Classification of Annotation Semirings over Query Containment

Egor V. Kostylev
University of Edinburgh
ekostyle@inf.ed.ac.uk

Juan L. Reutter
University of Edinburgh
juan.reutter@ed.ac.uk

András Z. Salamon
University of Edinburgh
andras.salamon@ed.ac.uk

ABSTRACT

We study the problem of query containment of (unions of) conjunctive queries over annotated databases. Annotations are typically attached to tuples and represent metadata such as probability, multiplicity, comments, or provenance. It is usually assumed that annotations are drawn from a commutative semiring. Such databases pose new challenges in query optimization, since many related fundamental tasks, such as query containment, have to be reconsidered in the presence of propagation of annotations.

We axiomatize several classes of semirings for each of which containment of conjunctive queries is equivalent to existence of a particular type of homomorphism. For each of these types we also specify all semirings for which existence of a corresponding homomorphism is a sufficient (or necessary) condition for the containment. We exploit these techniques to develop new decision procedures for containment of unions of conjunctive queries and axiomatize corresponding classes of semirings. This generalizes previous approaches and allows us to improve known complexity bounds.

Categories and Subject Descriptors

H.2.4 [**Database Management**]: Systems—*query processing*; H.2.1 [**Database Management**]: Data Models

Keywords

Annotation, Provenance, Query Optimization

1. INTRODUCTION

Relational database annotation is rapidly coming to market. The expressive power of curated [2] and probabilistic databases [12, 21], various forms of provenance [9, 3, 15], and even bag semantics as a way to model standard SQL [6], derives from an annotation attribute with special behaviour. In [15] it was observed that in all of these cases annotations propagate through queries as we expect if the domain of annotations has the structure of a *commutative semiring*.

PODS '12, May 21–23, 2012, Scottsdale, Arizona, USA.

Every application that supports annotations should be able to compare queries to perform standard tasks such as query rewriting and query optimization. However, as noted in [17, 6] for the particular case of bag semantics and quite generally in [14], the introduction of annotations requires a complete rethinking of these kinds of tasks: a pair of queries may behave differently when posed over ordinary relations or over annotated relations; the behaviour can be different even for different semirings. Hence a general theory is needed to explain how queries behave over annotated relations, and to provide query optimization and query rewriting techniques, regardless of the semiring chosen for annotations.

In this paper, we study the problem of *containment* of queries, specifically for the classes of *conjunctive queries* (CQs) and *unions of conjunctive queries* (UCQs). For this purpose we formally generalize the standard notion of containment for relational databases so that it subsumes previously studied containments for bag semantics [17, 6] and several naturally ordered semirings [14]. We study in our view the most general reasonable notion of containment, based on a few intuitive axioms which any containment should satisfy.

The ideal would be to obtain a decision procedure to decide containment of CQs or UCQs, for an arbitrary annotation semiring. However, there is evidence that obtaining such a procedure for all semirings is a truly challenging, if not impossible, task. Indeed, this would require solving containment for bag semantics, which is undecidable for UCQs [17] (and even for CQs with inequalities [18]), and is a long-standing open problem for CQs [6, 17, 1, 7]. With these observations in mind, we instead ask the following, narrower question: are there reasonable classes of semirings for which we can prove that containment of CQs or UCQs is decidable? In this paper we answer this question positively, by finding several such classes. Our main results generalize and extend previous work [14, 13] unifying how semantic properties of query containment link to syntactic properties of different types of *homomorphisms* between queries. We also show that these classes are of importance in practice, as they contain the majority of the annotation semirings that have been proposed.

In Sec. 3 we begin our study with containment of CQs. For standard relational databases (which can be modelled by a set semantics semiring consisting of just two elements **true** and **false**), query containment corresponds precisely to the NP-complete problem of deciding whether there exists a homomorphism between these queries [5]. Thus, the natural starting point of our search for decidable classes is to ask for which semirings the CQ containment problem coin-

cides with CQ containment for the usual set semantics. This question was partially answered in [17], where for semirings which are so called type A systems, containment was shown to be equivalent to the existence of a homomorphism. We show that it is possible to describe the class $\mathbf{C}_{\text{hom}}$ of *all* such semirings by two simple axioms: idempotence of multiplication and annihilation of the multiplicative identity. Interestingly, this class corresponds precisely to the class of type A′ systems [17], for which such a characterization was open.

Continuing our search for decidable classes, in Sec. 4 we consider those classes obtained by relaxing the axioms for $\mathbf{C}_{\text{hom}}$. In Sec. 4.1 to 4.4 we show that for each of these classes there exists a well-known natural type of homomorphism that characterizes the class, but only as a *sufficient condition* for containment of CQs. As an example, consider the class of semirings that satisfy only the annihilation axiom. In Sec. 4.2 we demonstrate that this class contains precisely all the semirings for which the existence of an *injective homomorphism* is a sufficient condition for containment of two CQs. A sufficient condition does not guarantee the decidability of the containment problem; one needs a necessary condition as well. For this purpose, we describe the *largest* class for which an injective homomorphism is necessary for containment of CQs. Thereby, we have that for all semirings in the intersection of these two classes, the existence of an injective homomorphism is both a necessary and sufficient condition for the containment of two CQs, resulting in a class of semirings for which containment is decidable.

We establish similar results for several other classes of semirings obtained by relaxing the axioms that define the class $\mathbf{C}_{\text{hom}}$, and show how these classes are characterized by other well known types of homomorphism. This gives us decision procedures for the corresponding classes of semirings. We provide tight complexity bounds for these procedures: all of them are NP-complete, just as for the class $\mathbf{C}_{\text{hom}}$ [5].

To describe some of these classes, we introduce the notion of CQ-*admissible polynomials*. A polynomial is CQ-admissible if it can be obtained by evaluating a CQ over a database annotated with variables. In Sec. 4.5 we give a syntactic characterization of these polynomials, which allows us to axiomatize several classes of semirings. This novel concept is of independent interest; e.g. in [19] the properties of such polynomials were used implicitly for effectively storing and manipulating the provenance of CQ results.

Moving beyond homomorphisms, in Sec. 4.6 we also find several semirings for which containment can be solved via a *small model property*, by looking for a small enough database witness for absence of containment. This results in new decision procedures to solve containment of CQs, for a wide range of semirings that had not been previously addressed.

In Sec. 5 we develop decision procedures to solve the containment of UCQs, which naturally extend the procedures we used for the case of CQs. In this respect, most of the existing positive results correspond to semirings for which one can decide containment of UCQs by checking their elements locally, one by one [14]. We identify the relationship of this property with idempotence of the semiring's addition, which gives us a decision procedure for containment of UCQs for several classes of semirings, including $\mathbf{C}_{\text{hom}}$. However, for other classes this simple local method does not work. To overcome its limitations, we introduce the notion of a *complete description* of a UCQ (inspired by [11] and [8]), which is a union of special CQs with inequalities, equivalent to the original UCQ. This allows us to identify classes of semirings for which a modified local method works, though applied not to UCQs themselves but to their complete descriptions.

Complete descriptions open up new possibilities for deciding containment of UCQs over different semantics. Our machinery allows us to devise in Sec. 5.2 a syntactic condition for checking containment of UCQs for the important semiring of provenance polynomials [15], for which only a small model property based approach was known [14]; and based on this, to improve the complexity upper bound. Also, complete descriptions allow us to improve the existing sufficient and necessary conditions for containment of UCQs over bag semantics.

2. PRELIMINARIES

Commutative semirings An algebraic structure $\mathcal{K} = \langle K, \oplus, \otimes, \mathbb{0}, \mathbb{1}\rangle$ with binary operations *sum* $\oplus$ and *product* $\otimes$ and constants $\mathbb{0}$ and $\mathbb{1}$ is a (*commutative*) *semiring* iff $\langle K, \oplus, \mathbb{0}\rangle$ and $\langle K, \otimes, \mathbb{1}\rangle$ are commutative monoids[1] with identities $\mathbb{0}$ and $\mathbb{1}$ respectively, $\otimes$ is distributive over $\oplus$, and $a \otimes \mathbb{0} = \mathbb{0}$ holds for each $a \in K$. It will be convenient for us to consider only *nontrivial* semirings, i.e. semirings such that $\mathbb{0} \neq \mathbb{1}$. We use the symbols $\sum$ and $\prod$ to denote sum and product of sets of semiring elements in the usual way.

$\mathcal{K}$-relations A *schema* $\mathbb{S}$ is a finite set of *relational symbols*, each of which is assigned a non-negative *arity*. For a semiring $\mathcal{K} = \langle K, \oplus, \otimes, \mathbb{0}, \mathbb{1}\rangle$ and an infinite domain $\mathbb{D}$ of constants, a $\mathcal{K}$-*instance* I over a schema $\mathbb{S}$ assigns to each relational symbol R from $\mathbb{S}$ of arity m a $\mathcal{K}$-*relation* R^I, which is a (total) function from the set of *tuples* $\mathbb{D}^m$ to K such that its *support*, i.e. the set $\{\mathbf{t} \mid \mathbf{t} \in \mathbb{D}^m, R^I(\mathbf{t})) \neq \mathbb{0}\}$, is finite. We call $R^I(\mathbf{t})$ the *annotation* of the tuple $\mathbf{t}$ in R^I.

Queries A *conjunctive query* (*CQ*) Q over a schema $\mathbb{S}$ is an expression of the form $\exists \mathbf{v}\ \phi(\mathbf{u}, \mathbf{v})$, where $\mathbf{u}$ is a list of *free* variables, $\mathbf{v}$ is a list of *existential* variables and $\phi(\mathbf{u}, \mathbf{v})$ is a multiset of relational atoms over $\mathbb{S}$ using variables $\mathbf{u} \cup \mathbf{v}$. As usual we write $\phi(\mathbf{u}, \mathbf{v}) = R_1(\mathbf{u}_1, \mathbf{v}_1), \ldots, R_n(\mathbf{u}_n, \mathbf{v}_n)$, where $\mathbf{u}_1 \cup \ldots \cup \mathbf{u}_n = \mathbf{u}$ and $\mathbf{v}_1 \cup \ldots \cup \mathbf{v}_n = \mathbf{v}$, keeping in mind that R_i and R_j in this expression can be the same symbol. A *union* of conjunctive queries (*UCQ*) $\mathbf{Q}$ is a multiset of CQs over the same schema and the same set of free variables.

Evaluations For CQ $Q = \exists \mathbf{v}\ R_1(\mathbf{u}_1, \mathbf{v}_1), \ldots, R_n(\mathbf{u}_n, \mathbf{v}_n)$ and a tuple $\mathbf{t}$, denote by $\mathcal{V}(Q, \mathbf{t})$ the set of all mappings f from $\mathbf{u} \cup \mathbf{v}$ to the domain $\mathbb{D}$ such that $f(\mathbf{u}) = \mathbf{t}$. Given a $\mathcal{K}$-instance I, the *evaluation* of Q on I for $\mathbf{t}$ is the value

$$Q^I(\mathbf{t}) = \sum_{f \in \mathcal{V}(Q,\mathbf{t})} \prod_{1 \leq i \leq n} R_i^I(f(\mathbf{u}_i, \mathbf{v}_i)).$$

Similarly, the *evaluation* of a UCQ $\mathbf{Q}$ on I for $\mathbf{t}$ is

$$\mathbf{Q}^I(\mathbf{t}) = \sum_{Q \in \mathbf{Q}} Q^I(\mathbf{t}).$$

From this definition it follows that if $\mathbf{Q} = \emptyset$ then $\mathbf{Q}^I(\mathbf{t}) = \mathbb{0}$.

3. GENERAL FRAMEWORK

3.1 $\mathcal{K}$-containment and positive semirings

As noted in [15], the introduction of annotations on relations requires a complete rethinking of the notions of query

[1] A commutative monoid is a set with an associative and commutative binary operation and an identity element.

optimization and query rewriting. In particular, it was first discovered in [6] that two queries that are equivalent when posed over ordinary relations may not be equivalent when evaluated on $\mathcal{K}$-relations. Furthermore, for two different semirings $\mathcal{K}_1$ and $\mathcal{K}_2$, two queries may be equivalent under $\mathcal{K}_1$-relations, but not equivalent under $\mathcal{K}_2$-relations.

