Reasoning in the Description Logic $\mathcal{BEL}$ using Bayesian Networks

İsmail İlkan Ceylan*
Theoretical Computer Science
TU Dresden, Germany
ceylan@tcs.inf.tu-dresden.de

Rafael Peñaloza†
Theoretical Computer Science, TU Dresden, Germany
Center for Advancing Electronics Dresden
penaloza@tcs.inf.tu-dresden.de

Abstract

We study the problem of reasoning in the probabilistic Description Logic $\mathcal{BEL}$. Using a novel structure, we show that probabilistic reasoning in this logic can be reduced in polynomial time to standard inferences over a Bayesian network. This reduction provides tight complexity bounds for probabilistic reasoning in $\mathcal{BEL}$.

1 Introduction

Description Logics (DLs) (Baader et al. 2007) are a family of knowledge representation formalisms tailored towards the representation of terminological knowledge in a formal manner. In their classical form, DLs are unable to handle the inherent uncertainty of many application domains. To overcome this issue, several probabilistic extensions of DLs have been proposed. The choice of a specific probabilistic DL over others depends on the intended application; these logics differ in their logical expressivity, their semantics, and their independence assumptions.

Recently, the DL $\mathcal{BEL}$ (Ceylan and Peñaloza 2014) was introduced as a means of describing certain knowledge that depends on an uncertain context, expressed by a Bayesian network (BN). An interesting property of this logic is that reasoning can be decoupled between the logical part and the 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.

In this paper 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. In a nutshell, a proof structure describes the class of contexts that entail the wanted consequence. A BN can be constructed 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 $\mathcal{BEL}$.

*Supported by DFG in the Research Training Group "RoSI" (GRK 1907).

†Partially supported by DFG within the Cluster of Excellence 'cfAED'.

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2 Proof Structures in $\mathcal{EL}$

$\mathcal{EL}$ 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. Formally, let $\mathsf{N_C}$ and $\mathsf{N_R}$ be disjoint sets of *concept names* and *role names*, respectively. $\mathcal{EL}$ concepts are defined through the syntactic 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 in terms of an *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where $\Delta^\mathcal{I}$ is a non-empty *domain* and $\cdot^\mathcal{I}$ is an *interpretation function* that maps every concept name A to a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ and every role name r to a set of binary relations $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. 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 $(\exists r.C)^\mathcal{I} := \{d \in \Delta^\mathcal{I} \mid \exists e \in \Delta^\mathcal{I} : (d,e) \in r^\mathcal{I} \wedge e \in C^\mathcal{I}\}$.

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 subsumption checking, i.e., deciding the sub-concept relations between given 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 $\mathcal{EL}$ in polynomial time by a completion algorithm (Baader, Brandt, and Lutz 2005). This algorithm requires the TBox to be in *normal form*; i.e., where all GCIs are 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 (Brandt 2004; Baader, Brandt, and Lutz 2005); for the rest of this paper, we assume that $\mathcal{T}$ is a TBox in normal form and GCI denotes a normalized subsumption relation.

In this paper, we are interested in deriving the GCIs in normal form that follow from $\mathcal{T}$; i.e. the *normalised logical closure* of $\mathcal{T}$. We introduce the *deduction rules* shown in Table 1 to produce the normalised logical closure of a TBox.

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

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

Algorithm 1 Construction of the pruned proof structure

Input: TBox $\mathcal{T}$
Output: $H = (W, F)$

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

Each rule maps a set of premises to a GCI that is implicitly encoded in the premises. It is easy to see that the sets of premises cover all pairwise combinations of GCIs in normal form and that the deduction rules produce the normalised logical closure of a TBox. Moreover, the given deduction rules introduce axioms only in normal form, and do not create any new concept or role name. Hence, if $n = |\mathsf{sig}(\mathcal{T})|$, the logical closure of $\mathcal{T}$ is computed after n^3 rule applications, at most.

Later on we will associate a probability to the GCIs in the TBox $\mathcal{T}$, and will be interested in computing the probability of a consequence. It will then be useful to be able not only to deduce the GCI, but also *all* the sub-TBoxes of $\mathcal{T}$ from which this GCI follows. Therefore, we store the traces of the deduction rules using a directed hypergraph.

