Semantic Width Revisited (Extended Abstract)

Georg Gottlob[1,2], Matthias Lanzinger[1], and Reinhard Pichler[1]

[1] TU Wien, Austria,
[2] University of Oxford, UK
{gottlob, mlanzing, pichler}@dbai.tuwien.ac.at

CEUR-WS.org/Vol-2369/short01.pdf

1 Introduction

Answering conjunctive queries (CQ) and, equivalently, solving constraint satisfaction problems (CSP), are among the most fundamental tasks in computer science. While these problems are NP-complete, there has been much success in identifying "islands of tractability". In this work, we want to further sharpen the image of this complexity landscape. In particular, we are interested in extending the pioneering work by Barceló, Pieris, Romero, and Vardi [1, 2], Dalmau, Kolaitis, and Vardi [4], and Grohe [6] on the deep connections between query minimization and structural decompositions. To this end, the following notion of *semantic widths* will be the central theme of our work. Note that from now on, we only concentrate on CQs. Clearly, all results hold for CSPs as well.

Definition 1. *Let $\mathcal{Q}$ be the class of all conjunctive queries and $w : \mathcal{Q} \to \mathbb{R}^+$ be some property of the query. We define* semantic w *as* $\texttt{sem-}w(q) := \inf\{w(q') \mid q' \simeq q\}$, *where $\simeq$ denotes homomorphic equivalence.*

Such semantic widths can be interpreted as measures of the inherent structural complexity of the underlying question posed by the query, whereas classical notions of width express the complexity of a specific way to pose the question.

So far, semantic notions of acyclicity [2], treewidth (*tw*) [4], and (generalized) hypertree width (*hw* and *ghw*, resp.) [1] have been investigated, while more powerful notions of width such as fractional hypertree width (*fhw*) [7], submodular width (*subw*) [9], and adaptive width (*adw*) [8] have been left unexplored. The goal of this work is to study semantic notions also of *fhw*, *adw*, and *subw*. Recall that the semantic notions of acyclicity, *tw*, and *ghw* can be characterized in terms of the *core* of the CQ. This naturally raises the question if such a characterization is also possible for *fhw*, *subw*, and *adw*. We will give an affirmative answer which, structurally, looks very similar to the previous results. However, some additional machinery will be needed to prove our new results.

In [9], *subw* was introduced to provide an in some sense "complete" characterization of the fixed-parameter tractability (FPT) of CQ Evaluation. Our investigation of the semantic version of *subw* will provide new input for such an FPT-characterization. More specifically, our new notion of $\texttt{sem-}subw$ will allow us to identify an FPT-fragment of CQ Evaluation which is *strictly bigger* than the fragment defined via *subw* in [9].

Strongly related to the search for (fixed-parameter) tractable fragments of CQ Evaluation is the complexity of the Check problem which, for given integer $k \geq 1$, is about deciding if a given CQ has width $\leq k$. Among the width notions mentioned above, this problem is tractable for *tw* and *hw* and NP-complete for *ghw* and *fhw*. For *subw* and *adw*, the complexity is expected to be even higher. When moving to semantic notions, the computation of the core introduces a further source of intractability. In case of *ghw* and *fhw*, several structural properties of the hypergraph underlying a CQ have been identified recently [5] to make the Check problem tractable. We will show that these properties may also be helpful for the computation of the semantic variants of *ghw* and *fhw*.

2 Preliminaries

We assume the reader to be familiar with basic concepts such as conjunctive queries (CQs) and their associated hypergraphs. We implicitly extend properties of hypergraphs to CQs through their associated hypergraphs. Due to space limitations, we refer to [9] for definitions of the fractional cover number (ρ^*), generalized hypertree width (*ghw*), fractional hypertree width (fhw), adaptive width (adw), and submodular width ($subw$).

We are mainly interested in two computational problems here. The first one is the usual query evaluation problem for a class of CQs $\mathcal{Q}$, which we denote as Eval($\mathcal{Q}$). The decision problem of checking, whether a query has width $\leq k$ for width notion w, will be referred to as Check(w, k).

The problem Check(w, k) with $w \in \{ghw, fhw\}$ has been shown NP-complete even for $k = 2$ [5]. On the positive side, also tractable fragments of this problem have been identified in [5] via the following hypergraph properties: for a hypergraph H, the *c-multi-intersection width* of H refers to the maximum cardinality of an intersection of any c distinct edges. For the special case $c = 2$, we simply use the term *intersection width*. The *degree* of H is the maximum number of edges a vertex of H occurs in. The *rank* of H is the maximum edge size in H.

