D-separation for the strong extension and the main natural extension of a credal network

Andrey G. Bronevich[1,2] **Igor N. Rozenberg**[2]

[1]JSC "Research and Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport", Russia
[2]Russian University of Transport, Russia

Abstract

In the paper, we consider two possible extensions of a credal network: the strong extension and the main natural extension. We prove that for both extensions the condition of the d-separation is preserved. The proof is based on some properties of conditional independence in such credal networks.

Keywords. Bayesian networks, credal networks, strong and natural extensions, d-separation

1. Introduction

Bayesian networks [8, 12, 13] proved their efficiency for representing knowledge in many applications of artificial intelligence. The description of any Bayesian network consists of two levels: the first level is described by an acyclic directed graph whose nodes are labeled by random variables with the arcs showing the dependences between them; the second level is described by conditional probability distributions within each node subject to random variables located in parent nodes. It is possible to generalize Bayesian networks, if we describe random variables by conditional credal sets. Such mathematical objects are called credal networks [2, 5, 6, 9]. While analyzing Bayesian networks or credal networks, there are two main problems. The first problem is how to describe the joint probability distribution of random variables represented by a network. The second problem is how to produce the approximate reasoning using such representations of our knowledge. The first problem is solved simply for Bayesian networks: the joint probability distribution is represented as the product of conditional probabilities from the nodes of a network. This problem is trickier for credal networks, and it is usually solved by mimicking independence or irrelevance relations of Bayesian networks. It is well known that in Bayesian networks it is possible to establish the conditional independence relation analyzing only their structural characteristics. It was shown [13] that two sets B and C of random variables in a Bayesian network are conditionally independent subject to set of random variables A if every path between B and C contains some special configurations. This relation on the sets of nodes in a Bayesian network is called d-separation.

There are two possible ways to preserve the d-separation relation when constructing a description of the joint probability distribution of random variables in credal networks. The first way is called the strong extension [2, 5] of a credal network. This construction is defined as the convex hull of all joint probability distributions of Bayesian networks compatible with a given credal network. This extension generalizes the strong independence of imprecisely described random variables, and we can derive it if we consider credal networks consisting of several connected components. The second way consists in finding the largest credal set compatible with a credal network and preserving the d-separation relation of conditional independence for random variables. In terms of imprecise probabilities, this credal set is called the natural extension [2, 5] of a credal network under the d-separation relation.

Although, these extensions look different, we should not consider them separately. It was shown in [5] that the d-separation relation is preserved for the strong extension of a credal network. From this point of view, the strong extension of a credal network gives us the smallest credal set under the d-separation relation, and every credal set between the strong extension and the natural extension under the d-separation relation of a credal network has this property of d-separation too. In our opinion, the natural extension of a credal network under d-separation has not yet been sufficiently studied in the literature comparing with the strong extension. There are several unsolved problems:

1. What is the smallest set of constraints linked with the independence (irrelevance) relations, which leads to the natural extension of a credal network under d-separation?

2. Is there an analytical expression for the natural extension of a credal network under d-separation?

3. What are the differences in inferences based on the

strong and the natural extensions of a credal network?

In the paper, we contribute to the solutions of the first two problems formulated above. For this goal, we introduce first a σ-natural extension of a credal network. Since every credal network is described by a directed graph without cycles, any such a graph defines a strict partial order ρ on the set of its nodes. Let σ be a linear strict order compatible with ρ, then a σ-natural extension does not preserve all possible irrelevances between random variables, it preserves only those of them, whose positions in the linear order are the same with the direction of irrelevance, i.e. if random variable x_i is irrelevant to random variable x_j in a credal network, then they preserve this property in its σ-natural extension if $(x_j, x_i) \in \sigma$. If we take the intersection of all σ-natural extensions of a credal network, then we get its main natural extension. The main result of this paper is that the main natural extension preserves d-separation of underlying random variables, i.e. the main natural extension coincides with the natural extension of a credal network under the d-separation relation. Thus, we also found the analytical expression for the natural extension of a credal network under the d-separation relation based on the σ-natural extensions. In the paper, we also propose the general scheme for proving the d-separation property that works for the strong extension and the main natural extension of a credal network.

The paper has the following structure. In Sections 2 and 3, we recall well known notions concerning Bayesian networks and concepts of independence of imprecisely described random variables. In Section 4, we introduce the notion of a credal network and consider its possible extensions, and finally in Section 5, we prove the property of d-separation for the strong extension and the main natural extensions of any credal network. The paper ends with concluding remarks.

2. Bayesian networks

A *Bayesian network* [8, 12] is a graphical representation for describing multivariate probability distributions. It is an acyclic graph that reflects the structure of the joint probability distribution constructed in such a way as to be optimal for solving a given problem. The nodes of this graph are labeled by random variables, and the arcs between the nodes describe their dependence. In a Bayesian network, every random variable x_i is described by the conditional distribution of x_i subject to known values of its parents. For example, consider how to describe the joint probability distributions for simple Bayesian networks depicted in Fig. 1.

For these Bayesian networks, the joint probability distributions are described as follows:

a) $p(x_1, x_2, w) = p(w|x_1, x_2)p(x_1, x_2) = p(w|x_1, x_2)p(x_2|x_1)p(x_1)$;

b) $p(x_1, x_2, w) = p(w|x_1, x_2)p(x_1, x_2) = p(w|x_1, x_2)p(x_1|x_2)p(x_2)$;

c) $p(x_1, x_2, w) = p(w|x_1, x_2)p(x_1, x_2) = p(w|x_1, x_2)p(x_1)p(x_2)$.

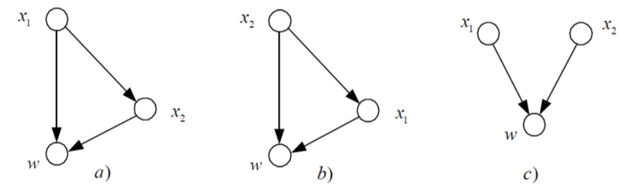

Figure 1. *Possible configurations of a Bayesian network*

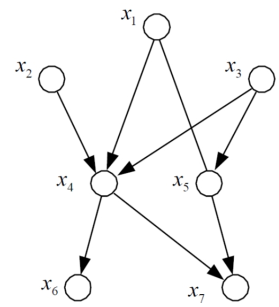

Figure 2. *An example of a Bayesian network with seven nodes*

Note that for cases a) and b) we do not impose any restriction on the joint probability distribution of random variables, however, in a view of an expert who extracts the knowledge, in the case a), the joint distribution of x_1 and x_2 is derived as $p(x_1, x_2) = p(x_2|x_1)p(x_1)$, which means that the value of x_1 affects more to a value of w than x_2. Similarly, in the case b), the random variable x_2 is more influential than x_1 in the analysis of w. In the case c), random variables x_1 and x_2 are independent, since $p(x_1, x_2) = p(x_1)p(x_2)$.

For an arbitrary Bayesian network with a set of nodes $X = \{x_1, ..., x_K\}$, the joint probability distribution of random variables is defined by

$$p(x_1, ..., x_K) = \prod_{k=1}^{K} p(x_k|pa_k),$$

where pa_k are the set of all parent nodes for x_k.

