Structure and Complexity of Bag Consistency

Albert Atserias
Universitat Politècnica de Catalunya
Barcelona, Catalonia
Spain
atserias@cs.upc.edu

Phokion G. Kolaitis
UC Santa Cruz and IBM Research
Santa Cruz, California
USA
kolaitis@ucsc.edu

ABSTRACT

Since the early days of relational databases, it was realized that acyclic hypergraphs give rise to database schemas with desirable structural and algorithmic properties. In a by-now classical paper, Beeri, Fagin, Maier, and Yannakakis established several different equivalent characterizations of acyclicity; in particular, they showed that the sets of attributes of a schema form an acyclic hypergraph if and only if the local-to-global consistency property for relations over that schema holds, which means that every collection of pairwise consistent relations over the schema is globally consistent. Even though real-life databases consist of bags (multisets), there has not been a study of the interplay between local consistency and global consistency for bags. We embark on such a study here and we first show that the sets of attributes of a schema form an acyclic hypergraph if and only if the local-to-global consistency property for bags over that schema holds. After this, we explore algorithmic aspects of global consistency for bags by analyzing the computational complexity of the global consistency problem for bags: given a collection of bags, are these bags globally consistent? We show that this problem is in NP, even when the schema is part of the input. We then establish the following dichotomy theorem for fixed schemas: if the schema is acyclic, then the global consistency problem for bags is solvable in polynomial time, while if the schema is cyclic, then the global consistency problem for bags is NP-complete. The latter result contrasts sharply with the state of affairs for relations, where, for each fixed schema, the global consistency problem for relations is solvable in polynomial time.

1. INTRODUCTION

This paper bring together two different strands of research in database theory: the study of global consistency and the study of bag semantics. Before presenting an overview of our main results, we provide some background to each of these two strands.

Minor revision of the paper with the same title published in the Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS 2021), ISBN978-1-4503-8381-3/21/06, June 20-25, 2021. DOI:10.1145/3452021.3458329

The study of global consistency in relational databases arose from the universal relation model, which is the assumption that all relations at hand are projections of a single relation, called the *universal relation*. Much of the work on database dependencies and normalization during the 1970s made this assumption first implicitly and then explicitly, as for instance in the paper by Beeri, Bernstein, and Goodman [8]. The universal relation model implies that occurrences of the same attribute in different relations have the same meaning; it also provides a framework to study dependencies across different relations. Furthermore, it has been argued that the universal relation model yields logical independence and access-path independence [22], thus it can be regarded as an early model of data integration. At times, the universal relation model was surrounded by controversy with arguments both against it [18] and in favor of it [24]. The controversy notwithstanding and instead of assuming the presence of a universal relation, researchers also investigated *when* a universal relation exists.

On the algorithmic side, the universal relation problem (also known as the global consistency problem) is the following decision problem: given relations $R_1, \ldots, R_m$, is there a relation R such that, for every $i \leq m$, the projection of R on the attributes of R_i is equal to R_i? If the answer is positive, then the relations $R_1, \ldots, R_m$ are said to be *globally consistent* and R is said to be a *universal relation* for them or a *witness* to their global consistency. Honeyman, Ladner, and Yannakakis [15] showed that the universal relation problem is NP-complete, even for relations of arity 2.

On the structural side, the problem is to characterize when a collection of relations is globally consistent. It is easy to see that if the relations $R_1, \ldots, R_m$ are globally consistent, then they are pairwise consistent (i.e., every two of them are globally consistent). As pointed out in [15], however, the converse does not hold in general; in other words, pairwise consistency is a necessary but not sufficient condition for global consistency. This state of affairs raised the question: can we identify the settings in which pairwise consistency is both a necessary and sufficient condition for global consistency? Let $R_1, \ldots, R_m$ be a collection of relations over a schema with $X_1, \ldots, X_m$ as the sets of attributes. The sets $X_1, \ldots, X_m$ can be viewed as the hyperedges of a hypergraph. Beeri et al. [9] showed that the sets of attributes of a schema form an acyclic hypergraph if and only if the local-to-global consistency property for relations over that schema holds, which means that every collection of pairwise consistent relations over the schema is globally consistent. Thus, for acyclic schemas, pairwise consistency is necessary and sufficient for global consistency. Consequently, the universal relation problem is solvable in polynomial time, if the sets of attributes of the schema form an acyclic hypergraph.

Much of the research in database theory assumes that relations are sets. In 1993, Chaudhuri and Vardi [12] pointed out that there is a gap between database theory and database practice because "real" databases use bags (multisets). They called for a re-examination of the foundations of databases where the fundamental concepts and algorithmic problems are investigated under bag semantics, instead of set semantics. In particular, Chaudhuri and Vardi [12] raised the question of the decidability of the conjunctive query containment problem under bags semantics (the same problem under set semantics is known to be NP-complete [11]). In spite of various efforts in the past and some recent progress [19, 20], this question remains unanswered at present.

It is perhaps surprising that a study of consistency notions under bag semantics has not been carried so far. Our main goal here is to embark on such a study and to explore both structural and algorithmic aspects of pairwise consistency and of global consistency under bag semantics. In this study, the consistency notions for bags are, of course, defined using bag semantics in the computation of projections.

Summary of Results In general, properties of relations need not carry over automatically to similar properties of bags. This phenomenon manifests itself in the context of consistency properties. Indeed, it is well known that if a collection of relations is globally consistent, then their relational join is a witness to their global consistency (see, e.g., [15]); in other words, their relational join is a universal relation for them and, in fact, it is the largest universal relation. In contrast, we point out that this property fails for bags, i.e., there is a collection of bags that is globally consistent but the bag-join of the bags in the collection is not a witness to their global consistency; furthermore, there may be no biggest witness to the consistency of these bags.

Our first result asserts that two bags are consistent if and only if they have the same projection on their common attributes. While the analogous fact for relations is rather trivial, here we need to bring in tools from the theory of linear programming and maximum flow problems. As a corollary, we obtain a polynomial-time algorithm for checking whether two given bags are consistent and returning a witness to their consistency, if they are consistent. After this, we establish our main result concerning the structure of bag consistency. Specifically, we show that the sets of attributes of a schema form an acyclic hypergraph if and only if the local-to-global consistency for bags over that schema holds. Thus, the main finding by Beeri et al. [9] about acyclicity and consistency extends to bags. The architecture of the proof, however, is different from that in [9]. In particular, if a schema is cyclic, we give an explicit construction of a collection of bags that are pairwise consistent, but not globally consistent; the inspiration for our construction comes from an earlier construction of hard-to-prove tautologies in propositional logic by Tseitin [23].

