Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Sister Conferences Best Papers Track

Shapley Value Computation in Ontology-Mediated Query Answering (Extended Abstract*)

Meghyn Bienvenu[1] , **Diego Figueira**[1] , **Pierre Lafourcade**[1]
[1] Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France
{meghyn.bienvenu, diego.figueira, pierre.lafourcade}@u-bordeaux.fr

Abstract

In this work, we explore the use of the Shapley value in ontology-mediated query answering (OMQA) and provide a detailed complexity analysis of Shapley value computation (SVC) in the OMQA setting. In particular, we establish a FP/#P-hard dichotomy for SVC for ontology-mediated queries $(\mathcal{T}, q)$ composed of an ontology $\mathcal{T}$ formulated in the description logic $\mathcal{ELHI}_\bot$ and a connected constant-free homomorphism-closed query q. We further strengthen the #P-hardness side of the dichotomy to cover possibly disconnected queries with constants. Our results exploit recently discovered connections between SVC and probabilistic query evaluation and allow us to generalize existing results on probabilistic OMQA.

1 Introduction

The Shapley value [Shapley, 1953] was originally proposed in the context of cooperative game theory as a method for fairly distributing the wealth of a coalition of players based upon their respective contributions, but has since found application in numerous domains. In artificial intelligence, the Shapley value has been utilized for defining inconsistency measures of propositional [Grant and Hunter, 2006; Hunter and Konieczny, 2010] and description logic knowledge bases [Deng *et al.*, 2007], and more recently for defining explanations of machine learning models [Lundberg and Lee, 2017]. The Shapley value has also gained attention in the database area [Bertossi *et al.*, 2023], where it has been employed both for defining inconsistency values of databases [Livshits and Kimelfeld, 2022] and also for providing quantitative explanations of query answers [Livshits *et al.*, 2021].

In general, Shapley value computation is known to be computationally challenging, being #P-hard in data complexity for common classes of queries, like conjunctive queries (CQs). This has motivated non-uniform complexity studies aimed at pinpointing which queries admit tractable Shapley value computation [Reshef *et al.*, 2020; Khalil and Kimelfeld, 2023], in particular, by establishing connections with probabilistic query evaluation and model counting [Deutch *et al.*, 2022; Kara *et al.*, 2024; Bienvenu *et al.*, 2024b].

In the present paper, we revisit the use of the Shapley value in the ontology setting, building upon these recent advances in the database area. We shall mostly focus on how the Shapley value can be employed for explaining answers in the context of ontology-mediated query answering (OMQA) [Poggi *et al.*, 2008; Calì *et al.*, 2011; Mugnier and Thomazo, 2014; Bienvenu and Ortiz, 2015; Xiao *et al.*, 2018]. We recall that the OMQA is used to improve access to incomplete and possibly heterogeneous data through the addition of ontology layer, which provides a user-friendly vocabulary for query formulation as well as domain knowledge that is taken into account when computing the query answers. With the increasing maturity and deployment of OMQA techniques, there is an acknowledged need to help users understand the query results. Various notions of explanations with different levels of detail can be considered for OMQA, ranging from providing proofs of how an answer can be derived [Borgida *et al.*, 2008; Alrabbaa *et al.*, 2022] to generating minimal subsets of the KB that suffice to obtain the answer or identifying the assertions and/or axioms that are relevant in the sense that they belong to such a minimal subset [Bienvenu *et al.*, 2019; Ceylan *et al.*, 2019; Ceylan *et al.*, 2020]. The Shapley value offers a more nuanced, quantitative version of the latter approach, by assigning the relevant assertions and axioms scores based upon their level of responsibility or importance in obtaining the considered query answer (or entailment).

For our study of Shapley value computation, we will work with description logic (DL) knowledge bases (KB), consisting of an ABox (dataset) and TBox (ontology). We introduce some natural ways of defining the Shapley value computation (SVC) problem in the DL setting, by varying what is to be explained (entailment of a TBox axiom, ABox assertion, or query answer), which parts of the KB are assigned values, and how the complexity is measured. To begin our study, we establish the #P-hardness of the Shapley value computation of a simple graph reachability query, which we then employ to show #P-hardness of several variants of the SVC problem, even for inexpressive DLs and atomic queries. In light of these initial negative results, we embark on a non-uniform complexity analysis, whose aim is to classify the data complexity of the Shapley value computation problems

*Extended abstract of [Bienvenu *et al.*, 2024a] from KR 2024.

