Tight Complexity Bounds for Reasoning in the Description Logic $\mathcal{BEL}$

İsmail İlkan Ceylan[1⋆] and Rafael Peñaloza[1,2⋆⋆]

[1] Theoretical Computer Science, TU Dresden, Germany
[2] Center for Advancing Electronics Dresden
{ceylan,penaloza}@tcs.inf.tu-dresden.de

Abstract. Recently, Bayesian extensions of Description Logics, and in particular the logic $\mathcal{BEL}$, were introduced as a means of representing certain knowledge that depends on an uncertain context. In this paper we introduce a novel structure, called *proof structure*, that encodes the contextual information required to deduce subsumption relations from a $\mathcal{BEL}$ knowledge base. Using this structure, we show that probabilistic reasoning in $\mathcal{BEL}$ can be reduced in polynomial time to standard Bayesian network inferences, thus obtaining tight complexity bounds for reasoning in $\mathcal{BEL}$.

1 Introduction

Description Logics (DLs) [2] are a family of knowledge representation formalisms that are characterized by their clear syntax, and formal, unambiguous semantics. DLs have been successfully employed for creating large knowledge bases, representing real application domains, prominently from the life sciences. Examples of such knowledge bases are SNOMED CT, GALEN, or the Gene Ontology.

A prominent missing feature of classical DLs is the capacity of specifying a *context* in which a portion of the knowledge holds. For instance, the behaviour of a system may depend on factors that are extrogenous to the domain, such as the weather conditions. For that reason, approaches for handling contexts in DLs have been studied; see e.g. [13, 14]. Since the specific context in which the ontology is being applied (e.g., the weather) may be uncertain, it is important to adapt context-based reasoning to consider also a probabilistic distribution over the contexts. Recently, $\mathcal{BEL}$ [7] and other probabilistic extensions of DLs [8] were introduced to describe certain knowledge that depends on an uncertain context, which is described by a Bayesian network (BN). Using these logics, one can represent knowledge that holds e.g., when it rains. Interestingly, reasoning in $\mathcal{BEL}$ can be decoupled between the logical part, and BN inferences. However, despite the logical component of this logic being decidable in polynomial time, the best known algorithm for probabilistic reasoning in $\mathcal{BEL}$ runs in exponential time.

⋆ Supported by DFG within the Research Training Group "RoSI" (GRK 1907).
⋆⋆ Partially supported by DFG within the Cluster of Excellence 'cfAED'.

We use a novel structure, called the *proof structure*, to reduce probabilistic reasoning for a $\mathcal{BEL}$ knowledge base to probabilistic inferences in a BN. Briefly, a proof structure describes all contexts that entail the wanted consequence. The BN can then be used to compute the probability of these contexts, which yields the probability of the entailment. Since this reduction can be done in polynomial time, it provides tight upper bounds for the complexity of reasoning in $\mathcal{BEL}$.

2 $\mathcal{EL}$ and Proof Structures

$\mathcal{EL}$ is a light-weight DL that allows for polynomial-time reasoning. It is based on *concepts* and *roles*, corresponding to unary and binary predicates from first-order logic, respectively. $\mathcal{EL}$ concepts are built inductively from disjoint, countably infinite sets $\mathsf{N_C}$ and $\mathsf{N_R}$ of *concept names* and *role names*, and applying the syntax rule $C ::= A \mid \top \mid C \sqcap C \mid \exists r.C$, where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$.

The *semantics* of $\mathcal{EL}$ is given by *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where $\Delta^\mathcal{I}$ is a non-empty *domain* and $\cdot^\mathcal{I}$ is an *interpretation function* that maps every $A \in \mathsf{N_C}$ to a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ and every role name r to a binary relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The interpretation function $\cdot^\mathcal{I}$ is extended to $\mathcal{EL}$ concepts by defining $\top^\mathcal{I} := \Delta^\mathcal{I}$, $(C \sqcap D)^\mathcal{I} := C^\mathcal{I} \cap D^\mathcal{I}$, and $(\exists r.C)^\mathcal{I} := \{d \in \Delta^\mathcal{I} \mid \exists e : (d, e) \in r^\mathcal{I} \wedge e \in C^\mathcal{I}\}$. The knowledge of a domain is represented through a set of axioms restricting the interpretation of the concepts.

Definition 1 (TBox). *A* general concept inclusion (GCI) *is an expression of the form $C \sqsubseteq D$, where C, D are concepts. A* TBox *$\mathcal{T}$ is a finite set of GCIs. The* signature *of $\mathcal{T}$ ($\mathsf{sig}(\mathcal{T})$) is the set of concept and role names appearing in $\mathcal{T}$. An interpretation $\mathcal{I}$ satisfies the GCI $C \sqsubseteq D$ iff $C^\mathcal{I} \subseteq D^\mathcal{I}$; $\mathcal{I}$ is a* model *of the TBox $\mathcal{T}$ iff it satisfies all the GCIs in $\mathcal{T}$.*

The main reasoning service in $\mathcal{EL}$ is deciding the subsumption relations between concepts based on their semantic definitions. A concept C is *subsumed by* D w.r.t. the TBox $\mathcal{T}$ ($\mathcal{T} \models C \sqsubseteq D$) iff $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{T}$.

It has been shown that subsumption can be decided in $\mathcal{EL}$ by a completion algorithm in polynomial time [1]. This algorithm requires the TBox to be in *normal form*; i.e., where all axioms in the TBox are of one of the forms $A \sqsubseteq B \mid A \sqcap B \sqsubseteq C \mid A \sqsubseteq \exists r.B \mid \exists r.B \sqsubseteq A$. It is well known that every TBox can be transformed into an equivalent one in normal form of linear size [1, 5]; for the rest of this paper, we assume that $\mathcal{T}$ is a TBox in normal form.

We are interested in deriving the subsumption relations in normal form that follow from $\mathcal{T}$; we call the set of all these subsumption relations the *normalised logical closure* of $\mathcal{T}$. This closure can be computed by an exhaustive application of the *deduction rules* from Table 1. Each rule maps a set of premises S to its consequence α; such a rule is *applicable* to a TBox $\mathcal{T}$ if $S \subseteq \mathcal{T}$ but $\alpha \notin \mathcal{T}$. In that case, its application adds α to $\mathcal{T}$. It is easy to see that these rules produce the normalised logical closure of the input TBox. Moreover, the deduction rules introduce only GCIs in normal form, and do not change the signature. Hence, if $n = |\mathsf{sig}(\mathcal{T})|$, the logical closure of $\mathcal{T}$ is found after at most n^3 rule applications.

Table 1: Deduction rules for $\mathcal{EL}$.