We then explore algorithmic aspects of global consistency for bags by analyzing the computational complexity of the global consistency problem for bags: given a collection of bags, are these bags globally consistent? Using a sparse-model property of integer programming that is reminiscent of Carathéodory's Theorem for conic hulls [13], we first show that this problem is in NP, even when the schema is part of the input. After this, we establish the following dichotomy theorem for fixed schemas: if the schema is acyclic, then the global consistency problem for bags is solvable in polynomial time, while if the schema is cyclic, then the global consistency problem for bags is NP-complete. The latter result contrasts sharply with the state of affairs for relations, where, for each fixed schema, the global consistency problem for relations is solvable in polynomial time. Our NP-hardness results build on an earlier NP-hardness result about three-dimensional statistical data tables by Irving and Jerrum [17], which was later on refined by De Loera and Onn [21]. Translated into our context, this result asserts the NP-hardness of the global consistency problem for bags over the triangle hypergraph, i.e., the hypergraph with hyperedges of the form $\{A_1, A_2\}, \{A_2, A_3\}, \{A_3, A_1\}$.

We conclude the paper with a brief overview of extensions of the results reported here to relations over semirings.

Related Work The interplay between local consistency and global consistency arises naturally in several different settings. Already in 1962, Vorob'ev [25] studied this interplay in the setting of probability distributions and characterized the local-to-global consistency property for probability distributions in terms of a structural property of hypergraphs that turned out to be equivalent to hypergraph acyclicity. It appears that Beeri et al. [9] were unaware of Vorob'ev work, but later on Vorob'ev's work was cited in a survey of database theory by Yannakakis [27]. In recent years, the interplay between local consistency and global consistency has been explored at great depth in the setting of quantum information by Abramsky and his collaborators (see, e.g., [3, 4, 5]). In that setting, the interest is in contextuality phenomena, which are situations where collections of measurements are locally consistent but globally inconsistent - Bell's celebrated theorem [10] is an instance of this. The similarities between these different settings (probability distributions, relational databases, and quantum mechanics) were pointed out explicitly by Abramsky [1, 2]. This also raised the question of developing a unifying framework in which, among other things, the results by Vorob'ev and the results by Beeri et al. are special cases of a single result. Using a relaxed notion of consistency, we established such a result for relations over semirings [6]. For the bag semiring, however, the relaxed notion of consistency that we studied in [6] is essentially equivalent to the consistency of probability distributions with rational values (and not to the consistency of bags). This left open the question of exploring the interplay between (the standard notions of) local consistency and global consistency for bags, which is what we set to do in the present paper.

2. RELATIONAL CONSISTENCY

Basic Notions An *attribute* A is a symbol with an associated set $\mathrm{Dom}(A)$ called its *domain*. If X is a finite set of attributes, then we write $\mathrm{Tup}(X)$ for the set of X*-tuples*; this means that $\mathrm{Tup}(X)$ is the set of functions that take each attribute $A \in X$ to an element of its domain $\mathrm{Dom}(A)$. Note that $\mathrm{Tup}(\emptyset)$ is non-empty as it contains the *empty tuple*, i.e., the unique function with empty domain. If $Y \subseteq X$ is a subset of attributes and t is an X-tuple, then the *projection of* t *on* Y, denoted by $t[Y]$, is the unique Y-tuple that agrees with t on Y. In particular, $t[\emptyset]$ is the empty tuple.

Let X be a set of attributes. A *relation* over X is a function $R : \mathrm{Tup}(X) \rightarrow \{0, 1\}$. We write $R(X)$ to emphasize the fact that R is a relation over *schema* X. The *support* $\mathrm{Supp}(R)$ of R is the set of X-tuples t with a non-zero value, i.e., $\mathrm{Supp}(R) := \{t \in \mathrm{Tup}(X) : R(t) \neq 0\}$. Whenever no confusion arises, we write R' to denote $\mathrm{Supp}(R)$. We say that R is *finite* if its support R' is a finite set. In what follows, we will make the blanket assumption that all relations considered are finite, so we will omit the term "finite". Every relation R can be identified with its support R', thus every relation R can be viewed as a finite set of X-tuples.

Let R be a relation over X and assume that $Z \subseteq X$. The *projection* $R[Z]$ of R on Z is the relation over Z consisting of all projections $t[Z]$, for $t \in R$. It is easy to see that for all $W \subseteq Z \subseteq X$, we have $R[Z][W] = R[W])$.

If X and Y are sets of attributes, then we write XY as shorthand for the union $X \cup Y$. Accordingly, if x is an X-tuple and y is a Y-tuple such that that $x[X \cap Y] = y[X \cap Y]$, then we write xy to denote the XY-tuple that agrees with x on X and on y on Y. We say that x *joins with* y, and that y *joins with* x, to *produce* the tuple xy.

The *join* $R \bowtie S$ of two relations $R(X)$ and $S(Y)$ is the relation over XY consisting of all XY-tuples t such that $t[X]$ is in R and $t[Y]$ is in S, i.e., all tuples of the form xy such that $x \in R$, $y \in S$, and x joins with y.

Consistency of Two Relations Assume that $R(X)$ and $S(X)$ are two relations over the schemas X and Y. We say that $R(X)$ and $S(Y)$ are *consistent* if there is a relation T over XY such that $T[X] = R$ and $T[Y] = S$. We also say that T *witnesses* the consistency of R and S. The next proposition, whose proof is straightforward, gives a useful criterion for the consistency of R and S.

PROPOSITION 1. *Let $R(X)$ and $S(Y)$ be two relations. The following statements are equivalent:*

1. *R and S are consistent.*
2. *$R[X \cap Y] = S[X \cap Y]$.*

Global Consistency of Relations Let $R_1(X_1), \ldots, R_m(X_m)$ be relations over the schemas $X_1, \ldots, X_m$. We say that the collection $R_1, \ldots, R_m$ is *globally consistent* if there is a relation T over $X_1 \cup \cdots \cup X_m$ such that $R_i = T[X_i]$ for all $i \in [m] = \{1, \ldots, m\}$. We say that T *witnesses* the global consistency of $R_1, \ldots, R_m$, and we call it a *universal relation* for $R_1, \ldots, R_m$. The next result presents well known and easy to prove facts about global consistency (see, e.g., [15]).

PROPOSITION 2. *Assume that $R_1(X_1), \ldots, R_m(X_m)$ are relations over the schemas $X_1, \ldots, X_m$.*

- *If T is a relation witnessing the global consistency of the relations $R_1, \ldots, R_m$, then $T \subseteq R_1 \bowtie \cdots \bowtie R_m$.*
- *The collection $R_1, \ldots, R_m$ is globally consistent if and only if $(R_1 \bowtie \cdots \bowtie R_m)[X_i] = R_i$ for all $i = 1, \ldots, m$.*

Consequently, if the collection $R_1, \ldots, R_m$ is globally consistent, then the join $R_1 \bowtie \cdots \bowtie R_m$ is the *largest* universal relation for $R_1, \ldots, R_m$.