SVC_Q associated with each ontology-mediated query (OMQ) $Q = (\mathcal{T}, q)$. By transferring recent results from the database setting, we establish a FP/#P-hard dichotomy result of Shapley value computation problem SVC_Q for OMQs $Q = (\mathcal{T}, q)$ where the TBox $\mathcal{T}$ is formulated in the Horn DL $\mathcal{ELHI}_\bot$ and q is a constant-free connected homomorphism-closed query. Our final and most technically challenging result shows that the #P-hardness part of the dichotomy can be strengthened to cover OMQs based upon a wider range of queries q. Specifically, we show that for any OMQ $Q = (\mathcal{T}, q)$ based upon a $\mathcal{ELHI}_\bot$ TBox and a UCQ q (which may be disconnected and/or contain constants), non-FO-rewritability of Q implies #P-hardness of SVC_Q. Due to the tight connections holding between SVC and probabilistic query evaluation, the proof of this result can be further used to obtain a FP/#P-hard dichotomy for probabilistic ontology-mediated queries from $(\mathcal{ELHI}_\bot, \mathrm{UCQ})$, substantially generalizing existing results.

2 Preliminaries

We very briefly recall here some key notions and terminology, referring readers to the full paper for more details.

Databases and Queries A *database* $\mathcal{D}$ is a finite set of *facts* $P(\vec{a})$, with P a k-ary relation symbol and $\vec{a}$ a k-ary tuple of *constants*. Results are formulated for *Boolean queries*, which evaluate to true or false in a given database (or first-order interpretation), with $\mathcal{D} \models q$ indicating q is true in $\mathcal{D}$. We consider well-known query classes like *conjunctive queries* (CQs) and *unions of conjunctive queries* (UCQs), given by (disjunctions of) existentially quantified conjunctions of relational atoms. We also consider *hom-closed* queries, i.e. q such that $\mathcal{D} \models q$ implies $\mathcal{D}' \models q$ for every $\mathcal{D}'$ that maps homomorphically into $\mathcal{D}$, as well as *connected* queries.

Knowledge Bases We assume that readers are familiar with the basics of description logics (DLs) [Baader *et al.*, 2017]. A DL *knowledge base* (*KB*) $\mathcal{K} = (\mathcal{A}, \mathcal{T})$ comprises an ABox (dataset) $\mathcal{A}$ and a TBox (ontology) $\mathcal{T}$, which are built from *concept names* (unary relations) and *role names* (binary relations) and individual constants. An *ABox* is a database whose facts take the form of *concept assertions* $A(c)$ and *role assertions* $r(c, d)$. A *TBox* is a finite set of *axioms*, whose form is dictated by the *DL* in question. We use $\mathcal{L}$ *TBox* to refer to a TBox formulated in the DL $\mathcal{L}$. Of particular interest to the present paper are the well-known lightweight DLs $\mathcal{EL}$ and DL-Lite$_{\mathsf{core}}$ and their common extension, the Horn DL $\mathcal{ELHI}_\bot$, see e.g. [Bienvenu and Ortiz, 2015]. The semantics of DL KBs is given in terms of first-order logic interpretations, with each satisfiable KB giving rise to a set of models.

Ontology-Mediated Query Answering We say a Boolean query q is entailed from a DL KB $\mathcal{K}$ if $\mathcal{I} \models q$ for every model $\mathcal{I}$ of $\mathcal{K}$. We may alternatively group $\mathcal{T}$ and q into an *ontology-mediated query* (*OMQ*) $Q = (\mathcal{T}, q)$, which is posed to the ABox $\mathcal{A}$, in which case we will write $\mathcal{A} \models Q$ to mean $(\mathcal{A}, \mathcal{T}) \models q$. We will use the notation $(\mathcal{L}, \mathcal{Q})$ to designate the class of all OMQs $(\mathcal{T}, q)$ such that $\mathcal{T}$ is formulated in the DL $\mathcal{L}$ and q is a query from the class of queries $\mathcal{Q}$.

