Bag Query Containment and Information Theory

Mahmoud Abo Khamis
relationalAI

Phokion G. Kolaitis
UC Santa Cruz and IBM Research - Almaden

Hung Q. Ngo
relationalAI

Dan Suciu
University of Washington

ABSTRACT

The query containment problem is a fundamental algorithmic problem in data management. While this problem is well understood under set semantics, it is by far less understood under bag semantics. In particular, it is a long-standing open question whether or not the conjunctive query containment problem under bag semantics is decidable. We unveil tight connections between information theory and the conjunctive query containment under bag semantics. These connections are established using information inequalities, which are considered to be the laws of information theory. Our first main result asserts that deciding the validity of a generalization of information inequalities is many-one equivalent to the restricted case of conjunctive query containment in which the containing query is acyclic; thus, either both these problems are decidable or both are undecidable. Our second main result identifies a new decidable case of the conjunctive query containment problem under bag semantics. Specifically, we give an exponential time algorithm for conjunctive query containment under bag semantics, provided the containing query is chordal and admits a simple junction tree.

KEYWORDS

Query containment; bag semantics; information theory; entropy

ACM Reference Format:
Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. 2020. Bag Query Containment and Information Theory. In *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS'20), June 14–19, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3375395.3387645

1 INTRODUCTION

Since the early days of relational databases, the query containment problem has been recognized as a fundamental algorithmic problem in data management. This problem asks: given two queries Q_1 and Q_2, is it true that $Q_1(\mathcal{D}) \subseteq Q_2(\mathcal{D})$, for every database $\mathcal{D}$? Here, $Q_i(\mathcal{D})$ is the result of evaluating the query Q_i on the database $\mathcal{D}$. Thus, the query containment problem has several different variants, depending on whether the evaluation uses set semantics or bag semantics, and whether $\mathcal{D}$ is a set database or a bag database. Query containment under set semantics on set databases is the most extensively studied and well understood such variant. In particular, Chandra and Merlin [8] showed that, for this variant, the containment problem for conjunctive queries is NP-complete.

Chaudhuri and Vardi [9] were the first to raise the importance of studying the query containment problem under bag semantics. In particular, they raised the question of the decidability of the containment problem for conjunctive queries under bag semantics. There are two variants of this problem: in the *bag-bag* variant, the evaluation uses bag semantics and the input database is a bag, while in the *bag-set* variant, the evaluation uses bag semantics and the input database is a set. It is known that for conjunctive queries, the bag-bag variant and the bag-set variant are polynomial-time reducible to each other (see, e.g., [17]); in particular, either both variants are decidable or both are undecidable. Which of the two is the case, however, remains an outstanding open question to date.

During the past twenty five years, the research on the query containment problem under bag semantics has produced a number of results about extensions of conjunctive queries and also about restricted classes of conjunctive queries. Specifically, using different reductions from Hilbert's 10th Problem, it has been shown that the containment problem under bag semantics is undecidable for both the class of unions of conjunctive queries [16] and the class of conjunctive queries with inequalities[17]. It should be noted that, under set semantics, the containment problem for these two classes of queries is decidable; in fact, it is NP-complete for unions of conjunctive queries [27], and it is Π_2^{P}-complete for conjunctive queries with inequalities [20, 28]. As regards to restricted classes of conjunctive queries, several decidable cases of the bag-bag variant were identified in [2], including the case where both Q_1 and Q_2 are *projection-free* conjunctive queries, i.e., no variable is existentially quantified. Quite recently, this decidability result was extended to the case where Q_1 is a projection-free conjunctive query and Q_2 is an arbitrary conjunctive query [21]; the proof is via a reduction to a decidable class of Diophantine inequalities. In a different direction, information-theoretic methods were used in [22] to study the *homomorphism domination exponent* problem, which generalizes the conjunctive query containment problem under bag semantics on graphs. In particular, it was shown in [22] that the conjunctive query containment problem under bag semantics is decidable when Q_1 is a series-parallel graph and Q_2 is a chordal graph. This was the first time that notions and techniques from information theory were applied to the study of the containment problem under bag semantics.

Notions and techniques from information theory have found a number of uses in other areas of database theory. For example, entropy and mutual information have been used to characterize database dependencies [23, 24] and normal forms in relational and XML databases [3]. More recently, information inequalities were

PODS'20, June 14–19, 2020, Portland, OR, USA

ACM ISBN 978-1-4503-7108-7/20/06…$15.00
https://doi.org/10.1145/3375395.3387645

used with much success to obtain tight bounds on the size of the output of a query on a given database [4, 14, 15, 18, 19], *and* to devise query plans for worst-case optimal join algorithms [18, 19].

This paper unveils deeper connections between information theory and the query containment problem under bag semantics. These connections are established through the systematic use of information inequalities, which have been called the "laws of information theory" [26] as they express constraints on the entropy and thus "govern the impossibilities in information theory" [31].

An *information inequality* is an inequality of the form

$$0 \leq \sum_{X \subseteq V} c_X h(X), \tag{1}$$

where V is a set of n random variables over finite domains, each coefficient c_X is a real number, i.e. $c = (c_X)_{X \subseteq V}$ is a 2^n-dimensional real vector, h is the *entropy function* of a joint distribution over V (V*-distribution* henceforth). In particular, $h(X)$ denotes the marginal entropy of the variables in the set $X \subseteq V$.

An information inequality may hold for the entropy function of some V-distribution, but may not hold for all V-distributions. Following [5], we say that an information inequality is *valid* if it holds for the entropy function of *every* V-distribution. This notion gives rise to the following natural decision problem, which we denote as IIP: given *integer* coefficients $c_X \in \mathbb{Z}$ for all $X \subseteq V$, is the information inequality (1) valid?[1]

In this paper, we will also study a generalization of this problem that involves taking maxima of linear combinations of entropies. A *max-information inequality* is an expression of the form

$$0 \leq \max_{\ell \in [k]} \sum_{X \subseteq V} c_{\ell,X} h(X), \tag{2}$$

where $V, X, h(X)$ are as before, and for each $\ell \in [k]$, $c_\ell := (c_{\ell,X})_{X \subseteq V}$ is a 2^n-dimensional real vector. We say that a max information inequality is *valid* if it holds for the entropy function of every V-distribution. We write Max-IIP to denote the following decision problem: given k integer vectors c_ℓ of dimension 2^n, is the max information inequality (2) valid? Clearly, IIP is the special case of Max-IIP in which $k = 1$.

Our first main result asserts that Max-IIP is *many-one equivalent* to the restricted case of the conjunctive query containment problem under bag semantics in which Q_1 is an arbitrary conjunctive query and Q_2 is an acyclic conjunctive query. In fact, we show that these two problems are reducible to each other via exponential-time many-one reductions. This result establishes a *new* and *tight* connection between information theory and database theory, showing that Max-IIP and the conjunctive query containment problem under bag semantics with acyclic Q_2 are equally hard.

To the best of our knowledge, it is not known whether Max-IIP is decidable. In fact, even IIP is not known to be decidable; in other words, it is not known if there is an algorithm for telling whether a given information inequality with integer coefficients is valid. Even though the decidability question about IIP and about Max-IIP does not seem to have been raised explicitly by researchers in information theory, we note that there is a growing body of research aiming to "characterize" all valid information inequalities; moreover, finding such a "characterization" is regarded as a central problem in modern information theory (see, e.g., the survey [5]). It is reasonable to expect that a "good characterization" of valid information inequalities will also give an algorithmic criterion for the validity of information inequalities. Thus, showing that IIP is undecidable would imply that no "good characterization" of valid information inequalities exists.

Our second main result identifies a new decidable case of the conjunctive query containment problem under bag semantics. Specifically, we show that there is an exponential-time algorithm for testing whether Q_1 is contained in Q_2 under bag semantics, where Q_1 is an arbitrary conjunctive query and Q_2 is a conjunctive query that is *chordal* and admits a *junction tree* that is *simple*. Here, a query is chordal if its Gaifman graph G is chordal, i.e., G admits a tree decomposition whose bags induce (maximal) cliques of G; such a tree decomposition is called a *junction tree*. A tree decomposition is *simple* if every pair of adjacent bags in the tree decomposition share at most one common variable. The result follows from a new class of decidable Max-IIP problems. Note that this result is incomparable to the aforementioned decidability result about series-parallel and chordal graphs in [22], in two ways. First, the result in [22] applies only to graphs (i.e., databases with a single binary relation symbol), while our result applies to arbitrary relational schemas. Second, our result imposes more restrictions on Q_2, but no restrictions on Q_1.

The work reported here reveals that the conjunctive query containment problem under bag semantics is tightly intertwined with the validity problem for information inequalities. Thus, our work sheds new light on both these problems and, in particular, implies that any progress made in one of these problems will translate to similar progress in the other.

2 DEFINITIONS

We describe here the two problems whose connection forms the main result of this paper.

2.1 Query Containment Under Bag Semantics

Homomorphisms between relational structures. We fix a *relational vocabulary*, which is a tuple $\mathcal{R} = (R_1, \ldots, R_m)$, where each symbol R_i has an associated arity a_i. A *relational structure* is $\mathcal{A} = (A, R_1^A, \ldots, R_m^A)$, where A is a finite set (called domain) and each R_i^A is a relation of arity a_i over the domain A. Given two relational structures $\mathcal{A}$ and $\mathcal{B}$ with domains A and B respectively, a homomorphism from $\mathcal{B}$ to $\mathcal{A}$ is a function $f : B \to A$ such that for all i, we have $f(R_i^B) \subseteq R_i^A$. We write $\mathsf{hom}(\mathcal{B}, \mathcal{A})$ for the set of all homomorphisms from $\mathcal{B}$ to $\mathcal{A}$, and denote by $|\mathsf{hom}(\mathcal{B}, \mathcal{A})|$ its cardinality.

Bag-Set Semantics. A *conjunctive query* Q with variables $\mathsf{vars}(Q)$ and atom set $\mathsf{atoms}(Q) = \{A_1, \ldots, A_k\}$ is a conjunction:

$$Q(\mathbf{x}) = A_1 \wedge A_2 \wedge \cdots \wedge A_k. \tag{3}$$

For each $j \in [k]$, the atom A_j is of the form $R_{i_j}(\mathbf{x}_j)$, where $\mathsf{rel}(A_j) \stackrel{\text{def}}{=} R_{i_j}$ is a relation name, and $\mathsf{vars}(A_j) \stackrel{\text{def}}{=} \mathbf{x}_j$ is a function,

$$\mathsf{vars}(A_j) : [\mathsf{arity}(\mathsf{rel}(A_j))] \to \mathsf{vars}(Q) \tag{4}$$

[1]Equivalently, one can allow the input coefficients to be rational numbers.

associating a variable to each attribute position of $\text{rel}(A_j)$. We allow repeated variables in an atom. The variables $\mathbf{x}$ are called *head variables*, and must occur in the body.

A *database instance* is a structure $\mathcal{D}$ with domain D. The answer of a query (3) with head variables $\mathbf{x}$ is a set of $\mathbf{x}$-tuples with multiplicities. Formally, for each $\mathbf{d} \in D^{\mathbf{x}}$, denote $Q(\mathcal{D})[\mathbf{d}] \stackrel{\text{def}}{=} \{f \in \text{hom}(Q, \mathcal{D}) \mid f(\mathbf{x}) = \mathbf{d}\}$. The *answer to Q on $\mathcal{D}$ under the bag-set semantics* is the mapping $\mathbf{d} \mapsto |Q(\mathcal{D})[\mathbf{d}]|$. The bag-set semantics corresponds to a `count(*)-groupby` query in SQL.

Given two queries Q_1, Q_2 with the same number of head variables, we say that Q_1 *is contained in* Q_2 *under bag-set semantics*, and denote with $Q_1 \preceq Q_2$, if for every $\mathcal{D}$, we have $Q_1(\mathcal{D}) \leq Q_2(\mathcal{D})$, where $\leq$ compares functions point-wise, $\forall \mathbf{d}$, $|Q_1(\mathcal{D})[\mathbf{d}]| \leq |Q_2(\mathcal{D})[\mathbf{d}]|$.

Problem 2.1 (Query containment problem under bag-set semantics)**.** Given Q_1, Q_2, check whether $Q_1 \preceq Q_2$.

A query Q is called a *Boolean query* if it has no head variables, $|\mathbf{x}| = 0$. It is known that the query containment problem under bag semantics can be reduced to that of Boolean queries under bag semantics. For completeness, we provide the proof in Appendix A, and only mention here that the reduction preserves all special properties discussed later in this paper: acyclicity, chordality, simplicity. For that reason, in this paper we only consider Boolean queries, and denote Problem 2.1 by `BagCQC`.

Bag-bag Semantics. In our setting the input database $\mathcal{D}$ is a set, only the query's output is a bag. This semantics is known under the term *bag-set* semantics. Query containment has also been studied under the *bag-bag* semantics, where the database may also have duplicates. This problem is known to be reducible to the containment problem under bag-set semantics [17], by adding a new attribute to each relation, and for that reason we do not consider it further in this paper. One aspect of the bag-bag semantics is that repeated atoms change the meaning of the query, while repeated atoms can be eliminated under bag-set semantics. For example $R(x) \wedge R(x) \wedge S(x, y)$ and $R(x) \wedge S(x, y)$ are different queries under bag-bag semantics, but represent the same query under bag-set semantics. Since we restrict to bag-set semantics we assume no repeated atoms in the query.

The Domination Problem. We briefly review two related problems that are equivalent to `BagCQC`. Given two relational structures $\mathcal{A}$ and $\mathcal{B}$, we say that $\mathcal{B}$ *dominates* $\mathcal{A}$, and write $\mathcal{A} \preceq \mathcal{B}$, if $\forall \mathcal{D}$, $|\text{hom}(\mathcal{A}, \mathcal{D})| \leq |\text{hom}(\mathcal{B}, \mathcal{D})|$.

Problem 2.2 (The domination problem, `DOM`)**.** Given a vocabulary $\mathcal{R}$, and two structures $\mathcal{A}, \mathcal{B}$, check if $\mathcal{B}$ dominates $\mathcal{A}$: $\mathcal{A} \preceq \mathcal{B}$.

`DOM` and `BagCQC` are essentially the same problem. Kopparty and Rossman [22] considered the following generalization:

Problem 2.3 (The exponent-domination problem)**.** Given a rational number $c \geq 0$ and two structures $\mathcal{A}, \mathcal{B}$, check whether $|\text{hom}(\mathcal{A}, \mathcal{D})|^c \leq |\text{hom}(\mathcal{B}, \mathcal{D})|$ for all structures $\mathcal{D}$.

This problem is equivalent to `DOM`, because it can be reduced to `DOM` by observing that $|\text{hom}(n \cdot \mathcal{A}, \mathcal{D})| = |\text{hom}(\mathcal{A}, \mathcal{D})|^n$, where $n \cdot \mathcal{A}$ represents n disjoint copies of $\mathcal{A}$ [22, Lemma 2.2]. Conversely, `DOM` is the special case $c = 1$.

2.2 Information Inequality Problems

In this paper all logarithms are in base 2. For a random variable X with values that are in a finite domain D, its (binary) *entropy* is defined by

$$H(X) := -\sum_{x \in D} \Pr[X = x] \cdot \log \Pr[X = x] \tag{5}$$

Note that in the above definition, X can be a tuple of random variables, in which case $H(X)$ is their joint entropy. The entropy $H(X)$ is a non-negative real number.

Let $V = \{X_1, \ldots, X_n\}$ be a set of n random variables jointly distributed over finite domains. For each $\alpha \subseteq [n]$, the joint distribution induces a marginal distribution for the tuple of variables $X_\alpha = (X_i : i \in \alpha)$. One can also equivalently think of X_α as a vector-valued random variable. Either way, the marginal entropy on X_α is defined by (5) too, where we replace X by X_α. Define the function $h : 2^{[n]} \to \mathbb{R}_+$ as $h(\alpha) \stackrel{\text{def}}{=} H(X_\alpha)$, for all $\alpha \subseteq [n]$. We call h an *entropic function* (associated with the joint distribution on V) and identify it with a vector $h \in \mathbb{R}_+^{2^n}$.

The set of all entropic functions is denoted[2] by $\Gamma_n^* \subseteq \mathbb{R}_+^{2^n}$. With some abuse, we blur the distinction between the set $[n]$ and the set of variables $V = \{X_1, \ldots, X_n\}$, and write $h(X_\alpha)$ instead of $h(\alpha)$.

An *information inequality*, or `II`, defined by a vector $c = (c_X)_{X \subseteq V} \in \mathbb{R}^{2^V}$, is an inequality of the form

$$0 \leq \sum_{X \subseteq V} c_X h(X) \tag{6}$$

The information inequality is *valid* if it holds for all $h \in \Gamma_n^*$ [5].

Problem 2.4 (`II`-Problem)**.** Given a set V and a collection of integers c_X, for $X \subseteq V$, check whether the information inequality (6) is valid.

A *max-information inequality*, or `Max-II`, is defined by k vectors $\mathbf{c}_\ell := (c_{\ell,X})_{X \subseteq V} \in \mathbb{R}^{2^V}$, $\ell \in [k]$, and is written as:

$$0 \leq \max_{\ell \in [k]} \sum_{X \subseteq V} c_{\ell,X} h(X) \tag{7}$$

The `Max-II` is *valid* if it holds for all entropic functions $h \in \Gamma_n^*$.

Problem 2.5 (`Max-II` Problem)**.** Given a set V and integers $c_{\ell,X}$, for $\ell \in [k]$ and $X \subseteq V$, check whether the `Max-II` (7) is valid.

We denote the `II`- and `Max-II` problems by `IIP` and `Max-IIP` respectively. Both are co-recursively enumerable and it is open if any of them is decidable.

3 MAIN RESULTS

3.1 Connecting `BagCQC` to Information Theory

We state our first main result, and defer its proofs to Sec. 4 and 5. Recall that a *many-one reduction* of a decision problem A to another decision problem B, denoted by $A \leq_m B$, is a computable function f such that for every input X, the yes/no answer to problem A on X is the same as the yes/no answer to the problem B on $f(X)$. This is a special case of a Turing reduction, $A \leq_T B$, which means an algorithm that solves A given access to an oracle that solves B. Two

[2] Most texts drop the component $h(\emptyset)$, which is always 0, and define $\Gamma_n^* \subseteq \mathbb{R}_+^{2^n - 1}$. We prefer to keep the $\emptyset$-coordinate to simplify notations.

problems are *many-one equivalent*, denoted by $A \equiv_m B$, if $A \leq_m B$ and $B \leq_m A$.

