Decision-Theoretic Specification of Credal Networks: A Unified Language for Uncertain Modeling with Sets of Bayesian Networks

Alessandro Antonucci   Marco Zaffalon

IDSIA, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale,
Galleria 2, 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. Credal nets appear to be 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 modeled by credal nets called *non-separately specified*. These, however, are still missing a graphical representation language and updating 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 outlooks 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. We explore this possibility for the 2U algorithm: an algorithm for exact updating of singly connected credal nets, which is extended by our results to a class of non-separately specified models. We also consider the problem of inference on Bayesian networks, when the reason that prevents some of the variables from being observed is unknown. The problem is first reformulated in the new graphical language, and then mapped into an equivalent problem on a separately specified net. This provides a first algorithmic approach to this kind of inference, which is also proved to be NP-hard by similar transformations based on our formalism.

Keywords: Probabilistic Graphical Models, Bayesian Networks, Credal Networks, Credal Sets, Imprecise Probabilities, Conservative Updating, Conservative Inference Rule.

1   Introduction

We focus on credal networks (Section 3) [1], 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 [2]. Using credal sets in the place of mass functions makes credal networks an *imprecise probability* model [3]. It can be shown, in particular, 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

1

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* (Section 5). 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 be far away from each other 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 to be an unfortunate gap in the literature as the non-separate specification seems to be the key to model many important problems (Sections 5.1–5.4). Separately specified credal nets, on the other hand, have been the subject of much algorithmic development (see [1] for an overview of the main results in this field).

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 6). This specification is called *decision-theoretic* being inspired, via the Cano-Cano-Moral (CCM) transformation [5], by the formalism of influence diagrams, and more generally of decision graphs [6]. In this language the graph of a credal network 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 modeled by non-separately specified ones; and hence to use *any* (both exact and approximate) existing algorithm for separately specified nets to solve such problems.

In Section 8 we explore this possibility in the case of the *2U algorithm* [7]: a polynomial-time algorithm for exact updating of singly connected credal networks with binary variables. We show that the algorithm, originally designed only for separately specified credal networks, can be extended to deal exactly and efficiently also with a class of non-separately specified models.

Our contributions also apply to the problem of belief updating on Bayesian networks by the *conservative inference rule* [8], which is a rule modeling situations where the reason that prevents some of the variables from being observed is unknown. The problem has been mapped onto a standard updating problem on a non-separately specified credal network [9], a result not straightforward to exploit in practice because of the lack of algorithms for non-separately specified credal networks. A feasible solution of this problem based on our formalism is presented in Section 9. First, we represent the problem by the new decision-theoretic language. Second, we use our transformation to reformulate the problem on a separately specified credal network defined over a larger domain. At this point, the problem can be solved by the existing algorithms for separately specified credal nets. Additionally, we also prove the NP-hardness of

---

[1]An exception is the classification algorithm developed for the *naive credal classifier* [4], but it is ad hoc for a very specific type of network. More generally speaking, it is not unlikely that some of the 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.

belief updating with this rule by similar transformations based on the results presented in this paper.

Some comments and perspectives for future developments are discussed in Section 10. The more technical parts of this paper are collected in Appendix A.

2   Basic Notation and Bayesian Networks

Let us first define some notation and the notion of Bayesian network. Let[2] $\mathbf{X} := (X_1, \ldots, X_n)$ 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 $X_i = x_i$. If X_i is a binary variable, we denote the elements of Ω_{X_i} by $\{x_i, \neg x_i\}$. 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 introduce the definition of Bayesian network:

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 formula:

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

for each $\mathbf{x} \in \Omega_{\mathbf{X}}$, where for each $i = 1, \ldots, n$ the values (x_i, π_i) are those 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* [2]. We follow [10] 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 $K(X)$ and the set of its extreme mass functions as $\text{ext}[K(X)]$. Given a *joint credal set* $K(X, Y)$, the *marginal credal set* for X is the credal set $K(X)$ obtained by the element-wise marginalization of Y from all the joint mass functions $P(X, Y) \in K(X, Y)$. We similarly denote as $K(X|y)$ a *conditional credal set* over X given a value y of Y, i.e., a credal set of conditional mass functions $P(X|y)$ obtained by element-wise application of Bayes rule. Finally, given a non-empty 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. We denote this set as $K_{\Omega'_X}(X)$. In the following we will use the well known fact that the vertices of $K_{\Omega'_X}(X)$ are the[3] $|\Omega'_X|$ degenerate mass functions assigning probability one to the single elements of Ω'_X.

---

[2]The symbol ":=" is used for definitions.
[3]The cardinality of a set Ω is denoted as $|\Omega|$.

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 specified in Definition 2.

The CN $\langle \mathcal{G}, \{\mathbb{P}_1, \ldots, \mathbb{P}_m\} \rangle$ determines the following credal set:[4]

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

where CH denotes the convex hull of the $|\Omega_{\mathbf{X}}|$-dimensional points represented by $\{P_j(\mathbf{X})\}_{j=1}^m$, which are the joint mass functions determined by the compatible BNs of the CN. This convexification is necessary to ensure consistency with Walley's theory of coherent lower previsions [3]. 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 bounds for the posterior expectation, of a given function of $\mathbf{X}$, considered for each $P(\mathbf{X}) \in K(\mathbf{X})$. This optimization on a continuous domain is in practice a combinatorial task, as lower and upper expectations can be equivalently computed considering only the vertices $\mathrm{ext}[K(\mathbf{X})]$ [3]. If the function of $\mathbf{X}$ to be considered is an indicator, inference is simply called *updating*. Other important inference problems can be also considered on CNs (e.g., dominance relations and non-dominated options computation [3, Sect. 3.9]). In this paper we provide some equivalent representations of the strong extension that can be therefore applied to general inference problems.

4 Separately Specified Credal Networks

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,[5] 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:

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 [10] is defined as the convex hull of the joint mass functions $P(\mathbf{X})$, with, for each $\mathbf{x} \in \Omega_{\mathbf{X}}$:

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

Here $K(X_i | \pi_i)$ can also be replaced by $\mathrm{ext}[K(X_i | \pi_i)]$ (see Proposition 1 in Appendix A).

---

[4]Generally speaking the fact that all the joint mass functions $\{P_j(\mathbf{X})\}_{j=1}^m$ in Equation (2) factorize as in Equation (1) does not imply that the every $P(\mathbf{X}) \in K(\mathbf{X})$ should do the same.

[5]Some authors use also the expression *locally defined* CNs [10].

5 Non-Separately Specified Credal Networks

Separately specified CNs are the most popular type of CN, but it is clearly possible to consider a CNs that cannot be formulated as in Definition 3. This corresponds to having relationships between the different specifications of the conditional credal sets, which means that the possible values for a given conditional mass function can be affected by the values of some other conditional mass functions. A CN of this kind is said to be *non-separately specified*.