In relational database theory, there has been an extensive study of both the structural and the algorithmic aspects of global consistency. We begin by surveying some of the results concerning the structural aspects of global consistency. The main problem is to characterize when a collection of relations is globally consistent.

We say that the relations $R_1(X_1), \ldots, R_m(X_m)$ are *pairwise consistent* if for every $i, j \in [m]$, the relations $R_i(X_i)$ and $R_j(X_j)$ are consistent. Clearly, if a relation T witnesses the global consistency of $R_1, \ldots, R_m$, then the relation $T[X_iX_j]$ witnesses the consistency of R_i and R_j, for every $i, j \in [m]$. Thus, if the collection $R_1, \ldots, R_m$ is globally consistent, then the relations $R_1, \ldots, R_m$ are pairwise consistent. The converse, however, is not true, in general. Indeed, let $X_1 = \{A_1, A_2\}$, $X_2 = \{A_2, A_3\}$, $X_3 = \{A_3, A_1\}$ and consider the relations $R_1(A_1A_2) = \{00, 11\}$, $R_2(A_2A_3) = \{01, 10\}$, $R_3(A_3A_1) = \{00, 11\}$. By Proposition 1, these relations are pairwise consistent. By Proposition 2, however, they are not globally consistent because $R_1 \bowtie R_2 \bowtie R_3 = \emptyset$.

Beeri, Fagin, Maier, and Yannakakis [9] characterized the set of schemas for which pairwise consistency is a necessary and sufficient condition for global consistency of relations. Their characterization involves notions from hypergraph theory that we now review.

Acyclic Hypergraphs A *hypergraph* is a pair $H = (V, E)$, where V is a set of *vertices* and E is a set of *hyperedges*, each of which is a non-empty subset of V. Every collection $X_1, \ldots, X_m$ of sets of attributes can be identified with a hypergraph $H = (V, E)$, where $V = X_1 \cup \cdots \cup X_m$ and $E = \{X_1, \ldots, X_m\}$. Conversely, every hypergraph $H = (V, E)$ gives rise to a collection $X_1, \ldots, X_m$ of sets of attributes, where $X_1, \ldots, X_m$ are the hyperedges of H. Thus, we can move from collections of sets of attributes to hypergraphs, and vice versa. The notion of an *acyclic* hypergraph generalizes the notion of an acyclic graph. Since we will not work directly with the definition of an acyclic hypergraph, we refer the reader to [9] for the precise definition. Instead, we focus on other notions that are equivalent to hypergraph acyclicity and will be of interest to us in the sequel.

Conformal and Chordal Hypergraphs The *primal* graph of a hypergraph $H = (V, E)$ is the undirected graph that has V as its set of vertices and has an edge between any two distinct vertices that appear together in at least one hyperedge of H. A hypergraph H is *conformal* if the set of vertices of every clique (i.e., complete subgraph) of the primal graph of H is contained in some hyperedge of H. A hypergraph H is *chordal* if its primal graph is chordal, that is, if every cycle of length at least four of the primal graph of H has a chord (i.e., an edge that connects two nodes on the cycle, but is not one of the edges of the cycle). To illustrate these concepts, let $V_n = \{A_1, \ldots, A_n\}$ be a set of n vertices and consider the hypergraphs

$$P_n = (V_n, \{A_1, A_2\}, \ldots, \{A_{n-1}, A_n\}) \quad (1)$$
$$C_n = (V_n, \{A_1, A_2\}, \ldots, \{A_{n-1}, A_n\}, \{A_n, A_1\}) \quad (2)$$
$$H_n = (V_n, \{V_n \setminus \{A_i\} : 1 \leq i \leq n\}) \quad (3)$$

If $n \geq 2$, then the hypergraph P_n is both conformal and chordal. The hypergraph $C_3 = H_3$ is chordal, but not conformal. For every $n \geq 4$, the hypergraph C_n is conformal, but not chordal, while the hypergraph H_n is chordal, but not conformal.

Running Intersection Property A hypergraph H has the *running intersection property* if there is a listing $X_1, \ldots, X_m$ of all hyperedges of H such that for every $i \in [m]$ with $i \geq 2$, there exists a $j < i$ such that $X_i \cap (X_1 \cup \cdots \cup X_{i-1}) \subseteq X_j$.

Join Tree A *join tree* for a hypergraph H is an undirected tree T with the set E of the hyperedges of H as its vertices and such that for every vertex v of H, the set of vertices of T containing v forms a subtree of T, i.e., if v belongs to two vertices X_i and X_j of T, then v belongs to every vertex of T in the unique simple path from X_i to X_j in T.

Local-to-Global Consistency Property for Relations Let H be a hypergraph and let $X_1, \ldots, X_m$ be a listing of all hyperedges of H. We say that H has the *local-to-global consistency property for relations* if every pairwise consistent collection $R_1(X_1), \ldots, R_m(X_m)$ of relations over the schemas $X_1, \ldots, X_m$ is globally consistent.

We are ready to state the main result in Beeri et al. [9].

Theorem 1 (Theorem 3.4 in [9]). *Let H be a hypergraph. The following statements are equivalent:*

(a) H is an acyclic hypergraph.

(b) H is a conformal and chordal hypergraph.

(c) H has the running intersection property.

(d) H has a join tree.

(e) H has the local-to-global consistency property for relations.

As an illustration, if $n \geq 2$, the hypergraph P_n is acyclic, hence it has the local-to-global consistency property for relations. In contrast, if $n \geq 3$, the hypergraphs C_n and H_n are cyclic, hence they do not have the local-to-global consistency property for relations.

Complexity of Global Consistency for Relations We now discuss the algorithmic aspects of global consistency. The *global consistency problem for relations* (also known as the *universal relation problem for relations*) asks: given a hypergraph $H = (V, \{X_1, \ldots, X_m\})$ and relations $R_1, \ldots, R_m$ over H, is the collection $R_1, \ldots, R_m$ globally consistent? Honeyman, Ladner, and Yannakakis [15] established the following result.

Theorem 2. *The global consistency problem for relations is* NP*-complete.*

The NP-hardness of the global consistency problem for relations is proved via a reduction from 3-Colorability in which each relation has arity 2 and consists of just six pairs. Specifically, each edge (u, v) in a given graph G gives rise to a relation of arity 2 with attributes u and v; the six pairs in the relation are the pairs of different colors chosen from the three colors "red", "blue", and "green". The membership in NP uses the observation that if a collection $R_1, \ldots, R_m$ of relations is globally consistent, then a witness W of this fact can be obtained as follows: for each $i \leq m$ and each tuple $t \in R_i$, pick a tuple in the join $R_1 \bowtie \cdots \bowtie R_m$ that extends t and insert it in W. In particular, the cardinality $|W|$ of W is bounded by the sum $\sum_{i=1}^{m} |R_i| \leq m \max\{|R_i| : i \in [m]\}$, and thus the size of W is bounded by a polynomial in the size of the input hypergraph H and the input relations $R_1, \ldots, R_m$.