Definition 2. 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 $v \in V$. A *path from S to v in H* is a sequence of hyperedges $(S_1, v_1), (S_2, v_2), \ldots, (S_n, v_n)$ such that $v_n = v$ and $S_i \subseteq S \cup \{v_j \mid 0 < j < i\}$ for every $i, 1 \leq i \leq n$. In this case, the path has *length* 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 that follow from $\mathcal{T}$ and $E_\mathcal{T} = \{(S, \alpha) \mid S \mapsto \alpha, S \subseteq V_\mathcal{T}\}$ where $\mapsto$ is the deduction relation defined in Table 1. We call this hypergraph the *proof structure* of $\mathcal{T}$. The following lemma follows from the correctness of the deduction rules.

Lemma 3. *Let $\mathcal{T}$ be a TBox in normal form, its proof structure $H_\mathcal{T} = (V_\mathcal{T}, E_\mathcal{T})$, $\mathcal{O} \subseteq \mathcal{T}$, and $C \sqsubseteq D \in V_\mathcal{T}$. There is a path from $\mathcal{O}$ to $C \sqsubseteq D$ in $H_\mathcal{T}$ iff $\mathcal{O} \models C \sqsubseteq D$.*

Intuitively, $H_\mathcal{T}$ is a compact representation of all the possible ways in which a GCI can be derived from the GCIs present in $\mathcal{T}$. Traversing this hypergraph backwards, from a GCI α being entailed by $\mathcal{T}$, it is possible to construct all proofs for α; hence the name "proof structure." As mentioned before, $|V_\mathcal{T}| \leq |\mathsf{sig}(\mathcal{T})|^3$; thus, it is enough 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 construct an *unfolded* version of this hypergraph by making different copies of each node in each level in order to avoid cycles. In this case, nodes are pairs of GCIs and labels, where the latter indicates to which level the nodes belong in the hypergraph. We write $S^i = \{(\alpha, i) \mid \alpha \in S\}$ to denote the i-labeled set of GCIs in S. Let $n := |\mathsf{sig}(\mathcal{T})|^3$, we start with the set $W_0 := \{(\alpha, 0) \mid \alpha \in \mathcal{T}\}$ and define the levels $0 \leq i < n$ inductively by

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

For each $i, 0 \leq i \leq n$, W_i contains all the consequences 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 $F_\mathcal{T} := \bigcup_{i=1}^n F_i$, with

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

The following is a simple consequence of our constructions and Lemma 3.

Theorem 4. *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 $\mathcal{T}$. 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; hence we obtain the following.

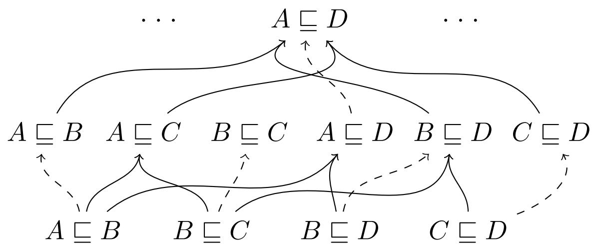

Figure 1: The first levels of an unfolded proof structure and the paths to $\langle A \sqsubseteq D \rangle$

Lemma 5. *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 6. Consider the following $\mathcal{EL}$ 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.[1] The first level V_0 of this hypergraph contains a representative for each GCI 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 ones (represented by a dashed line in Figure 1). Afterwards, we apply all possible deduction rules to the elements of V_0, and add a hyperedge from the premises at level V_0 to the conclusion at level V_1 (continuous lines). The same 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 briefly recall $\mathcal{BEL}$, a probabilistic extension of $\mathcal{EL}$ based on Bayesian networks (Ceylan and Peñaloza 2014), and use the construction of the (unfolded) proof structure to reduce reasoning in this logic, to standard Bayesian network inferences.

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 probability. To handle the joint probability distribution of the GCIs, these probabilities are encoded in a Bayesian network (Darwiche 2009). Formally, 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,[2] 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.

BNs encode a series of conditional independence assumptions between the random variables; more precisely, 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 (JPD) over V given by

$$P_{\mathcal{B}}(V) = \prod_{x \in V} P_{\mathcal{B}}(x \mid \pi(x)).$$

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 GCIs are annotated with a context, defined by a set of literals belonging to a BN.

Definition 7 (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 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.

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 $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.[3]

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 in the context 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 V-GCI.

Example 8. Let $V_0 = \{x, y, z\}$, and consider the V_0-TBox

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

The interpretation $\mathcal{I}_0 = (\{d\}, \cdot^{\mathcal{I}_0}, \mathcal{V}_0)$ given by $\mathcal{V}_0(\{x, \neg y, z\}) = 1$, $A^{\mathcal{I}_0} = \{d\}$, 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-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 (Ceylan 2013). Thus, we consider only $\mathcal{BEL}$ KBs in normal form in the following.

