Tractable Responsibility Measures for Ontology-Mediated Query Answering (Extended Abstract)

Meghyn Bienvenu[1], Diego Figueira[1] and Pierre Lafourcade[1]

[1]Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France

Abstract

Recent work on quantitative approaches to explaining query answers employs responsibility measures to assign scores to facts in order to quantify their respective contributions to obtaining a given answer. This extended abstract summarizes our KR 2025 paper on the complexity of computing such responsibility scores in ontology-mediated query answering, focusing on a very recently introduced family of Shapley-value-based responsibility measures defined in terms of weighted sums of minimal supports.

Keywords

Ontology-mediated query answering, Responsibility measures, Quantitative explanations of query answers, Shapley value, Complexity analysis

The question of how to explain query answers has received significant attention in both the database and ontology settings. Qualitative notions of explanation, based e.g. on minimal supports or proofs, have been more extensively explored, in particular in the ontology setting, cf. [1, 2, 3, 4, 5]. However, there has been recent interest in quantitative notions of explanation based upon *responsibility measures*, which assign scores to the dataset facts to quantify their respective contributions to obtaining a given answer. Prior work on responsibility measures for query answers has predominantly focused on the so-called 'drastic Shapley value' [6, 7, 8, 9, 10, 11, 12, 13]. This measure, which originates from cooperative game theory, was motivated by the appealing theoretical characterization of the Shapley value as being the only wealth distribution mechanism respecting certain guarantees, known as 'Shapley axioms' [14].

Unfortunately, the computation of the drastic Shapley value is generally intractable (#P-hard in data complexity), even in the absence of ontologies and for very simple (conjunctive) queries [6, 11]. Furthermore, it has recently been argued in [15] that: (i) not all Shapley axioms yield desirable properties when translated into the query answering setting, and (ii) the genuinely desirable properties for responsibility measures of query answers do not pinpoint a single best score function. In light of this, [15] has very recently proposed a family of responsibility measures, based on *weighted sums of minimal supports* (*WSMS*), where the score of a fact is defined as a weighted sum of the sizes of the query's minimal supports containing it. The cited work shows that WSMS satisfy several desirable properties and that they enjoy more favourable computational properties compared to the drastic Shapley value in the database setting. Further, WSMS can also be defined as the Shapley value of suitable cooperative games.

The positive results for WSMS in the database setting motivate us to investigate the complexity

DL 2025: 38th International Workshop on Description Logics, September 3-6, 2025, Opole, Poland

CEUR Workshop Proceedings (CEUR-WS.org)

of computing WSMS responsibility scores in the more challenging setting of *ontology-mediated query answering* (*OMQA*) [16, 17, 18]. For this first study of WSMS in the OMQA setting, we focus on description logic (DL) ontologies [19], paying particular attention to DLs of the DL-Lite family [20], which are the most commonly adopted in OMQA, due to their favourable computational properties. We thus consider *ontology-mediated queries (OMQs)* of the form $(\mathcal{T}, q)$, where $\mathcal{T}$ is formulated in a DL and q is a conjunctive query (CQ) or atomic query (AQ).

In what follows, we introduce the considered responsibility measures and briefly summarize the obtained results. We assume familiarity with basic notions in databases, DLs, and OMQA.

Responsibility Measures for Query Answers

Although we shall be interested in employing responsibility measures to quantify the contribution of facts to obtaining an answer $\vec{a}$ to a query $q(\vec{x})$, it will actually be more convenient to consider the equivalent task of quantifying contributions to satisfying the associated Boolean query $q(\vec{a})$ (obtained by instantiating the free variables $\vec{x}$ of q with $\vec{a}$).

We shall further focus on *monotone* Boolean queries, defined in the database setting as queries q such that $\mathcal{D}_1 \models q \Rightarrow \mathcal{D}_2 \models q$ whenever $\mathcal{D}_1 \subseteq \mathcal{D}_2$. Such queries notably include the class of *homomorphism-closed* queries, which covers most well-known classes of OMQs. Note that a natural qualitative approach to explaining why a monotone Boolean query q holds in a database $\mathcal{D}$ is to consider the set $\mathsf{MS}_q(\mathcal{D})$ of *minimal supports* of q in $\mathcal{D}$, defined as the inclusion-minimal subsets $\mathcal{D}' \subseteq \mathcal{D}$ such that $\mathcal{D}' \models q$.

