Information and Computation 222 (2013) 108–120

Contents lists available at SciVerse ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Generic expression hardness results for primitive positive formula comparison

Simone Bova [a], Hubie Chen [b,c,*], Matthew Valeriote [d]

[a] Department of Mathematics, Vanderbilt University, Nashville, USA
[b] Facultad de Informática, Universidad del País Vasco, E-20080, San Sebastián, Spain
[c] IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao, Spain
[d] Department of Mathematics & Statistics, McMaster University, Hamilton, Canada

ARTICLE INFO

Article history:
Available online 26 October 2012

Keywords:
Primitive positive formulas
Conjunctive queries
Equivalence
Computational complexity
Clones

ABSTRACT

We study the expression complexity of two basic problems involving the comparison of primitive positive formulas: equivalence and containment. In particular, we study the complexity of these problems relative to finite relational structures. We present two generic hardness results for the studied problems, and discuss evidence that they are optimal and yield, for each of the problems, a complexity trichotomy.

1. Introduction

Overview. A *primitive positive* (*pp*) formula is a first-order formula defined from atomic formulas and equality of variables using conjunction and existential quantification. The class of primitive positive formulas includes, and is essentially equivalent to, the class of *conjunctive queries*, which is well-established in relational database theory as a pertinent and useful class of queries, and which has been studied complexity-theoretically from a number of perspectives; see for example [1–3]. In this paper, we study the complexity of the following fundamental problems, each of which involves the comparison of two pp-formulas ϕ, ϕ' having the same free variables, over a relational structure.

- Equivalence: are the formulas ϕ, ϕ' equivalent – that is, do they have the same satisfying assignments – over the structure?
- Containment: are the satisfying assignments of ϕ contained in those of ϕ', over the structure?

We study the complexity of these computational problems with respect to various fixed structures. That is, we parameterize each of these problems with respect to the structure to obtain a family of problems, containing one member for each structure, and study the resulting families of problems. To employ the terminology of Vardi [4], we study the *expression complexity* of the presented comparison tasks. The suggestion here is that various relational structures – which may represent databases or knowledge bases, according to use – may possess structural characteristics that affect the complexity of the resulting problems, and our interest is in understanding this interplay. The present work focuses on relational structures that are finite (that is, have finite universe), and we assume that the structures under discussion are finite.

In this paper, we present two general expression hardness results on the problems of interest. In particular, each of our two main results provides a sufficient condition on a structure so that the problems are hard for certain complexity

* Corresponding author at: Facultad de Informática, Universidad del País Vasco, E-20080, San Sebastián, Spain.
E-mail addresses: simone.bova@vanderbilt.edu (S. Bova), hubie.chen@ehu.es (H. Chen), matt@math.mcmaster.ca (M. Valeriote).

0890-5401/$ – see front matter
http://dx.doi.org/10.1016/j.ic.2012.10.008

classes. Furthermore, we give evidence that our results are optimal, in that the conditions that they involve in fact describe dichotomies in the complexity of the studied problems; put together, our results indicate, for each of the studied problems, a complexity trichotomy.

Our study utilizes universal-algebraic tools that aid in understanding the set of primitive positively definable relations over a given structure. It is known that, relative to a structure, the set of relations that are definable by a primitive positive formula forms a robust algebraic object known as a *relational clone*; a known Galois correspondence associates, in a bijective manner, each such relational clone with a *clone*, a set of operations with certain closure properties. This correspondence provides a way to pass from a relational structure **B** to an algebra $\mathbb{A}_{\mathbf{B}}$ whose set of operations is the mentioned clone, in such a way that two structures having the same algebra have the same complexity (for each of the mentioned problems). In a previous paper by the present authors [5], we developed this correspondence and presented some basic complexity results for the problems at hand, including a classification of the complexity of the problems on all two-element structures.