Our main aim is to explore the problem of query containment over different $\mathcal{K}$-relations. First we need to formally specify what we mean by "equivalence" and "containment" of queries. The notion of equivalence is naturally formalised as follows: given a semiring $\mathcal{K}$, UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$ over the same schema are *$\mathcal{K}$-equivalent* (denoted $\mathbf{Q}_1 \equiv_\mathcal{K} \mathbf{Q}_2$) iff for every $\mathcal{K}$-instance I and tuple $\mathbf{t}$ it holds that $\mathbf{Q}_1^I(\mathbf{t}) = \mathbf{Q}_2^I(\mathbf{t})$. However, to study containment of queries over some semiring $\mathcal{K}$, we should be able to compare elements of $\mathcal{K}$ not only for equality. Therefore, we assume that the semiring $\mathcal{K}$ is equipped with a *partial order*[2] $\preceq_\mathcal{K}$. This allows us to define when a UCQ $\mathbf{Q}_1$ is *$\mathcal{K}$-contained* in a UCQ $\mathbf{Q}_2$, which we denote by $\mathbf{Q}_1 \subseteq_\mathcal{K} \mathbf{Q}_2$:

$$\mathbf{Q}_1 \subseteq_\mathcal{K} \mathbf{Q}_2 \iff \forall I\ \forall \mathbf{t}\ \mathbf{Q}_1^I(\mathbf{t}) \preceq_\mathcal{K} \mathbf{Q}_2^I(\mathbf{t}).$$

We always assume that $\preceq_\mathcal{K}$ is *minimal* with respect to $\subseteq_\mathcal{K}$, i.e. there is no subrelation of $\preceq_\mathcal{K}$ that produces the same $\mathcal{K}$-containment. However, for some partial orders the above definition results in a rather spartan notion of $\mathcal{K}$-containment. For example, by considering the usual order $\leq$ on the semiring $\mathbb{Z}$ of *integers*, one can easily verify that the empty UCQ is not $\mathbb{Z}$-contained in any non-empty UCQ.

Thus, we need to restrict the class of partially ordered semirings that we consider for our study. In order to do so, we list four intuitive requirements that, in our view, any definition of $\mathcal{K}$-containment should satisfy, and then identify all the semirings $\mathcal{K}$ equipped with partial orders $\preceq_\mathcal{K}$ for which the definition of $\mathcal{K}$-containment is guaranteed to satisfy our requirements. These requirements are as follows:

(C1) $\subseteq_\mathcal{K}$ is a *preorder*, i.e. reflexive and transitive;
(C2) $\mathbf{Q}_1 \equiv_\mathcal{K} \mathbf{Q}_2$ iff $\mathbf{Q}_1 \subseteq_\mathcal{K} \mathbf{Q}_2$ and $\mathbf{Q}_2 \subseteq_\mathcal{K} \mathbf{Q}_1$;
(C3) $\emptyset \subseteq_\mathcal{K} \mathbf{Q}$ holds for all $\mathbf{Q}$;
(C4) if $\mathbf{Q}_1 \subseteq_\mathcal{K} \mathbf{Q}_2$ then $\mathbf{Q}_1 \cup \mathbf{Q}_3 \subseteq_\mathcal{K} \mathbf{Q}_2 \cup \mathbf{Q}_3$ for any $\mathbf{Q}_3$.

Note that requirement (C3) rules out the example with $\mathbb{Z}$ and $\leq$. It turns out that we can easily axiomatize the class of semirings with partial orders that have $\mathcal{K}$-containments satisfying (C1) – (C4). The following proposition says that this class consists of all *positive* semirings, i.e. semirings $\mathcal{K} = \langle K, \oplus, \otimes, \mathbb{0}, \mathbb{1}\rangle$ equipped with a partial order $\preceq_\mathcal{K}$, such that

- $\mathbb{0} \preceq_\mathcal{K} a$ for all $a \in K$, and
- $a \preceq_\mathcal{K} b \Rightarrow a \oplus c \preceq_\mathcal{K} b \oplus c$ for all $a, b, c \in K$.

Proposition 3.1 *A semiring $\mathcal{K}$ equipped with a partial order $\preceq_\mathcal{K}$ is positive iff the corresponding $\mathcal{K}$-containment $\subseteq_\mathcal{K}$ satisfies requirements* (C1) – (C4).

We assume for the rest of the paper that all semirings are positive and denote the class of such semirings by $\mathbf{S}$.

We focus in this work on the following decision problems:

CQ $\mathcal{K}$-Containment	
Input:	CQs Q_1, Q_2
Question:	Is $Q_1 \subseteq_\mathcal{K} Q_2$?

UCQ $\mathcal{K}$-Containment	
Input:	UCQs $\mathbf{Q}_1$, $\mathbf{Q}_2$
Question:	Is $\mathbf{Q}_1 \subseteq_\mathcal{K} \mathbf{Q}_2$?

[2]A partial order is a transitive, reflexive and antisymmetric binary relation.

In particular, we are interested in classifying the semirings in $\mathbf{S}$ for which different conditions on CQs (and UCQs) are sufficient for $\mathcal{K}$-containment, and also for which semirings they are necessary. If for a semiring $\mathcal{K}$ such a condition is both sufficient and necessary, and it is possible to check the condition algorithmically, then we have a decision procedure for $\mathcal{K}$-containment.

3.2 Naturally ordered semirings and provenance polynomials

In [14] it was noted that in most semantics considered so far, including set and bag semantics, the notion of containment is based on natural orders of the semirings. A semiring $\mathcal{K} = \langle K, \oplus, \otimes, \mathbb{0}, \mathbb{1}\rangle$ is *naturally ordered* iff the preorder $\preceq_\mathcal{K}^{\text{nat}}$, defined as $a \preceq_\mathcal{K}^{\text{nat}} b \iff \exists c\ a \oplus c = b$, is a partial order. In principle, this condition appears to be too restrictive, and for this reason we have opted for the more general approach based on positive semirings. It is straightforward to show that any naturally ordered semiring is a positive semiring. Thus, our approach is general enough to include all previous work, as far as we are aware.

In [14] the problem of $\mathcal{K}$-containment of CQs and UCQs was considered for several naturally ordered semirings, including the one known as the semiring of *provenance polynomials*, $\mathcal{N}[X] = \langle \mathbb{N}[X], +, \times, 0, 1\rangle$. This is the set $\mathbb{N}[X]$ of polynomials over a set of variables X, with natural number coefficients, equipped with the usual operations $+$ and $\times$. In [15] it was pointed out that this semiring (without any order) is special among such semirings since it is "most general", i.e. possesses the *universal property*: for any (unordered) semiring $\mathcal{K} = \langle K, \oplus, \otimes, \mathbb{0}, \mathbb{1}\rangle$ any function $\nu\colon X \to K$ can be uniquely extended to a *morphism* $\mathsf{Eval}_\nu\colon \mathbb{N}[X] \to K$, i.e. a mapping between semirings which preserves all the operations and relations (including constants 0 and 1). In [14] it was shown that $\mathcal{N}[X]$, now with its natural order, is universal for all naturally ordered semirings. It turns out that this is also true for all (positive) semirings.

Proposition 3.2 *Given a set of variables X, $\mathcal{N}[X]$ is universal for the set $\mathbf{S}$ of all (positive) semirings.*

Based on this property, we can formulate different universal axioms on semirings, involving the order $\preceq_\mathcal{K}$, in terms of $\mathcal{N}[X]$. Given a semiring $\mathcal{K} = \langle K, \oplus, \otimes, \mathbb{0}, \mathbb{1}\rangle$ from $\mathbf{S}$, a set X of n variables, and polynomials P_1 and P_2 from $\mathbb{N}[X]$, we write $\mathsf{P}_1 \preceq_\mathcal{K} \mathsf{P}_2$ iff *for all* values $a_1, \ldots, a_n$ from K, the inequality $\mathsf{P}_1(a_1, \ldots, a_n) \preceq_\mathcal{K} \mathsf{P}_2(a_1, \ldots, a_n)$ holds. Here $\mathsf{P}_1(a_1, \ldots, a_n)$ and $\mathsf{P}_2(a_1, \ldots, a_n)$ denote the valuations of P_1 and P_2 over values $a_1, \ldots, a_n$. Since $\preceq_\mathcal{K}$ is a partial order, we can also write $\mathsf{P}_1 =_\mathcal{K} \mathsf{P}_2$ for $\mathsf{P}_1 \preceq_\mathcal{K} \mathsf{P}_2 \wedge \mathsf{P}_2 \preceq_\mathcal{K} \mathsf{P}_1$. We will use such polynomial notation throughout the paper.

3.3 Containment by homomorphisms

The study of query containment in the context of query optimization had begun for relational databases by the 1970s [5]. These databases can be naturally modelled by $\mathcal{B}$-relations, where $\mathcal{B} = \langle \{\mathbf{false}, \mathbf{true}\}, \vee, \wedge, \mathbf{false}, \mathbf{true}\rangle$ is the *set semantics* semiring. Here a tuple is annotated with **true** iff it is in the relation and **false** otherwise. For $\mathcal{B}$-containment the natural order $\preceq_\mathcal{B}$ is assumed, which is defined as **false** $\preceq_\mathcal{B}$ **true**. A CQ Q_1 is $\mathcal{B}$-contained in a CQ Q_2 iff one can find a homomorphism from Q_2 to Q_1, by the classical result of [5]. Given CQs $Q_1 = \exists \mathbf{v}_1\ \phi_1(\mathbf{u}_1, \mathbf{v}_1)$ and $Q_2 = \exists \mathbf{v}_2\ \phi_2(\mathbf{u}_2, \mathbf{v}_2)$, a *homomorphism* (also known as *containment mapping*) from

Q_2 to Q_1 is a function $h\colon \mathbf{u}_2 \cup \mathbf{v}_2 \to \mathbf{u}_1 \cup \mathbf{v}_1$ such that $h(\mathbf{u}_2) = \mathbf{u}_1$ and for each atom $R(\mathbf{u}, \mathbf{v})$ from $\phi_2(\mathbf{u}_2, \mathbf{v}_2)$, the atom $R(h(\mathbf{u}, \mathbf{v}))$ is in $\phi_1(\mathbf{u}_1, \mathbf{v}_1)$. A homomorphism extends to atoms and sets of atoms in the usual way. We write $Q_2 \to Q_1$ iff there exists a homomorphism from Q_2 to Q_1.

Based on the results of [14] or [17] it is straightforward to show that the existence of a homomorphism between CQs is necessary for their $\mathcal{K}$-containment over any semiring $\mathcal{K}$ from $\mathbf{S}$. Thus, a first natural question to ask is: which semirings behave the same as $\mathcal{B}$ with respect to containment of CQs, i.e. for which semirings $\mathcal{K}$ is it the case that $Q_2 \to Q_1$ is sufficient for $Q_1 \subseteq_{\mathcal{K}} Q_2$? This question was answered partially in [15, 14, 17]. In [13] it was shown that this correspondence holds if $\mathcal{K}$ is a distributive bilattice. As the main result of this section we show that it is possible to axiomatize the class of all semirings for which $\mathcal{K}$-containment of CQs coincides with the usual set semantics containment.

Denote by $\mathbf{C}_{\text{hom}}$ the class of semirings $\mathcal{K}$ that satisfy the following axioms (using the convenient polynomial notation introduced at the end of Sec. 3.2, i.e. assuming that all variables are universally quantified):

1. ($\otimes$-*idempotence*) $x \times x =_{\mathcal{K}} x$;
2. ($\mathbb{1}$-*annihilation*) $1 + x =_{\mathcal{K}} 1$.

Next we will see that $\mathbf{C}_{\text{hom}}$ contains exactly all semirings that behave like set semantics, w.r.t. $\mathcal{K}$-containment of CQs.

THEOREM 3.3 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{hom}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ iff $Q_2 \to Q_1$, for all CQs Q_1 and Q_2.*

Deciding the existence of a homomorphism between CQs is well known to be NP-complete [5]. We therefore obtain the following corollary.

COROLLARY 3.4 *If $\mathcal{K} \in \mathbf{C}_{\text{hom}}$ then* CQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete.*

Many semirings used for annotations are distributive lattices, and hence belong to $\mathbf{C}_{\text{hom}}$. Besides the set semantics $\mathcal{B}$, they include the semiring of positive boolean expressions $\texttt{PosBool}[X]$ described in [15], which is used in incomplete databases [16], and the probabilistic semiring $\mathcal{P}[\Omega]$ used in event tables [12, 21]. Also, the class $\mathbf{C}_{\text{hom}}$ corresponds precisely to the class of *type* A′ *systems* introduced in [17]; in fact, Thm. 3.3 answers the question from this paper, of what the decision procedure is for CQ containment over such systems. However, many annotation semirings do not belong to $\mathbf{C}_{\text{hom}}$, including provenance polynomials $\mathcal{N}[X]$, the why-provenance semiring $\texttt{Why}[X]$ from [3], or bag semantics $\mathcal{N}$ [6]. In the next section, we study what happens when we relax the conditions for $\mathbf{C}_{\text{hom}}$.

4. $\mathcal{K}$-CONTAINMENT OF CQS

From a practical point of view, it would be useful to have a decision procedure for $\mathcal{K}$-containment of CQs for an arbitrary semiring $\mathcal{K}$. However, there is evidence that obtaining such a procedure for all semirings not in $\mathbf{C}_{\text{hom}}$ is a truly challenging, if not impossible, task. The semiring $\mathcal{N} = \langle \mathbb{N}_0, +, \times, 0, 1\rangle$ of natural numbers with zero, with the usual arithmetic operations and the natural order, is used to model bag semantics [15]. A universal decision procedure for CQ $\mathcal{K}$-CONTAINMENT would thus require being able to solve this problem for the special case of bag semantics $\mathcal{N}$, which is a long-standing open problem [6, 17]. It is also not difficult to show that there are infinitely many semirings $\mathcal{K}$ for which the $\mathcal{K}$-containment of CQs is at least as hard as for bag semantics.