As an example, some authors considered so-called *extensive* specifications of CNs [11], 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 (see Figure 1). 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 set.

$$P(A) = \begin{bmatrix} .1 \\ .9 \end{bmatrix} \quad \text{(A)} \quad \begin{matrix} \text{(B)} & P(B|A) \in \left\{ \begin{bmatrix} .2 & .3 \\ .8 & .7 \end{bmatrix}, \begin{bmatrix} .4 & .5 \\ .6 & .5 \end{bmatrix} \right\} \\ \\ \text{(C)} & P(C|A) \in \left\{ \begin{bmatrix} .6 & .7 \\ .4 & .3 \end{bmatrix}, \begin{bmatrix} .8 & .9 \\ .2 & .1 \end{bmatrix} \right\} \end{matrix}$$
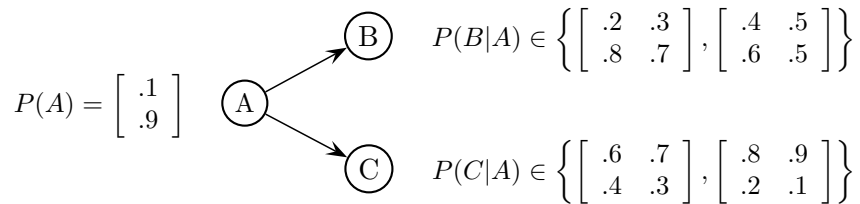
Figure 1: An extensive specification of a CN over three binary variables. The compatible BNs of the CN are those obtained by the four possible combinations of the probability tables $P(B|A)$ and $P(C|A)$. The network is non-separately specified, as the conditional mass functions over B, corresponding to the two columns of the conditional probability table $P(B|A)$, cannot vary one independently of the other (and similarly for C).

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

5.1 Conservative Inference Rule

The *conservative inference rule* (CIR) is a new rule for updating beliefs with incomplete observations [12]. CIR models the case of mixed knowledge about the process preventing the observation of some of the variables from being obtained: this incompleteness process is assumed to be *nearly unknown*[6] for some variables, while the others are assumed to be *missing at random* (MAR) [13]. This leads to an imprecise probability rule, where all the possible completions of the incomplete observations of the first type of variables are considered.

We consider an application of this rule to a BN over the set of variables $\mathbf{X}$, that assigns a positive probability to any joint state $\mathbf{x} \in \Omega_{\mathbf{X}}$. The set $\mathbf{X}$ is partitioned in the following four classes: (i) the queried variable X_q, (ii) the observed variables $\mathbf{X}_E$, (iii) the unobserved variables $\mathbf{X}_M$ for which the MAR assumption is known to hold, and (iv) the variables $\mathbf{X}_I$ made missing by a process that we basically ignore. CIR leads to the following credal set as our updated belief about the queried variable:

$$K(X_q||^{\mathbf{x}_I}\mathbf{x}_E) := \text{CH}\left\{P(X_q|\mathbf{x}_E, \mathbf{x}_I)\right\}_{\mathbf{x}_I \in \Omega_{X_I}}, \tag{4}$$

---

[6]The notion of nearly unknown incompleteness process is formalized in [12, Sect. 2.2] and is based on very weak assumptions about the mechanism that makes the observation incomplete.

where the superscript on the double conditioning bar is used to denote beliefs updated with CIR and to specify the set of missing variables $\mathbf{X}_I$ assumed to be non-MAR.

In [9] the computation of $K(X_q||^{\mathbf{x}_I}\mathbf{x}_E)$ has been mapped to a standard updating problem on a CN defined over a larger domain. The CN is obtained from the original BN by the following procedure:

Transformation 1 *Iterate, for each $X \in \mathbf{X}_I$, the following operations: (i) add an auxiliary node*[7] *X', which is a child of X and corresponds to a binary variable; (ii) specify the conditional probability table $P(X'|X)$ to belong to the following set of $|\Omega_X|$ tables:*

$$\left\{ \begin{bmatrix} 1\,0\,0\ldots0 \\ 0\,1\,1\ldots1 \end{bmatrix}, \ldots, \begin{bmatrix} 0\ldots0\,1\,0\ldots0 \\ 1\ldots1\,0\,1\ldots1 \end{bmatrix}, \ldots, \begin{bmatrix} 0\,0\,0\ldots0\,1 \\ 1\,1\,1\ldots1\,0 \end{bmatrix} \right\}. \tag{5}$$

As proved in [9], the conditional credal set in Equation (4) is equivalent to a credal set computed on the credal network returned by Transformation 1, for the same queried variable, conditional on the available evidence and on the joint state $\overline{\mathbf{x}}'_I \in \Omega_{\mathbf{X}'_I}$ of the array $\mathbf{X}'_I$ of the auxiliary nodes, for which all the variables $X' \in \mathbf{X}'_I$ are in the states corresponding to the first rows of the tables in Equation (5). This equivalence can be written as:

$$K(X_q||^{\mathbf{x}_I}\mathbf{x}_E) = K(X_q|\mathbf{x}_E, \overline{\mathbf{x}}'_I). \tag{6}$$

Note that Transformation 1 returns a CN which is clearly extensively specified and hence non-separately specified. This means that Equation (6) cannot be used in practice to solve CIR-based updating problems, because of lack of algorithms for non-separately specified CNs. In Section 9 this limitation is overcome by means of the results provided in Section 6 and Section 7.

5.2 Qualitative Networks

Qualitative probabilistic networks [15] can be regarded as an abstraction of BNs, where the probabilistic assessments are replaced by qualitative relations describing the influences or 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 binary variables A and B, that

$$P(a|b) \geq P(a|\neg b). \tag{7}$$

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 (7). An extensive specification for A should therefore be considered to model the positive influence of B [16].

5.3 Equivalent Graphs for Credal Networks

Remember that DAGs represent independencies between variables according to the Markov condition. Different DAGs describing the same independencies are said to be *equivalent* [17]. 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 [18].

---

[7]This transformation takes inspirations from Pearl's prescriptions about boundary conditions for propagation [14, Sect. 4.3].

Consider, for example, $A \to B$ and $B \to A$, 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.

