How Hard Is It to Decide if a Fact Is Relevant to a Query? (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

This extended abstract summarizes our recent work (to appear at KR 2026) on the complexity of deciding whether a given fact is relevant to a Boolean conjunctive query (CQ), i.e. whether the fact belongs to some minimal subset of the data that makes the query hold. Despite being of central importance to query answer explanation, the combined complexity of deciding relevance has not been studied in detail, leaving open what makes this problem hard, and which restrictions can yield lower complexity. Relevance has previously been shown to be harder than query evaluation: namely, Σ_2^p-complete for CQs, even over a binary signature. We further observe that NP-hardness applies already to (acyclic) chain CQs. Our work identifies self-joins (multiple atoms with the same relation) as the culprit. Indeed, we prove that if we forbid or bound the occurrence of self-joins, then relevance has the same complexity as query evaluation, namely, NP (without structural restrictions) and LogCFL (for bounded hypertreewidth classes). In the ontology setting, we establish an analogous result for ontology-mediated queries consisting of a CQ and DL-Lite$_\mathcal{R}$ ontology, namely that relevance is no harder than query answering provided that we bound the interaction width (which generalizes both self-join width and a recently introduced 'interaction-free' condition). Our results thus provide useful insights into what makes relevance harder than query evaluation and allow us to identify natural classes of queries which admit efficient relevance computation.

Keywords

relevance, query answer explanation, database queries, ontology-mediated query answering, description logics, DL-Lite, complexity analysis

1. Introduction

There has been considerable interest in both the database and knowledge representation communities on devising methods for explaining why a given query answer holds, resulting in a diversity of approaches. In the database area, proposals include qualitative notions based upon causality [1] and provenance [2], as well as quantitative notions of explanation based upon (causal or Shapley-value-based) responsibility measures, which involve assigning scores to facts based upon their contribution to making the query hold [3, 4, 5]. In the setting of ontology-mediated query answering (OMQA) [6, 7, 8], where query answers are computed by reasoning over the information contained in the data and the ontology, there have also been a diversity of explanation notions advanced in the literature. Some of these rely upon proofs which show step-by-step how a query answer is obtained [9, 10], others compute minimal subsets of facts that make the query hold [11, 12, 13, 14], while some recent work has begun to explore quantitative responsibility measures [15, 16].

In the present work, our focus will be on explanations given as *minimal supports* (aka causes, MinEXes), i.e. subset-minimal sets of facts that suffice to obtain the query answer. More precisely, we shall be interested in the fundamental problem of *(query) relevance*, which is to decide whether a given fact from the dataset is *relevant* to the query, in the sense that it belongs to at least one minimal support. Indeed, identifying the relevant facts can be considered a central task in query answer explanation, which is useful both for summarizing the minimal supports and for aiding users in debugging unexpected query results. Relevance is also closely related to recently explored Shapley-based responsibility measures. Indeed, for classes of monotone queries such as conjunctive queries (with or without an ontology), a fact will be assigned a positive score iff it is relevant (cf. discussions in [5]). Finally, it is worth noting that notions of relevance have also been considered for other kinds of explanations, including abductive explanations [17], justifications for ontology axiom entailment [18, 19], explanations for

DL 2026: 39th International Workshop on Description Logics, July 17–19, 2026, Lisbon, Portugal

query non-answers [20], and explanations for query (non)answers under various inconsistency-tolerant semantics [11].

2. Existing Results

We briefly summarize here what is known about the complexity of the query relevance task, which has been explored for ontology-mediated queries (OMQs) formulated in description logic and existential rule ontology languages [12, 13, 21, 14].

Data complexity For the data complexity measure, in which the size of the query and ontology are treated as fixed, the query relevance task is known to be tractable for (unions) of conjunctive queries (CQs), as well as for ontology-mediated queries that can be rewritten as UCQs (i.e. for so-called FO-rewritable ontology languages). Indeed, this is an easy consequence of the fact that the size of minimal supports is bounded by a constant, making it possible to solve the relevance problem by reduction to first-order query evaluation. By contrast, for (ontology-mediated) query languages for which the size of minimal supports is unbounded, the relevance problem becomes NP-hard in data complexity. This applies for example to OMQs consisting of an $\mathcal{EL}$ ontology and an atomic query.

