FIXED-PARAMETER TRACTABLE FRAGMENTS OF CONJUNCTIVE QUERIES AND CONSTRAINT SATISFACTION PROBLEMS

Georg Gottlob[a,b], Matthias Lanzinger[a], Reinhard Pichler[a]

[a]E192 - Institute of Logic and Computation

[b]University of Oxford, United Kingdom

INTRODUCTION

Conjunctive query (CQ) answering is one of the most fundamental problems in computer science. On the theoretical side, CQs are equivalent to constraint satisfaction problems (CSPs) and the homomorphism problem; the central problem of model theory. On the practical side, CQs are the foundation of modern database systems [1] and are widely used – often in the form of CSPs – to implement state of the art scheduling and planning solutions in industry applications.

It is therefore unfortunate that answering CQs belongs to the class of NP-complete problems; which means that all known methods of solving CQs and CSPs require computational effort – and therefore, time – that grows exponentially in the size of the problem, e.g., the database. Considering the scope of modern databases — which frequently exceeds multiple terabytes — it is of great importance for practical applications to study the theory of what precisely makes the problem hard and how to identify and evaluate easy cases. Identifying fragments of the problem, i.e., classes of conjunctive queries that exhibit some specific property, for which evaluation requires subexponential time is one of the main themes in this line of research in general and the presented work in particular.

FUNDAMENTALS OF THE PROBLEM

The problem has been heavily studied in the literature, starting with the well known treewidth which has found wide adoption throughout theoretical computer science as well as real-world applications [2]. Treewidth has some significant shortcomings in the context of CQs and more general methods, based around the idea of structurally decomposing the query, have subsequently been developed to address these shortcomings [3]. In particular, the study of (generalized) hypertree width and its fractional generalization has led to important results, identifying large fragments for which the CQ problem is indeed *tractable*, i.e., solvable in polynomial time.

Recent results have identified even more general fragments — identified by their *submodular width* [4] — for which *fixed-parameter tractable* (fpt) evaluation is possible. In fpt algorithms, the execution time may depend superpolynomially only on *part of the problem* (the namesake fixed parameter). In the context of database queries fixed-parameter tractability becomes very attractive in practice as the query itself is usually vanishingly small in comparison to the size of the database.

RESULTS AND DISCUSSION

In our ongoing research we bring together the principles of structural decomposition that have already been proven successful in this context with another classical theme of CQ research; query minimization and semantic equivalence of queries [5]. We call two queries *semantically equivalent* if they always return the same results over any database. The idea is then to compute the answer to a query by instead computing the answer of an equivalent query with the simplest structure. The main prob-

lem then is that there exist infinitely many equivalent queries, making it very difficult to find the best one. We formalize this plan by introducing the notions of *semantic fractional hypertree width* and *semantic submodular width* as the minimal respective widths over all equivalent queries. In our main result we are able to precisely identify those structures for which these widths are minimal and show that both properties identify significant and new fpt fragments of CQs. Semantic submodular width in particular is shown to induce the most general known fpt fragment, subsuming all other fragments identified in the literature.

CONCLUSION

Figure 1 illustrates our contributions in the context of other important properties from the literature that identify (fixed-parameter) tractable fragments of CQs. These new classes represent a significant advancement over previous results and are a significant step towards the ultimate goal — for theory and applications — of fully characterizing all fixed-parameter tractable conjunctive queries. Furthermore, our approach is the first to combine all the main threads of CQ research, providing a first holistic view of the problem and thereby laying the foundation for new fields of research.

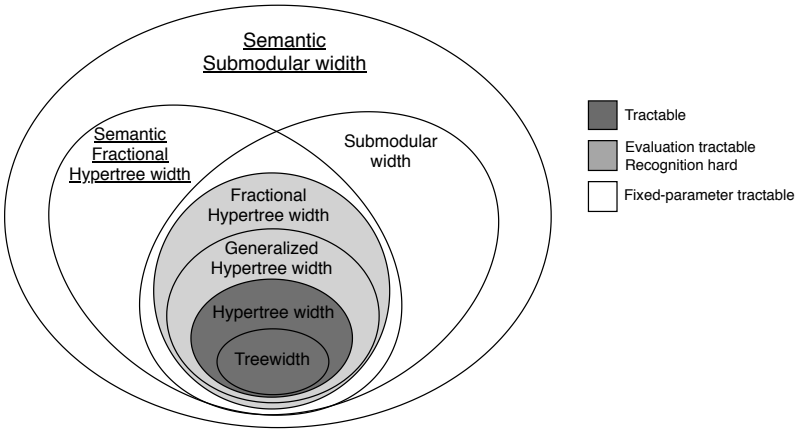

Figure 1: Properties used to identify (fixed-parameter) tractable fragments of CQs. New classes introduced in the presented work are underlined.

This work was supported by the Austrian Science Fund (FWF):P30930-N35.

REFERENCES

[1] S. Abiteboul, R. Hull, and V. Vianu, *Foundations of databases: the logical level*. Addison-Wesley Longman Publishing Co., Inc., 1995.

[2] T. Kloks, *Treewidth: computations and approximations*. Springer Science & Business Media, 1994, vol. 842.

[3] G. Gottlob, G. Greco, N. Leone, and F. Scarcello, "Hypertree decompositions: Questions and answers," in *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. ACM, 2016, pp. 57–74.

[4] D. Marx, "Tractable hypergraph properties for constraint satisfaction and conjunctive queries," *Journal of the ACM*, vol. 60, no. 6, p. 42, 2013.

[5] A. K. Chandra and P. M. Merlin, "Optimal implementation of conjunctive queries in relational data bases," in *Proceedings of the ninth annual ACM symposium on Theory of computing*. ACM, 1977, pp. 77–90.