As an example, let us consider the construction of joint probability distribution for the Bayesian network depicted in Fig. 2. In this case,

$$\begin{aligned} p(x_1, ..., x_7) = \; & p(x_7|x_4, x_5)p(x_6|x_4)p(x_5|x_1, x_3) \\ & p(x_4|x_1, x_2, x_3)p(x_1)p(x_2)p(x_3). \end{aligned}$$

Let us recall some definitions concerning the conditional independence of random variables. The set of random variables B does not depend on the set of random variables C provided that the set A of random variables is observed (where A,B,C are non-empty pairwise disjoint sets) if

$$p(B,C|\tilde{A}) = p(B|\tilde{A})p(C|\tilde{A}). \quad (1)$$

for all possible instantiations $\tilde{A}$ of the composite random variable A with the positive probability $p(\tilde{A})$. Since

$$p(B,C|A) = p(B|A,C)p(C|A),$$

$$p(B,C|A) = p(C|B,A)p(B|A).$$

the condition (1) also implies that $p(B|C,A) = p(B|A)$ and $p(C|B,A) = p(C|A)$.

The graphical representation of multivariate probability distributions by Bayesian networks implies the conditional independence of random variables that depends only on the structural characteristics of these networks. In a Bayesian network, sets of nodes B and C are *d-separated* by the set of nodes A if any path from the set of nodes B to the set of nodes C contains one of the following configurations:

1) a *chain* $b \to a \to c$ ($b \leftarrow a \leftarrow c$), where $a \in A$;
2) a *fork* $b \leftarrow a \to c$, where $a \in A$;
3) a *collider* $b \to u \leftarrow c$, where $u \notin A$ and u has no descendants in A.

It is possible to prove [13] that if the sets of nodes B and C are d-separated by A, then composite random variables B and C are conditionally independent given A.

3. Independence of imprecisely described random variables

In this section, a random variable will be described not by one probability distribution, but by some family of probability distributions presented by a credal set. Any such credal set can be defined by a set of densities $\mathbf{p}(x_1,...,x_n)$. Thus, if $\tilde{U}$ is a finite set of all possible values for a random vector $(x_1,...,x_n)$, then the following conditions hold for every $p(x_1,...,x_n) \in \mathbf{p}(x_1,...,x_n)$:

1) $p(\tilde{\mathbf{u}}) \geqslant 0$ for any $\tilde{\mathbf{u}} \in \tilde{U}$;
2) $\sum_{\tilde{\mathbf{u}}\in\tilde{U}} p(\tilde{\mathbf{u}}) = 1$.

In addition, since the set $\mathbf{p}(x_1,...,x_n)$ describes a credal set, it is convex and closed. The convexity implies that

3) $ap_1(x_1,...,x_n)+(1-a)p_2(x_1,...,x_n) \in \mathbf{p}(x_1,...,x_n)$ for all possible $a \in [0,1]$ and $p_i(x_1,...,x_n) \in \mathbf{p}(x_1,...,x_n)$, $i = 1,2$.

We will assume that all credal sets under consideration have only a finite number of extreme points, i.e. the following condition holds:

4) there are $p_i(x_1,...,x_n) \in \mathbf{p}(x_1,...,x_n)$, $i = 1,...,k$, such that every $p(x_1,...,x_n) \in \mathbf{p}(x_1,...,x_n)$ can be represented as $p(x_1,...,x_n) = \sum_{i=1}^{k} a_i p_i(x_1,...,x_n)$, where $a_i \geqslant 0$, $i = 1,...,k$, and $\sum_{i=1}^{k} a_i = 1$.

Conditional probability distributions of random variables are described similarly. Assume, for example, that $\mathbf{x}$ and $\mathbf{y}$ are random vectors such that they have no common variables. Then we can consider an imprecise description $\mathbf{p}(\mathbf{x}|\mathbf{y})$ of random vector $\mathbf{x}$ given $\mathbf{y}$. Assume that $\tilde{U}(\mathbf{y})$ is the set of all possible instantiations of $\mathbf{y}$, then $\mathbf{p}(\mathbf{x}|\tilde{\mathbf{y}})$ for every fixed $\tilde{\mathbf{y}} \in \tilde{U}(\mathbf{y})$ is a credal set satisfying conditions 1) - 4), i.e. we assume that the true probability distribution $p(\mathbf{x}|\tilde{\mathbf{y}})$ is not exactly known and in the set $\mathbf{p}(\mathbf{x}|\tilde{\mathbf{y}})$.

Let us consider how to define the description of conditional random variables if we know the description $\mathbf{p}(\mathbf{x},\mathbf{y})$ of the joint distribution of random variables $\mathbf{x}$ and $\mathbf{y}$. In this case, $\mathbf{p}(\mathbf{x}|\tilde{\mathbf{y}}) = \{p(\mathbf{x}|\tilde{\mathbf{y}})|p(\mathbf{x},\mathbf{y}) \in \mathbf{p}(\mathbf{x},\mathbf{y}), p(\tilde{\mathbf{y}}) > 0\}$, where $p(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$ in $p(\mathbf{x},\mathbf{y})$. The last formula is known in the literature as the *generalized Bayes rule*.

Remark 3.1. To avoid some further complicated constructions, we will assume that all results formulated below are given under the positivity assumption for lower bounds of probabilities. It implies that the generalized Bayes rule is produced in the case, when $p(\tilde{\mathbf{y}}) > 0$ for all $p(\mathbf{x},\mathbf{y}) \in \mathbf{p}(\mathbf{x},\mathbf{y})$. In this case, $\mathbf{p}(\mathbf{x}|\tilde{\mathbf{y}})$ is a convex credal set, whose extreme points are in the set $\{p^{(i)}(\mathbf{x}|\tilde{\mathbf{y}})\}_{i=1}^N$, where $\{p^{(i)}(\mathbf{x},\mathbf{y})\}_{i=1}^N$ are extreme points of $\mathbf{p}(\mathbf{x},\mathbf{y})$.

There are several ways [2–4, 6, 7, 14] to define independence in the theory of imprecise probabilities, however, we consider the following and most popular one.

Epistemic irrelevance: assume the joint distribution of random variables $\mathbf{x}$ and $\mathbf{y}$ is described by $\mathbf{p}(\mathbf{x},\mathbf{y})$, then the random variable $\mathbf{x}$ is *epistemically irrelevant* to $\mathbf{y}$ if for every instantiation $\tilde{\mathbf{y}} \in \tilde{U}(\mathbf{y})$, we have $\mathbf{p}(\mathbf{x}|\tilde{\mathbf{y}}) = \mathbf{p}(\mathbf{x})$, where $\mathbf{p}(\mathbf{x}) = \{p(\mathbf{x})|p(\mathbf{x},\mathbf{y}) \in \mathbf{p}(\mathbf{x},\mathbf{y})\}$.

Epistemic independence: random variables $\mathbf{x}$ and $\mathbf{y}$ are called *epistemically independent* if $\mathbf{x}$ is epistemically irrelevant to $\mathbf{y}$ and $\mathbf{y}$ is epistemically irrelevant to $\mathbf{x}$.

Consider the ways of defining $\mathbf{p}(\mathbf{x},\mathbf{y})$ through $\mathbf{p}(\mathbf{x})$ and $\mathbf{p}(\mathbf{y})$ if we know that random variable $\mathbf{x}$ is irrelevant to $\mathbf{y}$ or $\mathbf{x}$ and $\mathbf{y}$ are independent random variables.

Random variables $\mathbf{x}$ and $\mathbf{y}$ are called *strongly independent* iff their joint distribution is described by

$$\mathbf{p}(\mathbf{x},\mathbf{y}) = \mathbf{p}(\mathbf{x})\times_S\mathbf{p}(\mathbf{y}) = ch\{p(\mathbf{x})p(\mathbf{y})|p(\mathbf{x}) \in \mathbf{p}(\mathbf{x}), p(\mathbf{y}) \in \mathbf{p}(\mathbf{y})\},$$

where chS denotes the convex hull of the set S.