In order to see that, we consider the following specification of a CN over $A \to B$, where both A and B are binary variables: $\frac{1}{4} \leq P(a) \leq \frac{1}{2}$, $\frac{1}{2} \leq P(b|a) \leq \frac{3}{4}$ and $P(b|\neg a) = \frac{3}{4}$. As all the variables are binary, the computation of the credal set corresponding to these intervals is trivial. E.g., the vertices of $K(A)$ are clearly the two mass functions $[\frac{1}{4}, \frac{3}{4}]^T$ and $[\frac{1}{2}, \frac{1}{2}]^T$. Overall, we have a separately specified CN with four compatible BNs, corresponding to the possible combinations of the two vertices of $K(A)$ with the two vertices of $K(B|a)$. From the joint mass functions corresponding to these BNs, we can evaluate the conditional mass functions for the corresponding BNs over $B \to A$, which are those corresponding to the following probabilities:

$$
\begin{array}{llll}
P_1(b) = \frac{11}{16} & P_2(b) = \frac{3}{4} & P_3(b) = \frac{5}{8} & P_4(b) = \frac{3}{4} \\
P_1(a|b) = \frac{2}{11} & P_2(a|b) = \frac{1}{4} & P_3(a|b) = \frac{2}{5} & P_4(a|b) = \frac{1}{2} \\
P_1(a|\neg b) = \frac{2}{5} & P_2(a|\neg b) = \frac{1}{4}. & P_3(a|\neg b) = \frac{2}{3}. & P_4(a|\neg b) = \frac{1}{2}.
\end{array}
$$

According to Definition 2, these four distinct BN 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 $P(b) = \frac{5}{8}$, $P(a|b) = \frac{1}{2}$ and $P(a|\neg b) = \frac{2}{3}$, which would be clearly possible if the conditional credal sets were separately specified, leads to the inadmissible value $P(a) = \frac{9}{16} > \frac{1}{2}$.

It is useful to observe that general, non-separately specified, CNs do not suffer for these problems just because of their definition.

5.4   Learning from Incomplete Data

Given three binary 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 (i.e., the parameters) 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 in this case is 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 binary 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.[8] Let $P_1(A, B, C)$ and $P_2(A, B, C)$ be the joint mass functions obtained in this way, which define the same conditional mass functions for

$$
\begin{array}{l}
P_1(a) = P_2(a) = \frac{3}{4} \\
P_1(b|\neg a) = P_2(b|\neg a) = 0 \\
P_1(c|\neg b) = P_2(c|\neg b) = 1;
\end{array}
$$

_____

[8]We do this only for illustrative purposes, as there are arguably better ways to learn probabilities from data, such as the _imprecise Dirichlet model_ [19]. Yet, also these other methods would incur the same problem [20].

and different conditional mass functions for

$$P_1(b|a) = 1 \qquad P_2(b|a) = \frac{2}{3}$$
$$P_1(c|b) = \frac{2}{3} \qquad P_2(c|b) = \frac{1}{2}.$$

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 non-separately specified. To see that, just note that if the CN would be separately specified the values $P(b|a) = 1$ and $P(c|b) = \frac{1}{2}$ could be regarded as a possible instantiation of the conditional probabilities, despite the fact that there are no complete data sets leading to this combination of values.

6 Decision-Theoretic Specification of Credal Networks

We provide an alternative definition of CN with the same generality of Definition 2, but obtained through local specifications as in Definition 3. This result, which is inspired by the formalism of *decision networks* [6] via the CCM transform [5], is reported in Section 6.1.

Remarkably, both non-separately (Section 6.2) and separately specified CNs (Section 6.3) can be reformulated in accord to this definition by means of transformations taking only polynomial time. We can therefore regard the new definition as the basis for a graphical language to represent in a unified form CNs of any kind.

6.1 General Definition of Decision-Theoretic Specification

Definition 4 *A decision-theoretic specification of a credal network*[9] *over* $\mathbf{X}$ *is a triplet* $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$ *such that: (i)* $\mathcal{G}'$ *is a DAG over* $\mathbf{X}' := (\mathbf{X}_D, \mathbf{X})$; *(ii)* $\mathbb{O}$ *is a collection of non-empty sets* $\Omega_{X_i}^{\pi_i} \subseteq \Omega_{X_i}$, *one for each* $X_i \in \mathbf{X}_D$ *and* $\pi_i \in \Omega_{\Pi_i}$;[10] *(iii)* $\mathbb{P}'$ *is a set of conditional mass functions* $P'(X_j|\pi_j)$, *one for each* $X_j \in \mathbf{X}$ *and* $\pi_j \in \Omega_{\Pi_j}$.

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 displayed by circles, while those corresponding to $\mathbf{X}_D$ are said *decision nodes* and will be displayed by squares. Let us associate each decision node $X_i \in \mathbf{X}_D$ with its collection of so-called *decision functions*. For each $X_i \in \mathbf{X}_D$, the decision functions of X_i are all the possible maps $f_{X_i} : \Omega_{\Pi_i} \to \Omega_{X_i}$ returning an element of $\Omega_{X_i}^{\pi_i}$ for each $\pi_i \in \Omega_{\Pi_i}$. Note that the decision functions of a root node X_i are the single elements of Ω_{X_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, as illustrated below. A conditional mass function $P'(X_j|\pi_j)$ for each uncertain node $X_j \in \mathbf{X}$ and $\pi_j \in \Omega_{\Pi_j}$ is already specified by $\mathbb{P}'$. To determine a BN we have then to simply represent decision 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 function $P'_{\mathbf{s}}(X_i|\pi_i)$ assigning all the mass to the value $f_{X_i}(\pi_i) \in \Omega_{X_i}$, 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 $\mathbf{x}' = (\mathbf{x}_D, \mathbf{x}) \in \Omega_{\mathbf{X}'}$, the following factorization:

$$P'_{\mathbf{s}}(\mathbf{x}_D, \mathbf{x}) = \prod_{X_i \in \mathbf{X}_D} P'_{\mathbf{s}}(x_i|\pi_i) \cdot \prod_{X_j \in \mathbf{X}} P'(x_j|\pi_j). \tag{8}$$

---

[9]In [21] we use the expression "locally specified credal networks" to denote the same class of models.
[10]If X_i corresponds to a root node of $\mathcal{G}$, a single set equal to the whole Ω_{X_i} is considered.

The next step is then obvious: we want to define a CN over $\mathbf{X}$ by means of the set of BNs determined by all the possible strategies $\mathbf{s} \in \Omega_{\mathbf{S}}$. The question, at this point, is whether or not all these networks have the same DAG, as required by Definition 2. To show this we need to introduce the following transformation, that removes from $\mathcal{G}'$ the decision nodes by maintaining the dependence relations between the other nodes:

Transformation 2 *Given a decision-theoretic specification of a CN $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$, obtain a DAG $\mathcal{G}$ associated to the variables $\mathbf{X}$ iterating, for each decision node $X_i \in \mathbf{X}_D$, the following operations over $\mathcal{G}'$: (i) draw an arc from each parent of X_i to each child of X_i; (ii) remove the node X_i.*
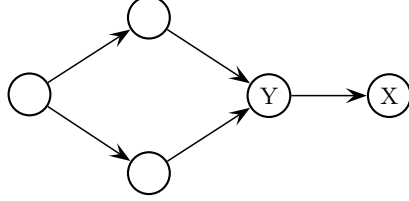

Figure 2: The DAG $\mathcal{G}$ returned by Transformation 2 given a decision-theoretic specification of a CN whose DAG is that in Figure 3 (or also Figure 4 or Figure 5 or Figure 6).

Figure 2 reports an example of the output of Transformation 2. The DAG $\mathcal{G}$ returned by Transformation 2 is considered by the next theorem.