With these observations in mind, we instead ask the following, narrower question: are there any reasonable classes of semirings for which we can prove that $\mathcal{K}$-containment of CQs is decidable? We have already pointed out that this is the case for the class $\mathbf{C}_{\text{hom}}$, since for all semirings $\mathcal{K}$ in $\mathbf{C}_{\text{hom}}$ the problem of $\mathcal{K}$-containment can be solved by deciding the existence of a homomorphism. A natural starting point for our search is therefore to relax the axioms of the class $\mathbf{C}_{\text{hom}}$. We thus obtain the class of semirings that satisfy the $\otimes$-idempotence axiom, that we denote by $\mathbf{S}_{\text{hcov}}$; the class of semirings that satisfy the $\mathbb{1}$-annihilation axiom, denoted by $\mathbf{S}_{\text{in}}$; and, if we relax both axioms, the class $\mathbf{S}$ of all (positive) semirings.

We show that for each of these classes there exists a natural type of homomorphism that characterizes the class, but only as a *sufficient condition* for $\mathcal{K}$-containment of CQs. In the search for classes similar to $\mathbf{C}_{\text{hom}}$, we then provide the *largest* class of semirings for which each of these conditions is *necessary* for $\mathcal{K}$-containment, resulting in analogues of Thm. 3.3 for different classes of semirings and different types of homomorphisms.

After $\mathbf{S}_{\text{hcov}}$, $\mathbf{S}_{\text{in}}$, and $\mathbf{S}$ we look at one more class, that we denote by $\mathbf{S}_{\text{sur}}$. This class lies "between" $\mathbf{S}_{\text{hcov}}$ and $\mathbf{S}$, in the sense that it can be obtained from $\mathbf{S}_{\text{hcov}}$ by a partial, instead of complete, relaxation of the $\otimes$-idempotence axiom. The class $\mathbf{S}_{\text{sur}}$ is interesting in its own right, since it can be characterized by the well studied notion of *surjective* homomorphism ([6, 17]) as yielding a sufficient condition for CQ $\mathcal{N}$-containment. In the same fashion, we identify the largest class of semirings for which this condition is also necessary.

Notice that, up to this point, we have only considered solving the $\mathcal{K}$-containment problem by means of finding different types of homomorphism between CQs. Thus, it is natural to ask whether there exists a different approach for solving this problem. We address this question at the end of this section, and show that there exists a large class of semirings which possesses a *small model property*: if a CQ Q_1 is not $\mathcal{K}$-contained in a CQ Q_2, then this is witnessed by a small enough $\mathcal{K}$-instance.

4.1 Containment by homomorphic covering

We begin with the class $\mathbf{S}_{\text{hcov}}$ of semirings satisfying the $\otimes$-idempotence axiom. For these semirings, we exploit the notion of homomorphic covering: given CQs Q_1 and Q_2, we say that Q_2 *homomorphically covers* Q_1, and write $Q_2 \rightrightarrows Q_1$, if for every atom $R(\mathbf{u}, \mathbf{v})$ in Q_1 there exists a homomorphism h from Q_2 to Q_1 with $R(\mathbf{u}, \mathbf{v})$ in the image of h.

This type of homomorphism arose in the context of query optimization as a necessary condition for $\mathcal{N}$-containment of CQs over bag semantics $\mathcal{N}$ [6]. It was also noted that existence of a homomorphic covering is not sufficient to guarantee $\mathcal{N}$-containment. Homomorphic coverings were also used in [14] to show that $Q_2 \rightrightarrows Q_1$ is both necessary and sufficient for $Q_1 \subseteq_{\texttt{Lin}[X]} Q_2$, where $\texttt{Lin}[X]$ is the lineage semiring [9].

Next we establish axiomatic bounds for semirings to have homomorphic covering as both a sufficient and necessary condition for $\mathcal{K}$-containment of CQs. Our next result shows that the class $\mathbf{S}_{\text{hcov}}$ captures precisely all semirings for which $Q_2 \rightrightarrows Q_1$ is a sufficient condition.

PROPOSITION 4.1 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{S}_{\text{hcov}}$;*
- *$Q_2 \rightrightarrows Q_1$ implies $Q_1 \subseteq_{\mathcal{K}} Q_2$, for all CQs Q_1, Q_2.*

Of course, a sufficient condition itself does not guarantee the decidability of the $\mathcal{K}$-containment problem; one needs such a condition to be necessary as well. Since one can easily find semirings in $\mathbf{S}_{\text{hcov}}$ for which the existence of a homomorphic covering is not a necessary condition (for example, any semiring in $\mathbf{C}_{\text{hom}}$), our only hope is to describe the *largest* class for which a homomorphic covering is necessary for $\mathcal{K}$-containment of CQs. Denote by $\mathbf{N}_{\text{hcov}}$ the class of semirings $\mathcal{K}$ such that for every $n, k \geq 1$ it holds that

$$x_1 \times \ldots \times x_n \times y \not\preceq_{\mathcal{K}} (x_1 + \ldots + x_n)^k$$

(again, assuming all variables to be universally quantified).

PROPOSITION 4.2 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{N}_{\text{hcov}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ implies $Q_2 \rightrightarrows Q_1$, for all CQs Q_1, Q_2.*

Therefore, bag semantics $\mathcal{N}$ is in $\mathbf{N}_{\text{hcov}}$, but not in $\mathbf{S}_{\text{hcov}}$. However, $\mathtt{Lin}[X]$ is in both, and we have the following result for the class $\mathbf{C}_{\text{hcov}} = \mathbf{S}_{\text{hcov}} \cap \mathbf{N}_{\text{hcov}}$[3] of all semirings which behave the same as $\mathtt{Lin}[X]$ w.r.t. $\mathcal{K}$-containment of CQs.

THEOREM 4.3 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{hcov}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ iff $Q_2 \rightrightarrows Q_1$, for all CQs Q_1 and Q_2.*

We also know that checking for homomorphic covering between CQs is an NP-complete problem [14]. This gives us the following result.

COROLLARY 4.4 *If $\mathcal{K} \in \mathbf{C}_{\text{hcov}}$ then* CQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete.*

4.2 Containment by injective homomorphism

In this section we consider the class $\mathbf{S}_{\text{in}}$ of semirings which satisfy the $\mathbb{1}$-annihilation axiom. This class was considered implicitly in previous studies of containment on $\mathcal{K}$-relations [15, 14, 17], and has notable applications. In the context of the Semantic Web it was shown in [4] that $\mathbf{S}_{\text{in}}$ is the class of all semirings which can be safely used as annotation domains for RDF data while respecting the inference system of RDFS. An extension of the SPARQL query language for querying annotated RDF data then followed in [22], entailing a need to solve optimization problems for this class of semirings. As an example of a semiring which is in $\mathbf{S}_{\text{in}}$, but not in $\mathbf{C}_{\text{hom}}$, we give the *tropical* semiring $\mathcal{T}^+ = \langle \mathbb{N}_0 \cup \{\infty\}, \min, +, \infty, 0\rangle$ (with its natural order).

To study the class $\mathbf{S}_{\text{in}}$, we introduce the notion of injective homomorphism: given CQs $Q_1 = \exists \mathbf{v}_1\ \phi_1(\mathbf{u}_1, \mathbf{v}_1)$ and $Q_2 = \exists \mathbf{v}_2\ \phi_2(\mathbf{u}_2, \mathbf{v}_2)$, a homomorphism h from Q_2 to Q_1 is *injective* (or *one-to-one*) if h is injective on atoms, i.e. the multiset of atoms $h(\phi_2(\mathbf{u}_2, \mathbf{v}_2))$ is contained in the multiset of atoms $\phi_1(\mathbf{u}_1, \mathbf{v}_1)$. We write $Q_2 \hookrightarrow Q_1$ iff there exists an injective homomorphism from Q_2 to Q_1. Similar to the case of $\mathbf{S}_{\text{hcov}}$, the following proposition shows that the class $\mathbf{S}_{\text{in}}$ is precisely the class of semirings for which the existence of an injective homomorphism is a sufficient condition for $\mathcal{K}$-containment of CQs.

[3]Since $\otimes$-idempotence defines $\mathbf{S}_{\text{hcov}}$, the exponent k may be omitted from the necessary condition of $\mathbf{C}_{\text{hcov}}$.

PROPOSITION 4.5 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{S}_{\text{in}}$;*
- *$Q_2 \hookrightarrow Q_1$ implies $Q_1 \subseteq_{\mathcal{K}} Q_2$, for all CQs Q_1, Q_2.*

Unfortunately, as shown in the following example, $Q_2 \hookrightarrow Q_1$ is just a sufficient, but not always necessary condition for CQ $\mathcal{K}$-containment for a semiring $\mathcal{K}$ from $\mathbf{S}_{\text{in}} \backslash \mathbf{C}_{\text{hom}}$.

EXAMPLE 4.6 Consider the conjunctive queries

$$Q_1 = \exists u, v, w\ R(u, v), R(u, w), \quad Q_2 = \exists u, v\ R(u, v), R(u, v).$$

We will see in Sec. 4.6 that Q_1 is $\mathcal{T}^+$-contained in Q_2. However, there is no injective homomorphism from Q_2 to Q_1.

To identify the largest class for which the existence of an injective homomorphism is a necessary condition for CQ $\mathcal{K}$-containment, we need the following definition.

DEFINITION 4.7 *A polynomial P from $\mathbb{N}[X]$ is* CQ-admissible *iff there exists a CQ Q, an $\mathcal{N}[X]$-instance I each tuple of which is annotated with either a unique variable from X or 0, and a tuple $\mathbf{t}$, such that $Q^I(\mathbf{t}) = \mathsf{P}$.*

Essentially, a polynomial is CQ-admissible if it is possible to obtain it by a CQ on an *abstractly tagged* instance ([15]).

We write $\mathbb{N}^{cq}[X]$ for the set of all CQ-admissible polynomials with variables X. We will use this notion intensively in the rest of this paper; for now, we have opted to give a non-constructive definition, but we will give an algebraic characterization of $\mathbb{N}^{cq}[X]$ in Sec. 4.5. Next we exploit the connection between CQ $\mathcal{K}$-containment and comparison of polynomials from $\mathbb{N}^{cq}[X]$ to define precisely the class of semirings for which an injective homomorphism is a corresponding necessary condition.

Denote by $\mathbf{N}_{\text{in}}$ the class of semirings $\mathcal{K}$ for which for every polynomial P from $\mathbb{N}^{cq}[X]$ and any set of variables $x_1, \ldots, x_n$, the inequality

$$x_1 \times \ldots \times x_n \preceq_{\mathcal{K}} \mathsf{P}$$

implies that there exists a subset $x_{i_1}, \ldots, x_{i_m}$ of the variables $x_1, \ldots, x_n$ such that P contains the monomial $x_{i_1} \times \ldots \times x_{i_m}$.

PROPOSITION 4.8 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{N}_{\text{in}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ implies $Q_2 \hookrightarrow Q_1$, for all CQs Q_1, Q_2.*

Thus, Prop. 4.5 and 4.8 give us decidability of CQ $\mathcal{K}$-CONTAINMENT for all semirings $\mathcal{K}$ from $\mathbf{C}_{\text{in}} = \mathbf{S}_{\text{in}} \cap \mathbf{N}_{\text{in}}$. Moreover, by showing that deciding the existence of an injective homomorphism between queries is NP-complete, we can say the same about $\mathcal{K}$-containment of CQs for any $\mathcal{K} \in \mathbf{C}_{\text{in}}$.

THEOREM 4.9
(1) The following are equivalent:

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{in}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ iff $Q_2 \hookrightarrow Q_1$, for all CQs Q_1 and Q_2.*

(2) If $\mathcal{K} \in \mathbf{C}_{\text{in}}$ then CQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete.*

Having this result, we however note that there are interesting semirings (including the tropical semiring $\mathcal{T}^+$), which lie in $\mathbf{S}_{\text{in}}$, but neither in $\mathbf{C}_{\text{hom}}$ nor in $\mathbf{C}_{\text{in}}$. In Sec. 4.6 we will see how to obtain decidability for some semirings in $\mathbf{S}_{\text{in}}$, but at the cost of higher complexity.

4.3 Containment by bijective homomorphism

We continue with the class obtained from $\mathbf{C}_{\text{hom}}$ by relaxing both of its axioms. This is just the class of all (positive) semirings $\mathbf{S}$. For this class we use the notion of bijective homomorphism. Given CQs $Q = \exists \mathbf{v}_1\ \phi_1(\mathbf{u}_1, \mathbf{v}_1)$ and $Q_2 = \exists \mathbf{v}_2\ \phi_2(\mathbf{u}_2, \mathbf{v}_2)$, we say that a homomorphism h from Q_2 to Q_1 is *bijective* (or *exact*) if it is a bijection on atoms, i.e. the multiset of atoms $h(\phi_2(\mathbf{u}_2, \mathbf{v}_2))$ is the same as the multiset of atoms $\phi_1(\mathbf{u}_1, \mathbf{v}_1)$. We write $Q_2 \hookrightarrow\!\!\!\!\rightarrow Q_1$ if there exists a bijective homomorphism from Q_2 to Q_1.