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

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

[3]When there is no danger of ambiguity, we will usually drop the prefix V and speak simply of e.g. a *TBox*, a *KB* or an *interpretation*.

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* $\langle C \sqsubseteq D : \emptyset \rangle$, denoted by $\mathcal{T} \models C \sqsubseteq D$, 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 can define a 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 9 (restriction)**.** Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB. The *restriction of* $\mathcal{T}$ to a valuation $\mathcal{W}$ of the variables in V is

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

To handle the probabilistic knowledge provided by the BN, we extend the semantics of $\mathcal{BEL}$ through multiple-world interpretations. Intuitively, a V-interpretation describes 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 10 (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 $\mathcal{I} \in \mathfrak{I}$. This probabilistic interpretation 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

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

It is a *model* of the KB $(\mathcal{B}, \mathcal{T})$ iff it is a (probabilistic) 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.

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? In the case of $\mathcal{BEL}$, we are also interested in finding the probability with which such a subsumption relation holds. For the rest of this section, we formally define this reasoning task, and provide a method for solving it, by reducing it to a decision problem in Bayesian networks.

3.1 Probabilistic Subsumption

Subsumption is one of the most basic decision problems in $\mathcal{EL}$. In $\mathcal{BEL}$, we generalize this problem to consider also the contexts and probabilities provided by the BN.

Definition 11 (p-subsumption)**.** Let C, D be two $\mathcal{BEL}$ concepts, κ a context, and $\mathcal{K}$ a $\mathcal{BEL}$ KB. 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* $\langle C \sqsubseteq D : \kappa \rangle$ *w.r.t.* $\mathcal{K}$ is defined as

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

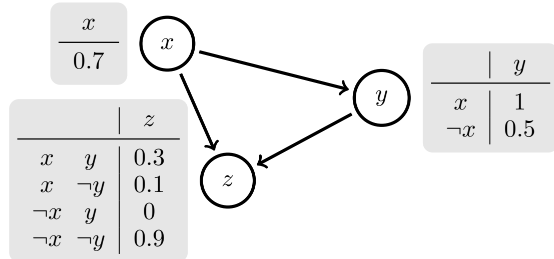

Figure 2: A simple BN

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

The following proposition was shown in (Ceylan and Peñaloza 2014).

Proposition 12. *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 13. 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 TBox from Example 8. It follows that $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \{x, y\} \rangle) = 1$ from the first 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 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 $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \emptyset \rangle) = 1$.

3.2 Deciding p-subsumption

We now show that deciding p-subsumption can be reduced to exact inference in Bayesian networks. This latter problem is known to be PP-complete (Roth 1996). Let $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ be an arbitrary but fixed $\mathcal{BEL}$ KB. From the V-TBox $\mathcal{T}$, we construct the classical $\mathcal{EL}$ TBox $\mathcal{T}' := \{\alpha \mid \langle \alpha : \kappa \rangle \in \mathcal{T}\}$; that is, $\mathcal{T}'$ 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 this TBox $\mathcal{T}'$. By construction, $H^u_\mathcal{T}$ is a directed acyclic hypergraph. Our goal is to transform this hypergraph into a DAG. Using this DAG, we will construct a BN, from which all the p-subsumption relations can be read through standard BN inferences. We explain this construction in two steps.

From Hypergraph to DAG Hypergraphs generalize graphs by allowing several vertices to be connected by a single edge. Intuitively, the hyperedges in a hypergraph encode a formula in disjunctive normal form. Indeed, an edge (S, v) expresses that if all the elements in S can be reached, then v is also reachable; this can be seen as an implication: $\bigwedge_{w \in S} w \Rightarrow v$. Suppose that there exist several edges sharing the same head $(S_1, v), (S_2, v), \ldots, (S_k, v)$ in the hypergraph. This situation 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 the upper part of Fig-

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**
    $E' \leftarrow E' \cup \{(\wedge_j, v)\}$
  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 $G = (V', E')$

ure 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 that a node has, we ensure that these auxiliary nodes, as well as the elements in $W_{\mathcal{T}}$ have at most two parent nodes.

