

RECENT RESULTS ON CARDINALITY ESTIMATION AND INFORMATION THEORETIC INEQUALITIES

HUNG Q. NGO

RelationalAI Inc.



Abstract. In the past 15 years or so, new and exciting connections between fundamental problems in database theory and information theory have emerged. There are several angles one can take to describe this connection. This paper takes one such angle, influenced by the author’s own bias and research results. In particular, we will describe how the cardinality estimation problem – a corner-stone problem for query optimizers – is deeply connected to information theoretic inequalities. Furthermore, we explain how inequalities can also be used to derive a couple of classic geometric inequalities such as the Loomis-Whitney inequality.

A purpose of the article is to introduce the reader to these new connections, where theory and practice meet in a wonderful way. Another objective is to point the reader to a research area with many new open questions.

Keywords. Shannon-type inequalities; Cardinality estimation; Conjunctive queries; Information theory.

1. INTRODUCTION

Database theory. Cardinality estimation [19] is a crucial component of the query optimization pipeline. The main question is “given a conjunctive query, return the best estimate of the output size as quickly as possible”. There are *many* approaches to cardinality estimation (see [10, 22] and references thereof).

One approach to cardinality estimation is to give a worst-case upper-bound on the cardinality estimates. Guaranteed worst-case cardinality bounds help make the query optimizer robust to outlier, avoiding a query plan which explodes in runtime under bad inputs. There are recent evidence that these “pessimistic” cost estimators work well in practice [10, 20].

Deriving tight upper-bounds on the output size of a conjunctive query this is where our story begins. We shall see a fascinating connection to information inequalities, geometric inequalities, and how they play a crucial role in the seemingly disjoint and different field of query optimization in databases. In fact, at Relational AI where the author works and leads the query optimization group, these worst-case size bounds are implemented in the

Dedicated to Professor Phan Dinh Dieu on the occasion of his 85th birth anniversary.

Corresponding author.

E-mail address: hung.ngo@relational.ai.

production system to great effect. Making real Mathematics work in practice is a wonderful feeling.

To briefly introduce the readers into these ideas, let us consider a simple example of a conjunctive query: given a directed graph $G = (V, E)$, how many directed triangles does it have? This is the simplest example of the conjunctive query cardinality estimation problem, because the “list all triangles of a graph” query can be expressed as a conjunctive query [1] of the form

$$Q_{\Delta}(a, b, c) \leftarrow E(a, b) \wedge E(b, c) \wedge E(c, a). \tag{1}$$

For the reader who is not familiar with database theory, the query (1) asks the following. There is an input relation E which is a *binary* relation, $(u, v) \in E$ if and only if (u, v) is an edge in the input graph. The *output* of the query, denoted by $Q_{\Delta}(a, b, c)$, includes all triples (a, b, c) for which $(a, b) \in E$, $(b, c) \in E$, and $(c, a) \in E$. The question is, if we know some statistics about the input graph, can we derive a meaningful bound/estimate of $|Q_{\Delta}|$?

Information theory. To answer the above question, here is an argument from an influential paper by Chung et al. [13] in extremal set theory. The paper proved the “Product Theorem” which uses the *entropy argument* connecting a count estimation problem to an entropic inequality. For the purpose of this introduction, we specialize the argument to Q_{Δ} , leaving the more general argument to Section 3.

Fix the input graph $G = (V, E)$. Let Q_{Δ} be the set of all triangles (a, b, c) in the graph. Construct a joint distribution on $V \times V \times V$ by selecting a triangle uniformly at random from Q_{Δ} . Let (A, B, C) denote the corresponding tuple of random variables. Let H denote the entropy of this particular distribution. Then, due to uniformity, the first fact we know is that

$$\log_2 |Q_{\Delta}| = H[A, B, C]. \tag{2}$$

Secondly, the support of the *marginal* distribution on (A, B) is a subset of E . Hence, the marginal entropy on (A, B) can be bounded by $H[A, B] \leq \log_2 |E|$. By symmetry, we have

$$H[A, B] \leq \log_2 |E|, \quad H[B, C] \leq \log_2 |E|, \quad H[A, C] \leq \log_2 |E|. \tag{3}$$

Finally, thanks to Shannon [14], we know that the function H is non-negative, monotone, and submodular. In particular, H satisfies the following basic Shannon inequalities

$$0 = H[\emptyset], \tag{4}$$

$$0 \leq H[X] \quad \forall X \subseteq \{A, B, C\}, \tag{5}$$

$$H[X] \leq H[Y] \quad \forall X, Y \text{ where } X \subseteq Y, \text{ and } X, Y \subseteq \{A, B, C\}, \tag{6}$$

$$H[X \cup Y|X] \leq H[Y|X \cap Y] \quad \forall X, Y \subseteq \{A, B, C\}. \tag{7}$$

In the above, for $Z \subseteq W$, $H[Z|W] := H[Z] - H[W]$ denotes the conditional entropy. Let Γ_3 denote the set of all functions $h : 2^{\{A, B, C\}} \rightarrow \mathbb{R}$ satisfying (5), (6), and (7), with $h(\emptyset) = 0$. Then, $H \in \Gamma_3$ and H satisfies (3) and (2). Hence, we can bound $|Q_{\Delta}|$ via $\log_2 |Q_{\Delta}|$ by

$$\begin{aligned} \log_2 |Q_{\Delta}| &\leq \max && h(A, B, C), \\ \text{s.t.} &&& h(A, B) \leq \log_2 |E|, \quad h(B, C) \leq \log_2 |E|, \quad h(C, A) \leq \log_2 |E|, \\ &&& h \in \Gamma_3. \end{aligned} \tag{8}$$

Note that (8) is simply a linear program over 7 variables $h(X)$, one for each non-empty subset $X \subseteq \{A, B, C\}$. Hence, solving for this linear program would be a way to obtain a bound. However, this is not completely satisfactory because algorithmic bounds do not give us insights into the magnitude of the attained bound.