As shown in [14], the condition $Q_2 \hookrightarrow\!\!\!\!\rightarrow Q_1$ is both sufficient and necessary for $\mathcal{N}[X]$-containment of CQs over the provenance polynomials semiring $\mathcal{N}[X]$. Since $\mathcal{N}[X]$ is universal for $\mathbf{S}$ by Prop. 3.2, we can conclude that this condition is sufficient for CQ $\mathcal{K}$-containment for an arbitrary semiring $\mathcal{K}$, i.e. $Q_2 \hookrightarrow\!\!\!\!\rightarrow Q_1$ always implies $Q_1 \subseteq_{\mathcal{K}} Q_2$.

From [14] we also know that existence of a bijective homomorphism is necessary for $\mathcal{B}[X]$-containment of CQs, where $\mathcal{B}[X] = \langle \mathbb{B}[X], +, \times, 0, 1\rangle$ is the semiring of *boolean provenance polynomials*, i.e. polynomials over X with boolean coefficients from $\mathbb{B} = \{\mathbf{false}, \mathbf{true}\}$. This means that $\mathcal{B}[X]$ behaves the same as $\mathcal{N}[X]$ w.r.t. $\mathcal{K}$-containment of CQs. As we have seen in previous sections, this is not the case for all semirings. Also, one can easily show that the existence of a bijective homomorphism is not necessary for bag semantics $\mathcal{N}$, or even for the semiring $\mathcal{R}^+$ of non-negative reals with the usual operations and order.

Our next aim is to identify all semirings which behave as $\mathcal{N}[X]$. To do so we again exploit the notion of CQ-admissible polynomials. Denote by $\mathbf{C}_{\text{bi}}$ the class of all semirings $\mathcal{K}$ for which for every polynomial P from $\mathbb{N}^{cq}[X]$ and any set of variables $x_1, \ldots, x_n$, the inequality

$$x_1 \times \ldots \times x_n \preceq_{\mathcal{K}} \mathsf{P}$$

implies that P contains the monomial $x_1 \times \ldots \times x_n$.

THEOREM 4.10 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{bi}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ iff $Q_2 \hookrightarrow\!\!\!\!\rightarrow Q_1$, for all CQs Q_1 and Q_2.*

In particular, notice that both $\mathcal{B}[X]$ and $\mathcal{N}[X]$ belong to $\mathbf{C}_{\text{bi}}$. Thus, this theorem can be seen as a generalization of the results of [14]. There it was also shown that $\mathcal{N}[X]$-containment of CQs is an NP-complete problem. We can now extend this result to the entire class $\mathbf{C}_{\text{bi}}$.

COROLLARY 4.11 *If $\mathcal{K} \in \mathbf{C}_{\text{bi}}$ then* CQ $\mathcal{K}$-CONTAINMENT *is NP-complete.*

This corollary completes our study of $\mathcal{K}$-containment of CQs for the classes of semirings obtained from $\mathbf{C}_{\text{hom}}$ by relaxing its axioms. Next we will look at another class, which corresponds to one more well-known type of homomorphism.

4.4 Containment by surjective homomorphism

Looking back to the bag semantics semiring $\mathcal{N}$, we know that it lies in the class $\mathbf{N}_{\text{hcov}}$ for which homomorphic covering is necessary, but neither in $\mathbf{C}_{\text{hcov}}$ nor in $\mathbf{C}_{\text{bi}}$. However, there does exist a well-known sufficient condition for $\mathcal{N}$-containment, other than just a bijective homomorphism. This condition is the existence of a surjective homomorphism ([6, 17]): given CQs $Q_1 = \exists \mathbf{v}_1\ \phi_1(\mathbf{u}_1, \mathbf{v}_1)$ and $Q_2 = \exists \mathbf{v}_2\ \phi_2(\mathbf{u}_2, \mathbf{v}_2)$ a homomorphism h from Q_2 to Q_1 is *surjective* (or *onto*) if h is a surjection on atoms, i.e. the multiset of atoms $\phi_1(\mathbf{u}_1, \mathbf{v}_1)$ is contained in the multiset of atoms $h(\phi_2(\mathbf{u}_2, \mathbf{v}_2))$. We write $Q_2 \twoheadrightarrow Q_1$ iff there exists a surjective homomorphism from Q_2 to Q_1.

It is therefore natural to ask for which semirings $Q_2 \twoheadrightarrow Q_1$ is sufficient for $\mathcal{K}$-containment of CQs, and for which this is necessary. Besides $\mathcal{N}$, this condition is sufficient for a larger class of semirings denoted *type* B *systems* [17]. From [14] it is known that $Q_2 \twoheadrightarrow Q_1$ is equivalent to $\mathtt{Why}[X]$- and $\mathtt{Trio}[X]$-containment of CQs, where $\mathtt{Why}[X]$ is a semiring capturing *why provenance* of [3], and $\mathtt{Trio}[X]$ is a semiring for the provenance model used in the Trio project [10]. However, the exact axiomatic bounds for these classes of semirings were not previously known.

As usual, we start by axiomatizing semirings which have $Q_2 \twoheadrightarrow Q_1$ as a sufficient condition. Denote by $\mathbf{S}_{\text{sur}}$ the class of semirings that satisfy the axiom:

1′. ($\otimes$-*semi-idempotence*) $x \times y \preceq_{\mathcal{K}} x \times x \times y$.

This class can be obtained by relaxing the $\otimes$-idempotence axiom of $\mathbf{S}_{\text{hcov}}$, but only partially, i.e. $\mathbf{S}_{\text{hcov}} \subset \mathbf{S}_{\text{sur}}$. Other than the semirings already mentioned as belonging to $\mathbf{S}_{\text{hcov}}$, it contains the semiring $\mathcal{T}^- = \langle \mathbb{N}_0 \cup \{-\infty\}, \max, +, -\infty, 0\rangle$ known as the *schedule* (or *max-plus*) *algebra* (with its natural order). As desired, the class $\mathbf{S}_{\text{sur}}$ corresponds to all the semirings for which the existence of a surjective homomorphism is a sufficient condition for $\mathcal{K}$-containment of CQs.

PROPOSITION 4.12 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{S}_{\text{sur}}$;*
- *$Q_2 \twoheadrightarrow Q_1$ implies $Q_1 \subseteq_{\mathcal{K}} Q_2$, for all CQs Q_1, Q_2.*

As we saw for the bag semantics semiring $\mathcal{N}$, the existence of a surjective homomorphism is not necessary for $\mathcal{N}$-containment, but homomorphic covering is. The same can be said about $\mathcal{T}^-$, but not $\mathtt{Why}[X]$ or $\mathtt{Trio}[X]$. Hence, again we need to axiomatize the class of semirings for which $Q_2 \twoheadrightarrow Q_1$ is necessary for $\mathcal{K}$-containment of CQs. For this we exploit once more the notion of CQ-admissible polynomials. Denote by $\mathbf{N}_{\text{sur}}$ the class of semirings $\mathcal{K}$ for which for every polynomial P from $\mathbb{N}^{cq}[X]$ and any set of variables $x_1, \ldots, x_n$, the inequality

$$x_1 \times \ldots \times x_n \preceq_{\mathcal{K}} \mathsf{P}$$

implies that there exist exponents $m_1, \ldots, m_n \geq 1$ such that P contains the monomial $x_1^{m_1} \times \ldots \times x_n^{m_n}$.

PROPOSITION 4.13 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{N}_{\text{sur}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ implies $Q_2 \twoheadrightarrow Q_1$, for all CQs Q_1, Q_2.*

For those semirings $\mathcal{K}$ that do belong to $\mathbf{C}_{\text{sur}} = \mathbf{S}_{\text{sur}} \cap \mathbf{N}_{\text{sur}}$ (like $\mathtt{Why}[X]$ and $\mathtt{Trio}[X]$), we have once again a decision procedure for $\mathcal{K}$-containment of CQs. This is summarized by the following theorem.

THEOREM 4.14 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{sur}}$;*
- *$Q_1 \subseteq_{\mathcal{K}} Q_2$ iff $Q_2 \twoheadrightarrow Q_1$, for all CQs Q_1 and Q_2.*

The complexity follows from the fact that checking a surjective homomorphism between CQs is NP-complete ([6]).

COROLLARY 4.15 *If $\mathcal{K} \in \mathbf{C}_{\text{sur}}$ then* CQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete.*

As mentioned in the introduction, we leave open the problem of finding decision procedures for all semirings that belong to $\mathbf{S}_{\text{sur}}$, but not to $\mathbf{N}_{\text{sur}}$, such as $\mathcal{N}$ or $\mathcal{T}^-$. In Sec. 4.6 we show that for some of these semirings, such as $\mathcal{T}^-$, the problem of $\mathcal{K}$-containment of CQs can be solved using a different approach, albeit with higher computational complexity.

We finish this section with a remark that a homomorphism is bijective iff it is both injective and surjective. Thus, we obtain that $\mathbf{C}_{\text{bi}} = \mathbf{N}_{\text{in}} \cap \mathbf{N}_{\text{sur}}$.

4.5 CQ-admissible polynomials

We defined the *evaluation* of a CQ $Q = \exists \mathbf{v}\ R_1(\mathbf{u}_1, \mathbf{v}_1), \ldots, R_n(\mathbf{u}_n, \mathbf{v}_n)$ on a $\mathcal{K}$-instance I for a tuple $\mathbf{t}$ as

$$Q^I(\mathbf{t}) = \sum_{f \in \mathcal{V}(Q,\mathbf{t})} \prod_{1 \le i \le n} R_i^I(f(\mathbf{u}_i, \mathbf{v}_i)).$$

Thus, the evaluation of a CQ on a $\mathcal{N}[X]$-instance with unique variables from the set X as annotations is a polynomial over X. In Def. 4.7 we called such polynomials *CQ-admissible*. We heavily used this notion in the definitions of the classes $\mathbf{N}_{\text{in}}$, $\mathbf{N}_{\text{bi}}$, and $\mathbf{N}_{\text{sur}}$. The goal of this section is to give a constructive algebraic characterization of the set $\mathbb{N}^{cq}[X]$ of all CQ-admissible polynomials. As we mentioned in the introduction, this notion is of independent interest: for instance, it was implicitly used in [19].

From the definition of evaluation we immediately obtain that every CQ-admissible polynomial must be homogeneous. Moreover, let Q be a CQ consisting of k atoms, and I be an $\mathcal{N}[X]$-instance with tuples annotated with variables $X = \{x_1, \ldots, x_n\}$, such that the mappings in the set $\mathcal{V}(Q, \mathbf{t})$ allow us to obtain any possible combination of images of the atoms of Q to non-zero annotated tuples of I. Then $Q^I(\mathbf{t}) = (x_1 + \ldots + x_n)^k$. Therefore, every P from $\mathbb{N}^{cq}[X]$ of degree k satisfies $\mathsf{P} \preceq_{\mathcal{N}[X]} (x_1 + \ldots + x_n)^k$. (This property allowed us to formulate the axiom for the class $\mathbf{N}_{\text{hcov}}$ in Sec. 4.1 without reference to CQ-admissible polynomials.) Hence polynomials such as $2x$ and $x^2 + y$ are not in $\mathbb{N}^{cq}[X]$.

The polynomials x^2, $2xy$ and $x + y$ satisfy the requirements above, and it is not difficult to construct CQs which admit them. Unfortunately, these are not the only requirements: the polynomial $x^2 + xy + y^2$ satisfies them, but can be proved not to be in $\mathbb{N}^{cq}[X]$. In order to present the precise characterization, we need an auxiliary notion: for a set of variables X an *ordered monomial of degree n* (or *o-monomial*) is a string from X^n. For an o-monomial $\vec{\mathsf{M}}$ we denote by $\vec{\mathsf{M}}[i]$ the variable appearing in its i-th position.

PROPOSITION 4.16 *A polynomial P is in $\mathbb{N}^{cq}[X]$ iff it can be represented in a form*

$$\vec{\mathsf{P}} = \sum_{1 \le \ell \le m} \vec{\mathsf{M}}_\ell, \text{ such that}$$

1. $\vec{\mathsf{M}}_\ell$, $1 \le \ell \le m$, are pairwise distinct o-monomials over X of the same degree n (here concatenation in $\vec{\mathsf{M}}_\ell$ as a string is interpreted as product in P), and

2. if for an o-monomial $\vec{\mathsf{M}}$ of degree n, and for each i, j with $1 \le i < j \le n$, the representation $\vec{\mathsf{P}}$ contains o-monomials (each of degree n) $\vec{\mathsf{M}}_1, \ldots, \vec{\mathsf{M}}_{2k+1}$, $k \ge 0$, such that

- *$\vec{\mathsf{M}}_1[i] = \vec{\mathsf{M}}[i]$, $\vec{\mathsf{M}}_{2k+1}[j] = \vec{\mathsf{M}}[j]$ and*
- *$\vec{\mathsf{M}}_{2\ell-1}[j] = \vec{\mathsf{M}}_{2\ell}[j]$, $\vec{\mathsf{M}}_{2\ell}[i] = \vec{\mathsf{M}}_{2\ell+1}[i]$ for all $1 \le \ell \le k$,*

then $\vec{\mathsf{M}}$ is contained in $\vec{\mathsf{P}}$.