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*

$$P_{\mathbf{s}}(\mathbf{x}) := \sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} P'_{\mathbf{s}}(\mathbf{x}_D, \mathbf{x}), \tag{9}$$

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 2.

From this, considering the BNs $\langle \mathcal{G}, \mathbb{P}_{\mathbf{s}} \rangle$ for each strategy $\mathbf{s} \in \Omega_{\mathbf{S}}$ as compatible BNs of a CN, it is possible to obtain the following result:

Corollary 1 *A decision-theoretic specification of a CN as in Definition 4 defines a CN over $\mathbf{X}$, based on the DAG $\mathcal{G}$ returned by Transformation 2.*

The strong extension of $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$ will be therefore intended as the strong extension $K(\mathbf{X})$ of the CN considered in Corollary 1. What we show in the next sections is how to provide decision-theoretic specifications of CNs, according to Definition 4, for both separately and non-separately specified CNs.

6.2 Decision-Theoretic Specification of Non-Separately Specified Credal Nets

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 3).

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 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 [5], is only seemingly local, because of the arcs connecting the decision node with all the uncertain nodes. However, in many cases, this is not the only way to provide a decision-theoretic specification of a CN.

Consider, for example, the class of extensively specified CNs introduced in Section 5. We can provide a decision-theoretic specification, as in Definition 4, of a CN of this kind by introducing a decision parent for each node of the original CN (Figure 4). 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. This means that, if X_i is 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, for each $\pi_i \in \Pi_i$, $P(X_i|d_i, \pi_i)$ is the mass function $P_{d_i}(X_i|\pi_i)$ associated to the d_i-th table of the extensive specification. E.g., the two tables defined in the extensive specification of the node B for the CN in Figure 1, can be indexed by a binary decision parent D:

$$P(B|a,d) = \begin{bmatrix} .2 \\ .8 \end{bmatrix} \quad P(B|\neg a, d) = \begin{bmatrix} .3 \\ .7 \end{bmatrix} \quad P(B|a, \neg d) = \begin{bmatrix} .4 \\ .6 \end{bmatrix} \quad P(B|\neg a, \neg d) = \begin{bmatrix} .5 \\ .5 \end{bmatrix}.$$

More generally, constraints for the specifications of conditional mass functions relative to different nodes can be similarly represented by decision nodes which are the parents of these nodes (see for example Figure 5).
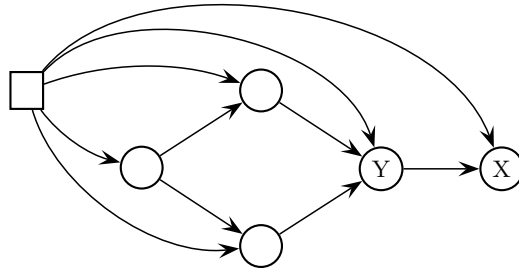

Figure 3: Decision-theoretic specification of a non-separately specified CN over the DAG in Figure 2. Remember that circles denote uncertain nodes, while the square is used for the decision node.
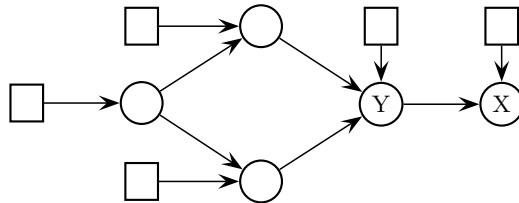

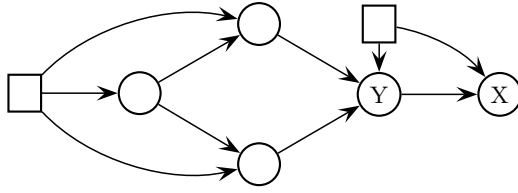Figure 4: Decision-theoretic specification of an extensive CN over the DAG in Figure 2.

Figure 5: Decision-theoretic specification of a non-separately specified CN over the DAG in Figure 2. Constraints between the specifications of the conditional credal sets of the nodes X and Y, and also between the three remaining nodes are assumed.

6.3 Decision-Theoretic Specification of Separately Specified Credal Nets

Finally, to provide a decision-theoretic specification, as required by Definition 4, of 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.

As an example, let us consider the node X of a separately specified CN defined over the DAG in Figure 2. Assume Y to be binary and both the credal sets $K(X|y)$ and $K(X|\neg y)$ to be made of three extreme mass functions. A requirement for the probability table $P(X|Y)$ to belong to a set of nine tables, obtained from all the possible combinations where the first column takes values in $\text{ext}[K(X|y)]$ and the second in $\text{ext}[K(X|\neg y)]$, is clearly equivalent to leave the conditional probability mass functions $P(X|y)$ and $P(X|\neg y)$ to vary in the relative credal sets.

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 (uncertain) node and its parents, according to the following graphical transformation.

Transformation 3 *Obtain a DAG $\mathcal{G}'$ from a DAG $\mathcal{G}$ over $\mathbf{X}$ by iterating, for each $X_i \in \mathbf{X}$, the following operations: (i) add a decision node D_i; (ii) draw an arc from each parent of X_i to D_i; (iii) delete the arcs connecting the parents of X_i with X_i; (iv) draw an arc from D_i to X_i.*

It is straightforward to check that Transformation 3 requires only a number of operations linear in the input size.

Given a separately specified CN $\langle \mathcal{G}, \mathbb{K} \rangle$ over $\mathbf{X}$, it is possible to consider a decision-theoretic specification of a CN $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$, where $\mathcal{G}'$ is the DAG returned by Transformation 3, $\mathbf{D}$ is the set of decision nodes (one for each node) added by the same transformation. To complete the decision-theoretic specification proceed as follows. For each uncertain node X_i, consider the set $\bigcup_{\pi_i \in \Omega_{\Pi_i}} \text{ext}[K(X_i|\pi_i)]$, i.e., the union of the extreme mass functions of *all* the conditional credal sets specified for X_i. Let the states $d_i \in \Omega_{D_i}$ of the corresponding decision node D_i to index the elements of this set. Accordingly, for each uncertain node X_i, the conditional mass function $P'(X_i|d_i)$ corresponds to 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 $K(X_i|\pi_i)$. For this approach, which is clearly polynomial in the overall number of vertices of the conditional credal sets in $\mathbb{K}$, we have taken inspiration from *probability trees* representations, as defined in [22].[11]

---

[11]It should be pointed out that the probability tree representation is different as it adds a variable for each
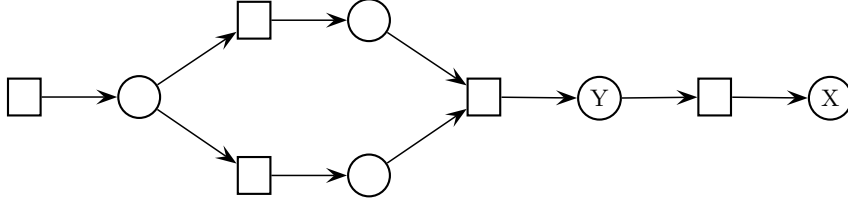
Figure 6: Decision-theoretic specification of a separately specified CN over the DAG in Figure 2.

