The Inferential Complexity of Bayesian and Credal Networks

Cassio Polpo de Campos[1,2] **Fabio Gagliardi Cozman**[1]
[1] Universidade de São Paulo - Escola Politécnica
[2] Pontifícia Universidade Católica de São Paulo
cassio@pucsp.br, fgcozman@usp.br

Abstract

This paper presents new results on the complexity of graph-theoretical models that represent probabilities (Bayesian networks) and that represent interval and set valued probabilities (credal networks). We define a new class of networks with bounded width, and introduce a new decision problem for Bayesian networks, the maximin a posteriori. We present new links between the Bayesian and credal networks, and present new results both for Bayesian networks (most probable explanation with observations, maximin a posteriori) and for credal networks (bounds on probabilities a posteriori, most probable explanation with and without observations, maximum a posteriori).

1 Introduction

This paper builds a picture of inferential complexity in graph-theoretical models of uncertainty that goes significantly beyond existing results. We focus on Bayesian and credal networks — the former is a purely probabilistic model, while the latter admits interval and set valued probabilities and generalizes several theories of uncertainty. There already is quite a solid understanding about the inferential complexity of Bayesian networks; we add to this picture a new class of networks that strengthens existing results, and a new type of problem, maximin a posteriori (MmAP), that can be of interest in game-theoretic settings. Our main contributions are related to credal networks, as little is known at this point concerning their inferential complexity: we present new results concerning computation of probability bounds, most probable explanations (MPE) and maximum a posteriori (MAP). We show that, rather surprisingly, the MPE problem without observations in credal networks of bounded width is polynomial, while the MPE problem with observations and the MAP problem are Σ_2^p-complete for credal networks.

We also strengthen the connection between Bayesian and credal networks by showing links between the computation of probability bounds, the MAP and the MmAP problems. Our results suggest that moving from point to interval/set valued probabilities takes us "one step up" in terms of complexity classes.

Section 2 presents a few definitions and, most importantly, tries to justify our interest in credal networks. Section 3 describes the problems we are interested in. Section 4 contains our new results; in particular, Table 1 offers a summary of our contributions. Section 5 concludes the paper and suggests future research.

2 Bayesian and Credal Networks

A Bayesian network (or `BN`) represents a *single* joint probability density over a collection of random variables. We assume throughout that variables are categorical; variables are uppercase and their assignments are lowercase.

Definition A Bayesian network is a pair $(G, \mathbb{P})$, where: $G = (V_G, E_G)$ is a directed acyclic graph, with V_G a collection of vertices associated to random variables $\mathbb{X}$ (a node per variable), and E_G a collection of arcs; $\mathbb{P}$ is a collection of conditional probability densities $p(X_i|\mathrm{pa}(X_i))$ where $\mathrm{pa}(X_i)$ denotes the parents of X_i in the graph ($\mathrm{pa}(X_i)$ may be empty), respecting the relations of E_G.

In a `BN` every variable is independent of its nondescendants nonparents given its parents (Markov condition). This structure induces a joint probability density by the expression $p(X_1, \ldots, X_n) = \prod_i p(X_i|\mathrm{pa}(X_i))$. Given a `BN`, E denotes the set of observed variables in the network (the *evidence*); e denotes the observed value of E.

We consider the class of networks with `bounded induced-width` (or `BIW`); that is, networks where the subjacent graph (obtained removing arc's directions) of G has maximum degree and induced-width bounded by $\log(f(s))$. We define $f(s)$ as a polynomial function in the size s of input (this size is evaluated over all information needed to specify the problem). We also consider the class of `polytrees` (or `PT`); that is, a `BIW` network where the subjacent graph of G has no cycles. Otherwise, a network is said `multiply-connected`. Note that our definition of `BIW` networks is not, as usually done, based on *fixed* induced-width; rather, we allow the induced-width to vary with the network size.

Credal networks generalize Bayesian networks by allowing each variable to be associated with *sets* of joint probability measures rather than single probability measures [Cano *et al.*, 1993; Cozman, 2000a]. Such graph-theoretical models can be viewed as Bayesian networks with relaxed numerical parameters; they can be used to study robustness of probabilistic models, to investigate the behavior of groups of experts, or to represent incomplete or vague knowledge about probabilities. Sets of probability measures are sufficiently powerful to represent belief functions, possibility measures, qualitative probabilities, and probabilistic logic statements, thus offering a general language that can convey many models of interest in artificial intelligence [Walley, 1996].

A set of probability distributions is called a *credal set* [Levi, 1980]. A credal set for X is denoted by $K(X)$. A conditional credal set is a set of conditional distributions, obtained applying Bayes rule to each distribution in a credal set of joint distributions [Walley, 1991]. Given a credal set $K(X)$, the *lower probability* and the *upper probability* of event A are defined respectively as $\underline{P}(A) = \min_{p(X)\in K(X)} P(A)$ and $\overline{P}(A) = \max_{p(X)\in K(X)} P(A)$. We assume that, given a credal set, finding a lower/upper probability is a polynomial operation.