Algorithm 2 describes the construction of such a DAG from a directed hypergraph. Essentially, 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. Additionally, 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. Clearly, the algorithm runs in polynomial time on the size of H, and if H is acyclic, then the resulting graph G is acyclic too. Moreover, all the nodes $v \in V$ that existed in the input hypergraph will 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 Bayesian Network 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 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 that $|\kappa| \leq 2$, to ensure that 0-level nodes have at most two parent nodes. This restriction can be easily removed by introducing conjunction 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 $G_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$. The DAG $G_{\mathcal{K}}$ is given by $G_{\mathcal{K}} = (V_{\mathcal{K}}, E_{\mathcal{K}})$, 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 given by: $P_{\mathcal{B}}((\alpha, 0) = \mathsf{true} \mid \mathcal{V}(\mathsf{var}(\kappa))) = \mathcal{V}(\kappa)$; 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, and 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 14. Consider the $\mathcal{BEL}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{B}_0)$, 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 upper part of the figure represents the DAG obtained from the unraveled proof structure of $\mathcal{T}$, while the lower 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 GCIs $A \sqsubseteq B$, $B \sqsubseteq C$, $C \sqsubseteq D$, or through the two GCIs $A \sqsubseteq B$, $B \sqsubseteq D$. The probability of any of these combinations of GCIs 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 4 we have that

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

which yields the following result.

Theorem 15. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{B})$ be a $\mathcal{BEL}$ KB, where $\mathcal{B}$ is over V, and $n = |\mathsf{sig}(\mathcal{T})|^3$. For a V-GCI $\langle \alpha : \kappa \rangle$, it holds that $P(\langle \alpha : \kappa \rangle) = 1 - P_{\mathcal{B}}(\kappa) + P_{\mathcal{B}_{\mathcal{K}}}((\alpha, n) \mid \kappa)$.*

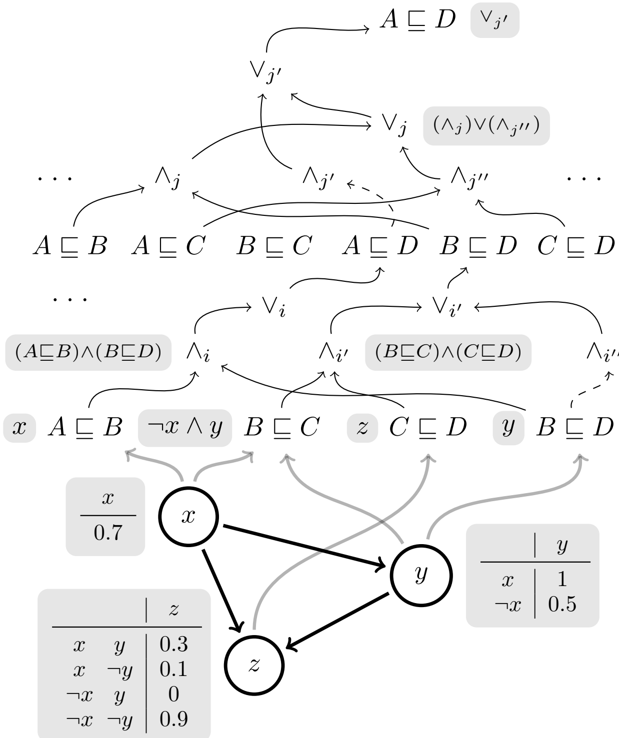

Figure 3: A portion of the constructed BN

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 the probability of query variables, given an evidence, which is known to be in PP (Roth 1996). Since p-subsumption is already PP-hard (Ceylan and Peñaloza 2014), we obtain the following result.

Corollary 16. *Deciding p-subsumption w.r.t. a $\mathcal{BEL}$ KB $\mathcal{K}$ is* PP*-complete on the size of $\mathcal{K}$.*

4 Related Work

An early attempt for combining BNs and DLs was P-CLASSIC (Koller, Levy, and Pfeffer 1997), which extends CLASSIC through probability distributions over the interpretation domain. In the same line, in PR-OWL (da Costa, Laskey, and Laskey 2008) the probabilistic component is interpreted by providing individuals with a probability distribution. As many others in the literature (see (Lukasiewicz and Straccia 2008) for a thorough survey on probabilistic DLs) these approaches differ from our multiple-world semantics, in which we consider a probability distribution over a set of classical DL interpretations.

DISPONTE (Riguzzi et al. 2012) is one representative of the approaches that consider a multiple-world semantics. The main difference with our approach is that in DISPONTE, the authors assume that all probabilities are 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 (d'Amato, Fanizzi, and Lukasiewicz 2008). Abstracting from the different logical component, BDL-Lite looks almost identical to ours. 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 (d'Amato, Fanizzi, and Lukasiewicz 2008), 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 $\mathcal{BEL}$.