3 Results

The following theorem is the basis of all our further considerations:

Theorem 1. *For every conjunctive query q:*
1. $\texttt{sem-}\rho^*(q) = \rho^*(Core(q))$
2. $\texttt{sem-}fhw(q) = fhw(Core(q))$
3. $\texttt{sem-}adw(q) = adw(Core(q))$
4. $\texttt{sem-}subw(q) = subw(Core(q))$

An interesting consequence of Theorem 1 is that bounded semantic width of a class $\mathcal{Q}$ of CQs, for the notions of width enumerated in the theorem, implies FPT of the Eval($\mathcal{Q}$) problem when parameterized by the query. This is due to the fact that core computation only depends on the query (not the data). This is of particular interest in the case of bounded $\texttt{sem-}subw$, because it properly generalizes bounded submodular width, and as such "escapes" the FPT-characterization theorem of Marx [9]. To see that the generalization is in fact proper, consider the class of "grid queries" $\mathcal{Q}_{\mathcal{G}_n}$, such that $\mathcal{Q}_{\mathcal{G}_n}$ asks if a given undirected graph contains a grid of size $\geq n$. The core of query $\mathcal{Q}_{\mathcal{G}_n}$ simply asks for the existence of a single (undirected) edge. However, the associated hypergraphs of the family $(\mathcal{Q}_{\mathcal{G}_n})_{n\geq 1}$ include grids of every dimension. Hence, $(\mathcal{Q}_{\mathcal{G}_n})_{n\geq 1}$ has unbounded treewidth. By considering the fractional independent set where every vertex has weight $1/rank(H)$, we get the inequality $subw(H) \geq adw(H) \geq (tw(H)+1)/rank(H)$. It is then clear that $\mathcal{Q}_{\mathcal{G}_n}$ has unbounded $subw$, which is in sharp contrast to $\texttt{sem-}subw(\mathcal{Q}_{\mathcal{G}_n}) = 1$.

Corollary 1. *Let $\mathcal{Q}$ be a class of CQs. Then bounded $\texttt{sem-}fhw$, $\texttt{sem-}subw$, and $\texttt{sem-}adw$ are sufficient conditions for the fixed-parameter tractability of* $\textsc{Eval}(\mathcal{Q})$ *(parameterized by the query). Furthermore, they properly subsume bounded fhw, subw, and adw, respectively.*

In [5], it was shown that the Check problem of *ghw* and *fhw* becomes tractable if the underlying hypergraphs satisfy certain properties such as bounded (multi-)intersection width, bounded degree, and/or bounded rank. Note that all these properties are hereditary in the sense that deletion of vertices and/or edges from a hypergraph does not destroy these properties. Thus, using Theorem 1, we can directly identify tractable fragments of Check for $\texttt{sem-}ghw$ and $\texttt{sem-}fhw$. We present Corollary 2 as an illustrative example for various similar results.

Corollary 2. *For a constant $k > 0$, let $\mathcal{Q}$ be a class of conjunctive queries with bounded fractional hypertree width and bounded degree, then* $\textsc{Check}(\texttt{sem-}fhw, k)$ *is tractable in $\mathcal{Q}$.*

There exist classes with bounded *fhw* but unbounded *ghw* [7]. However, little is known about the conditions under which this can occur. We show that for classes with either bounded degree or bounded intersection width, the properties of bounded *fhw* and bounded *ghw* (and, therefore, also bounded *hw*) in fact coincide. Furthermore, since $\texttt{sem-}fhw$ and $\texttt{sem-}ghw$(cf. [1]) are characterized by the width of the core, it is easy to generalize the result to the semantic case. As a direct consequence, the promise algorithm presented by Chen and Dalmau in [3] can be adapted to classes with bounded $\texttt{sem-}fhw$ and either bounded degree or bounded intersection width, making evaluation of these classes tractable.

