RC25157 (W1105-023) May 10, 2011
Computer Science

IBM Research Report

Trust Inference and Query Answering over Uncertain Information

Achille Fokoue, Mudhakar Srivatsa
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598 USA

Robert Young
Defense Science and Technology Labs.
United Kingdom

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

Trust Inference and Query Answering over Uncertain Information

Achille Fokoue[a], Mudhakar Srivatsa[a], Robert Young[b]

[a] *IBM T.J. Watson Research Center, USA*
[b] *Defence Science and Technology Labs, UK*

Abstract

On the Semantic Web, decision makers (humans or software agents alike) are faced with the challenge of examining large volumes of information originating from heterogeneous sources with the goal of ascertaining trust in various pieces of information. While previous work has focused on simple models for review and rating systems, we introduce a new trust model for rich, complex and uncertain information. We present the challenges raised by the new model, and solutions to support scalable trust-based query answering over uncertain information. An evaluation of the first prototype implementation under a variety of scenarios shows the robustness of our trust model, and the scalability of trust-based query answering over uncertain information.

1. Introduction

Decision makers (humans or software agents alike) relying on information available on the web are increasingly faced with the challenge of examining large volumes of information originating from heterogeneous sources with the goal of ascertaining trust in various pieces of information. Several authors have explored various trust computation models (e.g., eBay recommendation system [20], NetFlix movie ratings [18], EigenTrust [12], PeerTrust [21], etc.) to assess trust in various entities. A common data model subsumed by several trust computation models (as succinctly captured in Kuter and Golbeck [15]) is the ability of an entity to assign a *numeric* trust score to another entity (e.g., eBay recommendation, Netflix movie ratings, etc.). Such pair-wise numeric ratings contribute to a (dis)similarity score (e.g., based on $\mathcal{L}_1$ norm, $\mathcal{L}_2$ norm, cosine distance, etc.) which is used to compute personalized trust scores (as in PeerTrust) or recursively propagated throughout the network to compute global trust scores (as in EigenTrust).

A pair-wise numeric score based data model may impose severe limitations in several real-world applications. For example, let us suppose that information sources $\{S_1, S_2, S_3\}$ assert axioms $\phi_1 =$ *all men are mortal*, $\phi_2 =$ *Socrates is a man* and $\phi_3 =$ *Socrates is not mortal* respectively. While there is an obvious conflict when all the three axioms are put together, we note that: (i) there is no

pair-wise conflict, and (ii) there is no obvious numeric measure that captures (dis)similarity between two information sources.

This problem becomes even more challenging because of uncertainty associated with real-world data and applications. Uncertainty manifests itself in several diverse forms: from measurement errors (e.g., sensor readings) and stochasticity in physical processes (e.g., weather conditions) to reliability/trustworthiness of data sources; regardless of its nature, it is common to adopt a probabilistic measure for uncertainty. Reusing the *Socrates* example above, each information source S_i may assert the axiom ϕ_i with a certain probability $p_i = 0.6$. Further, probabilities associated with various axioms need not be (statistically) independent. In such situations, the key challenge is develop trust computation models for rich (beyond pair-wise numeric ratings) and uncertain (probabilistic) information.

The contributions of this paper are three fold. First, our approach offers a rich data model for trust. We allow information items to be encoded in inconsistency-tolerant extension of Bayesian Description Logics [2] (BDL) − a probabilistic extension of Description Logics, the foundation of OWL DL. The key idea behind trust inference is to leverage justifications of inconsistencies to compute trust scores for information sources; intuitively, an inconsistency corresponds to conflicts in information items reported by different information sources and the justification for an inconsistency traces back an inconsistency to a minimal set of information sources that are responsible for the conflict. This contribution builds on our previous work [7].

Second, our approach supports trust-based query answering over uncertain information. The key idea is to leverage inferred trust scores to refine a probabilistic database and support semantic query answering over the refined knowledge base. We show that past work on query answering on BDL [2] may sometimes result in counter-intuitive results; we extend past work by providing an intuitive query answering semantics for BDL and supporting inconsistency tolerant reasoning over a probabilistic knowledge base. To avoid the worst case exponential blow up typically associated with reasoning in most probabilistic extension of OWL DL, including BDL, we propose an error-bounded approximation algorithm for scalable probabilistic reasoning over large and very expressive knowledgebases.

Third, we have developed a prototype of our trust assessment and trust-based query answering system by implementing a probabilistic extension, PSHER, to our publicly available highly scalable DL reasoner SHER [6]. We empirically evaluate the efficacy of our scheme (on a publicly available UOBM dataset) when malicious sources use an oscillating behavior to milk the trust computation model and when honest sources are faced with measurement errors (high uncertainty) or commit honest mistakes. We also evaluate the scalability of query answering over large, expressive and probabilistic knowledge bases.

Figure 1 describes our architecture for trust inference and query answering over uncertain information. The remainder of the paper is organized as follows: After a brief introduction of Bayesian Description Logics (BDL) in Section 2, Section 3 describes an inconsistency-tolerant extension of BDL and presents

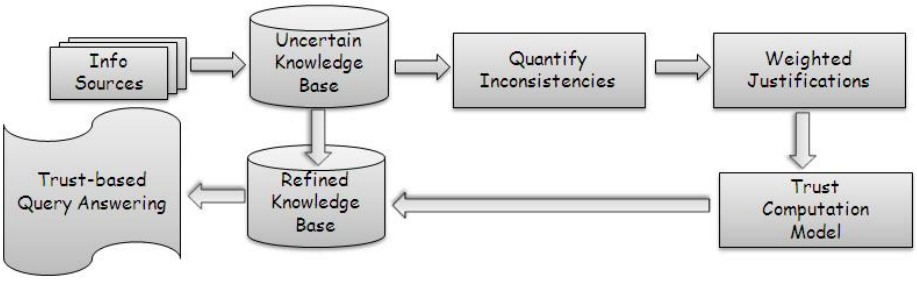

Figure 1: Trust Inference and Query Answering over Uncertain Information

solutions to effectively compute justifications (a proxy for (dis)similarity scores in our trust computation model). Section 4 describes our solutions for trust-based query answering. Section 5 presents an experimental evaluation of our system. We finally conclude in Section 7.

2. Background

In this section, we briefly describe our data model for uncertain information.

2.1. Bayesian Network Notation

A Bayesian Network [19] is a well-known probabilistic graphic model that encodes in a directed acyclic graph probabilistic dependencies between random variables. We briefly recall notations for a Bayesian Network, used in the remainder of the paper.

V: set of all random variables in a Bayesian network (e.g., $V = \{V_1, V_2\}$). $D(V_i)$ (for some variable $V_i \in V$): set of values that V_i can take (e.g., $D(V_1) = \{0, 1\}$ and $D(V_2) = \{0, 1\}$). v: assignment of all random variables to a possible value (e.g., $v = \{V_1 = 0, V_2 = 1\}$). $v|X$ (for some $X \subseteq V$): projection of v that only includes the random variables in X (e.g., $v|\{V_2\} = \{V_2 = 1\}$). $D(X)$ (for some $X \subseteq V$): Cartesian product of the domains $D(X_i)$ for all $X_i \in X$.

2.2. Bayesian Description Logics

Bayesian Description Logics [2] is a class of probabilistic description logic wherein each logical axiom is annotated with an event which is associated with a probability value via a Bayesian Network. In this section, we describe Bayesian DL at a syntactic level followed by a detailed example.

A probabilistic axiom over a Bayesian Network BN over a set of variables V is of the form $\phi : e$, where ϕ is a classical DL axiom, and the probabilistic annotation e is an expression of one of the following forms: $X = x$ or $X \neq x$ where $X \subseteq V$ and $x \in D(X)$. Intuitively, every probabilistic annotation represents a scenario (or an event) which is associated with the set of all value assignments $V = v$ with $v \in D(V)$ that are compatible with $X = x$ (resp. $X \neq$

x) (that is, $v|X = x$, resp. $v|X \neq x$), and their probability value $Pr_{BN}(V = v)$ in the Bayesian network BN over V. Simply put, the semantics of a probabilistic axiom $\phi : X = x$ is as follows: when event $X = x$ occurs then ϕ holds. $\phi : p$, where $p \in [0, 1]$, is often used to directly assign a probability value to an classical axiom ϕ. This is an abbreviation for $\phi : X_0 = true$, where X_0 is a boolean random variable which is independent from all other variables and such that $Pr_{BN}(X_0 = true) = p$. We abbreviate the probabilistic axiom of the form $\top : e$ (resp. $\phi : \top$) as e (resp. ϕ).

A probabilistic knowledge base (KB) $K = (\mathcal{A}, \mathcal{T}, BN)$ consists of : 1) a Bayesian Network BN over a set of random variables V, 2) a set of probabilistic Abox axioms $\mathcal{A}$ of the form $\phi : e$, where ϕ is a classical Abox axiom, and 3) a set of probabilistic Tbox axioms $\mathcal{T}$ of the form $\phi : e$, where ϕ is a classical Tbox axiom.

The following example illustrates how this formalism can be used to describe road conditions influenced by probabilistic events such as weather conditions:

$$\mathcal{T} = \{$$
$$SlipperyRoad \sqcap OpenedRoad \sqsubseteq HazardousCondition,$$
$$Road \sqsubseteq SlipperyRoad : Rain = true\}$$
$$\mathcal{A} = \{Road(route9A),$$
$$OpenedRoad(route9A) : TrustSource = true\}$$

In this example, the Bayesian network BN consists of three variables: $Rain$, a boolean variable which is true when it rains; $TrustSource$, a boolean variable which is true when the source of the axiom $OpenedRoad(route9A)$ can be trusted; and $Source$, a variable which indicates the provenance of the axiom $OpenedRoad(route9A)$. The probabilities specified by BN are as follows:

$$Pr_{BN}(TrustSource = true|Source = `Mary') = 0.8$$
$$Pr_{BN}(Rain = true) = 0.7$$
$$Pr_{BN}(TrustSource = true|Source = `John') = 0.5$$
$$Pr_{BN}(Source = `John') = 1$$

The first Tbox axiom asserts that a opened road that is slippery is a hazardous condition. The second Tbox axiom indicates that when it rains, roads are slippery. The Abox axioms assert that $route9A$ is a road and, assuming that the source of the statement $OpenedRoad(route9A)$ is trusted, $route9A$ is opened.

Informally, probability values computed through the Bayesian network 'propagate' to the 'DL side' as follows. Each assignment v of all random variables in BN (e.g.,v = {$Rain$ = true, $TrustSource$ = false, $Source$= 'John'}) corresponds to a primitive event ev (or a scenario). A primitive event ev is associated, through BN, to a probability value p_{ev} and a classical DL KB K_{ev} which consists of all classical axioms annotated with a compatible probabilistic annotation (e.g., $SlipperyRoad \sqcap OpenedRoad \sqsubseteq HazardousCondition, Road \sqsubseteq SlipperyRoad, Road(route9A)$). The probability value associated with the

statement ϕ (e.g., $\phi = HazardousCondition(route9A)$) is obtained by summing p_{ev} for all ev such that the classical KB K_{ev} entails ϕ
(e.g., $Pr(HazardousCondition(route9A)) = 0.35$).

3. Trust Inference

In this section we describe trust inference over uncertain information (encoded as BDL axioms).

3.1. Overview

A key element in our approach to inference trust in uncertain information is the ability to handle inconsistencies that often naturally arise when information from multiple sources is aggregated. Intuitively, our approach allows us to compute a degree of consistency ($0 \leq d \leq 1$) over an uncertain (probabilistic) knowledge base. We note that inconsistencies correspond to conflicts in information items reported by one or more information sources. Our approach assigns numeric weights to the degree of inconsistency using the *possible world* semantics. Revisiting the *Socrates* example (from Section 1), three probabilistic axioms $\phi_i : p_i$ correspond to eight possible worlds (the power set of the set of axioms without annotations), namely, $\{\{\phi_1, \phi_2, \phi_3\}, \{\phi_1, \phi_2\}, \cdots, \emptyset\}$. For instance, the possible world $\{\phi_1, \phi_2\}$ corresponds to a world wherein *all men are mortal*, and *Socrates is a man*. Each possible world has probability measure that can be derived from p_i (in general the probability measure can be computed as joint probability distributions over the random variables in the Bayesian Network). For instance, the probability of a possible world $\{\phi_1, \phi_2\}$ is given by $p_1 * p_2 * (1 - p_3) = 0.6 * 0.6 * (1 - 0.6) = 0.144$. Indeed, we observe that the world $\{\phi_1, \phi_2, \phi_3\}$ is inconsistent, while the remaining seven possible worlds are consistent. This allows us to compute the degree of inconsistency of a knowledge base as the sum of the probabilities associated with possible worlds that are inconsistent.