Our main result is that the Max-IIP is many-one equivalent to the query containment problem under bag semantics, when the containing query is restricted to be acyclic. We briefly review acyclic queries here (we only consider α-acyclicity in this paper [11]):

Definition 3.1. A *tree decomposition* of a query Q is a pair (T, χ) where T is an undirected forest[3] and $\chi : \text{nodes}(T) \rightarrow 2^{\text{vars}(Q)}$ satisfies (a) the running intersection property: $\forall x \in \text{vars}(Q)$, $\{t \in \text{nodes}(T) \mid x \in \chi(t)\}$ is connected in T, and (b) the coverage property: for every $A \in \text{atoms}(Q)$, there exists $t \in \text{nodes}(T)$ s.t. $\text{vars}(A) \subseteq \chi(t)$. The sets $\chi(t)$ are called the *bags*[4] of the tree decomposition. A query Q is *acyclic* if there exists a tree decomposition (T, χ) such that, for all $t \in \text{nodes}(T)$, $\chi(t) = \text{vars}(A)$ for some $A \in \text{atoms}(Q)$.

Theorem 3.2. *Let* BagCQC-A *denote the* BagCQC *problem* $Q_1 \leq Q_2$*, where* Q_2 *is restricted to acyclic queries. Then* Max-IIP $\equiv_m$ BagCQC-A.

The proof of the theorem consists of three steps. First, we describe in Sec. 4.1 a Max-IIP inequality that is sufficient for containment, which is quite similar to, and inspired by an inequality by Kopparty and Rossman [22]. Second, we prove in Sec. 4.2 that, when Q_2 is acyclic, then this inequality is also necessary, thus solving the conjecture in [22, Sec.3]; our proof is based on Chan-Yeung's group-characterizable entropic functions [6, 7]. In particular, BagCQC-A $\leq_m$ Max-IIP. We do not know if this can be strengthened to BagCQC and/or IIP respectively. Finally, we give the many-one reduction Max-IIP $\leq_m$ BagCQC-A in Sec. 5.

3.2 Novel Decidable Class of BagCQC

Our next two results consist of a novel decidable class of query containment under bag semantics, and, correspondingly, a novel decidable class of max-information inequalities. We state here the results, and defer their proofs to Appendix D.

We show that containment is decidable when Q_2 is *chordal* and admits a *simple* junction tree (decomposition); to formally state the result, we define chordality, simplicity, and junction tree next.

A query Q is said to be *chordal* if its Gaifman graph G is chordal, i.e., there is a tree decomposition of G in which every bag induces a clique of G. A tree decomposition of G (and thus of Q) where all bags induce *maximal cliques* of G is called a *junction tree* in the graphical models literature (see Def. 2.1 in [29]).

Fix a tree decomposition of a query Q, and let $t \in \text{nodes}(T)$. A tree decomposition is called *simple* if $\forall (t_1, t_2) \in \text{edges}(T)$, $|\chi(t_1) \cap \chi(t_2)| \leq 1$, and is called *totally disconnected* if[5] $\forall (t_1, t_2) \in \text{edges}(T)$, $\chi(t_1) \cap \chi(t_2) = \emptyset$.

Note that every acyclic query is chordal, but not necessarily simple; for example, the query $Q() \leftarrow R(a, b, c), S(b, c, e)$ is a non-simple acyclic query. Conversely a chordal query is not necessarily acyclic; for example, any k-clique query with $k \geq 3$ is chordal.

[3] We allow Q to be disconnected, in which case T can be a forest, but we continue to call it a tree decomposition.

[4] Not to be confused with the bag semantics.

[5] Equivalently, $\text{edges}(T) = \emptyset$, because any edge s.t. $\chi(t_1) \cap \chi(t_2) = \emptyset$ can be removed.

Theorem 3.3. *Checking* $Q_1 \leq Q_2$ *is decidable in exponential time when* Q_2 *is chordal and admits a simple junction tree.*

Next, we complement Theorem 3.3 by showing that, if $Q_1 \not\leq Q_2$ then there exists a "witness" with a simple structure. This result is similar in spirit to other results where a decision problem can be restricted to special databases: for example, query containment under set semantics holds iff it holds on the canonical database of Q_1 [8], and implication between functional dependencies holds iff it holds on all relations with two tuples.

Let Q_1 be a query and $V = \text{vars}(Q_1)$. A relation $P \subseteq D^V$ is called a V*-relation*. A V-relation P and Q_1 *induce* a database instance $\Pi_{Q_1}(P) \stackrel{\text{def}}{=} (D, R_1^D, \ldots, R_m^D)$ where,

$$\forall \ell \in [m] : \qquad R_\ell^D \stackrel{\text{def}}{=} \bigcup_{A \in \text{atoms}(Q_1) : \text{rel}(A) = R_\ell} \Pi_{\text{vars}(A)}(P) \qquad (8)$$

In other words, we project P on each atom, and define R_ℓ^D as the union of projections on atoms with relation name R_ℓ.

The notation $\Pi_{\text{vars}(A)}(P)$ requires some explanation, because the atom A may have repeated variables, thus $\text{vars}(A)$ is a function (described in (4)). Given a set of integer indices Y and a function $\varphi : Y \rightarrow V$, the *generalized projection* is $\Pi_\varphi(P) \stackrel{\text{def}}{=} \{f \circ \varphi \mid f \in D^V\}$. A tuple $f \in D^V$ is a function $V \rightarrow D$, hence $f \circ \varphi$ just denotes function composition. For example, if $Q_1 = R(x, x, y)$ and $P = \{(a, b)\}$, then $R^D = \Pi_{(x,x,y)}(P) = \{(a, a, b)\}$. Obviously $P \subseteq \text{hom}(Q_1, \Pi_{Q_1}(P))$, which means $|P| \subseteq |\text{hom}(Q_1, \Pi_{Q_1}(P))|$, and this implies:

Fact 3.4 (Witness)**.** If there exists a $\text{vars}(Q_1)$-relation P such that $|P| > |\text{hom}(Q_2, \Pi_{Q_1}(P))|$, then $Q_1 \not\leq Q_2$, in which case P is said to be a *witness* (for the fact that $Q_1 \not\leq Q_2$).

We next define two special types of relations (and witnesses). Let W be a set of integer indices. Fix $\psi : W \rightarrow 2^V$ and a tuple $f \in D^V$. For any index $y \in W$, we view $f(\psi(y))$ as an atomic value in the domain $D^{\psi(y)}$. Define the W-tuple $\psi \cdot f \stackrel{\text{def}}{=} (f(\psi(y)))_{y \in W}$; its components may belong to different domains.

Definition 3.5 (Product and normal relations)**.** A V-relation P is a *product relation* if $P = \prod_{x \in V} S_x$, where each S_x is a unary relation. A W-relation is called a *normal relation* if it is of the form $\{\psi \cdot f \mid f \in P\}$ where P is some product V-relation and $\psi : W \rightarrow 2^V$ is some function.

One can verify that every product relation is a normal relation. For a simple illustration, consider the case when $V = \{X_1, X_2\}$. A product relation on V is $\{(u, v) \mid u, v \in [N]\} = [N] \times [N]$. A normal relation with four attributes is $\{(uv, u, v, v) \mid u, v \in [N]\}$, where uv denotes the concatenation of u and v. This normal relation corresponds to the map $\psi : [4] \rightarrow 2^V$ where $\psi(1) = \{X_1, X_2\}$, $\psi(2) = \{X_1\}$, and $\psi(3) = \psi(4) = \{X_2\}$. In a product relation all attributes are independent, while a normal relation may have dependencies: in our example the first attribute uv is a key, and the last two attributes are equal.

Theorem 3.6. *Let* Q_2 *be chordal,*

(i) *If* Q_2 *admits a totally disconnected junction tree, then* $Q_1 \not\leq Q_2$ *if and only if there is a product witness.*
(ii) *If* Q_2 *admits a simple junction tree, then* $Q_1 \not\leq Q_2$ *if and only if there exists a normal witness.*

We prove both theorems in Appendix D, using the novel results on information-theoretic inequalities described next, in Sec. 3.3.

Example 3.7. We illustrate with the following queries:

$Q_1 = A(x_1, x_2) \wedge B(x_1, x_2) \wedge C(x_1, x_2) \wedge A(x_1', x_2') \wedge B(x_1', x_2') \wedge C(x_1', x_2')$

$Q_2 = A(y_1, y_2) \wedge B(y_1, y_3) \wedge C(y_4, y_2)$

Q_2 is acyclic with a simple junction tree: $\{y_1, y_3\} - \{y_1, y_2\} - \{y_2, y_4\}$. We prove that $Q_1 \npreceq Q_2$ has a normal witness:

$$P \stackrel{\text{def}}{=} \{(u, u, v, v) \mid u \in [n], v \in [n]\} \subseteq D^{\{x_1, x_2, x_1', x_2'\}}$$

P induces the database $\Pi_{Q_1}(P) = ([n], A^D, B^D, C^D)$, where $A^D = B^D = C^D = \{(u, u) \mid u \in [n]\}$, and $|P| = n^2 > |\text{hom}(Q_2, \Pi_{Q_1}(P))| = n$ when $n > 1$, proving $Q_1 \npreceq Q_2$.

On the other hand, there is no product relation P that can witness $Q_1 \npreceq Q_2$. Indeed, if $P = S_1 \times S_2 \times S_3 \times S_4$ where $S_1, \ldots, S_4$ are unary relations, then the associated database $\Pi_{Q_1}(P)$ has relations $A^D = B^D = C^D \stackrel{\text{def}}{=} (S_1 \times S_2) \cup (S_3 \times S_4)$, and therefore $|\text{hom}(Q_2, \Pi_{Q_1}(P))| \geq \max(|S_1 \times S_2|^2, |S_3 \times S_4|^2) \geq |S_1 \times S_2 \times S_3 \times S_4| = |P|$.

3.3 Novel Class of Shannon-Inequalities

Our decidability results are based on a new result on information-theoretic inequalities, proving that certain max-linear inequalities are essentially Shannon inequalities. To present it, we need to review some known facts about entropic functions. We refer to Appendix B and to [30] for additional information. Recall that the set of entropic functions over n variables is denoted $\Gamma_n^* \subseteq \mathbb{R}^{2^n}$, and that we blur the distinction between a set V of n variables and $[n]$.

We begin by discussing closure properties of entropic functions and then introduce certain special classes of entropic functions. For the benefit of the readers familiar with database theory, we give in Table 1 the mapping between some of the database concepts used in this paper and their information-theoretic counterparts. For our discussion, it is useful to define the notion of the *entropy of a relation*. Given a V-relation P, its *entropy* is the entropy of the joint distribution on V, uniform on the support of P (i.e., tuples in P).

First, the sum of two entropic functions is also an entropic function, that is, if $h_1, h_2 \in \Gamma_n^*$, then $h_1 + h_2 \in \Gamma_n^*$. It follows that if k is a positive integer and h is an entropic function, then the function $h' = kh$ is also entropic. However, if $c > 0$ is a positive real number and h is an entropic function, then the function $h' = ch$ need not be entropic, in general. In contrast, the function $h' = ch$ is entropic, if $c > 0$ is a positive real number and h is a *step function*, defined as follows. Let $W \subsetneq V$ be a proper subset of V. The *step function at* W, denoted by h_W, is the function

$$h_W(X) = \begin{cases} 0 & \text{if } X \subseteq W \\ 1 & \text{otherwise.} \end{cases}$$

Every step function h_W is entropic. To see this, consider the relation $P_W = \{f_1, f_2\} \subseteq \{1, 2\}^V$, where $f_1 = (1, 1, \ldots, 1)$ and $f_2 = (\underbrace{2, \ldots, 2}_{V-W}, \underbrace{1, \ldots, 1}_{W})$, that is, f_2 has 1's on the positions W and 2's on all other positions. It is not hard to verify that h_W is the entropy of the relation P_W, and thus the step function h_W is indeed entropic.

As mentioned above, if $c > 0$ is a positive real number and h_W is a step function, then the function $h' = ch_W$ is entropic; the proof of this fact is given in Appendix B. A *normal entropic* function, or simply *normal* function, is a non-negative linear combination of step functions, i.e., $\sum_{W \subsetneq V} c_W h_W$, for $c_W \geq 0$. We write $\mathcal{N}_n$ to denote the set of all normal functions. Since, as mentioned earlier, the sum of two entropic functions is entropic, it follows that every normal function is entropic; thus, we have that $\mathcal{N}_n \subseteq \Gamma_n^*$. In Appendix B, we show that the normal functions are precisely the entropic functions with a non-negative I-measure (defined by Yeung [30]). The term "normal" was introduced in [18]. One can check that the entropy of every normal relation (Def. 3.5) is a normal function.

Example 3.8. The *parity function* is the entropy of the following relation with 3 variables: $P = \{(X, Y, Z) \mid X, Y, Z \in \{0, 1\}, X \oplus Y \oplus Z = 0\}$. More precisely, the entropy is $h(X) = h(Y) = h(Z) = 1$, $h(XY) = h(XZ) = h(YZ) = h(XYZ) = 2$. We show in Sec. 6 that h is not normal.

A function $h : 2^V \to \mathbb{R}_+$ is called *modular* if it satisfies $h(X \cup Y) + h(X \cap Y) = h(X) + h(Y)$ for all $X, Y \subseteq V$, and $h(\emptyset) = 0$. It is easy to show that h is modular iff $h(X_\alpha) = \sum_{i \in \alpha} h(X_i)$ for all $\alpha \subseteq V$. It is immediate to check that the entropy of any product relation (Def. 3.5) is modular. We write $\mathcal{M}_n$ to denote the set of all modular functions. Every modular function is normal, hence it is also entropic. To see this, given a modular function h, for each $i \leq n$, define $W_i = V \setminus \{X_i\}$ and let h_{W_i} be the associated step function at W_i. It is now easy to verify that $h = \sum_{i=1}^n h(X_i) \cdot h_{W_i}$, thus h is a normal function. In summary, we have $\mathcal{M}_n \subseteq \mathcal{N}_n \subseteq \Gamma_n^*$.

All entropic functions satisfy Shannon's *basic* inequalities, called *monotonicity* and *submodularity*,

$$h(X) \leq h(X \cup Y) \qquad h(X \cup Y) + h(X \cap Y) \leq h(X) + h(Y) \tag{9}$$

for all $X, Y \subseteq V$. (Since $h(\emptyset) = 0$, monotonicity implies *non-negativity* too.) A function $h : 2^V \to \mathbb{R}_+$, $h(\emptyset) = 0$, that satisfies Eq.(9) is called a *polymatroid*, and the set of all polymatroids is denoted by Γ_n. Thus, $\Gamma_n^* \subseteq \Gamma_n$. Zhang and Yeung [32] showed that Γ_n^* is properly contained in Γ_n, for every $n \geq 4$. Any inequality derived by taking a non-negative linear combination of inequalities (9) is called a *Shannon inequality*. In a follow-up paper [33], Zhang and Yeung gave the first example of a 4-variable valid information inequality which is non-Shannon.

In summary, we have considered the chain of the following four sets: $\mathcal{M}_n \subsetneq \mathcal{N}_n \subsetneq \Gamma_n^* \subsetneq \Gamma_n$. Except for Γ_n^*, each of these sets is a polyhedral cone. Using basic linear programming, one can show that it is decidable whether a max-linear inequality holds on a polyhedral set. In contrast, (even) the topological closure of Γ_n^* is not polyhedral [25]; in fact, it is conjectured to not even be semi-algebraic [13], and it is an open question whether linear inequalities or max-linear inequalities on $\overline{\Gamma}_n^*$ are decidable.

For a given vector $(c_X)_{X \subseteq V} \subseteq \mathbb{R}^{2^n}$ where $c_\emptyset = 0$, we associate a *linear expression* E which is the linear function $E(h) \stackrel{\text{def}}{=} \sum_{X \subseteq V} c_X h(X)$. As stated earlier, a linear inequality $E(h) \geq 0$ that is valid for all $h \in \Gamma_n^*$ is called an *information inequality*; furthermore, a *max information inequality* is one of the form $\max_\ell E_\ell(h) \geq 0$, where $\forall \ell, E_\ell$ is a linear expression.

In this paper, for any variable sets $X, Y \subseteq V$, we write $h(XY)$ as a shorthand for $h(X \cup Y)$, and define the *conditional entropy*

Database Theory	Information Theory
$P \subseteq D^V$ A *relation* P over a set of n variables V, each of which has domain D	$h \in \Gamma_n^*$ An *entropic function* $h : 2^V \to \mathbb{R}_+$ over a set of n variables V. h is defined by a uniform probability distribution p over P.
$P = S_1 \times \cdots \times S_n \subseteq D^V$ A *product relation* P (Definition 3.5)	$h(X) = \sum_{i \in X} h(i), \quad \text{forall } X \subseteq V$ A *modular function* $h \in \mathcal{M}_n$
The set of product relations	The set of modular functions $\mathcal{M}_n$
$P = P_1 \otimes P_2, \quad \text{where } P_1 \subseteq D_1^V, P_2 \subseteq D_2^V, P \subseteq (D_1 \times D_2)^V$ A *domain product* P of two relations P_1, P_2, all of which are over the same variable set V (Definition B.1)	$h = h_1 + h_2, \quad \text{where } h, h_1, h_2 \in \Gamma_n^*$ A *sum* h of two entropic functions h_1, h_2, all of which are over n variables
$P_W \stackrel{\text{def}}{=} \{f_1, f_2\} \subseteq D^V$, for some $W \subseteq V$, where $f_1 \stackrel{\text{def}}{=} (1, 1, \ldots, 1)$, $f_2 \stackrel{\text{def}}{=} (\underbrace{2, \ldots, 2}_{V-W}, \underbrace{1, \ldots, 1}_{W})$, Given $W \subseteq V$, the relation P_W has two tuples f_1, f_2 differing only in positions $V - W$. (See Section 3.3)	$h_W(X) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } X \subseteq W \\ 1 & \text{otherwise} \end{cases}$ Given $W \subseteq V$, a *step function* h_W.
$P = P_{W_1} \otimes P_{W_2} \otimes \cdots \otimes P_{W_m}$ A *normal relation* P over variable set V is a domain product of m (not necessarily distinct) relations P_{W_i} for $W_i \subseteq V$ (Another way to phrase Definition 3.5)	$h = \sum_{W \subseteq V} c_W h_W, \quad \text{where } c_W \geq 0$ A *normal entropy* $h \in \mathcal{N}_n$ is a non-negative weighted sum of step functions h_W
The set of normal relations	The set of normal functions $\mathcal{N}_n \equiv$ the cone closure of step functions
P_W, when $\lvert V - W \rvert = 1$, becomes a product relation	h_W, when $\lvert V - W \rvert = 1$, becomes a modular function
Product relations are a proper subclass of normal relations	Modular functions are a proper subclass of normal functions $\mathcal{M}_n \subsetneq \mathcal{N}_n$
A *group-characterizable* relation [6] $P \stackrel{\text{def}}{=} \{(aG_1, \ldots, aG_n) \mid a \in G\}$, where G is a group and $G_1, \ldots, G_n$ are subgroups	An entropic function $h \in \Gamma_n^*$
The set of group-characterizable relations	Γ_n^*
–	$\Gamma_n - \Gamma_n^*$ Polymatroids that are not entropic have no analog in databases

Table 1: Translation between the database world and the information theory world.

to be $h(Y|X) \stackrel{\text{def}}{=} h(XY) - h(X)$. Despite its name, the mapping $Y \mapsto h(Y|X)$ is not always an entropic function (Appendix B), but it is always a limit of entropic functions. The submodularity law (9) can be written using conditional entropies as

$$h(XY|X) \leq h(Y|X \cap Y) \tag{10}$$

We call the term $h(Y|X)$ *simple* if $|X| \leq 1$. A simple term $h(Y|X)$ is *unconditioned* if $X = \emptyset$. A *conditional linear expression* is a linear expression E of the form $E(h) = \sum_{X \subseteq Y \subseteq V} d_{Y|X} \cdot h(Y|X)$, where $d_{Y|X}$ are non-negative coefficients. A conditional linear expression is said to be *simple* (respectively, unconditioned) if $d_{Y|X} > 0$ implies $h(Y|X)$ is simple (respectively, unconditioned).

