CSE599d: Advanced Query Processing

Lecture 17: Entropic Inequalities

Dan Suciu

University of Washington

Degree Sequencees

Overview

- AGM bound is $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})|$; tight up to a “rounding factor” $1/2^n$.
- In practice, bound is too loose, because cardinalities are insufficient.
- More general statistics: degree sequences.

Local Statistics

Cardinality of each table: $|R|, |S|, \ldots$
Domain sizes of attributes: $|R.X|, |R.Y|, \ldots, |S.Z|, \ldots$

Local Statistics

Cardinality of each table: $|R|, |S|, \ldots$
Domain sizes of attributes: $|R.X|, |R.Y|, \ldots, |S.Z|, \ldots$

Denote $\boxed{\deg_R(*|X) \stackrel{\text{def}}{=} (f_1, f_2, \ldots, f_n)}$ the frequencies (degrees) of $R.X$.

Local Statistics

Cardinality of each table: $|R|, |S|, \ldots$
Domain sizes of attributes: $|R.X|, |R.Y|, \ldots, |S.Z|, \ldots$

Denote $\boxed{\deg_R(*|X) \stackrel{\text{def}}{=} (f_1, f_2, \ldots, f_n)}$ the frequencies (degrees) of $R.X$.

Local statistics: a collection of frequency moments $\boxed{||\deg_R(*|X)||_p^p = \sum_i f_i^p}$

The Statistics: Examples

$R =$

X	Y	Z
1	a	10
1	b	20
1	b	30
2	a	40
2	b	50
3	b	60
3	c	70
4	d	80

Degree sequences:
$\text{deg}(*|X) = (3, 2, 2, 1)$

The Statistics: Examples

$R =$

X	Y	Z
1	a	10
1	b	20
1	b	30
2	a	40
2	b	50
3	b	60
3	c	70
4	d	80

Degree sequences:
$\text{deg}(*|X) = (3, 2, 2, 1)$

$\text{deg}(*|Y) = (4, 2, 1, 1)$.

$\text{deg}(*|XY) = (2, 1, 1, 1, 1, 1, 1)$

$\text{deg}(*|\emptyset) = (8)$

$\text{deg}(Y|X) = (2, 2, 2, 1)$

The Statistics: Examples

$R =$

X	Y	Z
1	a	10
1	b	20
1	b	30
2	a	40
2	b	50
3	b	60
3	c	70
4	d	80

$||\text{deg}_R(*|X)||_1$ = Cardinality $|R| = 8$

$||\text{deg}_R(X|\emptyset)||_1$ = Domain size $|R.X| = 4$

Degree sequences:

$\text{deg}(*|X) = (3, 2, 2, 1)$

$\text{deg}(*|Y) = (4, 2, 1, 1)$.

$\text{deg}(*|XY) = (2, 1, 1, 1, 1, 1, 1)$

$\text{deg}(*|\emptyset) = (8)$

$\text{deg}(Y|X) = (2, 2, 2, 1)$

The Statistics: Examples

$R =$

X	Y	Z
1	a	10
1	b	20
1	b	30
2	a	40
2	b	50
3	b	60
3	c	70
4	d	80

$||\text{deg}_R(*|X)||_1$ = Cardinality $\quad |R| = 8$

$||\text{deg}_R(X|\emptyset)||_1$ = Domain size $\quad |R.X| = 4$

$||\text{deg}(*|X)||_\infty$ = Max frequency $\quad 3$

Degree sequences:

$\text{deg}(*|X) = (3, 2, 2, 1)$

$\text{deg}(*|Y) = (4, 2, 1, 1).$

$\text{deg}(*|XY) = (2, 1, 1, 1, 1, 1, 1)$

$\text{deg}(*|\emptyset) = (8)$

$\text{deg}(Y|X) = (2, 2, 2, 1)$

The Statistics: Examples

$R =$

X	Y	Z
1	a	10
1	b	20
1	b	30
2	a	40
2	b	50
3	b	60
3	c	70
4	d	80

$||\text{deg}_R(*|X)||_1$ = Cardinality $|R| = 8$

$||\text{deg}_R(X|\emptyset)||_1$ = Domain size $|R.X| = 4$

$||\text{deg}(*|X)||_\infty$ = Max frequency 3

$||\text{deg}(*|X)||_2^2, ||\text{deg}(*|X)||_3^3$ = Various moments

Degree sequences:

$\text{deg}(*|X) = (3, 2, 2, 1)$

$\text{deg}(*|Y) = (4, 2, 1, 1)$.

$\text{deg}(*|XY) = (2, 1, 1, 1, 1, 1, 1)$

$\text{deg}(*|\emptyset) = (8)$

$\text{deg}(Y|X) = (2, 2, 2, 1)$

The Statistics: Examples

$R =$

X	Y	Z
1	a	10
1	b	20
1	b	30
2	a	40
2	b	50
3	b	60
3	c	70
4	d	80

$||\text{deg}_R(*|X)||_1$ = Cardinality $\quad |R| = 8$

$||\text{deg}_R(X|\emptyset)||_1$ = Domain size $\quad |R.X| = 4$

$||\text{deg}(*|X)||_\infty$ = Max frequency $\quad 3$

$||\text{deg}(*|X)||_2^2, ||\text{deg}(*|X)||_3^3$ = Various moments

Degree sequences:

$\text{deg}(*|X) = (3, 2, 2, 1)$

$\text{deg}(*|Y) = (4, 2, 1, 1)$.