As an example, consider the decision-theoretic specification of a separately specified CN over the DAG $\mathcal{G}$ in Figure 2. The output $\mathcal{G}'$ returned by Transformation 3 is the DAG in Figure 6. Regarding the quantification, consider for instance the procedure for the node X. Assume Y to be binary, and let $\text{ext}[K(X|y)] = \{P_1(X|y), P_2(X|y), P_3(X|y)\}$ and $\text{ext}[K(X|\neg y)] = \{P_4(X|\neg y), P_5(X|\neg y), P_6(X|\neg y)\}$. For the decision node D added in between X and Y, we set $\Omega_D := \{1, 2, 3, 4, 5, 6\}$, and, for the subsets of Ω_D corresponding to the possible values of Y, $\Omega_D^y := \{1, 2, 3\}$ and $\Omega_D^{\neg y} := \{4, 5, 6\}$, whereas, regarding X, we set $P'(X|D = d) := P_d(X|y_d)$ for each $d \in \Omega_D$ (where clearly $y_d := y$ if $d \in \Omega_D^y$ and $y_d := \neg y$ if $d \in \Omega_D^{\neg y}$).

7 From Decision-Theoretic to Separate Specification of Credal Networks

The transformations described in Section 6.2 and in Section 6.3 can be used to obtain in polynomial time a decision-theoretic specification of a CN of any kind.[12] In this section, we prove that any decision-theoretic specification of a CN over $\mathbf{X}$ can equivalently be regarded as a separate specification of a CN over $\mathbf{X}' := (\mathbf{X}_D, \mathbf{X})$. This transformation is technically straightforward: it is based on representing decision nodes by uncertain nodes (Figure 7) with vacuous conditional credal sets, as formalized below.

Transformation 4 *Given a decision-theoretic specification of a CN $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$ over $\mathbf{X}$, obtain a separately specified CN $\langle \mathcal{G}', \mathbb{K} \rangle$ over $\mathbf{X}' := (\mathbf{X}_D, \mathbf{X})$, where the conditional credal sets in $\mathbb{K}$ are as follows, for each $X_i \in \mathbf{X}$ and $\pi_i \in \Omega_{\Pi_i}$:*

$$
K(X_i|\pi_i) := \begin{cases} P'(X_i|\pi_i) & \text{if } X_i \in \mathbf{X} \\[2mm] K_{\Omega_{X_i}^{\pi_i}}(X_i) & \text{if } X_i \in \mathbf{X}_D, \end{cases} \tag{10}
$$

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 $\Omega_{X_i}^{\pi_i}$.

The (strong) relation between a decision-theoretic specification of a CN $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$ over $\mathbf{X}$ and the separately specified CN $\langle \mathcal{G}', \mathbb{K} \rangle$ over $\mathbf{X}' := (\mathbf{X}_D, \mathbf{X})$ returned by Transformation 4 is outlined by the following result:

---

configuration of the parents. Nevertheless the complexity of the representation does not increase as probability trees can represent asymmetrical irrelevance relationships.

[12] 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 decision-theoretic representation of a credal net.
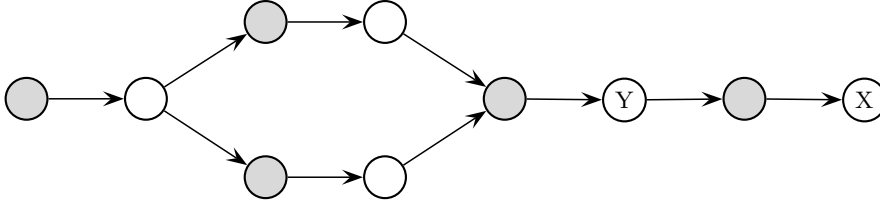
Figure 7: The DAG associated to the separately specified CN returned by Transformation 4, from the decision-theoretic specification of the CN based on the DAG in Fig. 6. The conditional credal sets of the white nodes (corresponding to the original uncertain nodes) are precisely specified, while the gray 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 $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$. Then:*

$$K(\mathbf{X}) = \tilde{K}(\mathbf{X}). \tag{11}$$

From Theorem 2, it is straightforward to obtain the following result:

Corollary 2 *Any inference problem on a CN obtained by a decision-theoretic specification can be equivalently solved in the separately specified CN returned by Transformation 4.*

Let us stress that Transformation 4 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 obtain a decision-theoretic specification using the prescriptions of Section 6.2, and apply Transformation 4 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. In the following sections, two examples of applications of this procedure will be presented. [13]

8 An Application: 2U for Extensive Specifications

The NP-hardness of CNs belief updating has been proved even for singly connected topologies [23]. Nevertheless, a singly connected CN with binary variables can be efficiently updated by the *2U algorithm* [7]. At the present moment, 2U is the only polynomial-time algorithm for exact updating of CNs, but it is designed only for separately specified CNs. Here we show how 2U can be readily extended to deal exactly and efficiently also with extensively specified CNs.

Consider a singly connected CN as in Figure 8.a, defined over a set of binary variables with extensive specification of the conditional probability tables. According to the discussion of Section 6.2, a decision-theoretic specification of this CN can obtained by simply adding to each node a decision parent indexing the different tables. Instead of a single decision parent, a set of binary decision parents whose joint states correspond to the tables can be equivalently adopted (Figure 8.b). This means that, for example, a set of four probability tables providing

---

[13] It should be pointed out that the transformation described in Section 6.3, returning a decision-theoretic specification of a separately specified CN, is not the inverse of Transformation 4, as the sequential application of the two transformations produces a model defined over a larger domain.

an extensive specification of a node can be indexed by a single decision parent with four states, or by the joint states of two binary decision parents. Clearly, if the number of tables is not an integer power of two, this procedure introduces a number of redundant joint states for the decision parents.

Finally, from the decision-theoretic specification, we obtain a separately specified CN through Transformation 4 (Figure 8.c). The overall procedure preserves the topology of the CN, which remains singly connected, and is still defined over binary variables only. We can therefore update the CN by 2U without making any change to the algorithm itself.
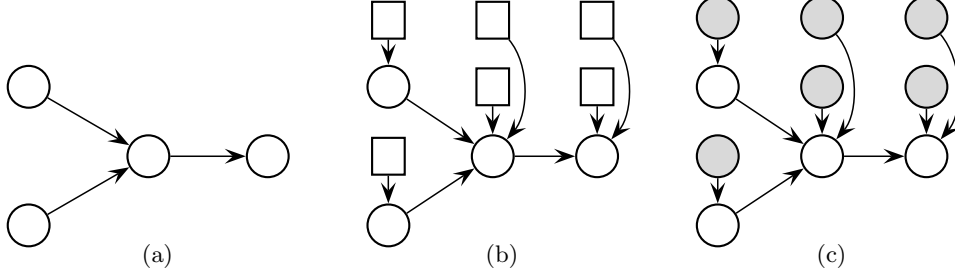

Figure 8: (a) A singly connected CN over four binary variables; (b) its decision-theoretic specification with binary decision parents, assuming extensive specifications by sets of two tables for the root nodes, and four tables for the others; (c) the separately specified CN returned by Transformation 4.

