

# Characterising Fixed Parameter Tractability of Query Evaluation Over Guarded TGDs (Extended Abstract)\*

Cristina Feier

University of Bremen

We consider the usual scenario of ontology mediated querying where queries are posed to a database whose knowledge is enriched by means of an ontology. A tuple  $(\mathcal{L}, \mathcal{Q})$ , where  $\mathcal{L}$  is an ontology language and  $\mathcal{Q}$  is a query language, is referred to as an *OMQ language*. One thoroughly explored fragment of FOL as a basis for ontology specification languages is that of *tuple generating dependencies* (TGDs). A *tgds* is a rule (logical implication) having as body and head conjunctions of atoms, where some variables occurring in head atoms might be existentially quantified (all other variables are universally quantified). As such, it potentially allows the derivation of atoms over fresh individuals (individuals not mentioned in the database). Answering (even atomic) queries with respect to sets of *tgds* is undecidable [7]. However, there has been lots of work on identifying decidable fragments [7, 1, 9]. A prominent such fragment is that of Guarded TGDs (GTGD) [7]: a *tgds* is *guarded* if all universally quantified variables occur as terms of some body atom, called *guard*. Many Description Logic languages like  $\mathcal{ELH}$  or  $\mathcal{ELHI}$  can be expressed using GTGDs.

While query answering with respect to GTGDs is decidable, the combined complexity of the problem is high: EXPTIME-complete for bounded arity schemas, and 2EXPTIME-complete in general. At the same time, while the fragment is tractable w.r.t. data complexity [6], the coefficient of the polynomial depends on the size of the ontology, which might be relatively high. By fixing the set of *tgds*, the combined complexity drops to NP for evaluating conjunctive queries (CQs), and to PTIME for evaluating atomic queries and CQs of bounded treewidth [8] which is similar to the complexity of query evaluation over databases [15]. A natural question is when can OMQs from (GTGD, UCQ) be evaluated efficiently?

The above question has been partially answered in [2]. There, the complexity of evaluating OMQs over bounded arity schemas is studied in a parameterized setting, with the parameter being the size of the OMQ. Under the assumption that  $\text{FPT} \neq \text{W}[1]$ , it is shown that a recursively enumerable (r.e.) class of OMQs over bounded arity schemas can be fixed parameter tractable (fpt) iff it has bounded semantic treewidth, i.e. there exists a  $k$  such that every OMQ from the class is equivalent to another OMQ from (GTGD, UCQ) whose UCQ has syntactic treewidth  $k$  [4]. The work generalizes previous results concerning

---

\* “Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).”

efficiency of evaluation of OMQs over bounded arity schemas based on the DL languages  $\mathcal{ELH}$  and  $\mathcal{ELHI}$  from [3].

The characterization from [2] is similar to the well-known result of Grohe [13], concerning the parameterized complexity of CQ evaluation over databases: the evaluation of a r.e. class of CQs is fpt iff the class has bounded semantic treewidth, i.e. there exists a class of equivalent CQs of bounded treewidth. The similarity is not coincidental, as the results build from [2] on the results of Grohe in a non-trivial way. In fact, the lower bound proof uses a central construction from [13], but has to employ sophisticated techniques to adapt this to OMQs.

The main research question we address in this paper is the following: *when is it possible to efficiently evaluate OMQs from (GTGD, UCQ) in the case where there is no restriction concerning schema arity?* We consider again a parameterized setting, where the parameter is the size of the OMQ. Given a r.e. class of OMQ  $\mathbb{Q}$  from (GTGD, UCQ), we denote with  $\text{p-OMQ}(\mathbb{Q})$  the parameterized problem of evaluating OMQs from  $\mathbb{Q}$ . As our first main result shows, in this case another structural measure for OMQs/UCQs plays a central role: *submodular width*. Similarly to the definition of semantic treewidth for OMQs, it is possible to define a notion of semantic submodular width of an OMQ. We obtain the following:

**Main Result 1.** Let  $\mathbb{Q}$  be a r. e. class of OMQs from (GTGD, UCQ). Assuming the Exponential Time Hypothesis,  $\text{p-OMQ}(\mathbb{Q})$  is fixed-parameter tractable iff  $\mathbb{Q}$  has bounded semantic submodular width.