Definition A credal network is a pair $(G, \mathbb{K})$, where: $G = (V_G, E_G)$ is a directed acyclic graph, with V_G a collection of vertices associated to random variables $\mathbb{X}$ (a node per variable), and E_G a collection of arcs; $\mathbb{K}$ is a collection of conditional credal sets $K(X_i|\mathrm{pa}(X_i))$, respecting the relations implied by E_G.

We assume a Markov condition in credal networks: every variable is independent of its nondescendants nonparents given its parents. In this paper we adopt the concept of *strong independence*; that is, the joint credal set represented by a credal network is a set where each vertex factorizes as a Bayesian network [Cozman, 2000b]. Thus we can really view a credal network as a set of Bayesian networks, all with identical graphs. Credal networks can also be classified as polytree, bounded induced-width or multiply-connected.

3 Reasoning

In this section we present decision versions of the inferences we focus in this paper.

Definition `BN-Pr`: given a Bayesian network $(G, \mathbb{P})$, evidence $E = e$ with $E \subseteq \mathbb{X}$, a query variable $Q \in \mathbb{X} \setminus E$ and its category q, and a rational number r, is $P(q|e) > r$?

Note the following distinction between MPE problems with or without evidence; the reason for this will be indicated later:

Definition `BN-MPE`: given a Bayesian network $(G, \mathbb{P})$, evidence $E = e$ with $E \subseteq \mathbb{X}$ and a rational number r, is there an instantiation x for $\mathbb{X} \setminus E$ such that $P(x, e) > r$?

Definition `BN-MPEe`: given a Bayesian network $(G, \mathbb{P})$, evidence $E = e$ with $E \subseteq \mathbb{X}$ and a rational number r, is there an instantiation x for $\mathbb{X} \setminus E$ such that $P(x|e) > r$?

Definition `BN-MAP`: given a Bayesian network $(G, \mathbb{P})$, evidence $E = e$ with $E \subseteq \mathbb{X}$, a set $Q \subseteq \mathbb{X} \setminus E$ and a rational number r, is there an instantiation q for Q such that $P(q|e) > r$?

We introduce the *maximin a posteriori* problem, which may be of interest in applications involving game-theoretic behavior with maximizers and minimizers [Kakade *et al.*, 2001]:

Definition `BN-MmAP`: given a Bayesian network $(G, \mathbb{P})$, some evidence $E = e$ with $E \subseteq \mathbb{X}$, the sets $A \subseteq \mathbb{X} \setminus E$ and $B \subseteq \mathbb{X} \setminus E$, with $A \cap B = \emptyset$ and a rational number r, is there an instantiation a for the A variables such that $\min_b P(a, b|e) > r$?

The corresponding problems in `CN` are now defined.

Definition `CN-Pr`: given a credal network $(G, \mathbb{K})$, evidence $E = e$ with $E \subseteq \mathbb{X}$, a query variable $Q \in \mathbb{X} \setminus E$ and its category q and a rational number r, is $\overline{P}(q|e) > r$?

In this definition we use upper queries, but lower queries may be of interest too. We can show that both queries lead to identical complexity results:

Lemma 1 *Evaluating marginal lower probabilities in* `CN` *is as hard as evaluating marginal upper probabilities.*

Proof Suppose we have a `CN-Pr` with marginal query $Q = q$. The calculation of $\underline{P}(q|e)$ can be done by inserting a binary child Q' to Q, where $P(q'|Q) = 1$ if $Q \neq q$ and 0 otherwise. Now $\overline{P}(q'|e) = \max \sum_{Q \neq q} P(Q|e) = 1 - \underline{P}(q|e)$. □

We now define *maximin* versions of MPE and MAP in credal networks (we could alternatively define maximax versions for these problems, with possibly different complexities).

Definition `CN-MAP`: given a credal network $(G, \mathbb{K})$, some evidence $E = e$ with $E \subseteq \mathbb{X}$, a set $Q \subseteq \mathbb{X} \setminus E$ and a rational number r, is there an instantiation q for the Q variables such that $\underline{P}(q|e) > r$?

`CN-MPEe` is obtained when `CN-MAP` has $Q = \mathbb{X} \setminus E$. `CN-MPE` is `CN-MPEe` without evidence, that is, $Q = \mathbb{X}$. We use abbreviations to refer to these problems (e.g. `PT-CN-Pr` is the belief updating problem in a polytree credal network).

