problems just because they are closed under equivalent changes in the structure of the DAG.

Learning from incomplete data Given three boolean random variables A, B and C, let the DAG $A \to B \to C$ express independencies between them. We want to learn the model probabilities for such a DAG from the incomplete data set in Table 1, assuming no information about the process making the observation of B missing in the last record of the data set. The most conservative approach is therefore to learn two distinct BNs from the two complete data sets corresponding to the possible values of the missing observation and consider indeed the CN made of these compatible BNs.

A	B	C
a	b	c
$\neg a$	$\neg b$	c
a	b	$\neg c$
a	$*$	c

Table 1: A data set about three boolean variables, $*$ denotes a missing observation.

To make things simple we compute the probabilities for the joint states by means of the relative frequencies in the complete data sets. Let $P_1(A,B,C)$ and $P_2(A,B,C)$ be the joint mass function obtained in this way, which define the same conditional mass functions for:

$$\begin{array}{l} P_1(A) = P_2(A) = [.75,.25] \\ P_1(B|\neg a) = P_2(B|\neg a) = [0,1] \\ P_1(C|\neg b) = P_2(C|\neg b) = [1,0]; \end{array}$$

and different conditional mass functions for:

$$\begin{array}{ll} P_1(B|a) = [1,0] & P_2(B|a) = [.67,.33] \\ P_1(C|b) = [.67,.33] & P_2(C|b) = [.5,.5]. \end{array}$$

We have therefore obtained two BNs over $A \to B \to C$, which can be regarded as the compatible BNs of a CN. Such a CN is clearly non-separately specified, because the two BNs specify different conditional mass functions for more than a variable.

7 From locally to separately specified credal nets

In this section, we prove that any locally specified CN over $\mathbf{X}'$ can equivalently be regarded as a separately specified CN over $\mathbf{X}$. The transformation is technically straightforward: it is based on representing decision nodes by uncertain nodes with vacuous conditional credal sets, as formalized below.

Transformation 2. *Given a locally specified CN $\langle \mathcal{G}, (\mathbf{X}_D, \mathbf{X}'), (\mathbb{O}, \mathbb{P}) \rangle$ over $\mathbf{X}'$, obtain a separately specified CN $\langle \mathcal{G}, \mathbb{K} \rangle$ over $\mathbf{X}$, where the conditional credal sets in $\mathbb{K}$ are for each $X_i \in \mathbf{X}$ and $\pi_i \in \Omega_{\Pi_i}$:*

$$K(X_i|\pi_i) := \begin{cases} P(X_i|\pi_i) & \text{if } X_i \in \mathbf{X}' \\ K_{\Omega_{X_i}^{\pi_i}}(X_i) & \text{if } X_i \in \mathbf{X}_D, \end{cases} \tag{7}$$

where $P(X_i|\pi_i)$ is the mass function specified in $\mathbb{P}$ and $K_{\Omega_{X_i}^{\pi_i}}(X_i)$ the vacuous *credal set for $\Omega_{X_i}^{\pi_i}$.*

Figure 5 reports an example of Transformation 2, which is clearly linear in the input size.

The (strong) relation between a locally specified CN $\langle \mathcal{G}, (\mathbf{X}_D, \mathbf{X}'), (\mathbb{O}, \mathbb{P}) \rangle$ over $\mathbf{X}'$ and the separately specified CN $\langle \mathcal{G}, \mathbb{K} \rangle$ returned by Transformation 2 is outlined by the following:

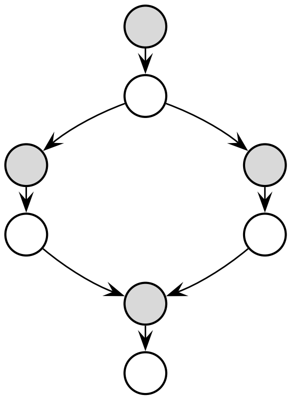

Figure 5: The DAG associated to the separately specified CN returned by Transformation 2, from the locally specified CN based on the DAG in Fig. 4. The conditional credal sets of the white nodes (corresponding to the original uncertain nodes) are precisely specified, while the grey nodes (i.e., new uncertain nodes corresponding to the former decision nodes) represent variables whose conditional credal sets are vacuous.