$\mapsto$	Premises (S)	Result (α)	$\mapsto$	Premises (S)	Result (α)
1	$\langle A \sqsubseteq B\rangle$, $\langle B \sqsubseteq C\rangle$	$\langle A \sqsubseteq C\rangle$	7	$\langle A \sqsubseteq \exists r.B\rangle$, $\langle \exists r.B \sqsubseteq C\rangle$	$\langle A \sqsubseteq C\rangle$
2	$\langle A \sqsubseteq \exists r.B\rangle$, $\langle B \sqsubseteq C\rangle$	$\langle A \sqsubseteq \exists r.C\rangle$	8	$\langle A \sqcap B \sqsubseteq C\rangle$, $\langle C \sqsubseteq X\rangle$	$\langle A \sqcap B \sqsubseteq X\rangle$
3	$\langle A \sqsubseteq \exists r.B\rangle$, $\langle C \sqsubseteq A\rangle$	$\langle C \sqsubseteq \exists r.B\rangle$	9	$\langle A \sqcap B \sqsubseteq C\rangle$, $\langle X \sqsubseteq A\rangle$	$\langle X \sqcap B \sqsubseteq C\rangle$
4	$\langle \exists r.A \sqsubseteq B\rangle$, $\langle B \sqsubseteq C\rangle$	$\langle \exists r.A \sqsubseteq C\rangle$	10	$\langle A \sqcap B \sqsubseteq C\rangle$, $\langle X \sqsubseteq B\rangle$	$\langle A \sqcap X \sqsubseteq C\rangle$
5	$\langle \exists r.A \sqsubseteq B\rangle$, $\langle C \sqsubseteq A\rangle$	$\langle \exists r.C \sqsubseteq B\rangle$	11	$\langle X \sqcap X \sqsubseteq C\rangle$	$\langle X \sqsubseteq C\rangle$
6	$\langle \exists r.A \sqsubseteq B\rangle$, $\langle B \sqsubseteq \exists r.C\rangle$	$\langle A \sqsubseteq C\rangle$			

We will later associate a probability to the GCIs in the TBox $\mathcal{T}$, and will be interested in computing the probability of a subsumption. It will then be useful to be able not only to derive the GCI, but also *all* the sub-TBoxes of $\mathcal{T}$ from which it follows. Therefore, we store the traces of the deduction rules using a directed hypergraph. A *directed hypergraph* is a tuple $H = (V, E)$ where V is a non-empty set of *vertices* and E is a set of *directed hyper-edges* of the form $e = (S, v)$ where $S \subseteq V$ and $v \in V$. Given $S \subseteq V$ and $v \in V$, a *path from S to v in H* of *length n* is a sequence of hyper-edges $(S_1, v_1), (S_2, v_2), \ldots, (S_n, v_n)$ where $v_n = v$ and $S_i \subseteq S \cup \{v_j \mid 0 < j < i\}$ for all $i, 1 \leq i \leq n$.

Given a TBox $\mathcal{T}$ in normal form, we build the hypergraph $H_\mathcal{T} = (V_\mathcal{T}, E_\mathcal{T})$, where $V_\mathcal{T}$ is the set of all GCIs in normal form that follow from $\mathcal{T}$ over the same signature and $E_\mathcal{T} = \{(S, \alpha) \mid S \mapsto \alpha, S \subseteq V_\mathcal{T}\}$, with $\mapsto$ the deduction relation defined in Table 1. We call this hypergraph the *proof structure* of $\mathcal{T}$. From the soundness and completeness of the deduction rules, we get the following lemma.

Lemma 2. *Let $\mathcal{T}$ be a TBox in normal form, $H_\mathcal{T} = (V_\mathcal{T}, E_\mathcal{T})$ its proof structure, $\mathcal{O} \subseteq \mathcal{T}$, and $C \sqsubseteq D \in V_\mathcal{T}$. There is a path from $\mathcal{O}$ to $C \sqsubseteq D$ in $H_\mathcal{T}$ iff $\mathcal{O} \models C \sqsubseteq D$.*

$H_\mathcal{T}$ is a compact representation of all the possible derivations of a GCI from the GCIs in $\mathcal{T}$ [3, 4]. Traversing this hypergraph backwards from a GCI α being entailed by $\mathcal{T}$, one constructs all proofs for α; hence the name "proof structure." Since $|V_\mathcal{T}| \leq |\mathsf{sig}(\mathcal{T})|^3$, it suffices to consider paths of length at most $|\mathsf{sig}(\mathcal{T})|^3$.

Clearly, the proof structure $H_\mathcal{T}$ can be cyclic. To simplify the process of finding the causes of a GCI being entailed, we build an *unfolded* version of this hypergraph by making different copies of each node. In this case, nodes are pairs of axioms and labels, where the latter indicates to which level the nodes belong in the hypergraph. Given a set of axioms S, and $i \geq 0$, $S^i := \{(\alpha, i) \mid \alpha \in S\}$ denotes the i-labeled set of GCIs in S. Let $n := |\mathsf{sig}(\mathcal{T})|^3$, we define the sets $W_i, 0 \leq i \leq n$ inductively by setting $W_0 := \{(\alpha, 0) \mid \alpha \in \mathcal{T}\}$ and for all i, $0 \leq i < n$

$$W_{i+1} := \{(\alpha, i+1) \mid S^i \subseteq W_i, S \mapsto \alpha\} \ \cup \ \{(\alpha, i+1) \mid (\alpha, i) \in W_i\}.$$

For each $i, 0 \leq i \leq n$, W_i contains all the GCIs that can be derived by at most i applications of the deduction rules from Table 1. The *unfolded proof structure* of $\mathcal{T}$ is the hypergraph $H^u_\mathcal{T} = (W_\mathcal{T}, F_\mathcal{T})$, where $W_\mathcal{T} := \bigcup_{i=0}^n W_i$ and $F_\mathcal{T} := \bigcup_{i=1}^n F_i$,

$$F_{i+1} := \{(S^i, (\alpha, i+1)) \mid S^i \subseteq W_i, S \mapsto \alpha\} \cup \{(\{(\alpha, i)\}, (\alpha, i+1)) \mid (\alpha, i) \in W_i\}.$$

Algorithm 1 Construction of the pruned proof structure

Input: TBox $\mathcal{T}$

Output: $H = (W, F)$ pruned proof structure for $\mathcal{T}$

$V_0 \leftarrow \mathcal{T}$, $E_0 \leftarrow \emptyset$, $i \leftarrow 0$
do
  $i \leftarrow i + 1$
  $V_i := V_{i-1} \cup \{\alpha \mid S \mapsto \alpha, S \subseteq V_{i-1}\}$
  $E_i = \{(S, \alpha) \mid S \mapsto \alpha, S \subseteq V_{i-1}\}$
while $V_i \neq V_{i-1}$ or $E_i \neq E_{i-1}$
$W := \{(\alpha, k) \mid \alpha \in V_k, 0 \leq k \leq i\}$
$E := \{(S, (\alpha, k)) \mid (S, \alpha) \in E_k, 0 \leq k \leq i\} \cup \{(\{(\alpha, k)\}, (\alpha, k{+}1)) \mid \alpha \in V_k, 0 \leq k < i\}$
return (W, E)

The following is a simple consequence of our constructions and Lemma 2.