4 Complexity results

Table 1 summarizes relevant complexity results.[1] We start by explaining the origin of the results in this table (numbering matches the numbers in the table):

1. [Dechter, 1996] describes algorithms with exponential time complexity on the induced-width of an elimination order; [Eyal, 2001] shows how to obtain a constant-factor approximation to optimal order in polynomial time for networks with bounded degree and bounded induced-width by $\log(f(s))$. Note that the result here is sightly different from the one in [Dechter, 1996]; we refer to the actual induced width of the graph, not the induced width of an ordering.
2. [Roth, 1996] shows complexity of functional version, [Litmman *et al.*, 2001] takes the decision version.
3. [Shimony, 1994] shows (by reduction from the vertex cover problem) that `BN-MPE` is NP-Complete, while Theorem 2 shows that `BN-MPEe` is PP-Complete.

[1] We assume that the reader is familiar with notions of complexity theory; for an introduction see [Papadimitriou, 1994]. Polynomial time means polynomial time in the size of input. A *reduction* means a polynomial time reduction; when a problem *is solvable by* another, there is a *reduction* from the former to the latter.

Problem	*Polytree*	*Bounded induced-width*	*Multiply-connected*
BN-Pr	Polynomial (1)	Polynomial (1)	PP-Complete (2)
BN-MPE	Polynomial (1)	Polynomial (1)	NP-Complete (3)
BN-MPEe	Polynomial (1)	Polynomial (1)	PP-Complete (3)
BN-MAP	NP-Complete (4)	NP-Complete (4)	NP^{PP}-Complete (5)
BN-MmAP	Σ_2^{P}-Complete (6)	Σ_2^{P}-Complete (6)	NP^{PP}-Hard (7)
CN-Pr	NP-Complete (8)	NP-Complete (8)	NP^{PP}-Complete (9)
CN-MPE	Polynomial (10)	Polynomial (10)	NP-Complete (11)
CN-MPEe	Σ_2^{P}-Complete (12)	Σ_2^{P}-Complete (12)	Σ_2^{P}-Hard and PP-Hard (13)
CN-MAP	Σ_2^{P}-Complete (12)	Σ_2^{P}-Complete (12)	NP^{PP}-Hard (7)

Table 1: Complexity results; numbers in parenthesis indicate the item that discusses the result.

4. [Park, 2002] reduces MAXSAT problem to BN-MAP in polytrees, a result that can be extended to BIW networks as both problems belong to NP (given the polynomial nature of BN-Pr in BIW networks).
5. [Park and Darwiche, 2004] by reduction from E-MAJSAT.
6. Theorem 8.
7. The complexity of BN-MAP implies it.
8. Theorem 3.
9. [Cozman *et al.*, 2004] by a reduction from E-MAJSAT.
10. Theorem 5.
11. Theorem 6.
12. Theorem 7.
13. BN-MPE and PT-CN-MPEe ensure it.

In the context of Bayesian networks, the difference between MPE and MPEe may seem academic, because any most probable explanation can be found with BN-MPE. However the same is not true for credal networks, where one cannot find a most probable explanation with evidence by simply running a version of MPE — and note that CN-MPE and CN-MPEe do display non-trivial differences. In fact, these differences were our motivation for differentiating MPE from MPEe. The following theorem clarifies the difference between these problems for Bayesian networks.

Theorem 2 BN-MPEe *is PP-Complete.*

Proof Pertinence is obtained from the fact that, after making a PP query to find $P(e)$ (this query is made once), we can decide whether a given instantiation x has $P(x|e) > r$ in linear time, by multiplying the probabilities. To show hardness, we reduce the decision problem #3SAT($\geq 2^{n/2}$) to it, which is PP-Complete [Bailey *et al.*, 2001] and can be stated as: *Given a set of boolean variables $X = \{X_1, \ldots, X_n\}$ and a 3CNF formula $\phi(X)$ with clauses $\{C_1, \ldots, C_m\}$, is $\phi(X)$ satisfied by at least $2^{n/2}$ of the instantiations of X?*

We construct a BN with binary nodes $X_1, \ldots, X_n$ (x_i and $\overline{x_i}$ are the categories) and $C_1, \ldots, C_m$ (c_i and $\overline{c_i}$), where X_i has no parents and uniform prior and C_i has three parents (the variables contained in the clause) with probabilities respecting the truth table for the clause. Furthermore, we insert a dummy binary node Y appearing non-negated in every clause (there will be 2^n instantiations with $\{Y = y\}$ satisfying ϕ; this ensures that the formula is satisfiable). Now we solve the BN-MPEe problem with queries $X_1, \ldots, X_n, Y$ and evidence $C_i = c_i$ for all $1 \leq i \leq m$ (indicating that all C_i are **true**). Then $P(X, Y|c_1, \ldots, c_m)$ is equal to

$$= \frac{P(c_1, \ldots, c_m|X_1, \ldots, X_n, Y)\, P(X_1, \ldots, X_n, Y)}{P(c_1, \ldots, c_m)}$$