Fortunately, there is a way to derive a more explicit bound from the above linear program. The following derivation line is due to Radhakrishnan [29]. From the basic Shannon inequalities that $h \in \Gamma_3$ satisfies, it follows that

$$\begin{aligned} h(A, B, C) &= h(A) + h(AB \mid A) + h(ABC \mid AB) \\ &= \frac{1}{2} (h(A) + h(AB \mid A) + h(A) + h(ABC \mid AB) + h(AB \mid A) + h(ABC \mid AB)) \\ &\leq \frac{1}{2} (h(A) + h(AB \mid A) + h(A) + h(AC \mid A) + h(B \mid \emptyset) + h(BC \mid B)) \\ &= \frac{1}{2} (h(A, B) + h(B, C) + h(B, C)). \end{aligned}$$

The inequality

$$h(A, B, C) \leq \frac{1}{2} (h(A, B) + h(B, C) + h(B, C)) \quad (9)$$

is a special case of the famous *Shearer's inequality* [13]. The inequality is a characterization of points in the polyhedron Γ_3 . Since $H \in \Gamma_3$, it satisfies Shearer's inequality, from which we can now derive an explicit upper-bound

$$\log_2 |Q_\Delta| = H[A, B, C], \quad (10)$$

$$\text{(from (9))} \leq \frac{1}{2} (H[A, B] + H[B, C] + H[A, C]) \quad (11)$$

$$\leq \frac{1}{2} (\log_2 |E| + \log_2 |E| + \log_2 |E|). \quad (12)$$

In other words, the number of triangles in a directed graph is bounded by $|Q_\Delta| \leq |E|^{3/2}$. This beautiful result has been rediscovered a few times [7] in different contexts (see [26] for a small survey).

Geometric inequalities. One way to think about the output size bound is to think of Q_Δ as containing points (a, b, c) in a three-dimensional space, whose projection onto the (A, B) -plane is contained in the two-dimensional point-set E , onto the (B, C) plane is contained in E , and onto the (A, C) -plane is contained in E . There is a known geometric inequality shown by Loomis and Whitney in 1949 [23] which addresses a more general problem: bound the volume of a convex body in space whose shadows on the coordinate hyperplanes have bounded areas. The triangle query above corresponds to the discrete measure case, where "volume" becomes "count". Specializing to the triangle, Loomis-Whitney states that

$$|Q_\Delta| \leq \sqrt{|E| \cdot |E| \cdot |E|} = |E|^{3/2}, \quad (13)$$

which is exactly what we derived while taking the information theoretic route.

Thus, while studied in a completely different context, Loomis-Whitney's inequality is our earliest known answer to determining the worst-case size bound of a special class of conjunctive queries. The class of queries coming from Loomis-Whitney setup are now referred to as *Loomis-Whitney queries* [27, 28].

Paper organization. We aim for the paper to be somewhat self-contained. However, that can only be accomplished given that the reader is willing to forego certain level of rigor, or that the reader is somewhat familiar with elementary database theory and information theory. Hence, we will be almost always brief in presenting materials, favoring to convey the main ideas instead of being correct the complete to the last detail. The reader is encouraged to refer to the references (such as [1, 14, 26, 28]) for more background and survey presentations.

The rest of the paper is organized as follows. Section 2 presents basic setup from database theory and information theory. Section 3 describes the information theoretic approach to the cardinality bound problem by deriving the bound using the entropy argument. This section generalizes the above Q_Δ example in two ways: the input query is arbitrary, and the set of constraints (such as $h(A, B) \leq \log |E|$ constraint) is much more general. We will explain what is known and not known about computing these bounds, and discuss the computational problem of computing the bound efficiently. More efficient formulations of the bounds are then discussed in Section 4. Section 5 concludes the paper.

2. BACKGROUND

Throughout the paper, we use the following convention. The non-negative reals, rationals, and integers are denoted by \mathbb{R}_+ , \mathbb{Q}_+ , and \mathbb{N} respectively. For a positive integer n , $[n]$ denotes the set $\{1, \dots, n\}$.

Functions \log without a base specified are base-2, i.e. $\log = \log_2$. Uppercase A_i denotes a variable/attribute, and lowercase a_i denotes a value in the discrete domain $\text{Dom}(A_i)$ of the variable. For any subset $S \subseteq [n]$, define $\mathbf{A}_S = (A_i)_{i \in S}$, $\mathbf{a}_S = (a_i)_{i \in S} \in \prod_{i \in S} \text{Dom}(A_i)$. In particular, \mathbf{A}_S is a tuple of variables and \mathbf{a}_S is a tuple of specific values with support S . We also use \mathbf{X}_S to denote variables and $\mathbf{x}_S, \mathbf{t}_S$ to denote value tuples in the same way.

A *multi-hypergraph* is a hypergraph where edges may occur more than once. Given a multi-hypergraph $\mathcal{H} = ([n], \mathcal{E})$, a vector $\boldsymbol{\delta} \in \mathbb{R}^{\mathcal{E}}$ is called a *fractional edge cover* of \mathcal{H} if it is non-negative (i.e. $\boldsymbol{\delta} \geq \mathbf{0}$ component-wise), and every vertex is “fractionally covered” at least 1, i.e. $\sum_{F: v \in F} \delta_F \geq 1$, for all $v \in [n]$.

2.1. Database theory

We associate a *full conjunctive query* Q to a multi-hypergraph $\mathcal{H} := ([n], \mathcal{E})$, $\mathcal{E} \subseteq 2^{[n]}$, the query is written as

$$Q(\mathbf{A}_{[n]}) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F), \quad (14)$$

with variables A_i , $i \in [n]$, and *atoms* R_F , $F \in \mathcal{E}$. The atoms R_F represent (relational) tables whose columns are variables \mathbf{A}_F . The answer to the query consists of all tuples $\mathbf{a}_{[n]}$ whose projections onto the coordinates $F \in \mathcal{E}$ belong to the corresponding table R_F , i.e. $\mathbf{a}_F \in R_F$ for all $F \in \mathcal{E}$.

In the triangle query example (1), the hypergraph has $n = 3$ vertices, and the query (1) is isomorphic to the following, presented in the form of the general conjunctive query (14)

$$Q_\Delta(A_1, A_2, A_3) \leftarrow R_{12}(A_1, A_2) \wedge R_{13}(A_1, A_3) \wedge R_{31}(A_3, A_1). \quad (15)$$

Each atom R_{ij} refers to the underlying “table” E which is the table of all edges of the graph that we want to count the number of triangles of.

When proving $|Q_\Delta| \leq |E|^{3/2}$ in the previous section, we use the query structure (1) and the fact that every input relation has size (at most) $|E|$. This is an example of a *constraint* that the input satisfies. We now generalize and formalize this notion of constraints, serving as inputs to our cardinality estimation problem.