In the presence of inconsistencies, our approach extracts justifications – minimal sets of axioms that together imply an inconsistency [11]. Our trust computation model essentially propagates the degree of inconsistency as blames (or penalties) to the axioms contributing to the inconsistency via justifications. This approach essentially allows us to compute trust in information at the granularity of an axiom. Indeed one may aggregate trust scores at different levels of granularity; e.g., axioms about a specific topic (e.g., birds), one information source (e.g., John), groups of information sources (e.g., all members affiliated with ACM), etc.

Intuitively, our trust computation model works as follows. First, we compute a probability measure for each justification as the sum of the probabilities associated with possible worlds in which the justification holds (namely, all the axioms in the justification are present). Justifications play the critical role of a proxy to the measure of dissimilarity between information sources, which is, as mentioned in section 1, a key element of most trust computation models.

Second, we partition the degree of inconsistency across all justifications; for instance, if a justification J_1 holds in 80% of the possible worlds then it is assigned four times the blame as a justification J_2 that holds in 20% of the possible worlds. Third, we partition the penalty associated with a justification across all axioms in the justification using a biased (on prior trust assessments) or an unbiased partitioning scheme. We note that there may be alternate approaches to derive trust scores from inconsistency measures and justifications; indeed, our approach is flexible and extensible to such trust computation models.

A naive implementation of our trust computation model requires *all* justifications. While computing a justification is an easy problem, exhaustively enumerating all possible justifications is known to be hard problem [11]. We formulate exhaustive enumeration of justifications as a tree traversal problem and develop an *importance sampling* approach to uniformly and randomly sample justifications without completely enumerating them. Unbiased sampling of justifications ensures that the malicious entities cannot game the trust computation model; say, selectively hide justifications that include axioms from malicious entities (and thus evade penalties) from the sampling process. For scalability reasons, our trust computation model operates on a random sample of justifications. A malicious entity may escape penalties due to incompleteness of justifications; however, across multiple inconsistency checks a malicious entity is likely to incur higher penalties (and thus lower trust score) than the honest entities.

In the following portions of this section we describe each of these steps: quantifying inconsistencies, extracting justifications and the trust computation model in detail.

3.2. Inconsistency and Justification

The ability to detect contradicting statements and measure the relative importance of the resulting conflict is a key prerequisite to estimate the (dis)similarity between information sources providing rich, complex and probabilistic assertions expressed as BDL axioms. Unfortunately, in the traditional BDL semantics [2], consistency is still categorically defined, i.e., a probabilistic KB is either completely satisfied or completely unsatisfied. In this section, we address this significant shortcoming by using a refined semantics which introduces the notion of degree of inconsistency. We start by presenting the traditional BDL semantics, which does not tolerate inconsistency.

Recall that BDL axioms ($\phi : e$) are extensions of classical axioms (ϕ) with a probabilistic annotation (e). BDL semantics defines an annotated interpretation as an extension of a first-order interpretation by assigning a value $v \in D(V)$ to V. An annotated interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, .^{\mathcal{I}})$ is defined in a similar way as a first-order interpretation except that the interpretation function $.^{\mathcal{I}}$ also maps the set of variables V in the Bayesian Network to a value $v \in D(V)$. An annotated interpretation $\mathcal{I}$ satisfies a probabilistic axiom $\phi : e$, denoted

$\mathcal{I} \models \phi : e$, iff $V^{\mathcal{I}} \models e \Rightarrow \mathcal{I} \models \phi$ [1]. Now, a probabilistic interpretation is defined as a probabilistic distribution over annotated interpretations.

Definition 1. *(From [2]) A* probabilistic interpretation *Pr is a probability function over the set of all annotated interpretations that associates only a finite number of annotated interpretations with a positive probability. The probability of a probabilistic axiom $\phi : e$ in Pr, denoted $Pr(\phi : e)$, is the sum of all $Pr(\mathcal{I})$ such that $\mathcal{I}$ is an annotated interpretation that satisfies $\phi : e$. A probabilistic interpretation Pr satisfies (or is a model of) a probabilistic axiom $\phi : e$ iff $Pr(\phi : e) = 1$. We say Pr satisfies (or is a model of) a set of probabilistic axioms F iff Pr satisfies all $f \in F$.*

Finally, we define the notion of consistency of a probabilistic knowledge base.

Definition 2. *(From [2]) The probabilistic interpretation Pr satisfies (or is a model of) a probabilistic knowledge base $K = (\mathcal{T}, \mathcal{A}, BN)$ iff (i) Pr is a model of $\mathcal{T} \cup \mathcal{A}$ and (ii) $Pr_{BN}(V = v) = \sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v} Pr(\mathcal{I})$ for all $v \in D(V)$. We say KB is consistent iff it has a model Pr.*

We note that condition (ii) in the previous definition ensures that the sum of probability values for annotated interpretations mapping V to $v \in D(V)$ is the same probability value assigned to $V = v$ by the Bayesian Network.

3.2.1. Degree of Inconsistency

In the previously presented traditional BDL semantics, consistency is still categorically defined. We now address this significant shortcoming for our trust application using a refined semantics which introduces the notion of degree of inconsistency.

First, we illustrate using a simple example the intuition behind the notion of degree of inconsistency for a KB. Let K be the probabilistic KB defined as follows: $K = (\mathcal{T}, \mathcal{A} \cup \{\top \sqsubseteq \bot : X = true\}, BN)$ where $\mathcal{T}$ is a classical Tbox and $\mathcal{A}$ is a classical Abox such that the classical KB $cK = (\mathcal{T}, \mathcal{A})$ is consistent; BN is a Bayesian Network over a single boolean random variable X, and the probability $Pr_{BN}(X = true) = 10^{-6}$ that X is true is extremely low. Under past probabilistic extensions to DL, the K is completely inconsistent, and nothing meaningful can be inferred from it. This stems from the fact that when X is true, the set of classical axioms that must hold (i.e., $\mathcal{T} \cup \mathcal{A} \cup \{\top \sqsubseteq \bot\}$) is inconsistent. However, the event $X = true$ is extremely unlikely, and, therefore, it is unreasonable to consider the whole probabilistic KB inconsistent. Intuitively, the likelihood of events, whose set of associated classical axioms is inconsistent, represents the degree of inconsistency of a probabilistic KB.

We now formally define a degree of inconsistency and present an inconsistency-tolerant refinement of the semantics of a Bayesian DL.

[1] This more expressive implication semantics differs from the equivalence semantics of [2]

Definition 3. *An annotated interpretation $\mathcal{I}$ is an annotated model of a probabilistic KB $K = (\mathcal{T}, \mathcal{A}, BN)$ where BN is a Bayesian Network over a set of variables V iff for each probabilistic axiom $\phi : e$, $\mathcal{I}$ satisfies $\phi : e$.*

In order, to measure the degree of inconsistency, we first need to find all primitive events v (i.e., elements of the domain $D(V)$ of the set of variables V) for which there are no annotated models $\mathcal{I}$ such that $V^{\mathcal{I}} = v$.

Definition 4. *For a probabilistic KB $K = (\mathcal{T}, \mathcal{A}, BN)$ where BN is a Bayesian Network over a set of variables V, the set of inconsistent primitive events, denoted $U(K)$, is the subset of $D(V)$, the domain of V, such that $v \in U(K)$ iff there is no annotated model $\mathcal{I}$ of K such that $V^{\mathcal{I}} = v$*

Finally, the degree of inconsistency of a probabilistic knowledge base is defined as the probability of occurrence of an inconsistent primitive event.

Definition 5. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic KB such that BN is a Bayesian Network over a set of variables V. The degree of inconsistency of K, denoted $DU(K)$, is a real number between 0 and 1 defined as follows:*

$$DU(K) = \sum_{v \in U(K)} Pr_{BN}(V = v)$$

A probabilistic interpretation Pr (as per Definition 1) satisfies (or is a model of) a probabilistic KB $K = (\mathcal{T}, \mathcal{A}, BN)$ to a degree d, $0 < d \leq 1$ iff.:

- *(i) Pr is a model as $\mathcal{T} \cup \mathcal{A}$ (same as in Definition 2)*
- *(ii) for $v \in V$,*

$$\sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v} Pr(\mathcal{I}) = \begin{cases} 0 & \text{if } v \in U(K) \\ \frac{Pr_{BN}(V=v)}{d} & \text{if } v \notin U(K) \end{cases}$$

- *(iii) $d = 1 - DU(K)$*

A probabilistic knowledge base $K = (\mathcal{T}, \mathcal{A}, BN)$ is consistent to the degree d, with $0 < d \leq 1$, iff there is a probabilistic interpretation that satisfies K to the degree d. It is completely inconsistent (or satisfiable to the degree 0), iff $DU(K) = 1$.

Informally, by assigning a zero probability value to all annotated interpretations corresponding to inconsistent primitive events, (ii) in Definition 5 removes them from consideration, and it requires that the sum of the probability value assigned to interpretations mapping V to v for $v \notin U(K)$ is the same as the joint probability distribution Pr_{BN} defined by BN with a normalization factor d.

In practice, computing the degree of inconsistency of a Bayesian DL KB can be reduced to classical description logics consistency check as illustrated by Theorem 1. First we introduce an important notation used in the remainder of the paper:

Notation 1. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic KB. For every $v \in D(V)$, let $\mathcal{T}_v$ (resp., $\mathcal{A}_v$) be the set of all axioms ϕ for which there exists a probabilistic axiom $\phi : e$ in $\mathcal{T}$ (resp., $\mathcal{A}$), such that $v \models e$. K_v denotes the classical KB $(\mathcal{T}_v, \mathcal{A}_v)$. Informally, K_v represents the classical KB that must hold when the primitive event v occurs. $K_\top$ denotes the classical KB obtained from K after removing all probabilistic annotations: $K_\top = (\cup_{v \in D(V)} \mathcal{T}_v . \cup_{v \in D(V)} \mathcal{A}_v)$.*

Theorem 1. *A probabilistic KB $K = (\mathcal{T}, \mathcal{A}, BN)$ is consistent to the degree d iff.*

$$d = 1 - \sum_{v \text{ s.t. } K_v \text{ inconsistent}} Pr_{BN}(V = v)$$

The proof of Theorem 1 is a consequence of Lemma 1 (detailed proofs for the Lemma and the Theorem are in Appendix A and Appendix B).

Lemma 1. *Let K be a probabilistic KB. $v \in U(K)$ iff K_v is inconsistent.*

3.2.2. Inconsistency Justification

A conflict or contradiction is formally captured by the notion of an inconsistency justification − minimal inconsistency preserving subset of the KB.

Definition 6. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic KB consistent to the degree d such that BN is a Bayesian Network over a set of variables V. $\mathcal{J}$ is an inconsistency justification iff. 1) $\mathcal{J} \subseteq (\mathcal{T}, \mathcal{A})$, 2) $(\mathcal{J}, BN)$ is probabilistic KB consistent to the degree d' such that $d' < 1$, and 3) for all $\mathcal{J}' \subset \mathcal{J}$, $(\mathcal{J}', BN)$ is probabilistic KB consistent to the degree 1 (i.e. $(\mathcal{J}', BN)$ is completely consistent). The degree $DU(\mathcal{J})$ of an inconsistency justification $\mathcal{J}$ is defined as the degree of inconsistency of the probabilistic KB made of its axioms: $DU(\mathcal{J}) = DU((\mathcal{J}, BN))$*

Justification computation in a probabilistic KB reduces to justification computation in classical KBs as shown by the following theorem, which is a direct consequence of Theorem 1 and Definition 6:

Theorem 2. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic KB, where BN is a Bayesian network over a set V of random variables. $\mathcal{J}$ is an inconsistency justification of K iff. there exists $v \in D(V)$ such that $Pr_{BN}(V = v) > 0$ and $\mathcal{J}_\top$, the classical KB obtained from $\mathcal{J}$ by removing all probabilistic annotations, is an inconsistency justification of K_v. Furthermore, the degree, $DU(\mathcal{J})$, of an inconsistency justification $\mathcal{J}$ is as follows:*

$$DU(\mathcal{J}) = \sum_{v \text{ s.t. } \mathcal{J}_\top \subseteq K_v} Pr_{BN}(V = v)$$

Thus, once we have found a classical justification in a classical KB K_v for $v \in D(V)$ using, for example, the scalable approach described in our previous work [4], the degree of the corresponding probabilistic justification can be obtained through simple set inclusion tests.