Theorem 2. *Let* $\tilde{K}(\mathbf{X}')$ *be the marginal for* $\mathbf{X}'$ *of the strong extension* $\tilde{K}(\mathbf{X})$ *of* $\langle \mathcal{G}, \mathbb{K} \rangle$ *and* $K(\mathbf{X}')$ *the strong extension of* $\langle \mathcal{G}, (\mathbf{X}_D, \mathbf{X}'), (\mathbb{O}, \mathbb{P}) \rangle$. *Then:*

$$K(\mathbf{X}') = \tilde{K}(\mathbf{X}'). \tag{8}$$

From Theorem 2, it is straightforward to conclude the following:

Corollary 2. *Any inference problem on a locally specified CN can be equivalently solved in the separately specified CN returned by Transformation 2.*

Let us stress that Transformation 2 is very simple, and it is surprising that it is presented here for the first time, as it is really the key to "separate" the credal sets of non-separately specified nets: in fact, given a non-separately specified CN, one can locally specify the CN, using the prescriptions of the second part of Section 5, and apply Transformation 2 to obtain a separately specified CN. According to Corollary 2, then, any inference problem on the original CN can equivalently be represented on this new separately specified CN. To make an example, this procedure can immediately solve the CIR updating problem described in Section 6, which is a notable and ready-to-use result.

8 Conclusions and outlooks

We have defined a new graphical language to formulate any type of credal network, both separately and non-separately specified. We have also showed that any net represented with the new language can be easily transformed into an equivalent separately specified credal net. This implies, in particular, that non-separately specified nets have an equivalent separately specified representation, for which solutions algorithms are available in the literature.

The transformation proposed also shows that a subclass of separately specified credal networks can be used to solve inference problems for arbitrary specified credal nets: this is the class of nets in which the credal sets are either vacuous or precise. It is worth noting that a recent development of the approximate L2U algorithm (Antonucci et al., 2006) seems to be particularly suited just for such a class, and should therefore be considered in future work.

Finally, the strong connection between the language for credal networks introduced in this paper and the formalism of decision networks (including influence diagrams), seems to be particularly worth exploring for cross-fertilization between the two fields.

Acknowledgements

This research was partially supported by the Swiss NSF grant 200020-109295/1.

A Proofs

The obvious proofs of Corollary 1 and Corollary 2 are omitted.

Proof of Theorem 1. Let us start the marginalization in Equation (5) from a decision node $X_j \in \mathbf{X}_D$. According to Equation (4), for each $\mathbf{x} \in \Omega_{\mathbf{X}}$:

$$\sum_{\mathbf{x}_j \in \Omega_{\mathbf{X}_j}} P_{\mathbf{s}}(\mathbf{x}) = \sum_{\mathbf{x}_j \in \Omega_{\mathbf{X}_j}} \left[\prod_{X_l \in \mathbf{X}_D} P_{\mathbf{s}}(x_l|\pi_l) \cdot \prod_{X_i \in \mathbf{X}'} P(x_i|\pi_i) \right]. \tag{9}$$

Thus, moving out of the sum the conditional probabilities which do not refer to the states of X_j (which are briefly denoted by Δ), Equation (9) becomes:

$$\Delta \cdot \sum_{x_j \in \Omega_{X_j}} \left[P_{\mathbf{s}}(x_j|\pi_j) \cdot \prod_{X_r \in \Gamma_{X_j}} P(x_r|x_j, \tilde{\pi}_r) \right], \tag{10}$$

where Γ_{X_j} denotes the children of X_j and, for each $X_r \in \Gamma_{X_j}$, $\tilde{\Pi}_r$ are the parents of X_r deprived of X_j. Therefore, considering that the mass function $P_{\mathbf{s}}(X_j|\pi_j)$ assigns all the mass to the value $f_{X_j}(\pi_j) \in \Omega_{X_j}$, where f_{X_j} is the decision function associated to $\mathbf{s}$, Equation (10) rewrites as

$$\Delta \cdot \prod_{X_r \in \Gamma_{X_j}} P(x_r|f_{X_j}(\pi_j), \tilde{\pi}_r). \tag{11}$$