Assume that random variable $\mathbf{x}$ is irrelevant to $\mathbf{y}$, then the largest credal set $\mathbf{p}(\mathbf{x},\mathbf{y})$ preserving this property is

$$\mathbf{p}(\mathbf{x},\mathbf{y}) = \left\{ p(\mathbf{x},\mathbf{y}) \middle| \begin{array}{l} \forall\tilde{\mathbf{y}} \in \tilde{U}(\mathbf{y}) : p(\mathbf{x}|\tilde{\mathbf{y}}) \in \mathbf{p}(\mathbf{x}), \\ p(\mathbf{y}) \in \mathbf{p}(\mathbf{y}) \end{array} \right\},$$

and denoted by $\mathbf{p}(\mathbf{x},\mathbf{y}) = \mathbf{p}(\mathbf{x})\times_I\mathbf{p}(\mathbf{y})$.

Assume that random variables $\mathbf{x}$ and $\mathbf{y}$ are independent, then the largest credal set $\mathbf{p}(\mathbf{x},\mathbf{y})$ preserving this property is

$$\mathbf{p}(\mathbf{x},\mathbf{y}) = ch \left\{ p(\mathbf{x},\mathbf{y}) \middle| \begin{array}{l} \forall\tilde{\mathbf{y}} \in \tilde{U}(\mathbf{y}) : p(\mathbf{x}|\tilde{\mathbf{y}}) \in \mathbf{p}(\mathbf{x}), \\ p(\mathbf{y}) \in \mathbf{p}(\mathbf{y}), \\ \forall\tilde{\mathbf{x}} \in \tilde{U}(\mathbf{x}) : p(\mathbf{y}|\tilde{\mathbf{x}}) \in \mathbf{p}(\mathbf{y}), \\ p(\mathbf{x}) \in \mathbf{p}(\mathbf{x}) \end{array} \right\}.$$

This set is denoted by $\mathbf{p}(\mathbf{x},\mathbf{y}) = \mathbf{p}(\mathbf{x}) \times \mathbf{p}(\mathbf{y})$, and, obviously,

$$\mathbf{p}(\mathbf{x}) \times \mathbf{p}(\mathbf{y}) = (\mathbf{p}(\mathbf{x}) \times_I \mathbf{p}(\mathbf{y})) \cap (\mathbf{p}(\mathbf{y}) \times_I \mathbf{p}(\mathbf{x})).$$

The following proposition (see Proposition 4 from [3]) shows the connections between introduced credal sets and introduced notions of irrelevance and independence of imprecisely described random variables.

Proposition 3.1. *Assume that the joint distribution of random variables* $\mathbf{x}$ *and* $\mathbf{y}$ *is described by* $\mathbf{p}(\mathbf{x},\mathbf{y})$*, then*

a) $\mathbf{x}$ *is irrelevant to* $\mathbf{y}$ *iff* $\mathbf{p}(\mathbf{x}) \times_S \mathbf{p}(\mathbf{y}) \subseteq \mathbf{p}(\mathbf{x},\mathbf{y}) \subseteq \mathbf{p}(\mathbf{x}) \times_I \mathbf{p}(\mathbf{y})$*;*

b) $\mathbf{x}$ *and* $\mathbf{y}$ *are independent iff* $\mathbf{p}(\mathbf{x}) \times_S \mathbf{p}(\mathbf{y}) \subseteq \mathbf{p}(\mathbf{x},\mathbf{y}) \subseteq \mathbf{p}(\mathbf{x}) \times \mathbf{p}(\mathbf{y})$*.*

4. Credal Networks and Their Extensions

Formally, a generalized Bayesian network based on imprecise probabilities [2, 5, 9], also called a *credal network*, is a directed acyclic graph whose vertices are labeled with random variables and arcs with conditional credal sets. Thus, a credal network is an ordinary Bayesian network if each credal set used in its description is a singleton. Now, to comprehend such a generalized Bayesian network, it remains to understand which family of probability measures describes the joint distribution of random variables of this credal network. To do this, we return to the case of an ordinary Bayesian network. Consider, for example, the Bayesian network in Fig.2. Then one among possible formulas describing the joint distribution is

$$\begin{aligned} p(x_1,...,x_7) = \quad & p(x_5|x_4)p(x_4|x_1,x_2)p(x_6|x_2,x_3) \\ & p(x_7|x_3)p(x_1)p(x_2)p(x_3). \end{aligned} \tag{2}$$

If we read the last formula from the end, we get that this $p(x_3)$ is the distribution of random variable x_3, $p(x_2)p(x_3)$ is the joint distribution of random variables x_2 and x_3 etc., $p(x_6|x_2,x_3)p(x_7|x_3)p(x_1)p(x_2)p(x_3)$ is the joint distribution of random variables x_6, x_7, x_1, x_2, x_3. Writing the joint distribution can be done in another way, for example,

$$\begin{aligned} p(x_1,...,x_7) = \quad & p(x_6|x_2,x_3)p(x_7|x_3)p(x_3)p(x_5|x_4) \\ & p(x_4|x_1,x_2)p(x_1)p(x_2). \end{aligned} \tag{3}$$

Note that formula (2) implies that random variable x_2 does not depend on x_3; x_1 does not depend on x_2 and x_3; x_7 does depend on x_1 and x_2 given x_3. Analogously, formula (3) implies that random variable x_3 does not depend on x_1, x_2, x_5, x_4.

Let us consider how to list all possible ways of writing the joint density of random variables for a Bayesian network. Recall that every Bayesian network determines the acyclic graph or the relation on the set of random variables $\{x_1,...,x_n\}$. If we take the transitive closure of this relation, then we get the strict order ρ. Note that this order is not linear, when, in particular, some random variables are independent. Assume that a permutation $\sigma : \{1,...,n\} \to \{1,...,n\}$ is consistent with the strict order ρ, if the strict linear order $\rho_\sigma = \{(\sigma(i),\sigma(j))|i < j\}$ is the superset of ρ. Then the joint density $p(x_1,...,x_n)$ can be derived using the recurrence equation:

$$\begin{aligned} & p_{i+1}(x_{\sigma(i+1)}, x_{\sigma(i)}, ..., x_{\sigma(1)}) = \\ & \qquad p(x_{\sigma(i+1)}|pa_{\sigma(i+1)})p(x_{\sigma(i)}, ..., x_{\sigma(1)}), \end{aligned}$$

where $i = 2, ..., n-1$. Next, consider possible ways to define the description of the joint distribution $\mathbf{p}(x_1,...,x_n)$, if we know $\mathbf{p}(x_1|A)$ and $\mathbf{p}(x_2,....,x_n)$, where $A \subseteq \{x_2,...,x_n\}$. We will describe the so called strong extension and natural extension of a Bayesian network [2, 5, 6]. The *strong extension* is defined as

$$\mathbf{p}(x_1,...,x_n) = \\ ch\left\{ p(x_1|A)p(x_2,...,x_n) \middle| \begin{matrix} p(x_1|A) \in \mathbf{p}(x_1|A), \\ p(x_2,...,x_n) \in \mathbf{p}(x_2,...,x_n) \end{matrix} \right\}.$$

where $ch\,S$ means that the convex hull of the set S is taken. We will denote this operation by $\mathbf{p}(x_1,...,x_n) = \mathbf{p}(x_1|A) \times_S \mathbf{p}(x_2,...,x_n)$. In particular, if $\mathbf{p}(x_1|A) = \mathbf{p}(x_1)$, i.e. $A = \emptyset$ and the random variable x_1 is irrelevant to $x_2,...,x_n$, then $\mathbf{p}(x_1,...,x_n) = \mathbf{p}(x_1) \times_S \mathbf{p}(x_2,...,x_n)$ is the product of credal sets $\mathbf{p}(x_1)$ and $\mathbf{p}(x_2,...,x_n)$, showing the strong independence between x_1 and $x_2,...,x_n$.