Several restricted cases of the global consistency problem for relations turn out to be solvable in polynomial time.

First, Proposition 1 implies that the consistency problem for two relations is solvable in polynomial time, since it amounts to checking that the two given relations $R(X)$ and $S(Y)$ have the same projection on $X \cap Y$.

Second, from the preceding fact and from Theorem 1, it follows that the global consistency problem for relations is solvable in polynomial time when restricted to acyclic hypergraphs, since, in this case, the global consistency of a collection of relations is equivalent to the pairwise consistency of the relations in the collection.

Finally, for every fixed hypergraph $H = (V, \{X_1, \ldots, X_m\})$ (be it cyclic or acyclic), the global consistency problem restricted to relations $R_1(X_1), \ldots, R_m(X_m)$ with sets of attributes $X_1, \ldots, X_m$ is also solvable in polynomial time. This is so because, by Proposition 2, one can first compute the join $J = R_1 \bowtie \cdots \bowtie R_m$ in polynomial time and then check whether $J[X_i] = R_i$ holds, for $i = 1, \ldots, m$. While the cardinality $|J|$ of this witness J can only be bounded by $\prod_{i=1}^{m} |R_i| \leq \max\{|R_i| : i \in [m]\}^m$, this cardinality is still polynomial in the size of the input because, in this case, the exponent m is fixed and not part of the input.

3. BAG CONSISTENCY

Basic Notions Let X be a set of attributes. A *bag* over X is a function $R : \mathrm{Tup}(X) \to \{0, 1, 2, \ldots\}$. As with relations, we write $R(X)$ to emphasize the fact that R is a bag over X; the support $\mathrm{Supp}(R)$ (also denoted by R') of R is the set of X-tuples t that are assigned non-zero value. We say that R is *finite* if its support R' is a finite set. In the sequel, we will assume that all bags are finite.

If R is a bag and t is an X-tuple, then the non-negative integer $R(t)$ is called the *multiplicity* of t in R; we write $t : R(t)$ to denote that the multiplicity of t in R is equal to $R(t)$. Every bag R can be viewed as a finite set of elements of the form $t : R(t)$, where $t \in R'$ and $R(t) \neq 0$. A bag can also be represented in tabular form. For example, the table

A	B	#
a_1	b_1 :	2
a_2	b_2 :	1
a_3	b_3 :	5

represents the bag $R = \{(a_1, b_1) : 2, (a_2, b_2) : 1, (a_3, b_3) : 5\}$. Let R be a bag over X and assume that $Z \subseteq X$. If t is a Z-tuple, then the *marginal of R over t* is defined by

$$R(t) := \sum_{\substack{r \in R': \\ r[Z]=t}} R(r). \tag{4}$$

Thus, every bag R over X induces a bag over Z, called the *marginal of R on Z* and denoted by $R[Z]$. It is easy to verify that for all $W \subseteq Z \subseteq X$, we have $R[Z][W] = R[W]$.

Let R be a bag over X and S a bag over Y. The *bag join* $R \bowtie_b S$ of R and S is the bag over XY having support $R' \bowtie S'$ and such that every XY-tuple $t \in R' \bowtie S'$ has multiplicity $(R \bowtie_b S)(t) = R(t[X]) \times S(t[Y])$.

Consistency of Two Bags Two bags $R(X)$ and $S(Y)$ are *consistent* if there is a bag $T(XY)$ such that $T[X] = R$ and $T[Y] = S$, where now the projections are computed according to Equation (4); we say that T *witnesses* the consistency of R and S. It is easy to see that if $R(X)$ and $S(Y)$ are consistent bags and T is a bag that witnesses their consistency, then $T' \subseteq R' \bowtie S'$, that is, the support of T is contained in the join of the supports of R and S.

By Proposition 1, if two relations $R(X)$ and $S(Y)$ are consistent, then their join $R \bowtie S$ witnesses their consistency; moreover, $R \bowtie S$ is the largest relation that has this property. In contrast, this is not true for bags because there are consistent bags $R(X)$ and $S(Y)$ such that the support T' of every bag T witnessing their consistency is a proper subset of $R' \bowtie S'$. For example, consider the bags $R_1(AB) = \{(1,2) : 1,\ (2,2) : 1\}$ and $S_1(BC) = \{(2,1) : 1,\ (2,2) : 1\}$; their consistency (as bags) is witnessed by the bags $T_1(ABC) = \{(1,2,2) : 1,\ (2,2,1) : 1\}$ and $T_2(ABC) = \{(1,2,1) : 1,\ (2,2,2) : 1\}$, but no other bag. This example can be extended as follows. For $n \geq 2$, let $R_{n-1}(A,B)$ and $S_{n-1}(B,C)$ be the bags

$$\{(1,2):1,\ (2,2):1,\ \ldots,\ (1,n):1,(n,n):1\}$$
$$\{(2,1):1,\ (2,2):1,\ \ldots,\ (n,1):1,(n,n):1\},$$

respectively. For every $n \geq 2$, the bags R_{n-1} and S_{n-1} are consistent. In fact, there are exactly 2^{n-1} bags witnessing their consistency; these witnesses are pairwise incomparable in the bag-containment sense and their supports are properly contained in the support $(R_{n-1} \Join_b S_{n-1})'$ of the bag join $R_{n-1} \Join_b S_{n-1}$. Note that the bags R_{n-1} and S_{n-1} are actually relations and that their join $R_{n-1} \Join S_{n-1}$ witnesses their consistency as relations, but not as bags.

By Proposition 1, two relations $R(X)$ and $S(Y)$ are consistent if and only if $R[X \cap Y] = S[X \cap Y]$. It is natural to ask if an analogous result holds true for bags. If two bags $R(X)$ and $S(Y)$ are consistent, then clearly $R[X \cap Y] = S[X \cap Y]$. The converse turns out to also be true, but its proof is far from obvious. We will establish the converse by bringing into the picture concepts from linear programming and from the theory of maximum flows.