$$= \frac{\frac{1}{2^{n+1}} P(c_1, \ldots, c_m|X_1, \ldots, X_n, Y)}{\sum_{X'_1, \ldots, X'_n, Y'} \left[P(c_1, \ldots, c_m|X'_1, \ldots, X'_n, Y')\, \frac{1}{2^{n+1}}\right]}$$

$$= \frac{1}{\#sats} \text{ if } X_1, \ldots, X_n, Y \text{ satisfies } \phi \text{ and } 0 \text{ otherwise,}$$

where $\#sats$ is the total number of satisfying instantiations. So, $\max P(X, Y|c_1, \ldots, c_m) \leq \frac{1}{2^n + 2^{n/2}}$ implies that formula $\phi(X)$ is satisfied by at least $2^{n/2}$ of all X instantiations. Note that if $\max P(X, Y|c_1, \ldots, c_m) = \alpha$ and $\alpha > \frac{1}{2^n + 2^{n/2}}$, then $P(X, Y|c_1, \ldots, c_m) = \alpha$ for all satisfying instantiations, which implies that there are less than $2^{n/2}$ instantiations of X satisfying $\phi(X)$. $\square$

The hardness of PT-CN-Pr was stated by [da Rocha and Cozman, 2002], but the proof there was flawed, and the central argument used zero probabilities and vertex-based description in an essential way. The following proof corrects these difficulties.

Theorem 3 PT-CN-Pr *and* BIW-CN-Pr *are NP-Complete.*

Proof Pertinence of BIW-CN-Pr (which ensures pertinence of PT-CN-Pr) is immediate, as choosing a vertex of each credal set we have a BIW-BN-Pr problem to solve, which is polynomial. To show hardness we reduce the MAX-3-SAT problem to PT-CN-Pr. It can be formulated as follows: *Given a set of boolean variables $\{X_1, \ldots, X_n\}$, a 3CNF formula with clauses $\{C_1, \ldots, C_m\}$ and an integer $0 \leq k \leq m$, is there an assignment for the variables that satisfies at least k clauses?* Initially we remove all clauses that have both x_i and $\overline{x_i}$ and decrement k for each elimination (because those clauses are already satisfied). For each variable X_i we construct two nodes, namely X_i and S_i. The former is binary, has no parents and represents the state of X_i; the probabilities $P(X_i = x_i)$ and $P(X_i = \overline{x_i})$ are in $[\varepsilon, 1-\varepsilon]$ ($0 < \varepsilon < \frac{1}{m+1}$ is a small constant). The latter may assume $m+1$ categories

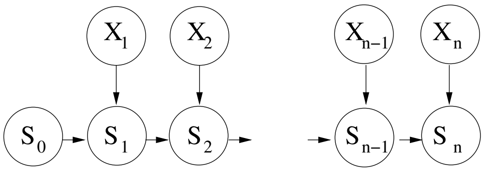

Figure 1: Polytree used in the network of Theorem 3.

(from 0 to m), has S_{i-1} and X_i as parents and is defined by

$$\begin{aligned} P(S_i = c|S_{i-1} = c, x_i) &= 0 \text{ if } x_i \in C_c, \text{ or 1 otherwise} \\ P(S_i = c|S_{i-1} = c, \overline{x_i}) &= 0 \text{ if } \overline{x_i} \in C_c, \text{ or 1 otherwise} \\ P(S_i = c|S_{i-1} \neq c, X_i) &= 0 \text{ for } X_i \in \{x_i, \overline{x_i}\}, \end{aligned}$$

for $c \neq 0$. When $S_i = 0$, we have:

$$\begin{aligned} P(S_i = 0|S_{i-1} = c, X_i) &= 1 - P(S_i = c|S_{i-1} = c, X_i) \\ P(S_i = 0|S_{i-1} = 0, X_i) &= 1 \text{ for } X_i \in \{x_i, \overline{x_i}\}. \end{aligned}$$

The rules above guarantee coherency in probabilities; note that we include a dummy node S_0 with $P(S_0 = c) = \frac{1}{m+1}$ for all c. Now, consider $P(S_n = c)$ for $c \neq 0$. Note that $P(S_n = c) = P(S_0 = c)\prod_i P(S_i = c|S_{i-1} = c)$ because, for $c \neq 0$, every time $\{S_i = c\}$ and $\{S_{i-1} \neq c\}$ appear together we are led to zero. Let $A_{c,i}$ be defined as follows:

$$A_{c,i} = \sum_{X_i \in \{x_i, \overline{x_i}\}} P(S_i = c|S_{i-1} = c, X_i)\, P(X_i)\,.$$