Our focus will be on providing *quantitative* explanations in the form of *responsibility measures*, which are functions that assign a score to every fact in the data, reflecting their contributions to making the query hold. Such measures have been formally defined, in the database setting, as ternary functions ϕ that take as input a database $\mathcal{D}$, a (Boolean) query q and a fact $\alpha \in \mathcal{D}$, and output a numerical value. As this definition is extremely permissive, [15, §4.1] identifies a set of desirable properties that ϕ ought to satisfy. While the formal definitions are rather technical and outside the scope of this paper, these properties intuitively state: (Sym-db) if two facts are interchangeable w.r.t. the query, they should have equal responsibility; (Null-db) if a fact $\alpha \in \mathcal{D}$ is irrelevant in the sense that $S \cup \{\alpha\} \models q$ iff $S \models q$ for all $S \subseteq \mathcal{D}$, then $\phi(\mathcal{D}, q, \alpha) = 0$, otherwise $\phi(\mathcal{D}, q, \alpha) > 0$; and (MS1) (resp. (MS2)) all other things being equal, a fact that appears in smaller (resp. more) minimal supports should have higher responsibility.

The notions of responsibility measures and minimal supports straightforwardly translate into the OMQA setting: take the ABox as the database and use an OMQ $(\mathcal{T}, q)$ for the query.

Shapley-Based Responsibility Measures

The responsibility measures considered in our work are based on the Shapley value. Originally defined in [14], it takes as input a cooperative game consisting of a finite set P of players and a wealth function $\xi \colon 2^P \to \mathbb{Q}$ that assigns a value to each coalition (ie, set) of players, with $\xi(\emptyset) = 0$. The *Shapley value* then assigns to each player $p \in P$ a value $\mathrm{Sh}(P, \xi, p)$ that should be seen as a 'fair share' of the total wealth $\xi(P)$ of the game that should be awarded to p based on the respective contributions of all players.

To obtain a responsibility measure from the Shapley value, one needs to model the input instance $(\mathcal{D}, q)$ as a cooperative game (P, ξ). The set P contains the elements that will receive a score, hence it should naturally be the set $\mathcal{D}$ itself. As for the wealth function, it must assign a numerical score to every database, reflecting in some way the satisfaction of the query. Formally, one needs to provide a *wealth function family* $\Xi^\star$ which associates a wealth function $\xi_q^\star$ with each query q. A responsibility measure can be straightforwardly obtained by applying the Shapley value to the game $(\mathcal{D}, \xi_q^\star)$: $\phi(\mathcal{D}, q, \alpha) := \mathrm{Sh}(\mathcal{D}, \xi_q^\star, \alpha)$.

The first wealth function family that was considered in the literature is Ξ^{dr}, defined by: $\xi_q^{\mathsf{dr}}(\mathcal{D}) := 1$ if $\mathcal{D} \models q$ and 0 otherwise [6], which gives rise to the *drastic Shapley value* $\mathrm{Sh}(\mathcal{D}, \xi_q^{\mathsf{dr}}, \alpha)$. In fact, Ξ^{dr} was until recently the only wealth function used to define Shapley-based responsibility measures for Boolean queries. Very recently, however, a new family of responsibility measures called *weighted sums of minimal supports* (WSMSs) has been defined as:

$$\phi^w_{\mathsf{wsms}}(\mathcal{D}, q, \alpha) := \sum_{\substack{S \in \mathsf{MS}_q(\mathcal{D}) \\ \alpha \in S}} w(|S|, |\mathcal{D}|)$$

based upon some weight function $w : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ [15]. It has been shown that all such measures can be equivalently defined via the Shapley value: for every weight function w, there exists a wealth function family Ξ^w such that $\phi^w_{\mathsf{wsms}}(\mathcal{D}, q, \alpha) = \mathrm{Sh}(\mathcal{D}, \xi_q^w, \alpha)$ [15, Proposition 4.4].