To prove the result, we use the fact that every OMQ from (GTGD, UCQ) can be rewritten into an OMQ from (GDLog, UCQ) [2], where GDLog stands for Guarded Datalog, the restriction of GTGD to rules with only universally quantified variables. For OMQs from (GDLog, UCQ), we construct equivalent OMQs called *covers* which witness bounded semantic submodular width. Covers are based on so-called *characteristic databases*, which are databases that entail an OMQ and which are sufficiently minimal with respect to the homomorphism order, in a very specific sense.

Typically, in a database setting, the database induced by a CQ (or its core) can be seen as a canonical database which entails the query and on which to base further constructions. However, in the case of OMQs which pose restrictions on the database schema this is no longer possible: a CQ might contain symbols which are not allowed to occur in a database. In fact, this is a typical usage of ontologies: to enrich the database schema with new terminology. As such, we identify other representative databases for OMQs. Based on these new concepts, for OMQs from (GDLog, UCQ) it is possible to obtain an alternative characterization for fpt evaluation based on syntactic measures:

**Main Result 2.** For  $\mathbb{Q}$  a r. e. class of OMQs from (GDLog, UCQ), let  $\mathbb{Q}_c$  and  $\mathbb{D}_{\mathbb{Q}}$  be the classes of covers and characteristic databases for OMQs from  $\mathbb{Q}$ , respectively. Under the Exponential Time Hypothesis, the following statements are equivalent:

1.  $\text{p-OMQ}(\mathbb{Q})$  is fixed-parameter tractable;

2.  $\mathbb{Q}_c$  has bounded submodular width;
3.  $\mathbb{D}_{\mathbb{Q}}$  has bounded submodular width.

The hardness result for the above characterization exploits the seminal result of Marx [14] concerning fpt evaluation of uniform CSPs closed under underlying hypergraphs. For a r.e. class of structures  $\mathbb{A}$ , the parameterized uniform CSP problem  $p\text{-CSP}(\mathbb{A}, -)$  asks whether there exists a homomorphism from some structure  $A$  in  $\mathbb{A}$  (with  $|A|$  being the parameter) to an arbitrary relational structure. The result of Marx has been lifted to arbitrary uniform CSPs in [11]:

**Theorem 1 (Theorem 1, [11]).** *Let  $\mathbb{C}$  be a r.e. class of structures. Assuming the Exponential Time Hypothesis,  $p\text{-CSP}(\mathbb{C}, -)$  is fixed-parameter tractable iff  $\mathbb{C}$  has bounded semantic submodular width.*

The problem of evaluating uniform CSPs can be seen as an alternative presentation of the problem of evaluating a r.e. class of Boolean CQs over databases. In fact, Grohe’s characterization for fpt evaluation of r. e. classes of CQs in the bounded arity case has been achieved via a uniform CSP detour [13]. To exploit the result from [11] concerning evaluation of uniform CSPs, we introduce an fpt-reduction from parameterized uniform CSP evaluation to parameterized OMQ evaluation which preserves semantic submodular width.

**Main Result 3.** For  $Q$  an OMQ from (GDLog, UCQ), there exists an fpt-reduction from  $p\text{-CSP}(\mathbb{D}_Q, -)$  to  $p\text{-OMQ}(\{Q\})$ .