Let $\mathcal{I}$ be a class of max-linear inequalities. We say that $\mathcal{I}$ is *essentially Shannon* if, for every inequality I in $\mathcal{I}$, I holds for every $h \in \Gamma_n^*$ if and only if I holds for every $h \in \Gamma_n$. Any essentially Shannon class is decidable, because Γ_n is polyhedral.

Theorem 3.9. *Consider a max-linear inequality of the following form, where $q > 0$, and E_ℓ are conditional linear expressions:*

$$q \cdot h(V) \leq \max_{\ell \in [k]} E_\ell(h) \tag{11}$$

(i) *Suppose that E_ℓ is unconditioned, $\forall \ell \in [k]$; then inequality* (11) *holds $\forall h \in \mathcal{M}_n$ if and only if it holds $\forall h \in \Gamma_n$.*
(ii) *Suppose that E_ℓ is simple, $\forall \ell \in [k]$; then, inequality* (11) *holds $\forall h \in \mathcal{N}_n$ if and only if it holds $\forall h \in \Gamma_n$.*

In particular, the class of inequalities (11), where each E_ℓ is simple, is essentially Shannon and decidable.

The proof of the theorem follows from a technical lemma, which is of independent interest:

Lemma 3.10. *Let $h : 2^{[n]} \to \mathbb{R}_+$ be any polymatroid. Then there exists a normal polymatroid $h' \in \mathcal{N}_n$ with the following properties:*

(1) *$h'(X) \leq h(X)$, forall $X \subseteq [n]$,*
(2) *$h'([n]) = h([n])$,*
(3) *$h'(\{i\}) = h(\{i\})$, forall $i \in [n]$.*

In addition, there exists a modular function $h'' \in \mathcal{M}_n$ that satisfies conditions (1) and (2).

This lemma says that every polymatroid h can be decreased to become a normal polymatroid h', while preserving the values at $[n]$ (all variables) and at all singletons $\{i\}$. If we drop the last condition, then the existence of a modular function h'' follows from the modularization lemma [19], which is based on Lovasz's monotonization of submodular functions:

$$h''(X) \stackrel{\text{def}}{=} \sum_{i \in X} h(\{i\}|[i-1])$$

The proof that one can also satisfy condition (3), by relaxing from a modular function to a normal one, is non-trivial and given in Sec. 6.

Proof of Theorem 3.9. We prove the second item. Let $E(h) \stackrel{\text{def}}{=} \max_\ell E_\ell(h) - q \cdot h(V)$, where each E_ℓ has the form $\sum_i h(Y_i|X_i)$ with $|X_i| \leq 1$. Let $h \in \Gamma_n$, and let $h' \in \mathcal{N}_n$ be the normal polymatroid in Lemma 3.10. For every ℓ, we have $E_\ell(h') = \sum_i h'(X_iY_i) - \sum_i h'(X_i) \leq \sum_i h(X_iY_i) - \sum_i h(X_i) = E_\ell(h)$, because $|X_i| \leq 1$ and therefore $h'(X_i) = h(X_i)$. Since $E(h') \geq 0$, we obtain $q \cdot h(V) = q \cdot h'(V) \leq \max_\ell E_\ell(h') \leq \max_\ell E_\ell(h)$ completing the proof. The first item of the theorem is proven similarly, and omitted. □

Example 3.11. We illustrate here with an inequality needed later in Ex. 4.3. Consider $h(X_1X_2X_3) \leq \max(E_1, E_2, E_3)$, where:

$$E_1 = h(X_1X_2) + h(X_2|X_1)$$
$$E_2 = h(X_2X_3) + h(X_3|X_2)$$
$$E_3 = h(X_1X_3) + h(X_1|X_3)$$

(Notice that all three expressions are simple, hence the theorem applies.) Using Shannon's submodularity law (10), we infer $E_1 = h(X_1X_2) + h(X_2|X_1) \geq h(X_1X_2) + h(X_2|X_1X_3)$ and, similarly for E_2, E_3; therefore,

$$\max(E_1, E_2, E_3) \geq \frac{1}{3}[E_1 + E_2 + E_3] \geq \frac{1}{3}\Big[h(X_1X_2) + h(X_2|X_1X_3) + h(X_2X_3) + h(X_3|X_1X_2) + h(X_1X_3) + h(X_1|X_2X_3)\Big] = h(X_1X_2X_3)$$

4 REDUCING `BagCQC-A` TO `Max-IIP`

This section proves that `BagCQC-A` $\leq_m$ `Max-IIP`, showing half of the equivalence claimed in Theorem 3.2. We start by associating to each query containment problem a max-information inequality. We then prove, two results: the inequality is always a sufficient condition for containment, and it is also necessary when the containing query is acyclic. From now on, we will use only upper case to denote variables, both random variables and query variables.

Before we begin, we need to introduce some notations. Fix a relation $P \subseteq D^V$ and a probability distribution with mass function $p : P \to [0, 1]$. If $X \subseteq V$ is a set of variables, and $\varphi : Y \to V$ is a function, then recall that $\Pi_X(P)$ and $\Pi_\varphi(P)$ denote the standard, and the generalized projections respectively. We write $\Pi_X(p)$ for the standard X-marginal of p, and write $\Pi_\varphi(p)$ for the *φ-pullback*[6]. The latter is a probability distribution on $\Pi_\varphi(P)$ defined as follows. Start from the standard marginal $\Pi_{\varphi(Y)}(p)$ on $\Pi_{\varphi(Y)}(P)$, then apply the isomorphism $\Pi_\varphi(P) \to \Pi_{\varphi(Y)}(P)$ defined as $\Pi_\varphi(f) \mapsto \Pi_{\varphi(Y)}(f)$, $\forall f \in P$. Finally, if $E = \sum_i c_i h(Y_i)$ is a linear expression of entropic terms, where each $Y_i \subseteq Y$, then we denote by $E \circ \varphi \stackrel{\text{def}}{=} \sum_i c_i h(\varphi(Y_i))$ the result of applying the substitution φ to each term in E.

[6]This is a generalization of the pullback in [22, Sec.4].

Example 4.1. Let $V = \{X_1, X_2, X_3\}$, $P \subseteq D^V$, $\varphi(Y_1) = X_1, \varphi(Y_2) = \varphi(Y_3) = X_2$. The generalized projection is $\Pi_\varphi(P) = \{(a, b, b) \mid (a, b, c) \in P\} \subseteq D^{\{Y_1, Y_2, Y_3\}}$. Its tuples are in 1-1 correspondence with the standard projection $\Pi_{\varphi(Y)}(P) = \Pi_{X_1X_2}(P) = \{(a, b) \mid (a, b, c) \in P\}$. If p is a distribution on P, then the φ-pullback is $\Pi_\varphi(p)(Y_1Y_2Y_3 = abb) \stackrel{\text{def}}{=} p(X_1X_2 = ab) = \sum_c p(X_1X_2X_3 = abc)$. Notice that we do not need to define the pullback for (a, b, c) where $b \neq c$, because $(a, b, c) \notin \Pi_\varphi(P)$. Consider now the linear expression $E = 3h(Y_1) + 4h(Y_2Y_3) - 6h(Y_3)$. Then $E \circ \varphi = 3h(X_1) + 4h(X_2) - 6h(X_2) = 3h(X_1) - 2h(X_2)$.

We will introduce now a fundamental expression, E_T, that connects query containment to information inequalities; we discuss its history in Sec. 7. Fix a tree decomposition (T, χ) of some query Q, and recall that T may be a forest. Choose a root node in each connected component, thus giving an orientation of T's edges, where each node t has a unique $\text{parent}(t)$. We associate to T the following linear expression of entropic terms:

$$E_{(T,\chi)}(h) \stackrel{\text{def}}{=} \sum_{t \in \text{nodes}(T)} h(\chi(t)|\chi(t) \cap \chi(\text{parent}(t))) \tag{12}$$

where $\chi(\text{parent}(t)) = \emptyset$ when t is a root node. We abbreviate $E_{(T,\chi)}$ with E_T when χ is clear from the context. Expression (12) is independent of the choice of the root nodes, because one can check that $E_T = \sum_{t \in \text{nodes}(T)} h(\chi(t)) - \sum_{(t_1,t_2) \in \text{edges}(T)} h(\chi(t_1) \cap \chi(t_2))$.

4.1 A Sufficient Condition

Henceforth, let $\text{TD}(Q)$ denote the set of all tree decompositions of a given query Q.

Theorem 4.2. *Let Q_1, Q_2 be two conjunctive queries, $n = |\text{vars}(Q_1)|$. If the following* `Max-II` *inequality holds $\forall h \in \Gamma_n^*$:*

$$h(\text{vars}(Q_1)) \leq \max_{(T,\chi) \in \text{TD}(Q_2)} \max_{\varphi \in \text{hom}(Q_2, Q_1)} (E_T \circ \varphi)(h) \tag{13}$$

then $Q_1 \preceq Q_2$.

The theorem is inspired by, and is similar to Theorem 3.1 by Kopparty and Rossman [22], with three differences. First, the result in [22] applies only to graphs (i.e., databases with a single binary relation symbol), while our result applies to arbitrary relational schemas. Second, we do not restrict Q_2 to be chordal. Finally, [22] restrict h to entropies satisfying the independence constraints defined by Q_1; while this restriction is not needed to prove Theorem 4.2, it was needed in [22] to prove necessity in a special case (Theorem 3.3 in [22]). We will prove necessity in the next section without needing this restriction. Our proof is an extension of the proof in [22], and deferred to Appendix C. The proof of both Theorem 4.2 and 4.4 below use the following notation. Give a node $t \in \texttt{nodes}(T)$ of tree decomposition of Q, we denote by Q_t the "subquery at t", consisting of all atoms $A \in \text{atoms}(Q)$ s.t. $\text{vars}(A) \subseteq \chi(t)$. We can assume w.l.o.g. (Appendix A) that $\text{vars}(Q_t) = \chi(t)$. We end this section with an example, also from [22].

Example 4.3. This example is attributed to Eric Vee in [22]:

$$Q_1 = R(X_1, X_2) \wedge R(X_2, X_3) \wedge R(X_3, X_1) \quad Q_2 = R(Y_1, Y_2) \wedge R(Y_1, Y_3)$$

We show that $Q_1 \preceq Q_2$. Query Q_2 is acyclic, and its tree decomposition T is $\{Y_1, Y_2\} - \{Y_1, Y_3\}$, therefore:

$$E_T = h(Y_1Y_2) + h(Y_3|Y_1) = h(Y_1Y_2) + h(Y_1Y_3) - h(Y_1)$$

There are three homomorphisms $\varphi : Q_2 \to Q_1$, hence inequality (13) becomes:

$$h(X_1X_2X_3) \leq \max(E_1, E_2, E_3) \tag{14}$$

where E_1, E_2, E_3 are the linear expressions in Example 3.11, where we have shown that the inequality holds forall entropic h. Theorem 4.2 implies $Q_1 \preceq Q_2$. In lieu of a general proof, we prove the theorem on this particular example. Consider any database $\mathcal{D}$, let $P_1 = \mathsf{hom}(Q_1, \mathcal{D})$, p_1 the uniform probability space on P_1, and h_1 its entropy. Since h_1 satisfies inequality (14), one of the three terms on the right is larger than the left, assume w.l.o.g. that this term corresponds to the homomorphism $\varphi(Y_1) = X_1$, $\varphi(Y_2) = \varphi(X_3) = X_2$. Thus, $h_1(X_1X_2X_3) \leq h_1(X_1X_2) + h_1(X_2|X_1)$. Let $P_2 = \mathsf{hom}(Q_2, \mathcal{D})$. This is a relation with attributes Y_1, Y_2, Y_3. We define a probability distribution p_2 on P_2 as follows: the marginal $p_2(Y_1, Y_2)$ is the same as $p_1(X_1, X_2)$, and the conditional $p_2(Y_3|Y_1)$ is the same as $p_1(X_2|X_1)$. In particular, its entropy h_2 satisfies $\log|P_2| \geq h_2(Y_1Y_2Y_3) = h_2(Y_1Y_2) + h_2(Y_3|Y_1) = h_1(X_1X_2) + h_1(X_2|X_1) \geq h_1(X_1, X_2, X_3) = \log|P_1|$ proving $Q_1 \preceq Q_2$.

4.2 A Necessary Condition

Next we prove that inequality (13) is also a necessary condition for containment $Q_1 \preceq Q_2$, when Q_2 is acyclic. Our result answers positively the conjecture by Kopparty and Rossman [22, Sect.3, Discussion 1], in the case when Q_2 is acyclic. To prove the theorem, we consider some entropy h on which Eq.(13) fails, and prove that the support of its probability distribution, P, is a witness for $Q_1 \not\preceq Q_2$. The key idea is to use Chan-Yeung's group-characterizable entropic functions [6, 7], and show that P can be chosen to be "totally uniform". This allows us to relate $|\mathsf{hom}(Q_2, \mathcal{D})|$ to the right-hand-side of Eq.(13). More precisely, we prove the following.

Theorem 4.4. *Let Q_2 be acyclic. If there exists an entropic function h such that* (13) *does not hold, namely,*

$$h(\mathsf{vars}(Q_1)) > \max_{(T,\chi)\in\mathsf{TD}(Q_2)} \max_{\varphi\in\mathsf{hom}(Q_2,Q_1)} (E_T \circ \varphi)(h) \tag{15}$$

then there exists a database $\mathcal{D}$ such that $|\mathsf{hom}(Q_1, \mathcal{D})| > |\mathsf{hom}(Q_2, \mathcal{D})|$.

Together, Theorems 4.2 and 4.4 prove that `BagCQC-A` $\leq_m$ `Max-IIP`. To prove Theorem 4.4, we need some definitions and lemmas, where we fix a relation $P \subseteq D^V$, for some set of variables V, let $p : P \to [0, 1]$ be its uniform distribution ($p(f) \stackrel{\text{def}}{=} 1/|P|$, for all $f \in P$), and $h : 2^V \to \mathbb{R}_+$ its entropy.

Definition 4.5. We call P *totally uniform* if every marginal of p is also uniform.

For any two sets $X, Y \subseteq V$, and any tuple $f_0 \in \Pi_X(P)$, define the Y-degree of f_0 as $\deg_P(Y|X = f_0) \stackrel{\text{def}}{=} |\{\Pi_Y(f) \mid f \in P, \Pi_X(f) = f_0\}|$.

Lemma 4.6. *Let P be totally uniform. Then, for any two sets $X, Y \subseteq V$, the following hold:*

1. $\deg_P(Y|X = f_0)$ *is independent of the choice of f_0, and we denote it by* $\deg_P(Y|X)$.
2. $\deg_P(Y|X) = |\Pi_{XY}(P)|/|\Pi_X(P)|$ *and* $h(Y|X) = \log(\deg_P(Y|X))$.

Proof. Item 1 follows from the fact that the X-marginal of p is uniform and, therefore, $p(X = f_0) = \deg(Y|X = f_0)/|\Pi_{XY}(P)|$ is independent of f_0. For item 2, $|\Pi_{XY}(P)| = \sum_{f_0\in\Pi_X(P)} \deg_P(Y|X = f_0) = |\Pi_X(P)|\cdot\deg_P(Y|X)$, and $h(Y|X) = h(XY) - h(X) = \log|\Pi_{XY}(P)| - \log|\Pi_X(P)| = \log(\deg_P(Y|X))$. □

Lemma 4.7. *If $P_1 \subseteq D^X$, $P_2 \subseteq D^Y$ and P_2 is totally uniform, then* $|P_1 \bowtie P_2| \leq |P_1| \cdot \deg_{P_2}(Y|X \cap Y)$.

Proof. $|P_1 \bowtie P_2| \leq \sum_{f\in P_1} \deg_{P_2}(Y|X \cap Y = \Pi_{X\cap Y}(f)) = |P_1| \deg_{P_2}(Y|X \cap Y)$. □

Lemma 4.8. *Suppose the* `Max-II` $\max_{i=1,q} E_i(h) \geq 0$ *fails for some entropic function h. Then, for every $\Delta > 0$, there exists a totally uniform relation P such that its entropy h satisfies* $\max_{i=1,q} E_i(h) + \Delta < 0$. *In other words, we can find a totally uniform witness that fails the inequality with an arbitrary large gap Δ.*

Proof. We use the following result on group-characterizable entropic functions [7]. Fix a group G. For every subgroup $G_1 \subseteq G$, denote $aG_1 \stackrel{\text{def}}{=} \{ab \mid b \in G_1\}$. An entropic function $h \in \Gamma_n^*$ is called *group-characterizable* if there exists a group G and subgroups $G_1, \ldots, G_n$ such that h is the entropy of the uniform probability distribution on $P \stackrel{\text{def}}{=} \{(aG_1, \ldots, aG_n) \mid a \in G\}$. Chan and Yeung [7] proved that the set of group-characterizable entropic functions is dense in Γ_n^*; in other words, every $h \in \Gamma_n^*$ is the limit of group-characterizable entropic functions. In particular, if a max-linear inequality is valid for all group-characterizable entropic functions, then it is also valid for all entropic functions.