4.6 Containment via small models

Up to now we have studied how to decide $\mathcal{K}$-containment of CQs by analyzing their structure, resulting in several classes of semirings for which the existence of a homomorphism of a corresponding type between the CQs is equivalent to their $\mathcal{K}$-containment. It is natural to ask whether the problem of decidability of CQ $\mathcal{K}$-CONTAINMENT can be solved by different techniques for some semirings which are not in any of these classes. Indeed, several other approaches have appeared in the literature. In [14] a PSPACE algorithm was suggested for checking $\mathcal{N}[X]$-containment of UCQs, based on the fact that if a UCQ Q_1 is not $\mathcal{N}[X]$-contained in a UCQ Q_2, then there exists a witnessing $\mathcal{N}[X]$-instance, with its size bounded by the size of Q_1 and Q_2. Another approach is to cast the problem of decidability of $\mathcal{K}$-containment as the problem of checking the corresponding order $\preceq_{\mathcal{K}}$ on polynomials, as done in [17] to show undecidability of UCQ $\mathcal{N}$-CONTAINMENT over bag semantics.

The main result of this section is that by combining these ideas one can obtain new decidability results for $\mathcal{K}$-containment of CQs over different semirings $\mathcal{K}$. In contrast with the rest of the paper, we identify several individual semirings for which our approach works, but leave a comprehensive description of such semirings for future research. In order to describe our algorithm we introduce some terminology.

A *CQ with inequalities* is a CQ with a set of inequalities $\neq$ on its existential variables. It is *complete* (a *CCQ*) if each pair of distinct variables is bounded by an inequality. A *complete description*[4] $\wr Q \wr$ of a CQ Q with existential variables $\mathbf{v}$ is the multiset of CCQs such that for every partition π of $\mathbf{v}$ it contains a CCQ obtained from Q by identifying all the variables from each equivalence class induced by π, and attaching an inequality for every pair of variables that remain different.

EXAMPLE 4.6 (CONTINUED) For CQ $Q_1 = \exists u, v, w\ \ R(u, v), R(u, w)$ we have $\wr Q_1 \wr = \{Q_{11}, Q_{12}, Q_{13}, Q_{14}, Q_{15}\}$, where

$$\begin{array}{lcl} Q_{11} & = & \exists u, v, w\ R(u, v), R(u, w), u \neq v, u \neq w, v \neq w; \\ Q_{12} & = & \exists u, v\ R(u, v), R(u, v), u \neq v; \\ Q_{13} & = & \exists u, v\ R(u, v), R(u, u), u \neq v; \\ Q_{14} & = & \exists u, w\ R(u, u), R(u, w), u \neq w; \text{ and} \\ Q_{15} & = & \exists u\ R(u, u), R(u, u). \end{array}$$

We will heavily use complete descriptions in Sec. 5, but for now we are interested in CCQs from $\wr Q \wr$ as a way of describing all possible images of a mapping from Q to a $\mathcal{K}$-instance. Formally, given a set of variables X, a *canonical instance* ([15]) $[\![Q]\!]$ of a CQ (or CCQ) Q is an $\mathcal{N}[X]$-instance with the same schema as Q and with the set of variables of Q as its domain, such that for every $\mathcal{N}[X]$-relation $R^{[\![Q]\!]}$ and for every tuple $\mathbf{u}, \mathbf{v}$ it holds that $R^{[\![Q]\!]}(\mathbf{u}, \mathbf{v}) = x_1 + \ldots + x_n$, where $n \ge 0$ is the number of atoms in Q of the form $R(\mathbf{u}, \mathbf{v})$, and $x_1, \ldots, x_n$ are unique (over all $[\![Q]\!]$) variables from X.

EXAMPLE 4.6 (CONTINUED) For Q_{11} and Q_{12} we have

$$R^{[\![Q_{11}]\!]}(u, v) = x_1, \quad R^{[\![Q_{11}]\!]}(u, w) = x_2,$$
$$R^{[\![Q_{12}]\!]}(u, v) = x_1 + x_2,$$

and all other tuples in $[\![Q_{11}]\!]$ and $[\![Q_{12}]\!]$ are annotated by 0.

Denote by $\mathbf{S}^1$ the set of $\oplus$-*idempotent* semirings, i.e. the semirings where $x =_{\mathcal{K}} x + x$ holds (this notation will be explained and generalized in Sec. 5). We have the following.

[4] This is similar to the one from [11] and linearization of [8].

THEOREM 4.17 *Given a semiring $\mathcal{K}$ from $\mathbf{S}^1$ and CQs Q_1 and Q_2, we have that $Q_1 \subseteq_{\mathcal{K}} Q_2$ iff $Q_1^{[\![Q]\!]}(\mathbf{t}) \preceq_{\mathcal{K}} Q_2^{[\![Q]\!]}(\mathbf{t})$ for every CCQ $Q \in \wr Q_1 \wr$ and every tuple $\mathbf{t}$ of variables of Q_1.*

This theorem shows that for $\mathcal{K} \in \mathbf{S}^1$, CQ $\mathcal{K}$-CONTAINMENT can be reduced to a small number of problems of checking the order $\preceq_{\mathcal{K}}$ between CQ-admissible polynomials.

COROLLARY 4.18 *If $\mathcal{K} \in \mathbf{S}^1$ and it is decidable to check if $\mathsf{P}_1 \preceq_{\mathcal{K}} \mathsf{P}_2$ for any pair of polynomials $\mathsf{P}_1, \mathsf{P}_2$ from $\mathbb{N}^{cq}[X]$, then* CQ $\mathcal{K}$-CONTAINMENT *is decidable.*

We do not investigate the decidability of $\mathsf{P}_1 \preceq_{\mathcal{K}} \mathsf{P}_2$ for the entire class $\mathbf{S}^1$, but do so for some of its most important members that do not have any corresponding type of homomorphism – the tropical semiring $\mathcal{T}^+$ and the schedule algebra $\mathcal{T}^-$. Since we can decide in PSPACE whether $\mathsf{P}_1 \preceq_{\mathcal{T}^+} \mathsf{P}_2$ and $\mathsf{P}_1 \preceq_{\mathcal{T}^-} \mathsf{P}_2$, we have the following result.

PROPOSITION 4.19 CQ $\mathcal{T}^+$- *and* $\mathcal{T}^-$-CONTAINMENT *are in* PSPACE.

We illustrate this proposition by an extension of Ex. 4.6.

EXAMPLE 4.6 (CONTINUED) We know that $\wr Q_1 \wr = \{Q_{11}, Q_{12}, Q_{13}, Q_{14}, Q_{15}\}$. Hence

$$Q_1^{[\![Q_{11}]\!]}() = x_1^2 + 2x_1x_2 + x_2^2, \text{ and } Q_2^{[\![Q_{11}]\!]}() = x_1^2 + x_2^2.$$

It is straightforward to see that

$$x_1^2 + 2x_1x_2 + x_2^2 =_{\mathcal{T}^+} x_1^2 + x_2^2.$$

The same can be shown for the $\mathcal{T}^+$-instances $[\![Q_{12}]\!]$, $[\![Q_{13}]\!]$, $[\![Q_{14}]\!]$, and $[\![Q_{15}]\!]$. By Thm. 4.17 we have that $Q_1 \subseteq_{\mathcal{T}^+} Q_2$.

5. $\mathcal{K}$-CONTAINMENT OF UCQS

In this section we look at $\mathcal{K}$-containment of UCQs, which generalizes the problem for CQs considered in Sec. 4. We examine both existing algorithms for deciding containment of UCQs and new ones developed here. All of these exploit the procedures used for CQs. Similarly to before, we identify classes of semirings corresponding to these algorithms, refining the classes from the previous sections.

Started for set semantics in [20], the study of $\mathcal{K}$-containment for UCQs continued for some particular semirings [17, 14], as well as classes of semirings such as type A systems [17] and distributive bilattices [13]. Generally, semirings from $\mathbf{S}^1$ (i.e. $\oplus$-idempotent semirings) were identified as *well behaved* w.r.t. containment of UCQs. Prior works show NP-completeness of checking $\mathcal{K}$-containment for some particular semirings $\mathcal{K}$ from $\mathbf{S}^1$, essentially relying on the following fact.

PROPOSITION 5.1 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{S}^1$;*
- *if for UCQs $\mathbf{Q}_1$, $\mathbf{Q}_2$ it holds that for each $Q_1 \in \mathbf{Q}_1$ there exists $Q_2 \in \mathbf{Q}_2$ with $Q_1 \subseteq_{\mathcal{K}} Q_2$, then $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$.*

This proposition says that $\mathcal{K}$ is $\oplus$-idempotent iff for $\mathcal{K}$-containment of UCQs it is sufficient to check CQs in $\mathbf{Q}_1$ locally, one at a time. Hence, if we have a sufficient condition for containment of CQs for some class inside $\mathbf{S}^1$, then we have one also for UCQs. Since $\mathbb{1}$-annihilation implies $\oplus$-idempotence, we have that $\mathbf{S}_{\text{in}} \subseteq \mathbf{S}^1$. Thus, next we study UCQ $\mathcal{K}$-containment for semirings in $\mathbf{S}_{\text{in}}$ and its subclasses.

5.1 Containment by homomorphism and injective homomorphism

We start with the classes $\mathbf{C}_{\text{hom}}$ and $\mathbf{C}_{\text{in}}$ from Sec. 4 and investigate for which semirings from these classes Prop. 5.1 can be used in decision procedures for UCQ $\mathcal{K}$-containment.

Recall that we write $Q_2 \to Q_1$ (and $Q_2 \hookrightarrow Q_1$) if there is a homomorphism (resp., injective homomorphism) from a CQ Q_2 to a CQ Q_1. We generalize these notions to unions as follows: given UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$ we write $\mathbf{Q}_2 \to \mathbf{Q}_1$ (and $\mathbf{Q}_2 \hookrightarrow \mathbf{Q}_1$) iff for each $Q_1 \in \mathbf{Q}_1$ there exists $Q_2 \in \mathbf{Q}_2$ such that $Q_2 \to Q_1$ (resp., $Q_2 \hookrightarrow Q_1$).

For the class $\mathbf{C}_{\text{hom}}$, we know that the existence of a homomorphism is a sufficient condition for containment of CQs, and thus by Prop. 5.1 we can conclude that $\mathbf{Q}_2 \to \mathbf{Q}_1$ implies $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$. It turns out, that for this class of semirings the "only if" direction of the second item of Prop. 5.1 holds as well, so this condition is also necessary. Hence, we can extend Prop. 5.1, and present the following theorem.

THEOREM 5.2 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{hom}}$;*
- *$\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ iff $\mathbf{Q}_2 \to \mathbf{Q}_1$, for all UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$.*

Thus, we have a decision procedure for $\mathcal{K}$-containment of UCQs, for all semirings $\mathcal{K}$ in $\mathbf{C}_{\text{hom}}$. The NP-completeness of this procedure was first obtained for the set semantics semiring $\mathcal{B}$ in [20], which we know to be in this class.

COROLLARY 5.3 *If $\mathcal{K} \in \mathbf{C}_{\text{hom}}$ then deciding $\mathcal{K}$-containment for UCQs is NP-complete.*

Unfortunately, once we move away from $\mathbf{C}_{\text{hom}}$, we cannot guarantee a similar result, since the "only if" direction of the second item of Prop. 5.1 does not hold for an arbitrary semiring in $\mathbf{S}^1$. This can be seen from the following example.

EXAMPLE 5.4 Consider again the tropical semiring $\mathcal{T}^+ = \langle \mathbb{N}_0 \cup \{\infty\}, \min, +, \infty, 0\rangle$, and UCQs $\mathbf{Q}_1 = \{Q_{11}\}$ and $\mathbf{Q}_2 = \{Q_{21}, Q_{22}\}$ over a schema with unary relations R, S, where

$$Q_{11} = \exists v\ R(v), S(v);$$
$$Q_{21} = \exists v\ R(v), R(v); \quad Q_{22} = \exists v\ S(v), S(v).$$

It is possible to show that $\mathbf{Q}_1 \subseteq_{\mathcal{T}^+} \mathbf{Q}_2$, but neither of the containments $Q_{11} \subseteq_{\mathcal{T}^+} Q_{21}$ nor $Q_{11} \subseteq_{\mathcal{T}^+} Q_{22}$ holds.