It is therefore sufficient to set $\Pi'_r := \Pi_j \cup \tilde{\Pi}_r$, and

$$P'_{\mathbf{s}}(X_r|\pi'_r) := P(X_r|f_{X_j}(\pi_j), \tilde{\pi}_r), \tag{12}$$

to regard Equation (11) as the joint mass function of a BN over $\mathbf{X} \setminus \{X_j\}$ based on the DAG returned by Transformation 1 considered for the single decision node $X_j \in \mathbf{X}_D$. The thesis therefore follows from a simple iteration over all the $X_j \in \mathbf{X}_D$. □

The following well-known and (relatively) intuitive proposition is required to obtain Theorem 2, and will be proved here because of the seemingly lack of its formal proof in the literature.

Proposition 1. *The vertices $\{\tilde{P}_j(\mathbf{X})\}_{j=1}^m$ of the strong extension $\tilde{K}(\mathbf{X})$ of a separately specified CN $\langle \mathcal{G}, \mathbb{K} \rangle$ are joint mass functions obtained by the combination of vertices of the separately specified conditional credal sets, i.e., for each $\mathbf{x} \in \Omega_{\mathbf{X}}$:*

$$\tilde{P}_j(\mathbf{x}) = \prod_{i=1}^n \tilde{P}_j(x_i|\pi_i), \tag{13}$$

for each $j = 1, \ldots, m$, where, for each $i = 1, \ldots, n$ and $\pi_i \in \Omega_{\Pi_i}$, $\tilde{P}_j(X_i|\pi_i)$ is a vertex of $K(X_i|\pi_i) \in \mathbb{K}$.

Proof of Proposition 1. We prove the proposition by a *reductio ad absurdum*, assuming that at least a vertex $\tilde{P}(\mathbf{X})$ of $\tilde{K}(\mathbf{X})$ is not obtained by a local combination of vertices of the conditional credal sets in $\mathbb{K}$. This means that, for each $\mathbf{x} \in \Omega_{\mathbf{X}}$, $\tilde{P}(\mathbf{x})$ factorizes as in Equation (13), but at least a conditional probability in this product comes from a conditional mass function which is not a vertex of the relative conditional credal set. This conditional mass function, say $P(X_t|\pi_t)$, can be expressed as a convex combination of vertices of $K(X_t|\pi_t)$, i.e., $P(X_t|\pi_t) = \sum_\alpha c_\alpha P_\alpha(X_t|\pi_t)$, with $\sum_\alpha c_\alpha = 1$ and, for each α, $c_\alpha \geq 0$ and $P_\alpha(X_t|\pi_t)$ is a vertex of $K(X_t|\pi_t)$. Thus, for each $\mathbf{x} \in \Omega_{\mathbf{X}}$,

$$\tilde{P}(\mathbf{x}) = \left[\sum_\alpha c_\alpha P_\alpha(x_t|\pi_t) \right] \cdot \prod_{i \neq t} P(x_t|\pi_t), \tag{14}$$

which can be easily reformulated as a convex combination. Thus, $\tilde{P}(\mathbf{X})$ is a convex combination of elements of the strong extension $\tilde{K}(\mathbf{X})$. This violates the assumption that $\tilde{P}(\mathbf{X})$ is a vertex of $\tilde{K}(\mathbf{X})$. □