We show that, if $\max_i E_i(h) \geq 0$ fails, then it fails with a gap $> \Delta$ on a group-characterizable entropy. Let h_0 be any entropic function witnessing the failure: $\max_{i=1,q} E_i(h_0) < 0$. Choose any $\delta > 0$ s.t. $\max_{i=1,q} E_i(h_0) + \delta < 0$, and define $k \stackrel{\text{def}}{=} \lceil\Delta/\delta\rceil + 1$. Since $h \stackrel{\text{def}}{=} k\cdot h_0 = h_0 + h_0 + \cdots + h_0$ is also entropic and $E_i(k\cdot h_0) = k\cdot E_i(h_0)$ for all i, we have that $\max_{i=1,q} E_i(h) + k \cdot \delta < 0$, and therefore $\max_{i=1,q} E_i(h) + \Delta < 0$. By Chan-Yeung's density result, we can assume that h is group-characterizable.

Finally, we prove that the set P defining a group-characterizable entropy is totally uniform. This follows immediately from the fact that, under the uniform distribution, every tuple $(aG_1, \ldots, aG_n) \in P$ has probability $|G_1 \cap \cdots \cap G_n|/|G|$, and the marginal probability of any tuple $(aG_{i_1}, \ldots, aG_{i_k}) \in \Pi_{i_1\cdots i_k}(P)$ has probability $|G_{i_1} \cap \cdots \cap G_{i_k}|/|G|$. (See Theorem 1 from [6].) □

Proof of Theorem 4.4. Let (T, χ) be a junction tree (decomposition) of Q_2, which exists because acyclic queries are chordal. Then,

$$h(\mathsf{vars}(Q_1)) > \max_{(T',\chi)\in\mathsf{TD}(Q_2)} \max_{\varphi\in\mathsf{hom}(Q_2,Q_1)} (E_{T'} \circ \varphi)(h) \tag{16}$$

$$\geq \max_{\varphi\in\mathsf{hom}(Q_2,Q_1)} (E_T \circ \varphi)(h). \tag{17}$$

Fix Δ such that $\Delta > \log|\mathsf{hom}(Q_2, Q_1)|$, and let $P \subseteq D^{\mathsf{vars}(Q_1)}$ be the totally uniform relation given by Lemma 4.8, whose entropy h satisfies:

$$\log|P| = h(\mathsf{vars}(Q_1)) > \Delta + \max_{\varphi\in\mathsf{hom}(Q_2,Q_1)} (E_T \circ \varphi)(h) \tag{18}$$

P's columns are in 1-1 correspondence with $\text{vars}(Q_1) = \{X_1, \ldots, X_n\}$. We annotate each value with the column name, thus a tuple $f = (c_1, c_2, \ldots, c_n) \in P$ becomes $f = ((\text{"}X_1\text{"}, c_1), (\text{"}X_2\text{"}, c_2), \ldots, (\text{"}X_n\text{"}, c_n))$; the annotated P is isomorphic with the original P, hence still totally uniform. Let $\mathcal{D} = \Pi_{Q_1}(P)$ be the database obtained by projecting the annotated P on the atoms of Q_1 (Eq.(8)). We have seen that $|\text{hom}(Q_1, \Pi_{Q_1}(P))| \geq |P|$. We will show that $|P| > |\text{hom}(Q_2, \mathcal{D})|$, thus P is a witness for $Q_1 \not\leq Q_2$. To do this we need to upper bound $|\text{hom}(Q_2, \mathcal{D})|$.

Let $e : \mathcal{D} \to Q_1$ be the homomorphism mapping every value $(\text{"}X\text{"}, c)$ to the variable X: this is a homomorphism[7] because, by the definition of $\mathcal{D}$, Eq.(8), each tuple $f_0 = R_i((\text{"}X_{j_1}\text{"}, c_1), (\text{"}X_{j_2}\text{"}, c_2), \ldots)$ in $\mathcal{D}$ is the projection of some $f \in P$ on the variables $\text{vars}(A)$ of some $A \in \text{atoms}(Q_1)$; then e maps f_0 to A. If we view a tuple $f \in P$ as a function $\text{vars}(Q_1) \to D$, where D is the domain, then $e \circ f$ is the identity function on $\text{vars}(Q_1)$. Fix $\varphi \in \text{hom}(Q_2, Q_1)$ and denote:

$$\text{hom}_\varphi(Q_2, \mathcal{D}) \stackrel{\text{def}}{=} \{g \in \text{hom}(Q_2, \mathcal{D}) \mid e \circ g = \varphi\}$$

We have

$$\text{hom}(Q_2, \mathcal{D}) = \bigcup_{\varphi \in \text{hom}(Q_2, Q_1)} \text{hom}_\varphi(Q_2, \mathcal{D})$$

$$|\text{hom}(Q_2, \mathcal{D})| = \sum_{\varphi \in \text{hom}(Q_2, Q_1)} |\text{hom}_\varphi(Q_2, \mathcal{D})| \tag{19}$$

We will compute an upper bound for $|\text{hom}_\varphi(Q_2, \mathcal{D})|$, for each homomorphism φ. We claim:

$$\text{hom}_\varphi(Q_2, \mathcal{D}) \subseteq \Join_{t \in \text{nodes}(T)} \Pi_{\varphi|_{\chi(t)}}(P) \tag{20}$$

where $\varphi|_{\chi(t)}$ is the restriction of φ to $\chi(t)$, and $\Pi_{\varphi|_{\chi(t)}}(P)$ is the generalized projection (Sec. 3.2), i.e. it is a relation with attributes $\chi(t)$. The reason for partitioning $\text{hom}(Q_2, \mathcal{D})$ into subsets $\text{hom}_\varphi(Q_2, \mathcal{D})$ is so we can apply inequality (20) to each set: notice that the right-hand-side depends on φ. To prove the claim (20), we first observe:

$$\text{hom}_\varphi(Q_2, \mathcal{D}) \subseteq \Join_{t \in \text{nodes}(T)} \text{hom}_{\varphi|_{\chi(t)}}(Q_t, \mathcal{D}) \tag{21}$$

This is a standard property of any join decomposition (not necessarily acyclic): every tuple $g \in \text{hom}(Q_2, \mathcal{D})$ is the join of its fragments $\Pi_{\chi(t)}(g) \in \text{hom}(Q_t, \mathcal{D})$, as long as the fragments cover all attributes of g. Next we prove the following *locality property*:

$$\text{hom}_{\varphi|_{\chi(t)}}(Q_t, \mathcal{D}) \subseteq \Pi_{\varphi|_{\chi(t)}}(P) \tag{22}$$

It says that every answer of Q_t on $\mathcal{D}$ can be found in a single row of P. Here we use the fact that Q_2 is acyclic therefore there exists some $B \in \text{atoms}(Q_2)$ s.t. $\text{vars}(B) = \chi(t)$. Then, any homomorphism $g_0 \in \text{hom}_{\varphi|_{\chi(t)}}(Q_t, \mathcal{D})$ maps B to some tuple $f_0 \in \mathcal{D}$. By construction of $\mathcal{D}$, there exists some $A \in \text{atoms}(Q_1)$ such that $f_0 \in \Pi_{\text{vars}(A)}(P)$; in particular, $f_0 = \Pi_{\text{vars}(A)}(f)$ for some $f \in P$. Thus g_0, when viewed as a tuple over variables $\chi(t)$, can be found in a single row $f \in P$, more precisely[8] $g_0 = \Pi_\psi(f)$, from some function $\psi : \chi(t) \to \text{vars}(Q_1)$. Noticed that we have used in an essential way the fact that $\chi(t)$ is covered by a single atom B: we will need to remove this restriction later when we prove Theorem 3.3 (Lemma D.1 in Appendix D). From here it is immediate to show that $\psi = \varphi|_{\chi(t)}$, by composing with e: $\varphi|_{\chi(t)} = e \circ g_0 = e \circ f \circ \psi = \psi$ because $e \circ f$ is the identity on $\text{vars}(Q_1)$. This completes the proof of Eq.(22), which, together with Eq.(21), proves the claim Eq.(20).

Finally, we will upper bound the size of the join in (20), by applying repeatedly Lemma 4.7. This is possible because each projection $\Pi_{\varphi|_{\chi(t)}}(P)$ is totally uniform. Formally, fix an order of $\texttt{nodes}(T)$, $t_1, t_2, \ldots, t_m$, such that every child occurs after its parent, and compute the join (20) inductively, applying Lemma 4.7 to each step. If $S_i \stackrel{\text{def}}{=} \Join_{j=1,i} \Pi_{\varphi|_{\chi(t_j)}}(P)$, then the lemma implies $|S_i| = |S_{i-1} \Join \Pi_{\varphi|_{\chi(t_i)}}(P)| \leq |S_{i-1}| \deg_{\Pi_{\varphi|_{\chi(t_i)}}(P)}(\chi(t_i)|\chi(t_i) \cap \chi(\text{parent}(t_i)))|$, and this proves:

$$|\Join_{t \in \text{nodes}(T)} \Pi_{\varphi|_{\chi(t)}}(P)| \leq \prod_{i=1,m} \deg_{\Pi_{\varphi|_{\chi(t_i)}}(P)}(\chi(t_i)|\chi(t_i) \cap \chi(\text{parent}(t_i))) \tag{23}$$

Let $p' \stackrel{\text{def}}{=} \Pi_{\varphi|_{\chi(t_i)}}(p)$ be the $\varphi|_{\chi(t_i)}$-pullback of p. Its entropy satisfies $h'(Z) = h(\varphi(Z)) = (h \circ \varphi)(Z)$ for all $Z \subseteq \chi(t_i)$, implying $\log \deg_{\Pi_{\varphi|_{\chi(t_i)}}(P)}(Y|Z) = (h \circ \varphi)(Y|Z)$. This observation, together with (20) and (23) allow us to relate $\text{hom}(Q_2, \mathcal{D})$ to $(E_T \circ \varphi)(h)$:

$$\begin{aligned}
&\log |\text{hom}_\varphi(Q_2, \mathcal{D})| \\
&\leq \sum_{i=1,m} \log \deg_{\Pi_{\varphi|_{\chi(t_i)}}(P)}(\chi(t_i)|\chi(t_i) \cap \chi(\text{parent}(t_i))) \\
&= \sum_{i=1,m} (h \circ \varphi)((\chi(t_i)|\chi(t_i) \cap \chi(\text{parent}(t_i))) = (E_T \circ \varphi)(h) \\
&< h(\text{vars}(Q_1)) - \Delta = \log |P| - \Delta \qquad \text{By Eq.(18)}
\end{aligned}$$

Equivalently, $|\text{hom}_\varphi(Q_2, \mathcal{D})| < |P|/2^\Delta$. We sum up (19):

$$|\text{hom}(Q_2, \mathcal{D})| < |\text{hom}(Q_2, Q_1)| \frac{|P|}{2^\Delta} < |P|$$

completing the proof. □

We remark that inequality (15) is slightly stronger than necessary to prove containment. In the proof, we only need the inequality to hold for some junction tree. Conversely, Theorem 4.2 can also be stated such that we only consider non-redundant tree decompositions, of which junction trees are a special case.

5 REDUCING Max-IIP TO BagCQC-A

The results of the previous section imply BagCQC-A $\leq_m$ Max-IIP. We now prove the converse, Max-IIP $\leq_m$ BagCQC-A; in other words we show that Max-IIP can be reduced to the containment problem $Q_1 \leq Q_2$, with acyclic Q_2.

Theorem 5.1. Max-IIP $\leq_m$ BagCQC-A.

The proof has two parts. First, we convert the Max-IIP in Eq. (7) into a form that resembles Eq.(13), then we construct Q_1 and Q_2.

[7]For example, let $Q_1 = R(X, X), R(X, Y), S(X, Y)$ and let P have a single tuple (a, a). First annotate P to $((X, a), (Y, a))$. Then $R^D = \{((X, a), (X, a)), ((X, a), (Y, a))\}$, $S^D = \{((X, a), (Y, a))\}$. Without the annotation, these relations would be $R^D = S^D = \{(a, a)\}$, and there is no homomorphsims to Q, since the tuple in S^D cannot be mapped anywhere.

[8]We include here the rigorous, but rather tedious argument. Since g_0 is a homomorphism it "maps" the atom B to the tuple f_0, meaning $(g_0 \circ \text{vars}(B)) = f_0 = (f \circ \text{vars}(A))$ (all are functions $[\text{arity}(B)] \to D$, where D is the domain). Since $\text{vars}(B) : [\text{arity}(B)] \to \chi(t)$ is surjective, it has a right inverse, which implies $g_0 = f \circ \psi$ for some ψ.

Example 5.2. We will illustrate the main idea on an IIP rather than a Max-IIP. Consider[9]:

$$0 \le h(X_1) + 2h(X_2) + h(X_3) - h(X_1X_2) - h(X_2X_3) \quad (24)$$

We start by rewriting the inequality as:

$$3h(X_1X_2X_3) \le h(X_1) + h(X_2) + h(X_2) + h(X_3) + h(X_1X_2X_3) + h(X_3|X_1X_2) + h(X_1|X_2X_3) \quad (25)$$

From the right-hand-side we derive two queries Q_1, Q_2. Query Q_1 has 9 variables, $X_i^{(\ell)}$, $i = 1, 3$, $\ell = 1, 3$, while Q_2 has 13 variables:

$$\begin{aligned}
& Q_1 = Q_1^{(1)} \wedge Q_1^{(2)} \wedge Q_1^{(3)} \\
\ell = 1,3: \quad & Q_1^{(\ell)} = S_1(X_1^{(\ell)}) \wedge S_2(X_2^{(\ell)}) \wedge S_3(X_2^{(\ell)}) \wedge S_4(X_3^{(\ell)}) \\
& \wedge R_1(X_1^{(\ell)}, X_2^{(\ell)}, X_3^{(\ell)}) \wedge R_2(X_1^{(\ell)}, X_2^{(\ell)}, X_1^{(\ell)}, X_2^{(\ell)}, X_3^{(\ell)}) \\
& \wedge R_3(X_2^{(\ell)}, X_3^{(\ell)}, X_1^{(\ell)}, X_2^{(\ell)}, X_3^{(\ell)}) \\
& Q_2 = S_1(U_1) \wedge S_2(U_2) \wedge S_3(U_3) \wedge S_4(U_4) \\
& \wedge R_1(Y_1^0, Y_2^0, Y_3^0) \wedge R_2(Y_1^0, Y_2^0, Y_1^1, Y_2^1, Y_3^1) \\
& \wedge R_3(Y_2^1, Y_3^1, Y_1^2, Y_2^2, Y_3^2)
\end{aligned}$$

We apply Eq.(13) to Q_1, Q_2. $\text{TD}(Q_2)$ has a single tree because Q_2 is acyclic (see the comment at the end of Sec. 4.1). Q_1 has 3 connected components, and Q_2 has 5, therefore there are 3^5 homomorphisms $Q_2 \to Q_1$. Eq.(13) becomes:

$$\begin{aligned}
& h(X_1^{(1)}X_2^{(1)}X_3^{(1)}X_1^{(2)}X_2^{(2)}X_3^{(2)}X_1^{(3)}X_2^{(3)}X_3^{(3)}) \le \\
& \max_{\ell_1,\ldots,\ell_5=1,3}(h(X_1^{(\ell_1)}) + h(X_2^{(\ell_2)}) + h(X_2^{(\ell_3)}) + h(X_3^{(\ell_4)}) + \\
& h(X_1^{(\ell_5)}X_2^{(\ell_5)}X_3^{(\ell_5)}) + h(X_3^{(\ell_5)}|X_1^{(\ell_5)}X_2^{(\ell_5)}) + h(X_1^{(\ell_5)}|X_2^{(\ell_5)}X_3^{(\ell_5)}))
\end{aligned}$$

We will prove in Lemma 5.4 that this Max-II is equivalent to the II in Eq.(25), completing the reduction. Our example only illustrated the reduction from IIP; Lemma 5.3 below addresses the challenges introduced by Max-IIP.

5.1 Max-IIP $\le_m$ Uniform-Max-IIP

Consider a general Max-IIP (Eq.(7)), which we repeat here:

$$0 \le \max_{\ell \in [k]} E_\ell(h) \quad (26)$$

where $E_\ell(h) \stackrel{\text{def}}{=} \sum_{X \subseteq V} c_{\ell,X} h(X)$. In order to reduce it to a query containment problem, we start by making the expressions E_ℓ uniform. More precisely, for fixed natural numbers n, p, q, we say that an expression E is (n, p, q)-uniform if:

$$E(h) = n \cdot h(U) + \sum_{j=0,p} h(Y_j|X_j) - q \cdot h(V) \quad (27)$$

where V is the set of all variables, U is a single variable called the *distinguished variable*, and X_j, Y_j, for $j = 0, p$, are (not necessarily distinct) sets of variables, satisfying the following conditions:

Chain condition $X_0 = \emptyset$ and $X_j \subseteq Y_{j-1} \cap Y_j$ for $j = 1, p$.
Connectedness $U \in X_j$ for $j = 1, p$

[9] This IIP holds, but our goal is not to check it, but to reduce it to BagCQC-A.