Lemma 4.1. *Let* $\mathbf{p}'(x_1|A)$ *be the set of extreme points of* $\mathbf{p}(x_1|A)$*, where* $A \subseteq \{x_2,...,x_n\}$*, and let* $\mathbf{p}'(x_2,...,x_n)$ *be the set of extreme points of* $\mathbf{p}(x_2,...,x_n)$*, then*

$$\mathbf{p}(x_1|A) \times_S \mathbf{p}(x_2,...,x_n) = \\ ch\left\{ p(x_1|A)p(x_2,...,x_n) \middle| \begin{matrix} p(x_1|A) \in \mathbf{p}'(x_1|A), \\ p(x_2,...,x_n) \in \mathbf{p}'(x_2,...,x_n) \end{matrix} \right\}.$$

Definition 4.1. Assume that G is a credal network defined on the set of random variables $\{x_1,...,x_n\}$, and σ is a permutation consistent with G. Then the joint distribution $\mathbf{p}_S(x_1,...,x_n)$, defined recursively by the rule:

$$\begin{aligned} & \mathbf{p}^\sigma_{i+1}(x_{\sigma(i+1)}, x_{\sigma(i)}, ..., x_{\sigma(1)}) = \\ & \qquad \mathbf{p}(x_{\sigma(i+1)}|pa_{\sigma(i+1)}) \times_S \mathbf{p}(x_{\sigma(i)}, ..., x_{\sigma(1)}), \end{aligned}$$

where $i = 2, ..., n-1$, for $i = n-1$, is called the *strong extension* of G.

Theorem 4.1. *The strong extension* $\mathbf{p}_S(x_1,...,x_n)$ *of every credal network* G *defined on the set* $\{x_1,...,x_n\}$ *of random variables does not depend on a possible permutation* σ *and can be computed as*

$$\mathbf{p}_S(x_1,...,x_n) = \\ ch\left\{ \prod_{k=1}^{n} p(x_k|pa_k) \,|\, p(x_k|pa_k) \in \mathbf{p}(x_k|pa_k) \right\}.$$

Remark 4.1. Note that, by Lemma 4.1, the calculation of the joint distribution obtained by the strong extension can be produced using extreme points. This implies Theorem 4.1 and the fact that the calculation of the strong extension of a credal network G from Theorem 4.1 can also be performed using extreme points of credal sets $\mathbf{p}(x_k|pa_k)$. The mentioned results are straightforward consequences of Proposition 9.1 formulated in [2].

Consider the construction of $\mathbf{p}(x_1,...,x_n)$ if we know $\mathbf{p}(x_1|A)$ and $\mathbf{p}(x_2,...,x_n)$, where $A \subseteq \{x_2,...,x_n\}$, using the natural extension. Let us notice the following. If we consider exact probabilities, then $p(x_1,...,x_n) = p(x_1|A)p(x_2,...,x_n)$, i.e. we assume that $p(x_1|A) = p(x_1|x_2,...,x_n)$. Let $\mathbf{p}(x_1|x_2,...,x_n) = \mathbf{p}(x_1|A)$, then the largest credal set $\mathbf{p}(x_1,...,x_n)$ describing the joint distribution of $x_1,...,x_n$ is defined as

$$\mathbf{p}(x_1,...,x_n) = \mathbf{p}(x_1|x_2,...,x_n)\times_I \mathbf{p}(x_2,...,x_n) = \left\{ p(x_1,...,x_n) \left| \begin{array}{l} p(x_1|x_2,...,x_n) \in \mathbf{p}(x_1|x_2,...,x_n), \\ p(x_2,...,x_n) \in \mathbf{p}(x_2,...,x_n) \end{array} \right. \right\}.$$

Consider how this credal set can be constructed using the next example.

Example 4.1. Let us calculate $\mathbf{p}(x_1,x_2) = \mathbf{p}(x_1|x_2)\times_I \mathbf{p}(x_2)$, where x_1 and x_2 are binary random variables taking their values in $\{0,1\}$. For clarity, we will interpret binary variables as follows: x_1 :="The student knows trigonometry", x_2 :="The student will perform an exercise in geometry". Assign $\mathbf{p}(x_1|x_2)$ and $\mathbf{p}(x_2)$ using lower probabilities. Assume that $\underline{p}(x_2=0)=0.3$, $\underline{p}(x_2=1)=0.5$, and $\mathbf{p}(x_1|x_2)$ are given in Table 1.

Table 1. *Conditional lower probabilities for Example 4.1.*

x_2	$\underline{p}(x_1=0\|x_2)$	$\underline{p}(x_1=1\|x_2)$
0	0.6	0.3
1	0.2	0.7

For the sake of brevity, let us denote $p_{ij} = p(x_1 = i, x_2 = j)$, $i,j=0,1$. Then $\mathbf{p}(x_1,x_2) = \mathbf{p}(x_1|x_2)\times_I \mathbf{p}(x_2)$ is described by the following system of linear inequalities:

$$\begin{cases} p_{00}+p_{10} \geqslant \underline{p}(x_2=0), \\ p_{01}+p_{11} \geqslant \underline{p}(x_2=1), \\ p_{00} \geqslant \underline{p}(x_1=0|x_2=0)(p_{00}+p_{10}), \\ p_{10} \geqslant \underline{p}(x_1=1|x_2=0)(p_{00}+p_{10}), \\ p_{01} \geqslant \underline{p}(x_1=0|x_2=1)(p_{01}+p_{11}), \\ p_{11} \geqslant \underline{p}(x_1=1|x_2=1)(p_{01}+p_{11}), \\ p_{00}+p_{01}+p_{10}+p_{11}=1, \\ p_{ij} \geqslant 0, \quad i,j=0,1. \end{cases}$$

Substituting the numerical values of lower probabilities, we get

$$\begin{cases} p_{00}+p_{10} \geqslant 0.3, \\ p_{01}+p_{11} \geqslant 0.5, \\ 0.4p_{00}-0.6p_{10} \geqslant 0, \\ 0.7p_{10}-0.3p_{00} \geqslant 0, \\ 0.8p_{01}-0.2p_{11} \geqslant 0, \\ 0.3p_{11}-0.7p_{01} \geqslant 0, \\ p_{00}+p_{01}+p_{10}+p_{11}=1, \\ p_{ij} \geqslant 0, \quad i,j=0,1. \end{cases} \tag{4}$$

Note that the inequalities $p_{ij} \geqslant 0$, $i,j=0,1$, in system (4) are redundant. For example, the inequality $p_{00} \geqslant 0$ follows from the first and third inequalities in (4), since $0.6(p_{00}+p_{10}) + (0.4p_{00}-0.6p_{10}) = p_{00} \geqslant 0.18$. Next, we simplify system (4), excluding the variable $p_{11} = 1-p_{00}-p_{01}-p_{10}$. As the result, we get

$$\begin{cases} 0.3 \leqslant p_{00}+p_{10} \leqslant 0.5, \\ 0.4p_{00}-0.6p_{10} \geqslant 0, \\ 0.7p_{10}-0.3p_{00} \geqslant 0, \\ 0.2(1-p_{00}-p_{10}) \leqslant p_{01} \leqslant 0.3(1-p_{00}-p_{10}). \end{cases}$$

Let us solve the first three inequalities graphically. The solution is the trapezium with vertices $A=(0.18,0.12)$, $B=(0.3,0.2)$, $C=(0.35,0.15)$, $D=(0.21,0.09)$, shown in Fig. 3.