The wealth function family $\Xi^{\mathsf{ms}} := \Xi^w$ induced by the inverse weight function $w \colon (n, k) \mapsto 1/n$ is of particular interest as its wealth function $\xi_q^{\mathsf{ms}}(\mathcal{D})$ is simply the number of minimal supports for q in $\mathcal{D}$, which constitutes a very natural measure of how 'robust' the entailment $\mathcal{D} \models q$ is. Observe however that the weight function w can be adjusted to suit the needs of particular settings by giving more or less weight to the size of the minimal supports relative to their numbers (intuitively tuning the relative importance of (MS1) and (MS2)).

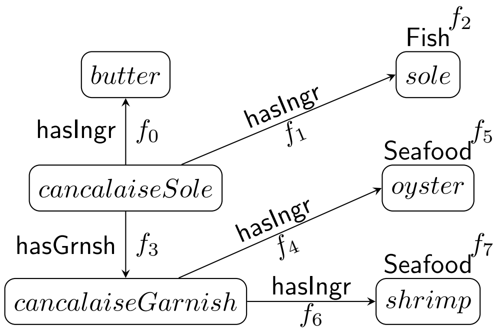
The Shapley values obtained from Ξ^{dr} and from Ξ^w (for any positive and non-decreasing w) yield responsibility measures that satisfy the properties (Sym-db)–(MS2) [15, Propositions B.1 and B.2]. As the following example illustrates, however, these measures do not always coincide, as the properties do not identify a unique 'reasonable' responsibility measure.

Example 1. *Consider the DL-Lite$_{\mathsf{core}}$ KB $(\mathcal{A}, \mathcal{T})$ whose TBox $\mathcal{T}$ contains the following axioms:*

$$\exists \mathsf{hasIng}.\mathsf{FishBased} \sqsubseteq \mathsf{FishBased},\ \mathsf{hasGrnsh} \sqsubseteq \mathsf{hasIng},\ \mathsf{Seafood} \sqsubseteq \mathsf{FishBased},\ \mathsf{Fish} \sqsubseteq \mathsf{FishBased}$$

and whose ABox $\mathcal{A}$ is depicted in Figure 1. We take the query $q := \mathsf{FishBased}(cancalaiseSole)$. There are 3 minimal supports for the OMQ $Q := (\mathcal{T}, q)$ in $\mathcal{A}$: $\{f_1, f_2\}$, $\{f_3, f_4, f_5\}$ and $\{f_3, f_6, f_7\}$. Although the properties (Sym-db)–(MS2) enforce many conditions, they do not restrict the relative values of f_1 and f_3. Indeed, we can observe in Figure 1 that $\mathrm{Sh}(\mathcal{D}, \xi_q^{dr}, f_1) > \mathrm{Sh}(\mathcal{D}, \xi_q^{dr}, f_3)$, but $\mathrm{Sh}(\mathcal{D}, \xi_q^{ms}, f_1) < \mathrm{Sh}(\mathcal{D}, \xi_q^{ms}, f_3)$. Note for example that $\mathrm{Sh}(\mathcal{D}, \xi_q^{ms}, f_3) = 1/3 + 1/3$ since f_3 is in two minimal supports, both of size 3, and hence each contributing $1/3$.

Following [15], for any wealth function family $\Xi^\star$ and class $\mathcal{C}$ of queries, we denote by $\mathrm{SVC}^\star_{\mathcal{C}}$ the problem of computing $\mathrm{Sh}(\mathcal{D}, \xi_q^\star, \alpha)$ given any database $\mathcal{D}$, fact $\alpha \in \mathcal{D}$, and query $q \in \mathcal{C}$. We also consider the problem $\mathrm{SVC}^\star_q$ associated with a single fixed query q. Our focus in this paper will be on the case $\Xi^\star = \Xi^w$ for some weight function w, in particular Ξ^{ms}, in which case we will speak of *WSMS computation*. Moreover, we shall study these tasks in the OMQA setting, so $\mathcal{C}$ will be a class $(\mathcal{L}, \mathcal{Q})$ of OMQs, and q will be a particular OMQ Q.