Theorem 3. *Let $\mathcal{T}$ be a TBox, and $H_\mathcal{T} = (V_\mathcal{T}, E_\mathcal{T})$ and $H^u_\mathcal{T} = (W_\mathcal{T}, F_\mathcal{T})$ the proof structure and unfolded proof structure of $\mathcal{T}$, respectively. Then,*

1. *for all $C \sqsubseteq D \in V_\mathcal{T}$ and all $\mathcal{O} \subseteq \mathcal{T}$, $\mathcal{O} \models C \sqsubseteq D$ iff there is a path from $\{(\alpha, 0) \mid \alpha \in \mathcal{O}\}$ to $(C \sqsubseteq D, n)$ in $H^u_\mathcal{T}$, and*
2. *$(S, \alpha) \in E_\mathcal{T}$ iff $(S^{n-1}, (\alpha, n)) \in F_\mathcal{T}$.*

The unfolded proof structure of a TBox $\mathcal{T}$ is thus guaranteed to contain the information of all possible causes for a GCI to follow from $\mathcal{T}$. Moreover, this hypergraph is acyclic, and has polynomially many nodes, on the size of $\mathcal{T}$, by construction. Yet, this hypergraph may contain many redundant nodes. Indeed, it can be the case that all the simple paths in $H_\mathcal{T}$ starting from a subset of $\mathcal{T}$ are of length $k < n$. In that case, $W_i = W_{i+1}$ and $F_i = F_{i+1}$ hold for all $i \geq k$, modulo the second component. It thus suffices to consider the sub-hypergraph of $H^u_\mathcal{T}$ that contains only the nodes $\bigcup_{i=0}^{k} W_i$. Algorithm 1 describes a method for computing this pruned hypergraph. In the worst case, this algorithm will produce the whole unfolded proof structure of $\mathcal{T}$, but will stop the unfolding procedure earlier if possible. The **do-while** loop is executed at most $|\mathsf{sig}(\mathcal{T})|^3$ times, and each of these loops requires at most $|\mathsf{sig}(\mathcal{T})|^3$ steps.

Lemma 4. *Algorithm 1 terminates in time polynomial on the size of $\mathcal{T}$.*

We briefly illustrate the execution of Algorithm 1 on a simple TBox.

Example 5. Consider the $\mathcal{EL}$ TBox $\mathcal{T} = \{A \sqsubseteq B, B \sqsubseteq C, B \sqsubseteq D, C \sqsubseteq D\}$. The first levels of the unfolded proof structure of $\mathcal{T}$ are shown in Figure 1.[3] The first level V_0 of this hypergraph contains a representative for each axiom in $\mathcal{T}$. To construct the second level, we first copy all the GCIs in V_0 to V_1, and add a hyperedge joining the equivalent GCIs (represented by dashed lines in Figure 1). Then, we apply all possible deduction rules to the elements of V_0, and add a

[3] For the illustrations we drop the second component of the nodes, but visually make the level information explicit.

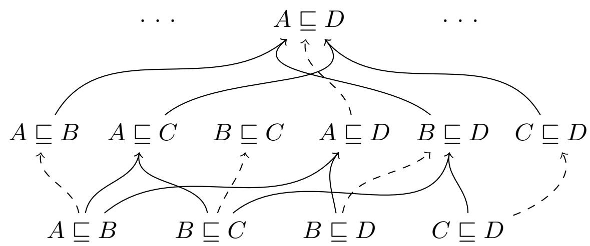

Fig. 1: The first levels of an unfolded proof structure and the paths to $\langle A \sqsubseteq D \rangle$

hyperedge from the premises at level V_0 to the conclusion at level V_1 (continuous lines). This procedure is repeated at each subsequent level. Notice that the set of GCIs at each level is monotonically increasing. Additionally, for each GCI, the in-degree of each representative monotonically increases throughout the levels.

In the next section, we recall $\mathcal{BEL}$, a probabilistic extension of $\mathcal{EL}$ based on Bayesian networks [7], and use the construction of the (unfolded) proof structure to provide tight complexity bounds for reasoning in this logic.

3 The Bayesian Description Logic $\mathcal{BEL}$

The probabilistic Description Logic $\mathcal{BEL}$ extends $\mathcal{EL}$ by associating every GCI in a TBox with a probabilistic context. The joint probability distribution of the contexts is encoded in a Bayesian network [12]. A *Bayesian network* (BN) is a pair $\mathcal{B} = (G, \Phi)$, where $G = (V, E)$ is a finite directed acyclic graph (DAG) whose nodes represent Boolean random variables,[4] and Φ contains, for every node $x \in V$, a conditional probability distribution $P_{\mathcal{B}}(x \mid \pi(x))$ of x given its parents $\pi(x)$. If V is the set of nodes in G, we say that $\mathcal{B}$ is a BN *over* V.

Intuitively, $G = (V, E)$ encodes a series of conditional independence assumptions between the random variables: every variable $x \in V$ is conditionally independent of its non-descendants given its parents. Thus, every BN $\mathcal{B}$ defines a unique joint probability distribution over V where $P_{\mathcal{B}}(V) = \prod_{x \in V} P_{\mathcal{B}}(x \mid \pi(x))$. As with classical DLs, the main building blocks in $\mathcal{BEL}$ are concepts, which are syntactically built as $\mathcal{EL}$ concepts. The domain knowledge is encoded by a generalization of TBoxes, where axioms are annotated with a context, defined by a set of literals belonging to a BN.

Definition 6 (KB). *Let V be a finite set of Boolean variables. A V-*literal *is an expression of the form x or $\neg x$, where $x \in V$; a V-*context *is a consistent set of V-literals.*

*A V-*restricted general concept inclusion *(V-GCI) is of the form $\langle C \sqsubseteq D : \kappa \rangle$ where C and D are $\mathcal{BEL}$ concepts and κ is a V-context. A V-*TBox *is a finite*

[4] In their general form, BNs allow for arbitrary discrete random variables. We restrict w.l.o.g. to Boolean variables for ease of presentation.

set of V-GCIs. A $\mathcal{BEL}$ knowledge base *(KB) over V is a pair $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ where $\mathcal{B}$ is a BN over V and $\mathcal{T}$ is a V-TBox.*[5]

The semantics of $\mathcal{BEL}$ extends the semantics of $\mathcal{EL}$ by additionally evaluating the random variables from the BN. Given a finite set of Boolean variables V, a *V-interpretation* is a tuple $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$ where $\Delta^{\mathcal{I}}$ is a non-empty set called the *domain*, $\mathcal{V}^{\mathcal{I}} : V \to \{0, 1\}$ is a *valuation* of the variables in V, and $\cdot^{\mathcal{I}}$ is an *interpretation function* that maps every concept name A to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and every role name r to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

The interpretation function $\cdot^{\mathcal{I}}$ is extended to arbitrary $\mathcal{BEL}$ concepts as in $\mathcal{EL}$ and the valuation $\mathcal{V}^{\mathcal{I}}$ is extended to contexts by defining, for every $x \in V$, $\mathcal{V}^{\mathcal{I}}(\neg x) = 1 - \mathcal{V}^{\mathcal{I}}(x)$, and for every context κ, $\mathcal{V}^{\mathcal{I}}(\kappa) = \min_{\ell \in \kappa} \mathcal{V}^{\mathcal{I}}(\ell)$, where $\mathcal{V}^{\mathcal{I}}(\emptyset) := 1$. Intuitively, a context κ can be thought as a conjunction of literals, which is evaluated to 1 iff each literal is evaluated to 1.