Combined complexity With respect to combined complexity, the relevance task has been proven Σ_2^p-complete for conjunctive queries, even in the absence of an ontology and with only binary relations [21]. Additional combined complexity results have been established for various kinds of OMQs, with complexities ranging from Σ_2^p to 2EXPTIME, depending on the expressivity of the ontology language.

While the preceding results show that relevance is generally harder than query evaluation, to the best of our knowledge, there has not been any exploration of what precisely makes it harder, nor the related question of which restrictions on query structure can lower the complexity.

3. Contributions

Our aim is to pinpoint the sources of difficulty and identify classes of queries for which relevance is no harder than query evaluation, in particular, leading to the identification of tractable classes. To this end, we have conducted a detailed study of the combined complexity of query relevance in both the database and ontology settings.

Database setting Our first contribution is to show that for conjunctive queries (without an ontology), it is the presence of self-joins (i.e. multiple query atoms using the same relation) that make relevance harder than query evaluation. We start by observing that for 'well-behaved' classes of conjunctive queries, like acyclic CQs and CQs of bounded hypertreewidth [22], for which query evaluation is tractable, the relevance problem is NP-complete. In fact, NP-hardness holds even for acyclic chain CQs of the form $R(x_1, x_2) \wedge R(x_2, x_3) \wedge \cdots \wedge R(x_{n-1}, x_n)$ on a single relation, which demonstrates that tractability cannot be regained by simplifying the query structure. However, if we restrict to CQs with 'few' self-joins (via a notion of self-join width, recently introduced in [5]), then the complexity of relevance matches that of query evaluation, namely, it is NP-complete for general CQs and LogCFL-complete for classes of CQs with bounded hypertree width. These results are summarized in Table 1.

Ontology setting We next consider the relevance problem for OMQs consisting of a CQ and a description logic (DL) ontology. After observing that both conjunction ($\sqcap$) and qualified existential restrictions ($\exists R.C$) can make relevance hard even for atomic queries, we focus on ontologies formulated in DL-Lite$_\mathcal{R}$, a prominent lightweight DL often used for OMQA [23]. We introduce a novel notion of interaction width, which can be seen as generalizing both self-join width and a recently introduced 'interaction-free' property [16]. Very roughly speaking, two query atoms are said to *interact* if there

	Query evaluation	Relevance	
		sjw<∞	sjw ∞
hw<∞	LogCFL-c	LogCFL-c	NP-c†
hw ∞	NP-c	NP-c	Σ_2^p-c‡

Table 1
Complexity results for classes of conjunctive queries with (un)bounded hypertree width (hw) and self-join width (sjw). Shaded cells indicate new results. † NP-hardness applies to chain CQs over a single binary relation. ‡ General Σ_2^p-hardness was previously shown [14], we prove it holds even for CQs with a single binary relation.

	OMQ evaluation		Relevance	
	iw<∞	iw ∞	iw<∞	iw ∞
tw=1, ℓ<∞	NL-c	LogCFL-c	NL-c	NP-c
tw <∞	LogCFL-c	NP-c	LogCFL-c	NP-h
tw ∞	NP-c		NP-c	Σ_2^p-c

Table 2
Complexity results for classes of ontology-mediated queries in (DL-Lite$_{\mathcal{R}}$, CQ), with (un)bounded treewidth (tw), interaction width (iw), or number of leaves (ℓ) in acyclic CQs (tw $= 1$). Shaded cells indicate new results.

exists a single assertion that can make each of the atoms (individually) hold, and interaction width is then defined as the number of such interacting query atoms. Our second main contribution is then to show that for OMQs consisting of a CQ and DL-Lite$_{\mathcal{R}}$ ontology, bounding the interaction width makes relevance no harder than query evaluation. To show this, we exploit several properties of DL-Lite$_{\mathcal{R}}$, in particular the so-called "singleton support" property (= only one fact needed to satisfy a query atom), the structure of DL-Lite$_{\mathcal{R}}$ canonical models, and existing results on the impact of restricting query and/or ontology structure. The obtained results, summarized in Table 2, show that simultaneously bounding the interaction width and treewidth enables efficient (LogCFL) relevance computation, and for acyclic bounded-leaf CQs, relevance drops to NL.