First-order (FO) rewriting is a prominent OMQA technique which rewrites the input OMQ into another query that can be directly evaluated by a database system. Formally, a Boolean FO-query q^* is an *FO-rewriting of Boolean OMQ* $Q = (\mathcal{T}, q)$ if for every ABox $\mathcal{A}$: $\mathcal{A} \models (\mathcal{T}, q)$ iff $\mathcal{A} \models q^*$. If we modify the definition to only quantify over ABoxes that are satisfiable w.r.t. $\mathcal{T}$, we speak instead of a *rewriting w.r.t. consistent ABoxes*. If an OMQ Q possesses an FO-rewriting, we say that Q is *FO-rewritable*, else it is called *non-FO-rewritable*.

3 Shapley Value

We briefly recall the definition of Shapley value in games and how it has been employed in the database setting, before illustrating its application in the ontology setting.

3.1 Shapley Value in Cooperative Games

The Shapley value [Shapley, 1953] was introduced as a means to fairly distribute wealth amongst players in a a cooperative game, based upon their respective contributions. A *cooperative game* consists of a finite set of players P and a *wealth function* $\mathbf{v} : \wp(P) \to \mathbb{Q}$ that assigns a value to each coalition (*i.e.*, set) of players, with $\mathbf{v}(\emptyset) = 0$. Picture a scenario where the players arrive one by one in a random order, and each one earns what she added to the current coalition's wealth on arrival. The *Shapley value* of a player $p \in P$ is defined as her expected earnings in this scenario, which can be expressed as:

$$\mathrm{Sh}(P, \mathbf{v}, p) := \frac{1}{|P|!} \sum_{\sigma \in \mathfrak{S}(P)} (\mathbf{v}(\sigma_{\leqslant p}) - \mathbf{v}(\sigma_{<p})) \qquad (1)$$

where $\mathfrak{S}(P)$ denotes the set of permutations of P and $\sigma_{<p}$ (resp. $\sigma_{\leqslant p}$) the set of players that appear before p (resp. before or at p) in the permutation σ. Intuitively, we take the average marginal contribution $\mathbf{v}(\sigma_{\leqslant p}) - \mathbf{v}(\sigma_{<p})$ of p, across all possible orderings σ of the players.

3.2 Shapley Value for Database Queries

There has been significant interest lately in the problem of computing the Shapley value of database facts as a means of quantifying their contributions to a query answer. The formal setting is as follows: the database $\mathcal{D}$ is *partitioned* into *endogenous* and *exogenous facts*, $\mathcal{D} = \mathcal{D}_\mathsf{n} \uplus \mathcal{D}_\mathsf{x}$, the players of the cooperative game are the endogenous facts $\mathcal{D}_\mathsf{n}$, and the wealth function of a given Boolean query q is defined, for every subset $B \subseteq \mathcal{D}_\mathsf{n}$ of endogenous facts, as $\mathbf{v}_q(B) = v_B - v_\mathsf{x}$ where $v_B = 1$ (resp. $v_\mathsf{x} = 1$) if $\mathcal{D}_\mathsf{x} \cup B \models q$ (resp. if $\mathcal{D}_\mathsf{x} \models q$), and 0 otherwise. We use SVC_q to denote the problem of computing the Shapley value $\mathrm{Sh}(\mathcal{D}_\mathsf{n}, \mathbf{v}_q, \alpha)$ for the input partitioned database $\mathcal{D}_\mathsf{n} \uplus \mathcal{D}_\mathsf{x}$ and fact $\alpha \in \mathcal{D}_\mathsf{n}$.

In [Livshits *et al.*, 2021], a FP/#P-hard dichotomy was established for the class of *self-join-free* CQs (i.e. not having two atoms with the same relation name), showing that for every such query q, SVC_q is either tractable (in FP) or #P-hard. Interestingly, the dichotomy coincides with the FP/#P-hard dichotomy previously established for probabilistic query evaluation (PQE) [Dalvi and Suciu, 2004]. We recall that the problem PQE_q is to compute, given a tuple-independent probabilistic database $\mathcal{D} = (S, \pi)$ (consisting of a database S and probability assignment $\pi : S \to (0, 1]$), the probability of Boolean query q being true, where each assertion

$\alpha \in S$ has independent probability $\pi(\alpha)$ of being present in the database. The PQE dichotomy extends to the more general class of UCQs [Dalvi and Suciu, 2012], where the queries for which PQE is tractable are known as *safe* UCQs. However, it is an open problem whether UCQs (or even CQs with self-joins) also enjoy a dichotomy for SVC.