A *degree constraint* is a triple $(X, Y, N_{Y|X})$, where $X \subsetneq Y \subseteq [n]$ and $N_{Y|X} \in \mathbb{N}$. The relation R_F is said to *guard* the degree constraint $(X, Y, N_{Y|X})$ if $Y \subseteq F$ and

$$\deg_F(\mathbf{A}_Y | \mathbf{A}_X) := \max_{\mathbf{t}} |\pi_{\mathbf{A}_Y}(\sigma_{\mathbf{A}_X = \mathbf{t}}(R_F))| \leq N_{Y|X}. \quad (16)$$

Here, π denotes the “projection operator” and σ the “selection operator” in relational algebra [1]. In plain language, the degree constraint states that: “in the relation R_F , for every fixed binding $\mathbf{A}_X = \mathbf{t}$, there are at most $N_{Y|X}$ bindings of \mathbf{A}_Y .”

Note that a given relation may guard multiple degree constraints. Let DC denote a set of degree constraints. The input database \mathbf{D} is said to *satisfy* DC if every constraint in DC has a guard, in which case we write $\mathbf{D} \models \text{DC}$. (The database assigns, to each relation symbol R_F , an actual relation typically denoted by $R_F^{\mathbf{D}}$. To avoid cumbersome notations we avoid adding super-scripts.)

A *cardinality constraint* is an assertion of the form $|R_F| \leq N_F$, for some $F \in \mathcal{E}$; it is exactly the degree constraint $(\emptyset, F, N_{F|\emptyset})$ guarded by R_F . A *functional dependency* $\mathbf{A}_X \rightarrow \mathbf{A}_Y$ is a degree constraint with $N_{X \cup Y|X} = 1$. In particular, degree constraints strictly generalize both cardinality constraints and functional dependencies.

In the triangle query (15), suppose in addition to knowing that $|E| = N$ we also know that the out-degree of every vertex is bounded by D . Then this database (i.e. the graph G) satisfies the following degree constraints: $(\emptyset, \{u, v\}, N)$ for every pair $u, v \in [3]$. and $(\{1\}, \{1, 2\}, D)$ $(\{2\}, \{2, 3\}, D)$ and $(\{3\}, \{3, 1\}, D)$.

Our problem setting is general, where we are given a query of the form (14) and a set DC of degree constraints satisfied by the input database \mathbf{D} . The first task is to find a good upper bound of, or to determine exactly the quantity $\sup_{\mathbf{D} \models \text{DC}} |Q(\mathbf{D})|$, the worst-case output size of the query given that the input satisfies the degree constraints. The second task is to design an algorithm running in time as close to the bound as possible.

2.2. Information theory

We present simple facts about entropies that are needed in the rest of this paper. (See [14, 30] for a more formal treatment.)

Consider a joint probability distribution \mathcal{D} on n discrete variables $\mathbf{A} = (A_i)_{i \in [n]}$ and a probability mass function Pr . The *entropy function* associated with \mathcal{D} is a function $H : 2^{\mathbf{A}} \rightarrow \mathbb{R}_+$, where

$$H[\mathbf{A}_F] := \sum_{\mathbf{a}_F \in \prod_{i \in F} \text{Dom}(A_i)} \text{Pr}[\mathbf{A}_F = \mathbf{a}_F] \log \frac{1}{\text{Pr}[\mathbf{A}_F = \mathbf{a}_F]} \quad (17)$$

is the entropy of the marginal distribution on \mathbf{A}_F . To simplify notations, we will also write $H[F]$ for $H[\mathbf{A}_F]$, turning H into a set function $H : 2^{[n]} \rightarrow \mathbb{R}_+$. For any $F \subseteq [n]$, define the “support” of the marginal distribution on \mathbf{A}_F to be

$$\text{supp}_F(\mathcal{D}) := \left\{ \mathbf{x}_F \in \prod_{i \in F} \text{Dom}(A_i) \mid \text{Pr}[\mathbf{A}_F = \mathbf{x}_F] > 0 \right\}. \quad (18)$$

Given $X \subseteq Y \subseteq [n]$, define the *conditional entropy* to be

$$H[Y \mid X] := H[Y] - H[X]. \tag{19}$$

This is also known as the *chain rule* of entropy, which – for the purpose for this paper – we took as the *definition* of conditional entropy. Note that we chose to define conditional entropies (19) only for $X \subseteq Y$. This is without loss of generality and for the brevity of later sections. In general, the chain rule for conditional entropies can be written as, for arbitrary X, Z

$$H[Z \mid X] := H[Z \cup X \mid X].$$

In our sense (19), we can pick $Y = Z \cup X$ to recover the above more standard expression for conditional entropies.

The following facts are basic and fundamental in information theory

$$H[\emptyset] = 0, \tag{20}$$

$$H[X] \leq \log |\text{supp}_X(\mathcal{D})| \quad \forall X \subseteq [n], \tag{21}$$

$$H[X] \leq H[Y] \quad \forall X \subseteq Y \subseteq [n], \tag{22}$$

$$H[X \cup Y \mid Y] \leq H[X \mid X \cap Y] \quad \forall X, Y \subseteq [n]. \tag{23}$$

Inequality (21) follows from Jensen’s inequality and the concavity of the entropy function. Equality holds if and only if the marginal distribution on X is *uniform*. Entropy measures the “amount of uncertainty” we have: The more uniform the distribution, the less certain we are about where a random point is in the space. Inequality (22) is the *monotonicity* property: Adding more variables *increases* uncertainty. Inequality (23) is the *submodularity* property: Conditioning on more variables *reduces* uncertainty.¹

For any three sets of variables X, Y, Z , the *conditional mutual information* $I(X; Y \mid Z)$ is defined by

$$I(X; Y \mid Z) := H[XZ \mid Z] - H[XYZ \mid YZ]. \tag{24}$$

From submodularity of H , it follows that conditional mutual information is non-negative.

When writing entropies, it is customary to drop the union \cup , writing $H[XY \mid Z]$ instead of $H[X \cup Y \mid Z]$, for example. The formula (24) above used this convention.

2.3. Set functions and polymatroids

A function $f : 2^{\mathcal{V}} \rightarrow \mathbb{R}_+$ is called a (non-negative) *set function* on \mathcal{V} . A set function f on \mathcal{V} is *modular* if $f(S) = \sum_{v \in S} f(\{v\})$ for all $S \subseteq \mathcal{V}$, is *monotone* if $f(X) \leq f(Y)$ whenever $X \subseteq Y$, and is *submodular* if $f(X \cup Y) + f(X \cap Y) \leq f(X) + f(Y)$ for all $X, Y \subseteq \mathcal{V}$.