The V-interpretation $\mathcal{I}$ is a *model* of the V-GCI $\langle C \sqsubseteq D : \kappa \rangle$, denoted as $\mathcal{I} \models \langle C \sqsubseteq D : \kappa \rangle$, iff (i) $\mathcal{V}^{\mathcal{I}}(\kappa) = 0$, or (ii) $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. It is a *model* of the V-TBox $\mathcal{T}$ iff it is a model of all the V-GCIs in $\mathcal{T}$. The idea is that the restriction $C \sqsubseteq D$ is only required to hold whenever the context κ is satisfied. Thus, any interpretation that violates the context trivially satisfies the whole V-GCI.

Example 7. Let $V_0 = \{x, y, z\}$, and consider the V_0-TBox

$$\mathcal{T}_0 := \{\langle A \sqsubseteq C : \{x, y\} \rangle\, ,\ \langle A \sqsubseteq B : \{\neg x\} \rangle\, ,\ \langle B \sqsubseteq C : \{\neg x\} \rangle\}.$$

The V_0-interpretation $\mathcal{I}_0 = (\{d\}, \cdot^{\mathcal{I}_0}, \mathcal{V}_0)$ with $\mathcal{V}_0(\{x, \neg y, z\}) = 1$, $A^{\mathcal{I}_0} = \{d\}$, and $B^{\mathcal{I}_0} = C^{\mathcal{I}_0} = \emptyset$ is a model of $\mathcal{T}_0$, but is not a model of the V_0-GCI $\langle A \sqsubseteq B : \{x\} \rangle$, since $\mathcal{V}_0(\{x\}) = 1$ but $A^{\mathcal{I}_0} \not\subseteq B^{\mathcal{I}_0}$.

A V-TBox $\mathcal{T}$ is in *normal form* if for each V-GCI $\langle \alpha : \kappa \rangle \in \mathcal{T}$, α is an $\mathcal{EL}$ GCI in normal form. A $\mathcal{BEL}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ is in *normal form* if $\mathcal{T}$ is in normal form. As for $\mathcal{EL}$, every $\mathcal{BEL}$ KB can be transformed into an equivalent one in normal form in polynomial time [6]. Thus, we consider only $\mathcal{BEL}$ KBs in normal form in the following. The DL $\mathcal{EL}$ is a special case of $\mathcal{BEL}$ in which all V-GCIs are of the form $\langle C \sqsubseteq D : \emptyset \rangle$. Notice that every valuation satisfies the empty context $\emptyset$; thus, a V-interpretation $\mathcal{I}$ satisfies the V-GCI $\langle C \sqsubseteq D : \emptyset \rangle$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. We say that $\mathcal{T}$ *entails* $\langle C \sqsubseteq D : \emptyset \rangle$ ($\mathcal{T} \models C \sqsubseteq D$), if every model of $\mathcal{T}$ is also a model of $\langle C \sqsubseteq D : \emptyset \rangle$. For a valuation $\mathcal{W}$ of the variables in V, we define the TBox containing all axioms that must be satisfied in any V-interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$ with $\mathcal{V}^{\mathcal{I}} = \mathcal{W}$.

Definition 8 (restriction). *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a $\mathcal{BEL}$ KB. The* restriction of *$\mathcal{T}$ to a valuation $\mathcal{W}$ of the variables in V is the V-TBox*

$$\mathcal{T}_{\mathcal{W}} := \{\langle C \sqsubseteq D : \emptyset \rangle \mid \langle C \sqsubseteq D : \kappa \rangle \in \mathcal{T}, \mathcal{W}(\kappa) = 1\}.$$

To handle the probabilistic knowledge provided by the BN, we extend the semantics of $\mathcal{BEL}$ through multiple-world interpretations. A V-interpretation describes

[5] Unless stated otherwise, we assume that $\mathcal{K}$ is over V in the rest of the paper.

a possible world; by assigning a probabilistic distribution over these interpretations, we describe the required probabilities, which should be consistent with the BN provided in the knowledge base.

Definition 9 (probabilistic model). *A* probabilistic interpretation *is a pair* $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$*, where* $\mathfrak{I}$ *is a set of* V*-interpretations and* $P_{\mathfrak{I}}$ *is a probability distribution over* $\mathfrak{I}$ *such that* $P_{\mathfrak{I}}(\mathcal{I}) > 0$ *only for finitely many interpretations* $\mathcal{I} \in \mathfrak{I}$*.* $\mathcal{P}$ *is a* model *of the TBox* $\mathcal{T}$ *if every* $\mathcal{I} \in \mathfrak{I}$ *is a model of* $\mathcal{T}$*.* $\mathcal{P}$ *is* consistent *with the BN* $\mathcal{B}$ *if for every possible valuation* $\mathcal{W}$ *of the variables in* V *it holds that*

$$\sum_{\mathcal{I} \in \mathfrak{I},\, \mathcal{V}^{\mathcal{I}} = \mathcal{W}} P_{\mathfrak{I}}(\mathcal{I}) = P_{\mathcal{B}}(\mathcal{W}).$$

$\mathcal{P}$ *is a* model *of the KB* $(\mathcal{B}, \mathcal{T})$ *iff it is a model of* $\mathcal{T}$ *and consistent with* $\mathcal{B}$*.*

One simple consequence of this semantics is that probabilistic models preserve the probability distribution of $\mathcal{B}$ for contexts; the probability of a context κ is the sum of the probabilities of all valuations that extend κ.

3.1 Contextual Subsumption

Just as in classical DLs, we want to extract the information that is implicitly encoded in a $\mathcal{BEL}$ KB. In particular, we are interested in solving different reasoning tasks for this logic. One of the fundamental reasoning problems in $\mathcal{EL}$ is subsumption: is a concept C always interpreted as a subconcept of D? This problem is extended to also consider the contexts in $\mathcal{BEL}$.

Definition 10 (contextual subsumption). *Let* $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ *be a* $\mathcal{BEL}$ *KB,* C, D *be two* $\mathcal{BEL}$ *concepts, and* κ *a* V*-context.* C is contextually subsumed by D in κ w.r.t. $\mathcal{K}$*, denoted* $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle$*, if every probabilistic model of* $\mathcal{K}$ *is also a model of* $\{\langle C \sqsubseteq D : \kappa \rangle\}$*.*

Contextual subsumption depends only on the contexts, and not on their associated probabilities. It was shown in [7] that contextual subsumption is coNP-hard, even if considering only the empty context. To show that the problem is in fact coNP-complete, we use the following lemma also shown in [7].