After that using the inequalities for p_{01} we calculate extreme measures presenting them as points $P_i = (p_{00},p_{10},p_{01},p_{11})$. We see that point A generates two extreme measures:

$$P_1 = (0.18,0.12,0.14,0.56), \quad P_2 = (0.18,0.12,0.21,0.49).$$

Analogously, we find other extreme measures using points B,C,D:

$$P_3 = (0.3,0.2,0.1,0.4), \quad P_4 = (0.3,0.2,0.15,0.35),$$

$$P_5 = (0.35,0.15,0.1,0.4), \quad P_6 = (0.35,0.15,0.15,0.35),$$

$$P_7 = (0.21,0.09,0.14,0.56), \quad P_8 = (0.21,0.09,0.21,0.49).$$

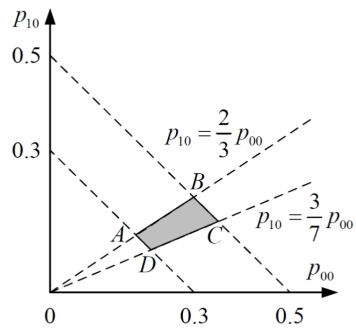

Figure 3. *Graphical solution of (4)*

Example 4.2. Let us calculate $\mathbf{p}(x_1, x_2) = \mathbf{p}(x_1|x_2) \times_S \mathbf{p}(x_2)$ for $\mathbf{p}(x_1|x_2)$ and $\mathbf{p}(x_2)$ from Example 4.1. We see that $\mathbf{p}(x_1|x_2)$ and $\mathbf{p}(x_2)$ are described by credal sets:

$$\mathbf{p}(x_2) = \{ap^{(1)}(x_2) + (1-a)p^{(2)}(x_2)|a \in [0,1]\},$$

$$\mathbf{p}(x_1|x_2 = 0) = \{ap^{(1)}(x_1|x_2 = 0) + (1-a)p^{(2)}(x_1|x_2 = 0)|a \in [0,1]\},$$

$$\mathbf{p}(x_1|x_2 = 1) = \{ap^{(1)}(x_1|x_2 = 1) + (1-a)p^{(2)}(x_1|x_2 = 1)|a \in [0,1]\},$$

where $p^{(1)}(x_2 = 0) = 0.3$, $p^{(2)}(x_2 = 1) = 0.5$, and values

$$p^{(k,n)}(x_1|x_2) = \begin{cases} p^{(k)}(x_1|x_2 = 0), & x_2 = 0, \\ p^{(n)}(x_1|x_2 = 1), & x_2 = 1, \end{cases}$$

are shown in Table 2. (It is sufficient to know $p^{(k,n)}(x_1 = 0|x_2)$, since $p^{(k,n)}(x_1 = 1|x_2) = 1 - p^{(k,n)}(x_1 = 0|x_2)$.)

Table 2. *Values of $p^{(k,n)}(x_1 = 0|x_2)$ for Example 4.2*

x_2	$p^{(1,1)}(x_1 = 0\|x_2)$	$p^{(2,2)}(x_1 = 0\|x_2)$	$p^{(1,2)}(x_1 = 0\|x_2)$	$p^{(2,1)}(x_1 = 0\|x_2)$
0	0.6	0.7	0.6	0.7
1	0.2	0.3	0.3	0.2

Let us find extreme points of $\mathbf{p}(x_1, x_2) = \mathbf{p}(x_1|x_2) \times_S \mathbf{p}(x_2)$. These are conditional densities, obtained by $p^{(k,n,l)}(x_1, x_2) = p^{(k,n)}(x_1|x_2)p^{(l)}(x_2)$, $k, n, l = 1, 2$. The values of these densities are given in Table 3.

The computation results (see Examples 4.1 and 4.2) show that in this case $\mathbf{p}(x_1|x_2) \times_S \mathbf{p}(x_2) = \mathbf{p}(x_1|x_2) \times_I \mathbf{p}(x_2)$. The same equality is true for the more general case considered in

Table 3. *Extreme points of $\mathbf{p}(x_1|x_2) \times_S \mathbf{p}(x_2)$ for Example 4.2*

x_2	$p^{(1,1)}(x_1 = 0\|x_2)$	$p^{(2,2)}(x_1 = 0\|x_2)$	$p^{(1,2)}(x_1 = 0\|x_2)$	$p^{(2,1)}(x_1 = 0\|x_2)$
1,1,1	0.18	0.12	0.14	0.56
2,2,1	0.21	0.09	0.49	0.21
1,1,2	0.3	0.2	0.1	0.4
2,2,2	0.35	0.15	0.15	0.35
1,2,1	0.18	0.12	0.21	0.49
2,1,1	0.21	0.09	0.14	0.56
1,2,2	0.3	0.2	0.15	0.35
2,1,2	0.35	0.15	0.1	0.4

Lemma 4.2. *The following equation is valid:*

$$\mathbf{p}(x_1|x_2, ..., x_n) \times_I \mathbf{p}(x_2, ..., x_n) = \mathbf{p}(x_1|x_2, ..., x_n) \times_S \mathbf{p}(x_2, ..., x_n).$$

Remark 4.2. Clearly, if $A \subseteq \{x_2, ..., x_n\}$, then

$$\mathbf{p}(x_1|A) \times_S \mathbf{p}(x_2, ..., x_n) \subseteq \mathbf{p}(x_1|A) \times_I \mathbf{p}(x_2, ..., x_n),$$

and the equality is fulfilled only in special cases, when $A = \{x_2, ..., x_n\}$ (see Lemma 4.2), or credal sets $\mathbf{p}(x_1|A)$, $\mathbf{p}(x_2, ..., x_n)$ are singletons.

Let us start to formulate the definition of the natural extension of a credal network G.

Definition 4.2. Let G be a credal network defined on the set of random variables $\{x_1, ..., x_n\}$, and σ is a permutation, which is consistent with G. Then the joint probability distribution $\mathbf{p}^{\sigma}(x_1, ..., x_n)$ given by the rule:

$$\mathbf{p}^{\sigma}_{i+1}(x_{\sigma(i+1)}, x_{\sigma(i)}, ..., x_{\sigma(1)}) = \mathbf{p}(x_{\sigma(i+1)}|pa_{\sigma(i+1)}) \times_I \mathbf{p}(x_{\sigma(i)}, ..., x_{\sigma(1)}),$$

where $i = 2, ..., n-1$, for $i = n-1$ is called the *natural σ-extension* of G.

Note that using different consistent permutations we can get various natural σ-extensions of G. Taking this into account, we introduce the following definition of the main natural extension for a credal network.

Definition 4.3. Let G be a credal network defined on the set of random variables $\{x_1, ..., x_n\}$, and let Σ be a set of all permutations consistent with G. Then the *main natural extension* of G is

$$\mathbf{p}(x_1, ..., x_n) = \bigcap_{\sigma \in \Sigma} \mathbf{p}^{\sigma}(x_1, ..., x_n).$$

Example 4.3. Assume that a credal network G consists of two isolated nodes x_1 and x_2. Then the strong extension of G is described by $\mathbf{p}(x_1) \times_S \mathbf{p}(x_2)$, and the main natural extension of G is described by the density $\mathbf{p}(x_1, x_2) = \mathbf{p}(x_1) \times_I \mathbf{p}(x_2) \cap \mathbf{p}(x_2) \times_I \mathbf{p}(x_1)$. A reader can show that these two extensions are not the same in general.

5. The concept of d-separation for credal networks

It would be natural to require that extensions of credal networks have the same properties of conditional independence for random variables within the theory of imprecise probabilities based on the d-separation as for Bayesian networks. We will see later that the strong extension and the main natural extension of a credal network have this property.