9 Application to Conservative Inference Rule

As a more involved application of the results in Section 6 and Section 7, let us consider the CIR-based updating problem detailed in Section 5.1. In Section 9.1, the problem is mapped into a standard updating problem on a separately specified CN, which can be updated by standard techniques. The result follows from a general equivalence relation regarding CIR-based inference in general. In the special case of updating, we also obtain, by similar transformations, a hardness proof in Section 9.2.

9.1 Algorithms for CIR-based Inference

Consider a BN over $\mathbf{X} := (\tilde{\mathbf{X}}, \mathbf{X}_I)$, assigning positive probability to any joint state, where $\mathbf{X}_I$ are the variables missing by a process we do not know. As shown in Equation (4) in the special case of updating, CIR-based inference requires the evaluation of all the possible completions of the missing variables $\mathbf{X}_I$. This is equivalent to making inferences using the following credal set:

$$K_{\mathbf{X}_I}(\tilde{\mathbf{X}}) := \mathrm{CH}\{P(\tilde{\mathbf{X}}|\mathbf{x}_I)\}_{\mathbf{x}_I \in \Omega_{\mathbf{X}_I}}, \tag{12}$$

where the conditional mass function $P(\tilde{\mathbf{X}}|\mathbf{x}_I)$ is obtained from the joint mass function $P(\mathbf{X})$ associated to the BN.

The BN becomes an extensively specified CN over $(\tilde{\mathbf{X}}, \mathbf{X}_I, \mathbf{X}'_I)$ after Transformation 1. A decision-theoretic specification of this non-separately specified CN can be indeed obtained by simply adding to each node $X' \in \mathbf{X}'_I$ a decision parent of X', say X'', indexing the different tables. Such decision-theoretic CN specification corresponding to the CIR-based inference problem is considered in the following theorem.

14

Theorem 3 *The following equivalence between credal sets holds:*

$$K_{\mathbf{X}_I}(\tilde{\mathbf{X}}) = K(\tilde{\mathbf{X}}|\overline{\mathbf{x}}'_I), \tag{13}$$

where the conditional credal set on the right-hand side is obtained from the strong extension of the decision-theoretic CN specification corresponding to the CIR-based inference problem.

Equation (13) can be used to map general CIR-based inference problems in BNs into corresponding standard inferences on CNs. The equivalence with respect to CIR-based updating in Equation (6) can be regarded as an obvious corollary of Theorem 3.

Finally, according to Theorem 2, the conditional credal set on the right-hand side of Equation (13) can be equivalently obtained from the strong extension of the separately specified CN returned by Transformation 4. Overall, this procedure, which is illustrated in Figure 9, maps CIR-based inference problems in BNs into corresponding problems in separately specified CNs, for which existing algorithms can be employed. An analogous procedure could be developed to address CIR-based inference problems on CNs.
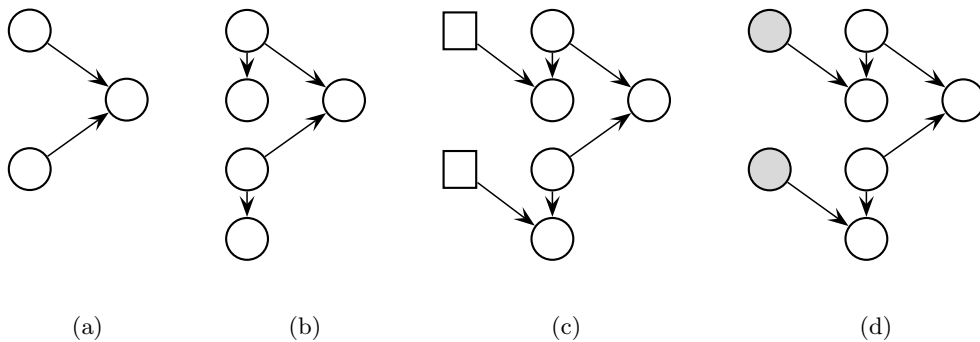

(a)   (b)   (c)   (d)

Figure 9: (a) A CIR-based inference problem on a BN where the missing variables $\mathbf{X}_I$ are those corresponding to the root nodes; (b) the corresponding extensive CN returned by Transformation 1; (c) the decision-theoretic specification of this extensive CN; (d) the separately specified CN returned by Transformation 4.

9.2 Hardness of CIR-based Updating

In a recent work [24], CIR-based classification on BNs has been proved to be NP-hard.[14] In this section we obtain a similar result for the updating problem.

To this extent, consider the class of separately specified CNs such that the specification of the non-root nodes is precise, i.e., the corresponding conditional credal sets are reduced to a single conditional probability mass function. These CNs are said to be *with precise non-root nodes* and are considered in the following theorem.

Theorem 4 *Any updating problem on a CN with precise non-root nodes can be mapped into a CIR-based updating problem on a BN in linear time.*

---

[14]The proof in [24] considers classification problems based on the *conservative updating rule* [8], a special case of CIR, that assumes the set of variables missing in a MAR way to be empty.

Remarkably, the hardness proof of updating with CNs reported in [23, Theorem 3] is based on the reduction of a *Boolean satisfiability* [25] problem to the updating of a (singly-connected) CN with precise non-root nodes. According to Theorem 4, we obtain therefore the following result:

Corollary 3 *CIR-based updating on BNs is NP-hard.*

10 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 updating algorithms are available in the literature.

Two examples of applications of this procedure have been detailed: the generalization of the 2U algorithm to the extensive case, and a general algorithmic procedure to solve CIR-based inference on Bayesian networks. Additionally, we have also exploited our formalism to prove the NP-hardness of CIR-based updating on Bayesian networks.

Many other developments seem to be possible. First of all it is important to note that the proposed transformation 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. A recent development of the approximate L2U algorithm [26] is 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.

Acknowledgments

A Proofs of the Theorems

The proofs of Corollary 1, Corollary 2 and Corollary 3 are obvious and hence omitted.