Theorems 1 and 2 provide a concrete mechanism to compute degree of inconsistency of a probabilistic KB, and a degree of inconsistency of a justification. However, they are highly intractable since they require an exponential number, in the number of variables in BN, of corresponding classical tasks. We will address this issue in the next section.

3.2.3. Error-Bounded Approximate Reasoning

A Bayesian network based approach lends itself to fast Monte Carlo sampling algorithms for scalable partial consistency checks and query answering over a large probabilistic KB. In particular, we use a *forward sampling* approach described in [1] to estimate $pr = \sum_{v \in \Pi} Pr_{BN}(V = v)$ (recall theorem 1 and 2). The forward sampling approach generates a set of samples $v_1, \cdots, v_n$ from BN (each sample is generated in time that is linear in the size of BN) such that the probability pr can be estimated as $\widehat{pr_n} = \frac{1}{n} * \sum_{i=1}^{n} I(v_i \in \Pi)$, where $I(z) = 1$ if z is true; 0 otherwise. One can show that $\widehat{pr_n}$ is an unbiased estimator of pr such that $\lim_{n\to\infty} \sqrt{n} * (\widehat{pr_n} - pr) \rightarrow \mathcal{N}(0, \sigma_z^2)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 and σ_z^2 denotes the variance of $I(z)$ for a boolean variable z. Hence, the sample size n which guarantees an absolute error of ϵ or less with a confidence level η is given by the following formula: $n = \frac{2*(erf^{-1}(\eta))^2 * \sigma_{z_{max}}^2}{\epsilon^2}$, where erf^{-1} denotes the inverse Gauss error function ($\sigma_{z_{max}}^2 = 0.25$ for a boolean random variable). For example, to compute the degree of consistency of a probabilistic KB within $\pm 5\%$ error margin with a 95% confidence, the sample size $n = 396$ is necessary.

3.2.4. Sampling Justifications in a Classical KB

Ideally, it is desirable to find all classical justifications. Computing a single justification can be done fairly efficiently by 1) using tracing technique to obtain a significantly small set S of axioms that is responsible for an inconsistency discovered by a single consistency test, and 2) performing additional $|S|$ consistency check on KBs of size at most $|S| - 1$ to remove extraneous elements from S. In our previous work [4], we presented a scalable approach to efficiently compute a large number of − but not all − justifications in large and expressive KBs through the technique of summarization and refinement [5]. The idea consists in looking for patterns of justifications in a dramatically reduced summary of the KB, and retrieve concrete instances of these patterns in the real KB.

Unfortunately, computing all justifications is well known to be intractable even for small and medium size expressive KBs [11]. [11] establishes a connection between the problem of finding all justifications and the hitting set problem (i.e., given n sets S_i, find sets that intersect each S_i). The intuition behind this result is the fact that in order to make an inconsistent KB consistent at least one axiom from each justification must be removed. Therefore, starting from a single justification a Reiter's Hitting Tree can be constructed in order to get all justifications as illustrated in Figure 2 from [11]: Starting from the first justification $J = \{2, 3, 4\}$ computed in the KB K (J is set to be the root v_0 of the tree), the algorithm arbitrary selects an axiom in J, say 2, and creates

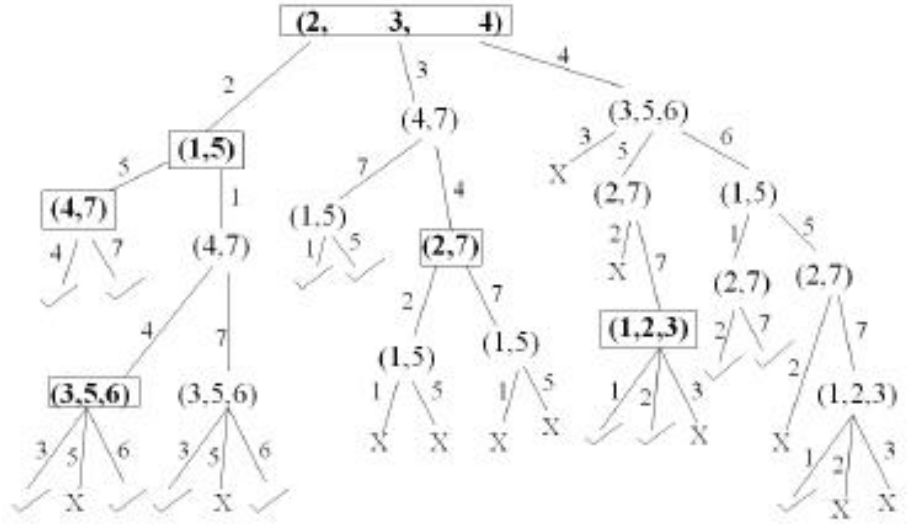

Figure 2: Computing all justifications using Reiter's Hitting Set Tree Algorithm from [11]

a new node w with an empty label in the tree and a new edge $< v_0, w >$ with axiom 2 in its label. The algorithm then tests the consistency of the $K - \{2\}$. If it is inconsistent, as in this case, a justification J' is obtained for $K - \{2\}$, say $\{1, 5\}$, and it is inserted in the label of the new node w. This process is repeated until the consistency test is positive in which case the new node is marked with a check mark. As an important optimization, we stop exploring super set of path discovered earlier and marked the node with 'X'.

In order to avoid the high cost associated with exploring the whole Hitting Set Tree to find all conflicts. One can find the first K conflicts by exploring the Reiter's Hitting Set Tree (HST) until K distinct justifications are found. The problem with this approach is that nodes in the HST are not equally likely to be selected with such a scheme: the probability $\pi(v_d)$ of a node v_d in a path $< v_0 v_1 ... v_d >$ to be selected is $\pi(v_d) = \prod_{0 \leq i < d}(1/|v_i|)$, where $|v_i|$ denotes the number of axioms in the justification v_i. As a result, a malicious source can use the bias in the sampling to 'hide' its conflicts.

However, since the bias can be precisely quantified, one can obtain an unbiased sample as follows. We select K nodes in the HST by exploring the HST in the normal way, but each time a node v_i is encountered, it is selected iff. a random number r generated uniformly from [0,1] is such that $r \leq \min(\beta/\pi(v_i), 1)$, where β is a strictly positive real number. The following Proposition shows that, in this approach, for a sample of K HST nodes, if β is properly chosen, then the expected number of time a node is selected is identical for all nodes.

Proposition 1. *Let N_v denotes the random variable representing the number of time the node v appears in a HST sample of size K. The expected value*

$E(N_v)$ *of* N_v *is:*

$$E(N_v) = \begin{cases} K * \pi(v) & \text{if } \beta \geq \pi(v) \\ K * \beta & \text{if } 0 < \beta < \pi(v) \end{cases}$$

Thus, if β is chosen such that $0 < \beta < \min_{v \in HST}(\pi(v))$, then we obtain an unbiased sample from the HST. Unfortunately, the minimum value of $\pi(v)$ depends on the tree structure (branching factor and maximum depth), and cannot be computed precisely without exploring the whole HST. In practice, we use the following sampling approach to select K nodes (the trade-off between computation cost and bias in the sample is controlled by a parameter of the algorithm, α):

1. Let *visited* denote the set of visited nodes. Set *visited* to $\emptyset$,
2. Traverse the HST in any order, and add the first $\max(K - |visited|, 1)$ nodes visited to *visited*
3. Let π_{min} be the minimum value of $\pi(v)$ for $v \in visited$.
4. Set $\beta = \pi_{min}/\alpha$, where $\alpha > 1$ is a parameter of the sampling algorithm which controls the trade-off between computation cost and biased in the sampling. Higher values of α, while reducing the bias in our sampling, increase the computation cost by reducing the probability of a node selection − hence, increasing the length of tree traversal.
5. For each $v \in visited$, add it to the result set RS with a probability of $\beta/\pi(v)$
6. If $|RS| < K$ and the HST has not been completely explored, then set $RS = \emptyset$ and continue the exploration from step 2; otherwise return RS

3.3. Trust Computation Model

We now briefly formalize the problem of assessing trust in a set IS consisting of n information sources. The trust value assumed or known prior to any statement made by an information source i is specified by a probability distribution $PrTV(i)$ over the domain [0, 1]. For example, a uniform distribution is often assumed for new information source for which we have no prior knowledge. Statements made by each source i is specified in the form of a probabilistic KB $\mathcal{K}^i = (\mathcal{T}^i, \mathcal{A}^i, BN^i)$. The knowledge function C maps an information source i to the probabilistic KB $\mathcal{K}^i$ capturing all its statements. The trust update problem is a triple $(IS, PrTV, C)$ whose solution yields a posterior trust value function $PoTV$. $PoTV$ maps an information source i to a probability distribution over the domain [0, 1], which represents our new belief in the trustworthiness of i after processing statements in $\bigcup_{j \in IS} C(j)$.

In this paper, we only focus on trust computation based on direct observations, that is, on statements directly conveyed to us by the information sources. Inferring trust from indirect observations (e.g., statements conveyed to us from IS_1 via IS_2) is an orthogonal problem; one could leverage solutions proposed in [12], [21], [15] to infer trust from indirect observations.

We model prior and posterior trust of a source i ($PrTV(i)$ and $PoTV(i)$) using a beta distribution $\mathcal{B}(\alpha, \beta)$ as proposed in several other trust computation models including [10]. Intuitively, the reward parameter α and the penalty parameter β correspond to good (non-conflicting) and bad (conflicting) axioms

contributed to an information source respectively. The trust assessment problem now reduces to that of (periodically) updating the parameters α and β based on the axioms contributed by the information sources. One may bootstrap the model by setting $PrTV(i)$ to $\mathcal{B}(1,1)$ − a uniform and random distribution over [0, 1], when we have no prior knowledge. In the rest of this section we focus on computing the reward (α) and penalty (β) parameters.

We use a simple reward structure wherein an information source receives unit reward for every axiom it contributes if the axiom is not in a justification for inconsistency [2]. We use a scaling parameter $\triangle$ to control the relative contribution of reward and penalty to the overall trust assessment; we typically set $\triangle > 1$, that is, penalty has higher impact on trust assessment than the reward. The rest of this section focuses on computing penalties from justifications for inconsistency.

Section 3.2.4 describes solutions to construct (a random sample of) justifications that explain inconsistencies in the KB; further, a justification J is associated with a weight $DU(J)$ that corresponds to the possible worlds in which the justification J holds (see section 3.2.2 for formal definition of $DU(J)$ and an algorithm to compute it). For each justification J_i we associate a penalty $\triangle(J_i) = \triangle * DU(J_i)$. The trust computation model traces a justification J_i, to conflicting information sources $\mathcal{S} = \{S_{i_1}, \cdots, S_{i_n}\}$ (for some $n \geq 2$) that contributed to the axioms in J_i. In this paper we examine three solutions to partition $\triangle(J_i)$ units of penalty amongst the contributing information sources as shown below. We use t_{i_j} to denote the expectation of $PrTV(i_j)$ for an information source i_j, that is, $t_{i_j} = \frac{\alpha_{i_j}}{\alpha_{i_j}+\beta_{i_j}}$.

$$\triangle(S_{i_j}) = \begin{cases} \frac{\triangle(J_i)}{n} & \textit{unbiased} \\ \frac{\triangle(J_i)}{n-1} * (1 - \frac{t_{i_j}}{\sum_{k=1}^{n} t_{i_k}}) & \textit{biased} \\ \textit{by trust in other sources} & \\ \triangle(J_i) * \frac{\frac{1}{t_{i_j}}}{\sum_{k=1}^{n} \frac{1}{t_{i_k}}} & \textit{biased} \\ \textit{by inverse self trust} & \end{cases}$$

The unbiased version distributes penalty for a justification equally across all conflicting information sources; the biased versions tend to penalize less trustworthy sources more. One possible approach is to weigh the penalty for a source S_{i_j} by the sum of the expected prior trust values for all the other conflicting sources, namely, $\mathcal{S} - \{S_{i_j}\}$. For instance, if we have three information sources S_{i_1}, S_{i_2} and S_{i_3} with expected prior trust $t_{i_1} = 0.1$ and $t_{i_2} = t_{i_3} = 0.9$ then the penalty for source i_1 must be weighted by $\frac{1}{2} * \frac{0.9+0.9}{0.1+0.9+0.9} = 0.47$, while that of sources i_2 and i_3 must be weighted by 0.265. Clearly, this approach penalizes the less trustworthy source more than the trusted sources; however, we note that even when the prior trust in i_1 is arbitrarily close to zero, the penalty for the honest source i_2 and i_2 is weighted by 0.25. A close observation reveals that

[2] A preprocessing step weeds out trivial axioms (e.g., *sun rises in the east*)

a malicious source (with very low prior trust) may penalize honest nodes (with high prior trust) by simply injecting conflicts that involve the honest nodes; for instance, if sources i_2 and i_3 assert axioms ϕ_2 and ϕ_3 respectively, then the malicious source i_1 can assert an axiom $\phi_1 = \neg\phi_2 \vee \neg\phi_3$ and introduce an inconsistency whose justification spans all the three sources. To overcome this problem, this paper uses a third scheme that weights penalties for justifications by the inverse value of prior trust in the information source.