Proof of Theorem 2. According to Theorem 1, the strong extension $K(\mathbf{X}')$ of $\langle \mathcal{G}, (\mathbf{X}_D, \mathbf{X}'), (\mathbb{O}, \mathbb{P}) \rangle$ can be regarded as the marginal for $\mathbf{X}'$ of

$$K(\mathbf{X}) = \mathrm{CH}\{P_{\mathbf{s}}(\mathbf{X})\}_{\mathbf{s} \in \Omega_{\mathbf{S}}}, \tag{15}$$

where for each $\mathbf{s} \in \Omega_{\mathbf{S}}$, $P_{\mathbf{s}}(\mathbf{X})$ is the joint mass function associated to $\langle \mathcal{G}, \mathbb{P}_{\mathbf{s}} \rangle$. For each $X_d \in \mathbf{X}_D$, the conditional mass functions $P_{\mathbf{s}}(X_d|\pi_d)$, specified for each $\pi_d \in \Omega_{\Pi_d}$ in $\mathbb{P}_{\mathbf{s}}$, assign all the mass to the single state $f_{X_d}(\pi_d) \in \Omega_{X_d}^{\pi_d}$, where $f_{X_d}^{\pi_d}$ are the decision functions associated to the strategy $\mathbf{s} \in \Omega_{\mathbf{S}}$, and represent therefore a vertex of the vacuous conditional credal set $K_{\Omega_{X_d}}^{\pi_d}(X_d) \in \mathbb{K}$ specified in Equation (7). Thus, $P_{\mathbf{s}}(\mathbf{X})$ is a vertex of $\tilde{K}(\mathbf{X})$ because of Proposition 1. As $\mathbf{s}$ varies in $\Omega_{\mathbf{S}}$, all the vertices of $\tilde{K}(\mathbf{X})$ are obtained and therefore $\tilde{K}(\mathbf{X}) = K(\mathbf{X})$, from which the thesis follows marginalizing. □

References

A. Antonucci and M. Zaffalon. 2006. Equivalence between Bayesian and credal nets on an updating problem. In *Proceedings of third international conference on Soft Methods in Probability and Statistics*. To appear.

A. Antonucci, M. Zaffalon, J.S. Ide, and F.G. Cozman. 2006. Binarization algorithms for approximate updating in credal nets. In *Proceedings of the third European Starting AI Researcher Symposium*. To appear.

A. Cano and S. Moral. 2002. Using probability trees to compute marginals with imprecise probabilities. *Int. J. Approx. Reasoning*, 29(1):1–46.

A. Cano, J. Cano, and S. Moral. 1994. Convex sets of probabilities propagation by simulated annealing on a tree of cliques. In *Proceedings of 5th International Conference on Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 4–8.

F. G. Cozman, C. P. de Campos, J. S. Ide, and J. C. Ferreira da Rocha. 2004. Propositional and relational Bayesian networks associated with imprecise and qualitative probabilistic assessments. In *Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence*, pages 104–111. AUAI Press.

F. G. Cozman. 2000. Credal networks. *Artificial Intelligence*, 120:199–233.

F.G. Cozman. 2005. Graphical models for imprecise probabilities. *Int. J. Approx. Reasoning*, 39(2-3):167–184.

J. C. Ferreira da Rocha and F. G. Cozman. 2002. Inference with separately specified sets of probabilities in credal networks. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 430–437. Morgan Kaufmann.

I. Levi. 1980. *The Enterprise of Knowledge*. MIT Press, London.

S. Moral and A. Cano. 2002. Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence*, 35(1-4):295–321.

T. Verma and J. Pearl. 1990. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270. Elsevier.

P. Walley. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York.

M. P. Wellman. 1990. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44(3):257–303.

M. Zaffalon. 2001. Statistical inference of the naive credal classifier. In *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 384–393. Shaker.

M. Zaffalon. 2005. Conservative rules for predictive inference with incomplete data. In *Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, pages 406–415. SIPTA.

N. Zhang, R. Qi, and D. Poole. 1993. A computational theory of decision networks. *Int. J. Approx. Reasoning*, 11(2):83–158.