With each pair of bags $R(X)$ and $S(Y)$, we associate the following linear program $P(R,S)$. Let $J = R' \Join S'$ be the join of the supports of R and S. For each $t \in J$, there is a variable x_t. For each $t \in J$ and $r \in R'$, define $a_{r,t} = 1$ if $t[X] = r$ and $a_{r,t} = 0$ if $t[X] \neq r$. Similarly, for each $t \in J$ and $s \in S'$, define $a_{s,t} = 1$ if $t[Y] = s$ and $a_{s,t} = 0$ if $t[Y] \neq s$. The constraints of $P(R,S)$ are:

$$\begin{array}{ll} \sum_{t \in J} a_{r,t} x_t = R(r) & \text{for } r \in R', \\ \sum_{t \in J} a_{s,t} x_t = S(s) & \text{for } s \in S', \\ x_t \geq 0 & \text{for } t \in J. \end{array} \tag{5}$$

The linear program $P(R,S)$ can be viewed as the set of the flow constraints of an instance of the max-flow problem. A *network* $N = (V,E,c,s,t)$ is a directed graph $G = (V,E)$ with a non-negative weight $c(u,v)$, called the *capacity*, assigned to each edge $(u,v) \in E$, and two distinguished vertices $s,t \in V$, called the *source* and the *sink*. A *flow* for the network is an assignment of non-negative weights $f(u,v)$ on the edges $(u,v) \in E$ so that both the capacity constraints and the flow constraints are respected, that is,

$$\begin{array}{lr} f(u,v) \leq c(u,v) & \text{for } (u,v) \in E, \\ \sum_{v \in N^-(u)} f(v,u) = \sum_{w \in N^+(u)} f(u,w) & \text{for } u \in V \setminus \{s,t\}, \end{array}$$

where $N^-(u)$ and $N^+(u)$ denote the sets of in-neighbors and out-neighbors of u in G. The *value* of such a flow is the quantity $\sum_{w \in N^+(s)} f(s,w) = \sum_{v \in N^-(t)} f(v,t)$, where the equality follows from the flow constraints. In the *max-flow problem*, the goal is to find a flow of maximum value. A flow is *saturated* if $f(s,w) = c(s,w)$ for every $w \in N^+(s)$ and $f(v,t) = c(v,t)$ for every $v \in N^-(t)$. It is obvious that if a saturated flow exists, then every max flow is saturated.

With each pair $R(X)$ and $S(Y)$ of bags, we associate the following network $N(R,S)$. The network has $1+|R'|+|S'|+1$ vertices: one source vertex s^*, one vertex for each tuple r in the support R' of R, one vertex for each tuple s in the support S' of S, and one target vertex t^*. There is an arc of capacity $R(r)$ from s^* to r for each $r \in R'$, an arc of capacity $S(s)$ from s to t^* for each $s \in S'$, and an arc of unbounded (i.e., very large) capacity from $t[X]$ to $t[Y]$ for each $t \in R' \Join S'$.

The next result yields several different characterizations of the consistency of two bags.

LEMMA 1. *Let $R(X)$ and $S(Y)$ be two bags. The following statements are equivalent:*

1. *$R(X)$ and $S(Y)$ are consistent.*
2. *$R[X \cap Y] = S[X \cap Y]$.*
3. *$P(R,S)$ is feasible over the rationals.*
4. *$P(R,S)$ is feasible over the integers.*
5. *$N(R,S)$ admits a saturated flow.*

PROOF. (*Sketch*) The equivalence of the statements (1) and (4) is immediate from the definitions. As discussed earlier, (1) implies (2). To show that (2) implies (3), we assume that $R[Z] = S[Z]$ and show that $P(R,S)$ is feasible over the rationals. For each $t \in J = R' \Join S'$, we set $x_t := R(t[X])S(t[Y])/R(t[Z]) = R(t[X])S(t[Y])/S(t[Z])$ (where the equality follows from the assumption that $R[Z] = S[Z]$) and verify that this is a rational solution of $P(R,S)$. For (3) implies (5), let $x^* = (x^*_t)_{t \in J}$ be a rational solution for $P(R,S)$ and let f be the following assignment for $N(R,S)$:

$$\begin{array}{ll} f(s^*,r) := c(s^*,r) = R(r) & \text{for each } r \in R'; \\ f(t[X],t[Y]) := x^*_t & \text{for each } t \in J; \\ f(s,t^*) := c(s,t^*) = S(s) & \text{for each } s \in S'. \end{array}$$

This assignment is a flow since the equations of $P(R,S)$ say that the flow-constraints are satisfied; furthermore, it is a saturated flow by construction. For (5) implies (1), let g be a saturated flow for $N(R,S)$; in particular, this is a max flow for $N(R,S)$. Since all capacities in $N(R,S)$ are integers, the integrality theorem for the max-flow problem asserts that there is a max flow f consisting of integers (see, e.g., [26]), which, of course, is also a saturated flow. Let $T(XY)$ be the bag defined by setting $T(t) := f(t[X],t[Y])$ for each $t \in R' \Join S'$. Since f is saturated, we have that $f(s^*,r) = c(s^*,r) = R(r)$ for each $r \in R'$ and $f(s,t^*) = c(s,t^*) = S(s)$ for each $s \in S'$. This means that the flow-constraints imply that T witnesses the consistency of R and S. Thus, the statements (1), (2), (3), and (5) are equivalent. □

The equivalence of statements (1) and (2) in Lemma 1 yields a simple polynomial-time test to determine the consistency of two bags, namely, given two bags $R(X)$ and $S(y)$, check whether or not $R[X \cap Y] = S[X \cap Y]$. Later on, we will see that the equivalence of statements (1) and (5) implies that there is a polynomial-time algorithm for constructing a witness to the consistency of two consistent bags.

Global Consistency for Bags Let $R_1(X_1), \ldots, R_m(X_m)$ be bags over the schemas $X_1, \ldots, X_m$. We say that the collection $R_1, \ldots, R_m$ is *globally consistent* if there a bag T over $X_1 \cup \cdots \cup X_m$ such that $T_i[X_i] = R_i$ for all $i \in [m]$. We say that the bag T *witnesses* the global consistency of the bags $R_1, \ldots, R_m$. As with relations, pairwise consistency of a collection of bags is a necessary, but not sufficient, condition for the global consistency of the collection. Let H be a hypergraph and let $X_1, \ldots, X_m$ be a listing of all hyperedges of H. We say that H has the *local-to-global consistency property for bags* if every pairwise consistent collection $R_1(X_1), \dots, R_m(X_m)$ of bags over the schemas $X_1, \dots, X_m$ is globally consistent. The main structural result of this paper asserts that the acyclic hypergraphs are precisely the hypergraphs for which the local-to-global consistency property for bags holds.

THEOREM 3. *Let H be a hypergraph. The following statements are equivalent:*

(a) H is an acyclic hypergraph.

(b) H is a conformal and chordal hypergraph.

(c) H has the running intersection property.

(d) H has a join tree.