4. Trust-based Query Answering

In this paper we have so far described solutions to infer trust in uncertain information. Now, we proceed towards leveraging the inferred trust scores to refine the knowledge base and supporting trust-based query answering over the refined knowledge base. In this paper we adopt a simple approach to refine the knowledge base: first, for each information source s we add a random variable T_s to the Bayesian network such that T_s is the expectation of the beta distribution whose parameters α and β are inferred by the trust model (see Section 3); for each axiom $\phi : X$ in the original knowledge base, we include an axiom $\phi : X \wedge T_{source} = true$.

In the following portions of this section we focus on query answering over the refined knowledge base (encoded as BDL axioms).

4.1. Query Answering Semantics

First, we briefly recall the query answering semantics defined in [2] and explain its shortcomings:

Definition 7. *(From [2]) An annotated interpretation $\mathcal{I}$ satisfies (or is a model of) a ground query $\psi : e$, denoted $\mathcal{I} \models \psi : e$, iff $V^{\mathcal{I}} \models e$ and $\mathcal{I} \models \psi$. The probability of a ground query $\psi : e$ in Pr, denoted $Pr(\psi : e)$, is the sum of all $Pr(\mathcal{I})$ such that $\mathcal{I}$ is an annotated interpretation that satisfies $\psi : e$. An answer for a probabilistic query $Q = \psi : e$ to a probabilistic knowledge base $K = (\mathcal{T}, \mathcal{A}, BN)$ is a pair (θ, pr) consisting of a ground substitution θ for the variables in ψ and some $pr \in [0, 1]$ such that $Pr(\psi\theta : e) = pr$ for all models Pr of K. An answer (θ, pr) for Q to K is positive iff $pr > 0$.*

This definition, which requires all probabilistic models Pr of K to assign the exact same probability $Pr(\psi\theta : e) = pr$ to an answer (θ, pr), is counter-intuitive[3] as illustrated in the following example.

Example 1. *Let $K = (\mathcal{A}, \mathcal{T}, BN)$ be probabilistic knowledge base defined as follows:*

$$\begin{aligned} \mathcal{A} &= \{A(a) : X = true, D(a) : X = false\} \\ \mathcal{T} &= \{A \sqsubseteq C : X = true\} \end{aligned}$$

[3] under both the implication and equivalence semantics

and BN is a Bayesian Network over a single boolean variable X such that $Pr_{BN}(X = true) = Pr_{BN}(X = false) = 0.5$. *Consider the query* $C(y)$ *asking for all the instances* y *of* C.

The classical knowledge base $cK_1 = (\mathcal{A} = \{A(a)\}, \mathcal{T} = \{A \sqsubseteq C\})$ associated with the primitive event $X = true$ entails $C(a)$, and the classical knowledge base $cK_2 = (\mathcal{A} = \{D(a)\}, \mathcal{T} = \emptyset)$ associated with the primitive event $X = false$ does not entail $C(a)$. One would expect $(y \rightarrow a, 0.5)$ to be a solution to the query. Unfortunately, according to the query answering semantics, a cannot be a solution, because there exist two probabilistic models[4] Pr and Pr' of K such that $Pr(C(a)) \neq Pr'(C(a))$:

- Pr assigns non-zero probability to only two annotated interpretations $\mathcal{I}_T$ and $\mathcal{I}_F$ and $Pr(\mathcal{I}_T) = Pr(\mathcal{I}_F) = 0.5$. $\mathcal{I}_T$ is defined as follows : its domain $\Delta^{\mathcal{I}_T} = \{\alpha\}$, $A^{\mathcal{I}_T} = C^{\mathcal{I}_T} = \{\alpha\}, D^{\mathcal{I}_T} = \emptyset, a^{\mathcal{I}_T} = \alpha, V^{\mathcal{I}_T} = (X = true)$. $\mathcal{I}_F$ is defined as follows : its domain $\Delta^{\mathcal{I}_F} = \{\alpha, \beta\}$, $A^{\mathcal{I}_F} = \{\beta\}, C^{\mathcal{I}_F} = \{\alpha\}, D^{\mathcal{I}_F} = \{\alpha\}, a^{\mathcal{I}_F} = \alpha, V^{\mathcal{I}_F} = (X = false)$.
- Pr' assigns non-zero probability to only two annotated interpretations $\mathcal{I}'_T$ and $\mathcal{I}'_F$ and $Pr'(\mathcal{I}'_T) = Pr'(\mathcal{I}'_F) = 0.5$ $\mathcal{I}'_T = \mathcal{I}_T$. $\mathcal{I}'_T$ is the same as $\mathcal{I}_T$ previously defined. $\mathcal{I}'_F$ differs from $\mathcal{I}_F$ only in that it maps C to an empty set instead of $\{\alpha\}$: $\Delta^{\mathcal{I}'_F} = \{\alpha, \beta\}$, $A^{\mathcal{I}'_F} = \{\beta\}, C^{\mathcal{I}'_F} = \emptyset, D^{\mathcal{I}'_F} = \{\alpha\}, a^{\mathcal{I}'_F} = \alpha, V^{\mathcal{I}'_F} = (X = false)$.

The annotated interpretation $\mathcal{I}_T$ and $\mathcal{I}'_T$ obviously satisfies $C(a)$ - they have to otherwise Pr and Pr' would not be models of K. One the other hand, an annotated interpretation with non-zero probability and which maps X to $false$ does not have to satisfy $C(a)$. $\mathcal{I}_F$ does satisfy $C(a)$, making $Pr(C(a)) = 1$, while $\mathcal{I}'_F$ does not satisfy it, making $Pr'(C(a)) = 0.5$. So, we have two models which disagree on the probability value to assign to $(y \rightarrow a)$ as an answer, therefore according to the query answering semantics of Definition 7 $(y \rightarrow a)$ cannot be an answer, which is counterintuitive.

Furthermore, this example also serves as a counter-example to the main computational result of [2] [5].

4.1.1. Refined Query Answering Semantics

Requiring all probabilistic models of K to agree on the probability value assigned to each ground query is too constraining. A more appropriate semantics, which aligns better with our intuition, consists in defining the probability value associated with an variable assignment θ answer to a query $\psi : e$ as the infimum of $Pr(\psi\theta : e)$ for all models Pr of K. The following definition formally introduces the notion of meaningful answers.

[4] These are models under both implication and equivalence semantics

[5] It shows that the simple computation of Theorem 6 in [2] is incompatible with their semantics. In fact, for a query Q, their semantics requires to perform not only the classical query Q answering against all classical KB K_v, but also classical query $\neg Q$. Our proposed semantics avoids this issue (no $\neg Q$ introduced) and is proven sound (see Appendix).

Definition 8. *An* meaningful answer *for a probabilistic query $Q = \psi : e$ to a probabilistic KB $K = (\mathcal{T}, \mathcal{A}, BN)$ satisfiable to a degree d ($d \neq 0$), is a pair (θ, pr) consisting of a ground substitution θ for the variables in ψ and some $pr \in [0, 1]$ such that $pr = \inf\{Pr(\psi\theta : e) | Pr$ is model of $K\}$, where $\inf S$ denotes the infimum of the set S .*

For partially unsatisfiable probabilistic knowledge bases the *meaningful* query answering semantics of Definition 8, considers only answers for satisfiable primitive events v (i.e., $v \notin U(K)$) since, by definition, for $v \in U(K)$ and an interpretation $\mathcal{I}$ s.t. $V^{\mathcal{I}} = v$, $Pr(\mathcal{I}) = 0$. Intuitively, for a solution (θ, pr) to a query $\psi : e$, pr represents the weighted fraction (weighted by $Pr_{BN}(V = v)$) of satisfiable primitive events v s.t. $v \models e$ and θ is a solution to ψ in the classical knowledge base that must hold given v. This intuition is confirmed by Theorem 3 in the next section.

In this semantics, unsatisfiable primitive events do not contribute any solution. Consequently, it does not properly extend the query answering semantics of classical description logics, where any substitution θ is a solution to any query against any unsatisfiable classical knowledge base.

4.1.2. Extended Classical DL Semantics

To properly extend the classical query answering semantics, for a query $Q = \psi : e$ against a probabilistic knowledge base K s.t. $DU(K) \neq 1$, we need to:

- factor in the contribution of all unsatisfiable primitive events by considering that any substitution θ is a solution to a KB K associated with the unsatisfiable event. This contribution is as follows:

$$\sum_{v \in U(K) \text{ and } v \models e} Pr_{BN}(V = v)$$

- properly weight the contribution of satisfiable primitive event v to a substitution θ by denormalizing $Pr(\psi\theta : e)$. The modified weight is $(1 - DU(K))Pr(\psi\theta : e)$

Definition 9. *A* classically extended answer *for a probabilistic query $Q = \psi : e$ to a probabilistic knowledge base $K = (\mathcal{T}, \mathcal{A}, BN)$ is a pair (θ, pr) consisting of a ground substitution θ for the variables in Q and some $pr \in [0, 1]$ such that (we set $\inf \emptyset = 0$)*

$$pr =$$
$$(1 - DU(K)) \times \inf\{Pr(\psi\theta : e) | Pr \text{ is model of } K\}$$
$$+ \sum_{v \in U(K) \text{ and } v \models e} Pr_{BN}(V = v)$$

Intuitively, in Definition 9, the probability pr associated with an answer θ to a query Q represents the likelihood that, if an primitive event e is selected randomly according to the probability distribution specified by the BN, θ will be a solution of Q over the classical set of axioms that must hold when e occurs (with the understanding that any substitution is a solution for any query against any unsatisfiable classical knowledge base). In the next section, Theorem 3 formalizes this intuition.

In the rest of the paper we consider only the *classically extended answer* semantics.

4.2. Computational Properties

In this section, we introduce key theorems that demonstrate how reasoning tasks can be carried out on partially consistent knowledge bases by reducing them to reasoning over classical knowledge bases. All detailed proofs are presented in the appendix.

A key aspect of the query answering semantics presented in Definition 9 is the ability to compute the infimum of the probability of a grounded query over the set of all probabilistic models. The following critical lemma shows how such infimum can be computed precisely:

Lemma 2. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base satisfiable to the degree d ($d \neq 0$). Let $Q_g = \psi : e$ be a grounded query.*

$$\inf\{Pr(\psi : e) | Pr \text{ is a probabilistic model of } K\}$$

$$= \frac{1}{d} \sum_{v \in \Omega} Pr_{BN}(v)$$

with $\Omega = \{v | K_v$ is satisfiable and $K_v \models \psi$ and $v \models e\}$

The following theorem, which is a direct consequence of Lemma 2, shows how query answering against a partially satisfiable probabilistic knowledge base can be reduced to query answering against classical knowledge bases.

Theorem 3. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base satisfiable to the degree d, and let $Q = \psi : e$ be a probabilistic query against K. Let θ be a ground substitution for the variables in Q and let $pr \in [0, 1]$. (θ, pr) is an answer to Q iff*

$$pr = \sum_{v \in \Gamma} Pr_{BN}(V = v)$$

with $\Gamma = \{v | K_v \models \psi\theta$ and $v \models e\}$ (Note that any substitution θ is a answer to any query against an unsatisfiable classical KB).

Theorems 1 and 3 provide a concrete mechanism to check for partial satisfiability and perform query answering against a partially satisfiable probabilistic KB. However, they are highly intractable since they require an exponential number, in the number of variables in BN, of classical query answering tasks. We will address this issue in the next section.

4.3. Error-Bounded Approximate Reasoning

The structure of pr from Theorem 3 is identical to that of d (degree of inconsistency) in Theorem 1. Hence, scalable solutions to compute d described in Section 3.2.3 are applicable here.