Lemma 11. *Let* $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ *be a KB. Then* $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle$ *iff for every valuation* $\mathcal{W}$ *with* $\mathcal{W}(\kappa) = 1$*, it holds that* $\mathcal{T}_{\mathcal{W}} \models C \sqsubseteq D$*.*

Using this lemma, it is easy to see that contextual subsumption is in coNP: to decide that the subsumption does not hold, we simply guess a valuation $\mathcal{W}$ and verify in polynomial time that $\mathcal{W}(\kappa) = 1$ and $\mathcal{T}_{\mathcal{W}} \not\models C \sqsubseteq D$.

Corollary 12. *Contextual subsumption is co*NP*-complete.*

In $\mathcal{BEL}$ one might be interested in finding the probability with which such a consequence holds, or given a subsumption relation, computing the most probable context in which it holds. For the rest of this section, we formally define these reasoning tasks, and provide a method for solving them based on Bayesian networks inferences.

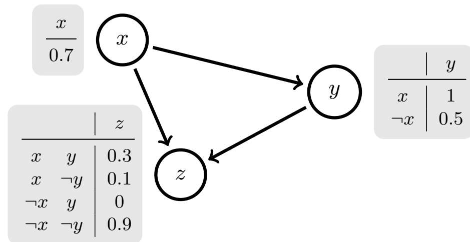

Fig. 2: A simple BN

3.2 Probabilistic Subsumption

We generalize subsumption between concepts to consider also the probabilities provided by the BN.

Definition 13 (p-subsumption). *Let $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ be a $\mathcal{BEL}$ KB, C, D two $\mathcal{BEL}$ concepts, and κ a V-context. For a probabilistic interpretation $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$, we define $P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa\rangle) := \sum_{\mathcal{I}\in\mathfrak{I},\mathcal{I}\models\langle C\sqsubseteq D:\kappa\rangle} P_{\mathfrak{I}}(\mathcal{I})$. The* probability of the V-GCI $\langle C \sqsubseteq D : \kappa\rangle$ w.r.t. $\mathcal{K}$ *is defined as*

$$P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) := \inf_{\mathcal{P}\models\mathcal{K}} P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa\rangle).$$

We say that C is p-subsumed *by D in κ, for $p \in (0, 1]$ if $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) \geq p$.*

Proposition 14 ([7]). *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB. Then*

$$P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) = 1 - P_{\mathcal{B}}(\kappa) + \sum_{\substack{\mathcal{T}_{\mathcal{W}}\models C\sqsubseteq D\\ \mathcal{W}(\kappa)=1}} P_{\mathcal{B}}(\mathcal{W}).$$

Example 15. Consider the KB $\mathcal{K}_0 = (\mathcal{B}_0, \mathcal{T}_0)$, where $\mathcal{B}_0$ is the BN from Figure 2 and $\mathcal{T}_0$ the V_0-TBox from Example 7. It follows that $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \{x, y\}\rangle) = 1$ from the first V-GCI in $\mathcal{T}$ and $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \{\neg x\}\rangle) = 1$ from the others since any model of $\mathcal{K}_0$ needs to satisfy the V-GCIs asserted in $\mathcal{T}$ by definition. Notice that $A \sqsubseteq C$ does not hold in context $\{x, \neg y\}$, but $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \{x, \neg y\}\rangle) = 1$. Since this describes all contexts, we conclude $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \emptyset\rangle) = 1$.

Deciding p-subsumption We show that deciding p-subsumption can be reduced to deciding the D-PR problem over a Bayesian network. Given a BN $\mathcal{B} = (G, \Phi)$ over V and a V-context κ, the *D-PR problem* consists on deciding whether $P_{\mathcal{B}}(\kappa) > p$. This problem is known to be PP-complete [9, 22].

Let $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ be an arbitrary but fixed $\mathcal{BEL}$ KB. From the labelled V-TBox $\mathcal{T}$, we construct the $\mathcal{EL}$ TBox $\mathcal{T}' := \{\alpha \mid \langle\alpha : \kappa\rangle \in \mathcal{T}\}$. $\mathcal{T}'$ contains the same axioms as $\mathcal{T}$, but ignores the contextual information encoded in their labels. Let now $H^u_{\mathcal{T}}$ be the (pruned) unraveled proof structure for $\mathcal{T}'$. By construction, $H^u_{\mathcal{T}}$ is a directed acyclic hypergraph. Our goal is to transform this hypergraph into a DAG and construct a BN, from which all the p-subsumption relations can be read through standard BN inferences. We explain this construction in two steps.

Algorithm 2 Construction of a DAG from a hypergraph

Input: $H = (V, E)$ directed acyclic hypergraph
Output: $G = (V', E')$ directed acyclic graph

$V' \leftarrow V$, $i, j \leftarrow 0$
for each $v \in V$ **do**
  $\mathfrak{S} \leftarrow \{S \mid (S, v) \in E\}$, $j \leftarrow i$
  for each $S \in \mathfrak{S}$ **do**
    $V' \leftarrow V' \cup \{\wedge_i\}$, $E' \leftarrow E' \cup \{(u, \wedge_i) \mid u \in S\}$
    if $i > j$ **then**
      $V' \leftarrow V' \cup \{\vee_i\}$, $E' \leftarrow E' \cup \{(\wedge_i, \vee_i)\}$
    $i \leftarrow i + 1$
  if $i = j + 1$ **then** ▷ If the GCI has only one explanation
    $E' \leftarrow E' \cup \{(\wedge_j, v)\}$
  else
    $E' \leftarrow E' \cup \{(\vee_k, \vee_{k+1}) \mid j < k < i - 1\} \cup \{(\vee_{i-1}, v), (\wedge_j, \vee_{j+1})\}$
return $G = (V', E')$

From Hypergraph to DAG Hypergraphs generalize graphs by allowing edges to connect many vertices. These hyperedges can be seen as an encoding of a formula in disjunctive normal form. An edge (S, v) expresses that if all the elements in S can be reached, then v is also reachable; we see this as an implication: $\bigwedge_{w \in S} w \Rightarrow v$. Several edges sharing the same head $(S_1, v), (S_2, v), \ldots, (S_k, v)$ in the hypergraph can be described through the implication $\bigvee_{i=1}^{k} (\bigwedge_{w \in S_i} w) \Rightarrow v$. We can thus rewrite any directed acyclic hypergraph into a DAG by introducing auxiliary conjunctive and disjunctive nodes (see Figure 3); the proper semantics of these nodes will be guaranteed by the conditional probability distribution defined later. Since the space needed for describing the conditional probability tables in a BN is exponential on the number of parents of the node, we ensure that all the nodes in this DAG have at most two parent nodes.