A Uniform-Max-IIP is a Max-IIP, Eq.(26), such that there exist numbers n, p, q and a variable U s.t. all expressions E_ℓ in Eq.(26) are (n, p, q)-uniform, and have U as a distinguished variable. Notice that n, p, q and U are the same in all expressions E_ℓ. Clearly, a Uniform-Max-IIP is a special case of a Max-IIP. We prove:

Lemma 5.3. Max-IIP $\le_m$ Uniform-Max-IIP. *Moreover, the reduction can be done in polynomial time.*

Proof. Every E_ℓ in Eq.(26) has the form $\sum_{X \subseteq V} c_{\ell,X} h(X)$. By expanding each positive coefficient as $c_{\ell,X} = 1 + 1 + \cdots$ and each negative coefficient as $c_{\ell,X} = -1 - 1 - \cdots$, we can write:

$$E_\ell(h) = \sum_{i=1}^{m_\ell} h(Y_i) - \sum_{j=1}^{n_\ell} h(X_j) = \sum_{i=1}^{m_\ell} h(Y_i) + \sum_{j=1}^{n_\ell} h(V|X_j) - n_\ell \cdot h(V)$$

Define $X_0 \stackrel{\text{def}}{=} \emptyset$ and add $h(V|X_0) - h(V)$ $(= 0)$ to E_ℓ:

$$E_\ell(h) = \sum_{i=1}^{m_\ell} h(Y_i) + \sum_{j=0}^{n_\ell} h(V|X_j) - (n_\ell + 1) \cdot h(V) \quad (28)$$

The second sum is a chain, because $X_0 = \emptyset$ and every X_j is contained in V. Let $n \stackrel{\text{def}}{=} \max_\ell n_\ell$. We add $n - n_\ell$ terms $h(V) - h(V)$ to the expression E_ℓ, resulting in two changes to the expression (28): the term $-(n_\ell + 1) \cdot h(V)$ is replaced by $-(n + 1) \cdot h(V)$, and the sum $\sum_{i=1,m_\ell} h(Y_i)$ becomes $\sum_{i=1,m_\ell+n-n_\ell} h(Y_i)$ where the $n - n_\ell$ new terms are $Y_i \stackrel{\text{def}}{=} V$. We combine the two sums $\sum_i h(Y_i) + \sum_j h(V|X_j)$ into a single sum by writing $h(Y_i)$ as $h(Y_i|\emptyset)$, and thus E_ℓ becomes:

$$E_\ell(h) = \sum_{j=0}^{p_\ell} h(Y_j|X_j) - (n + 1) \cdot h(V) \quad (29)$$

Notice that Eq.(29) still satisfies the chain condition: $X_0 = \emptyset$, and $X_j \subseteq Y_{j-1} \cap Y_j$ for $j = 1, p_\ell$. Our next step is to enforce the connectedness condition.

Let U be a fresh variable. We will denote by h an entropic function over the variables V, and by h' an entropic function over the variables UV. For $\ell \in [k]$, denote by E'_ℓ the following expression:

$$E'_\ell(h') = (n + 1) \cdot h'(U) + \sum_{j=0}^{p_\ell} h'(UY_j|UX_j) - (n + 1) \cdot h'(UV) \quad (30)$$

We claim: $\forall h, 0 \le \max_\ell E_\ell(h)$ iff $\forall h', 0 \le \max_\ell E'_\ell(h')$. For the $\Leftarrow$ direction, assume $\forall h' : 0 \le \max_\ell E'_\ell(h')$ and let h be any entropic function over the variables V. We extended it to an entropic function h' over the variables UV, by defining U to be a constant random variable. In other words, $h'(X) \stackrel{\text{def}}{=} h(X - \{U\})$ forall $X \subseteq UV$; in particular $h'(U) = 0$. Then $E'_\ell(h') = E_\ell(h)$, forall $\ell \in [k]$, and the claim follows from $0 \le \max_\ell E'_\ell(h') = \max_\ell E_\ell(h)$. For the $\Rightarrow$ direction, let h' be any entropic function over the variables UV, and denote $h(-) \stackrel{\text{def}}{=} h'(-|U)$ the conditional entropy. The conditional entropy h is not necessarily entropic, but it is the limit of entropic functions (see Appendix B), hence it satisfies $0 \le \max_\ell E_\ell(h)$. Then, $E'_\ell(h') = \sum_{j=0}^{p_\ell} h'(UY_j|UX_j) - (n+1) \cdot h'(UV|U) = \sum_{j=0}^{p_\ell} h(Y_j|X_j) - (n+1) \cdot h(V) = E_\ell(h)$, and the claim follows from $0 \le \max_\ell E_\ell(h) = \max_\ell E'_\ell(h')$.

To enforce $X_0 = \emptyset$ in the chain condition, we write E'_ℓ as:

$$E'_\ell(h') = n \cdot h'(U) + \Big(h'(U) + \sum_{j=0}^{p_\ell} h'(UY_j|UX_j)\Big) - (n+1) \cdot h'(UV)$$

Finally, we need to ensure that all numbers p_ℓ are equal, and, for that, we set $p \stackrel{\text{def}}{=} 1 + \max_\ell p_\ell$ and add $p - p_\ell - 1$ terms $h'(U|U)$ to $E'_\ell(h')$. Comparing it with Eq.(27), the new E'_ℓ is an $(n, p, n+1)$-uniform expression, proving the lemma. □

5.2 A Technical Lemma

The `Uniform-Max-IIP` has some arbitrary q, while Eq.(13) has $q = 1$. We prove here a technical lemma showing that an (n, p, q)-uniform `Max-IIP` is equivalent to some `Uniform-Max-IIP` with $q = 1$. We do this by introducing new random variables.

Let V be a set of variables. For each variable $Z \in V$, we create q fresh copies $Z^{(\ell)}$, $\ell = 1 \dots q$, called adornments of Z. If X is a set of variables, then $X^{(\ell)}$ is the set where all variables are adorned with ℓ. We will denote by h an entropic function over the original variables V, and by h' an entropic function over the adorned variables $V^{(1)} \cdots V^{(q)}$. If $F = \sum_i c_i h'(X_i^{(\ell_i)})$ is a linear expression over adorned variables, then its erasure, $\epsilon(F) \stackrel{\text{def}}{=} \sum_i c_i h(X_i)$, is defined as the expression obtained by erasing every adornment; we also say that F is an adornment of $\epsilon(F)$. Conversely, if $E = \sum_i c_i h(X_i)$ is an expression over the original variables, then a constant adornment is an expression of the form $E^{(\ell)} = \sum_i c_i h'(X_i^{(\ell)})$, i.e. all terms are adorned by the same ℓ; clearly $\epsilon(E^{(\ell)}) = E$.

Lemma 5.4. *Let $E_1, \dots, E_k$ be linear expressions over variables V, and $F_1, \dots, F_m$ be linear expressions over adorned variables $V^{(1)}, \dots, V^{(q)}$ for some $q \geq 1$, such that (a) each F_j is an adornment of some E_i, i.e. $\epsilon(F_j) = E_i$, and (b) all constant adornments are included, i.e for every E_i and every ℓ there exists $F_j = E_i^{(\ell)}$. Then the following two statements are equivalent:*

$$\forall h: \; q \cdot h(V) \leq \max_{i \in [k]} E_i(h) \tag{31}$$

$$\forall h': \; h'(V^{(1)} \cdots V^{(q)}) \leq \max_{j=1,m} F_j(h') \tag{32}$$

Proof. (31) ⇒ (32) follows from:

$$\begin{aligned} h'(V^{(1)} \cdots V^{(q)}) &\leq \sum_{\ell=1,q} h'(V^{(\ell)}) && \leq q \max_{\ell=1,q} h'(V^{(\ell)}) \\ &\leq \max_{\ell=1,q} \max_{i \in [k]} E_i^{(\ell)}(h') && \text{Eq.(31) applied to } V^{(\ell)} \\ &\leq \max_{j=1,m} F_j(h') && \text{Assumption (b)} \end{aligned}$$

(32) ⇒ (31) Let h be an entropic function over variables V. That means that there exists a joint distribution over random variables V whose entropy is given by h. For each random variable Z, create q i.i.d. copies $Z^{(\ell)}$, for $\ell = 1, q$, and denote by h' the entropy function of the new random variables $V^{(1)}, \dots, V^{(q)}$. Thus, for any adorned set $X^{(\ell)}$, $h'(X^{(\ell)}) = h(X)$, and, if $E_i = \epsilon(F_j)$, then $E_i(h) = F_j(h')$. The claim follows from:

$$\begin{aligned} q \cdot h(V) &= h'(V^{(1)}) + \cdots + h'(V^{(q)}) && \text{By } h(V) = h'(V^{(\ell)}) \text{, forall } \ell \\ &= h'(V^{(1)} \cdots V^{(q)}) && \text{Independence} \\ &\leq \max_{j=1,m} F_j(h') && \text{Eq.(32)} \\ &\leq \max_{i \in [k]} E_i(h) && \text{Assumption (a)} \end{aligned}$$

□

5.3 `Uniform-Max-IIP` $\leq_m$ `BagCQC-A`

Given an (n, p, q)-uniform `Max-IIP` problem (31), $q \cdot h(V) \leq \max_i E_i$, where

$$E_i = n \cdot h(U) + \sum_{j=0,p} h(Y_{ij}|X_{ij}), \tag{33}$$

we will construct two queries Q_1, Q_2 such that $Q_1 \preceq Q_2$ iff condition (32) holds, which we have proven is equivalent to (31). Recall that the distinguished variable U occurs everywhere, except in the sets X_{i0} which, by definition, are $\emptyset$. We first substitute everywhere the single variable U with two variables, $U = U_1U_2$. This does not affect the `Max-IIP`, since we can simply treat U_1U_2 as a joint variable.

The query Q_2 will have one atom for each term of the expression E_i in (33), which is possible because, by uniformity, all expressions E_i have the same number of terms. In particular, there will be an atom R_j corresponding to the term $h(Y_{ij}|X_{ij})$, however, the number of variables Y_{ij} depends on i. For that reason, we consider their disjoint union, as follows. For each variable $V \in V$ and each i, j, let V^{ij} be a fresh copy of V; if $W = \{V_1, V_2, \dots\}$ is a set, then we denote by $W^{ij} \stackrel{\text{def}}{=} \{V_1^{ij}, V_2^{ij}, \dots\}$. We define $\tilde{Y}_j \stackrel{\text{def}}{=} \bigcup_{i \in [k]} Y_{ij}^{ij}$, for $j = 0, p$, and $\tilde{X}_j \stackrel{\text{def}}{=} \bigcup_{i \in [k]} X_{ij}^{i(j-1)}$, for $j = 1, p$, and $\tilde{X}_0 \stackrel{\text{def}}{=} \emptyset$. We notice that $|\tilde{Y}_j| = \sum_i |Y_{ij}|$, the sets $\tilde{Y}_0, \dots, \tilde{Y}_p$ are disjoint, and, since the chain condition $X_{ij} \subseteq Y_{i(j-1)}$ holds in (33), we also have $\tilde{X}_j \subseteq \tilde{Y}_{j-1}$; of course, $\tilde{X}_j$ is disjoint from $\tilde{Y}_j$. We define Q_2 as:

$$Q_2 = S_1(\tilde{U}_1) \wedge \cdots \wedge S_n(\tilde{U}_n) \wedge R_0(\tilde{X}_0\tilde{Y}_0\tilde{Z}) \wedge \cdots \wedge R_p(\tilde{X}_p\tilde{Y}_p\tilde{Z})$$

All relation symbols are distinct. The relations $S_1, \dots, S_n$ are binary, and $\tilde{U}_1, \dots, \tilde{U}_n$ are disjoint sets of two fresh variables each, and $\tilde{Z}$ is a fresh set of k variables. Thus, each relation R_j has arity $(\sum_i(|X_{ij}| + |Y_{ij}|)) + k$. All variables occurring in R_j are distinct (since $\tilde{X}_j \subseteq \tilde{Y}_{j-1}$, which is disjoint from $\tilde{Y}_j$) and they occur in the order that corresponds to the order $X_{1j} \dots X_{kj} Y_{1j} \dots Y_{kj}$ of the original variables, followed by the k variables $\tilde{Z}$. Any two consecutive atoms R_{j-1}, R_j share the variables $\tilde{X}_j$ and $\tilde{Z}$, and therefore the tree decomposition of Q_2 consists of n isolated components plus a chain:

$$T: \quad \{\tilde{U}_1\} \;\; \dots \;\; \{\tilde{U}_n\} \tag{34}$$

$$\{\tilde{X}_0, \tilde{Y}_0, \tilde{Z}\} \stackrel{\tilde{X}_1, \tilde{Z}}{-} \{\tilde{X}_1, \tilde{Y}_1, \tilde{Z}\} \stackrel{\tilde{X}_2, \tilde{Z}}{-} \{\tilde{X}_2\tilde{Y}_2, \tilde{Z}\} \cdots \stackrel{\tilde{X}_p, \tilde{Z}}{-} \{\tilde{X}_p, \tilde{Y}_p, \tilde{Z}\}$$

The query Q_1 consists of q isomorphic sub-queries:

$$Q_1 = Q_1^{(1)} \wedge \cdots \wedge Q_1^{(q)}$$

which have disjoint sets of variables. We describe now the subquery $Q_1^{(\ell)}$. Its variables consist of adorned copies $V^{(\ell)}$ of the variables V, and the query is in turn a conjunction of k sub-queries (which are no longer disjoint):

$$Q_1^{(\ell)} = Q_{1,1}^{(\ell)} \wedge \cdots \wedge Q_{1,k}^{(\ell)}$$

To define its atoms, we need some notations. Recall that the distinguished variables $U_1 U_2$ occur everywhere (except X_{i0} which is empty). Then, for every i, we define the the following sequences of variables:

$$\hat{X}_{ij}^{(\ell)} = \underbrace{U_1^{(\ell)} \cdots U_1^{(\ell)}}_{|X_{1j}|} \cdots \underbrace{X_{ij}^{(\ell)}}_{|X_{ij}|} \cdots \underbrace{U_1^{(\ell)} \cdots U_1^{(\ell)}}_{|X_{kj}|}$$

$$\hat{Y}_{ij}^{(\ell)} = \underbrace{U_1^{(\ell)} \cdots U_1^{(\ell)}}_{|Y_{1j}|} \cdots \underbrace{Y_{ij}^{(\ell)}}_{|Y_{ij}|} \cdots \underbrace{U_1^{(\ell)} \cdots U_1^{(\ell)}}_{|Y_{kj}|}$$

$$\hat{Z}_{i}^{(\ell)} = \underbrace{U_1^{(\ell)}}_{1} \cdots \underbrace{U_1^{(\ell)}}_{i-1} \underbrace{U_2^{(\ell)}}_{i} \underbrace{U_1^{(\ell)}}_{i+1} \cdots \underbrace{U_1^{(\ell)}}_{k}$$

That is, the length of $\hat{X}_{ij}^{(\ell)}$ is the same as that of the concatenation $X_{1j} X_{2j} \dots X_{kj}$, and has the distinguished variables $U_1^{(\ell)}$ on all positions except the positions of X_{ij}, where it has the adornment $X_{ij}^{(\ell)}$. (As a special case, $\hat{X}_{i0}^{(\ell)} = \emptyset$.) Note that the length of $\hat{X}_{ij}^{(\ell)}$ is independent of i, and $|\hat{X}_{ij}^{(\ell)}| = |\tilde{X}_j|$ (the variables from Q_2). Similarly for $\hat{Y}_{ij}^{(\ell)}$. The sequence $\hat{Z}_i$ has length k and contains $U_1^{(\ell)}$ everywhere except for position i where it has $U_2^{(\ell)}$. Then, query $Q_{1,i}^{(\ell)}$ is:

$$Q_{1,i}^{(\ell)} = S_1(U^{(\ell)}) \wedge \cdots \wedge S_n(U^{(\ell)}) \wedge$$
$$R_0(\hat{X}_{i0}^{(\ell)} \hat{Y}_{i0}^{(\ell)} \hat{Z}_i^{(\ell)}) \wedge R_1(\hat{X}_{i1}^{(\ell)} \hat{Y}_{i1}^{(\ell)} \hat{Z}_i^{(\ell)}) \wedge \cdots \wedge R_p(\hat{X}_{ip}^{(\ell)} \hat{Y}_{ip}^{(\ell)} \hat{Z}_i^{(\ell)})$$

Notice that the variables of the atom R_j are just $Y_{ij}^{(\ell)}$ (which contains $U_1^{(\ell)}, U_2^{(\ell)}$, and $X_{ij}^{(\ell)}$), and some variables are repeated several times.

We start by noticing that every homomorphism $\varphi : Q_2 \to Q_1$ must map all atoms in the chain $R_0 \cdots R_p$ to the same sub-query $Q_1^{(\ell)}$: this is because the chain is connected and, if one atom is mapped to an atom whose variables are adorned with ℓ, then all atoms must be mapped to atoms adorned similarly with ℓ. We claim something stronger, that φ maps the entire chain to the same sub-query $Q_{1,i}^{(\ell)}$. This is enforced by the variables $\tilde{Z}$ of Q_2: if one atoms is mapped to the sub-query $Q_{1,i}^{(\ell)}$, then $\varphi(\tilde{Z}_i) = U_2^{(\ell)}$ and $\varphi(\tilde{Z}_{i'}) = U_1^{(\ell)}$ forall $i' \neq i$, implying that all other atoms are mapped to the same sub-query.