Graph setting As a third and last contribution, we investigate a more abstract notion of relevance based on graph homomorphisms, which underlies the CQ relevance problem. Concretely, we study the question of, given two graphs G and G', whether a given edge belongs to a minimal homomorphic image of G in G'. We consider both directed and undirected variants of this problem, finding that they are both Σ_2^p-complete, thereby providing natural and simple Σ_2^p-complete problems of independent interest. As a corollary, we obtain that Σ_2^p-hardness of the relevance problem already holds for CQs over signatures consisting of a single binary relation.

4. Discussion and Future Work

The main takeaway from our study is that the difficulty of the relevance problem for (DL-Lite$_{\mathcal{R}}$-mediated) CQs hinges fundamentally on the number of atoms (or of variables from different atoms) that may interact with a fact in the following sense: if we restrict the number of self-join variables or interacting atoms, then the complexity of relevance is essentially as good or bad as that of query evaluation. These insights allowed us to identify natural classes of queries for which relevance can be efficiently decided, and we expect that they will also prove useful when designing practical algorithms for the relevance problem. Our work also opens up a number of interesting directions for future research.

Complexity dichotomies & parameterized complexity It is a natural question whether bounding the self-join width or interaction width is a necessary condition for obtaining tractability or lower complexity. More precisely, one might try to prove a dichotomy result along the following lines: "For every recursively enumerable class C of conjunctive queries, if C has bounded self-join width and bounded (hyper)treewidth, then relevance is in polynomial time, and otherwise it is NP-hard".

Unfortunately, it is not at all clear that such a dichotomy holds, as there is currently no such PTIME / NP-hard dichotomy established for the combined complexity of CQ evaluation, and existing research provides strong evidence against the existence of such a dichotomy [24]. By contrast, an FPT / W[1]-hard dichotomy has been proven for the *parameterized* evaluation of conjunctive queries, using the treewidth (modulo cores) as the parameter [24]. Hence, a promising direction, which we plan to explore in future work, is to investigate the parameterized complexity of the relevance problem.

Other ontology languages We conjecture that the notion of interaction width we introduced for $\mathsf{DL\text{-}Lite}_{\mathcal{R}}$ can be fruitfully applied also to linear existential rules, as they similarly enjoy the singleton support property (i.e. only one fact is needed to satisfy a query atom). For ontology languages which admit conjunction, and hence do not enjoy the singleton support property (since multiple facts may be needed to infer a query atom), such as $\mathsf{DL\text{-}Lite}_{\mathsf{Horn}}$, new restrictions will be needed to define well-behaved OMQ classes admitting tractable relevance computation. Note however that we cannot expect any positive results for classes of $\mathsf{DL\text{-}Lite}_{\mathsf{Horn}}$ OMQs obtained via bounding parameters which are dominated by the query size, since relevance for atomic queries is already NP-hard. Moreover, for non-first-order-rewritable DLs like $\mathcal{EL}$, the intractability result for atomic CQs holds even w.r.t. data complexity, thereby precluding any tractability results for classes of OMQs defined by bounding any combination of width notions or other parameters that assign a finite value to each OMQ. It would still be worthwhile, however, to develop pragmatic algorithms for deciding relevance in such logics.

Relevance w.r.t. inconsistent KBs When studying relevance of OMQs, we have focused on *consistent* knowledge bases. Since explaining the inconsistency of a given KB can be seen as explaining the Boolean query 'is the KB consistent?', we could naturally consider the relevance problem for inconsistency. We should be able to transfer upper bounds from the query setting, but not necessarily lower bounds, since such 'inconsistency queries' may be restricted in ways that lower the complexity. Indeed, in $\mathsf{DL\text{-}Lite}_{\mathcal{R}}$, relevance of inconsistency is easily shown to be in NL as minimal inconsistent subsets are of size at most 2. Alternatively, one may instead compute query answers under inconsistency-tolerant semantics and study the relevance problem for explanations under such semantics. Interestingly, existing proposals of explanations for repair-based semantics [11] are defined via minimal supports, so we expect that our results will prove useful also in this setting.