Algorithm 2 constructs such a DAG from a directed hypergraph. The algorithm adds a new node $\wedge_i$ for each hyperedge (S, v) in the input hypergraph H, and connects it with all the nodes in S. If there are k hyperedges that lead to a single node v, it creates $k - 1$ nodes $\vee_i$. These are used to represent the binary disjunctions among all the hyperedges leading to v. The algorithm runs in polynomial time on the size of H, and if H is acyclic, the resulting graph G is acyclic too. Moreover, all the nodes $v \in V$ that existed in the input hypergraph have at most one parent node after the translation; every $\vee_i$ node has exactly two parents, and the number of parents of a node $\wedge_i$ is given by the set S from the hyperedge $(S, v) \in E$ that generated it. In particular, if the input hypergraph is the unraveled proof structure for a TBox $\mathcal{T}$, then the size of the generated graph G is polynomial on the size of $\mathcal{T}$, and each node has at most two parent nodes.

From DAG to BN The next step is to build a BN that preserves the probabilistic entailments of a $\mathcal{BEL}$ KB. Let $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ be such a KB, with $\mathcal{B} = (G, \Phi)$, and let $G_{\mathcal{T}}$ be the DAG obtained from the unraveled proof structure of $\mathcal{T}$ using

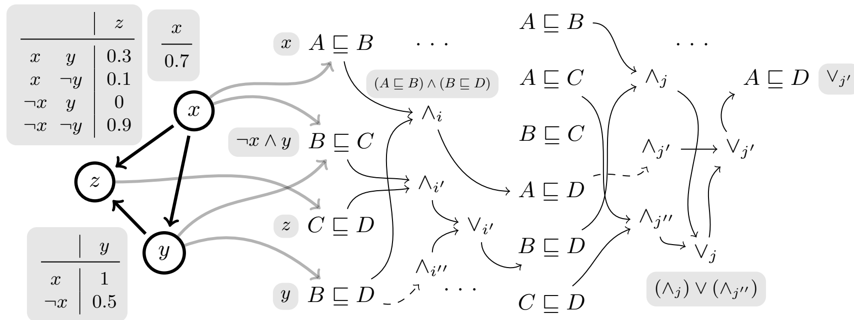

Fig. 3: A portion of the constructed BN

Algorithm 2. Recall that the nodes of $G_{\mathcal{T}}$ are either (i) pairs of the form (α, i), where α is a GCI in normal form built from the signature of $\mathcal{T}$, or (ii) an auxiliary disjunction ($\vee_i$) or conjunction ($\wedge_i$) node introduced by Algorithm 2. Moreover, $(\alpha, 0)$ is a node of $G_{\mathcal{T}}$ iff there is a context κ with $\langle \alpha : \kappa \rangle \in \mathcal{T}$. We assume w.l.o.g. that for node $(\alpha, 0)$ there is *exactly one* such context. If there were more than one, then we could extend the BN $\mathcal{B}$ with an additional variable which describes the disjunctions of these contexts, similarly to the construction of Algorithm 2. Similarly, we assume w.l.o.g. that each context κ appearing in $\mathcal{T}$ contains at most two literals, which is a restriction that can be easily removed by introducing auxiliary nodes as before. For a context κ, let $\mathsf{var}(\kappa)$ denote the set of all variables appearing in κ. We construct a new BN $\mathcal{B}_{\mathcal{K}}$ as follows.

Let $G = (V, E)$ and $G_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$. Construct the graph $G_{\mathcal{K}} = (V_{\mathcal{K}}, E_{\mathcal{K}})$, where $V_{\mathcal{K}} := V \cup V_{\mathcal{T}}$ and $E_{\mathcal{K}} := E \cup E_{\mathcal{T}} \cup \{(x, (\alpha, 0)) \mid \langle \alpha : \kappa \rangle \in \mathcal{T}, x \in \mathsf{var}(\kappa)\}$. Clearly, $G_{\mathcal{K}}$ is a DAG. We now need only to define the conditional probability tables for the nodes in $V_{\mathcal{T}}$ given their parents in $G_{\mathcal{K}}$; notice that the structure of the graph G remains unchanged for the construction of $G_{\mathcal{K}}$. For every node $(\alpha, 0) \in V_{\mathcal{T}}$, there is a κ such that $\langle \alpha : \kappa \rangle \in \mathcal{T}$; the parents of $(\alpha, 0)$ in $G_{\mathcal{K}}$ are then $\mathsf{var}(\kappa) \subseteq V$. The conditional probability of $(\alpha, 0)$ given its parents is defined, for every valuation $\mathcal{V}$ of $\mathsf{var}(\kappa)$ as $P_{\mathcal{B}}((\alpha, 0) = \mathsf{true} \mid \mathcal{V}) = \mathcal{V}(\kappa)$; that is, the probability of $(\alpha, 0)$ being true given a valuation of its parents is 1 if the valuation makes the context κ true; otherwise, it is 0. Each auxiliary node has at most two parents. The conditional probability of a conjunction node $\wedge_i$ being true is 1 iff all parents are true, and the conditional probability of a disjunction node $\vee_i$ being true is 1 iff at least one parent is true. Finally, every (α, i) with $i > 0$ has exactly one parent node v; (α, i) is true with probability 1 iff v is true.

Example 16. Consider the $\mathcal{BEL}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{B}_0)$ over $V = \{x, y, z\}$ where

$$\mathcal{T} = \{\langle A \sqsubseteq B : \{x\}\rangle, \langle B \sqsubseteq C : \{\neg x, y\}\rangle, \langle C \sqsubseteq D : \{z\}\rangle, \langle B \sqsubseteq D : \{y\}\rangle\}.$$

The BN obtained from this KB is depicted in Figure 3. The DAG obtained from the unraveled proof structure of $\mathcal{T}$ appears on the right, while the left part shows the original BN $\mathcal{B}_0$. The gray arrows depict the connection between these two

DAGs, which is given by the labels in the V-GCIs in $\mathcal{T}$. The gray boxes denote the conditional probability of the different nodes given their parents.

Suppose that we are interested in $P(\langle A \sqsubseteq_{\mathcal{K}} D : \emptyset \rangle)$. From the unraveled proof structure, we can see that $A \sqsubseteq D$ can be deduced either using the axioms $A \sqsubseteq B$, $B \sqsubseteq C$, $C \sqsubseteq D$, or through the two axioms $A \sqsubseteq B$, $B \sqsubseteq D$. The probability of any of these combinations of axioms to appear is given by $\mathcal{B}_0$ and the contextual connection to the axioms at the lower level of the proof structure. Thus, to deduce $P(\langle A \sqsubseteq_{\mathcal{K}} D : \emptyset \rangle)$ we need only to compute the probability of the node $(A \sqsubseteq D, n)$, where n is the last level.