Other probabilistic extensions of $\mathcal{EL}$ are (Lutz and Schröder 2010) and (Niepert, Noessner, and Stuckenschmidt 2011). 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.

5 Conclusions

We have described the probabilistic DL $\mathcal{BEL}$, which extends the light-weight DL $\mathcal{EL}$ with uncertain contexts. We have shown that it is possible to construct, from a given $\mathcal{BEL}$ KB $\mathcal{K}$, a BN $\mathcal{B}_\mathcal{K}$ that encodes all the probabilistic and logical knowledge of $\mathcal{K}$ w.r.t. to the signature of the KB. Moreover, the size of $\mathcal{B}_\mathcal{K}$ is polynomial on the size of $\mathcal{K}$. We obtain that probabilistic reasoning over $\mathcal{K}$ is at most as hard as deciding inferences in $\mathcal{B}_\mathcal{K}$, which yields a tight complexity bound for deciding p-subsumption in this logic.

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 (Pan, McMichael, and Lendjel 1998); 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.

Clearly, once we have constructed the associated BN $\mathcal{B}_\mathcal{K}$ from a given $\mathcal{BEL}$ KB $\mathcal{K}$, this can be used for additional inferences, beyond deciding subsumption. We think that reasoning tasks such as *contextual subsumption* and finding the *most likely context*, defined in (Ceylan and Peñaloza 2014) can be solved analogously. Studying this and other reasoning problems is also a task of future work.

Finally, 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. (Simancik, Kazakov, and Horrocks 2011). It would also be interesting to generalize the probabilistic component to consider other kinds of probabilistic graphical models (Koller and Friedman 2009).

References

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition.

Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *Proc. IJCAI-05*. Morgan-Kaufmann.

Brandt, S. 2004. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In *Proc. ECAI-2004*, 298–302. IOS Press.

Ceylan, İ. İ., and Peñaloza, R. 2014. The Bayesian Description Logic $\mathcal{BEL}$. In *Proc. IJCAR 2014*. To appear.

Ceylan, İ. İ. 2013. Context-sensitive bayesian description logics. Master's thesis, Dresden University of Technology, Germany.

da Costa, P. C. G.; Laskey, K. B.; and Laskey, K. J. 2008. Pr-owl: A bayesian ontology language for the semantic web. In *Uncertainty Reasoning for the Semantic Web I, URSW 2005-2007*, volume 5327 of *LNCS*, 88–107. Springer.

d'Amato, C.; Fanizzi, N.; and Lukasiewicz, T. 2008. Tractable reasoning with bayesian description logics. In *Proc. Second International Conference on Scalable Uncertainty Management (SUM 2008)*, volume 5291 of *LNCS*, 146–159. Springer.

Darwiche, A. 2009. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press.

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models - Principles and Techniques*. MIT Press.

Koller, D.; Levy, A. Y.; and Pfeffer, A. 1997. P-classic: A tractable probablistic description logic. In *Proc. 14th National Conference on Artificial Intelligence (AAAI-97)*, 390–397. AAAI Press.

Lukasiewicz, T., and Straccia, U. 2008. Managing uncertainty and vagueness in description logics for the semantic web. *J. of Web Semantics* 6(4):291–308.

Lutz, C., and Schröder, L. 2010. Probabilistic description logics for subjective uncertainty. In Lin, F.; Sattler, U.; and Truszczynski, M., eds., *KR*. AAAI Press.

Niepert, M.; Noessner, J.; and Stuckenschmidt, H. 2011. Log-linear description logics. In Walsh, T., ed., *IJCAI*, 2153–2158. IJCAI/AAAI.

Pan, H.; McMichael, D.; and Lendjel, M. 1998. Inference algorithms in bayesian networks and the probanet system. *Digital Signal Processing* 8(4):231 – 243.

Riguzzi, F.; Bellodi, E.; Lamma, E.; and Zese, R. 2012. Epistemic and statistical probabilistic ontologies. In *Proc. 8th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW-12)*, volume 900, 3–14. CEUR-WS.

Roth, D. 1996. On the hardness of approximate reasoning. *Artif. Intel.* 82(1-2):273–302.

Simancik, F.; Kazakov, Y.; and Horrocks, I. 2011. Consequence-based reasoning beyond horn ontologies. In *Proc. IJCAI-11*, 1093–1098. IJCAI/AAAI.