We get $P(S_n = c) = \frac{1}{m+1}\prod_{i\in\{1,\ldots,n\}} A_{c,i}$. Note that $A_{c,i}$ may assume three values: $A_{c,i} = 1$ if X_i does not influence C_c; $A_{c,i} = \varepsilon$ if X_i satisfies C_c and $A_{c,i} = 1 - \varepsilon$ if X_i does not satisfy C_c. We may conclude that if $P(S_n = c) \leq \alpha = (1-\varepsilon)^2\varepsilon$, then some X_i satisfied C_c. Furthermore, we know that if C_c was not satisfied, then $P(S_n = c) = \beta = (1-\varepsilon)^3$. Note that $\alpha < \beta$. To find out how many clauses were not satisfied, we have to sum over all categories of S_n, obtaining $P(S_n = 0) = 1 - \sum_{c\in\{1,\ldots,m\}} P(S_n = c)$ and thus $\overline{P}(S_n = 0)$ minimizes this sum. We define

$$r_h = [(m+1)(1 - \overline{P}(S_n = 0)) - h\beta]/(m-h).$$

and then calculate $r_0, r_1, \ldots$ until $r_h \leq \alpha$ or $h = m-1$. We know that $1 - \overline{P}(S_n = 0)$ is the minimum sum of all $P(S_n = c)$, for $c \neq 0$. This sum is composed by two types of terms: those which are equal to β and those which are less than or equal to α. So what we are verifying with r_h is whether there are more than h terms of the sum that are equal to β or not. The last thing should be noted is that $\varepsilon < \frac{1}{m+1}$ ensures that if just one term of the sum equals to β, then the sum is greater than $m\alpha$, that is, if the sum is composed by $m-1$ terms equal to ε^3 and just one equal to β, it must sum greater than $m\alpha$, because one clause was not satisfied. $\varepsilon < \frac{1}{m+1}$ ensures that everywhere it is necessary (for any h). Thus the inference $\overline{P}(S_n = 0)$ solves `MAX-3-SAT` problem. If $r_h > \alpha$ for all $0 \leq h \leq m-1$, then no clause was satisfied. Otherwise h counts how many clauses were not satisfied. □

Because of the reduction from `MAX-3-SAT`, we can state that there is no polynomial time approximation scheme for `PT-CN-Pr` (or `BIW-CN-Pr`) unless P=NP.

Remark The proof still holds if we substitute all ε by zero; the proof becomes simpler but depends on events of zero probability, which may be inconvenient, as pointed out by [Zaffalon, 2003]. Note also that the proof can be rewritten using inequalities instead of vertices, because all credal sets are in binary nodes (pertinence and hardness still hold).

It is known that `CN-Pr` is solvable by `BN-MAP`, by conducting a *CCM* transform in a credal network [Cozman, 2000a]. The following lemma presents the reverse connection between inferences in Bayesian and credal networks.

Lemma 4 `BN-MAP` *is solvable by* `CN-Pr` *with joint queries without changing the topology of the network among the three types defined.*

Proof Suppose $X_1, \ldots, X_n$ are the MAP variables. Add a binary child X'_i to each X_i with $P(X'_i|X_i) \in [0,1]$ and the constraint $\sum_{X_i} P(x'_i|X_i) = 1$. Now we have

$$\max_{X_1,\ldots,X_n} P(X_1, \ldots, X_n|e) = \overline{P}(x'_1, \ldots, x'_n|e)$$

After evaluating $\overline{P}(x'_1, \ldots, x'_n|e)$, we just have to look at each X'_i node and set X_i according to which of the $P(x'_i|X_i)$ is equal to one (exactly one will be). □

It should be noted that joint queries are not more difficult than single marginal queries: we have that `CN-Pr` with joint queries is still NP^{PP}-Complete and `BIW-CN-Pr` with joint queries is still NP-Complete.

Theorem 5 `PT-CN-MPE` *and* `BIW-CN-MPE` *are solvable in polynomial time.*

Proof It is enough to realize that the inner min of the `BIW-CN-MPE` query $\max_x \min_{P\in K(\mathbb{X})} \prod_i P(x_i|\text{pa}(x_i))$ factorizes, as the network is locally specified (note that x_i's are consistent with the observation e). The optimization becomes $\max_x \prod_i \underline{P}(x_i|\text{pa}(x_i))$, which is equivalent to `BIW-BN-MPE`. □

Theorem 6 `CN-MPE` *is NP-Complete.*

Proof Hardness is immediate, because `BN-MPE` is NP-Complete and can be trivially transformed to a `CN-MPE` (we just have to use credal networks composed by single probability densities). Pertinence is reached because, given an assignment x to the variables, the value of $\underline{P}(x)$ is given by $\prod_i \underline{P}(x_i|\text{pa}(x_i))$. This holds because each credal set $K(x_i|\text{pa}(x_i))$ is locally specified. □

Theorem 7 `PT-CN-MPEe`, `BIW-CN-MPEe`, `PT-CN-MAP` *and* `BIW-CN-MAP` *are all* Σ_2^P*-Complete.*