From the properties of proof structures and Theorem 3 we have that

$$P_{\mathcal{B}_{\mathcal{K}}}((\alpha, n) \mid \kappa) = \sum_{\mathcal{V}(\kappa)=1} P_{\mathcal{B}_{\mathcal{K}}}((\alpha, n) \mid \mathcal{V}) P_{\mathcal{B}_{\mathcal{K}}}(\mathcal{V}) = \sum_{\substack{\mathcal{T}_{\mathcal{W}} \models \alpha \\ \mathcal{W}(\kappa)=1}} P_{\mathcal{B}_{\mathcal{K}}}(\mathcal{W}).$$

which yields the following result.

Theorem 17. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ be a $\mathcal{BEL}$ KB, C, D two $\mathcal{BEL}$ concepts, κ a V-context and $n = |\mathsf{sig}(\mathcal{T})|^3$. For a V-GCI $\langle C \sqsubseteq D : \kappa \rangle$, the following holds: $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle) = 1 - P_{\mathcal{B}}(\kappa) + P_{\mathcal{B}_{\mathcal{K}}}((C \sqsubseteq D, n) \mid \kappa)$.*

This theorem states that we can reduce the problem of p-subsumption w.r.t. the $\mathcal{BEL}$ KB $\mathcal{K}$ to a probabilistic inference in the BN $\mathcal{B}_{\mathcal{K}}$. Notice that the size of $\mathcal{B}_{\mathcal{K}}$ is polynomial on the size of $\mathcal{K}$. This means that p-subsumption is at most as hard as deciding D-PR problems over the BN $\mathcal{B}_{\mathcal{K}}$ which is in PP [22]. Since p-subsumption is also PP-hard [7], we get the following.

Theorem 18. *Deciding p-subsumption is* PP*-complete in the size of the KB.*

3.3 Most Likely Context

Finding the most likely context for a consequence can be seen as the dual of computing the probability of this consequence. Intuitively, we are interested in finding the most likely explanation for an event; if a consequence holds, we want to find the context for which this consequence is most likely to occur.

Definition 19 (most likely context). *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB, C, D two $\mathcal{BEL}$ concepts. A V-context κ is a* most likely context *(mlc) for $C \sqsubseteq D$ if (i) $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle$ and (ii) for all contexts κ' with $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa' \rangle$, $P_{\mathcal{B}}(\kappa) \geq P_{\mathcal{B}}(\kappa')$.*

Computing all most likely contexts can be done in exponential time. Moreover, it is not possible to lower this bound since a GCI may have exponentially many mlcs. Here we are interested in finding one most likely context, or more precisely, on its associated decision problem: given a context κ, decide whether κ is an mlc for $C \sqsubseteq D$ w.r.t. $\mathcal{K}$. This problem is clearly in coNP^{PP}: to show that κ is not a mlc, we can guess a V-context κ', and check with a PP oracle that $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa' \rangle$ and $P_{\mathcal{B}}(\kappa') > p$ hold, using the construction from Section 3.2.

To show that it is also coNP$^{\text{PP}}$-hard, we provide a reduction from D-MAP, which corresponds to finding a valuation that maximizes the probability of an event. Formally, the D-MAP problem consists of deciding, given a BN $\mathcal{B}$ over V, a set $Q \subseteq V$ a V-context κ, and $p > 0$, whether there exists a valuation λ of the variables in Q such that $P_{\mathcal{B}}(\kappa \cup \lambda) > p$.

Let $\mathcal{B} = ((V, E), \Phi)$ be a BN, κ a V-context, $Q = \{x_1, \ldots, x_k\} \subseteq V$, and $p > 0$. Define $V' = V \uplus \{x^+, x^- \mid x \in Q\} \uplus \{z\}$, where $\uplus$ denotes the disjoint union, and $E' = E \cup \{(x, x^+), (x, x^-) \mid x \in Q\}$. We construct $\mathcal{B}' = ((V', E'), \Phi')$ where Φ' contains $P_{\mathcal{B}'}(v \mid \pi(v)) = P_{\mathcal{B}}(v \mid \pi(x))$ for all $v \in V$, and $P_{\mathcal{B}'}(z) = p$, $P_{\mathcal{B}'}(x^+ \mid x) = 1$, $P_{\mathcal{B}'}(x^+ \mid \neg x) = 0$, $P_{\mathcal{B}'}(x^- \mid x) = 0$, and $P_{\mathcal{B}'}(x^- \mid \neg x) = 1$ for all $x \in Q$. Let now

$$\begin{aligned}\mathcal{T} = \{&\langle A_{i-1} \sqsubseteq A_i : x_i^+ \rangle, \langle A_{i-1} \sqsubseteq A_i : x_i^- \rangle \mid 1 \leq i \leq k\} \cup \\ &\{\langle A_k \sqsubseteq B : \kappa \rangle, \langle A_0 \sqsubseteq B : z \rangle\},\end{aligned}$$

and $\mathcal{K} = (\mathcal{B}', \mathcal{T})$. It is easy to see that for any V'-context κ', if $\langle A_0 \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$ and $z \notin \kappa'$, then $\kappa \subseteq \kappa'$ and for every $x \in Q, \{x^+, x^-\} \cap \kappa' \neq \emptyset$. Moreover, by construction $P_{\mathcal{B}}(z) = p$ and $P_{\mathcal{B}}(x^+, x^-) = 0$ for all $x \in Q$.

Theorem 20. *Let $\mathcal{B}$ be a BN over V, κ a V-context, $Q \subseteq V$, $p > 0$ and $\mathcal{K}$ the KB built as described above. There is a valuation λ of the variables in Q such that $P_{\mathcal{B}}(\lambda \cup \kappa) > p$ iff $\{z\}$ is not an mlc for $A_0 \sqsubseteq B$ w.r.t. $\mathcal{K}$.*

From this theorem, and the upper bound described above, we obtain a tight complexity bound for deciding a most likely context.

Corollary 21. *Deciding whether κ is a most likely context is co*NP$^{\text{PP}}$*-complete.*

If the context κ is a complete valuation, then the complexity of this problem reduces to NP-complete. This is an immediate result of applying the standard chain rule for exact inference, which is in PTime, and reducing the *most probable explanation* (D-MPE) problem in BNs, which is NP-complete [23].

4 Related Work

The amount of work combining DLs with probabilities is too vast to enumerate here. We mention only the work that relates the closest to our approach, and refer the interested reader to a thorough, although slightly outdated survey [17].

An early attempt for combining BNs and DLs was P-Classic [16], which extends Classic through probability distributions over the interpretation domain. In the same line, in PR-OWL [10] the probabilistic component is interpreted by providing individuals with a probability distribution. As many others in the literature, these approaches differ from our multiple-world semantics, in which we consider a probability distribution over a set of classical DL interpretations. Other probabilistic extensions of $\mathcal{EL}$ are [18] and [19]. The former introduces probabilities as a concept constructor, while in the latter the probabilities of axioms, which are always assumed to be independent, are implicitly encoded through a weighting function, which is interpreted with a log-linear model. Thus, both formalisms differ greatly from our approach.