Our hardness results. Our first hardness result (Section 3) yields that for any structure **B** whose associated algebra $\mathbb{A}_{\mathbf{B}}$ gives rise to a variety $\mathcal{V}(\mathbb{A}_{\mathbf{B}})$ that *admits the unary type*, both the equivalence and containment problems are Π_2^p-hard. Note that this is the maximal complexity possible for these problems, as the problems are contained in the class Π_2^p. The condition of admitting the unary type originates from tame congruence theory, a theory developed to understand the structure of finite algebras [6]. We observe that this result implies a dichotomy in the complexity of the studied problems under the *G-set conjecture* for the constraint satisfaction problem (CSP), a conjecture put forth by Bulatov, Jeavons, and Krokhin [7] which predicts exactly where the tractability/intractability dichotomy lies for the CSP. (Recall that the CSP can be formulated as the problem of deciding, given a structure **B** and a primitive positive sentence ϕ, whether or not ϕ holds on **B**.) In particular, under the G-set conjecture, the structures not obeying the described condition have equivalence and containment problems in coNP. The resolution of the G-set conjecture, on which there has been focused and steady progress over the past decade [8–11], would thus, in combination with our hardness result, yield a coNP/Π_2^p-complete dichotomy for the equivalence and containment problems. In fact, our hardness result already unconditionally implies dichotomies for our problems for all classes of structures where the G-set conjecture has already been established, including the class of three-element structures [8], and the class of conservative structures [12].

One formulation of the G-set conjecture is that, for a structure **B** whose associated algebra $\mathbb{A}_{\mathbf{B}}$ is idempotent, the absence of the unary type in the variety generated by $\mathbb{A}_{\mathbf{B}}$ implies that the CSP over **B** is polynomial-time tractable. The presence of the unary type is a known sufficient condition for intractability in the idempotent case [7,10], and this conjecture predicts exactly where the tractability/intractability dichotomy lies for the CSP. It should be noted, however, that the boundary that is suggested by our hardness result for the equivalence and containment problems is *not* the same as the boundary suggested by the G-set conjecture for the CSP. The G-set conjecture, which is typically phrased on idempotent algebras, yields a prediction on the CSP complexity of all structures via a theorem [7] showing that each structure **B** has the same CSP complexity as a structure $\mathbf{B}'$ whose associated algebra is idempotent. The mapping from **B** to $\mathbf{B}'$ does not preserve the complexity of the problems studied here, and indeed, there are examples of two-element structures **B** such that our hardness result applies to **B** – the equivalence and containment problems on **B** are Π_2^p-complete – but $\mathbf{B}'$ does not admit the unary type and indeed has a polynomial-time tractable CSP [5]. Our new result requires establishing a deeper understanding of the identified algebras' structure, some of which admit a tractable CSP, in order to obtain hardness.

Our second hardness result (Section 4) shows that for any structure **B**, if the variety $\mathcal{V}(\mathbb{A}_{\mathbf{B}})$ is not congruence modular, then the equivalence and containment problems are coNP-hard. Previous work identified one most general condition for the tractability of the equivalence and containment problems: if the algebra *has few subpowers* – a combinatorial condition [13,9] involving the number of subalgebras of powers of an algebra – then these problems are polynomial-time tractable [5, Theorem 7]. This second hardness result appears to perfectly complement this tractability result: there are no known examples of algebras $\mathbb{A}_{\mathbf{B}}$ (of structures **B** having finitely many relations) that are not covered by one of these results, and in fact the *Edinburgh conjecture* predicts that none exist, stating that every such algebra $\mathbb{A}_{\mathbf{B}}$ that generates a congruence modular variety also has few subpowers. Concerning this conjecture, it should be pointed out that the resolution of the *Zadóri conjecture*, a closely related conjecture of which the Edinburgh conjecture is a generalization, was recently announced by Libor Barto [14]. The Edinburgh conjecture is of current interest, with recent work presented by Ralph McKenzie and colleagues. Added while in press. Libor Barto has recently announced a proof of the Edinburgh conjecture. We also point out that this conjecture (as with the Zadóri conjecture) is purely algebraic, making no references to notions of computation.

In summary, up to polynomial-time computation, **we completely resolve the complexity of the studied problems on all finite structures, showing a P/coNP-complete/Π_2^p-complete trichotomy – modulo two conjectures;** one is computational and one is algebraic, and for each there is both highly nontrivial supporting evidence and current investigation. The coNP/Π_2^p-complete dichotomy is presented in Section 3 (see Theorems 4 and 5) and the P/coNP-hard dichotomy is presented in Section 4 (see Theorems 7 and 8).

2. Preliminaries

Here, a *signature* is a set of relation symbols, each having an associated arity; we assume that all signatures are of finite size. A *relational structure* over a signature σ consists of a universe B and, for each relation symbol $R \in \sigma$, a relation $R^{\mathbf{B}} \subseteq B^k$ where k is the arity of R. We assume that all relational structures under discussion have universes of finite size. A *primitive*