The conditional independence (irrelevance) for imprecisely described random variables is defined similarly like for precisely described random variables. A random variable X is irrelevant to random variable Y given random variable Z, if random variable $X|\tilde{Z}$ is irrelevant to random variable $Y|\tilde{Z}$ for all possible instantiations $\tilde{Z}$ of random variable Z. Analogously, X and Y are independent given Z if $X|\tilde{Z}$ and $Y|\tilde{Z}$ are independent for all possible instantiations $\tilde{Z}$ of random variable Z. Assume the joint distribution of random variables X, Y and Z is described by a credal set $\mathbf{p}(X,Y,Z)$, then for checking irrelevance or independence relations between random variables X and Y given Z, we need to find credal sets $\mathbf{p}(X,Y|\tilde{Z})$ and check the validness of irrelevance or independence relation for every instantiation $\tilde{Z}$ of random variable Z.

We will consider next some special cases, when the conditional independence (irrelevance) of random variables is satisfied.

Proposition 5.1. *Let $\mathbf{p}(x_2,...,x_n)$ be a credal set describing the joint distribution of random variables $x_2,...,x_n$, $\mathbf{p}(x_1|B)$ is the conditional credal set given $B \subseteq \{x_2,...,x_n\}$ and*

$$\mathbf{p}(x_1,...,x_n) = \mathbf{p}(x_1|B)\times_S \mathbf{p}(x_2,...,x_n),$$

then x_1 does not depend on random variables in $A = \{x_2,...,x_n\}\backslash B$.

Remark 5.1. Note that the independence relation for random variables is symmetric, but not transitive in general. The independence relation has the same properties for imprecisely given random variables. So Proposition 5.1 says that the random variable x_i does not depend on the random variables A, but the dependence within the set A can be any. Note that in the theory of imprecise probabilities, we consider a weaker version of independence, defined by the irrelevance relation. This relation is not symmetric in general.

Proposition 5.2. *Let $\mathbf{p}(x_2,...,x_n)$ be a credal set describing the joint distribution of random variables $x_2,...,x_n$, $\mathbf{p}(x_1|B)$ is the conditional credal set given $B \subseteq \{x_2,...,x_n\}$, and*

$$\mathbf{p}(x_1,...,x_n) = \mathbf{p}(x_1|B)\times_I \mathbf{p}(x_2,...,x_n),$$

then x_1 is irrelevant to random variables in $A = \{x_2,...,x_n\}\backslash B$.

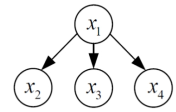

Figure 4. *Simple credal network*

Proposition 5.3. *For disjoint non-empty sets of random variables B, C and D, the following equalities hold:*

$$[\mathbf{p}(D|B)\times_S \mathbf{p}(B)]\times_S \mathbf{p}(C) = \mathbf{p}(D|B)\times_S [\mathbf{p}(B)\times_S \mathbf{p}(C)], \quad (5)$$

$$[\mathbf{p}(D|B)\times_I \mathbf{p}(B)]\times_I \mathbf{p}(C) = \mathbf{p}(D|B)\times_I [\mathbf{p}(B)\times_I \mathbf{p}(C)]. \quad (6)$$

Let us analyze the introduced concepts for the credal network shown in Fig. 4. Consider first the strong extension of this network. Then

$$\mathbf{p}_S(x_1,...,x_4) = \mathbf{p}(x_4|x_1)\times_S \mathbf{p}(x_3|x_1)\times_S \mathbf{p}(x_2|x_1)\times_S \mathbf{p}(x_1),$$

The last expression and Proposition 5.1 imply that x_4 is irrelevant to x_2 and x_3 given x_1. Analogously, considering the representation

$$\mathbf{p}_S(x_1,...,x_4) = \mathbf{p}(x_3|x_1)\times_S \mathbf{p}(x_4|x_1)\times_S \mathbf{p}(x_2|x_1)\times_S \mathbf{p}(x_1),$$

we find that x_3 is irrelevant to x_2 and x_4 given x_1. After that in the similar way we prove that x_2 is irrelevant to x_3 and x_4 given x_1. Therefore, according to the definition random variables x_2, x_3, x_4 are independent given x_1.

Let us show how the same result can be derived for the natural extension of this credal network. Let us consider one of these σ-natural extensions.

$$\mathbf{p}^\sigma(x_1,...,x_4) = \mathbf{p}(x_4|x_1)\times_I \mathbf{p}(x_3|x_1)\times_I \mathbf{p}(x_2|x_1)\times_I \mathbf{p}(x_1).$$

The last expression and Proposition 5.2 imply that x_4 is irrelevant to x_2 and x_3 given x_1 if the joint distribution of $x_1,...,x_4$ is described by $\mathbf{p}^\sigma(x_1,...,x_4)$. Let $\mathbf{p}(x_1,...,x_4)$ be the main natural extension of this network, then

$$\mathbf{p}_S(x_1,...,x_4) \subseteq \mathbf{p}(x_1,...,x_4) \subseteq \mathbf{p}^\sigma(x_1,...,x_4).$$

Since x_4 is irrelevant to x_2 and x_3 given x_1 for the both credal sets $\mathbf{p}_S(x_1,...,x_4)$ and $\mathbf{p}^\sigma(x_1,...,x_4)$ (you can see above the proof of this property for $\mathbf{p}_s$), then x_4 is also irrelevant to x_2 and x_3 given x_1 for the main natural extension $\mathbf{p}(x_1,...,x_4)$. After that we can prove similarly that x_3 is irrelevant to x_2 and x_4 given x_1 , and x_2 is irrelevant to x_3 and x_4 given x_1 for $\mathbf{p}(x_1,...,x_4)$. This implies that x_2, x_3, x_4 are independent given x_1 for the main natural extension of this credal network.

It is possible to prove that x_2, x_3, x_4 are independent given x_1 using another way. Assume that x_1 is an instantiated node, and we denote it by $\tilde{x}_1$. Then the credal network in Fig. 4 given $\tilde{x}_1$ is equivalent to the network with three unconnected nodes x_2, x_3, x_4, in which $p(x_i) := p(x_i|\tilde{x}_1)$, $i = 2, 3, 4$. This implies that random variables $x_i|\tilde{x}_1$, $i = 2, 3, 4$, are independent both for the strong extension and the natural extension of the network. Using Proposition 5.3, one can prove similarly the following.

Proposition 5.4. *Assume that the acyclic graph G of a credal network splits into connected components $G_1,\ldots,G_k$, and the set of nodes in G_i, $i = 1, ..., k$, is denoted by X_i. Then multivariate random variables X_i, $i = 1, ..., k$, are independent both for the strong extension and the main natural extension of the network.*

The result formulated in Proposition 5.4 can be extended to the conditional independence. Next propositions establish some special graphoid properties of the strong extension and the main natural extension of any credal network.

Proposition 5.5. *Assume that X, Y, Z_1, Z_2 are multivariate pairwise disjoint random variables and the joint distribution of X, Z_1 is described by $\mathbf{p}_1(X, Z_1)$, the joint distribution of Y, Z_2 is described by $\mathbf{p}_2(Y, Z_2)$, and the joint distribution of X, Y, Z_1, Z_2 is $\mathbf{p}(X, Y, Z_1, Z_2) = \mathbf{p}_1(X, Z_1)\times_S\mathbf{p}_2(Y, Z_2)$. Then*

1) $\mathbf{p}(X|Z) = \mathbf{p}_1(X|Z_1)$, $\mathbf{p}(Y|Z) = \mathbf{p}_2(Y|Z_2)$,

2) $\mathbf{p}(X, Y|Z) = \mathbf{p}_1(X|Z_1)\times_S\mathbf{p}_2(Y|Z_2)$,

where $Z = Z_1 \cup Z_2$.