From Prop. 5.1 and 4.5 we know that $\mathbf{Q}_2 \hookrightarrow \mathbf{Q}_1$ is a sufficient condition for $\mathcal{K}$-containment of UCQs, for all semirings in $\mathbf{S}_{\text{in}}$. However, the example above shows that this condition may not be necessary for all semirings from $\mathbf{S}_{\text{in}}$. Next we identify the semirings for which it is. Denote by $\mathbf{N}_{\text{in}}^1$ the class of semirings $\mathcal{K}$ for which for every polynomial $\mathsf{P} \in \mathbb{N}[X]$ without a constant term and any set of variables $x_1, \ldots, x_n$ from X, the inequality

$$x_1 \times \ldots \times x_n \preceq_{\mathcal{K}} \mathsf{P}$$

implies that there exists a subset $x_{i_1}, \ldots, x_{i_m}$ of variables $x_1, \ldots, x_n$ such that P contains the monomial $x_{i_1} \times \ldots \times x_{i_m}$. Note that this is the same condition as the one for $\mathbf{N}_{\text{in}}$, but it is required to hold not only for CQ-admissible, but for all polynomials without a constant term. This definition is justified by the fact that any such polynomial can be obtained on an $\mathcal{N}[X]$-instance with tuples annotated with unique variables. For $\mathbf{N}_{\text{in}}^1$ we have the desired proposition.

PROPOSITION 5.5 *The following two statements are equivalent:*

- *semiring* $\mathcal{K}$ *belongs to* $\mathbf{N}^1_{\text{in}}$*;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *implies* $\mathbf{Q}_2 \hookrightarrow \mathbf{Q}_1$*, for all UCQs* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$*.*

For the class $\mathbf{C}^1_{\text{in}} = \mathbf{S}_{\text{in}} \cap \mathbf{N}^1_{\text{in}}$ we have the following result.

THEOREM 5.6
(1) The following are equivalent:
- *semiring* $\mathcal{K}$ *belongs to* $\mathbf{C}^1_{\text{in}}$*;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *iff* $\mathbf{Q}_2 \hookrightarrow \mathbf{Q}_1$*, for all UCQs* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$*.*

(2) If $\mathcal{K} \in \mathbf{C}^1_{\text{in}}$ *then* UCQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete.*

In the following sections we will see that the classes $\mathbf{C}_{\text{bi}}$, $\mathbf{C}_{\text{sur}}$, and $\mathbf{C}_{\text{hcov}}$, for which we have decision procedures for containment of CQs, do not lie inside $\mathbf{S}^1$ but have non-empty intersections with it. For these intersections, Prop. 5.1 gives a sufficient condition for $\mathcal{K}$-containment of UCQs. However, new techniques will have to be developed to handle semirings outside these intersections.

5.2 Containment by bijective homomorphism

In Sec. 4.3 we argued that the existence of a bijective homomorphism is a sufficient condition for $\mathcal{K}$-containment of two CQs, for any semiring $\mathcal{K}$. Thus, Prop. 5.1 gives us a sufficient condition for $\mathcal{K}$-containment of *UCQs* for any semiring $\mathcal{K}$ in $\mathbf{S}^1$: to affirm that $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ it suffices for every CQ from $\mathbf{Q}_1$ to find a bijective homomorphism to it from some CQ of $\mathbf{Q}_2$. In [14] it was shown that this condition is also necessary for the semiring of boolean provenance polynomials $\mathcal{B}[X]$ (which is universal for $\mathbf{S}^1$). At the end of this section we will find a subclass $\mathbf{C}^1_{\text{bi}}$ of $\mathbf{S}^1$ consisting of all semirings which behave the same as $\mathcal{B}[X]$ w.r.t. UCQs.

However, not much is known about decision procedures and complexity for containment of UCQs for semirings outside $\mathbf{S}^1$, other than the classic result that this problem is undecidable under bag semantics $\mathcal{N}$ [17]. As was observed in [14], if for UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$ it holds that for every $Q_1 \in \mathbf{Q}_1$ there exists a *unique* $Q_2 \in \mathbf{Q}_2$ such that $Q_1 \subseteq_{\mathcal{K}} Q_2$, then $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ for any semiring $\mathcal{K}$, i.e. this condition is sufficient for containment of UCQs for any semantics. However, it was also shown that this condition is not necessary for provenance polynomials $\mathcal{N}[X]$, and therefore for any semiring. This is shown in the following example.

EXAMPLE 5.7 Consider a schema with a binary relation R, and UCQs $\mathbf{Q}_1 = \{Q_{11}, Q_{12}\}$, $\mathbf{Q}_2 = \{Q_{21}, Q_{22}\}$, where

$$Q_{11} = \exists u, v\ R(u,v), R(u,u); \quad Q_{12} = \exists u, v\ R(u,v), R(v,v);$$
$$Q_{21} = \exists u, v, w\ R(u,v), R(w,w); \quad Q_{22} = \exists u\ R(u,u), R(u,u).$$

We cannot find for every CQ Q in $\mathbf{Q}_1$ a unique CQ in $\mathbf{Q}_2$ containing Q. Later we will demonstrate that $\mathbf{Q}_1 \subseteq_{\mathcal{N}[X]} \mathbf{Q}_2$.

A PSPACE algorithm was suggested in [14] for deciding $\mathcal{N}[X]$-containment of UCQs. This algorithm involves *guessing* a small enough $\mathcal{N}[X]$-instance as a counterexample, and then posing the queries over it. However, the possibility of solving it using some type of homomorphism was left open.

In what follows, we use the techniques developed in Sec. 4.3 to devise a syntactic criterion to decide $\mathcal{N}[X]$-containment of unions of conjunctive queries. We also give a procedure for checking this criterion which allows us to improve the PSPACE upper bound given in [14]. Afterwards we shall see which semirings behave just as $\mathcal{N}[X]$ w.r.t. containment of UCQs (clearly, all such semirings lie in $\mathbf{C}_{\text{bi}}$). But first, in order to study these problems, we revisit the notion of complete description from Sec. 4.6 and extend it to UCQs.

A *complete description* $\wr\mathbf{Q}\wr$ of a UCQ $\mathbf{Q}$ is a union of complete descriptions of its elements.[5] The semantics of a CCQ Q is the same as that of a CQ (given in the preliminaries), except that $\mathcal{V}(Q, \mathbf{t})$ for any $\mathbf{t}$ contains only mappings preserving the inequalities. Similarly, homomorphisms of all types considered in Sec. 4 between CCQs should preserve the inequalities. Hence, for any UCQ $\mathbf{Q}$ and semiring $\mathcal{K}$ we have that $\mathbf{Q} \equiv_{\mathcal{K}} \wr\mathbf{Q}\wr$, i.e. complete descriptions are just explicit representations of UCQs.

All endomorphisms of CCQs are automorphisms. This key property allows us to use CCQs in the condition for $\mathcal{N}[X]$-containment of UCQs. Further, for CCQs Q_1 and Q_2, this property implies that $Q_2 \twoheadrightarrow Q_1$ iff Q_1 and Q_2 are *isomorphic*, i.e. coincide up to renaming of existential variables. Given a UCQ $\mathbf{Q}$ and a CQ Q we write $\mathbf{Q}[Q^{\simeq}]$ for the number of CQs in $\mathbf{Q}$ that are isomorphic to Q.

DEFINITION 5.8 *Given UCQs* $\mathbf{Q}_1$*,* $\mathbf{Q}_2$*, we write* $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$ *iff for each CCQ* Q *it holds that*

$$\wr\mathbf{Q}_1\wr[Q^{\simeq}] \leq \wr\mathbf{Q}_2\wr[Q^{\simeq}].$$

If $\mathbf{Q}_1 = \{Q_1\}$ and $\mathbf{Q}_2 = \{Q_2\}$ consist of single CQs then $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$ is equivalent to $Q_2 \twoheadrightarrow Q_1$, so the definition above extends bijective homomorphisms. We are ready to state the decision procedure for $\mathcal{N}[X]$-containment of UCQs.

PROPOSITION 5.9 *For any UCQs* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$ *it holds that* $\mathbf{Q}_1 \subseteq_{\mathcal{N}[X]} \mathbf{Q}_2$ *iff* $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$.

Next we continue Ex. 5.7 and show that $\mathbf{Q}_1 \subseteq_{\mathcal{N}[X]} \mathbf{Q}_2$.

EXAMPLE 5.7 (CONTINUED) Having that $\wr\mathbf{Q}_1\wr = \{Q'_{11}, Q'_{12}, Q'_{22}, Q'_{22}\}$ and $\wr\mathbf{Q}_2\wr = \{Q'_{21}, Q'_{11}, Q'_{12}, Q'_{22}, Q'_{22}\}$, where

$$\begin{aligned}
Q'_{11} &= \exists u, v\ R(u,v), R(u,u), u \neq v;\\
Q'_{12} &= \exists u, v\ R(u,v), R(v,v), u \neq v;\\
Q'_{21} &= \exists u, v, w\ R(u,v), R(w,w), u \neq v, v \neq w, w \neq u;\\
Q'_{22} &= \exists u\ R(u,u), R(u,u),
\end{aligned}$$

we conclude that $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$, and hence $\mathbf{Q}_1 \subseteq_{\mathcal{N}[X]} \mathbf{Q}_2$.

Therefore, the condition $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$ on UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$ is equivalent to their $\mathcal{N}[X]$-containment. Using Prop. 3.2, we conclude that it is sufficient for $\mathcal{K}$-containment over any semiring $\mathcal{K}$. We leave the question of the complexity of checking this condition to the end of this section, but look now at semirings which behave the same as $\mathcal{N}[X]$, i.e. for which the condition is also necessary. Again we exploit the relationship between queries and polynomials.

Denote by $\mathbf{C}^\infty_{\text{bi}}$ the class of all semirings $\mathcal{K}$ such that for every coefficient $\ell > 0$, polynomial $\mathsf{P} \in \mathbb{N}[X]$ without a constant term and monomial M over X, the inequality

$$\ell\,\mathsf{M} \preceq_{\mathcal{K}} \mathsf{P}$$

implies that M has a coefficient at least ℓ in P.

Notice that $\mathbf{C}^\infty_{\text{bi}}$ is a subclass of $\mathbf{C}_{\text{bi}}$. Also, as desired, $\mathcal{N}[X]$ is in $\mathbf{C}^\infty_{\text{bi}}$, i.e. this class contains semirings $\mathcal{K}$ which behave the same as $\mathcal{N}[X]$ w.r.t. $\mathcal{K}$-containment of UCQs. It turns out that it contains all such semirings.

PROPOSITION 5.10 *The following are equivalent:*
- *semiring* $\mathcal{K}$ *belongs to* $\mathbf{C}^\infty_{\text{bi}}$*;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *iff* $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$*, for all UCQs* $\mathbf{Q}_1$*,* $\mathbf{Q}_2$*.*

[5] Recall here, that we assume that CCQs are multisets, i.e. this union is always disjoint.

The above proposition gives us a decision procedure for $\mathcal{K}$-containment of UCQs, for all semirings $\mathcal{K}$ in $\mathbf{C}_{\text{bi}}^{\infty}$. For these semirings the equality $kx =_{\mathcal{K}} \ell x$ holds only when $k = \ell$. Coming back to $\oplus$-idempotent semirings $\mathcal{K}$ from the class $\mathbf{S}^1$, we have that $kx =_{\mathcal{K}} \ell x$ holds for all $k, \ell \in \mathbb{N}$. What happens for those semirings that lie "in between" these classes? Such semirings satisfy $kx =_{\mathcal{K}} \ell x$ not for all, but just for some $k \neq \ell$. In what follows, we will classify them and parameterize Prop. 5.10 over this classification.

A semiring $\mathcal{K}$ has *offset* k iff for all $\ell \geq k$ it holds that $kx =_{\mathcal{K}} \ell x$. In particular, $\oplus$-idempotent semirings from $\mathbf{S}^1$ have offset 1. The following proposition says that the smallest offset of a semiring $\mathcal{K}$ identifies all its axioms $kx =_{\mathcal{K}} \ell x$.

PROPOSITION 5.11 *Suppose $\mathcal{K}$ is a (positive) semiring. If $kx =_{\mathcal{K}} \ell x$ holds for some $1 \leq k < \ell$, then $\mathcal{K}$ has offset k.*

Based on this fact, we consider the classes $\mathbf{S}^k$, $k \in \mathbb{N}$, of semirings with offset k. We have already seen one such class, $\mathbf{S}^1$. Note, that for all $k \geq 1$ we have $\mathbf{S}^k \subset \mathbf{S}^{k+1}$. Our aim is to obtain a sufficient condition for containment of UCQs for each $\mathbf{S}^k$. The following example gives an idea how to do it.

EXAMPLE 5.7 (CONTINUED) Coming back to UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$, we know that $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ for any semiring $\mathcal{K}$. However, if we take $\mathbf{Q}_1' = \mathbf{Q}_1 \cup \{Q_{22}\}$ we have that now $\wr\mathbf{Q}_1'\wr = \wr\mathbf{Q}_1\wr \cup \{Q_{22}'\}$ has not two, but three CCQs isomorphic to Q_{22}'. Since $\wr\mathbf{Q}_2\wr$ has only two of them, we have that $\wr\mathbf{Q}_2\wr \not\twoheadrightarrow_{\infty} \wr\mathbf{Q}_1'\wr$ and thereby $\mathbf{Q}_1' \not\subseteq_{\mathcal{N}[X]} \mathbf{Q}_2$. At the same time, we see that $\mathbf{Q}_1' \subseteq_{\mathcal{K}} \mathbf{Q}_2$ for any semiring $\mathcal{K}$ with offset 2. The reason is that we can dismiss the third copy of Q_{22}' in $\wr\mathbf{Q}_1'\wr$, since it is made *redundant* by the offset 2 of $\mathcal{K}$, i.e. by removing it we do not alter the result of the query over any $\mathcal{K}$-instance.