(e) H has the local-to-global consistency property for bags.

PROOF. (*Outline*) Let H be a hypergraph. By Theorem 1, statements (a), (b), (c), and (d) are equivalent, because these statements express "structural" properties of hypergraphs, i.e., they involve only the vertices and the hyperedges of the hypergraph at hand. So, we only have to show that statement (e), which involves "semantic" notions about bags, is equivalent to (one of) the other three statements. This will be achieved in two steps. First, we show that statement (c) implies statement (e), i.e., if H has the running intersection property, then H has the local-to-global consistency property for bags. Second, we show that statement (e) implies statement (b) by showing the contrapositive: if H is not conformal or H is not chordal, then H does not have the local-to-global consistency property for bags.

Step 1. If the hypergraph H has the running intersection property, then there is a listing $X_1, \dots, X_m$ of its hyperedges such that for every $i \in [m]$ with $i \geq 2$, there is a $j \in [i-1]$ such that $X_i \cap (X_1 \cup \cdots \cup X_{i-1}) \subseteq X_j$. Let $R_1(X_1), \dots, R_m(X_m)$ be a collection of pairwise consistent bags over the schemas $X_1, \dots, X_m$. By induction on $i = 1, \dots, m$, we show that there is a bag T_i over $X_1 \cup \cdots \cup X_i$ that witnesses the global consistency of the bags $R_1, \dots, R_i$. The claim is obvious for the base case $i = 1$. Assume that $i \geq 2$ and that the claim is true for all smaller indices. Let $X := X_1 \cup \cdots \cup X_{i-1}$ and, by the running intersection property, let $j \in [i-1]$ be such that $X_i \cap X \subseteq X_j$. By induction hypothesis, there is a bag T_{i-1} over X that witnesses the global consistency of $R_1, \dots, R_{i-1}$. We show that T_{i-1} and R_i are consistent by showing that $T_{i-1}[X \cap X_i] = R_i[X \cap X_i]$ and invoking Lemma 1. After this, we show that if T_i is a bag that witnesses the consistency of the bags T_{i-1} and R_i, then T_i witnesses the global consistency of $R_1, \dots, R_i$.

Step 2. We have to show that if if H is not conformal or H is not chordal, then H does not have the local-to-global consistency property for bags. We first establish that it is enough to show that certain "minimal" hypergraphs do not have the local-to-global consistency property for bags. Specifically, it is enough to show the following two statements:

1. No hypergraph $H_n = (V_n, \{V_n \setminus \{A_i\} : 1 \leq i \leq n\})$ with $V_n = \{A_1, \dots, A_n\}$ and $n \geq 3$ has the local-to-global consistency property for bags. Recall that H_n is not conformal.

2. No hypergraph $C_n = (V_n, \{\{A_i, A_{i+1}\} : i \in [n]\})$ with $V_n = \{A_1, \dots, A_n\}$, $A_{n+1} := A_1$, and $n \geq 4$ has the local-to-global consistency property for bags. Recall that C_n is not chordal.

The preceding "minimal" non-conformal and non-chordal hypergraphs share the following properties: all their hyperedges have the same number of vertices and all their vertices appear in the same number of hyperedges. Let $H^* = (V^*, E^*)$ be a hypergraph and let d and k be positive integers. The hypergraph H^* is called *k-uniform* if every hyperedge of H^* has exactly k vertices. It is called *d-regular* if every vertex of H^* appears in exactly d hyperedges of H. Thus, the "minimal" non-conformal hypergraph H_n is $(n-1)$-uniform and $(n-1)$-regular. Likewise, the "minimal" non-chordal hypergraph C_n is 2-uniform and 2-regular.

Assume that H^* is a k-uniform and d-regular hypergraph with $d \geq 2$ and with hyperedges $E^* = \{X_1, \dots, X_m\}$. We construct a collection $C(H^*) := \{R_1(X_1), \dots, R_m(X_m)\}$ of bags and show that the bags in this collection are pairwise consistent but are not globally consistent. This will imply that the local-to-global consistency property for bags fails for the hypergraphs H_n and C_n above.

For each $i \in [m]$ with $i \neq m$, let R_i be the bag over X_i defined as follows: (a) the support R'_i of R_i consists of all tuples $t : X_i \to \{0, \dots, d-1\}$ whose total sum $\sum_{C \in X_i} t(C)$ is congruent to 0 mod d; (b) $R_i(t) := 1$ for each such X_i-tuple, and $R_i(t) := 0$ for every other X_i-tuple.

For $i = m$, let R_m be the bag over X_m defined as follows: (a) the support R'_m of R_m consists of all tuples $t : X_m \to \{0, \dots, d-1\}$ whose total sum $\sum_{C \in X_m} t(C)$ is congruent to 1 mod d; (b) $R_m(t) := 1$ for each such X_m-tuple, and $R_m(t) := 0$ for every other X_m-tuple.

To show that the bags $R_1, \dots, R_m$ are pairwise consistent, it suffices (by Lemma 1) to show that for distinct $i, j \in [m]$, we have $R_i[Z] \equiv R_j[Z]$, where $Z := X_i \cap X_j$. In turn, this follows from the claim that for every $i \in [m]$ and every Z-tuple $t : Z \to \{0, \dots, d-1\}$, we have $R_i(t) = d^{k-|Z|-1}$. Indeed, since by k-uniformity every hyperedge of H has exactly k vertices, for every $u \in \{0, \dots, d-1\}$, there are exactly $d^{k-|Z|-1}$ many X_i-tuples $t_{i,u,1}, \dots, t_{i,u,d^{k-|Z|-1}}$ that extend t and have total sum congruent to u mod d. It follows then that $R_i[Z] = R_j[Z]$ for every two distinct $i, j \in [m]$, regardless of whether $m \in \{i, j\}$ or $m \notin \{i, j\}$.

To show that the relations $R_1, \dots, R_m$ are not globally consistent, we proceed by contradiction. If T were a bag that witnesses their consistency, then T would be non-empty and its support would contain a tuple t such that the projections $t[X_i]$ belong to the supports R'_i of the R_i, for each $i \in [m]$. In turn this means that

$$\textstyle\sum_{C \in X_i} t(C) \equiv 0 \bmod d, \qquad \text{for } i \neq m \tag{6}$$

$$\textstyle\sum_{C \in X_i} t(C) \equiv 1 \bmod d, \qquad \text{for } i = m. \tag{7}$$

Since by d-regularity each $C \in V$ belongs to exactly d many sets X_i, adding up all the equations in (6) and (7) gives

$$\textstyle\sum_{C \in V} dt(C) \equiv 1 \bmod d, \tag{8}$$

which is absurd since the left-hand side is congruent to 0 mod d and the right-hand side is congruent to 1 mod d. ☐

It should be pointed out that the proof of Theorem 1 in [9] has a different architecture than the proof of our Theorem 3. In particular, to prove the equivalence between the local-to-global consistency property for relations and acyclicity, Beeri et al. make use of Graham's algorithm, which is an algorithm for testing if a given hypergraph is acyclic. More importantly, for every cyclic hypergraph H, the proof of

Theorem 1 in [9] yields a collection of relations over H that are pairwise consistent but not globally consistent; these relations, however, are not pairwise consistent as bags, therefore they cannot be used to prove Theorem 3.