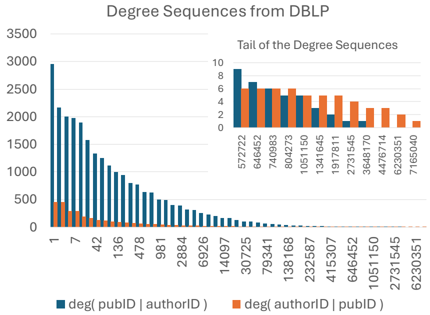
$\text{deg}(*|XY) = (2, 1, 1, 1, 1, 1, 1)$

$\text{deg}(*|\emptyset) = (8)$

$\text{deg}(Y|X) = (2, 2, 2, 1)$

Local statistics: $p = 1, 2, \ldots, 10, \infty$ moments

Frequencies Matter



DBLP in 2024:

- $|\text{DBLP}| = 24 \cdot 10^6$
- Most prolific author: `H. Vincent Poor` 2951 pubs
- Most authors per paper: 450 [Srivastava et al., 2023]

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$$|Q| \leq |R| \cdot |S|$$

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$$|Q| \leq |R| \cdot |S|$$

$S.Y$ is a key: $|Q| \leq |R|$

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$$|Q| \leq |R| \cdot |S|$$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\text{deg}_S(Z|Y)||_\infty$

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$$|Q| \leq |R| \cdot |S|$$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\text{deg}_S(Z|Y)||_\infty$

Cauchy-Schwartz:

$$|Q| \leq ||\text{deg}_R(X|Y)||_2 \cdot ||\text{deg}_S(Z|Y)||_2$$

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$$|Q| \leq |R| \cdot |S|$$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\text{deg}_S(Z|Y)||_\infty$

Cauchy-Schwartz:

$$|Q| \leq ||\text{deg}_R(X|Y)||_2 \cdot ||\text{deg}_S(Z|Y)||_2$$

$Q = R(X, Y) \bowtie S(Y, Z) \bowtie T(Z, X)$

Upper bounds on $|Q|$:

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$|Q| \leq |R| \cdot |S|$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\text{deg}_S(Z|Y)||_\infty$

Cauchy-Schwartz:
$|Q| \leq ||\text{deg}_R(X|Y)||_2 \cdot ||\text{deg}_S(Z|Y)||_2$

$Q = R(X, Y) \bowtie S(Y, Z) \bowtie T(Z, X)$

Upper bounds on $|Q|$:

$(|R| \cdot |S| \cdot |T|)^{1/2}$ (AGM bound)

Upper Bounds: Examples

$Q = R(X,Y) \bowtie S(Y,Z)$

$$|Q| \leq |R| \cdot |S|$$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\text{deg}_S(Z|Y)||_\infty$

Cauchy-Schwartz:
$$|Q| \leq ||\text{deg}_R(X|Y)||_2 \cdot ||\text{deg}_S(Z|Y)||_2$$

$Q = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$

Upper bounds on $|Q|$:

$$(|R| \cdot |S| \cdot |T|)^{1/2}$$ (AGM bound)

$$\left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$$

Upper Bounds: Examples

$Q = R(X,Y) \bowtie S(Y,Z)$

$$|Q| \leq |R| \cdot |S|$$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\text{deg}_S(Z|Y)||_\infty$

Cauchy-Schwartz:
$|Q| \leq ||\text{deg}_R(X|Y)||_2 \cdot ||\text{deg}_S(Z|Y)||_2$

$Q = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$

Upper bounds on $|Q|$:

$$(|R| \cdot |S| \cdot |T|)^{1/2} \qquad \text{(AGM bound)}$$

$$\left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$$

$$\left(||\text{deg}_R(Y|X)||_3^3 \cdot ||\text{deg}_S(Y|Z)||_3^3 \cdot |T|^5\right)^{1/6}$$

Upper Bounds: Examples

$Q = R(X, Y) \bowtie S(Y, Z)$

$|Q| \leq |R| \cdot |S|$

$S.Y$ is a key: $|Q| \leq |R|$

General: $|Q| \leq |R| \cdot ||\deg_S(Z|Y)||_\infty$

Cauchy-Schwartz:
$|Q| \leq ||\deg_R(X|Y)||_2 \cdot ||\deg_S(Z|Y)||_2$

$Q = R(X, Y) \bowtie S(Y, Z) \bowtie T(Z, X)$

Upper bounds on $|Q|$:

$(|R| \cdot |S| \cdot |T|)^{1/2}$ (AGM bound)

$\left(||\deg_R(Y|X)||_2^2 \cdot ||\deg_S(Z|Y)||_2^2 \cdot ||\deg_T(X|Z)||_2^2\right)^{1/3}$

$\left(||\deg_R(Y|X)||_3^3 \cdot ||\deg_S(Y|Z)||_3^3 \cdot |T|^5\right)^{1/6}$

Many others Take the min!