The above example illustrates the intuition that to obtain a "tight" sufficient condition for some semiring $\mathcal{K}$, one should take into account its smallest offset and, if it is greater than 1, split UCQs to their complete descriptions. Hence, the desired sufficient condition for such semirings is likely to resemble $\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$. For the sake of uniformity we extend the notion of offset and say that any semiring has offset ∞; that is, $\mathbf{S}^{\infty} = \mathbf{S}$, and write $\mathbb{N}_{\infty}$ for $\mathbb{N} \cup \{\infty\}$. Using the ideas above, we can extend the condition stated in Def. 5.8 to each $k \in \mathbb{N}_{\infty}$, so that the criterion $\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_k \wr\mathbf{Q}_1\wr$ also generalizes Prop. 5.1 for values of k other than 1. The definition is rather technical (it needs to take into account *automorphisms* of CCQs), and for space reasons is deferred to the full version. Using this criterion we state the following fact.

PROPOSITION 5.12 *For each $k \in \mathbb{N}_{\infty}$ the following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{S}^k$;*
- *$\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_k \wr\mathbf{Q}_1\wr$ implies $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$, for all UCQs $\mathbf{Q}_1$, $\mathbf{Q}_2$.*

Similarly to the homomorphism types from Sec. 4, for every $k \in \mathbb{N}$ we can axiomatize a class $\mathbf{N}_{\text{bi}}^k$ for which $\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_k \wr\mathbf{Q}_1\wr$ is necessary for containment of UCQs. All the axioms are similar to the one for $\mathbf{C}_{\text{bi}}^{\infty}$ and are omitted. The semirings from each intersection $\mathbf{C}_{\text{bi}}^k = \mathbf{S}^k \cap \mathbf{N}_{\text{bi}}^k$ have the same smallest offset k. For these classes we have $\mathbf{C}_{\text{bi}}^k \subset \mathbf{C}_{\text{bi}}$, $k \geq 1$. The following theorem extends Prop. 5.10 to the classes $\mathbf{C}_{\text{bi}}^k$ and uniformly establishes complexity bounds for all the procedures from this section.

THEOREM 5.13 *Let $k \in \mathbb{N}_{\infty}$.*
(1) The following are equivalent:

- *semiring $\mathcal{K}$ belongs to $\mathbf{C}_{\text{bi}}^k$;*
- *$\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ iff $\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_k \wr\mathbf{Q}_1\wr$, for all UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$.*

(2) If $\mathcal{K} \in \mathbf{C}_{\text{bi}}^k$ then UCQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete if $k = 1$, in Π_2^p if $2 \leq k < \infty$, and in* coNP$^{\#P}$ *if $k = \infty$.*

The complexity for $\mathbf{C}_{\text{bi}}^1$ is lower, since the counterpart of Prop. 5.1 holds and splitting UCQs to complete descriptions is redundant. This agrees with the result of [14] for $\mathcal{B}[X]$. For the case of $k = \infty$ the coNP$^{\#P}$ upper bound improves the result given there for $\mathcal{N}[X]$.[6] For the intermediate cases of $2 \leq k < \infty$, the complexity drops since the number of CCQs in the search space is bounded by k.

5.3 Containment by surjective homomorphism

In this section we look at the problem of containment of UCQs over semirings from the class $\mathbf{S}_{\text{sur}}$, for which the existence of a surjective homomorphism is sufficient for containment of CQs. We develop a syntactic condition similar to the condition $\hookrightarrow\!\!\!\!\twoheadrightarrow_{\infty}$ from Sec. 5.2, which is sufficient for UCQ $\mathcal{K}$-containment for all semirings $\mathcal{K}$ from $\mathbf{S}_{\text{sur}}$, and necessary for some of them, including the universal semirings of $\mathbf{S}_{\text{sur}}$.

The naive approach is to state such a condition by requiring for every Q_1 in $\mathbf{Q}_1$ the existence of a unique CQ Q_2 in $\mathbf{Q}_2$ such that $Q_2 \twoheadrightarrow Q_1$. However, this condition suffers from the same problem as the similar condition for the class $\mathbf{S}$: it is sufficient for $\mathcal{K}$-containment of $\mathbf{Q}_1$ in $\mathbf{Q}_2$ over every $\mathcal{K}$ from $\mathbf{S}_{\text{sur}}$, but modifying Ex. 5.7 it is possible to show that it is not necessary for any such semiring. Thus, such a condition doesn't suit our purposes, because it doesn't aid in our search for decision procedures for containment of UCQs. This leaves open the possibility of finding a stricter criterion, which is still sufficient for all semirings in $\mathbf{S}_{\text{sur}}$. Using the power of complete descriptions, we devise such a criterion.

DEFINITION 5.14 *Given two UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$ we write $\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$ iff for every CCQ Q_1 from $\wr\mathbf{Q}_1\wr$ there exists a* unique *CCQ Q_2 in $\wr\mathbf{Q}_2\wr$ such that $Q_2 \twoheadrightarrow Q_1$.*

With this condition in hand we can state the proposition.

PROPOSITION 5.15 *The following are equivalent:*

- *semiring $\mathcal{K}$ belongs to $\mathbf{S}_{\text{sur}}$;*
- *$\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$ implies $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$, for all UCQs $\mathbf{Q}_1, \mathbf{Q}_2$.*

Notice that since bag semantics $\mathcal{N}$ is in $\mathbf{S}_{\text{sur}}$, this proposition gives us a new sufficient condition for $\mathcal{N}$-containment of UCQs, which improves previous results of [6, 17].

COROLLARY 5.16 *If $\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$ then $\mathbf{Q}_1 \subseteq_{\mathcal{N}} \mathbf{Q}_2$.*

While it is possible to axiomatize the class $\mathbf{N}_{\text{sur}}^{\infty}$ of semirings for which the condition $\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$ is necessary for a UCQ $\mathbf{Q}_1$ to be $\mathcal{K}$-contained in a UCQ $\mathbf{Q}_2$, the definition is somewhat technical, and we defer the reader to the full version. By the results of Sec. 4.4, this condition is not necessary for any semiring that is not in $\mathbf{N}_{\text{sur}}$, like the bag semantics semiring ($\mathcal{N}$-containment of UCQs is in general undecidable [17]). One can also show that it is not necessary for any semiring from $\mathbf{N}_{\text{sur}}$ with finite smallest offset.

As intended, $\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$ leads to a decision procedure for UCQ $\mathcal{K}$-containment for the class $\mathbf{C}_{\text{sur}}^{\infty} = \mathbf{S}_{\text{sur}} \cap \mathbf{N}_{\text{sur}}^{\infty}$.

THEOREM 5.17 *The following two statements are equivalent:*

[6] This bound coincides with the best known bound for deciding bag-equivalence of CQs with inequalities $<$ (see [8]).

- *semiring $\mathcal{K}$ belongs to* $\mathbf{C}^{\infty}_{\text{sur}}$*;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *iff* $\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$*, for all UCQs* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$*.*

The proof exploits Hall's marriage theorem and the fact that, for CCQs Q_1 and Q_2 we have $Q_2 \twoheadrightarrow Q_1$ iff Q_2 is isomorphic to a CCQ which contains exactly the same atoms as Q_1 but with greater or equal multiplicities. Checking $\wr\mathbf{Q}_2\wr \twoheadrightarrow_{\infty} \wr\mathbf{Q}_1\wr$ can clearly be done in EXPTIME. We leave the issue of exact complexity open.

Finally, we analyze necessary conditions for the classes $\mathbf{S}^k_{\text{sur}} = \mathbf{S}_{\text{sur}} \cap \mathbf{S}^k$ of semirings having finite offsets. For semirings with minimal offsets $k \geq 2$ the straightforward extension on the base of complete descriptions does not work, and in order to find such a criterion one needs to use even more elaborate representations of UCQs. Formulating this criterion is possible, but extremely technical, and thus it appears unlikely that it would have any practical applicability.

Instead, we concentrate on the case $k = 1$, i.e. $\oplus$-idempotent semirings. Given UCQs $\mathbf{Q}_1$, $\mathbf{Q}_2$ we write $\mathbf{Q}_2 \twoheadrightarrow_1 \mathbf{Q}_1$ if for each $Q_1 \in \mathbf{Q}_1$ there exists $Q_2 \in \mathbf{Q}_2$ such that $Q_2 \twoheadrightarrow Q_1$.

It immediately follows from Prop. 5.1 and 4.12 that the condition $\mathbf{Q}_2 \twoheadrightarrow_1 \mathbf{Q}_1$ implies that $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$. Moreover, it was noted in [14] that this condition is also necessary for $\mathtt{Why}[X]$-containment of $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Next we identify all the semirings that behave just as $\mathtt{Why}[X]$.

Denote by $\mathbf{N}^1_{\text{sur}}$ the class of semirings $\mathcal{K}$ for which for every polynomial $\mathsf{P} \in \mathbb{N}[X]$ without a constant term and any set of variables $x_1, \ldots, x_n$, the inequality

$$x_1 \times \ldots \times x_n \preceq_{\mathcal{K}} \mathsf{P}$$

implies that there exist exponents $m_1, \ldots, m_n \geq 1$ such that P contains the monomial $x_1^{m_1} \times \ldots \times x_n^{m_n}$. Similarly to the class $\mathbf{N}^1_{\text{in}}$, this condition is the same as the condition for $\mathbf{N}_{\text{sur}}$, but should hold not only for CQ-admissible, but for all polynomials without constant terms. Given that this is a stronger requirement, one expects that $\mathbf{N}^1_{\text{sur}} \subset \mathbf{N}_{\text{sur}}$. This is indeed the case, since for example the semiring $\mathtt{Trio}[X]$ is not in $\mathbf{N}^1_{\text{sur}}$, but is in $\mathbf{N}_{\text{sur}}$.

The class $\mathbf{N}^1_{\text{sur}}$ corresponds precisely to the class of semirings $\mathcal{K}$ for which $\mathbf{Q}_2 \twoheadrightarrow_1 \mathbf{Q}_1$ is necessary for $\mathcal{K}$-containment of UCQs. For the intersection $\mathbf{C}^1_{\text{sur}} = \mathbf{S}_{\text{sur}} \cap \mathbf{N}^1_{\text{sur}}$ the following corollary holds. The complexity was first shown for the $\mathtt{Why}[X]$ semiring in [14].

Corollary 5.18
(1) The following are equivalent:
- *semiring $\mathcal{K}$ belongs to* $\mathbf{C}^1_{\text{sur}}$*;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *iff* $\mathbf{Q}_2 \twoheadrightarrow_1 \mathbf{Q}_1$*, for all UCQs* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$*.*

(2) If $\mathcal{K} \in \mathbf{C}^1_{\text{sur}}$ *then* UCQ $\mathcal{K}$-Containment *is* NP*-complete.*

5.4 Containment by homomorphic covering

So far we have generalized to UCQs all the types of homomorphisms from Sec. 4, except homomorphic covering. This section closes the remaining gap. We have left it to the end of the paper, since, when compared to the previous results, the case of $\mathbf{S}_{\text{hcov}}$ is rather specific. Nevertheless, we identify a sufficient condition for UCQ containment for the class $\mathbf{S}_{\text{hcov}}$ and show that for some semirings it is also necessary. It again is based on the concept of complete descriptions.

It is important that all semirings in $\mathbf{S}_{\text{hcov}}$ have offset 2.

Proposition 5.19 *The following holds:* $\mathbf{S}_{\text{hcov}} \subseteq \mathbf{S}^2$*.*

For the $\oplus$-idempotent semirings from $\mathbf{S}^1_{\text{hcov}} = \mathbf{S}^1 \cap \mathbf{S}_{\text{hcov}}$ Prop. 5.1 holds, as usual. This time, however, the condition requiring checking CQs in UCQs $\mathbf{Q}_1$ and $\mathbf{Q}_2$ only pairwise is never necessary, as shown in the following example.

Example 5.20 Consider UCQs $\mathbf{Q}_1 = \{Q_{11}\}$ and $\mathbf{Q}_2 = \{Q_{21}, Q_{22}\}$ over a schema with unary relations R, S, where

$$Q_{11} = \exists v\, R(v), S(v); \qquad Q_{21} = \exists v\, R(v); \quad Q_{22} = \exists v\, S(v).$$

It is not difficult to show that $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$, over any semiring $\mathcal{K} \in \mathbf{S}_{\text{hcov}}$. However, neither $Q_{21} \rightrightarrows Q_{11}$ nor $Q_{22} \rightrightarrows Q_{11}$.