DISPONTE [21] considers a multiple-world semantics. The main difference with our approach is that in DISPONTE, all probabilities are assumed to be independent, while we provide a joint probability distribution through the BN. Another minor difference is that $\mathcal{BEL}$ allows for classical consequences whereas DISPONTE does not. Closest to our approach is perhaps the Bayesian extension of DL-Lite called BDL-Lite [11]. Abstracting from the different logical component, BDL-Lite looks almost identical to $\mathcal{BEL}$. There is, however, a subtle but important difference. In our approach, an interpretation $\mathcal{I}$ satisfies a V-GCI $\langle C \sqsubseteq D : \kappa \rangle$ if $\mathcal{V}^{\mathcal{I}}(\kappa) = 1$ implies $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. In [11], the authors employ a closed-world assumption over the contexts, where this implication is substituted for an equivalence; i.e., $\mathcal{V}^{\mathcal{I}}(\kappa) = 0$ also implies $C^{\mathcal{I}} \not\subseteq D^{\mathcal{I}}$. The use of such semantics can easily produce inconsistent KBs, which is impossible in $\mathcal{BEL}$.

5 Conclusions

We studied the probabilistic DL $\mathcal{BEL}$, which extends $\mathcal{EL}$ with uncertain contexts based on a BN. Given $\mathcal{BEL}$ KB $\mathcal{K}$, we construct in polynomial time a BN $\mathcal{B}_{\mathcal{K}}$ that encodes all the probabilistic and logical knowledge of $\mathcal{K}$ w.r.t. the signature of the KB. This construction is based on the proof structure, a hypergraph representation of all the traces of any consequence derivation. As a result, we obtain that (i) deciding p-subsumption in $\mathcal{BEL}$ can be reduced to exact inference in $\mathcal{B}_{\mathcal{K}}$ and (ii) one most likely context can be found by computing a valuation of a subset of the variables in $\mathcal{B}_{\mathcal{K}}$ that maximizes the probability of an event. These provide tight complexity bounds for both of the reasoning problems.

While the construction is polynomial on the input KB, the obtained DAG might not preserve all the desired properties of the original BN. For instance, it is known that the efficiency of the BN inference engines depends on the treewidth of the underlying DAG [20]; however, the proof structure used by our construction may increase the treewidth of the graph. One direction of future research will be to try to optimize the reduction by bounding the treewidth and reducing the ammount of nodes added to the graph.

Finally, it should be clear that our construction does not depend on the chosen DL $\mathcal{EL}$, but rather on the fact that a simple polynomial-time consequence-based method can be used to reason with it. It should thus be a simple task to generalize the approach to other consequence-based methods, e.g. [24]. It would also be interesting to generalize the probabilistic component to consider other kinds of probabilistic graphical models [15].

References

1. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. IJCAI-05. Morgan-Kaufmann (2005)

2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2nd edn. (2007)
3. Baader, F., Peñaloza, R.: Automata-based axiom pinpointing. J. of Automated Reasoning 45(2), 91–129 (August 2010)
4. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic $\mathcal{EL}^+$. In: Proc. 30th German Conference on Artificial Intelligence (KI2007). LNAI, vol. 4667, pp. 52–67. Springer, Osnabrück, Germany (2007)
5. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In: Proc. ECAI-2004. pp. 298–302. IOS Press (2004)
6. Ceylan, İ.İ.: Context-Sensitive Bayesian Description Logics. Master's thesis, Dresden University of Technology, Germany (2013)
7. Ceylan, İ.İ., Peñaloza, R.: The Bayesian Description Logic $\mathcal{BEL}$. In: Demri, S., Kapur, D., Weidenbach, C. (eds.) Proceedings of the 7th International Joint Conference on Automated Reasoning (IJCAR'14). Lecture Notes in Computer Science, vol. 8562, pp. 480–494. Springer International Publishing (2014)
8. Ceylan, İ.İ., Peñaloza, R.: Bayesian Description Logics. In: Bienvenu, M., Ortiz, M., Rosati, R., Simkus, M. (eds.) Proceedings of the 27th International Workshop on Description Logics (DL'14). CEUR Workshop Proceedings, vol. 1193, pp. 447–458. CEUR-WS (2014)
9. Cooper, G.F.: The computational complexity of probabilistic inference using bayesian belief networks (research note). Artif. Intel. 42(2-3), 393–405 (Mar 1990)
10. da Costa, P.C.G., Laskey, K.B., Laskey, K.J.: Pr-owl: A bayesian ontology language for the semantic web. In: Uncertainty Reasoning for the Semantic Web I, URSW 2005-2007. LNCS, vol. 5327, pp. 88–107. Springer (2008)
11. d'Amato, C., Fanizzi, N., Lukasiewicz, T.: Tractable reasoning with bayesian description logics. In: Proc. Second International Conference on Scalable Uncertainty Management (SUM 2008). LNCS, vol. 5291, pp. 146–159. Springer (2008)
12. Darwiche, A.: Modeling and Reasoning with Bayesian Networks. Cambridge University Press (2009)
13. Homola, M., Serafini, L.: Contextualized knowledge repositories for the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 12(0) (2012)
14. Klarman, S., Gutirrez-Basulto, V.: $\mathcal{ALC}_{\mathcal{ALC}}$: A context description logic. In: JELIA. vol. 6341, pp. 208–220 (2010)
15. Koller, D., Friedman, N.: Probabilistic Graphical Models - Principles and Techniques. MIT Press (2009)
16. Koller, D., Levy, A.Y., Pfeffer, A.: P-classic: A tractable probablistic description logic. In: Proc. 14th National Conference on Artificial Intelligence (AAAI-97). pp. 390–397. AAAI Press (1997)
17. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the semantic web. J. of Web Semantics 6(4), 291–308 (2008)
18. Lutz, C., Schröder, L.: Probabilistic description logics for subjective uncertainty. In: Lin, F., Sattler, U., Truszczynski, M. (eds.) KR. AAAI Press (2010)
19. Niepert, M., Noessner, J., Stuckenschmidt, H.: Log-linear description logics. In: Walsh, T. (ed.) IJCAI. pp. 2153–2158. IJCAI/AAAI (2011)
20. Pan, H., McMichael, D., Lendjel, M.: Inference algorithms in bayesian networks and the probanet system. Digital Signal Processing 8(4), 231 – 243 (1998)

21. Riguzzi, F., Bellodi, E., Lamma, E., Zese, R.: Epistemic and statistical probabilistic ontologies. In: Proc. 8th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW-12). vol. 900, pp. 3–14. CEUR-WS (2012)
22. Roth, D.: On the hardness of approximate reasoning. Artif. Intel. 82(1-2), 273–302 (1996)
23. Shimony, E.S.: Finding maps for belief networks is np-hard. Artif. Intell. 68(2), 399–410 (Aug 1994)
24. Simancik, F., Kazakov, Y., Horrocks, I.: Consequence-based reasoning beyond horn ontologies. In: Proc. IJCAI-11. pp. 1093–1098. IJCAI/AAAI (2011)