Interaction width beyond relevance We expect that our new notion of interaction width may prove useful for identifying tractable OMQ classes for other explanation-related tasks. In particular, it was recently shown that the class of interaction-free OMQs admits efficient counting of minimal supports [16], and we are cautiously optimistic that this positive result can be extended to classes of OMQs in $\mathsf{DL\text{-}Lite}_{\mathcal{R}}$ with bounded interaction width (in line with an analogous result shown in the database setting for classes of CQs with bounded self-join width [5]). Counting minimal supports has applications in query answer explanation as the number of minimal supports can be used to assign numeric scores to facts based upon their contribution to making the query hold, as detailed in [5, 16].

Acknowledgments

This work has been partially supported by ANR grants INTENDED (ANR-19-CHIA-0014) and EXPAND (ANR-25-CE23-1215).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] A. Meliou, W. Gatterbauer, K. F. Moore, D. Suciu, The complexity of causality and responsibility for query answers and non-answers, Proc. VLDB Endow. 4 (2010) 34–45. doi:10.14778/1880172.1880176.

[2] T. J. Green, V. Tannen, The semiring framework for database provenance, in: ACM Symposium on Principles of Database Systems (PODS), ACM Press, 2017, pp. 93–99. doi:10.1145/3034786.3056125.

[3] E. Livshits, L. Bertossi, B. Kimelfeld, M. Sebag, The Shapley value of tuples in query answering, Logical Methods in Computer Science (LMCS) Volume 17, Issue 3 (2021) 6942. doi:10.46298/lmcs-17(3:22)2021.

[4] 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) 1–24. doi:10.1145/3651606.

[5] M. Bienvenu, D. Figueira, P. Lafourcade, Shapley revisited: Tractable responsibility measures for query answers, Proceedings of the ACM on Management of Data (PACMMOD) 3 (2025) 1–26. doi:10.1145/3725249.

[6] 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. doi:10.1007/978-3-540-77688-8_5.

[7] 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. doi:10.1007/978-3-319-21768-0_9.

[8] 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. doi:10.24963/ijcai.2018/777.

[9] 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. doi:10.1007/978-3-540-88873-4_35.

[10] 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. doi:10.1007/978-3-031-21541-4_11.

[11] 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. doi:10.1613/JAIR.1.11395.

[12] 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. doi:10.24963/IJCAI.2019/227.

[13] 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. doi:10.3233/FAIA200153.

[14] İ. İ. Ceylan, T. Lukasiewicz, E. Malizia, A. Vaicenavicius, Explanations for query answers under existential rules, Artif. Intell. 341 (2025) 104294. doi:10.1016/J.ARTINT.2025.104294.

[15] 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.

[16] M. Bienvenu, D. Figueira, P. Lafourcade, Tractable responsibility measures for ontology-mediated query answering, in: Principles of Knowledge Representation and Reasoning (KR), 2025. doi:10.24963/KR.2025/16.

[17] T. Eiter, G. Gottlob, The complexity of logic-based abduction, Journal of the ACM 42 (1995) 3–42. doi:10.1145/200836.200838.

[18] R. Peñaloza, B. Sertkaya, Understanding the complexity of axiom pinpointing in lightweight

description logics, Artif. Intell. 250 (2017) 80–104. doi:10.1016/J.ARTINT.2017.06.002.
[19] J. Chen, Y. Ma, R. Peñaloza, H. Yang, Union and intersection of all justifications, in: International Conference on Semantic Web (ESWC), volume 13261 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 56–73. doi:10.1007/978-3-031-06981-9_4.
[20] D. Calvanese, M. Ortiz, M. Simkus, G. Stefanoni, Reasoning about explanations for negative query answers in DL-Lite, Journal of Artificial Intelligence Research (JAIR) 48 (2013) 635–669. doi:10.1613/JAIR.3870.
[21] A. Vaicenavičius, Explanations for ontology-mediated query answers, Ph.D. thesis, University of Oxford, UK, 2020. URL: https://ora.ox.ac.uk/objects/uuid:e032d5cb-e8f8-4aaa-b1b5-b12dda4981c4/files/djd472w643.
[22] G. Gottlob, N. Leone, F. Scarcello, Hypertree decompositions and tractable queries, Journal of Computer and System Sciences (JCSS) 64 (2002) 579–627. doi:10.1006/JCSS.2001.1809.
[23] 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). doi:10.1007/s10817-007-9078-x.
[24] M. Grohe, The complexity of homomorphism and constraint satisfaction problems seen from the other side, J. ACM 54 (2007) 1:1–1:24. doi:10.1145/1206035.1206036.