By Theorems 4.2 and 4.4, we have:

$$Q_1 \preceq Q_2 \quad \text{iff} \quad \forall h', h'(\text{vars}(Q_1)) \leq \max_{\varphi \in \text{hom}(Q_2, Q_1)} (E_T \circ \varphi)(h') \quad (35)$$

We claim that the following are equivalent:

$$\forall h', h'(\text{vars}(Q_1)) \leq \max_{\varphi \in \text{hom}(Q_2, Q_1)} (E_T \circ \varphi)(h') \qquad \text{iff}$$
$$\forall h, q \cdot h(V) \leq \max_i E_i(h), \quad (36)$$

where E_i is given by (33). The claim implies the theorem: $Q_1 \preceq Q_2$ iff $\forall h, h(V) \leq \max_i E_i(h)$. To prove the claim, we will use Lemma 5.4, and, for that, we need to verify the conditions of the lemma. We start by applying the definition of E_T (Eq. (12)), where T is the tree decomposition of Q_2, Eq.(34), and obtain (recall that $\tilde{X}_0 = \emptyset$):

$$E_T = h(\tilde{U}_1) + \cdots + h(\tilde{U}_n) + h(\tilde{Y}_0 \tilde{Z}) + \sum_{j=1,p} h(\tilde{X}_j \tilde{Y}_j \tilde{Z} | \tilde{X}_j \tilde{Z})$$

Consider a homomorphism $\varphi \in \text{hom}(Q_2, Q_1)$. By the previous discussion, it maps all atoms in the chain to the same subquery $Q_{1,i}^{(\ell)}$ for some ℓ and i. We illustrate it by showing Q_2 and $\varphi(Q_2)$ next to each other:

$$Q_2 = S_1(\tilde{U}_1) \wedge \cdots \wedge S_n(\tilde{U}_n) \wedge R_0(\tilde{X}_0 \tilde{Y}_0 \tilde{Z}) \wedge \cdots \wedge R_p(\tilde{X}_p \tilde{Y}_p \tilde{Z})$$
$$\varphi(Q_2) = S_1(U^{(\ell_1)}) \wedge \cdots \wedge S_n(U^{(\ell_n)})$$
$$\wedge R_0(\hat{X}_{i0}^{(\ell)} \hat{Y}_{i0}^{(\ell)} \hat{Z}_i^{(\ell)}) \wedge \cdots \wedge R_p(\hat{X}_{ip}^{(\ell)} \hat{Y}_{ip}^{(\ell)} \hat{Z}_i^{(\ell)})$$

Next, we apply the substitution φ to E_T to obtain $E_T \circ \varphi$. Since each of the original expressions E_i in Eq.(33) was (n, p, q)-uniform, U occurs in every set Y_{ij} and X_{ij} (except for X_{i0}). By construction, $\hat{Z}_i^{(\ell)}$ is a sequence consisting only of the variables $U_1^{(\ell)}$ and $U_2^{(\ell)}$, thus the following set inclusions hold (except for $\hat{Z}_i^{(\ell)} \subseteq \hat{X}_{i0}^{(\ell)}$): $\hat{Z}_i^{(\ell)} \subseteq \hat{X}_{ij}^{(\ell)} \subseteq \hat{Y}_{ij}^{(\ell)}$, and we obtain:

$$E_T \circ \varphi = h(U^{(\ell_1)}) + \cdots + h(U^{(\ell_n)}) + h(\hat{Y}_{i0}^{(\ell)} \hat{Z}_i^{(\ell)}) +$$
$$\sum_{j=1,p} h(\hat{X}_{ij}^{(\ell)} \hat{Y}_{ij}^{(\ell)} \hat{Z}_i^{(\ell)} | \hat{X}_{ij}^{(\ell)} \hat{Z}_i^{(\ell)})$$
$$= h(U^{(\ell_1)}) + \cdots + h(U^{(\ell_n)}) + h(Y_{i0}^{(\ell)}) + \sum_{j=1,p} h(Y_{ij}^{(\ell)} | X_{ij}^{(\ell)})$$

Clearly its erasure is precisely $\epsilon(E_T \circ \varphi) = E_i$ from Eq. (33) (recall that $X_{i0} = \emptyset$), proving condition (a) of the lemma. Conversely, for each adornment $E_i^{(\ell)}$ there exists a homomorphism $\varphi : Q_2 \to Q_1$ such that $E_T \circ \varphi = E_i^{(\ell)}$, which proves condition (b), completing the proof of Th. 5.1.

6 PROOF OF LEMMA 3.10

Recall that we blurred the distinction between a set of n variables V and the set $[n]$. In this section we will use only $[n]$. Let $L \stackrel{\text{def}}{=} 2^{[n]}$ be the lattice of subsets of $[n]$. Given a function $h : L \to \mathbb{R}_+$, we define its dual $g : L \to \mathbb{R}_+$ as its Möbious inverse [18]:

$$\forall X : \quad h(X) = \sum_{Y : Y \supseteq X} g(Y), \quad g(X) = \sum_{Y : Y \supseteq X} (-1)^{|Y - X|} h(Y) \quad (37)$$

For any set $S \subseteq L$ we define:

$$g(S) \stackrel{\text{def}}{=} \sum_{X \in S} g(X) \quad (38)$$

Notice that $g(L) = h(\emptyset)$.

Fact 6.1. Let $h : L \to \mathbb{R}_+$ be any function. Then h is a polymatroid (i.e. $h \in \mathcal{N}_n$) iff its Möbius inverse g satisfies: $g(L) = 0$, $g([n]) \geq 0$ and $g(X) \leq 0$ forall $X \neq [n]$.

Proof. First we check that the Möbius inverse of a step function h_W satisfies the required properties, for $W \subsetneq V$:

$$h_W(X) = \begin{cases} 0 & \text{if } X \subseteq W \\ 1 & \text{otherwise} \end{cases} \qquad g_W(X) = \begin{cases} 1 & \text{if } X = V \\ -1 & \text{if } X = W \\ 0 & \text{otherwise} \end{cases}$$

The converse follows by observing that every g with the required properties is a non-negative linear combination of the g_W's: $g = \sum_{W \subsetneq [n]} (-g(W)) \cdot g_W$, therefore $h = \sum_{W \subsetneq [n]} (-g(W)) \cdot h_W$. □

Fact 6.1 can be used, for example, to show that the parity function h (Example 3.8) is not normal. Indeed, it's Möbius inverse given by Eq.(37) at $\emptyset$ is $g(\emptyset) = 1$, which implies that h is not normal. Fact. 6.1 will be our key ingredient to prove Lemma 3.10: in order to construct the required normal polymatroid h', we will instead construct its dual g' and check that it satisfies the conditions in Fact. 6.1. We also need a technical lemma:

Lemma 6.2. *Let $a_1, \ldots, a_n \geq 0$ be n non-negative numbers. Define:*

$$h(X) = \max\{a_i \mid i \in X\} \tag{39}$$

Then h is a normal polymatroid.

Proof. Assume w.l.o.g. $a_1 \leq a_2 \leq \cdots \leq a_n$ and define $\delta_i = a_{i+1} - a_i$ for $i = 0, 1, \ldots, n-1$, where $a_0 = 0$. Define $g : 2^{[n]} \to \mathbb{R}$:

$$g(X) \stackrel{\text{def}}{=} \begin{cases} a_n & \text{if } X = [n] \\ -\delta_i & \text{if } X = [i], (= \{1, 2, \ldots, i\}), \text{ for some } i < n \\ 0 & \text{otherwise} \end{cases}$$

We check that g is the dual of h by verifying:

$$h(X) = a_{\max(X)} = -\delta_{\max(X)} - \delta_{\max(X)+1} - \cdots - \delta_{n-1} + a_n = \sum_{Y: X \subseteq Y} g(Y)$$

We assumed above that $\max(\emptyset) = 0$. □

Finally, we need to recall the definitions of the *conditional entropy* and the *conditional mutual information*:

$$h(i|X) = h(\{i\} \cup X) - h(X)$$
$$I(i; j|X) = h(\{i\} \cup X) + h(\{j\} \cup Y) - h(X) - h(\{i, j\} \cup X) \tag{40}$$

and observe that, denoting $[X, Y] \stackrel{\text{def}}{=} \{Z \mid X \subseteq Z \subseteq Y\}$, we have:

$$h(X) = g([X, [n]]) \tag{41}$$
$$h(i|X) = -g([X, [n] - \{i\}]) \tag{42}$$
$$I(i; j|X) = -g([X, [n] - \{i, j\}]) \tag{43}$$

We can now prove Lemma 3.10, and will proceed by induction on n. Split the lattice $L = 2^{[n]}$ into two disjoint sets $L = L_1 \cup L_2$ where:

$$L_1 = [\emptyset, [n-1]] \qquad L_2 = [\{n\}, [n]]$$

In other words, L_1 contains all subsets without n, while L_2 contains all subsets that include n. Then:

- $g(L_2) = h(\{n\})$. It follows $g(L_1) = -h(\{n\})$.
- Subtract $h(\{n\})$ from $g([n])$ and add it to $g([n-1])$, and call g_1, g_2 the new functions on L_1, L_2 respectively. Formally:

$$g_1(X) = \begin{cases} g([n-1]) + h(\{n\}) & \text{if } X = [n-1] \\ g(X) & \text{if } X \subset [n-1] \end{cases}$$

$$g_2(X \cup \{n\}) = \begin{cases} g([n]) - h(\{n\}) & \text{if } X = [n-1] \\ g(X \cup \{n\}) & \text{if } X \subset [n-1] \end{cases}$$

  Notice that $g_1(L_1) = 0$ and $g_2(L_2) = 0$.

- One can check that the dual[10] of g_2 is the *conditional* polymatroid[11], defined as $h_2 : L_2 \to \mathbb{R}$:

$$\forall X \in L_2 : h_2(X) \stackrel{\text{def}}{=} h(X|\{n\})$$

- We apply induction to h_2 and obtain a normal polymatroid $h_2' : L_2 \to \mathbb{R}$ satisfying (1), (2), (3):

$$h_2'(X) \leq h_2(X) = h(X|\{n\})$$
$$h_2'([n]) = h_2([n]) = h([n]|\{n\})$$
$$h_2'(\{i, n\}) = h_2(\{i, n\}) = h(\{i\}|\{n\}) \qquad \text{since } \{i, n\} \text{ is an atom in } L_2$$

  Notice that $h_2'(\{n\}) = 0$, since $\{n\}$ is the bottom of L_2. Let g_2' be the dual of h_2', thus $g_2'(X) \leq 0$ forall $X \neq [n]$ (because h_2' is normal).
- One can check that the dual of g_1 is the function[12]

$$h_1(X) \stackrel{\text{def}}{=} I(X; \{n\})$$

  This is no longer a polymatroid. Instead, here we use Lemma 6.2 and define the normal polymatroid $h_1' : L_1 \to \mathbb{R}$:

$$h_1'(X) \stackrel{\text{def}}{=} \max_{i \in X} h_1(\{i\}) = \max_{i \in X} I(\{i\}; \{n\})$$

  Let $g_1' : L_1 \to \mathbb{R}$ be its dual. Thus, $g_1'(X) \leq 0$ forall $X \neq [n-1]$, and $g_1'([n-1]) = \max_{i \in [n-1]} I(\{i\}; \{n\})$.
- We combine g_1', g_2' into a single function $g' : L(= L_1 \cup L_2) \to \mathbb{R}$ as follows. g' agrees with g_1' on L_1 and with g_2' on L_2 except that we subtract a mass of $h(\{n\})$ from $g_1'([n-1])$ and add it to $g_2'([n])$. Formally:

$$g'(X) \stackrel{\text{def}}{=} \begin{cases} g_2'([n]) + h(\{n\}) & \text{if } X = [n] \\ g_1'([n-1]) - h(\{n\}) & \text{if } X = [n-1] \\ g_1'(X) & \text{if } X \in L_1, X \neq [n-1] \\ g_2'(X) & \text{if } X \in L_2, X \neq [n] \end{cases}$$

- We claim that for every $X \neq [n]$, $g'(X) \leq 0$. This is obvious for all cases above (since g_1', g_2' are normal), except when $X = [n-1]$. Here we check: $g'([n-1]) = g_1'([n-1]) - h(\{n\}) = \max_{i \in [n-1]} I(\{i\}; \{n\}) - h(\{n\}) \leq 0$ because $I(\{i\}; \{n\}) \leq h(\{n\})$.
- Denote $h' : L(= L_1 \cup L_2) \to \mathbb{R}$ the dual of g'; we have established that h' is a normal polymatroid. The following hold:

$$\forall x \in L_1 : h'(X) = \sum_{Y: X \subseteq Y \subseteq [n]} g'(Y)$$
$$= \sum_{Y: X \subseteq Y \subseteq [n-1]} g'(Y) + \sum_{Y: X \subseteq Y \subseteq [n-1]} g'(Y \cup \{n\})$$

[10] Strictly speaking we cannot talk about the dual of g_2 because we defined the dual only for functions $g : 2^{[m]} \to \mathbb{R}$. However, with some abuse, we identify the lattice L_2 with $2^{[n-1]}$, and in that sense the dual of $g_2 : L_2 \to \mathbb{R}$ is a function $h_2 : L_2 \to \mathbb{R}$.

[11] Proof: $h_2(X) = \sum_{Y: X \subseteq Y \subseteq [n]} g_2(Y) = \sum_{Y: X \subseteq Y \subseteq [n]} g(Y) - h(\{n\}) = h(X) - h(\{n\}) = h(X|\{n\})$.

[12] Proof:

$$h_1(X) = \sum_{Y: X \subseteq Y \subseteq [n-1]} g_1(Y) = h(\{n\}) + \sum_{Y: X \subseteq Y \subseteq [n-1]} g(Y)$$
$$= h(\{n\}) + \sum_{Y: X \subseteq Y \subseteq [n]} g(Y) - \sum_{Y: X \subseteq Y \subseteq [n-1]} g(Y \cup \{n\})$$
$$= h(\{n\}) + h(X) - h(X \cup \{n\}) = I(X; \{n\})$$

$$
\begin{aligned}
&= \sum_{Y:X\subseteq Y\subseteq[n-1]} g_1'(Y) + \sum_{Y:X\subseteq Y\subseteq[n-1]} g_2'(Y\cup\{n\}) \\
&= h_1'(X) + h_2'(X\cup\{n\}) \qquad (44)
\end{aligned}
$$

$$
\begin{aligned}
\forall X \in L_2 : h'(X) &= \sum_{Y:X\subseteq Y\subseteq[n]} g'(Y) \\
&= h(\{n\}) + \sum_{Y:X\subseteq Y\subseteq[n]} g_2'(Y) = h(\{n\}) + h_2'(X) \qquad (45)
\end{aligned}
$$

- We check that h' satisfies properties (1), (2), (3).

$$
\begin{aligned}
\forall X \in L_1 : h'(X) &= h_1'(X) + h_2'(X\cup\{n\}) && \text{by Eq.(44)}\\
&\le h_1(X) + h_2(X\cup\{n\})\\
&= I(X;\{n\}) + h(X|\{n\}) = h(X)\\
\forall X \in L_2 : h'(X) &= h(\{n\}) + h_2'(X) && \text{by Eq.(45)}\\
&\le h(\{n\}) + h_2(X)\\
&= h(\{n\}) + h(X|\{n\}) = h(X)\\
h'([n]) &= h(\{n\}) + h_2'([n]) && \text{by Eq.(45)}\\
&= h(\{n\}) + h_2([n])\\
&= h(\{n\}) + h([n]|\{n\}) = h([n])\\
\forall i \in [n-1] : h'(\{i\}) &= h_1'(\{i\}) + h_2'(\{i,n\}) && \text{by Eq.(44)}\\
&= h_1(\{i\}) + h_2(\{i,n\})\\
&= I(\{i\};\{n\}) + h(\{i\}|\{n\}) = h(\{i\})\\
h'(\{n\}) &= h(\{n\}) + h_2'(\{n\}) = h(\{n\}) + 0 && \text{by Eq.(45)}
\end{aligned}
$$

This completes the proof. We illustrate with an extensive example in the full version of the paper [1].

7 CONCLUSION AND DISCUSSION

In this paper we established a fundamental connection between information inequalities and query containment under bag semantics. In particular, we proved that the max-information-inequality problem is many-one equivalent to the query containment where the containing query is acyclic. It is open whether these problems are decidable. Our results help in the sense that, progress on one of these open questions will immediately carry over to the other. We end with a discussion of our results and a list of open problems.

Beyond Chordal Our results showed that the query containment problem $Q_1 \preceq Q_2$ is equivalent to a `Max-IIP` when Q_2 is either acyclic, or when it is chordal *and* has a simple junction tree. In all other cases, condition (13) is only sufficient, and we do not know if it is also necessary.

Repeated Variables, Unbounded Arities Our reduction form `Max-IIP` to query containment constructs two queries Q_1, Q_2 where the atoms have repeated variables, and the arities of some of the relation names depend on the size of the `Max-IIP`. We leave open the question whether the reduction can be strengthened to atoms without repeated variables, and/or queries over vocabularies of bounded arity.

Max-Linear Information Inequalities Linear information inequalities have been studied extensively in the literature, while Max-linear ones much less. Our result proves the equivalence of `BagCQC-A` and `Max-IIP`, and this raises the question of whether `IIP` and `Max-IIP` are different. In the full version of the paper [1], we provide some evidence suggesting that they might be the same.

The remarkable formula E_T (Eq.(12)) The first to introduce the expression E_T was Tony Lee [23]. This early paper established several fundamental connections between the entropy h of the uniform distribution of a relation P, and constraints on P: it showed that an FD $X \to Y$ holds iff $h(Y|X) = 0$, that an MVD $X \twoheadrightarrow Y$ holds iff $I(Y; V - (X \cup Y)|X) = 0$, and, finally, that P admits an acyclic join decomposition given by a tree T iff $E_T(h) = h(V)$. It also proved that E_T is equivalent to an inclusion-exclusion expression, which, in our notation becomes:

$$E_t = \sum_{S\subseteq \mathsf{nodes}(T)} (-1)^{|S|+1} CC(T\cap S)\cdot h(\chi(S)) \qquad (46)$$

where $\chi(S) \stackrel{\text{def}}{=} \bigcap_{t\in S}\chi(t)$, and $CC(T \cap S)$ denotes the number of connected components of the subgraph of T consisting of the nodes $\{t \mid t \in \mathsf{nodes}(T), \chi(t) \cap \bigcup_{t'\in S}\chi(t') \neq \emptyset\}$.

Discussion of Kopparty and Rossman [22] We now re-state the results in [22] using the notions introduced in this paper in order to describe their connection. Theorem 3.1 in [22] essentially states that Eq.(13) is sufficient for containment, thus it is a special case of our Theorem 4.2 for graph queries; they use an inclusion-exclusion formula for E_T, similar to (46), but given for chordal queries only. Theorem 3.2 in [22] essentially states that, if Eq.(13) fails on a normal polymatroid, then there exists a database $\mathcal{D}$ witnessing $Q_1 \not\preceq Q_2$, thus it is a special case of our Lemma D.1 for the case when the queries are graphs; they use a different expression for E_T, based on the Möbious inversion of h. This inversion is precisely the I-measure of h, as we explain in Appendix B. Finally, Theorem 3.3 in [22] proves essentially that Eq.(13) is necessary and sufficient when Q_1 is series-parallel and Q_2 is chordal. This differs from our Theorem 3.3 in that it imposes more restrictions on Q_1 and fewer on Q_2. The proof of our Theorem 3.3 relies on the fact that any counterexample of Eq.(13) is a normal entropic function, but this does not hold in the setting of Theorem 3.3 [22]; however, the only exception is given by the parity function (Appendix B), a case that [22] handles directly.

ACKNOWLEDGMENTS

Suciu was partially supported by NSF grants III-1703281, III-1614738, IIS-1907997, AitF-1535565. Kolaitis was partially supported by NSF Grant IIS-1814152.

REFERENCES

[1] Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. 2019. Bag Query Containment and Information Theory. *CoRR* abs/1906.09727 (2019).

[2] Foto N. Afrati, Matthew Damigos, and Manolis Gergatsoulis. 2010. Query containment under bag and bag-set semantics. *Inf. Process. Lett.* 110, 10 (2010), 360–369.

[3] Marcelo Arenas and Leonid Libkin. 2005. An information-theoretic approach to normal forms for relational and XML data. *J. ACM* 52, 2 (2005), 246–283.