Some recent results have helped clarify the relation between the two dichotomies. First, [Deutch *et al.*, 2022] showed how to reduce SVC to PQE. More recently, it has been shown that SVC and SPPQE (a restriction of PQE to probabilistic databases where $\pi(\alpha) \in \{p, 1\}$ for a single $p \in (0, 1)$) are polynomial-time inter-reducible for many fragments of hom-closed queries [Bienvenu *et al.*, 2024b], in particular for connected queries without constants. This yielded a much broader dichotomy result for the class of graph databases (i.e. having only unary and binary relations):

Theorem 1. *[Bienvenu* et al.*, 2024b, Corollaries 4.1 and 4.2] For every connected hom-closed Boolean query q,* $\mathrm{SPPQE}_q \equiv_{\mathsf{P}} \mathrm{SVC}_q$*; further, on graph databases,* SVC_q *is in* FP *if q is equivalent to a safe UCQ and* $\#\mathsf{P}$*-hard otherwise.*

3.3 Shapley Value in the Ontology Setting

We present a running example which illustrates the use of the Shapley value in the ontology setting.

Example 2. *Consider the $\mathcal{ELHI}_\perp$ KB defined in Figure 1, where the ABox (bottom half) contains information on some ingredients and recipes, and the TBox (top half) defines more complex notions such as a 'land-sea recipe'. For instance,* $\exists\mathsf{HasIngr}.\mathsf{FishBased} \sqsubseteq \mathsf{FishBased}$ *intuitively translates as 'anything that has a fish-based ingredient is fish-based'.*

A user of this KB might obtain poulardeNantua as an answer to the query $\mathsf{LandSea}(x)$*, and wonder which ingredients are the most responsible for this fact. She can thus set everything but the role assertions (which specify ingredients) as exogenous and compute the Shapley values for the Boolean query* $\mathsf{LandSea}(poulardeNantua)$.

We can compute the values via Eq. (1)*. There are $4! = 24$ possible permutations over the 4 endogenous role assertions $\{e_1, e_2, e_3, e_4\}$. 14 out of 24 permutations are s.t. $\mathbf{v}(\sigma_{\leqslant e_1}) - \mathbf{v}(\sigma_{<e_1}) = 1$, and similarly 6 for e_2, 2 for e_3, and 2 for e_4, making the respective Shapley values: 14/24, 6/24, 2/24 and 2/24.*

As expected, e_1 has the highest responsibility because it is necessary to satisfy the query, then comes e_2 that only needs to be combined with e_1 and finally e_3 and e_4 that must be used together in combination with e_1.

As we shall see in Section 4, there are several different Shapley value computation tasks that can be considered in the ontology setting. Indeed, we may vary the reasoning task (e.g. query answering or axiom entailment), which parts of the knowledge base are assigned Shapley values (only assertions, only axioms, or both?), how the complexity is measured (is the query fixed or part of the input?), and the languages used to formulate the TBox and the queries.

4 Complexity Results for Ontology Setting

In this section, we briefly present the results of our complexity analysis of different Shapley value computation tasks.

$\mathsf{FishBased} \sqcap \mathsf{MeatBased} \sqsubseteq \mathsf{LandSea}$	$\mathsf{Fish} \sqsubseteq \mathsf{FishBased}$
$\exists\mathsf{HasIngr}.\mathsf{FishBased} \sqsubseteq \mathsf{FishBased}$	$\mathsf{Seafood} \sqsubseteq \mathsf{FishBased}$
$\exists\mathsf{HasIngr}.\mathsf{MeatBased} \sqsubseteq \mathsf{MeatBased}$	$\mathsf{Crustacean} \sqsubseteq \mathsf{SeaFood}$
$\mathsf{HasSauce} \sqsubseteq \mathsf{HasIngr}$	$\mathsf{Meat} \sqsubseteq \mathsf{MeatBased}$

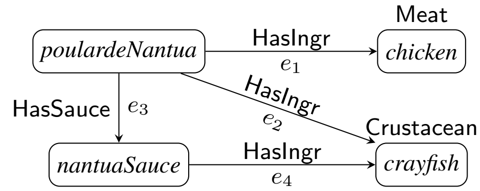

Figure 1: An example KB, with data and knowledge about a recipe from [Escoffier, 1903]. The arrows represent role assertions and labels on top of boxes (*e.g.* Meat) represent concept assertions.