Let n be a positive integer. A function $h : 2^{[n]} \rightarrow \mathbb{R}_+$ is said to be *entropic* if there is a joint distribution on $\mathbf{A}_{[n]}$ with entropy function H such that $h(S) = H[S]$ for all $S \subseteq [n]$. We will write $h(S)$ and $h(\mathbf{A}_S)$ interchangeably, depending on context.

Unless specified otherwise, we will only consider *non-negative* and *monotone* set functions f for which $f(\emptyset) = 0$; this assumption will be implicit in the entire paper. Furthermore, for $X \subseteq Y$, we will write

$$h(Y \mid X) := h(Y) - h(X) \tag{25}$$

¹ $H[X \mid X \cap Y] \geq H[X \mid (X \cap Y) \cup (Y \setminus X)] = H[X \mid Y] = H[X \cup Y \mid Y]$.

for all our set functions h .

Let M_n , and Γ_n denote the set of all (non-negative and monotone) modular and submodular set functions on \mathcal{V} , respectively. The set Γ_n is called the set of *polymatroidal functions*, or simply *polymatroids*. Let Γ_n^* denote the set of all entropic functions on n variables, and $\bar{\Gamma}_n^*$ denote its topological closure.

The notations $\Gamma_n, \Gamma_n^*, \bar{\Gamma}_n^*$ are standard in information theory. It is known [30] that Γ_n^* is a cone which is not topologically closed. And hence, when optimizing over this cone we take its topological closure $\bar{\Gamma}_n^*$, which is convex. It is easy to see that M_n and Γ_n are *polyhedral cones*. (Note that we can view them as either functions or vectors in \mathbb{R}^{2^n-1} .)

There is another interesting class of set functions called *normal functions* [2, 5], defined as follows. For every $\emptyset \neq W \subseteq [n]$, a *step function* $s_W : 2^{[n]} \rightarrow \mathbb{R}_+$ is defined by

$$s_W(X) = \begin{cases} 0 & X \subseteq W, \\ 1 & \text{otherwise.} \end{cases} \tag{26}$$

A function is normal if it is a non-negative linear combination of step functions. Let N_n denote the set of normal functions on $[n]$.

As mentioned above, entropic functions satisfy non-negativity, monotonicity, and submodularity. Linear inequalities regarding entropic functions derived from these three properties are called *Shannon-type inequalities*. For a very long time, it was widely believed that Shannon-type inequalities form a complete set of linear inequalities satisfied by entropic functions, namely $\bar{\Gamma}_n^* = \Gamma_n$. This indeed holds for $n \leq 3$, for example. However, in 1998, in a breakthrough paper in information theory, Zhang and Yeung [31] presented a new inequality which cannot be inferred from Shannon-type inequalities. Their result proved that, $\bar{\Gamma}_n^* \subsetneq \Gamma_n$ for any $n \geq 4$.

The following inclusion chain can be found in a combination of [2, 30].

Theorem 2.1. *The following chain of inclusion holds*

$$M_n \subseteq N_n \subseteq \Gamma_n^* \subseteq \bar{\Gamma}_n^* \subseteq \Gamma_n. \tag{27}$$

When $n \geq 4$, all of the containments are strict.

We intuitively explain the first two inclusions. Fix a non-empty set $W \subseteq [n]$. Let $\mathbf{b} = (b_i)_{i \in [n]}$ denote the binary vector defined by $b_i = \mathbf{1}_{i \in W}$. Let $\mathbf{1}$ denote the all-1 vector. Consider the n -dimensional distribution constructed by picking \mathbf{b} or $\mathbf{1}$ uniformly with probability 0.5. For any set X of variables, the marginal entropy on X is 0 if $X \subseteq W$, and is 1 otherwise. Thus, the step function s_W is entropic! The fact that all normal functions are entropic follows from this, with a bit of extra complication [2].

To explain the first inclusion, consider a modular function h , where

$$h(X) = \sum_{i \in X} h(\{i\}) \quad \forall X.$$

Define the normal function g by

$$g(X) := \sum_{i \in [n]} h(\{i\}) \cdot s_{[n]-\{i\}}(X), \quad X \subseteq [n]. \tag{28}$$

Since $s_{[n]-\{i\}}(X) = 1$ iff $i \in X$, it follows that $g \equiv h$, which means h is normal.

3. INFORMATION THEORETIC BOUNDS

Given a set of degree constraints DC . Recall that we write $\mathbf{D} \models \text{DC}$ to denote the fact that the database \mathbf{D} satisfies the degree constraints DC . The *cardinality bound problem* is to determine the quantity

$$\sup_{\mathbf{D} \models \text{DC}} |Q(\mathbf{D})|. \quad (29)$$

This quantity is called the (worst-case output size) of the query, over databases satisfying the input degree constraints. Algorithms evaluating Q running in time $\tilde{O}(|\mathbf{D}| + \sup_{\mathbf{D} \models \text{DC}} |Q(\mathbf{D})|)$ are called *worst-case optimal join algorithms* [26].

To obtain a bound in the general case, we employ the entropy argument, which by now is widely used in extremal combinatorics [13, 21, 29]. (The argument applied to the triangle query was already presented in Section 1.) The particular argument below can be found in the first paper mentioning Shearer's inequality [13], and a line of follow-up work [5, 6, 17, 18, 29].

Define the collection HDC of set functions satisfying the degree constraints DC

$$\text{HDC} := \{h \mid h : 2^{[n]} \rightarrow \mathbb{R}, h(Y) - h(X) \leq \log N_{Y|X}, \forall (X, Y, N_{Y|X}) \in \text{DC}\}. \quad (30)$$

Then, the entropy argument immediately gives the following result, first explicitly formulated in [6].

Theorem 3.1. (From [5, 6]) *Let Q be a conjunctive query and DC be a given set of degree constraints, then for any database \mathbf{D} satisfying DC , we have*

$$\sup_{\mathbf{D} \models \text{DC}} \log |Q(\mathbf{D})| = \max_{h \in \bar{\Gamma}_n^* \cap \text{HDC}} h([n]) \quad (\text{entropic bound}) \quad (31)$$

$$\leq \max_{h \in \Gamma_n \cap \text{HDC}} h([n]). \quad (\text{polymatroid bound}) \quad (32)$$

Furthermore, the entropic bound is asymptotically tight (equality (31) is asymptotic) and the polymatroid bound is not.