As an immediate consequence of Theorems 1 and 3, we obtain the following result.

COROLLARY 1. *Let H be a hypergraph. The following statements are equivalent:*

(a) H has the local-to-global consistency property for relations.

(b) H has the local-to-global consistency property for bags.

Complexity of Global Consistency for Bags The *global consistency problem for bags* asks: given a hypergraph $H = (V, \{X_1, \ldots, X_m\})$ and bags $R_1, \ldots, R_m$ over H, is the collection $R_1, \ldots, R_m$ globally consistent? Using an integral version of Carathéodory's Theorem due to Eisenbrand and Shmonin [13], we can show that this problem is in NP.

At the end of Section 2, we saw that for every fixed hypergraph H, the global consistency problem for relations over the hyperedges of H is solvable in polynomial time. As we shall see next, the state of affairs is by far more nuanced for bags. Every fixed hypergraph H gives rise to the decision problem $\mathrm{GCPB}(H)$, which asks: given bags $R_1, \ldots, R_m$ over H, is the collection $R_1, \ldots, R_m$ globally consistent? The next result is a dichotomy theorem that classifies the complexity of all decision problems $\mathrm{GCPB}(H)$, where H is a hypergraph.

THEOREM 4. *Let $H = (V, \{X_1, \ldots, X_m\})$ be a hypergraph. Then the following statements are true.*

1. *If H is acyclic, then* $\mathrm{GCPB}(H)$ *is in* P.
2. *If H is cyclic, then* $\mathrm{GCPB}(H)$ *is* NP*-complete.*

PROOF. (*Hint*) The first part of the theorem follows from Lemma 1 and Theorem 3. For the second part of the theorem, NP-hardness is proved via a series of reductions.

We first show the NP-hardness of each of the problems $\mathrm{GCPB}(C_n)$ and $\mathrm{GCPB}(H_n)$, where $n \geq 3$, as follows.

The problem $\mathrm{GCPB}(C_3)$ generalizes the consistency problem for 3-dimensional contingency tables (3DCT): given a positive integer n and, for each $i, j, k \in [n]$, non-negative integers $R(i,k)$, $C(j,k)$, $F(i,j)$, is there an $n \times n \times n$ table of non-negative integers $X(i,j,k)$ such that $\sum_{q=1}^{n} X(i,q,k) = R(i,k)$, $\sum_{q=1}^{n} X(q,j,k) = C(j,k)$, $\sum_{q=1}^{n} X(i,j,q) = F(i,j)$ for all indices $i, j, k \in [n]$? This problem was shown to be NP-complete in [17]. To see that $\mathrm{GCPB}(C_3)$ generalizes the consistency problem for 3DCT, let X, Y, Z be three attributes with domain $[n]$, and let $R(XZ)$, $C(YZ)$, $F(XY)$ be the bags given by the three tables $R(i,k)$, $C(j,k)$, $F(i,j)$. Therefore, $\mathrm{GCPB}(C_3)$ is NP-hard. For $n \geq 4$, we show that there is a polynomial time reduction from $\mathrm{GCPB}(C_{n-1})$ to $\mathrm{GCPB}(C_n)$. As for the problems $\mathrm{GCPB}(H_n)$, the problem $\mathrm{GCPB}(H_3)$ is NP-hard because $H_3 = C_3$; after this, we show that for every $n \geq 4$, there is a polynomial-time reduction from $\mathrm{GCPB}(H_{n-1})$ to $\mathrm{GCPB}(H_n)$.

Finally, if H is a cyclic hypergraph, then we show that there exists some $n \geq 4$ such that $\mathrm{GCPB}(C_n)$ or $\mathrm{GCPB}(H_n)$ reduces in polynomial time to $\mathrm{GCPB}(H)$. □

Table 1 compares the structural and algorithmic aspects of global consistency for relations vs. those for bags.

4. RELATIONS OVER SEMIRINGS

What do relations and bags have in common? For quite some time, it has been realized that relations and bags can be viewed as different instances of a single generalized concept of a relation in which tuples have "labels" that come from the domain of some algebraic structure.

Ioannidis and Ramakrishnan [16] considered relations over *labeled systems* and studied the query containment problem for relations over such systems. Later on Green, Karvounarakis, and Tannen [14] considered relations over semirings and studied the provenance of query answers. A *semiring* is an algebraic structure of the form $K = (A, +, \times, 0, 1)$ such that $(A, +, 0)$ is a commutative monoid, $(A, \times, 1)$ is a monoid, $\times$ distributes over $+$, and $a \times 0 = 0 \times a = 0$, for every $a \in A$. A semiring K is *positive* if the following two properties hold: (i) if $a + b = 0$, then $a = 0$ and $b = 0$; (ii) if $a \times b = 0$, then $a = 0$ or $b = 0$ (i.e., K has no *zero divisors*). If K is a semiring and X is a set of attributes, then a *K-relation over X* is a function $R : \mathrm{Tup}(X) \to A$. Thus, relations are $\mathbb{B}$-relations where $\mathbb{B} = (\{0,1\}, \vee, \wedge, 0, 1)$ is the Boolean semiring, while bags are $\mathbb{N}$-relations, where $\mathbb{N} = (\{0,1,2,\ldots\}, +, \times, 0, 1)$ is the semiring of non-negative integers with the standard arithmetic operations.

In the PODS 2021 proceedings version of the present paper [7], we raised the question of whether or not the results about the global consistency for bags extend to K-relations, where K is a positive semiring. In particular, does the analog of Theorem 3 for K-relations hold, where K is an arbitrary positive semiring? If not, are there broad classes of semirings for which the analog of Theorem 3 for K-relations holds? Since that time, we have obtained fairly complete answers to these questions that we summarize next; these results will appear in a forthcoming paper.

Our first finding asserts that if K is an arbitrary positive semiring and H is a hypergraph such that the local-to-global consistency property for K-relations holds, then H must be acyclic. Thus, one of the two directions in Theorem 3 holds for arbitrary positive semirings. Our second finding, however, reveals that the reverse direction does not hold for arbitrary positive semirings. For this, we consider the semiring $\mathbb{R}_1 = (\{0\} \cup [1, \infty], +, \times, 0, 1)$ of real numbers that are either 0 or at least 1 and the acyclic hypergraph

$$H = (\{A, B, C, D), \{\{A, D\}, \{B, D\}, \{C, D\}\}).$$

We show that there are three $\mathbb{R}_1$-relations T_1, T_2, T_3 over H that are pairwise consistent but not globally consistent.

