Extended Abstract: Tractable Query Answering for Expressive Ontologies and Existential Rules[⋆]

David Carral, Irina Dragoste, Markus Krötzsch

Center for Advancing Electronics Dresden (cfaed), TU Dresden, Germany

Answering conjunctive queries (CQs) over knowledge bases (KBs) containing disjunctive existential rules is a relevant reasoning task which can be addressed using the *disjunctive chase algorithm*—a sound and complete materialisation-based procedure where all relevant consequences are pre-computed, allowing queries to be directly evaluated over materialised sets of facts—and *acyclicity notions* [7,8]—sufficient conditions that guarantee termination of this procedure [2]. As shown in [5,6], acyclicity notions can be used to determine that the chase will indeed terminate over a large subset of real-world ontologies. Nevertheless, even if a KB is characterised as acyclic, CQ answering still remains a problem of high theoretical complexity: CQ answering over acyclic programs with disjunctive existential rules is coN2ExpTime-complete [3]. For acyclic Horn-$\mathcal{SROIQ}$ ontologies, it is ExpTime-complete [6]. Moreover, CQ answering becomes even harder if we consider non-deterministic ontologies.

Example 1. Let $\mathcal{R}_n = \{D_{i-1}(x) \rightarrow \exists y_i.L_i(x, y_i) \wedge D_i(y_i), D_{i-1}(x) \rightarrow \exists z_i.R_i(x, z_i) \wedge D_i(z_i) \mid i = 1, \ldots, n\}$ with $n \geq 0$. The chase of the program $\mathcal{P} = \langle \mathcal{R}_n, \{D_0(c)\}\rangle$ is exponentially large in n. Note that $\mathcal{P}$ is acyclic with respect to all notions described in [6] and can be expressed in most DL fragments.

We study the limits of tractable reasoning using the chase and propose a series of restrictions that, if combined, prevent the exponential blow-up highlighted in the previous example. An important concept for predicting the behaviour of the chase procedure is the *dependency graph* of a rule set, defined next. Our definition refers to the *skolem chase*, which uses functional terms to denote fresh elements.

Definition 2. *Consider a rule set $\mathcal{R}$ where (without loss of generality) rules do not share variables. The* dependency graph $G(\mathcal{R})$ *of $\mathcal{R}$ has the existentially quantified variables in $\mathcal{R}$ as nodes, and an edge $y \rightarrow z$ if the skolem chase of some program $\langle \mathcal{R}, \mathcal{I}\rangle$ contains terms of the form $f_z(\vec{t})$ and $f_y(\vec{s})$, where f_z and f_y are the skolem functions for z and y, respectively, and $f_y(\vec{s})$ occurs among the terms in $\vec{t}$.*

Intuitively, $y \rightarrow z$ means that a domain element created for the existential variable y was involved in an application of the rule of z (to instantiate a variable that occurred in body *and* head). Let $\mathcal{R}_n$ be the rule set from Example 1. Then, $G(\mathcal{R}_n) = \emptyset$ if $n \leq 1$, and $G(\mathcal{R}_n) = \{y_{i-1} \rightarrow y_i, y_{i-1} \rightarrow z_i, z_{i-1} \rightarrow y_i, z_{i-1} \rightarrow z_i \mid i = 2, \ldots, n\}$ otherwise.

The key to our tractability results is the notion of a *braid*, which, intuitively speaking, consists of a possibly large number of intertwined paths.

[⋆] The full version of this paper was published at ISWC 2017 [4].

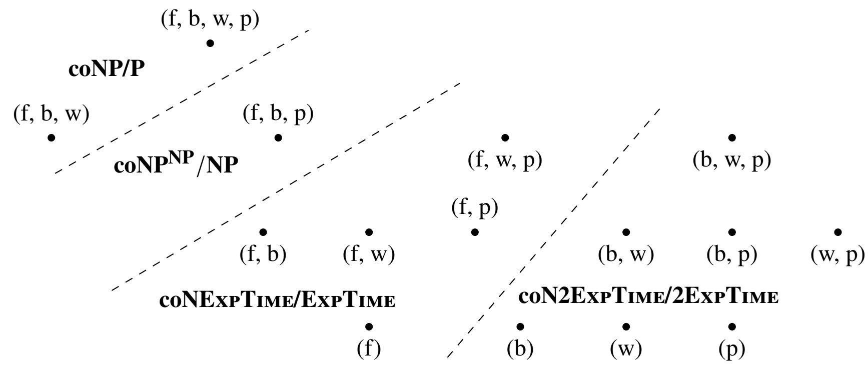

Fig. 1. Combined complexity of Boolean CQ entailment (non-deterministic/deterministic rules) with respect to the size of a rule set that satisfies (a) and a combination of (f), (b), (w) and (p)