4.1 Initial Intractability Results

A first application of the Shapley value to ontologies is to focus solely on the TBox and determine which axioms are most responsible for a given TBox entailment, *e.g.* to find out why crustaceans count as fish-based. Unfortunately, we can show $\#\mathsf{P}$-hardness already for the simplest possible DL $\mathcal{L}_{\min}$ containing only concept name inclusions. The proof relies upon first establishing $\#\mathsf{P}$-hardness of a Shapley value computation task involving reachability in directed graphs, then using axiom entailment to encode reachability.

Proposition 3. *For every TBox $\mathcal{T}$ and $S \subseteq \mathcal{T}$ and pair of concept names A, B, define $\mathbf{v}_{A\sqsubseteq B}(S) := 1$ if $S \models A \sqsubseteq B$, and 0 otherwise. The problem of computing, given a $\mathcal{L}_{\min}$ TBox $\mathcal{T}$, concept names A, B, and an axioms $\mu \in \mathcal{T}$, the Shapley value of μ in the game $(\mathcal{T}, \mathbf{v}_{A\sqsubseteq B})$ is $\#\mathsf{P}$-hard.*

We next consider how to use the Shapley value to explain query entailment *w.r.t.* a DL KB. In our example, a user may ask why there is a fish-based recipe in the KB of Figure 1 if there is no Fish assertion. She could therefore compute the Shapley values for the CQ $\exists x.\mathsf{FishBased}(x)$ and discover that the *crayfish* is considered fish-based because of the assertion Crustacean(*crayfish*) and the axioms $\mathsf{Crustacean} \sqsubseteq \mathsf{SeaFood} \sqsubseteq \mathsf{FishBased}$. The input will now be a *partitioned KB* $\mathcal{K} = (\mathcal{A}_\mathsf{n} \uplus \mathcal{A}_\mathsf{x}, \mathcal{T}_\mathsf{n} \uplus \mathcal{T}_\mathsf{x})$ and the task is to compute the Shapley value of the statements in $\mathcal{A}_\mathsf{n} \cup \mathcal{T}_\mathsf{n}$, in order to quantify their responsibility in $\mathcal{K}$ entailing q. Concretely, for any Boolean query q we consider the cooperative game with players $\mathcal{A}_\mathsf{n} \cup \mathcal{T}_\mathsf{n}$ and $\mathbf{v}_q(S) := v_S - v_\mathsf{x}$ where $v_S = 1$ (resp. $v_\mathsf{x} = 1$) if $(\mathcal{A}_\mathsf{x} \cup (S \cap \mathcal{A}_\mathsf{n}), \mathcal{T}_\mathsf{x} \cup (S \cap \mathcal{T}_\mathsf{n})) \models q$ (resp. if $(\mathcal{A}_\mathsf{x}, \mathcal{T}_\mathsf{x}) \models q$), and 0 otherwise.

Proposition 4. *The problem of computing Shapley values for Boolean CQs over partitioned KBs on $\mathcal{L}_{\min}$ is $\#\mathsf{P}$-hard. Hardness holds even for queries given as ABox assertions.*

Propositions 3 and 4 show that computing Shapley values of axioms is inevitably intractable, as reasoning on concept and role inclusions naturally involves reachability. One idea to sidestep this issue would be to treat the whole TBox as exogenous. Conceptually, this corresponds to treating TBox axioms as given or obvious, thereby focusing on explanations in terms of the ABox assertions. In our running example of

Figure 1, the user may wonder what ingredients make the Poularde Nantua recipe fish-based, leading her to compute Shapley values for the CQ $\mathsf{FishBased}(\mathit{poulardeNantua})$, setting everything but the $\mathsf{HasIngr}$-roles as exogenous.

In terms of complexity, the problem will once again be $\#\mathsf{P}$-hard due to reachability, which this time is expressed within the data itself. The source of difficulty are axioms such as $\exists\mathsf{HasIngr}.\mathsf{FishBased} \sqsubseteq \mathsf{FishBased}$, which can be found in any DL at least as expressive as $\mathcal{EL}$.