Proof of Theorem 1 *Let us start the marginalization in Equation (9) from a decision node $X_j \in \mathbf{X}_D$. According to Equation (8), for each $\mathbf{x}' \in \Omega_{\mathbf{X}'}$:*

$$\sum_{\mathbf{x}_j \in \Omega_{\mathbf{X}_j}} P'_{\mathbf{s}}(\mathbf{x}') = \sum_{\mathbf{x}_j \in \Omega_{\mathbf{X}_j}} \left[\prod_{X_l \in \mathbf{X}_D} P'_{\mathbf{s}}(x_l|\pi_l) \cdot \prod_{X_i \in \mathbf{X}} P'(x_i|\pi_i) \right]. \tag{14}$$

Thus, moving out of the sum the conditional probabilities that do not refer to the states of X_j (which are briefly denoted by Δ), Equation (14) becomes:

$$\Delta \cdot \sum_{x_j \in \Omega_{X_j}} \left[P'_{\mathbf{s}}(x_j|\pi_j) \cdot \prod_{X_r \in \Gamma_{X_j}} P'(x_r|x_j, \tilde{\pi}_r) \right], \tag{15}$$

16

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 (15) rewrites as

$$\Delta \cdot \prod_{X_r \in \Gamma_{X_j}} P'(x_r|f_{X_j}(\pi_j), \tilde{\pi}_r). \tag{16}$$

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

$$P_{\mathbf{s}}(X_r|\overline{\pi}_r) := P'(X_r|f_{X_j}(\pi_j), \tilde{\pi}_r), \tag{17}$$

to regard Equation (16) as the joint mass function of a BN over $\mathbf{X}' \setminus \{X_j\}$ based on the DAG returned by Transformation 2 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 intuitive proposition is required to obtain Theorem 2. It is proved here only 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 product of vertices of the separately specified conditional credal sets, i.e., for each $\mathbf{x} \in \Omega_{\mathbf{X}}$:*

$$\tilde{P}_j(\mathbf{x}) = \prod_{i=1}^n \tilde{P}_j(x_i|\pi_i), \tag{18}$$

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 *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 product 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 (18), 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 $\mathbf{x} \in \Omega_{\mathbf{X}}$,*

$$\tilde{P}(\mathbf{x}) = \left[\sum_\alpha c_\alpha P_\alpha(x_t|\pi_t) \right] \cdot \prod_{i \neq t} P(x_i|\pi_i), \tag{19}$$

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 *Consider a vertex of the strong extension of $\langle \mathcal{G}', \mathbb{K} \rangle$, i.e., a joint mass function $\tilde{P}(\mathbf{X}') \in \text{ext}[\tilde{K}(\mathbf{X}')]$. According to Proposition 1, $\tilde{P}(\mathbf{X}')$ can be obtained by the product, as in Equation (18), of a combination of vertices of the conditional credal sets in $\mathbb{K}$. Thus, for each $X_i \in \mathbf{X}_D$ and $\pi_i \in \Omega_{\Pi_i}$, $\tilde{P}(X_i|\pi_i)$ is a vertex of the vacuous credal set $K_{\Omega_{X_i}^{\pi_i}}(X_i)$, i.e., a degenerate mass function over X_i assigning all the mass to a single $\tilde{x}_i \in \Omega_{X_i}$. Consider, on $\langle \mathcal{G}', \mathbb{O}, \mathbb{P}' \rangle$, the decision function f_{X_i} of X_i such that $f_{X_i}(\pi_i) = \tilde{x}_i$ for each $\pi_i \in \Omega_{\Pi_i}$. Let $\tilde{\mathbf{s}} \in \Omega_{\mathbf{S}}$ be the strategy corresponding to the array of decision functions selected in this way for each*

$X_i \in \mathbf{X}_D$. *Clearly $P_{\tilde{\mathbf{s}}}(\mathbf{X}') = \tilde{P}(\mathbf{X}')$. Thus, considering all the vertices of $\tilde{K}(\mathbf{X}')$, we conclude $\tilde{K}(\mathbf{X}') \subseteq K(\mathbf{X}')$.*

In order to prove the inverse inclusion, given a strategy $\mathbf{s} \in \Omega_{\mathbf{S}}$, consider the joint mass function $P_{\mathbf{s}}(\mathbf{X}')$ associated to the BN $\langle \mathcal{G}', \mathbb{P}'_{\mathbf{s}} \rangle$. As shown in Section 6.1, the elements of $\mathbb{P}'_{\mathbf{s}}$ corresponding to the nodes $X_i \in \mathbf{X}$ are just the same conditional probability mass functions $P'(X_i|\pi_i)$ specified in Equation (10). On the other side, the elements of $\mathbb{P}'_{\mathbf{s}}$ corresponding to the nodes $X_i \in \mathbf{X}_D$ are (degenerate) mass functions reproducing the decision functions of $\mathbf{s}$, and should therefore belong to the vacuous credal sets in Equation (10). Thus, $P_{\mathbf{s}}(\mathbf{X}') \in \tilde{K}(\mathbf{X}')$, and hence $K(\mathbf{X}') \subseteq \tilde{K}(\mathbf{X}')$.

Overall we have $K(\mathbf{X}') = \tilde{K}(\mathbf{X}')$, from which the thesis follows by a simple marginalization.

Proof of Theorem 3 *Let $P(\tilde{\mathbf{X}}, \mathbf{X}_I)$ be the joint probability mass function associated to the BN. Let also $\mathbf{X}''_I$ denote the array of decision nodes of the CN. For each $X \in \mathbf{X}_I$, the corresponding elements of $\mathbf{X}'_I$ and $\mathbf{X}''_I$ are indicated as X' and X'', i.e., X' is the auxiliary child added to X according to Transformation 1, and X'' is the decision parent added to X' according to the prescriptions in Section 6.2. Let also $\overline{x}' \in \Omega_{X'}$ denote the component corresponding to X' of $\overline{\mathbf{x}}'_I \in \Omega_{\mathbf{X}'_I}$ (i.e., the value of the binary variable X' corresponding to the first row of the tables $P(X'|X)$ in Equation (5)). Note that X'' indexes the set of probability tables $P(X'|X)$ in Equation (5), which are in correspondence with the elements of Ω_X. We can therefore set $\Omega_{X''} := \Omega_X$ and regard the state $X'' = x$, for each $x \in \Omega_X$, as the index of the table in Equation (5) such that $P(X' = \overline{x}'|X = x) = 1$. Thus, the elements of $\Omega_{\mathbf{X}''_I}$, indexing the compatible BNs of the CN, can be identified with those of $\Omega_{\mathbf{X}_I}$. Let $P_{\mathbf{X}''_I = \hat{\mathbf{x}}_I}(\tilde{\mathbf{X}}, \mathbf{X}_I, \mathbf{X}'_I)$ denote, for each $\hat{\mathbf{x}}_I \in \Omega_{\mathbf{X}''_I}$, the joint probability mass function of the compatible BN corresponding to $\hat{\mathbf{x}}_I$. The following factorization clearly holds:*

$$P_{\mathbf{X}''_I = \hat{\mathbf{x}}_I}(\tilde{\mathbf{x}}, \mathbf{x}_I, \mathbf{x}'_I) = P(\tilde{\mathbf{x}}, \mathbf{x}_I) \cdot \prod_{X \in \mathbf{X}_I} \left[P_{\mathbf{X}''_I = \hat{\mathbf{x}}_I}(x'|x) \right]. \tag{20}$$

According to Equation (5), we have:

$$P_{\hat{\mathbf{X}}''_I = \hat{\mathbf{x}}_I}(\overline{x}'|x) = \delta_{x,\hat{x}}, \tag{21}$$

where $\delta_{x,\hat{x}}$ is equal to one if and only if $x = \hat{x}$ and zero otherwise. Thus, from Equation (20) and Equation (21), it follows that:

$$\sum_{\mathbf{x}_I \in \Omega_{\mathbf{X}_I}} P_{\mathbf{X}''_I = \hat{\mathbf{x}}_I}(\tilde{\mathbf{x}}, \mathbf{x}_I, \overline{\mathbf{x}}'_I) = P(\tilde{\mathbf{x}}, \hat{\mathbf{x}}_I). \tag{22}$$

From this we obtain

$$P_{\mathbf{X}''_I = \hat{\mathbf{x}}_I}(\tilde{\mathbf{X}}|\overline{\mathbf{x}}'_I) = P(\tilde{\mathbf{X}}|\mathbf{X}_I = \hat{\mathbf{x}}_I). \tag{23}$$

The thesis follows by simply considering Equation (23) for each $\hat{\mathbf{x}}_I \in \Omega_{\mathbf{X}_I}$.