[4] Albert Atserias, Martin Grohe, and Dániel Marx. 2013. Size Bounds and Query Plans for Relational Joins. *SIAM J. Comput.* 42, 4 (2013), 1737–1767.

[5] Terence Chan. 2011. Recent Progresses in Characterising Information Inequalities. *Entropy* 13, 2 (2011), 379–401.

[6] Terence H. Chan. 2007. Group characterizable entropy functions. In *IEEE International Symposium on Information Theory, ISIT 2007, Nice, France, June 24-29, 2007.* IEEE, 506–510.

[7] Terence H. Chan and Raymond W. Yeung. 2002. On a relation between information inequalities and group theory. *IEEE Transactions on Information Theory* 48, 7 (2002), 1992–1995.

[8] Ashok K. Chandra and Philip M. Merlin. 1977. Optimal Implementation of Conjunctive Queries in Relational Data Bases. In *ACM STOC*. 77–90.
[9] Surajit Chaudhuri and Moshe Y. Vardi. 1993. Optimization of *Real* Conjunctive Queries. In *ACM PODS, 1993*. 59–70.
[10] Ronald Fagin. 1982. Horn clauses and database dependencies. *J. ACM* 29, 4 (1982), 952–985.
[11] Ronald Fagin. 1983. Degrees of Acyclicity for Hypergraphs and Relational Database Schemes. *J. ACM* 30, 3 (1983), 514–550.
[12] Dan Geiger and Judea Pearl. 1993. Logical and Algorithmic Properties of Conditional Independence and Graphical Models. *The Annals of Statistics* 21, 4 (1993), 2001–2021.
[13] Arley Gomez, Carolina Mejía Corredor, and J. Andres Montoya. 2017. Defining the almost-entropic regions by algebraic inequalities. *IJICoT* 4, 1 (2017), 1–18.
[14] Georg Gottlob, Stephanie Tien Lee, Gregory Valiant, and Paul Valiant. 2012. Size and Treewidth Bounds for Conjunctive Queries. *J. ACM* 59, 3 (2012), 16:1–16:35.
[15] Martin Grohe and Dániel Marx. 2014. Constraint Solving via Fractional Edge Covers. *ACM Trans. Algorithms* 11, 1 (2014), 4:1–4:20.
[16] Yannis E. Ioannidis and Raghu Ramakrishnan. 1995. Containment of Conjunctive Queries: Beyond Relations as Sets. *ACM Trans. Database Syst.* 20, 3 (1995), 288–324.
[17] T. S. Jayram, Phokion G. Kolaitis, and Erik Vee. 2006. The containment problem for REAL conjunctive queries with inequalities. In *ACM PODS, 2006*. 80–89.
[18] Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu. 2016. Computing Join Queries with Functional Dependencies. In *ACM PODS, 2016*. 327–342.
[19] Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu. 2017. What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?. In *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017*. 429–444.
[20] Anthony C. Klug. 1988. On conjunctive queries containing inequalities. *J. ACM* 35, 1 (1988), 146–160.
[21] George Konstantinidis and Fabio Mogavero. 2019. Attacking Diophantus: Solving a Special Case of Bag Containment. In *ACM PODS, 2019*.
[22] Swastik Kopparty and Benjamin Rossman. 2011. The homomorphism domination exponent. *European Journal of Combinatorics* 32, 7 (2011), 1097 – 1114. Homomorphisms and Limits.
[23] Tony T. Lee. 1987. An Information-Theoretic Analysis of Relational Databases - Part I: Data Dependencies and Information Metric. *IEEE Trans. Software Eng.* 13, 10 (1987), 1049–1061.
[24] Tony T. Lee. 1987. An Information-Theoretic Analysis of Relational Databases - Part II: Information Structures of Database Schemas. *IEEE Trans. Software Eng.* 13, 10 (1987), 1061–1072.
[25] Frantisek Matús. 2007. Infinitely Many Information Inequalities. In *IEEE International Symposium on Information Theory, ISIT 2007, Nice, France, June 24-29, 2007*. 41–44.
[26] Nicholas Pippenger. 1986. What are the laws of information theory. In *1986 Special Problems on Communication and Computation Conference*. 3–5.
[27] Yehoshua Sagiv and Mihalis Yannakakis. 1980. Equivalences Among Relational Expressions with the Union and Difference Operators. *J. ACM* 27, 4 (1980), 633–655.
[28] Ron van der Meyden. 1997. The Complexity of Querying Indefinite Data about Linearly Ordered Domains. *J. Comput. Syst. Sci.* 54, 1 (1997), 113–135.
[29] Martin J. Wainwright and Michael I. Jordan. 2008. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning* 1, 1-2 (2008), 1–305. https://doi.org/10.1561/2200000001
[30] Raymond W. Yeung. 2008. *Information Theory and Network Coding* (1 ed.). Springer Publishing Company, Incorporated.
[31] Raymond W Yeung. 2012. *A first course in information theory.* Springer Science & Business Media.
[32] Zhen Zhang and Raymond W. Yeung. 1997. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Information Theory* 43, 6 (1997), 1982–1986.
[33] Zhen Zhang and Raymond W Yeung. 1998. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory* 44, 4 (1998), 1440–1452.

A BACKGROUND ON CQ'S

We prove the following in the full version of the paper [1]:

Lemma A.1. *The containment problem under bag-set semantics $Q_1 \preceq Q_2$ is reducible in polynomial time to the containment problem under bag-set semantics for Boolean queries, $Q_1' \preceq Q_2'$. Moreover, this reduction preserves any property of queries discussed in this paper: acyclicity, chordality, simplicity.*

We prove now a claim that we made in Sec 4.1, namely that, for any node t of a tree decomposition, we can assume $\text{vars}(Q_t) = \chi(t)$, where Q_t is the query obtained by taking the conjunction of all atoms with $\text{vars}(A) \subseteq \chi(t)$.

Fact A.2. (Informal) Let (T, χ) be a tree decomposition of some query Q, and, forall $t \in \texttt{nodes}(T)$, let Q_t denote the conjunction of $A \in \text{atoms}(Q)$ s.t. $\text{vars}(A) \subseteq \chi(t)$. Then, for the purpose of query containment, we can assume that $\text{vars}(Q_t) = \chi(t)$, for every $t \in \texttt{nodes}(T)$. More specifically, we can assume that for every $t \in \texttt{nodes}(T)$ and every $A \in \text{atoms}(Q)$ such that $\text{vars}(A) \cap \chi(t) \neq \emptyset$, there exists $A' \in \text{atoms}(Q)$ such that $\text{vars}(A') = \text{vars}(A) \cap \chi(t)$, hence $A' \in \text{atoms}(Q_t)$.

Proof. To see an example where this property fails, consider $Q = R(x, y, u) \wedge S(y, z) \wedge R(x, z, v)$. Let T be the tree decomposition $\{x, y, u\} - \{x, y, z\} - \{x, z, v\}$, and let t be the middle node, $\chi(t) = \{x, y, z\}$. Then $Q_t = S(y, z)$ and its variables do not cover $\chi(t)$.

We prove that the property can be satisfied w.l.o.g. We first modify the vocabulary, by adding for each relation name R of arity a and for each $S \subset [a]$, a new relation name R_S of arity $|S|$. Similarly, we modify a query Q by adding, for each atom $R(X_1, \ldots, X_a)$ and for each $S \subset [a]$, a new atom $R_S(x_S)$, where $x_S \stackrel{\text{def}}{=} (X_i)_{i \in S}$. Denote by $\hat{Q}$ the modified query. Obviously $\hat{Q}$ satisfies the desired property. We claim that this change does not affect query containment, more precisely $Q_1 \preceq Q_2 \Leftrightarrow \hat{Q}_1 \preceq \hat{Q}_2$. The $\Leftarrow$ direction follows by expanding an input database $\mathcal{D}$ for Q_1, Q_2 with extra predicates $R_S^D \stackrel{\text{def}}{=} \Pi_S(R^D)$ for every relation symbol R and every $S \subset [a]$ where a is the arity of R. The $\Rightarrow$ direction follows from modifying an input database $\mathcal{D}$ for $\hat{Q}_1, \hat{Q}_2$ by replacing every (a-ary) relation R^D by $R^D \ltimes \left(\Join_{S \subset [a]} R_S^D\right)$. □

B BACKGROUND ON INFORMATION THEORY

Next, we review some additional background in information theory used in this paper, continuing the brief introduction in Sec. 3.3. Before we start, we review a basic concept, which we call "domain-product", first introduced by Fagin [10] to prove the existence of an Armstrong relation for constraints defined by Horn clauses, and later used by Geiger and Pearl [12] to prove that Conditional Independence constraints on probability distributions also admit an Armstrong relation. The same construction appears under the name "fibered product" in [22].

Definition B.1. Fix two domains D_1, D_2. For any two tuples $f \in D_1^V$, $g \in D_2^V$ we define $f \otimes g \in (D_1 \times D_2)^V$ as the function $(f \otimes g)(x) \stackrel{\text{def}}{=} (f(x), g(x))$ for all $x \in V$. The *domain product* of two relations $P_1 \subseteq D_1^V$, $P_2 \subseteq D_2^V$ is $P_1 \otimes P_2 \stackrel{\text{def}}{=} \{f \otimes g \mid f \in P_1, g \in P_2\}$. If p_1, p_2 are probability distributions on P_1, P_2 respectively, then their *product* $p_1 \cdot p_2$ is the probability distribution $(p_1 \cdot p_2)(f, g) \stackrel{\text{def}}{=} p_1(f) \cdot p_2(g)$ on $P_1 \otimes P_2$.

We start with a basic fact: if h_1, h_2 are two entropic functions, then $h_1 + h_2$ is also entropic. Indeed, if h_i is the entropy of $p_i : P_i \to [0, 1]$, then $h_1 + h_2$ is the entropy of $p_1 \cdot p_2 : P_1 \otimes P_2 \to [0, 1]$, where $P_1 \otimes P_2$ is the domain product.

Fact B.2. If $n = 1$ (i.e. there is a single random variable) and h is entropic, then $c \cdot h$ is also entropic for every $c > 0$.

Proof. Start with a distribution p whose entropy is $\lceil c \rceil \cdot h$. Let n be the number of outcomes, and $p_1, \ldots, p_n$ their probabilities. For each $\lambda \in [0, 1]$ define $p^{(\lambda)}$ to be the distribution $p_1^{(\lambda)} = p_1 + (1 - p_1)(1-\lambda), p_i^{(\lambda)} = p_i \cdot \lambda$ for $i > 1$, and $h^{(\lambda)}$ its entropy. Then $h^{(0)} = 0$, $h^{(1)} = \lceil c \rceil \cdot h$, and, by continuity, there exists λ s.t. $h^{(\lambda)} = c \cdot h$. □

Corollary B.3. *For every $W \subsetneq V$ and every $c > 0$, the function $c \cdot h_W$ is entropic, where h_W is the step function. It follows that every normal function is entropic (because it is a sum $\sum_W c_W h(W)$ and $c_W h(W)$ is entropic).*

Proof. By the previous fact, there exists a random variable Z whose entropy is $h_0(Z) = c$. Let h be the entropy of the following n random variables: forall $U \in V - W$, define $U \stackrel{\text{def}}{=} Z$ (hence, forall $X \subseteq V - W$, $h(X) = h_0(Z) = c$), and for every $U \in W$, define U to be a constant (hence for every $X \subseteq W$, $h(X) = 0$). Therefore, $h = c \cdot h_W$. □

However, when $n \geq 3$, then Zhang and Yeung [32] proved that $c \cdot h$ is not necessarily entropic. Their proof is based on the *parity function*, introduced in Example 3.8.

Fact B.4. Γ_3^* is not convex.

Proof. Zhang and Yeung [32] prove this fact as follows. Let h be the entropy of the parity function in Example 3.8. For every $c > 0$, consider the function $h' = c \cdot h$. They prove that h' is entropic iff $c = \log M$, for some integer M, which implies that Γ_3^* is not convex. We include here their proof for completeness. Assuming h' is entropic let p' be its probability distribution, then the following independence constraints hold: $X \perp Y$, because $h'(XY) = h'(X) + h'(Y)$, and similarly $X \perp Z$ and $Y \perp Z$. The following functional dependencies also hold: $XY \to Z$ (because $h'(XY) = h'(XYZ)$) and similarly $XZ \to Y$, $YZ \to X$. Let x, y, z be any three values s.t. $p'(x, y, z) > 0$. Then $p'(x, y, z) = p'(x, y) = p'(x)p'(y)$. Similarly $p'(x, y, z) = p'(y)p'(z)$, which implies $p'(x) = p'(z)$. Therefore, for any other value x', $p'(x') = p'(z)$. This means that the variable X is uniformly distributed, because $p'(x) = p'(x')$ forall x, x', hence $p'(x) = 1/M$ where M is the size of the domain of X. It follows that $h'(X) = \log M$, proving the claim. □

Yeung [30] proves that the topological closure $\bar{\Gamma}_n^*$ is a convex set, for every n. Thus, $\Gamma_n^* \subseteq \bar{\Gamma}_n^*$ and the inclusion is strict for $n \geq 3$. The elements of $\bar{\Gamma}_n^*$ are called *almost entropic functions*. We note that if a linear information inequality, or a max-linear information inequality is valid forall entropic functions $h \in \Gamma_n^*$, then, by continuity, it is also valid forall almost entropic functions $h \in \bar{\Gamma}_n^*$.

Let h be an entropic function, and $X, Y \subseteq V$ two sets of variables. For every outcome $X = x$, we denote by $h(Y|X = x)$ the entropy of Y conditioned on $X = x$. The function $Y \mapsto h(Y|X = x)$ is an entropic function (by definition). Recall that we have defined $h(Y|X) \stackrel{\text{def}}{=} h(XY) - h(X)$. It can be shown by direct calculation that $h(Y|X) = \sum_x h(Y|X = x) \cdot p(X = x)$, in other words it is a convex combination of entropic functions. Thus, $h(Y|X)$ is the expectation, over the outcomes x, of $h(Y|X = x)$, justifying the name "conditional entropy".

Fact B.5. In general, the mapping $Y \mapsto h(Y|X)$ is not entropic.

Proof. To see an example, consider two probability spaces on X, Y, Z, with probabilities p, p' and entropies h, h' such that h is the entropy of the parity (Example 3.8) and $h' = 2h$. Consider a 4'th variable U, whose outcomes are $U = 0$ or $U = 1$ with probabilities 1/2, and consider the mixture model: if $U = 0$ then sample X, Y, Z using p, if $U = 1$ then sample X, Y, Z using p'. Let h'' be the entropy over the variables X, Y, Z, U. Then the conditional entropy $h''(W|U) = 3/2h(W)$, for all $W \subseteq \{X, Y, Z\}$, and thus it is not entropic. □

Yeung [30] defines the I-measure as follows. Fix a set of variables V, which we identify with $[n]$. Let $\Omega = 2^{[n]} - \{\emptyset\}$. An I-measure is any function $\mu : 2^\Omega \to \mathbb{R}$ such that $\mu(X \cup Y) = \mu(X) + \mu(Y)$ whenever $X \cap Y = \emptyset$. Notice that μ is not necessarily positive. For each variable $V_i \in V$ we denote by $\hat{V}_i \stackrel{\text{def}}{=} \{\omega \in \Omega \mid i \in \omega\} \subseteq \Omega$, and extend this notation to sets $X \subseteq V$ by setting $\hat{X} \stackrel{\text{def}}{=} \bigcup_{V \in X} \hat{V}$. For each variable V_i denote $\hat{V}_i^1 \stackrel{\text{def}}{=} \hat{V}_i$ and $\hat{V}_i^0 \stackrel{\text{def}}{=}$ the complement of $\hat{V}_i$. An *atomic cell* is an intersection $C \stackrel{\text{def}}{=} \bigcap_{j=1,n} \hat{V}_j^{\varepsilon_j}$, where $\varepsilon_j \in \{0, 1\}$ forall j, where at least one $\varepsilon_j = 1$. Obviously, μ is uniquely defined by its values on the atomic cells.

Given $h \in \mathbb{R}^{2^n}$ (not necessarily entropic), the *I-measure associated* to h is the unique μ satisfying the following, forall $X \subseteq V$:

$$h(X) = \sum_{C: C \subseteq \hat{X}} \mu(C) \tag{47}$$

The normal entropic functions $\mathcal{N}_n$ are precisely those with a non-negative I-measure. This can be seen immediately by observing that, for any step function h_W, it's I-measure μ_W assigns the value 1 to the cell $(\bigcap_{V \notin W} V^1) \cap (\bigcap_{V \in W} V^0)$, and 0 to everything else. In fact, there is a tight connection between the I-measure μ and the Möbious inverse function g (Eq.(37) in Sec. 6), which we explain next. First, we notice that Equation (37) implies:

$$h(X) = -\sum_{Y: Y \not\supseteq X} g(Y) \tag{48}$$

The connection between μ and g follows by a careful inspection of Eq. (47) and Eq (48). Each atomic cell C in Eq. (47) is uniquely defined by the set of its negatively occurring variables, denote this by $\mathsf{neg}(C)$. Then, $C \subseteq \hat{X}$ iff $X \not\subseteq \mathsf{neg}(C)$. Define the function $g : 2^V \to \mathbb{R}$ as $g(\mathsf{neg}(C)) \stackrel{\text{def}}{=} -\mu(C)$ and $g(V) = h(V)$ (recall that $\mathsf{neg}(C) \neq V$). Then Eq.(47) becomes $h(X) = \sum_{C: X \not\subseteq \mathsf{neg}(C)} \mu(C) = -\sum_{Y: X \not\subseteq Y} g(Y)$ which is precisely Eq.(48).

C PROOF OF THEOREM 4.2

We give here the proof of Theorem 4.2. This generalizes the proof of Theorem 3.1 by Kopparty and Rossman [22], whose main idea is illustrated in Example 4.3. Recall the theorem:

Theorem 4.2 (A sufficient condition). *Let Q_1, Q_2 be two conjunctive queries, $n = |\mathsf{vars}(Q_1)|$, and let $\mathsf{TD}(Q_2)$ denote the set of tree decompositions of Q_2. If the following* Max-II *inequality holds $\forall h \in \Gamma_n^*$:*

$$h(\mathsf{vars}(Q_1)) \leq \max_{(T,\chi) \in \mathsf{TD}(Q_2)} \max_{\varphi \in \mathsf{hom}(Q_2, Q_1)} (E_T \circ \varphi)(h) \tag{49}$$

then $Q_1 \preceq Q_2$.