Proposition 5. *Let q be a CQ of the form $A(c)$, for $A \in \mathsf{N_C}$ and $c \in \mathsf{N_I}$. Then the problem of computing Shapley values for q over partitioned KBs on $\mathcal{EL}$ is $\#\mathsf{P}$-hard, even if we assume that the TBox only contains exogenous axioms.*

4.2 A Dichotomy for OMQs in $\mathcal{ELHI}_\bot$

The results of Section 4.1 show that allowing TBoxes to be part of the input makes the Shapley value computation problems $\#\mathsf{P}$-hard. This suggests the interest of analyzing the complexity of Shapley value computation at the level of individual *ontology-mediated queries* $Q = (\mathcal{T}, q)$, only taking data as input. Such a non-uniform approach to complexity analysis has previously been undertaken for several OMQA settings, and in particular in the context of probabilistic OMQA [Jung and Lutz, 2012]. Moreover, this perspective aligns nicely with the formulation of Shapley value computation for database queries, enabling transfer of results.

In order to be able to apply Theorem 1 in a ‘black-box’ fashion, we need to identify a class of OMQs that is hom-closed and connected. To this end, we prove the following lemma, which shows that the addition of an $\mathcal{ELHI}_\bot$ ontology preserves the connectedness of hom-closed queries.

Lemma 6. *Let q be a connected C-hom-closed[1] query and $\mathcal{T}$ an $\mathcal{ELHI}_\bot$ ontology. Then the OMQ $Q := (\mathcal{T}, q)$ is a connected C-hom-closed query.*

Lemma 6 allows us to exploit the equivalence of SVC with probabilistic database evaluation from [Bienvenu *et al.*, 2024b] (recalled in Theorem 1) to obtain a dichotomy for OMQs whose base query is constant-free and connected.

Theorem 7. *For every connected (constant-free) hom-closed query q and $\mathcal{ELHI}_\bot$ ontology $\mathcal{T}$, we have $\mathrm{SVC}_{(\mathcal{T},q)} \equiv_{\mathsf{P}} \mathrm{SPPQE}_{(\mathcal{T},q)}$. Further, the problem is in FP if the OMQ $(\mathcal{T}, q)$ can be rewritten into a safe UCQ and $\#\mathsf{P}$-hard otherwise.*

It is decidable whether or not a given UCQ is safe [Dalvi and Suciu, 2012, implicit]. It therefore follows that the dichotomy given by Theorem 7 is effective whenever first-order rewritability is decidable for the considered class of OMQs, and that a first-order rewriting can always be effectively computed when there exists one. This is in particular true for ($\mathcal{ELHI}_\bot$,CQ) [Bienvenu *et al.*, 2016, Theorem 5].

Furthermore, [Jung and Lutz, 2012, Theorem 5] gives a syntactic characterisation of which constant-free connected OMQs in $(\text{DL-Lite}_{\mathsf{core}}, CQ)$ are equivalent to safe UCQs. By Theorem 7, it also characterizes constant-free connected OMQs $Q \in (\text{DL-Lite}_{\mathsf{core}}, CQ)$ that are *s.t.* $\mathrm{SVC}_Q \in \mathsf{FP}$.

[1]To accommodate queries with constants (Theorems 8 and 9), we must consider here C-hom-closed queries, defined using C-homomorphisms, which fix the interpretation of the constants in C.

4.3 Strengthening the $\#\mathsf{P}$-Hardness Result

The dichotomy of Theorem 7 is limited in two respects:

- The result only covers connected constant-free queries.
- When an OMQ is seen as an abstract query, the distinction between consistent and inconsistent ABoxes is lost. However, one might be interested in explaining answers to a query only over consistent ABoxes.

When considering OMQs in $(\mathcal{ELHI}_\bot, \mathrm{UCQ})$, both of these points can be improved upon by carefully exploiting the properties of OMQs that are non-FO-rewritable.

Theorem 8. *Let q be a UCQ and $\mathcal{T}$ an $\mathcal{ELHI}_\bot$ ontology. If the OMQ $Q := (\mathcal{T}, q)$ is non-FO-rewritable w.r.t. consistent ABoxes, then SVC_Q and $\mathrm{PQE}_Q(1/2; 1)$[2] on consistent ABoxes are both $\#\mathsf{P}$-hard.*

As a consequence of Theorem 8, we obtain the following dichotomy for the probabilistic evaluation of OMQs, which extends [Jung and Lutz, 2012, Theorem 7] by generalizing from a subset of $(\mathcal{ELI}, CQ)$ to $(\mathcal{ELHI}_\bot, UCQ)$.