Proof Pertinence of them is immediate as `BIW-CN-MAP` belongs to Σ_2^P (given the MAP variables, we get a `BIW-CN-Pr` to solve). To see hardness we reduce to `PT-CN-MPEe` a version of $\texttt{QSAT}_2$ that is Σ_2^P-Complete: *Given a set of variables $X_1, \ldots, X_n$, an integer $0 < k \leq n$ and a boolean 3CNF formula $\phi(X)$ over these variables, is it true that, for all instantiations to the first k variables, there is an instantiation of the remaining $n-k$ that satisfies $\phi(X)$?*

Initially we construct a network similar to that of Theorem 3. The variables $X_1, \ldots, X_k$ are defined the same way as there. The variables $X_{k+1}, \ldots, X_n$ become ternary, assuming the categories $x_i, \overline{x_i}, o_i$. Their probabilities are: $P(X_i = o_i) = \varepsilon$ and $P(X_i = x_i)$, $P(X_i = \overline{x_i})$ belong to $[0, 1-\varepsilon]$, where $0 < \varepsilon < \frac{1}{m+2}$ is a small constant. The probabilities of S_i given its parents are the same as there, except when $i > k$. In these cases we have $P(S_i = c | S_{i-1} = c, o_i) = 0$, for $c \neq 0$ (the case when $c = 0$ remains the same, that is, equals to 1).

Furthermore, we add a dummy node Q with S_n as parent and $P(q|S_n = c) = 0$ for $c \neq 0$ and 1 otherwise. We will solve the PT-CN-MPEe problem $\max \underline{P}(X, S|q)$, where $S = \{S_0, \ldots, S_n\}$ and $X = \{X_1, \ldots, X_n\}$. Let X^- be $\{X_1, \ldots, X_k\}$ and X^+ be $\{X_{k+1}, \ldots, X_n\}$; then

$$P\big(X^-, X^+, S|q\big) = \frac{P(q|S)\,P(S|X)\,P(X^-)\,P(X^+)}{P(q)}.$$

First note that the given q forces $S_n = 0$ to get a non-zero probability. Furthermore, for all instantiations x^-, x^+, s of the variables, there is another instantiation with $s' = \{\forall i S_i = 0\}$ that has its probability greater than the former, that is, $\underline{P}(x^-, x^+, s'|q) > \underline{P}(x^-, x^+, s|q)$. This holds because the S_i nodes are not credal and the conditional probabilities $P(S_i = 0|S_{i-1} = 0)$ are equal to 1, for all i. Thus we know that the solution of the MPEe problem will be attained in an instantiation where all S_i are set to 0. Besides that, we have that choosing the category o_i for all X^+ variables lead us to greater probabilities than if we choose any other. That is,

$$\max_{X^-, X^+, S} \underline{P}\big(X^-, X^+, S|q\big) = \max_{X^-} \underline{P}\big(X^-, o, s|q\big)\,,$$

where o denotes $\{X_i = o_i \text{ for } i \in \{k+1, \ldots, n\}\}$ and s denotes $\{\forall i\ S_i = 0\}$. If we choose a category different from o_i whenever possible, the maximum probability would not reach the same value (in fact it will be zero, because $P(X_i \neq o_i) \in [0, 1-\varepsilon]$, and thus it may assume value zero).

$$P\big(X^-, o, s|q\big) = \frac{\frac{1}{m+1}\varepsilon^{n-k} P(X^-)}{1 - \sum_{c\neq 0} P(S_n = c)}.$$

When finding $\underline{P}(X^-, o, s|q)$, the numerator $P(X^-)$ will become ε^k, because any solution that does not make $P(X^-) = \varepsilon^k$ will not be a minimum for $\underline{P}(X^-, o, s|q)$ (remember that $\varepsilon < \frac{1}{m+2}$). This holds because just one X^- variable using the extreme point $(1-\varepsilon)$ instead of ε is enough to make $\underline{P}(X^-, o, s|q)$ too large:

$$\frac{\frac{1}{m+1}\varepsilon^{n-k}\varepsilon^{k-1}(1-\varepsilon)}{1-R} > \frac{\frac{1}{m+1}\varepsilon^{n-k}\varepsilon^{k}}{1-R'}$$

where R and R' are any possible values for the sum $\sum_{c\neq 0} P(S_n = c)$ that appears in the denominator (note that these sums are restricted in $[0, \frac{m}{m+1}]$ by the probabilities of the network).