The reduction is important also as a stand-alone result as it can be used as a black-box tool to port results from the uniform CSP/database realm to the OMQ one. To further showcase this, we revisit the problem of evaluating r.e. classes of OMQs over bounded arity schemas, previously addressed in [2] and establish analogues of Main Result 1 and Main Result 2. In this case, submodular width is replaced with treewidth and the Exponential Time Hypothesis with the assumption that  $\text{FPT} \neq \text{W}[1]$ . This is due to the fact that the set of characteristic databases of an OMQ has the same treewidth as its cover. As such, we retrieve the semantic characterization from [2] without going into details of the construction employed by Grohe in [13], and also establish an alternative syntactic characterization for fpt evaluation of OMQs from (GDLog, UCQ) over bounded arity schemas using the newly introduced notions of cover and characteristic databases.

So far we have only considered Boolean OMQs, as the corresponding results for CQ evaluation in the unbounded arity case have also only been established in the Boolean case. As future work, we plan to extend our work to the non-Boolean case. Many results from the database world concerning efficiency of performing tasks like counting [12], enumeration [5, 10], and so on, use structural measures on the queries similar to the ones involved in the characterizations for evaluation. Thus, by generalizing our constructs to the non-Boolean case, depending on which structural measures are preserved when transitioning from sets of characteristic databases to covers of OMQs, we might be able to lift such results to the OMQ world.

## References

1. Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artif. Intell.*, 175(9-10):1620–1654, 2011.
2. Pablo Barceló, Victor Dalmau, Cristina Feier, Carsten Lutz, and Andreas Pieris. The limits of efficiency for open- and closed-world query evaluation under guarded tgds. In *PODS*, pages 259–270, 2020.
3. Pablo Barceló, Cristina Feier, Carsten Lutz, and Andreas Pieris. When is ontology-mediated querying efficient? In *LICS*, pages 1–13, 2019.
4. Pablo Barceló, Diego Figueira, Georg Gottlob, and Andreas Pieris. Semantic optimization of conjunctive queries. 2019. Under submission, available at <http://homepages.inf.ed.ac.uk/apieris/BFGP.pdf>.
5. Christoph Berkholz and Nicole Schweikardt. Constant delay enumeration with fpt-preprocessing for conjunctive queries of bounded submodular width. In *MFCS 2019*, volume 138, pages 58:1–58:15, 2019.
6. Andrea Cali, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *Journal of Web Semantics*, 14:57 – 83, 2012.
7. Andrea Cali, Georg Gottlob, and Michael Kifer. Taming the infinite chase: Query answering under expressive relational constraints. *J. Artif. Intell. Res.*, 48:115–174, 2013.
8. Andrea Cali, Georg Gottlob, and Thomas Lukasiewicz. Datalog $\pm$ : a unified approach to ontologies and integrity constraints. In *ICDT*, pages 14–30, 2009.
9. Andrea Cali, Georg Gottlob, and Andreas Pieris. Towards more expressive ontology languages: The query answering problem. *Artif. Intell.*, 193:87–128, 2012.
10. Nofar Carmeli and Markus Kröll. Enumeration Complexity of Conjunctive Queries with Functional Dependencies. In *(ICDT 2018)*, volume 98, pages 11:1–11:17, 2018.
11. Hubie Chen, Georg Gottlob, Matthias Lanzinger, and Reinhard Pichler. Semantic width and the fixed-parameter tractability of constraint satisfaction problems. In Christian Bessiere, editor, *IJCAI*, pages 1726–1733, 2020.
12. Hubie Chen and Stefan Mengel. A Trichotomy in the Complexity of Counting Answers to Conjunctive Queries. In *(ICDT 2015)*, volume 31, pages 110–126, 2015.
13. Martin Grohe. The complexity of homomorphism and constraint satisfaction problems seen from the other side. *J. ACM*, 54(1):1:1–1:24, 2007.
14. Dániel Marx. Tractable hypergraph properties for constraint satisfaction and conjunctive queries. In *STOC*, pages 735–744, 2010.
15. Mihalis Yannakakis. Algorithms for acyclic database schemes. In *VLDB*, pages 82–94, 1981.