Theorem 9. *Let Q be a $(\mathcal{ELHI}_\bot, \mathrm{UCQ})$ OMQ. Then $\mathrm{PQE}_Q(1/2; 1)$, SPPQE_Q and PQE_Q are all in FP if Q is FO-rewritable into a safe UCQ, and $\#\mathsf{P}$-hard otherwise. Further, this dichotomy is effective if $Q \in (\mathcal{ELHI}_\bot, CQ)$.*

5 Conclusion & Discussion

This paper summarizes the results presented in [Bienvenu *et al.*, 2024a] on using Shapley values to provide quantitative explanations in the setting of ontology-mediated query answering. By exploiting recent results on SVC from the database setting, we obtained complexity dichotomies for OMQs whose TBoxes are formulated in the well-known Horn DL $\mathcal{ELHI}_\bot$. In particular, Theorem 7 identifies classes of OMQs for which the Shapley value can be computed in FP, while Theorem 8 provides a general $\#\mathsf{P}$-hardness result for non-FO-rewritable OMQs. Using the same techniques, and leveraging known connections between SVC and probabilistic query evaluation, we were further able to obtain a general dichotomy result for probabilistic OMQA (Theorem 9).

The dichotomy result for probabilistic OMQA is stronger than the one for SVC, which is to be expected as existing results on SVC in the pure database setting do not cover UCQs that are disconnected or with constants. It is an important open question whether there is a dichotomy for SVC_q for all Boolean UCQs q. Note that any progress on this question can be immediately transferred to the OMQA setting.

An interesting but challenging direction for future work is to study SVC for ontologies formulated using existential rules. We expect that the extension of our results to such ontologies will be non-trivial, due to the presence of higher-arity predicates and the lack of forest-shaped models. Indeed, the dichotomy for SPPQE that we exploited to obtain Theorem 7 is currently only known for arity 2 relations, and the proof of Theorem 8 relies upon the existence of a unique path between ‘anonymous’ elements, ensured by the forest-shaped structure of canonical models of Horn DL ontologies.

[2]$\mathrm{PQE}_Q(1/2; 1)$ is obtained by restricting PQE (or SPPQE) by only allowing 1 and $\frac{1}{2}$ as probability values.

Acknowledgements

This work was partially supported by ANR AI Chair INTENDED (ANR-19-CHIA-0014) and ANR project CQFD (ANR-18-CE23-0003).

References

[Alrabbaa *et al.*, 2022] Christian Alrabbaa, Stefan Borgwardt, Patrick Koopmann, and Alisa Kovtunova. Explaining ontology-mediated query answers using proofs over universal models. In *Proceedings of the International Joint Conference on Rules and Reasoning (RuleML+RR)*, pages 167–182, 2022.

[Baader *et al.*, 2017] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.

[Bertossi *et al.*, 2023] Leopoldo E. Bertossi, Benny Kimelfeld, Ester Livshits, and Mikaël Monet. The Shapley value in database management. *SIGMOD Record*, 52(2):6–17, 2023.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-Mediated Query Answering with Data-Tractable Description Logics. In *Tutorial Lectures of the Reasoning Web (RW) Summer School*, volume 9203 of *Lecture Notes in Computer Science*, pages 218–307. 2015.

[Bienvenu *et al.*, 2016] Meghyn Bienvenu, Peter Hansen, Carsten Lutz, and Frank Wolter. First order-rewritability and containment of conjunctive queries in Horn description logics. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 965–971, 2016.

[Bienvenu *et al.*, 2019] Meghyn Bienvenu, Camille Bourgaux, and François Goasdoué. Computing and explaining query answers over inconsistent DL-Lite knowledge bases. *Journal of Artificial Intelligence Research (JAIR)*, 64:563–644, 2019.

[Bienvenu *et al.*, 2024a] Meghyn Bienvenu, Diego Figueira, and Pierre Lafourcade. Shapley value computation in ontology-mediated query answering. In *Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning (KR)*, 2024.

[Bienvenu *et al.*, 2024b] Meghyn Bienvenu, Diego Figueira, and Pierre Lafourcade. When is Shapley value computation a matter of counting? In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, 2024.

[Borgida *et al.*, 2008] Alexander Borgida, Diego Calvanese, and Mariano Rodriguez-Muro. Explanation in the DL-Lite family of description logics. In *Proceedings of the International Conference: On the Move to Meaningful Internet Systems (OTM)*, pages 1440–1457, 2008.