Proof. (Sketch of proof) Let $\mathbf{D} \models \text{DC}$ be any database instance satisfying the input degree constraints. Construct a distribution \mathcal{D} on $\prod_{i \in [n]} \text{Dom}(A_i)$ by picking uniformly a tuple $\mathbf{a}_{[n]}$ from the output $Q(\mathbf{D})$. Let H denote the corresponding entropy function. Then, due to uniformity we have $\log_2 |Q(\mathbf{D})| = H([n])$. Now, consider any degree constraint $(X, Y, N_{Y|X}) \in \text{DC}$. From (21) it follows that $H[Y | X] \leq \log N_{Y|X}$. Since $H \in \bar{\Gamma}_n^*$, this proves the entropic bound. The polymatroid relaxation follows from the chain of inclusion (27).

The proof that the entropic bound is asymptotically tight (i.e. the equality in (31)) is more involved, requiring a detour to group-theoretic constructions. The connection between information theory and group theory first observed in Chan and Yeung [12]. Basically, given any entropic function $h \in \bar{\Gamma}_n^* \cap \text{HDC}$, one can construct a database instance \mathbf{D} which satisfies all degree constraints DC and $\log |Q(\mathbf{D})| \geq h([n])$. The database instance is constructed from a system of (algebraic) groups derived from the entropic function. The reader is referred to [6] for more details.

The looseness of the polymatroid bound follows from exploiting the breakthrough result from Zhang and Yeung [31] on non-Shannon-type inequalities, which we already mentioned in Section 2.2. They showed that the following inequality is non-Shannon-type,

$$2I(A; B) \leq I(C; D) + I(C; A, B) + 3I(A; B | C) + I(A; B | D), \quad (33)$$

which means that *every* distribution on four variables (A, B, C, D) satisfies (33), and that there is a polymatroid in Γ_4 which does not. Before describing the polymatroid *not* satisfying (33), we first use the definition of (conditional) mutual information (24) to rearrange (33) into a form that is conducive for us to finish the proof

$$\begin{aligned} 2(h(A) - h(AB | B)) &\leq (h(C) - h(CD | D)) + (h(C) - h(ABC | AB)) \\ &\quad + 3(h(AC | C) - h(ABC | BC)) + (h(AD | D) - h(ABD | BD)), \end{aligned}$$

which is equivalent to

$$\begin{aligned} 0 &\leq 3h(AC) + 3h(AB) + 3h(BC) + h(AD) + h(BD) \\ &\quad - 2h(A) - 2h(B) - h(C) - 4h(ABC) - h(CD) - h(ABD). \end{aligned}$$

We now convert the inequality into a form we can use in the proof, by adding $11h(ABCD)$ to both sides

$$\begin{aligned} 11h(ABCD) &\leq 3h(AC) + 3h(AB) + 3h(BC) + h(AD) + h(BD) \\ &\quad + 2h(ABCD | A) + 2h(ABCD | B) + h(ABCD | C) \\ &\quad + 4h(ABCD | ABC) + h(ABCD | CD) + h(ABCD | ABD). \end{aligned} \quad (34)$$

Note again that (34) is completely equivalent to (33), which means every $h \in \Gamma_n^*$ satisfies (34). Furthermore, they showed that, for any $\theta > 0$, the following function in Γ_4 does *not* satisfy (34)

$$h_\theta(X) = 2\theta \quad \forall X \in \{A, B, C, D\}, \quad (35)$$

$$h_\theta(CD) = 3\theta, \quad (36)$$

$$h_\theta(XY) = 2\theta \quad \forall X, Y \in \{A, B, C, D\}, |\{X, Y\}| = 2, XY \neq CD, \quad (37)$$

$$h_\theta(ABCD) = h_\theta(XYZ) = 4\theta \quad \forall X, Y, Z \in \{A, B, C, D\}, |\{X, Y, Z\}| = 3. \quad (38)$$

It is not hard, although somewhat tedious, to show that h_θ is a polymatroid.

Now, to complete the proof that the polymatroid bound is not tight, consider the following set of degree constraints, which are read off from (34) *and* from the values of h_θ :

- Cardinality constraints (\emptyset, XY, N^2) , for $XY \in \{AC, AB, BC, AD, BD\}$,
- Degree constraints $(X, ABCD, N^2)$, for $X \in \{A, B, C\}$,
- Degree constraint $(CD, ABCD, N)$.
- Functional dependency constraints $(ABC, ABCD, 1)$ and $(ABD, ABCD, 1)$.

For this set of constraints, from (31) and (34) we have

$$\log |Q| = H[ABCD] \leq \frac{1}{11} (22 \times \log N + 10 \times \log N + \log N) = 3 \log N. \quad (39)$$

On the other hand, the function h_θ with $\theta = \log N$ is a solution to (32) with objective value $h_\theta(ABCD) = 4 \log N$, proving the gap. As $N \rightarrow \infty$, the gap can grow arbitrary large. ■

The entropic bound is tight; unfortunately it is not known whether it is even computable [3]. The main reason is that the geometric object $\bar{\Gamma}_n^*$ is highly complex, requiring an infinite number of inequalities to characterize [24]. The characterization of $\bar{\Gamma}_n^*$ is an active and fascinating research topic [11, 15].

The polymatroid bound is computable, because (32) is just a linear program. However, it raises three natural questions:

- What is the exact computational complexity of computing the polymatroid bound?
- For which class of input degree constraints can we compute the polymatroid bound more efficiently than solving the linear program (32)? Preferrably, we'd like to compute it in polynomial time.
- For which class of input degree constraints is the polymatroid bound (32) tight?

For the first question, it is not known whether the problem is NP-hard. We do know partial answers to the next two questions, in the sense that there are sufficient conditions under which we can compute the polymatroid bound in poly-time *and* for which the bound are is tight. We explore these partial answers in the next section.

4. COMPUTATIONAL ISSUES

This section focuses on cases where we can compute the polymatroid bound (32) more efficiently. While we can expect the number of degree constraints to be small (in practice), say a polynomial in n , the number of submodularity constraints is exponential. We can reduce the number of constraints a little, by observing that monotonicity and submodularity can be replaced equivalently by the following

$$h(X) \leq h(X \cup \{i\}) \quad \forall X \in 2^{[n]} \text{ and } i \in [n] \setminus X, \quad (40)$$

$$h(X \cup \{i, j\}) + h(X) \leq h(X \cup \{i\}) + h(X \cup \{j\}) \quad \forall X \in 2^{[n]}, \text{ and } i, j \in [n] \setminus X. \quad (41)$$

(These imply all other monotonicity and submodularity constraints). Even with the reduction, there still are $O(2^n n^2)$ many constraints. We explore next some sufficient conditions under which we can reduce the number of constraints further.