In addition, for some applications it may be sufficient to return ground substitutions θ whose probability pr exceeds a threshold thr. In this case, we argue that it is easier to decide if pr exceeds a threshold thr, rather than accurately estimate the true value of pr itself. With slight abuse of notation, we overload pr to denote a random variable that represents our belief in the true probability associated with a ground substitution θ. Hence, our goal is to select answers (θ, pr) such that $\Pr(pr > thr) > \eta$ (where η is the confidence level); symmetrically, one can also discard answers (θ, pr) such that $\Pr(pr < thr) > \eta$. We note that since $\widehat{pr_n} = pr + X$, where $X \approx \mathcal{N}(0, \frac{\sigma^2_{z_{max}}}{n})$, $\Pr(pr > thr) = \Pr(X < \widehat{pr_n} - thr) = \frac{1}{2}\left(1 + erf\left(\frac{\sqrt{n}*(\widehat{pr_n} - thr)}{\sqrt{2}*\sigma_{z_{max}}}\right)\right)$, where erf denotes the Gaussian error function. Setting $\Delta_{n_{min}} = erf^{-1}(2\eta - 1)*\frac{\sqrt{2}*\sigma_{z_{max}}}{\sqrt{n}}$, θ is a valid (resp. invalid) substitution if $\widehat{pr_n} - thr > \Delta_{n_{min}}$ (resp. $\widehat{pr_n} - thr < -\Delta_{n_{min}}$).

Using $\eta = 95\%$ and $n = 25$ (resp. $n = 100$), we require $\Delta_{n_{min}} = 0.17$ (resp. $\Delta_{n_{min}} = 0.085$). For instance, let thr= 0.75 and consider an answer θ whose true probability $pr > 0.92$; while it requires $n = 396$ samples to estimate the true probability pr with 95% confidence and under 5% error, one can conclude that the true probability exceeds $thr = 0.75$ with 95% confidence using only $n = 25$ samples. Similarly, one can eliminate answers whose true probability is below 0.57 using only $n = 25$ samples. Hence, one can initially accept (resp. reject) answers whose probability significantly exceeds (resp. falls below) the threshold thr; only answers whose probability is in the neighborhood of the threshold thr requires more samples to validate them. One can also show that the expected value of n in order to attain the desired confidence level η is given by $\hat{n} = \frac{2*\sigma^2_{z_{max}}*(erf^{-1}(2\eta-1))^2}{(pr-thr)^2}$. For example, setting $\eta = 95\%$, $thr = 0.75$, a substitution θ with $pr = 0.85$ may be adjudged as a valid answer using sample size $n = 60$.

5. Experimental Evaluation

To evaluate our approach, we have developed a prototype implementation, PSHER, that extends SHER reasoner [6] to support Bayesian $\mathcal{SHIN}$ (the core of OWL 1.0 DL) reasoning. SHER was chosen for its unique ability to scale reasoning to very large and expressive KBs [5], and to efficiently detect large number of inconsistency justifications in a scalable way [4]. PSHER uses the results of sections 3.2.1, 3.2.2 and 3.2.3 to reduce the problem of computing justifications on a probabilistic KB to detecting those justifications on classical KBs using SHER.

Axioms asserted by various information sources in our experiments were taken from the UOBM benchmark [17] which was modified to $\mathcal{SHIN}$ expressivity, and its Abox was modified by randomly annotating half of the axioms

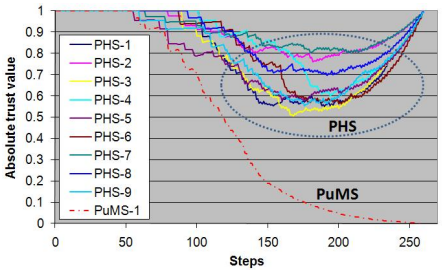

Figure 3: Trust under single PuMS attack (No duplication)

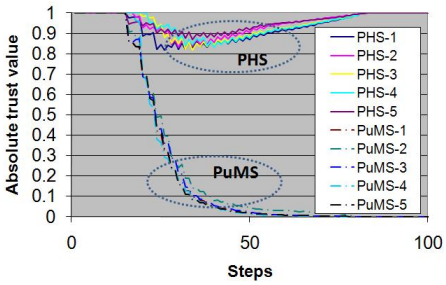

Figure 4: Trust under 50% PuMS attack (No duplication)

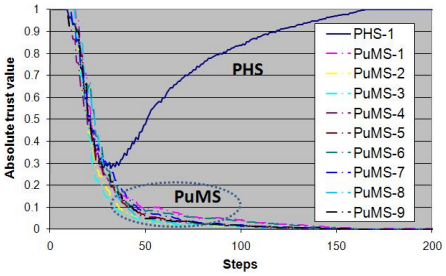

Figure 5: Trust under 90% PuMS attack (No duplication)

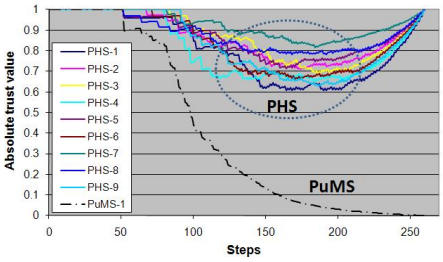

Figure 6: Trust under single PuMS attack (25% duplication)

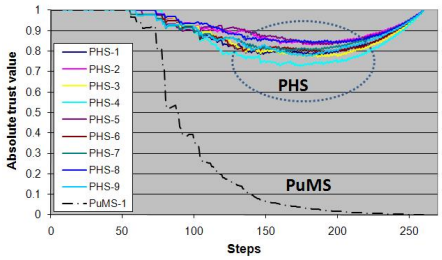

Figure 7: Trust under single PuMS attack (50% duplication)

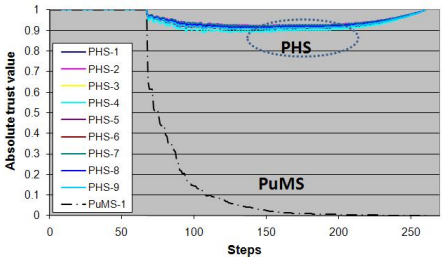

Figure 8: Trust under single PuMS attack (100% duplication)

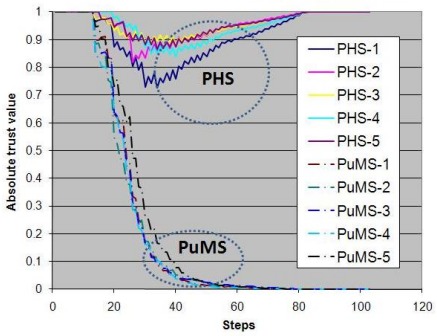

Figure 9: Trust under 50% PuMS attack (25% duplication)

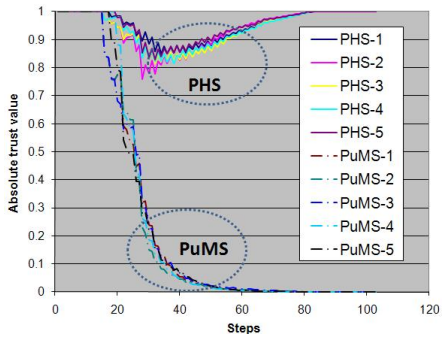

Figure 10: Trust under 50% PuMS attack (50% duplication)

with probability values. Furthermore, we inserted additional Abox assertions in order to create inconsistencies involving axioms in the original UOBM KB. Note that not all axioms in the original UOBM KB end up being part of an inconsistency, which introduces an asymmetry in information source's knowledge (e.g., a malicious source is not assumed to have complete knowledge of all axioms asserted by other sources).

5.1. Trust Inference

We evaluate the robustness of our trust model under various attack scenarios. In our experiments, we considered 4 types of information sources:

- Perfect honest sources (PHS) whose axioms are taken from the UOBM KB before the introduction of inconsistencies.
- Purely malicious sources (PuMS) whose axioms are selected from the ones added to UOBM KB in order to create inconsistencies.
- Imperfect honest sources (IHS) have the majority of their axioms (more than 90%) from the UOBM KB before the introduction of inconsistencies. They allow us to simulate the behavior of our approach when honest sources are faced with measurement errors or commit honest mistakes.

- Partially malicious sources (PaMS) are such that between 10% to 90% of their axioms are selected from the axioms added to UOBM KB to create inconsistency. They are primarily used to simulate the behavior of our approach when malicious sources use an oscillating behavior to milk our trust computation scheme.

Axioms were randomly assigned to various sources without violating the proportion of conflicting vs. non-conflicting axioms for each type of source.

Our first experiment (Figure 3) measures the impact of a single purely malicious source (PuMS) on the trust values of 9 perfect honest sources. The PuMS asserts more and more incorrect axioms contradicting PHS's axioms (at each steps, each source asserts about 100 additional statements until all their axioms have been asserted) while the PHSs continue to assert more of what we consider as correct axioms. Axioms asserted by the PuMS do not necessarily yield an inconsistency in the same step in which they are asserted, but, by the end of the simulation, they are guaranteed to generate an inconsistency. For this experiment, there is no duplication of axioms across sources, and we do not assume any prior knowledge about the trustworthiness of the sources. Since each justification creates by the malicious source also involves at least one PuMS, initially, it manages to drop significantly the absolute trust value of some PHSs (up to 50% for PHS-3). However, a PuMS hurts its trust value significantly more than he hurts those of other sources. As a result of the fact that our scheme is such that less trustworthy sources get assigned a large portion of the penalty for a justification, the single PuMS eventually ends up receiving almost all the penalty for its inconsistencies, which allows the trust values of honest sources to recover. Due to information asymmetry (malicious sources do no have complete knowledge of informations in other sources and thus cannot contradict all the statements of an PHS), our scheme remains robust, in the sense that honest sources would recover, even when the proportion of PuMS increases (see Fig. 4 where 50% of the sources are PuMS and Fig. 5 where 90% of sources are PuMS).

In the previous experiments, although honest sources manage to recover from the attack, they can still be severely hurt before the credibility of the malicious sources decreased enough to enable a recovery for honest sources. This problem can be addressed in two ways: 1) by increasing the degree of redundancy between sources as illustrated in Figures 6, 7, 8, 9 and 10; and 2) by taking into account a priori knowledge of each source as illustrated in Figure 11.

In case of moderate to high redundancy between sources (Figures 6, 7, 8, 9 and 10), a justification generated by a malicious source to compromise a honest source is likely to hurt the malicious much more than the honest source because the axioms in the justification coming from the honest source are likely to be confirmed by (i.e. duplicated in) other honest sources. Therefore, the malicious source will be involved in as many justifications as there are corroborating honest sources, while each corroborating source will be involved in a single justification.

In Figure 11, we assume that we have a high a priori confidence in the trustworthiness of the honest sources: the prior distribution of the trust value of PHS in that experiment is a beta distribution with parameter $\alpha = 2000$ and

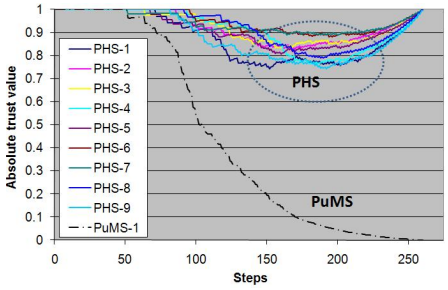

Figure 11: Trust under single PuMS attack: No duplication - Prior = B(2000,1)

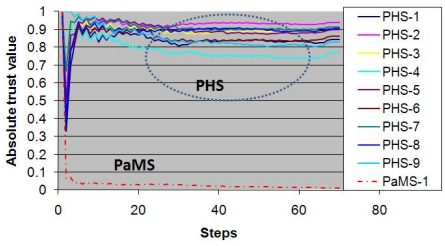

Figure 12: Oscillating experiment - 90% PHS & 10% PaMS (No duplication)

$\beta = 1$. As expected, in Figure 11, the damage inflicted by the malicious source is significantly reduced compared to Figure 3 where no prior knowledge about the source trustworthiness was taken into account.

The next experiment evaluates the behavior of our scheme when partially malicious sources use an oscillating behavior. They alternate periods where they assert incorrect axioms, contradicting axioms asserted in the same period by other sources, with periods in which they assert only correct axioms. As opposed to previous experiments where malicious axioms asserted in a step were not guaranteed to yield an inconsistency in the same step, in the oscillation experiments, the inconsistency is observed at the same step. As shown in Figure 12 and 13, in absence of prior knowledge, the trust values of partially malicious sources (PaMS) and honest sources drop significantly at the first period in which incorrect axioms are stated. However, malicious sources, which due to information asymmetry, can only contradict limited set of statements from honest sources, never recover significantly, while honest sources quickly improve their trust values by asserting more axioms not involved in conflicts. As in the previous non-oscillating experiments, the negative impact on honest sources can be reduced considerably through axiom duplication and prior strong confidence in their trustworthiness.

The last experiment simulates an oscillating scenario where all four types of sources are present: 30% PHS, 20% PuMS, 30% IHS and 20%PaMS. Figure 14 shows how our scheme correctly separates the 4 types of sources as expected.