Summarizing, all S variables in the solution of MPEe are set to zero, all X^+ variables are set to o, and the instantiation chosen for X^- makes $P(X^-) = \varepsilon^k$, which implies that if x_i belongs to the instantiation of X^-, then $P(x_i) = \varepsilon$ and $P(\overline{x_i}) = 1 - \varepsilon$ (the opposite case is analogous). This means that the vertices of the credal sets of the X^- nodes are completely fixed by the instantiation chosen. Thus the only credal sets that can float in the denominator are the X^+ variables (S_i variables are not credal and X^- are already fixed as indicated). So, processing the MPEe we have

$$\underline{P}\big(X^-, o, s|q\big) = \frac{\frac{1}{m+1}\varepsilon^n}{1 - \min\left(\sum_{c\neq 0} P(S_n = c)\right)},$$

where $\min \sum_{c\neq 0} P(S_n = c)$ is evaluated over all possible vertices of the X^+ credal sets (we know that $P(X_i = o_i)$ is set to ε, but the probabilities $P(X_i = x_i)$ and $P(X_i = \overline{x_i})$ may vary between 0 and $1 - \varepsilon$).

The formula $\phi(X)$ will be satisfied by X^- and X^+ if the sum $\sum_{c\neq 0} P(S_n = c)$ is less than $\delta_1 = \frac{m}{m+1}(1-\varepsilon)^2\varepsilon(1-\varepsilon)^{n-k}$ (this is the maximum value that a satisfied $\phi(X)$ may assume). All unsatisfied formulas lead to greater values. In fact the smallest value that a unsatisfied formula implies in the sum is greater than $\delta_2 = \frac{m}{m+1}(1-\varepsilon)^{n-k+3}$. Note that $\delta_2 > \delta_1$.

If we query the PT-CN-MPEe problem $\max \underline{P}(X, S|q)$ with $r = \frac{1}{m+1}\frac{\varepsilon^n}{1-\delta_1}$ and get a negative answer, then for all instantiations to X^-, there is an instantiation to X^+ that satisfy $\phi(X)$ (that is, the sum is bounded by δ_1). If the answer is positive, then there does exist an instantiation to X^- where no instantiation to X^+ can make $\phi(X)$ satisfied. □

Theorem 8 PT-BN-MmAP *and* BIW-BN-MmAP *are* Σ_2^P-*Complete.*

Proof Pertinence of BIW-BN-MmAP (which ensures pertinence of PT-BN-MmAP) is trivial. Given a instantiation for the MAP variables, we need to solve a minimization over the Y variables, which is NP-Complete (see it as a BIW-CN-Pr using Lemma 4). Hardness of PT-BN-MmAP (which ensures hardness of BIW-BN-MmAP) is reached by a reduction from another version of QSAT$_2$ that is Σ_2^P-Complete: *Given a set of variables $X_1, \ldots, X_n$, an integer $0 < k \leq n$ and a boolean 3DNF formula $\phi(X)$ over these variables, is there an instantiation to the first k variables such that, for all instantiations of the remaining $n - k$, $\phi(X)$ is satisfied?*

We construct again a network following the ideas of Theorem 3. It includes a binary node to each X_i, without parents and with uniform prior. There are n nodes S_i with parents S_{i-1} and X_i. They have $m + 1$ categories and are defined as follows (for $c \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, n\}$):

$$P(S_i = c|S_{i-1} = c, x_i) = \begin{cases} 1 & \text{if } x_i \in \text{clause } c \\ 0 & \text{if } \overline{x_i} \in \text{clause } c \\ \frac{1}{2} & \text{otherwise} \end{cases}$$
$$P(S_i = c|S_{i-1} \neq c, x_i) = 0.$$

The conditional probabilities of S_i given $\overline{x_i}$ are defined analogously. The probabilities of $P(S_i = 0|S_{i-1}, X_i)$ ensure that they sum exactly 1, as done in Theorem 3. Furthermore, $P(S_0)$ has uniform prior and we insert an additional binary node Q with S_n as parent, having $P(q|S_n = c) = 1$ if $c \neq 0$ and 0 otherwise.

So, we have that $P(X|q) = \frac{\frac{1}{2^n}P(q|X)}{P(q)}$ is equal to

$$= \frac{\frac{1}{(m+1)2^n}\sum_{c\neq 0}\prod_i A_{c,i}}{\frac{m}{(m+1)}\left(\frac{1}{2}\right)^n} = \frac{\sum_{c\neq 0}\prod_i A_{c,i}}{m}$$

where $A_{c,i} = \sum_{X_i} P(S_i = c|S_{i-1} = c, X_i)\, P(X_i)$ for $i \in \{1, \ldots, n\}$. $A_{c,i}$ assumes value zero only when the variable it represents denies the clause c. The numerator of $P(X|q)$ sums how many clauses are satisfied by the instantiation of X variables. Let $X^- = \{X_1, \ldots, X_k\}$ and $X^+ = \{X_{k+1}, \ldots, X_n\}$. We have $\min_{X^+} P(X^-, X^+|q) = 0$ if, given the instantiation for the X^- variables, there is an assignment to X^+ variables that can deny $\phi(X)$. This way, questioning if PT-BN-MmAP problem with MAP variables X^- and evidence q (which maximizes $\min_{X^+} P(X|q)$) has non-zero answer is enough to solve the QSAT_2 problem. Given this X^- instantiation, all X^+ will satisfy $\phi(X)$. □

5 Conclusion

We can summarize the contributions of this paper as follows. Concerning Bayesian networks, we have first introduced a more general definition of bounded induced-width networks (demonstrating that many problems where induced-width actually grows with the network remain polynomial), and we have shown the difference between the BN-MPE and the BN-MPEe problems. More importantly, we have introduced the MmAP problem and presented its complexity. A possible improvement of our results would be to present a completeness result for multiply-connected networks.