The above example captures the intuition that both Q_{21} and Q_{22} should be used *at the same time* to produce a covering for Q_{11}. Next we define a condition, that generalizes this intuition. Given UCQs $\mathbf{Q}_1$, $\mathbf{Q}_2$, we write $\mathbf{Q}_2 \rightrightarrows_1 \mathbf{Q}_1$ if for each $Q_1 \in \mathbf{Q}_1$ and every atom $R(\mathbf{u}, \mathbf{v})$ in Q_1 there is a homomorphism h from some $Q_2 \in \mathbf{Q}_2$ to Q_1 with $R(\mathbf{u}, \mathbf{v})$ in the image of h.

We shall see below that this condition is only adequate for those semirings in $\mathbf{S}_{\text{hcov}}$ which have offset 1. To define a general condition we need to use complete descriptions. First we extend the definition of $\rightrightarrows_1$ to complete descriptions, in the expected way: the homomorphisms from CCQs of $\wr\mathbf{Q}_2\wr$ covering all atoms of all CCQs in $\wr\mathbf{Q}_1\wr$ should preserve inequalities. Notice that $\mathbf{Q}_2 \rightrightarrows_1 \mathbf{Q}_1$ iff $\wr\mathbf{Q}_2\wr \rightrightarrows_1 \wr\mathbf{Q}_1\wr$.

In the condition $\wr\mathbf{Q}_2\wr \rightrightarrows_2 \wr\mathbf{Q}_1\wr$ for semirings with offset 2, we also require that every CCQ without automorphisms having multiplicity more than one in $\wr\mathbf{Q}_1\wr$ has to be covered by two CCQs in $\wr\mathbf{Q}_2\wr$. Formally, we have that $\wr\mathbf{Q}_2\wr \rightrightarrows_2 \wr\mathbf{Q}_1\wr$, if (1) $\wr\mathbf{Q}_2\wr \rightrightarrows_1 \wr\mathbf{Q}_1\wr$, and (2) for every CCQ Q_1 in $\wr\mathbf{Q}_1\wr$ without nontrivial automorphisms (preserving inequalities)
- either there exist two CCQs[7] $Q_2', Q_2'' \in \wr\mathbf{Q}_2\wr$ such that $Q_2' \to Q_1$ and $Q_2'' \to Q_1$,
- or $\min(\wr\mathbf{Q}_1\wr[Q_1^{\approx}], 2) \leq \wr\mathbf{Q}_2\wr[Q_1^{\approx}]$.

Finally, we can present the desired characterization.

Proposition 5.21 *For* $k = 1, 2$ *the following are equivalent:*
- *semiring $\mathcal{K}$ belongs to* $\mathbf{S}^k_{\text{hcov}}$*;*
- $\wr\mathbf{Q}_2\wr \rightrightarrows_k \wr\mathbf{Q}_1\wr$ *implies* $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$*, for all UCQs* $\mathbf{Q}_1$*,* $\mathbf{Q}_2$*.*

Similarly to the previous conditions, it is possible to identify classes for which $\rightrightarrows_k$ is necessary for containment of UCQs. For $k = 1, 2$ denote by $\mathbf{N}^k_{\text{hcov}}$ the class of all semirings $\mathcal{K}$ such that for every coefficient ℓ and polynomial $\mathsf{P} \in \mathbb{N}[X]$ without a constant term, the inequality

$$\ell(x_1 \times \ldots \times x_n) \preceq_{\mathcal{K}} \mathsf{P}$$

implies that P uses all the variables $x_1, \ldots, x_n$ and has no less than $\min(\ell, k)$ monomials.

Proposition 5.22 *For every* $k = 1, 2$ *the following are equivalent:*
- *semiring $\mathcal{K}$ belongs to* $\mathbf{N}^k_{\text{hcov}}$*;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *implies* $\wr\mathbf{Q}_2\wr \rightrightarrows_k \wr\mathbf{Q}_1\wr$*, for all UCQs* $\mathbf{Q}_1$*,* $\mathbf{Q}_2$*.*

Notice that the bag semantics semiring $\mathcal{N}$ belongs to $\mathbf{N}^2_{\text{hcov}}$. Thus the condition $\wr\mathbf{Q}_2\wr \rightrightarrows_2 \wr\mathbf{Q}_1\wr$ is of particular interest, since it is a new necessary condition for $\mathcal{N}$-containment of UCQs. This improves on conditions known previously [6].

Corollary 5.23 *If* $\mathbf{Q}_1 \subseteq_{\mathcal{N}} \mathbf{Q}_2$ *then* $\wr\mathbf{Q}_2\wr \rightrightarrows_2 \wr\mathbf{Q}_1\wr$*.*

As usual, for the intersections $\mathbf{C}^k_{\text{hcov}} = \mathbf{S}^k_{\text{hcov}} \cap \mathbf{N}^k_{\text{hcov}}$ we have the following theorem. The NP-completeness of UCQ

[7] CQs Q_2' and Q_2'' still may be isomorphic or even coincide.

$\mathcal{K}$-containment of CQs				$\mathcal{K}$-containment of UCQs			
class	key axioms	homomorphism type	compl.	sub-class	extra axiom	homomorphism type	compl.
$\mathbf{C}_{\text{hom}}$	$\otimes$-idempotence $\mathbb{1}$-annihilation	$Q_2 \to Q_1$ (usual)	NP-c†	$\mathbf{C}_{\text{hom}}$	—	$\mathbf{Q}_2 \to \mathbf{Q}_1$	NP-c†
$\mathbf{C}_{\text{hcov}}$	$\otimes$-idempotence	$Q_2 \rightrightarrows Q_1$ (hom. cov.)	NP-c†	$\mathbf{C}^1_{\text{hcov}}$ $\mathbf{C}^2_{\text{hcov}}$	offset 1 —	$\mathbf{Q}_2 \rightrightarrows_1 \mathbf{Q}_1$ $\wr\mathbf{Q}_2\wr \rightrightarrows_2 \wr\mathbf{Q}_1\wr$	NP-c† in Π^p_2
$\mathbf{C}_{\text{in}}$	$\mathbb{1}$-annihilation	$Q_2 \hookrightarrow Q_1$ (injective)	NP-c	$\mathbf{C}^1_{\text{in}}$	—	$\mathbf{Q}_2 \hookrightarrow \mathbf{Q}_1$	NP-c
$\mathbf{C}_{\text{sur}}$	$\otimes$-semi-idempotence	$Q_2 \twoheadrightarrow Q_1$ (surjective)	NP-c†	$\mathbf{C}^1_{\text{sur}}$ $\mathbf{C}^\infty_{\text{sur}}$	offset 1 —	$\mathbf{Q}_2 \twoheadrightarrow_1 \mathbf{Q}_1$ $\wr\mathbf{Q}_2\wr \twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$	NP-c† in EXPTIME
$\mathbf{C}_{\text{bi}}$	—	$Q_2 \hookrightarrow\!\!\!\!\twoheadrightarrow Q_1$ (bijective)	NP-c†	$\mathbf{C}^1_{\text{bi}}$ $\mathbf{C}^{k>1}_{\text{bi}}$ $\mathbf{C}^\infty_{\text{bi}}$	offset 1 offset k —	$\mathbf{Q}_2 \hookrightarrow\!\!\!\!\twoheadrightarrow_1 \mathbf{Q}_1$ $\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_k \wr\mathbf{Q}_1\wr$ $\wr\mathbf{Q}_2\wr \hookrightarrow\!\!\!\!\twoheadrightarrow_\infty \wr\mathbf{Q}_1\wr$	NP-c† in Π^p_2 in coNP$^{\#P}$

Table 1: Summary of semiring classes and complexity (results known before are marked by †). Key axioms define the corresponding sufficient classes; the axioms for the necessary classes are omitted for clarity.

$\mathcal{K}$-containment for semirings from $\mathbf{C}^1_{\text{hcov}}$ was first provided in [14] for the case of the lineage semiring $\mathtt{Lin}[X]$.

THEOREM 5.24
(1) Given a number $k = 1, 2$, the following are equivalent:
- *semiring $\mathcal{K}$ belongs to $\mathbf{C}^k_{\text{hcov}}$;*
- $\mathbf{Q}_1 \subseteq_{\mathcal{K}} \mathbf{Q}_2$ *iff* $\wr\mathbf{Q}_2\wr \rightrightarrows_k \wr\mathbf{Q}_1\wr$, *for all UCQs* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$.

(2) UCQ $\mathcal{K}$-CONTAINMENT *is* NP*-complete if* $\mathcal{K} \in \mathbf{C}^1_{\text{hcov}}$, *and in* Π^p_2 *if* $\mathcal{K} \in \mathbf{C}^2_{\text{hcov}}$.

6. CONCLUSION

We have studied containment of CQs and UCQs over annotated relations. We have established several interesting classes of semirings for which these problems are decidable by means of different syntactic criteria, developed by modifying and extending the well-known notion of homomorphism between CQs. Our work extends previous results on the subject and should have practical implications, since most semirings used for annotations in the literature fall into one of these well-behaved classes. Tab. 1 provides a summary, with complexity bounds for checking the associated criteria. For semirings that do not fall into these classes, we have extended the range of available machinery for query optimization problems, by providing generalized or improved necessary and sufficient conditions. For some of these semirings we also suggest new decision procedures based on small model properties.

Many problems remain open. In particular, we would like to continue studying the *small model property* approach, either proving or disproving that such methods can work for semirings with non-idempotent addition. It is also interesting to study *CQ-admissible polynomials* on their own, and in particular how to decide containment over them. We believe that this study may have consequences for solving some of the fundamental open problems in the area of query optimization. Finally, we anticipate that the concept of *complete descriptions* opens new possibilities to solve containment and equivalence problems over different semantics for not only CQs and UCQs, but for a much wider range of queries.

Acknowledgements We thank Peter Buneman, Jeff Egger, Diego Figueira and Tony Tan for useful discussions, and Todd J. Green for comments on previous results. Support provided by FET-Open Project FoX, grant agreement 233599; and EPSRC grants F028288/1 and G049165.

7. REFERENCES

[1] F.N. Afrati, M. Damigos, M. Gergatsoulis. Query containment under bag and bag-set semantics. *IPL* **110**(10), 2010.
[2] P. Buneman, J. Cheney, W.C. Tan, S. Vansummeren. Curated databases. *PODS* 2008, 1–12.
[3] P. Buneman, S. Khanna, W.C. Tan. Why and Where: a characterization of data provenance. *ICDT* 2001, 316–330.
[4] P. Buneman, E.V. Kostylev. Annotation Algebras for RDFS. *SWPM* 2010. CEUR Workshop Proc.
[5] A.K. Chandra, P.M. Merlin. Optimal implementation of conjunctive queries in relational data bases. *STOC* 1977.
[6] S. Chaudhuri, M.Y. Vardi. Optimization of *real* conjunctive queries. *PODS* 1993, 59–70.
[7] R. Chirkova. Equivalence and minimization of conjunctive queries under combined semantics. *ICDT* 2012.
[8] S. Cohen, W. Nutt, Y. Sagiv. Deciding equivalences among conjunctive aggregate queries. *JACM* **54**(2), 2007.
[9] Y. Cui, J. Widom, J.L. Wiener. Tracing the lineage of view data in a warehousing environment. *ACM ToDS* **25**(2), 179–227, 2000.
[10] A. Das Sarma, M. Theobald, J. Widom. Exploiting lineage for confidence computation in uncertain and probabilistic databases. *ICDE* 2008, 1023–1032.
[11] R. Fagin, P.G. Kolaitis, L. Popa, W.C. Tan. Quasi-inverses of schema mappings. *ACM ToDS* **33**(2), 2008.
[12] N. Fuhr, T. Rölleke. A probabilistic relational algebra for the integration of information retrieval and database systems. *ACM ToIS* **15**(1), 32–66, 1997.
[13] G. Grahne, N. Spyratos, D. Stamate. Semantics and containment of queries with internal and external conjunctions. *ICDT* 1997, *LNCS* 1186, 71–82.
[14] T. Green. Containment of conjunctive queries on annotated relations. *Th. Comp. Syst.* **49**(2), 429–459, 2011.
[15] T.J. Green, G. Karvounarakis, V. Tannen. Provenance semirings. *PODS* 2007, 31–40.
[16] T. Imieliński, W. Lipski, Jr. Incomplete information in relational databases. *JACM* **31**(4), 761–791, 1984.
[17] Y.E. Ioannidis, R. Ramakrishnan. Containment of conjunctive queries: beyond relations as sets. *ACM ToDS* **20**(3), 1995.
[18] T.S. Jayram, P.G. Kolaitis, E. Vee. The containment problem for *real* conjunctive queries with inequalities. *PODS* 2006.
[19] D. Olteanu, J. Závodný. Factorised representations of query results: size bounds and readability. *ICDT* 2012.
[20] Y. Sagiv, M. Yannakakis. Equivalences among relational expressions with the union and difference operators. *JACM* **27**(4), 633–655, 1980.
[21] E. Zimányi. Query evaluation in probabilistic relational databases. *TCS* **171**(1–2), 179–219, 1997.
[22] A. Zimmermann, N. Lopes, A. Polleres, U. Straccia. A general framework for representing, reasoning and querying with annotated Semantic Web data. *Web Semantics*. In press.