5.2. Trust-based Query Answering

To illustrate the benefits of a trust-based query answering approach, we evaluate how query answering completeness, as measured by recall, and soundness, as measured by precision, are affected by either the lack of trust inference or imprecision in the trust inference computation. For that purpose, we assume that only half of the axioms in UOBM-1 can be trusted, while the others are assumed to be incorrect. A classical query answering against the classical knowledge base $\mathcal{K}_g$ consisting of the randomly selected trustworthy axioms of UOBM-1 provides the ground truth for our evaluation. Next, we define the following refinement

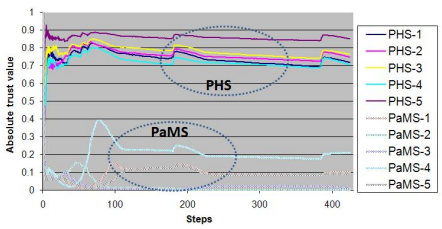

Figure 13: Oscillating experiment - 50% PHS & 50% PaMS (No duplication)

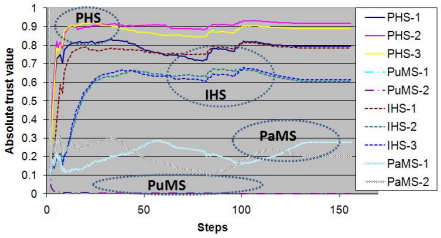

Figure 14: Oscillating experiment - 30% PHS, 20% PuMS, 30% IHS & 20% PaMS

of UOBM-1 to reflect the absence of trust computation and varying levels of precision in the trust computation module:

- $(\mathcal{K}_{ptc}, BN_{ptc})$ is a refinement of UOBM-1 where trusted axioms (i.e., axioms in $\mathcal{K}_g$) are assigned a trust value of 1 and all other axioms have a trust value of 0. $(\mathcal{K}_{ptc}, BN_{ptc})$ corresponds to the refinement performed based on the results of a perfect trust computation. Formally,

$$\mathcal{K}_{ptc} = \{\phi : T_c = \textbf{true} | \phi \in \mathcal{K}_g\}$$

$$\bigcup \{\phi : T_f = \textbf{true} | \phi \in (UOBM_1 - \mathcal{K}_g)\}$$

  where the Bayesian Network BN_{ptc} consists of two independent boolean variables T_c and T_f such that $Pr_{BN}(T_c = \textbf{true}) = 1$ and $Pr_{BN}(T_f = \textbf{true}) = 0$

- $(\mathcal{K}_{ntc}, BN_{ntc})$ represents the knowledge base obtained when no trust computation is performed. Basically all axioms are taken at face value. Formally, $\mathcal{K}_{ntc}$ is exactly the same as $\mathcal{K}_{ptc}$ defined previously, and BN_{ntc} consists of the two independent boolean variables T_c and T_f such that $Pr_{BN}(T_c = \textbf{true}) = 1$ and $Pr_{BN}(T_f = \textbf{true}) = 1$.

- $(\mathcal{K}^e_{tc}, BN^e_{tc})$ is a refinement of UOBM-1 performed based on the results of a trust inference computation with a uniform error rate of e%, where $e \in \{10, 30, 50\}$. Formally, $\mathcal{K}^e_{tc}$ is exactly the same as $\mathcal{K}_{ptc}$ defined previously, and BN^e_{tc} consists of the two independent boolean variables T_c and T_f such that $Pr_{BN}(T_c = \textbf{true}) = 1 - e/100$ and $Pr_{BN}(T_f = \textbf{true}) = e/100$.

Figures 15 and 16 show the weighted average [6] of precision and recall, respectively, for membership query answering performed on 15 concepts [7] with an absolute error +/-10% and a confidence of 90%. The results are grouped by the threshold *thr* used in selecting answers, and all the previously defined trust-based refinements of UOBM-1 are considered.

[6] each query contribution was weighted by the size of the expected number of answers
[7] Those concepts were chosen because of the complexity of the reasoning involved

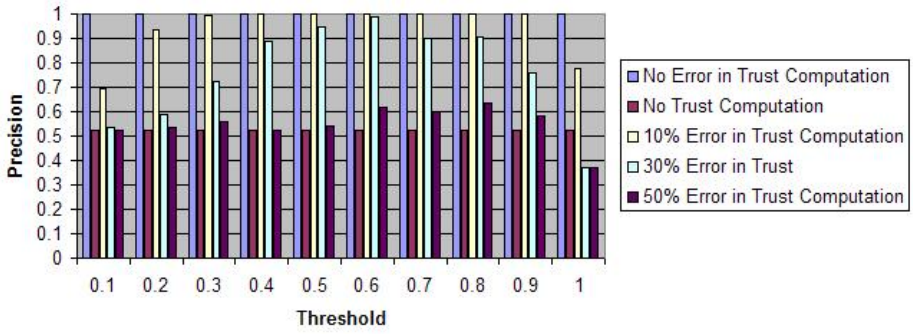

Figure 15: Precision

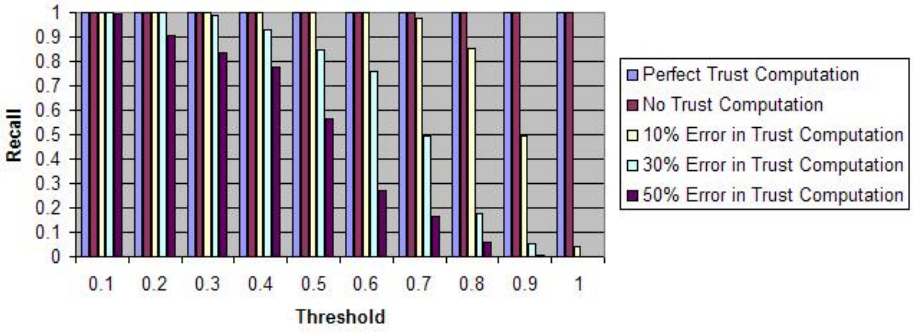

Figure 16: Recall

Dataset	type axioms	role axioms	avg (mins)	stdev (mins)	range (mins)
UOBM-1	25,453	214,177	4	1	3-5
UOBM-10	224,879	1,816,153	22	10	14-34
UOBM-30	709,159	6,494,950	61	26	41-94

Table 1: PSHER on probabilistic UOBM KBs

In Figure 15, as expected, query answering precision decreases with the error rate of the trust inference computation. However, except for the very high threshold of 1, it is always better than query answering without trust inference, which has a constant precision of about 50%, even in case of a high error rate of 50%. For recall, taking every axiom at face value (i.e., in absence of trust inference computation), already guarantees a recall of 100% at all thresholds. Such recall can only be matched at all thresholds in the case of perfect trust inference computation as illustrated in Figure 16. In practice, recall is more sensitive (especially at high thresholds) to the error rate of the trust inference computation (e.g. in Figure 16, at threshold 0.9, recall is almost equal to zero for refinements based on trust inference computation with error rate greater than or equal to 30%).

In terms of combined F-Score, for a trust inference computation with an error rate of less than 10%, which is the typical range for our implementation, trust-based query answering significantly outperforms query without trust for all thresholds that are less than or equal to 0.8. At thresholds greater than 0.9, trust-based query answering with trust inference error rate of 10% becomes inferior to query answering without trust because the threshold clearly falls into the margin of error of both the trust inference module and, in our experiment, the accuracy of our error-bounded probabilistic query answering module.

The main lesson from the results presented in this section is that, provided that the query answering threshold is properly set with regards to the error rate of the trust inference computation module and the query answering module, trust-based query answering significantly outperforms in terms of F-Score query answering without taking into account axiom trustworthiness.

5.3. Performance and Scalability

In this section we describe performance and scalability results on PSHER. We issued instance retrieval queries for a subset of concepts[8] in the ontology. The results are reported for 1, 10, and 30 universities, which are referred to as UOBM-1, UOBM-10 and UOBM-30. The runs were made on a 64 bit 2.4 GHz AMD 4-core processor 16G RAM Linux machine (only 2 cores were used: one for the SHER process, and the other for the DB2 process). The Abox was stored in DB2.

As expected, Table 1 shows that PSHER preserves SHER scalability characteristics: it still scales sublinearly. For the experiments whose results are

[8] non-atomic complex concepts

reported in Table 1, we computed the probability values for all answers, without any threshold, with an absolute error of 0.1 and an confidence of 0.9. The resulting large number of classical KBs (72) to consider for each query explains the relatively high absolute runtime in Table 1.

As described in the previous section, one can enhance the performance of PSHER using an application specified threshold (thr) for selecting answers. In that scenario, for an instance retrieval query and a given confidence level η, at every stage of the evaluation, each individual in the Abox is in one of three states : rejected when its probability is already known to be below thr; accepted when its probability is already known to be above thr; or unknown when not enough samples have been processed to conclude with confidence η. We conducted instance retrieval experiments with thresholds set at 0.6 and 0.8. As shown in Figure 17, the overwhelming majority of individuals are quickly in a known state (rejected or accepted). For example, for the concept 'Chair' with threshold 0.8, 99.95% of the individuals are in a known state in less than 3.3 mins. Most of the rest of the time is spent computing the status of a small number of individuals whose probability is close to thr.

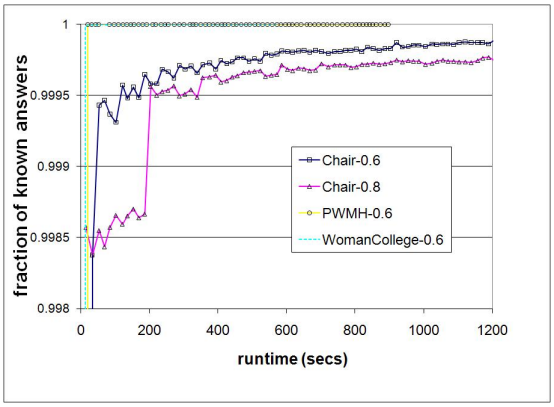

Figure 17: Query Answering with threshold on UOBM10

Scalability of PSHER is achieved through parallelism since each probabilistic reasoning task performed by PSHER is reduced to n corresponding classical tasks evaluated using SHER, where n depends on the desired precision as explained in Section 3.2.3.

6. Related Work

Our work advances the state-of-the-art in two different domains, namely, trust inference and probabilistic extension of Description Logics.

Several authors have explored various trust computation models (e.g., eBay recommendation system [20], NetFlix movie ratings [18], EigenTrust [12], PeerTrust [21], etc.) to assess trust in various entities. A common data model subsumed

by several trust computation models (as succinctly captured in Kuter and Golbeck [15]) is the ability of an entity to assign a *numeric* trust score to another entity (e.g., eBay recommendation, Netflix movie ratings, etc.). Such pair-wise numeric ratings contribute to a (dis)similarity score (e.g., based on $\mathcal{L}_1$ norm, $\mathcal{L}_2$ norm, cosine distance, etc.) which is used to compute personalized trust scores (as in PeerTrust) or recursively propagated throughout the network to compute global trust scores (as in EigenTrust). The main differences between our approach and prior work are twofold. First, our data model is far more complex than the typical pairwise numeric scores. Second, our trust computation model directly takes into account uncertainty in information asserted by various sources.

Various probabilistic extensions of Description Logics (DL), the theoretical foundation of OWL, have been proposed to enable reasoning under uncertainty. By and large, these approaches provide a tight integration of a particular DL (in most cases, a inexpressive one: $\mathcal{ALC}$ in [8] and [9], *Classic* in [14], $\mathcal{FL}$ in [23], etc.) with a given probabilistic reasoning formalism (e.g., Bayesian Network in [14], cross entropy minimization in [9]). The tightness of integration often imposes severe limitations on the underlying DL in order to assure decidability or scalability, and makes it harder to extend the approach to new DLs. Also, these approaches typically do not support uncertainty in the assertional part of the knowledge base, or impose some important restrictions on it. Probabilistic extensions of expressive DL formalisms often add an exponential overhead: in [16], for example, a probabilistic reasoning task reduces to an exponential number of corresponding classical reasoning tasks.

In the semantic web community, two of the most important threads of work in representing and reasoning with uncertainty in OWL have relied on Bayesian Networks [22] [3] or sophisticated lexicographic entailment from default reasoning [16] as their underlying probabilistic reasoning formalism. The first approaches present a syntax for encoding prior and conditional probability values for various OWL constructs as well as a translation rules to generate a Bayesian Network out of the annotations. One of their major shortcomings is their lack of formal semantics. Furthermore, [22] also lacks a general approach for query answering, whereas [3] is limited to taxonomical reasoning. On the other hand, approaches such as [16] specify a well-founded semantics, but are unfortunately highly intractable in practice and are less amenable to the form of approximation presented in this paper. Pronto the [13], the only implementation of this approach, can barely scale to a couple of dozens of probabilistic axioms.