According to Proposition 1 and to Lemma 1, both relations and bags have the following property: two relations $R(X)$, $S(Y)$ (or two bags $R(X)$, $S(Y)$) are consistent if and only if $R[X \cap Y] = S[X \cap Y]$. We say that a semiring K has the *inner consistency* property if the preceding property holds for all pairs of K-relations. Our third finding tells that if K is a positive semiring with the inner consistency property and if H is an acyclic hypergraph, then the local-to-global consistency property holds for H. Thus, for positive semirings with the inner consistency property, the acyclicity of a hypergraph H is equivalent to the local-to-global consistency property for H. This result provides a common generalization of Theorem 1 for relations and of Theorem 3 for bags.

Finally, we identify several different sufficient conditions for a semiring to have the inner consistency property. As a result, we establish that the equivalence between acyclic-

Table 1: Relational Consistency vs. Bag Consistency

	Relations	Bags
Witness of global consistency	Join is a witnesses	Join need *not* be a witness
Local-to-global consistency property for H	H is acyclic	H is acyclic
Global Consistency Problem for acyclic H	in P	in P
Global Consistency Problem for cyclic H	in P	NP-complete

ity and the local-to-global consistency property holds for a plethora of semirings, including the tropical semirings, the log semirings, Łukasiewicsz' semiring, and every semiring that is a bounded distributive lattice.

Acknowledgments

The research of Albert Atserias was partially supported by MICIN project PID2019-109137GB-C22 (PROOFS). The research of Phokion Kolaitis was partially supported by NSF Award No. 1814152.

5. REFERENCES

[1] S. Abramsky. Relational databases and Bell's theorem. In *In Search of Elegance in the Theory and Practice of Computation - Essays Dedicated to Peter Buneman*, volume 8000 of *Lecture Notes in Computer Science*, pages 13–35. Springer, 2013.

[2] S. Abramsky. Contextual semantics: From quantum mechanics to logic, databases, constraints, and complexity. *Bull. EATCS*, 113, 2014.

[3] S. Abramsky, R. S. Barbosa, K. Kishida, R. Lal, and S. Mansfield. Contextuality, cohomology and paradox. In *24th EACSL Annual Conference on Computer Science Logic, CSL 2015*, pages 211–228, 2015.

[4] S. Abramsky and A. Brandenburger. A unified sheaf-theoretic account of non-locality and contextuality. *CoRR*, abs/1102.0264, 2011.

[5] S. Abramsky, S. Mansfield, and R. S. Barbosa. The cohomology of non-locality and contextuality. In *Proc. 8th Intern. Workshop on Quantum Physics and Logic, QPL 2011*, volume 95 of *EPTCS*, pages 1–14, 2011.

[6] A. Atserias and P. G. Kolaitis. Consistency, acyclicity, and positive semirings. *CoRR*, abs/2009.09488, 2020. To appear in a volume dedicated to Samson Abramsky in the series *Outstanding Contributions to Logic*.

[7] A. Atserias and P. G. Kolaitis. Structure and complexity of bag consistency. In *PODS'21: Proc. of the 40th ACM Symposium on Principles of Database Systems*, pages 247–259. ACM, 2021.

[8] C. Beeri, P. A. Bernstein, and N. Goodman. A sophisticate's introduction to database normalization theory. In *Fourth Intern. Conf. on Very Large Data Bases*, pages 113–124. IEEE Computer Society, 1978.

[9] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the desirability of acyclic database schemes. *J. ACM*, 30(3):479–513, July 1983.

[10] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[11] A. K. Chandra and P. M. Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proc. of the 9th Annual ACM Symp. on Theory of Computing*, pages 77–90, 1977.

[12] S. Chaudhuri and M. Y. Vardi. Optimization of *Real* conjunctive queries. In *Proc. of the Twelfth ACM Symposium on Principles of Database Systems*, pages 59–70, 1993.

[13] F. Eisenbrand and G. Shmonin. Carathéodory bounds for integer cones. *Operations Research Letters*, 34(5):564 – 568, 2006.

[14] T. J. Green, G. Karvounarakis, and V. Tannen. Provenance semirings. In *Proc. of the Twenty-Sixth ACM Symposium on Principles of Database Systems*, pages 31–40, 2007.

[15] P. Honeyman, R. E. Ladner, and M. Yannakakis. Testing the universal instance assumption. *Inf. Process. Lett.*, 10(1):14–19, 1980.

[16] Y. E. Ioannidis and R. Ramakrishnan. Containment of conjunctive queries: Beyond relations as sets. *ACM Trans. Database Syst.*, 20(3):288–324, 1995.

[17] R. W. Irving and M. Jerrum. Three-dimensional statistical data security problems. *SIAM J. Comput.*, 23(1):170–184, 1994.

[18] W. Kent. Consequences of assuming a universal relation. *ACM Trans. Database Syst.*, 6(4):539–556, 1981.

[19] M. A. Khamis, P. G. Kolaitis, H. Q. Ngo, and D. Suciu. Bag query containment and information theory. In *Proc. of the 39th ACM Symposium on Principles of Database Systems, PODS 2020*, pages 95–112, 2020.

[20] G. Konstantinidis and F. Mogavero. Attacking Diophantus: Solving a special case of bag containment. In *Proc. of the 38th ACM Symposium on Principles of Database Systems, PODS*, pages 399–413, 2019.

[21] J. A. D. Loera and S. Onn. The complexity of three-way statistical tables. *SIAM J. Comput.*, 33(4):819–836, 2004.

[22] D. Maier, J. D. Ullman, and M. Y. Vardi. On the foundations of the universal relation model. *ACM Trans. Database Syst.*, 9(2):283–308, 1984.

[23] G. S. Tseitin. On the complexity of derivation in propositional calculus. *Structures in Constructive Mathematics and Mathematical Logic*, pages 115–125, 1968.

[24] J. D. Ullman. The U.R. strikes back. In *Proc. of the ACM Symposium on Principles of Database Systems*, pages 10–22, 1982.

[25] N. N. Vorob'ev. Consistent families of measures and their extensions. *Theory of Probability & Its Applications*, 7(2):147–163, 1962.

[26] H. S. Wilf. *Algorithms and complexity (2. ed.)*. A K Peters, 2002.

[27] M. Yannakakis. Perspectives on database theory. *SIGACT News*, 27(3):25–49, 1996.