4.1. Useful sub-classes of polymatroids

From the chain of inclusion (27), we know the following

$$\sup\{h([n]) \mid h \in \text{HDC} \cap \Gamma_n\} \quad \text{polymatroid bound, not tight!} \quad (42)$$

$$\geq \sup\{h([n]) \mid h \in \text{HDC} \cap \Gamma_n^*\} \quad \text{entropic bound, tight!} \quad (43)$$

$$\geq \sup\{h([n]) \mid h \in \text{HDC} \cap \mathbf{N}_n\} \quad (44)$$

$$\geq \sup\{h([n]) \mid h \in \text{HDC} \cap \mathbf{M}_n\}. \quad (45)$$

The nice thing about the sets \mathbf{N}_n and \mathbf{M}_n is that they require less number of (linear) constraints to describe. Furthermore, if we can show that (44) and (42) collapse, then the polymatroid bound is tight! Similarly, if (45) and (42) collapse, then the bound is tight. We next prove two sufficient conditions on the input degree constraints so that these bounds collapse.

The first result is implicit in [2]. A set DC of degree constraints is *simple* if $(X, Y, N_{Y|X}) \in \text{DC}$ implies $|X| \leq 1$.

Theorem 4.1. (From [2]) *Let DC be a set of simple degree constraints, then*

$$\sup\{h([n]) \mid h \in \text{HDC} \cap \Gamma_n\} = \sup\{h([n]) \mid h \in \text{HDC} \cap \mathbf{N}_n\} \quad (46)$$

In particular, the polymatroid bound is tight for this class of degree constraints.

Proof. (Sketch of proof) Let h be an arbitrary polymatroid. In [2] we showed that we can always construct a normal function g for which

$$g([n]) = h([n]), \tag{47}$$

$$g(\{i\}) = h(\{i\}) \quad \forall i \in [n], \tag{48}$$

$$g(X) \leq h(X) \quad \forall X \subseteq [n]. \tag{49}$$

Thus, if $h \in \text{HDC}$ then $g \in \text{HDC}$ because $g(Y | X) \leq h(Y | X)$ for all $|X| \leq 1$. This means, for any solution h to the polymatroid linear program, we can construct a solution g to the normal linear program with the same cost. ■

How much did we save by assuming DC is simple? The description of members of \mathbf{N}_n requires $O(2^n)$ non-negative coefficients λ_W , where

$$h(X) = \sum_{W \subseteq [n]} \lambda_W \cdot s_W(X) = \sum_{X \not\subseteq W} \lambda_W. \tag{50}$$

Hence, computing the normal bound (44) is to solve the linear program

$$\max \quad \lambda_{[n]} \tag{51}$$

$$\text{s.t.} \quad \sum_{W: Y \not\subseteq W, X \subseteq W} \lambda_W \leq \log_2 N_{Y|X}, \quad (X, Y, N_{Y|X}) \in \text{DC}, \tag{52}$$

$$\boldsymbol{\lambda} \geq \mathbf{0}. \tag{53}$$

This linear program has size roughly $O(2^n)$ (modulo the degree constraints). This is an improvement over $O(2^n n^2)$, but probably not enough for practical usage.

The second result is implicit in [26]. A set DC of degree constraints is *acyclic* if there exists a permutation π of $[n]$ for which every member of X precedes every member of $Y - X$ in the permutation, for every degree constraint $(X, Y, N_{Y|X}) \in \text{DC}$. In that case we say that DC is *compatible* with π .

Theorem 4.2. (From [26]) *Let DC be a set of acyclic degree constraints, then*

$$\sup\{h([n]) \mid h \in \text{HDC} \cap \Gamma_n\} = \sup\{h([n]) \mid h \in \text{HDC} \cap \mathbf{M}_n\}. \tag{54}$$

In particular, the polymatroid bound is tight for acyclic degree constraints.

Proof. Let h be an arbitrary polymatroid. It is sufficient to construct a modular function g for which the following hold

$$g([n]) = h([n]), \tag{55}$$

$$g(Y|X) \leq h(Y|X) \quad \forall (X, Y, N_{Y|X}) \in \text{DC}. \tag{56}$$

Without loss of generality, assume the identity permutation is compatible with DC, i.e. for every $(X, Y, N_{Y|X}) \in \text{DC}$, we have $x < y$ for all $x \in X$ and $y \in Y - X$. Define

$$g(S) := \begin{cases} 0 & \text{if } S = \emptyset, \\ h([i]) - h([i-1]) & \text{if } S = \{i\}, i \in [n], \\ \sum_{i \in S} g(i) & \text{if } S \subseteq [n], |S| > 1. \end{cases} \tag{57}$$

The function g is clearly modular because h is monotone. The fact that $g([n]) = h([n])$ follows from the telescoping sum. We will show by induction on $|Y - X|$ that $g(Y|X) \leq h(Y|X)$ for any degree constraint $(X, Y, N_{Y|X}) \in \text{DC}$. The base case when $Y = X$ holds

trivially. Let $(X, Y, N_{Y|X})$ be any degree constraint in DC where $|Y - X| > 0$. Let j be the largest integer in $Y - X$. We have

$$g(Y | X) = h([j] | [j - 1]) + g(Y - \{j\} | X) \quad (58)$$

$$\text{(induction hypothesis)} \leq h([j] | [j - 1]) + h(Y - \{j\} | X) \quad (59)$$

$$= h([j - 1] \cup Y | [j - 1]) + h(Y - \{j\} | X) \quad (60)$$

$$\text{(submodularity of } h) \leq h(Y | Y \cap [j - 1]) + h(Y - \{j\} | X) \quad (61)$$

$$= h(Y | Y - \{j\}) + h(Y - \{j\} | X) \quad (62)$$

$$= h(Y | X). \quad (63)$$

■

Acyclic degree constraints are thus very nice, because the size of the modular linear program (45) is polynomial in n . Explicitly, setting $\nu_i = h(\{i\})$ the linear program is

$$\begin{aligned} \max \quad & \sum_{i \in [n]} \nu_i & (64) \\ \text{s.t.} \quad & \sum_{i \in Y \setminus X} \nu_i \leq \log_2 N_{Y|X} & (X, Y, N_{Y|X}) \in \text{DC} \\ & \boldsymbol{\nu} \geq \mathbf{0}. \end{aligned}$$

Note that if the set DC contains only cardinality constraints, i.e. $X = \emptyset$ for all $(X, Y, N_{Y|X}) \in \text{DC}$, then DC is trivially acyclic. The triangle query example (1) has acyclic constraints, because all we had was the cardinality constraint on $|E|$.