$\star$	f_0	f_1, f_2	f_3	f_4, f_5, f_6, f_7
dr	0	$\frac{1224}{5040}$	$\frac{1056}{5040}$	$\frac{384}{5040}$
ms	0	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$

Figure 1: Left: Example KB about a recipe from [21]. Arrows represent role assertions and box labels (e.g. Fish) indicate concept assertions. Right: Values of $\mathrm{Sh}(\mathcal{D}, \xi_q^\star, \cdot)$ for assertions in the example KB.

Summary of Contributions

Our results show that the good computational behaviour of WSMS in the database setting [15] extends to some relevant classes of OMQs. This is in sharp contrast to the intractability of the drastic Shapley measure considered in the database [6, 11] and ontology [13] settings. More precisely, WSMS computation is tractable in data complexity for $\mathsf{UCQ}^{\neq}$-rewritable OMQs:

Theorem 1. $\mathsf{SVC}_Q^w \in \mathsf{FP}$ *for every tractable weight function* w *and every Boolean OMQ* Q *that is* $\mathsf{UCQ}^{\neq}$*-rewritable. In particular,* $\mathsf{SVC}^w_{(DL\text{-}Lite_\mathcal{R}, \mathsf{UCQ})}$ *enjoys* FP *data complexity.*

We show in fact that WSMS computation for such OMQs can be implemented using relational database systems via simple SQL queries.

We also identify DL constructs that render WSMS computation intractable. In particular, we show that the data complexity becomes $\#\mathsf{P}$-hard for classes of OMQs capturing reachability:

Theorem 2. *Let* w *be a reversible tractable weight function, and* $\mathcal{L}$ *be any DL that can express the axiom* $\exists r.A \sqsubseteq A$*. Then, there exists an OMQ* $Q \in (\mathcal{L}, \mathsf{AQ})$ *such that* SVC_Q^w *is* $\#\mathsf{P}$*-hard.*

Furthermore, the presence of concept conjunction, present in lightweight DLs like as $\mathcal{EL}$ and Horn dialects of DL-Lite, leads to $\#\mathsf{P}$-hardness in *combined complexity*, again already for AQs.

For common DL-Lite dialects that do not admit conjunction, we obtain tractable combined complexity for OMQs based upon atomic queries. Furthermore, by means of careful analysis, we are able to identify classes of structurally restricted conjunctive queries that admit tractable WSMS computation, via reduction to the atomic case. We omit the formal definition of *interaction-free* OMQs, but intuitively they disallow undesirable interactions between query atoms and suitably generalize the self-join-free condition employed in the database setting.

Theorem 3. *Let* w *be a tractable weight function and* $\mathcal{C}$ *be a subclass of interaction-free OMQs from* $(DL\text{-}Lite_\mathcal{R}, \mathsf{CQ})$ *such that* $\{q \mid (\mathcal{T}, q) \in \mathcal{C}\}$ *has bounded treewidth. Then* SVC_Q^w *is in* FP *for combined complexity.*

The preceding theorem cannot be obtained by simply rewriting the OMQ and applying results from the database setting (indeed, there are no known tractability results for UCQs). Instead, it is necessary to exploit properties of canonical models of DL-Lite$_\mathcal{R}$ KBs. It is an interesting open question whether Theorem 3 can be extended to linear existential rule ontologies.

Acknowledgements

The authors have been partially supported by ANR AI Chair INTENDED (ANR-19-CHIA-0014).

References

[1] A. Borgida, D. Calvanese, M. 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), 2008, pp. 1440–1457.

[2] C. Alrabbaa, S. Borgwardt, P. Koopmann, A. Kovtunova, Explaining ontology-mediated query answers using proofs over universal models, in: Proceedings of the International Joint Conference on Rules and Reasoning (RuleML+RR), 2022, pp. 167–182.

[3] M. Bienvenu, C. Bourgaux, F. Goasdoué, Computing and explaining query answers over inconsistent DL-Lite knowledge bases, Journal of Artificial Intelligence Research (JAIR) 64 (2019) 563–644. URL: https://doi.org/10.1613/jair.1.11395. doi:10.1613/JAIR.1.11395.