Our most significant results pertain to credal networks, with direct implications to models that handle interval and set probabilities, belief functions, possibility measures, qualitative probabilities, and families of probabilistic logic. We have clarified the so far unexplored complexity of CN-MPE, CN-MPEe, and CN-MAP. The polynomial character of CN-MPE in some cases is rather surprising.

There are several interesting problems still to be explored. For example, binary networks (Bayesian and credal) could display lower complexity than their non-binary counterparts in problems such as PT-BN-MAP, BIW-BN-MAP and BIW-CN-Pr. These problems belong to NP and are clearly related, but are there polynomial time algorithms to solve them? We conjecture there are not, even as we note that there is a polynomial algorithm to solve CN-Pr in binary polytrees [Fagiuoli and Zaffalon, 1998].

Acknowledgments

We thank Aritanan Gruber for indicating relevant complexity results, Carlos Guestrin for pointing out a flaw in [da Rocha and Cozman, 2002], Thomas Lukasiewicz for initial thoughts on the complexity of credal networks, and a reviewer for suggesting a short proof for Theorem 5. This research received support from HP Brazil R&D; the second author was partially supported by CNPq, Brazil.

References

[Bailey *et al.*, 2001] D. D. Bailey, V. Dalmau, and P. G. Kolaitis. Phase transitions of PP-complete satisfiability problems. In *Int. Joint Conf. on Artificial Intelligence*, pp. 183–192, Seattle, Washington, 2001.

[Cano *et al.*, 1993] J. Cano, M. Delgado, and S. Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *Int. Journal of Approximate Reasoning*, 8:253–280, 1993.

[Cozman *et al.*, 2004] F. G. Cozman, C. P. de Campos, J. S. Ide, and J. C. F. da Rocha. Propositional and relational Bayesian networks associated with imprecise and qualitative probabilistic assessments. In *XX Conf. on Uncertainty in Artificial Intelligence*, pp. 104–111, Banff, 2004. AUAI Press.

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

[Cozman, 2000b] F. G. Cozman. Separation properties of sets of probabilities. In *XVI Conf. on Uncertainty in Artificial Intelligence*, pp. 107–115, San Francisco, 2000. Morgan Kaufmann.

[da Rocha and Cozman, 2002] J. C. F. da Rocha and F. G. Cozman. Inference with separately specified sets of probabilities in credal networks. In *XVIII Conf. on Uncertainty in Artificial Intelligence*, pp. 430–437, San Francisco, 2002. Morgan Kaufmann.

[Dechter, 1996] R. Dechter. Bucket elimination: A unifying framework for probabilistic inference. In *XII Conf. on Uncertainty in Artificial Intelligence*, pp. 211–219, San Francisco, 1996. Morgan Kaufmann.

[Eyal, 2001] A. Eyal. *Dividing and Conquering Logic*. PhD thesis, Stanford University, 2001.

[Kakade *et al.*, 2001] S. Kakade, and L. Ortiz. Graphical economics. *COLT*, 2004.

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

[Levi, 1980] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts, 1980.

[Litmman *et al.*, 2001] M. Litmman, S. M. Majercik, and T. Pitassi. Stochastic boolean satisfiability. *Journal of Automated Reasoning*, 27(3):251–296, 2001.

[Papadimitriou, 1994] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, San Diego, 1994.

[Park and Darwiche, 2004] J. D. Park and A. Darwiche. Complexity results and approximation strategies for MAP explanations. *Journal of Artificial Intelligence Research*, 21:101–133, 2004.

[Park, 2002] J. D. Park. MAP complexity results and approximation methods. In *XVIII Conf. in Uncertainty in Artificial Intelligence*, University of Alberta, Edmonton, Alberta, Canada, 2002. Morgan Kaufmann.

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

[Roth, 1996] D. Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1-2):273–302, 1996.

[Shimony, 1994] S. E. Shimony. Finding MAPs for belief networks is NP-hard. *Artificial Intelligence*, 68(2):399–410, 1994.

[Walley, 1991] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[Walley, 1996] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

[Zaffalon, 2003] M. Zaffalon and E. Fagiuoli. Tree-based credal networks for classification. *Reliable Computing*, 9(6):487–509, 2003.