[Calì *et al.*, 2011] Andrea Calì, Georg Gottlob, Thomas Lukasiewicz, and Andreas Pieris. A logical toolbox for ontological reasoning. *SIGMOD Record*, 40(3):5–14, 2011.

[Ceylan *et al.*, 2019] İsmail İlkan Ceylan, Thomas Lukasiewicz, Enrico Malizia, and Andrius Vaicenavicius. Explanations for query answers under existential rules. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1639–1646, 2019.

[Ceylan *et al.*, 2020] İsmail İlkan Ceylan, Thomas Lukasiewicz, Enrico Malizia, and Andrius Vaicenavicius. Explanations for ontology-mediated query answering in description logics. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 672–679, 2020.

[Dalvi and Suciu, 2004] Nilesh N. Dalvi and Dan Suciu. Efficient query evaluation on probabilistic databases. In *Proceedings of the International Conference on Very Large Data Bases (VLDB)*, pages 864–875, 2004.

[Dalvi and Suciu, 2012] Nilesh Dalvi and Dan Suciu. The dichotomy of probabilistic inference for unions of conjunctive queries. *Journal of the ACM*, 59(6):1–87, December 2012.

[Deng *et al.*, 2007] Xi Deng, Volker Haarslev, and Nematollaah Shiri. Measuring Inconsistencies in Ontologies. In *Proceedings of the European Semantic Web Conference*, pages 326–340, 2007.

[Deutch *et al.*, 2022] Daniel Deutch, Nave Frost, Benny Kimelfeld, and Mikaël Monet. Computing the Shapley value of facts in query answering. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 1570–1583, 2022.

[Escoffier, 1903] Auguste Escoffier. *Le guide culinaire : aide-mémoire de cuisine pratique*. Bibliothèque Professionnelle, Paris, 1903.

[Grant and Hunter, 2006] John Grant and Anthony Hunter. Measuring inconsistency in knowledge bases. *Journal of Intelligent Information Systems*, 27(2):159–184, 2006.

[Hunter and Konieczny, 2010] Anthony Hunter and Sébastien Konieczny. On the measure of conflicts: Shapley inconsistency values. *Artificial Intelligence (AIJ)*, 174(14):1007–1026, 2010.

[Jung and Lutz, 2012] Jean Christoph Jung and Carsten Lutz. Ontology-based access to probabilistic data with OWL QL. In *Proceedings of the International Semantic Web Conference (ISWC)*, pages 182–197, 2012.

[Kara *et al.*, 2024] Ahmet Kara, Dan Olteanu, and Dan Suciu. From Shapley value to model counting and back. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, 2024.

[Khalil and Kimelfeld, 2023] Majd Khalil and Benny Kimelfeld. The complexity of the Shapley value for regular path queries. In *Proceedings of the International Conference on Database Theory (ICDT)*, pages 11:1–11:19, 2023.

[Livshits and Kimelfeld, 2022] Ester Livshits and Benny Kimelfeld. The Shapley value of inconsistency measures for functional dependencies. *Logical Methods in Computer Science (LMCS)*, 18(2), 2022.

[Livshits *et al.*, 2021] Ester Livshits, Leopoldo Bertossi, Benny Kimelfeld, and Moshe Sebag. The Shapley value of tuples in query answering. *Logical Methods in Computer Science (LMCS)*, Volume 17, Issue 3:6942, 2021.

[Lundberg and Lee, 2017] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[Mugnier and Thomazo, 2014] Marie-Laure Mugnier and Michaël Thomazo. An introduction to ontology-based query answering with existential rules. In *Tutorial Lectures of the International Reasoning Web (RW) Summer School*, volume 8714 of *Lecture Notes in Computer Science (LNCS)*, pages 245–278. 2014.

[Poggi *et al.*, 2008] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. In *Journal on Data Semantics X*, pages 133–173, 2008.

[Reshef *et al.*, 2020] Alon Reshef, Benny Kimelfeld, and Ester Livshits. The impact of negation on the complexity of the Shapley value in conjunctive queries. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 285–297, 2020.

[Shapley, 1953] Lloyd S Shapley. A value for n-person games. In *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, 1953.

[Xiao *et al.*, 2018] Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-based data access: A survey. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5511–5519, 2018.