Our work is influenced and closely aligned with [2] which introduces a flexible, loosely coupled and orthogonal integration of Description Logics and Bayesian Network formalisms without limiting, as previous integrations [23] [14] did, the expressiveness of the underlying description logics. In particular, [2] supports uncertainty in both terminological and assertional part of the knowledge base, and can be extended without any fundamental change to any decided subset of first order logic. Our approach differs from [2] in three important ways. First, we use a more intuitive semantics for query answering which is better

aligned with the spirit of the semantic web where complete information cannot be assumed. The difference between our query answering semantics (based on infimum probability value over all probabilistic models) and the previous semantics, which requires all models to agree on a probability value for an answer, is significant. As illustrated by Example 1 of section 4, the previous semantics is not only counter-intuitive; it also does not produce the desirable computational properties claimed in [2]. This paper is the first to propose an intuitive semantics for Bayesian DLs and to correctly establish the kinds of flexibility and computational properties claimed in [2] accompanied by a detailed proof of correctness in Appendix C and Appendix D. Second, we extended [2] to better tolerate inconsistencies that could naturally arise from the aggregation of uncertain information coming from different sources with a varying reliability. Finally, while the approach outlined in [2] was tractable with respect to data complexity, it still required, for each probabilistic reasoning task, an exponential number, in the number of random variables in the Bayesian Network, of similar classical reasoning tasks. This made the approach intractable in practice for Bayesian Network with a medium to large number of variables. Our approach addresses this issue by using a sampling technique to approximate as precisely as needed the probability value associated with each answer to a query against a probabilistic knowledge base.

7. Conclusion

In this paper, we have introduced a new trust framework for rich, complex and uncertain information by leveraging the expressiveness of Bayesian Description Logics. We have demonstrated the robustness of the proposed framework under a variety of scenarios, and shown how duplication of assertions across different sources as well as prior knowledge of the trustworthiness of sources can further enhance it. We have also shown techniques to enable scalable trust-based query answering over uncertain knowledge base.

Acknowledgements. Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorised to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] J. Cheng and M. J. Druzdzel. AIS-BN: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks. In *Journal of AI Research*, 2000.

[2] C. D'Amato, N. Fanizzi, and T. Lukasiewicz. Tractable reasoning with bayesian description logics. In *Scalable Uncertainty Management (SUM08)*, pages 146–159, 2008.
[3] Z. Ding, Y. Peng, and R. Pan. BayesOWL: Uncertainty modeling in semantic web ontologies. *Studies in Fuzziness and Soft Computing*, 204:3, 2006.
[4] J. Dolby, J. Fan, A. Fokoue, A. Kalyanpur, A. Kershenbaum, L. Ma, J. W. Murdock, K. Srinivas, and C. A. Welty. Scalable cleanup of information extraction data using ontologies. In *ISWC/ASWC*, pages 100–113, 2007.
[5] J. Dolby, A. Fokoue, A. Kalyanpur, A. Kershenbaum, E. Schonberg, K. Srinivas, and L. Ma. Scalable semantic retrieval through summarization and refinement. In *AAAI*, pages 299–304, 2007.
[6] J. Dolby, A. Fokoue, A. Kalyanpur, E. Schonberg, and K. Srinivas. Scalable highly expressive reasoner (sher). *J. Web Sem.*, 7(4):357–361, 2009.
[7] A. Fokoue, M. Srivatsa, and R. Young. Assessing trust in uncertain information. In *International Semantic Web Conference*, 2010.
[8] J. Heinsohn. Probabilistic description logics. In *UAI-1994*, 1994.
[9] M. Jaeger. Probabilistic reasoning in terminological logics. In *KR-94*.
[10] A. Josang and R. Ismail. The beta reputation system. In *15th Conference on Electronic Commerce*, 2002.
[11] A. Kalyanpur. *Debugging and Repair of OWL-DL Ontologies*. PhD thesis, University of Maryland, 2006.
[12] S. Kamvar, M. Schlosser, and H. Garcia-Molina. EigenTrust: Reputation management in P2P networks. In *WWW Conference*, 2003.
[13] P. Klinov. Pronto: A non-monotonic probabilistic description logic reasoner. In *ESWC*, pages 822–826, 2008.
[14] D. Koller, A. Levy, and A. Pfeffer. P-classic: A tractable probabilistic description logic. In *AAAI-97*, pages 390–397, 1997.
[15] U. Kuter and J. Golbeck. SUNNY: A New Algorithm for Trust Inference in Social Networks, using Probabilistic Confidence Models. In *AAAI-07*, 2007.
[16] T. Lukasiewicz. Expressive probabilistic description logics. *Artif. Intell.*, 172(6-7):852–883, 2008.
[17] L. Ma, Y. Yang, Z. Qiu, G. Xie, and Y. Pan. Towards a complete owl ontology benchmark. In *ESWC, 2006*, pages 124–139, 2006.
[18] Netflix. Netflix Prize. http://www.netflixprize.com/.
[19] L. G. Neuberg. Causality: Models, reasoning, and inference, by judea pearl, cambridge university press, 2000. *Econometric Theory*, 19(04):675–685, August 2003.
[20] J. B. Schafer, J. Konstan, and J. Riedl. Recommender Systems in E-Commerce. In *ACM Conference on Electronic Commerce*, 1999.
[21] L. Xiong and L. Liu. Supporting reputation based trust in peer-to-peer communities. In *IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 71, 16(7)*, July 2004.
[22] Y. Yang and J. Calmet. Ontobayes: An ontology-driven uncertainty model. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, volume 1, pages 457–463, 2005.
[23] P. M. Yelland. An alternative combination of bayesian networks and description logics. In *KR*, pages 225–234, 2000.

Appendix A. Proof of Lemma 1

We prove that both directions of the equivalence in Lemma 1 hold.

Appendix A.1. $v \in U(K) \Rightarrow K_v$ *is inconsistent*

Let $v \in U(K)$. Assuming K_v is consistent, it follows that it has a model I_v, which we extend to be an annotated interpretation by setting $V^{I_v} = v$.

Let $\phi : e \in K$, we examine the following two cases:

- if $v \models e$, then, by definition of K_v, $\phi \in K_v$. Since I_v is a model of K_v, it follows that $I_v \models \phi$. Therefore I_v is a model of the annotated axiom $\phi : e$ of K
- if $v \models e$ does not hold, then, by definition of the implication semantics, I_v is a model of the annotated axiom $\phi : e$ of K.

So I_v is a model of every annotated axiom $\phi : e$ of K and $V^{I_v} = v$, which, by definition of $U(K)$, implies $v \notin U(K)$. This contradicts our initial hypothesis. Therefore, ($v \in U(K) \Rightarrow K_v$ is inconsistent) must hold

Appendix A.2. K_v *is inconsistent* $\Rightarrow v \in U(K)$

Let v be such that K_v is inconsistent. Assuming that $v \notin U(K)$, it follows that there is an annotated model I_v of K such that $V^{I_v} = v$.

Let $\phi \in K_v$, by definition of K_v, there is at least one annotated axiom $\phi : e$ in K s.t. $v \models e$. The fact that I_v is an annotated model of K s.t. $V^{I_v} = v$ implies that $I_v \models \phi$. So I_v is a model of every axiom $\phi \in K_v$, which means that I_v is a model of K_v and K_v is consistent. This contradicts our initial hypothesis. Therefore (K_v is inconsistent $\Rightarrow v \in U(K)$) must hold.

Appendix B. Proof of Theorem 1

We start by introducing some important definitions needed for the proof.

Definition 10. *Let* $K = (\mathcal{T}, \mathcal{A}, BN)$ *be a probabilistic knowledge base, where* BN *is a Bayesian Network over a set of variables* V*. An* annotated interpretation mapping choice τ *is a complete function from the set* $D(V) - U(K)$ *of consistent primitive events to the set* $\hat{\mathcal{I}}$ *of all annotated interpretations of* K *such that, for* $v \in D(V) - U(K)$*:*

- $\tau(v)$ *is an annotated interpretation such that* $V^{\tau(v)} = v$*, and*
- $\tau(v)$ *is a model of* K_v[9]*.*

[9] More precisely, the classical interpretation which $\tau(v)$ extends by assigning $V^{\tau(v)}$ to v is a model of K_v.

A direct consequence of Lemma 1 is that, if $DU(K) \neq 1$, at least one annotated interpretation mapping choice exists. It can be constructed by mapping every consistent primitive event v to an annotated interpretation I_v which extends a model of K_v by a assigning V^{I_v} to v. From the definition, it is obvious that an annotated interpretation mapping choice is an injection.

Now, for a given annotated interpretation mapping choice τ, we introduce the notion of canonical probabilistic interpretation of K:

Definition 11. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base such that its degree of inconsistency $DU(K) \neq 1$ and BN is a Bayesian Network over a set of variables V. Let τ be an annotated interpretation mapping choice. The canonical probabilistic interpretation of K given τ, denoted Pr^τ, is the probabilistic interpretation defined as follows: for an annotated interpretation $\mathcal{I}$,*

$$Pr^\tau(\mathcal{I}) = \begin{cases} \frac{Pr_{BN}(V=v)}{1-DU(K)} & \text{if } \mathcal{I} = \tau(v) \\ 0 & \text{if } \forall v \in D(V) - U(V), \\ & \tau(v) \neq \mathcal{I} \end{cases}$$

Next, we present and prove an important Lemma needed to establish Theorem 1:

Lemma 3. *Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base such that its degree of inconsistency $DU(K) \neq 1$. Let τ be an annotated interpretation mapping choice. The canonical probabilistic interpretation Pr^τ of K given τ is a model of K to the degree $d = 1 - DU(K)$. We hereafter refer to it as the* canonical probabilistic model *given τ.*

Appendix B.1. Proof of Lemma 3

Condition (iii) of the definition of a probabilistic model (see Definition 5) is obviously satisfied.

We now prove that condition (ii) of Definition 5 is satisfied. For $v \in V$, we show that $P_v = \sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}}=v} Pr^\tau(\mathcal{I})$ satisfies (ii). We conisder the following two cases:

- Case 1: $v \in U(K)$. Since, by definition of Pr^τ, $Pr^\tau(\mathcal{I}) = 0$ for any annotated interpretation $\mathcal{I}$ such that $V^{\mathcal{I}} = v$ with $v \in U(K)$, it follows that $\sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}}=v} Pr^\tau(\mathcal{I}) = 0$.
- Case 2: $v \notin U(K)$. By definition of Pr^τ, for all $\mathcal{I} \in \hat{\mathcal{I}}$ such that $V^{\mathcal{I}} = v$ and $\mathcal{I} \neq \tau(v)$, $Pr^\tau(\mathcal{I}) = 0$. Therefore, $P_v = Pr^\tau(\tau(v))$, and, by definition of Pr^τ, $Pr^\tau(\tau(v)) = \frac{Pr_{BN}(V=v)}{1-DU(K)}$, which establishes $\sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}}=v} Pr^\tau(\mathcal{I}) = \frac{Pr_{BN}(V=v)}{1-DU(K)}$

In both cases, condition (ii) is satisfied.

Finally, we prove that condition (i) of Definition 5 is satisfied. Let $\phi : e$ be an axiom in K. We want to establish that $Pr^\tau(\phi : e) = 1$.

By definition, $Pr^{\tau}(\phi : e) = \sum_{\mathcal{I} \ s.t. \ \mathcal{I} \models \phi:e} Pr(\mathcal{I})$. It follows that $Pr^{\tau}(\phi : e) = 1$ iff. for all $\mathcal{I}$ such that $Pr^{\tau}(\mathcal{I}) > 0$, $\mathcal{I} \models \phi : e$ (i.e. $\forall v \in D(V) - U(K)$ s.t. $Pr_{BN}(V = v) > 0$, $\tau(v) \models \phi : e$).

We now prove that $\forall v \in D(V) - U(K)$, $\tau(v) \models \phi : e$. Let $v \in D(V) - U(K)$. We consider the following two cases:

- Case 1: $v \models e$. In this case, by definition of K_v, ϕ must be an axiom of K_v. Since, by definition of an annotated interpretation mapping choice, $\tau(v)$ is a model of K_v, it follows that $\tau(v) \models \phi$. This establishes that $\tau(v) \models \phi : e$
- Case 2: $v \models e$ does not hold. In this case, by definition of the implication semantics, $\tau(v) \models \phi : e$ trivially holds.