Definition 3. *Consider a directed graph G. A* path *is a sequence of nodes $\alpha_1, \ldots, \alpha_n$ with $\alpha_i \to \alpha_{i+1} \in G$ for all $i = 1, \ldots, n-1$. The graph G is* acyclic *if, for every path $\alpha_1, \ldots, \alpha_n$ with $n \geq 2$, $\alpha_1 \neq \alpha_n$. A* simple path *is a path which does not contain two occurrences of the same node. A* braid *is a sequence of nodes $\alpha_1, \ldots, \alpha_n$ such that, for all $i = 1, \ldots, n-1$, there are at least two different simple paths from α_i to α_{i+1}.*

As our main theoretical contribution, we study the complexity of reasoning over a rule set $\mathcal{R}$ that satisfies some combination of the following restrictions:

(a) The graph $G(\mathcal{R})$ is acyclic.
(f) The arity of all function symbols in $sk(\mathcal{R})$ is at most 1.
(b) The length of the braids in $G(\mathcal{R})$ is bounded.
(w) The treewidth of the rules in $\mathcal{R}$ is bounded.
(p) The arity of the predicates in $\mathcal{R}$ is bounded.

We summarise our findings in Figure 1, where we assume that the rule set satisfies (a), as otherwise reasoning becomes undecidable. All the complexity results are tight.

We empirically study the generality of these restrictions using (deterministic and non-deterministic) real-world ontologies without equality from the MOWL Corp (MC) [9] and Oxford Ontology Library (OOL) [1]. To do so, we transform the DL ontologies in these corpora into KBs with disjunctive existential rules, and then check how many of them satisfy (a-p). Since rule sets resulting from the transformation of normalised DL ontologies satisfy (f), (w), and (p), we check how many satisfy (a) and (b). We found that 61.8% (974) of the ontologies from MC and (76%) (171) from OOL satisfy (a). Moreover, 78.3% of the acyclic ontologies from MC contain braids of length at most 1, 90.8% of length at most 2, 95.5% at most 3, 98.8% at most 4 and 99% at most 5. In the OOL, 51.4% of the acyclic ontologies feature braids of length at most 1, 69.5% at most 2, 81.2% at most 3, 92.3% at most 4, 97.6 % at most 5, and 98.2 % at most 6. Our work therefore suggests a new approach to efficient CQ answering that might be applicable to many real-world ontologies.

Acknowledgements This work was supported by the DFG within the cfaed Cluster of Excellence, CRC 912 (HAEC), and Emmy Noether grant KR 4381/1-1.

References

1. The oxford ontology library. Available at `https://www.cs.ox.ac.uk/isg/ontologies/`
2. Bourhis, P., Morak, M., Pieris, A.: The impact of disjunction on query answering under guarded-based existential rules. In: Proc. 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI'13). pp. 796–802. AAAI Press/IJCAI (2013)
3. Carral, D., Dragoste, I., Krötzsch, M.: Restricted chase (non)termination for existential rules with disjunctions. In: Sierra, C. (ed.) Proc. 26th Int. Joint Conf. on Artificial Intelligence (IJCAI'17). pp. 922–928. IJCAI (2017)
4. Carral, D., Dragoste, I., Krötzsch, M.: Tractable query answering for expressive ontologies and existential rules. In: Proc. 16th Int. Semantic Web Conf. (ISWC'17). LNCS, vol. 10587, pp. 156–172 (2017)
5. Carral, D., Feier, C., Hitzler, P.: A practical acyclicity notion for query answering over Horn-SRIQ ontologies. In: The 15th International Semantic Web Conference, Kobe, Japan, 2016, Proceedings, Part I. LNCS, vol. 9981, pp. 70–85 (2016)
6. Cuenca Grau, B., Horrocks, I., Krötzsch, M., Kupke, C., Magka, D., Motik, B., Wang, Z.: Acyclicity notions for existential rules and their application to query answering in ontologies. J. of Artificial Intelligence Research 47, 741–808 (2013)
7. Krötzsch, M., Rudolph, S.: Extending decidable existential rules by joining acyclicity and guardedness. In: Proc. 22nd Int. Conf. on Artificial Intelligence (IJCAI'11). pp. 963–968. IJCAI (2011)
8. Marnette, B.: Generalized schema-mappings: from termination to tractability. In: Paredaens, J., Su, J. (eds.) Proc. 28th Symposium on Principles of Database Systems (PODS'09). pp. 13–22. ACM (2009)
9. Matentzoglu, N., Bail, S., Parsia, B.: A snapshot of the OWL Web. In: Proc. 12th Int. Semantic Web Conf. (ISWC'13). LNCS, vol. 8218, pp. 331–346. Springer (2013)