[4] İ. İ. Ceylan, T. Lukasiewicz, E. Malizia, A. Vaicenavicius, Explanations for query answers under existential rules, in: International Joint Conference on Artificial Intelligence (IJCAI), 2019, pp. 1639–1646.

[5] İ. İ. Ceylan, T. Lukasiewicz, E. Malizia, A. Vaicenavicius, Explanations for ontology-mediated query answering in description logics, in: Proceedings of the European Conference on Artificial Intelligence (ECAI), 2020, pp. 672–679.

[6] E. Livshits, L. E. Bertossi, B. Kimelfeld, M. Sebag, The Shapley value of tuples in query answering, Logical Methods in Computer Science (LMCS) 17 (2021). doi:10.46298/LMCS-17(3:22)2021.

[7] D. Deutch, N. Frost, B. Kimelfeld, M. Monet, Computing the shapley value of facts in query answering, in: ACM SIGMOD International Conference on Management of Data (SIGMOD), ACM, 2022, pp. 1570–1583. URL: https://doi.org/10.1145/3514221.3517912. doi:10.1145/3514221.3517912.

[8] M. Khalil, B. Kimelfeld, The complexity of the shapley value for regular path queries, in: International Conference on Database Theory (ICDT), volume 255 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023, pp. 11:1–11:19. URL: https://doi.org/10.4230/LIPIcs.ICDT.2023.11. doi:10.4230/LIPICS.ICDT.2023.11.

[9] A. Kara, D. Olteanu, D. Suciu, From shapley value to model counting and back, Proceedings of the ACM on Management of Data (PACMMOD) 2 (2024) 79. URL: https://doi.org/10.1145/3651142. doi:10.1145/3651142.

[10] A. Reshef, B. Kimelfeld, E. Livshits, The impact of negation on the complexity of the shapley value in conjunctive queries, in: ACM Symposium on Principles of Database Systems (PODS), ACM, 2020, pp. 285–297. URL: https://doi.org/10.1145/3375395.3387664. doi:10.1145/3375395.3387664.

[11] M. Bienvenu, D. Figueira, P. Lafourcade, When is Shapley value computation a matter of counting?, Proceedings of the ACM on Management of Data (PACMMOD) 2 (2024) 105. URL: https://doi.org/10.1145/3651606. doi:10.1145/3651606.

[12] P. Karmakar, M. Monet, P. Senellart, S. Bressan, Expected Shapley-Like Scores of Boolean

functions: Complexity and Applications to Probabilistic Databases, Proc. ACM Manag. Data 2 (2024) 92:1–92:26. doi:10.1145/3651593.
[13] M. Bienvenu, D. Figueira, P. Lafourcade, Shapley value computation in ontology-mediated query answering, in: Principles of Knowledge Representation and Reasoning (KR), 2024. doi:10.24963/KR.2024/15.
[14] L. S. Shapley, A value for n-person games, in: Contributions to the Theory of Games II, Princeton University Press, 1953, pp. 307–317.
[15] M. Bienvenu, D. Figueira, P. Lafourcade, Shapley revisited: Tractable responsibility measures for query answers, 2025. URL: https://arxiv.org/abs/2503.22358. arXiv:2503.22358, preprint available at https://arxiv.org/abs/2503.22358. To appear in PODS 2025.
[16] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, in: Journal on Data Semantics X, 2008, pp. 133–173.
[17] M. Bienvenu, M. 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*, 2015, pp. 218–307.
[18] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: A survey, in: International Joint Conference on Artificial Intelligence (IJCAI), 2018, pp. 5511–5519.
[19] F. Baader, I. Horrocks, C. Lutz, U. Sattler, An Introduction to Description Logic, Cambridge University Press, 2017.
[20] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati, Tractable reasoning and efficient query answering in description logics: The DL-Lite family, Journal of Automated Reasoning (JAR) (2007).
[21] A. Escoffier, Le guide culinaire : aide-mémoire de cuisine pratique, Bibliothèque Professionnelle, 1903.