To prove the theorem we need three lemmas. The first lemma is folklore, and represents the main property of tree decomposition used for query evaluation. If $f \in D^X$, $g \in D^Y$ agree on $X \cap Y$, then $f \bowtie g$ is the unique tuple $\in D^{X \cup Y}$ that extends both f and g. If $P_1 \subseteq D^X, P_2 \subseteq D^Y$, then $P_1 \bowtie P_2 = \{f \bowtie g \mid f \in P_1, g \in P_2\}$.

Lemma C.1. *Let (T, χ) be a tree decomposition for Q and recall that $Q \equiv \bigwedge_{t \in \text{nodes}(T)} Q_t$ where Q_t is a conjunction of atoms A s.t. $\text{vars}(A) \subseteq \chi(t)$. Then, for every $\mathcal{D}$, $\text{hom}(Q, \mathcal{D}) = \bowtie_{t \in \text{nodes}(t)} \text{hom}(Q_t, \mathcal{D})$.*

Lemma C.2. *Fix a homomorphism $\varphi : Q_2 \to Q_1$, let (T, χ) be a tree decomposition of Q_2, $\mathcal{D}$ be a database instance, and $P = \text{hom}(Q_1, \mathcal{D})$. Then, for every node $t \in \text{nodes}(T)$, denoting $P'_t \stackrel{\text{def}}{=} \Pi_{\varphi|_{\chi(t)}}(P)$ we have:*

$$P'_t \subseteq \text{hom}(Q_t, \mathcal{D}) \tag{50}$$

Proof. Every tuple in $\Pi_{\varphi|_{\chi(t)}}(P)$ is the composition $f \circ \varphi|_{\chi(t)}$ for some $f \in P$. The lemma follows from the fact that both $\varphi|_{\chi(t)} : Q_t \to Q_1$ and $f : Q_1 \to \mathcal{D}$ are homomorphism. □

Lemma C.3. *(1) Let $p : P(\subseteq D^V) \to [0,1]$ be a probability distribution, and $h : 2^V \to \mathbb{R}_+$ be its entropy. If $\varphi : Y \to V$ and $Z \subseteq Y$, then the $\varphi|_Z$-pullback of p, $\Pi_{\varphi|_Z}(p)$, is equal to the Z-marginal of $\Pi_\varphi(p)$. In particular, if $h' : 2^Y \to \mathbb{R}_+$ is the entropy of $\Pi_\varphi(p)$, then, $\forall Z \subseteq Y$, $h'(Z) = h(\varphi(Z))$. (2) If $\varphi : V' \to V$ and $Y_1, Y_2 \subseteq V'$, then the pull-back distributions $\Pi_{\varphi|_{Y_1}}(p)$ and $\Pi_{\varphi|_{Y_2}}(p)$ agree on the common variables $Y_1 \cap Y_2$.*

Proof. (1) The φ-pullback $\Pi_\varphi(p)$ is defined to be the same as the $\varphi(Y)$-marginal of p. Therefore its Z-marginal is the $\varphi(Z)$-marginal of p. By definition, $\Pi_{\varphi|_Z}(p)$ is also the $\varphi(Z)$-marginal of p, hence they are equal. Formally, given $f \in P$:

$$\Pi_\varphi(p)(Z = \Pi_Z(\Pi_\varphi(f))) = \sum_{f' : \Pi_Z(\Pi_\varphi(f')) = \Pi_Z(\Pi_\varphi(f))} p(f')$$

$$= \sum_{f' : \Pi_{\varphi(Z)}(f') = \Pi_{\varphi(Z)}(f)} p(f') = \Pi_{\varphi|_Z}(p)(Z = \Pi_{\varphi|_Z}(f))$$

because $\Pi_Z \circ \Pi_\varphi = \Pi_{\varphi|_Z}$. This discussion immediately implies that $h'(Z) = h(\varphi(Z))$, forall Z.

(2) Let $Z = Y_1 \cap Y_2$. By claim (1), the Z-marginal of $\Pi_{\varphi|_{Y_1}}(p)$ is $\Pi_{\varphi|_Z}(p)$ and similarly for the Z-marginal of $\Pi_{\varphi|_{Y_2}}(p)$, hence they are equal. □

Proof. (Of Theorem 4.2) Let $\mathcal{D}$ be any database with domain D, and let $P = \text{hom}(Q_1, \mathcal{D})$. Consider the uniform probability distribution $p : P \to [0,1]$, defined as $p(f) = 1/|P|$ for all tuples $f \in P$, and let h be its entropy. We have $h = \log|P|$ because p is uniform. By assumption of the theorem, there exists a homomorphism $\varphi : Q_2 \to Q_1$ and a tree decomposition (T, χ) of Q_2 such that:

$$\log|P| = h(\text{vars}(Q_1)) \leq (E_T \circ \varphi)(h) \tag{51}$$

For each $t \in \text{nodes}(T)$, consider the projections of P and p on $\chi(t)$:

$$P'_t \stackrel{\text{def}}{=} \Pi_{\varphi|_{\chi(t)}}(P), \qquad p'_t \stackrel{\text{def}}{=} \Pi_{\varphi|_{\chi(t)}}(p)$$

Lemma C.1 and Lemma C.2 imply:

$$P' \stackrel{\text{def}}{=} \bowtie_{t \in \text{nodes}(T)} P'_t \subseteq \bowtie_{t \in \text{nodes}(T)} \text{hom}(Q_t, \mathcal{D})$$

$$= \text{hom}(Q_2, \mathcal{D}) \tag{52}$$

We will construct a probability distribution $p' : P' \to [0,1]$, with entropy function $h' : 2^{\text{vars}(Q_2)} \to \mathbb{R}_+$, such that the following hold:

$$h'(\text{vars}(Q_2)) = E_T(h') \tag{53}$$

$$E_T(h') = (E_T \circ \varphi)(h) \tag{54}$$

We will construct p' by stitching together the pull-back distributions p'_t, for $t \in \text{nodes}(T)$; this is possible because, by Lemma C.3 (2), any two induced probabilities p'_{t_1}, p'_{t_2} agree on the common variables $\chi(t_1) \cap \chi(t_2)$.

Formally, we start by listing $\text{nodes}(T)$ in some order, $t_1, t_2, \ldots, t_m$, such that each child is listed after its parent. Let $P'_i \stackrel{\text{def}}{=} \bowtie_{j=1,i} P'_{t_j}$, let T_i be the subtree induced by the nodes $\{t_1, \ldots, t_i\}$, and $\text{vars}(T_i) = \bigcup_{j=1,i} \chi(t_i)$ its variables. We construct by induction on i a probability distribution $p'_i : P'_i \to [0,1]$ such it agrees with $p'_{t_1}, \ldots, p'_{t_i}$ on $\chi(t_1), \ldots, \chi(t_i)$ respectively, and it's entropy function $h'_i : 2^{\text{vars}(T_i)} \to \mathbb{R}_+$ satisfies:

$$h'_i(\text{vars}(T_i)) = E_{T_i}(h'_i) \tag{55}$$

$$E_{T_i}(h'_i) = (E_{T_i} \circ \varphi)(h) \tag{56}$$

To define p'_i, we need to extend p'_{i-1} to the variables $\text{vars}(T_i) - \text{vars}(T_{i-1}) = \chi(t_i) - \chi(\text{parent}(t_i))$. We define p'_i to satisfy the following: (1) p'_i agrees with p'_{t_i} on $\chi(t_i)$, (2) p'_i agrees with p'_{i-1} on the $\text{vars}(T_{i-1})$, and (3) $\chi(t_i)$ is independent of $\text{vars}(T_{i-1})$ given $\chi(t_i) \cap \chi(\text{parent}(t_i))$. Notice that (1) and (2) are not conflicting because p'_{t_i} agrees with any other p'_j on their common variables. Formally, we define p'_i through a sequence of three equations:

$$p'_i(\chi(t_i) | \chi(t_i) \cap \chi(\text{parent}(t_i))) \stackrel{\text{def}}{=} p'_{t_i}(\chi(t_i) | \chi(t_i) \cap \chi(\text{parent}(t_i))) \tag{57}$$

$$p'_i(\chi(t_i) | \text{vars}(T_{i-1})) \stackrel{\text{def}}{=} p'_i(\chi(t_i) | \chi(t_i) \cap \chi(\text{parent}(t_i))) \tag{58}$$

$$p'_i(\text{vars}(T_i)) \stackrel{\text{def}}{=} p'_i(\chi(t_i) | \text{vars}(T_{i-1})) p'_{i-1}(\text{vars}(T_{i-1})) \tag{59}$$

We check Eq.(55):

$$\begin{aligned} h'_i(\text{vars}(T_i)) =& h'_i(\chi(t_i) | \text{vars}(T_{i-1})) + h'_{i-1}(\text{vars}(T_{i-1})) \\ & \text{(by Eq.(59))} \\ =& h'_i(\chi(t_i) | \text{vars}(T_{i-1})) + E_{T_{i-1}}(h'_{i-1}) \\ & \text{(Induction)} \\ =& h'_i(\chi(t_i) | \text{vars}(T_{i-1})) + E_{T_{i-1}}(h'_i) \\ & (h'_i \text{ is identical to } h'_{i-1} \text{ on } \text{vars}(T_{i-1})) \\ =& h'_i(\chi(t_i) | \chi(t_i) \cap \chi(\text{parent}(t_i))) + E_{T_{i-1}}(h'_i) \\ & \text{(by Eq.(58))} \\ =& E_{T_i}(h') \qquad \text{(Definition of } E_T) \end{aligned}$$

We check Eq.(56).

$$\begin{aligned} E_{T_i}(h'_i) =& h'_i(\chi(t_i) | \chi(t_i) \cap \chi(\text{parent}(t_i))) + E_{T_{i-1}}(h'_i) \\ & \text{(Definition of } E_T) \\ =& h'_i(\chi(t_i) | \chi(t_i) \cap \chi(\text{parent}(t_i))) + (E_{T_{i-1}} \circ \varphi)(h) \\ & \text{(Induction)} \end{aligned}$$

$$=h'_{t_i}(\chi(t_i)|\chi(t_i)\cap\chi(\text{parent}(t_i))) + (E_{T_{i-1}}\circ\varphi)(h) \quad \text{(by Eq.(57))}$$
$$=h(\varphi(\chi(t_i))|\varphi(\chi(t_i)\cap\chi(\text{parent}(t_i)))) + (E_{T_{i-1}}\circ\varphi)(h) \quad \text{(Lemma C.3 (1))}$$
$$=(E_{T_i}\circ\varphi)(h) \qquad \text{(Definition of } E_T\text{)}$$

This completes the inductive proof.

By setting $i = m$ (the number of nodes in T) in Eq.(55) and (56) we derive Eq.(53) and (54). The proof of the theorem follows from:

$$\begin{aligned}
\log|\text{hom}(Q_1,\mathcal{D})| &= \log|P| \\
&= h(\text{vars}(Q_1)) \le (E_T\circ\varphi)(h) && \text{(by Eq. (51))}\\
&= E_T(h') && \text{(by Eq.(54))}\\
&= h'(\text{vars}(Q_2)) && \text{(by Eq.(53))}\\
&\le \log|P'| && \text{(Since } P' \text{ is the support of } h')\\
&\le \log|\text{hom}(Q_2,\mathcal{D})| && \text{(By Eq (52))}
\end{aligned}$$

□

D PROOF OF THEOREM 3.3 AND 3.6

In Th.4.4 we proved that, when Q_2 is acyclic and Eq.(13) fails, then $Q_1 \not\preceq Q_2$. We prove here a variation of that result: when Q_2 is chordal and Eq.(13) fails on a normal entropic function, then $Q_1 \not\preceq Q_2$. Recall that a junction tree is a special tree decomposition.

Lemma D.1. *Let Q_2 be chordal and admit a simple junction tree T, and let E_T be its linear expression, Eq.(12). If there exists a normal entropic function h (i.e. with a non-negative I-measure) such that:*

$$h(\text{vars}(Q_1)) > \max_{\varphi\in\text{hom}(Q_2,Q_1)} (E_T\circ\varphi)(h) \tag{60}$$

then there exists a database instance $\mathcal{D}$ such that $|\text{hom}(Q_1,\mathcal{D})| > |\text{hom}(Q_2,\mathcal{D})|$.

We first show how to use the lemma and the essentially-Shannon inequalities in Th 3.9 to prove Theorems 3.3 and 3.6. Assume Q_2 is chordal and has a simple junction tree T. We prove: $Q_1 \preceq Q_2$ iff Eq.(13) holds. It suffices to prove that Eq.(13) is necessary, because sufficiency follows from Th. 4.2. Suppose Eq.(13) fails. Then there exists an entropic function h such that (60) holds where T in (60) is a simple junction tree of Q_2. Since T is simple, the conditional linear expressions on the right-hand-side of (60) are also simple. By Th 3.9, there exists a *normal* entropic function h such that (60) holds. Then, by Lemma D.1, $Q_1 \not\preceq Q_2$. This proves that Eq.(13) is necessary and sufficient for containment. Furthermore, Eq.(13) is decidable, since it is an essentially-Shannon inequality, and this completes the proof of Theorems 3.3. The proof of Theorem 3.6 follows immediately from the fact that the set of normal entropic functions $\mathcal{N}_n$ is the cone generated by the entropies of normal relations, and the set of modular functions $\mathcal{M}_n$ is the cone generated by the entropies of product relations.

It remains to prove Lemma D.1; the lemma generalizes Theorem 3.2 of [22] to arbitrary vocabularies (beyond graphs). To prove the theorem, we will update the proof of Theorem 4.4, where we used acyclicity of Q_2: more precisely we need to re-prove the locality property, Eq.(22). We repeat it here:

$$\text{hom}_{\varphi|_{\chi(t)}}(Q_t,\mathcal{D}) \subseteq \Pi_{\varphi|_{\chi(t)}}(P)$$

We start by observing that this property fails in general.

Example D.2. Let $Q_1 = R(X_1,X_2), S(X_2,X_3), T(X_3,X_1)$ and $Q_2 = R(Y_1,Y_2), S(Y_2,Y_3), T(Y_3,Y_1)$ (they are identical). Consider the parity function in Example 3.8; more precisely, this is the entropy of the relation $P = \{(X_1,X_2,X_3) \mid X_1,X_2,X_3 \in \{0,1\}, X_1\oplus X_2\oplus X_3 = 0\}$, which we show here for clarity:

$P =$

0	0	0
0	1	1
1	0	1
1	1	0

Recall that the entropy of P is not a normal entropic function (Sec. 6). This relation is perfectly uniform (in fact it is a group characterization). Computing $\mathcal{D} = \Pi_{Q_1}(P)$ we obtain $R^D = S^D = T^D = \{(0,0),(0,1),(1,0),(1,1)\}$. Q_2 is a clique, with a bag $Q_t = Q_2$, and $\text{hom}(Q_t,\mathcal{D})$ contains one extra triangle, $(1,1,1)$, which is in no single row of P.

The example shows that we need to use in a critical way the fact that the counterexample h is a normal entropic function, $h \in \mathcal{N}_n$. To use this fact, we will describe a class of relations whose entropic functions generate precisely the cone $\mathcal{N}_n$, and prove that these are precisely the normal relations (Def. 3.5).

Consider the normal entropic function h given by Lemma D.1. We can assume w.l.o.g. that h is a sum of step functions[13], $h = \sum_i h_{W_i}$, where each h_{W_i} is a step function (not necessarily distinct). Recall from Section 3.3 that P_{W_i} is the 2-tuple relation whose entropy is h_{W_i}; to reduce clutter we denote here P_{W_i} by P_i. Then h is the entropy of their domain-product (Def B.1), $P = P_1 \otimes P_2 \otimes \cdots \otimes P_m$. One can check that P is totally uniform (it is even a group realization). We now prove the locality property, Eq.(22), using the fact that P is a domain product, which allows us to rewrite Eq.(22) as:

$$\text{hom}_{\varphi|_{\chi(t)}}(Q_t,\mathcal{D}_1\otimes\cdots\otimes\mathcal{D}_m) \subseteq \Pi_{\varphi|_{\chi(t)}}(P_1\otimes\cdots\otimes P_m)$$

It suffices prove that $\text{hom}_{\varphi|_{\chi(t)}}(Q_t,\mathcal{D}_i) \subseteq \Pi_{\varphi|_{\chi(t)}}(P_i)$ for each i. Recall that P_i has two tuples, $P_i = \{f_1, f_2\}$, where $f_1 = (1,1,\ldots,1)$ and f_2 has values 1 on positions $\in W$ and values 2 on positions $\notin W$, for some set of attributes W. Fix a tuple $g \in \text{hom}_{\varphi|_{\chi(t)}}(Q_t,\mathcal{D}_i)$; we must prove that either $g \in \Pi_{\varphi|_{\chi(t)}}(f_1)$ or $g \in \Pi_{\varphi|_{\chi(t)}}(f_2)$. If g maps every variable in $\text{vars}(Q_t)$ to 1, then the first condition holds, so assume that g maps some variable $Y \in \text{vars}(Q_t)$ to 2; in particular, $\varphi(Y) \notin W$. We must prove that, for every variable Y', if $\varphi(Y') \notin W$ then $g(Y') = 2$. Here we use the fact that Q_2 is chordal, hence Q_t is a clique, thanks to Fact A.2. Therefore, there exists $B \in \text{atoms}(Q_t)$ that contains both Y and Y'. Since g is a homomorphism, it maps B to some tuple in $\Pi_{\varphi(\text{vars}(B))}(P)$; since both $\varphi(Y), \varphi(Y') \notin W$, this tuple must have the value 2 on both positions (they can be identical: $\varphi(Y) = \varphi(Y')$). It follows that all variables Y' s.t. $\varphi(Y') \notin W$ are mapped to 2, proving that $g \in \Pi_{\varphi|_{\chi(t)}}(f_2)$. This proves the local property, Eq.(22). The rest of the proof of Theorem 4.4 remains unchanged, and this completes the proof of Lemma D.1.

[13] Suppose the contrary, that the inequality holds for all functions h that are sums of step functions. Then it holds for all linear combinations $\sum_W c_W h_W$ where $c_W \ge 0$ are integer coefficients. If an inequality holds for h, then it also holds for $\lambda \cdot h$ for any constant $\lambda > 0$; it follows that the inequality holds for all linear combinations $\sum_W c_W h_W$ where $c_W \ge 0$ are rationals. The topological closure of these expressions is $\mathcal{N}_n$, contradicting the fact that the inequality fails on some $h \in \mathcal{N}_n$.