Proof of Theorem 4 *Consider a (separately specified) CN over $\mathbf{X}$ with precise non-root nodes. A decision-theoretic specification of this CN can be obtained by simply adding to each root node a decision parent node indexing the vertices of the unconditional credal set associated to this node. Let $\mathbf{X}_D$ be the decision nodes added to the CN by the transformation. For each $\mathbf{x}_D \in \Omega_{\mathbf{X}_D}$, let $P_{\mathbf{x}_D}(\mathbf{X})$ be the joint probability mass function associated to the compatible BN indexed by $\mathbf{x}_D$.*

Obtain a BN from the decision-theoretic specification of the CN, by simply regarding the decision root nodes $\mathbf{X}_D$ as uncertain nodes for which uniform unconditional mass functions

are specified. Let $\tilde{P}(\mathbf{X}, \mathbf{X}_D)$ be the joint probability mass function associated to this BN. It is straightforward to check that, for each $\mathbf{x}_D \in \Omega_{X_D}$,

$$\tilde{P}(\mathbf{x}|\mathbf{x}_D) = P_{\mathbf{x}_D}(\mathbf{x}). \tag{24}$$

Thus, a generic updating problem on the CN can be mapped into a CIR-based updating problem on this BN, by simply assuming the variables $\mathbf{X}_D$ to be missing by an unknown mechanism.

References

[1] F. G. Cozman, Graphical models for imprecise probabilities, International Journal of Approximate Reasoning 39 (2–3) (2005) 167–184.

[2] I. Levi, The Enterprise of Knowledge, MIT Press, London, 1980.

[3] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, New York, 1991.

[4] M. Zaffalon, Statistical inference of the naive credal classifier, in: G. de Cooman, T. Fine, T. Seidenfeld (Eds.), Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications (ISIPTA '01), Shaker Publishing, The Netherlands, 2001, pp. 384–393.

[5] A. Cano, J. Cano, S. Moral, Convex sets of probabilities propagation by simulated annealing on a tree of cliques, in: Proceedings of Fifth International Conference on Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU '94), 1994, pp. 4–8.

[6] N. Zhang, R. Qi, D. Poole, A computational theory of decision networks, International Journal of Approximate Reasoning 11 (2) (1993) 83–158.

[7] E. Fagiuoli, M. Zaffalon, 2U: an exact interval propagation algorithm for polytrees with binary variables, Artificial Intelligence 106 (1) (1998) 77–107.

[8] G. de Cooman, M. Zaffalon, Updating beliefs with incomplete observations, Artificial Intelligence 159 (2004) 75–125.

[9] A. Antonucci, M. Zaffalon, Equivalence between Bayesian and credal nets on an updating problem, in: J. Lawry, E. Miranda, A. Bugarin, S. Li, M. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Integrated Uncertainty Modelling (Proceedings of the third international conference on Soft Methods in Probability and Statistics: SMPS 2006), Springer, 2006, pp. 223–230.

[10] F. G. Cozman, Credal networks, Artificial Intelligence 120 (2000) 199–233.

[11] J. C. Ferreira da Rocha, F. G. Cozman, Inference with separately specified sets of probabilities in credal networks, in: A. Darwiche, N. Friedman (Eds.), Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI 2002), Morgan Kaufmann, San Francisco, 2002, pp. 430–437.

[12] M. Zaffalon, Conservative rules for predictive inference with incomplete data, in: F. Cozman, R. Nau, T. Seidenfeld (Eds.), Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA '05), SIPTA, 2005, pp. 406–415.

[13] R. J. A. Little, D. B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.

[14] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, 1988.

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

[16] F. G. Cozman, C. P. de Campos, J. S. Ide, J. C. Ferreira da Rocha, Propositional and relational Bayesian networks associated with imprecise and qualitative probabilistic assessments, in: M. Chickering, J. Halpern (Eds.), Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence (UAI 2004), AUAI Press, Arlington, Virginia, 2004, pp. 104–111.

[17] T. Verma, J. Pearl, Equivalence and synthesis of causal models, in: P. Bonissone, M. Henrion, L. Kanal, J. Lemmer (Eds.), Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence (UAI 1990), Elsevier, New York, 1991, pp. 255–270.

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

[19] P. Walley, Inferences from multinomial data: learning about a bag of marbles, J. R. Statist. Soc. B 58 (1) (1996) 3–57.

[20] A. Cano, M. Gómez-Olmedo, S. Moral, Credal nets with probabilities estimated with an extreme imprecise dirichlet model, in: G. de Cooman, I. Vejnarová, M. Zaffalon (Eds.), Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '07), Action M Agency, Prague, 2007, pp. 57–66.

[21] A. Antonucci, M. Zaffalon, Locally specified credal networks, in: M. Studený, J. Vomlel (Eds.), Proceedings of the third European Workshop on Probabilistic Graphical Models (PGM-2006), Action M Agency, Prague, 2006, pp. 25–34.

[22] A. Cano, S. Moral, Using probability trees to compute marginals with imprecise probabilities, International Journal of Approximate Reasoning 29 (1) (2002) 1–46.

[23] C. P. de Campos, F. G. Cozman, The inferential complexity of Bayesian and credal networks, in: Proceedings of the International Joint Conference on Artificial Intelligence, Edinburgh, 2005, pp. 1313–1318.

[24] A. Antonucci, M. Zaffalon, Fast algorithms for robust classification with Bayesian nets, International Journal of Approximate Reasoning 44 (3) (2007) 200–223.

[25] M. R. Garey, D. S. Johnson, Computers and Intractability; a Guide to the Theory of NP-completeness, Freeman, San Francisco, 1979.

[26] A. Antonucci, M. Zaffalon, J. Ide, F. Cozman, Binarization algorithms for approximate updating in credal nets, in: L. Penserini, P. Peppas, A. Perini (Eds.), Proceedings of the third European Starting AI Researcher Symposium (STAIRS '06), IOS Press, Amsterdam, 2006, pp. 120–131.