Theorem 2. *Let q be a conjunctive query and let its associated hypergraph have degree d and intersection width i. The following statements hold:*

- $fhw(q) \leq ghw(q) \leq d\, fhw(q)$ *(Implicit in [5])*
- $\texttt{sem-}fhw(q) \leq \texttt{sem-}ghw(q) \leq d\, \texttt{sem-}fhw(q)$
- $fhw(q) \leq ghw(q) \leq 2i\, fhw(q)^2 + 2\, fhw(q)$
- $\texttt{sem-}fhw(q) \leq \texttt{sem-}ghw(q) \leq 2i\, \texttt{sem-}fhw(q)^2 + 2\, \texttt{sem-}fhw(q)$

4 Conclusion

So far, we have given a characterization of the semantic variants of ρ^*, fhw, adw, and $subw$. From this we are able to derive new insights into the complexity of CQs. Figure 1 illustrates our current view of the complexity landscape of CQs.

Many consequences of Theorem 1 are yet to be explored. Of particular interest is the role of sem-$subw$. Marx has shown that bounded $subw$ characterizes those hypergraphs for which evaluation of the associated CQs is fixed-parameter tractable [9]. The natural next step is to characterize precisely those CQs for which the evaluation is fixed-parameter tractable. Semantic submodular width is a natural candidate step in this direction but the question whether it provides a necessary condition for fixed-parameter tractable CQ evaluation remains an important open problem for future work.

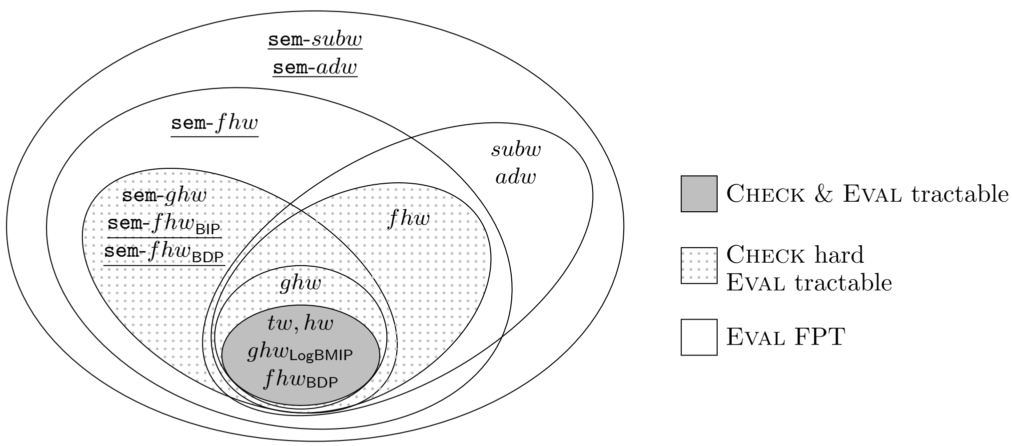

Fig. 1. CQ Complexity Landscape. Contributions presented in this paper are underlined. We write w_P to denote a width notion w on classes constrained to a property P where P is one of bounded degree (BDP), bounded intersection (BIP), or logarithmically bounded multi-intersection (LogBMIP).

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

References

1. Barceló, P., Pieris, A., Romero, M.: Semantic optimization in tractable classes of conjunctive queries. SIGMOD Rec. **46**(2), 5–17 (Sep 2017)
2. Barceló, P., Romero, M., Vardi, M.Y.: Semantic acyclicity on graph databases. SIAM Journal on Computing **45**(4), 1339–1376 (2016)
3. Chen, H., Dalmau, V.: Beyond hypertree width: Decomposition methods without decompositions. In: Proc. CP 2005. pp. 167–181. Springer (2005)

4. Dalmau, V., Kolaitis, P.G., Vardi, M.Y.: Constraint satisfaction, bounded treewidth, and finite-variable logics. In: Proc. CP 2002. pp. 310–326. Springer (2002)
5. Fischl, W., Gottlob, G., Pichler, R.: General and fractional hypertree decompositions: Hard and easy cases. In: Proc. PODS 2018. pp. 17–32. ACM (2018)
6. Grohe, M.: The complexity of homomorphism and constraint satisfaction problems seen from the other side. Journal of the ACM (JACM) **54**(1), 1 (2007)
7. Grohe, M., Marx, D.: Constraint solving via fractional edge covers. ACM Trans. Algorithms **11**(1), 4:1–4:20 (2014)
8. Marx, D.: Tractable structures for constraint satisfaction with truth tables. Theory of Computing Systems **48**(3), 444–464 (2011)
9. Marx, D.: Tractable hypergraph properties for constraint satisfaction and conjunctive queries. Journal of the ACM **60**(6), 42 (2013)