This complete the proof that condition (i) of Definition 5 is satisfied.

We have shown that the canonical interpretation Pr^{τ} satisfies all the conditions of Definition 5. It is therefore a model of K to the degree $d = 1 - DU(K)$.

Now, we are ready to prove that both directions of the equivalence of Theorem 1 hold.

Appendix B.2. Proof of Theorem 1: Necessary Condition

In this section, we establish the following: (K is consistent to the degree d) $\Rightarrow (d = 1 - \sum_{v \ s.t. \ K_v \ inconsistent} Pr_{BN}(V = v))$.

Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base consistent to the degree d. By condition (iii) of Definition 5, $d = 1 - DU(K)$. Since $DU(K) = \sum_{v \ s.t. \ K_v \ inconsistent} Pr_{BN}(V = v)$, it follows that $d = 1 - \sum_{v \ s.t. \ K_v \ inconsistent} Pr_{BN}(V = v)$.

Appendix B.3. Proof of Theorem 1: Sufficient Condition

In this section, we establish the following: $(d = 1 - \sum_{v \ s.t. \ K_v \ inconsistent} Pr_{BN}(V = v))$ $\Rightarrow$ (K is consistent to the degree d).

Let d be such that $d = 1 - \sum_{v \ s.t. \ K_v \ inconsistent} Pr_{BN}(V = v)$. Then, it follows that $d = 1 - DU(K)$ (or $DU(K) = 1 - d$).

If $DU(K) \neq 1$, as already observed, a direct consequence of Lemma 1 is that at least one annotated interpretation mapping choice exists. Let τ be an annotated interpretation mapping choice for K. According to Lemma 3, the canonical probabilistic interpretation Pr^{τ} of K given τ is a model of K to the degree $d = 1 - DU(K)$, which establishes that K is consistent to the degree d.

If $DU(K) = 1$, by Definition 5, K is consistent to the degree $d = 0$ (i.e. $d = 1 - DU(K)$).

Appendix C. Proof of Lemma 2

Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base consistent to the degree d ($d \neq 0$). Let $Q_g = \psi : e$ be a grounded query. Let $\Omega = \{v | K_v \textit{ is consistent and } K_v \models \psi \textit{ and } v \models e\}$.

There are two important steps in the proof of Lemma 2:

1. First, we need to show that for every probabilistic model Pr of K, $Pr(\psi : e) \geq \frac{1}{d} \sum_{v \in \Omega} Pr_{BN}(v)$
2. Next, we show that there exists a special probabilistic model Pr^0 of K such that $Pr^0(\psi : e) = \frac{1}{d} \sum_{v \in \Omega} Pr_{BN}(v)$

Together, these two steps prove that Lemma 2 holds.

Appendix C.1. $\forall Pr \text{ s.t. } Pr \models K,\ Pr(\psi : e) \geq \frac{1}{d} \sum_{v \in \Omega} Pr_{BN}(v)$

Let Pr be a model of K. By Definition 9,

$$Pr(\psi : e) = \sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}} \models e \text{ and } \mathcal{I} \models \psi} Pr(\mathcal{I})$$

By reordering the terms of the sum by grouping together annotated interpretations assigning the same value to V, we have the following:

$$Pr(\psi : e) = \sum_{(v \text{ s.t. } v \models e)} \sum_{(\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v \text{ and } \mathcal{I} \models \psi)} Pr(\mathcal{I})$$

Adding the condition $Pr(\mathcal{I}) \neq 0$ obviously does not change the previous equality, which leads to the following equality, hereafter referred to as (I):

$$Pr(\psi : e) =$$

$$\sum_{(v \text{ s.t. } v \models e)} \sum_{(\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v \text{ and } \mathcal{I} \models \psi \text{ and } Pr(\mathcal{I}) \neq 0)} Pr(\mathcal{I})$$

We obtain the following inequality, hereafter referred to as (II), by restricting the terms of the outer summation to v s.t. K_v is consistent and $K_v \models \psi$:

$$Pr(\psi : e) \geq$$

$$\sum_{(v \text{ s.t. } v \models e \text{ and } K_v \text{ consistent and } K_v \models \psi)} S(v, \psi)$$

where

$$S(v, \psi) = \sum_{(\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v \text{ and } \mathcal{I} \models \psi \text{ and } Pr(\mathcal{I}) \neq 0)} Pr(\mathcal{I})$$

Next, we show that for v such that $v \models e$ and K_v is consistent and $K_v \models \psi$, the following holds:

$$\sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v} Pr(\mathcal{I}) =$$

$$\sum_{(\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v \text{ and } \mathcal{I} \models \psi \text{ and } Pr(\mathcal{I}) \neq 0)} Pr(\mathcal{I})$$

To simplify the notation, we define P_v as $P_v = \sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v} Pr^{\tau}(\mathcal{I})$
We obviously have: $P_v = \sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}} = v \text{ and } Pr(\mathcal{I}) \neq 0} Pr(\mathcal{I})$

Let v be such that $v \models e$ and K_v is consistent and $K_v \models \psi$. Let $\mathcal{I}$ be an annotated interpretation of K such that $V^{\mathcal{I}} = v$ and $Pr(\mathcal{I}) \neq 0$. Since $Pr(\mathcal{I}) \neq 0$ and Pr is a model, it follows that $\mathcal{I}$ is a model of all annotated axioms (this is a direct consequence of $Pr(\phi : e') = 1$ for all $\phi : e'$ axioms of K), which means that $\mathcal{I}$ is a model of K_v. Since $K_v \models \psi$, it follows that $\mathcal{I} \models \psi$. This result demonstrates that adding the condition $\mathcal{I} \models \psi$ in the following summation does not change it for v such that $v \models e$ and K_v is consistent and $K_v \models \psi$:

$$P_v = \sum_{\mathcal{I} \text{ s.t. } V^{\mathcal{I}}=v \text{ and } Pr(\mathcal{I})\neq 0} Pr(\mathcal{I})$$

Therefore, for v such that $v \models e$ and K_v is consistent and $K_v \models \psi$,

$$P_v = \sum_{(\mathcal{I} \text{ s.t. } V^{\mathcal{I}}=v \text{ and } \mathcal{I}\models\psi \text{ and } Pr(\mathcal{I})\neq 0)} Pr(\mathcal{I})$$

The inequality (II) becomes:

$$Pr(\psi : e) \geq \sum_{(v \text{ s.t. } v\models e \text{ and } K_v \text{ consistent and } K_v\models\psi)} P_v$$

By definition of a model of Pr, for $v \notin U(K)$ (i.e. K_v is satifiable according to Lemma 1), $P_v = \frac{1}{d} Pr_{BN}(V = v)$.

Hence,

$$Pr(\psi : e) \geq \frac{1}{d} \sum_{v\in\Omega} Pr_{BN}(v)$$

This complete the first step of proof of Lemma 2.

Appendix C.2. $\exists Pr^0$ *s.t.* $Pr^0 \models K$ *and* $Pr^0(\psi : e) = \frac{1}{d}\sum_{v\in\Omega} Pr_{BN}(v)$

Now, we show that there exists a special probabilistic model Pr^0 of K such that $Pr^0(\psi : e) = \frac{1}{d}\sum_{v\in\Omega} Pr_{BN}(v)$.

Let τ^0 be an annotated interpretation mapping choice defined as follows: For $v \in D(V) - U(K)$,

- If v is such that $K_v \models \psi$, then, for $\tau^0(v)$, we choose a classical model $\mathcal{I}$ of K_v and extend it to become an annotated interpretation of K by assigning $V^{\mathcal{I}} = v$. $\tau^0(v)$ is set to the resulting extension. By definition of logical entailment, $\tau^0(v) \models \psi$ obviously holds.

- If v is such that $K_v \models \psi$ does not hold, then, by definition of logical entailment, there exists at least one classical model $\mathcal{I}$ of K_v such that $\mathcal{I}$ does not satisfy ψ. We extend $\mathcal{I}$ to become an annotated interpretation of K by assignining $V^{\mathcal{I}} = v$. $\tau^0(v)$ is chosen to be the resulting extension. So, $\tau^0(v)$ does not satisfy ψ

Now, we consider the canonical model Pr^{τ^0} given τ^0.
By equality (I) in section Appendix C.1,

$$Pr^{\tau^0}(\psi : e) =$$

$$\sum_{(v \ s.t. \ v \models e)} \ \sum_{(\mathcal{I} \ s.t. \ V^{\mathcal{I}} = v \ and \ \mathcal{I} \models \psi \ and \ Pr(\mathcal{I}) \neq 0)} Pr^{\tau^0}(\mathcal{I})$$

By defintion of Pr^{τ^0} and τ^0,

$$Pr^{\tau^0}(\psi : e) =$$

$$\sum_{(v \ s.t. \ v \models e \ and \ K_v \ consistent \ and \ K_v \models \psi)} Pr^{\tau^0}(\tau^0(v))$$

By definition of Pr^{τ^0}:

$$Pr^{\tau^0}(\psi : e) = \frac{1}{d} \sum_{v \in \Omega} Pr_{BN}(v)$$

So Pr^{τ^0} is that special model we were looking for.

Appendix D. Proof of Theorem 3

Let $K = (\mathcal{T}, \mathcal{A}, BN)$ be a probabilistic knowledge base consistent to the degree d, and let $Q = \psi : e$ be a probabilistic query against K. Let θ be a ground substitution for the variables in Q and let $pr \ in [0, 1]$.

First, let us consider the case where $d \neq 0$.

By definition, (θ, pr) is a solution to Q in K iff.

$$pr =$$

$$(1 - DU(K)) \times \inf\{Pr(\psi\theta : e) | Pr \ is \ model \ of \ K\}$$
$$+ \sum_{v \in U(K) \ and \ v \models e} Pr_{BN}(V = v)$$

So, by Lemma 2, the following equivalence (hereafter referred to as (A)) must hold: (θ, pr) is a solution to Q in K iff.

$$pr =$$

$$\delta \times \sum_{v \ s.t. \ K_v \models \psi\theta \ and \ v \models e \ and \ K_v \ consistent} Pr_{BN}(V = v)$$
$$+ \sum_{v \in U(K) \ and \ v \models e} Pr_{BN}(V = v)$$

where

$$\delta = \frac{(1 - DU(K))}{d}$$

Since, by Lemma 1, $v \in U(K) \Leftrightarrow K_v$ is inconsistent, and an inconsistent classical KB entails everything, the following equality, hereafter referred to as (B), holds:

$$\sum_{v \in U(K) \ and \ v \models e} Pr_{BN}(V = v) =$$

$$\sum_{v \in U(K) \ and \ v \models e \ and \ K_v \models \psi\theta} Pr_{BN}(V = v)$$

Finally, by definition of partial consistency of K, $d = 1 - DU(K)$ (this equality is hereafter referred to as (C))

From (A), (B) and (C), the following holds: (θ, pr) is a solution to Q in K iff.

$$pr =$$

$$\sum_{v \ s.t. \ K_v \models \psi\theta \ and \ v \models e \ and \ K_v \ consistent} Pr_{BN}(V = v)$$

$$+ \sum_{v \in U(K) \ and \ v \models e \ and \ K_v \models \psi\theta} Pr_{BN}(V = v)$$

Since, by Lemma 1, $v \notin U(K) \Leftrightarrow K_v$ is consistent, it follows that: (θ, pr) is a solution to Q in K iff.

$$pr =$$

$$\sum_{v \ s.t. \ K_v \models \psi\theta \ and \ v \models e \ and \ v \notin U(K)} Pr_{BN}(V = v)$$

$$+ \sum_{v \in U(K) \ and \ v \models e \ and \ K_v \models \psi\theta} Pr_{BN}(V = v)$$

which means that : (θ, pr) is a solution to Q in K iff.

$$pr = \sum_{v \models e \ and \ K_v \models \psi\theta} Pr_{BN}(V = v)$$

Now, for the case where $d = 0$. By definition, (θ, pr) is a solution to Q in K iff.

$$pr = \sum_{v \models e \ and \ v \in U(K)} Pr_{BN}(V = v) + 0$$

Since $v \in U(K)$ iff. K_v is inconsistent and an inconsistent KB entails everyhting, (θ, pr) is a solution to Q in K iff.

$$pr = \sum_{v \ s.t. \ v \models e \ and \ v \in U(K) \ and \ K_v \models \psi\theta} Pr_{BN}(V = v)$$

Since $U(K) = D(V)$ for $d = 0$, it follows that (θ, pr) is a solution to Q in K iff.

$$pr = \sum_{v \ s.t. \ v \models e \ and \ K_v \models \psi\theta} Pr_{BN}(V = v)$$