Proposition 5.6. *Assume that X, Y, Z_1, Z_2 are multivariate pairwise disjoint random variables and the joint distribution of X, Z_1 is described by $\mathbf{p}_1(X, Z_1)$, the joint distribution of Y, Z_2 is described by $\mathbf{p}_2(Y, Z_2)$, and the joint distribution of X, Y, Z_1, Z_2 is $\mathbf{p}(X, Y, Z_1, Z_2) = \mathbf{p}_1(X, Z_1) \times \mathbf{p}_2(Y, Z_2)$. Then*

1) $\mathbf{p}(X|Z) = \mathbf{p}_1(X|Z_1)$, $\mathbf{p}(Y|Z) = \mathbf{p}_2(Y|Z_2)$,

2) $\mathbf{p}_1(X|Z_1)\times_S\mathbf{p}_2(Y|Z_2) \subseteq \mathbf{p}(X, Y|Z) \subseteq \mathbf{p}_1(X|Z_1) \times \mathbf{p}_2(Y|Z_2)$,

where $Z = Z_1 \cup Z_2$.

Consider a credal network G with the set of nodes X. Let us analyze when multivariate random variables $B, C \subseteq X$ are independent given $\tilde{A} \subseteq X$. Here we assume that $\tilde{A}, B, C$ pairwise disjoint sets. At the first step, we can remove outgoing arcs from nodes in $\tilde{A}$, i.e. if the fragment of G is like in Fig. 5, then we should delete arcs connecting $\tilde{a}_1, \tilde{a}_2, \tilde{a}_3$ with x_2. In a new credal network G' we assign $\mathbf{p}'(x_2|x_1) = \mathbf{p}(x_2|x_1, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3)$. Clearly, the new credal network G' has the same independence relation as the credal network G given $\tilde{A}$. In the next theorem, we will use the following notation. Assume that we consider a credal network G, then we write $a \leqslant b$ for nodes in G if there is a path in G such that $a \to ... \to b$ assuming that $a \leqslant a$ by definition, i.e. $\leqslant$ is a non-strict partial order that corresponds to G.

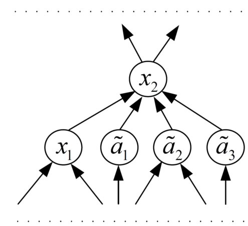

Figure 5. *The fragment of a credal network*

Theorem 5.1. *Assume that G is a credal network with the set of nodes X and a set of nodes $\tilde{A} \subseteq X$ is such that there are no arcs outgoing from $\tilde{A}$. Then multivariate random variables $B, C \subseteq X$ are independent given $\tilde{A}$ both for the strong extension and the main natural extensions of G if there are credal subnetworks G_1 and G_2 of G with the sets of nodes X_1 and X_2 with $X_1 \cap X_2 = \emptyset$ such that*

1) $\tilde{A}_1 = \tilde{A} \cap X_1$, $B \subseteq X_1$, and

- *a) $x \in X_1$ and $y \leqslant x$ implies $y \in X_1$;*
- *b) $x \in X_1$, $a \in \tilde{A}$, and $x \leqslant a$ implies $a \in \tilde{A}_1$;*

2) $\tilde{A}_2 = \tilde{A} \cap X_2$, $C \subseteq X_2$, and

- *a) $x \in X_2$ and $y \leqslant x$ implies $y \in X_2$;*
- *b) $x \in X_2$, $a \in \tilde{A}$, and $x \leqslant a$ implies $a \in \tilde{A}_2$.*

We will describe credal subnetworks described in Theorem 5.1 using the following auxiliary lemma.

Lemma 5.1. *Assume that G is a directed acyclic graph with the set of nodes X that defines the partial order $\leqslant$ on X. Assume also that $\tilde{A}$ is the set of all maximal elements w.r.t. $\leqslant$. Then a subgraph G_1 of G with the set of nodes X_1 and such that*

a) $x \in X_1$ and $y \leqslant x$ implies $y \in X_1$;

b) $x \in X_1$, $a \in \tilde{A}$, and $x \leqslant a$ implies $a \in X_1$;

is minimal if G_1 is a connected component of G. Conversely, every connected component of G has properties a) and b).

Based on Lemma 5.1, we propose the following algorithm for checking conditional independence.

Algorithm 1. *Input data*: a credal network G with the set of nodes X; $\tilde{A}$,B,C are disjoint subsets of X with B,$C \neq \emptyset$.

Output data: the solution whether B and C are independent given $\tilde{A}$ (if $\tilde{A} = \emptyset$, then the result provide us information about the unconditional independence of B and C).

Step 1. Remove arcs outgoing from the set $\tilde{A}$. We will denote the partial order in a new credal network by $\leqslant_{\tilde{A}}$.

Step 2. Create a subnetwork $G^{(1)}$ with the sets of nodes

$$X^{(1)} = \{x \in X | \exists y \in \tilde{A} \cup B \cup C \,:\, x \leqslant_{\tilde{A}} y\}.$$

Step 3. To find the connected components of $G^{(1)}$. If each found connected component does not contain

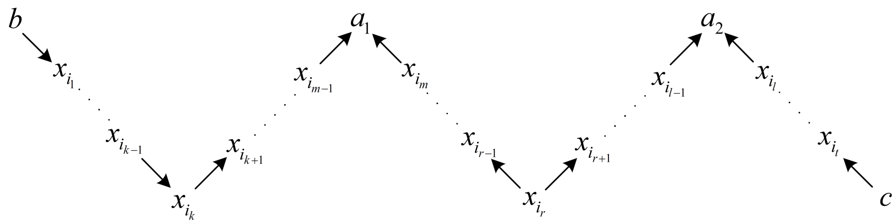

Figure 6. *An example of a chain described in Lemma 5.2*

nodes both from B and C, then multivariate random variables B and C are independent given $\tilde{A}$.

The justification of Algorithm 1. Assume that all found connected components does not contain nodes both from B and C. Let us denote by $\mathbf{G}^{(1)}$ the set of such connected components. Consider the following networks:

$$G_1 = \bigcup_{K_b \in \mathbf{G}^{(1)} | b \in B} K_b, G_2 = \bigcup_{K_c \in \mathbf{G}^{(1)} | c \in C} K_c,$$

where lower indices b and c in K_b and K_c symbolize that the corresponding connected components contain nodes b and c, respectively. Clearly, by Lemma 5.1, networks G_1 and G_2 have properties from Theorem 5.1, i.e. $X_1 \cap X_2 = \emptyset$, and items 1) and 2). Therefore, random variables B and C are independent given $\tilde{A}$.

Let us show that if there is a connected component $K \in \mathbf{G}^{(1)}$, which contains nodes $b \in B$ and $c \in C$, then G_1 and G_2 with the properties from Theorem 5.1 do not exist. Assume to the contrary that such networks G_1 and G_2 exist, however, we found a K in $\mathbf{G}^{(1)}$ with the above properties. Since K is a connected component of $G^{(1)}$, there is a chain γ in K with the head $b \in B$ and the tail $c \in C$. Obviously, the node b belongs to G_1, and let us consequently enumerate in γ nodes as $x_1 = b$, x_2, $\dots$, x_{i-1},x_i, $\dots$,$x_l = c$. Since by our assumption only a part of γ belongs to G_1, there is an $i \in \{1, ..., l\}$ such that $x_{i-1} \in G_1$ and $x_i \notin G_1$. This situation is possible only if $x_{i-1} \to x_i$ in G_1. On the other hand, since $x_i \in G^{(1)}$, then there are three possibilities: 1) $x_i \lessdot_{\tilde{A}} b'$ where $b' \in B$; 2) $x_i \lessdot_{\tilde{A}} a$, where $a \in \tilde{A}$; 3) $x_i \lessdot_{\tilde{A}} c'$, where $c' \in C$. The first two possibilities contradict our assumption that $x_i \notin G_1$. If $x_i \lessdot_{\tilde{A}} c'$, then $x_{i-1}, x_i \in G_2$ and this contradicts our assumption, that networks G_1 and G_2 do not contain common nodes.