4.2. Connection to geometric inequalities and information inequalities

Studying the linear program (64) leads to further interesting connections. Associate a dual variable $\delta_{Y|X}$ for every $(X, Y, N_{Y|X}) \in \text{DC}$. In what follows for brevity we sometimes write $(X, Y) \in \text{DC}$ instead of the lengthier $(X, Y, N_{Y|X}) \in \text{DC}$. The dual LP of (64) is

$$\begin{aligned} \min \quad & \sum_{(X, Y, N_{Y|X}) \in \text{DC}} \delta_{Y|X} \log_2 N_{Y|X} & (65) \\ \text{s.t.} \quad & \sum_{\substack{(X, Y) \in \text{DC} \\ i \in Y - X}} \delta_{Y|X} \geq 1 & \forall i \in [n] \\ & \delta_{Y|X} \geq 0 & \forall (X, Y) \in \text{DC}. \end{aligned}$$

Weak duality and Theorem 4.1 leads to the following simple observation.

Proposition 4.1. *Given a query Q with a set DC of acyclic degree constraints, then*

$$|Q| \leq \prod_{(X, Y, N_{Y|X}) \in \text{DC}} N_{Y|X}^{\delta_{Y|X}} \quad (66)$$

for any vector $\boldsymbol{\delta}$ which is feasible to (65).

When DC contains only cardinality constraint (and thus it is acyclic), the above proposition specialises to the following:

- The Bollobás-Thomason inequality [9] for discrete measures. The statement of the bound is simpler when $X = \emptyset$ for all degree constraints. We write N_Y instead of $N_{Y|\emptyset}$, and δ_Y instead of $\delta_{Y|\emptyset}$. Let $\mathcal{H} = ([n], \mathcal{E})$ be the hypergraph whose edges are exactly

the Y from the degree constraints (\emptyset, Y, N_Y) . Then, δ is a feasible solution to (65) if and only if it is a fractional edge cover of \mathcal{H} . Bollobás-Thomason proved that

$$|Q| \leq \prod_{Y \in \mathcal{E}} N_Y^{\delta_Y}, \quad (67)$$

whenever δ is a fractional edge cover for \mathcal{E} . The statement in their paper is a little different, but can be shown easily to be equivalent to the above.

- Furthermore, if we restrict the instance to degree constraints of the form (\emptyset, Y, N_Y) , with $|Y| = n - 1$, then we get the Loomis-Whitney inequality for discrete measures. The Loomis-Whitney bound is often stated with a specific fractional edge cover $\delta_Y = 1/(n - 1)$ for all $Y \in \binom{[n]}{n-1}$.
- The bound (67) is now known as the AGM-bound in database theory [8].

5. CONCLUSIONS

We surveyed recent developments in the connection between information theory and database theory. There are many exciting discoveries which we were not able to cover within the scope of this article. There also are many interesting open problems arising naturally from the setup presented here.

Our work [5, 6] developed the family of information-theoretic bounds for a much more general class of queries called *disjunctive datalog* queries. The connection to information inequalities such as Shearer's lemma and its generalizations is established in a systematic manner. Information theoretic inequalities also have other applications in databases [2] and beyond in Computer Science. In [3] we formulated a large class of (decision) problems in information theory with a wide range of applications in various areas of Computer Science.

The problems related to efficiently computing the polymatroid bound are open. It is not known whether computing the bound is NP-hard. We believe it is NP-hard. Then, the problems are to identify sub-classes of input degree constraints for which the bound can be efficiently computed. We took the first steps in Section 4., but much is left to be desired.

In addition to deriving cardinality bounds, query optimization and evaluation algorithms meeting those bounds are deep and exciting subjects of algorithmic research, where there is great synergy between statistics (graphical models), constraint satisfaction, logic, and database. The reader is referred to [4, 26] and references thereof for more information.

The entropy argument we used is only one form of entropy argument. By now this argument is widely used in extremal combinatorics to prove a wide variety of results [21]. Most closely related to our setting is an inequality from FriedGut which uses a more refined argument – moving beyond the uniform distribution – but otherwise similar to the one we presented [16].

These results are used in the implementation of a production-grade query optimizer, illustrating a beautiful application of Mathematics to a production system. I am certain that Prof. Phan would be proud of such applications.

REFERENCES

- [1] S. Abiteboul, R. Hull, and V. Vianu, *Foundations of Databases*. Addison-Wesley, 1995.

- [2] M. Abo Khamis, P. G. Kolaitis, H. Q. Ngo, and D. Suciu, “Bag query containment and information theory,” in *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020*, D. Suciu, Y. Tao, and Z. Wei, Eds. ACM, 2020, pp. 95–112. [Online]. Available: <https://doi.org/10.1145/3375395.3387645>
- [3] —, “Decision problems in information theory,” in *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, ser. LIPIcs, A. Czumaj, A. Dawar, and E. Merelli, Eds., vol. 168. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 106:1–106:20. [Online]. Available: <https://doi.org/10.4230/LIPIcs.ICALP.2020.106>
- [4] M. Abo Khamis, H. Q. Ngo, and A. Rudra, “FAQ: questions asked frequently,” in *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, T. Milo and W. Tan, Eds. ACM, 2016, pp. 13–28. [Online]. Available: <http://doi.acm.org/10.1145/2902251.2902280>
- [5] M. Abo Khamis, H. Q. Ngo, and D. Suciu, “Computing join queries with functional dependencies,” in *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, T. Milo and W. Tan, Eds. ACM, 2016, pp. 327–342. [Online]. Available: <http://doi.acm.org/10.1145/2902251.2902289>
- [6] —, “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, Chicago, IL, USA, May 14-19, 2017*, E. Sallinger, J. V. den Bussche, and F. Geerts, Eds. ACM, 2017, pp. 429–444. [Online]. Available: <http://doi.acm.org/10.1145/3034786.3056105>
- [7] N. Alon, “On the number of subgraphs of prescribed type of graphs with a given number of edges,” *Israel J. Math.*, vol. 38, no. 1-2, pp. 116–130, 1981. [Online]. Available: <http://dx.doi.org.gate.lib.buffalo.edu/10.1007/BF02761855>
- [8] A. Atserias, M. Grohe, and D. Marx, “Size bounds and query plans for relational joins,” in *FOCS*. IEEE Computer Society, 2008, pp. 739–748.
- [9] B. Bollobás and A. Thomason, “Projections of bodies and hereditary properties of hypergraphs,” *Bull. London Math. Soc.*, vol. 27, no. 5, pp. 417–424, 1995. [Online]. Available: <https://doi-org.gate.lib.buffalo.edu/10.1112/blms/27.5.417>
- [10] W. Cai, M. Balazinska, and D. Suciu, “Pessimistic cardinality estimation: Tighter upper bounds for intermediate join cardinalities,” in *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*, P. A. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, and T. Kraska, Eds. ACM, 2019, pp. 18–35. [Online]. Available: <https://doi.org/10.1145/3299869.3319894>
- [11] T. Chan, “Recent progresses in characterising information inequalities,” *Entropy*, vol. 13, no. 2, pp. 379–401, 2011. [Online]. Available: <https://doi-org.gate.lib.buffalo.edu/10.3390/e13020379>
- [12] T. H. Chan and R. W. Yeung, “On a relation between information inequalities and group theory,” *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1992–1995, 2002. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2002.1013138>

- [13] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer, “Some intersection theorems for ordered sets and graphs,” *J. Combin. Theory Ser. A*, vol. 43, no. 1, pp. 23–37, 1986. [Online]. Available: [http://dx.doi.org.gate.lib.buffalo.edu/10.1016/0097-3165\(86\)90019-1](http://dx.doi.org.gate.lib.buffalo.edu/10.1016/0097-3165(86)90019-1)
- [14] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2006.
- [15] I. Csiszár, “Axiomatic characterizations of information measures,” *Entropy*, vol. 10, no. 3, pp. 261–273, 2008. [Online]. Available: <https://doi.org/10.3390/e10030261>
- [16] E. Friedgut, “Hypergraphs, entropy, and inequalities,” *Amer. Math. Monthly*, vol. 111, no. 9, pp. 749–760, 2004. [Online]. Available: <https://doi-org.gate.lib.buffalo.edu/10.2307/4145187>
- [17] E. Friedgut and J. Kahn, “On the number of copies of one hypergraph in another,” *Israel J. Math.*, vol. 105, pp. 251–256, 1998. [Online]. Available: <http://dx.doi.org.gate.lib.buffalo.edu/10.1007/BF02780332>
- [18] G. Gottlob, S. T. Lee, G. Valiant, and P. Valiant, “Size and treewidth bounds for conjunctive queries,” *J. ACM*, vol. 59, no. 3, p. 16, 2012. [Online]. Available: <http://doi.acm.org/10.1145/2220357.2220363>
- [19] H. Harmouch and F. Naumann, “Cardinality estimation: An experimental survey,” *Proc. VLDB Endow.*, vol. 11, no. 4, pp. 499–512, 2017. [Online]. Available: <http://www.vldb.org/pvldb/vol11/p499-harmouch.pdf>
- [20] A. Hertzschuch, C. Hartmann, D. Habich, and W. Lehner, “Simplicity done right for join ordering,” in *11th Conference on Innovative Data Systems Research, CIDR 2021, Virtual Event, January 11-15, 2021, Online Proceedings*. www.cidrdb.org, 2021. [Online]. Available: http://cidrdb.org/cidr2021/papers/cidr2021_paper01.pdf
- [21] S. Jukna, *Extremal combinatorics*, 2nd ed., ser. Texts in Theoretical Computer Science. An EATCS Series. Springer, Heidelberg, 2011, with applications in computer science. [Online]. Available: <https://doi-org.gate.lib.buffalo.edu/10.1007/978-3-642-17364-6>
- [22] V. Leis, B. Radke, A. Gubichev, A. Kemper, and T. Neumann, “Cardinality estimation done right: Index-based join sampling,” in *8th Biennial Conference on Innovative Data Systems Research, CIDR 2017, Chaminade, CA, USA, January 8-11, 2017, Online Proceedings*. www.cidrdb.org, 2017. [Online]. Available: <http://cidrdb.org/cidr2017/papers/p9-leis-cidr17.pdf>
- [23] L. H. Loomis and H. Whitney, “An inequality related to the isoperimetric inequality,” *Bull. Amer. Math. Soc.*, vol. 55, pp. 961–962, 1949.
- [24] F. Matús, “Infinitely many information inequalities,” in *IEEE International Symposium on Information Theory, ISIT 2007, Nice, France, June 24-29, 2007*. IEEE, 2007, pp. 41–44. [Online]. Available: <https://doi.org/10.1109/ISIT.2007.4557201>
- [25] T. Milo and W. Tan, Eds., *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. ACM, 2016. [Online]. Available: <http://doi.acm.org/10.1145/2902251>
- [26] H. Q. Ngo, “Worst-case optimal join algorithms: Techniques, results, and open problems,” in *Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Houston, TX, USA, June 10-15, 2018*, J. V. den Bussche and M. Arenas, Eds. ACM, 2018, pp. 111–124. [Online]. Available: <https://doi.org/10.1145/3196959.3196990>

- [27] H. Q. Ngo, E. Porat, C. Ré, and A. Rudra, “Worst-case optimal join algorithms: [extended abstract],” in *PODS*, 2012, pp. 37–48.
- [28] H. Q. Ngo, C. Ré, and A. Rudra, “Skew strikes back: new developments in the theory of join algorithms,” *SIGMOD Record*, vol. 42, no. 4, pp. 5–16, 2013. [Online]. Available: <http://doi.acm.org/10.1145/2590989.2590991>
- [29] J. Radhakrishnan, “Entropy and counting,” in *Computational Mathematics, Modelling and Algorithms*, J. C. Misra, Ed. Narosa Pub House, 2003, pp. 146–168.
- [30] R. W. Yeung, *Information Theory and Network Coding*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [31] Z. Zhang and R. W. Yeung, “On characterization of entropy function via information inequalities,” *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1440–1452, 1998. [Online]. Available: <http://dx.doi.org/10.1109/18.681320>

Received on June 08, 2021
Accepted on August 18, 2021