Remark 5.2. . Algorithm 1 resembles Theorem 4.1 formulated in [8]. Here, we translate it in terms of directed acyclic graphs considered in our paper.

Theorem 4.1 *Assume that G be a directed acyclic graph with the set of nodes X and $\tilde{A}, B, C$ are disjoint subsets of X with $B, C \neq \emptyset$. Assume that G' is obtained from G using the following procedure:*

1. Delete any leaf node x from G as long as x does not belong to $\tilde{A} \cup B \cup C$. This step is repeated until no more nodes can be deleted.

2. Delete all arcs outgoing from nodes in $\tilde{A}$.

Then B, C are d-separated by $\tilde{A}$ iff sets B, C are disconnected in G'.

Our next goal is to prove that Algorithm 1 exactly checks the property of d-separation. We think that it is possible to prove this using Theorem 4,1. However, we prefer to prove this fact directly as given in the next results.

Lemma 5.2. *Assume that G be a credal network with the set of nodes X, and $\tilde{A}$, $B \neq \emptyset$, $C \neq \emptyset$ are disjoint subsets in X. Then sets B and C are not d-separated by $\tilde{A}$ iff there is a chain γ with a tail $b \in B$ and a head $c \in C$ whose other nodes belongs to $X \backslash (B \cup C)$ and such that the possible configurations of $x \in \gamma$ ($x \neq b$ and $x \neq c$) and its neighbors $y, z \in \gamma$ are*

a) $y \to x \to z$, $y \leftarrow x \leftarrow z$, $y \leftarrow x \to z$, if $x, y, z \notin \tilde{A}$;
b) $y \to x \leftarrow z$ if $x \in \tilde{A}$ and $y, z \notin \tilde{A}$.

Remark 5.3. The description of a chain from Lemma 5.2 has the simple geometrical interpretation. If we depict this chain by a curve like in Fig. 6, then it starts at the node $b \in B$ ($c \in C$) and ends at the node $c \in C$ ($b \in B$) every maximum of this curve is in the set $\tilde{A}$, and other nodes belong to $X \backslash (\tilde{A} \cup B \cup C)$. In Fig. 6, $b \in B$, $c \in C$, $a_1, a_2 \in \tilde{A}$, and $x_{i_1}, ..., x_{i_{k-1}}, x_{i_k}, x_{i_{k+1}}, ..., x_{i_{m-1}}, x_{i_m}, ..., x_{i_{r-1}}, x_{i_r}, x_{i_{r+1}}, ..., x_{i_{l-1}}, x_{i_l}, x_{i_t} \in X \backslash (\tilde{A} \cup B \cup C)$.

Proposition 5.7. *Assume that G is a credal network with the set of nodes X, $\tilde{A}$, B, C are disjoint subsets of X such that B, $C \neq \emptyset$. Then sets B, C are d-separated by $\tilde{A}$ iff random variables B, C are independent given $\tilde{A}$ according to Algorithm 1.*

6. Concluding Remarks

The obtained results allow us to provide inferences in a credal network using its natural extension under the d-separation relation using only irrelevances of random variables, which are described in its σ-natural extensions. It is interesting to evaluate how many of constraints should be involved in this case comparing with the case, when we use the d-separation relation directly. Note that if a credal network G whose every pair of nodes is connected by an arc, then it defines the strict linear order on its nodes and by Lemma 4.2, the main natural extension of G coincides with its strong extension.

Our next goal for the future research is approximate reasoning in credal networks under the d-separation assumption. We think that for this purpose the eliminating nodes algorithms known for Bayesian network may be useful. Note that efficient algorithms for finding exact bounds of conditional probabilities under the d-separation assumption are known for a strong extension of a credal network, whose nodes describe binary random variables. The first such algorithm was proposed in [1] and it generalizes the algorithm [12]. For multi-connected credal networks, methods of finding approximate solutions can be used, for example, based on the iterative application of the algorithm for a singly-connected network or polytree [11].

It is also interesting to apply σ-natural extensions of a credal network, in which only irrelevance relations of random variables are assumed [9], and these relations do not accumulate all errelevances inherent to d-separation. So, for example, for oriented trees, random variables are assumed to be conditionally irrelevant, lying below and above a given node. Under such assumptions, it is possible to construct probabilistic inference algorithms of linear complexity w.r.t. the number of nodes in a tree [10].

Additional author information

Acknowledgements. The work was carried out within the state assignment of the Russian Federation No. 103-00001-25-02 dated 20.03.2025. The authors would like to express their great gratitude to the anonymous reviewers for their helpful comments.

References

[1] Alessandro Antonucci and Marco Zaffalon. "Decision-theoretic specification of credal networks: A unified language for uncertain modeling with sets of Bayesian networks". In: *International Journal of Approximate Reasoning* 49.2 (2008), pp. 345–361. DOI: 10.1016/j.ijar.2008.02.005.

[2] T. Augustin, F.P.A. Coolen, G. de Cooman, and M.C.M. Troffaes. *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. Wiley, 2014. DOI: 10.1002/9781118763117.

[3] Andrey G. Bronevich and George J. Klir. "Measures of uncertainty for imprecise probabilities: An axiomatic approach". In: *International Journal of Approximate Reasoning* 51.4 (2010), pp. 365–390. DOI: 10.1016/j.ijar.2009.11.003.

[4] Inés Couso, Serafín Moral, and Peter Walley. "Examples of independence for imprecise probabilities." In: *Proceedings of the First International Symposium on Imprecise Probability: Theories and Applications*. Ed. by Gert De Cooman, Fábio G. Cozman, Serafín Moral, and Peter Walley. 1999, pp. 121–130.

[5] Fabio G. Cozman. "Credal networks". In: *Artificial Intelligence* 120.2 (2000), pp. 199–233. DOI: 10.1016/S0004-3702(00)00029-1.

[6] Fabio G. Cozman. "Graphical models for imprecise probabilities". In: *International Journal of Approximate Reasoning* 39.2 (2005), pp. 167–184. DOI: 10.1016/j.ijar.2004.10.00.

[7] Fabio G. Cozman and Peter Walley. "Graphoid properties of epistemic irrelevance and independence". In: *Ann Math Artif Intell* 45.2 (2005), pp. 173–195. DOI: 10.1007/s10472-005-9004-z.

[8] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511811357.

[9] Jasper De Bock. "Credal networks under epistemic irrelevance". In: *International Journal of Approximate Reasoning* 85 (2017), pp. 107–138. DOI: 10.1016/j.ijar.2017.03.012.

[10] Gert de Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon. "Epistemic irrelevance in credal nets: The case of imprecise Markov trees". In: *International Journal of Approximate Reasoning* 51.9 (2010), pp. 1029–1052. DOI: 10.1016/j.ijar.2010.08.011.

[11] Jaime S. Ide and Fabio G. Cozman. "Approximate algorithms for credal networks with binary variables". In: *International Journal of Approximate Reasoning* 48.1 (2008), pp. 275–296. DOI: 10.1016/j.ijar.2007.09.003.

[12] Judea Pearl. "Fusion, propagation, and structuring in belief networks". In: *Artificial Intelligence* 29.3 (1986), pp. 241–288. DOI: 10.1016/0004-3702(86)90072-X.

[13] Thomas Verma and Judea Pearl. "Causal networks: semantics and expressiveness". In: *Uncertainty in Artificial Intelligence*. Ed. by Ross D. Shachter, Tod S. Levitt, Laveen N. Kanal, and John F. Lemmer. Vol. 9. Machine Intelligence and Pattern Recognition. North-Holland, 1990, pp. 69–76. DOI: 10.1016/B978-0-444-88650-7.50011-1.

[14] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1991.