Discussion

Pessimistic Cardinality Estimator:

- One-sided theoretical guarantee.[1]
- Uses local statistics on input tables.
- Easy to combine: just take their min.

How do we derive and compute these upper bounds?

[1] Why call it "estimator"? Term coined by [Cai et al., 2019].

Entropic Inequalities

Entropic Vectors

Random variable X with finite outcomes $X \in D$. Its entropy is:

$$\boxed{h(X) \stackrel{\text{def}}{=} -\sum_{x \in D} p(x) \log p(x)} \in \mathbb{R}_+$$

Entropic Vectors

Random variable X with finite outcomes $X \in D$. Its entropy is:

$$\boxed{h(X) \stackrel{\text{def}}{=} -\sum_{x \in D} p(x) \log p(x)} \in \mathbb{R}_+$$

$0 \leq h(X) \leq \log |D|$

Entropic Vectors

Random variable X with finite outcomes $X \in D$. Its entropy is:

$$\boxed{h(X) \stackrel{\text{def}}{=} -\sum_{x \in D} p(x) \log p(x)} \in \mathbb{R}_+$$

$0 \leq h(X) \leq \log |D|$

Random vars $X_1, \ldots, X_n$. Their *entropic vector*: $\boxed{\boldsymbol{h} = (h(X_\alpha))_{\alpha \subseteq [n]} \in \mathbb{R}_+^{2^n}}$

Entropic Vectors

Random variable X with finite outcomes $X \in D$. Its entropy is:

$$\boxed{h(X) \stackrel{\text{def}}{=} -\sum_{x \in D} p(x) \log p(x)} \in \mathbb{R}_+$$

$0 \leq h(X) \leq \log |D|$

Random vars $X_1, \ldots, X_n$. Their *entropic vector*: $\boxed{\boldsymbol{h} = (h(X_\alpha))_{\alpha \subseteq [n]} \in \mathbb{R}_+^{2^n}}$

X	Y
a	p
a	q
b	q
a	m

Entropic Vectors

Random variable X with finite outcomes $X \in D$. Its entropy is:

$$\boxed{h(X) \stackrel{\text{def}}{=} -\sum_{x \in D} p(x) \log p(x)} \in \mathbb{R}_+$$

$0 \leq h(X) \leq \log |D|$

Random vars $X_1, \ldots, X_n$. Their *entropic vector*: $\boxed{\boldsymbol{h} = (h(X_\alpha))_{\alpha \subseteq [n]} \in \mathbb{R}_+^{2^n}}$

X	Y	p
a	p	$1/4$
a	q	$1/4$
b	q	$1/4$
a	m	$1/4$

$h(XY) = \log 4$

Entropic Vectors

Random variable X with finite outcomes $X \in D$. Its entropy is:

$$\boxed{h(X) \stackrel{\text{def}}{=} -\sum_{x \in D} p(x) \log p(x)} \in \mathbb{R}_+$$

$0 \leq h(X) \leq \log |D|$

Random vars $X_1, \ldots, X_n$. Their *entropic vector*: $\boxed{\boldsymbol{h} = (h(X_\alpha))_{\alpha \subseteq [n]} \in \mathbb{R}_+^{2^n}}$

X	Y	p
a	p	$1/4$
a	q	$1/4$
b	q	$1/4$
a	m	$1/4$

$h(XY) = \log 4$

X	p
a	$3/4$
b	$1/4$

$h(X) \leq \log 2$

Y	p
p	$1/4$
q	$2/4$
m	$1/4$

$h(Y) \leq \log 3$

$\emptyset$	p
	1

$h(\emptyset) = 0$

Shannon Inequalities

Basic Shannon Inequalities

$$h(U \cup V) \geq h(U) \qquad \text{Monotonicity}$$

$$h(U) + h(V) \geq h(U \cup V) + h(U \cap V) \qquad \text{Submodularity}$$

A Shannon inequality is a consequence of these inequalities.

Example

$$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$$

Example

$$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$$

$$h(XY) + h(YZ) + h(XZ)$$

Example

$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$

$$\underline{h(XY) + h(YZ)} + h(XZ)$$

Example

$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$

$$\underline{h(XY) + h(YZ)} + h(XZ)$$
$$\geq h(XYZ) + h(Y) + h(XZ)$$

Example

$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$

$$\underline{h(XY) + h(YZ)} + h(XZ) \geq h(XYZ) + \underline{h(Y) + h(XZ)}$$

Example

$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$

$$
\begin{aligned}
\underline{h(XY) + h(YZ)} &+ h(XZ) \\
&\geq h(XYZ) + \underline{h(Y) + h(XZ)} \\
&\geq h(XYZ) + h(XYZ) + h(\emptyset)
\end{aligned}
$$

Example

$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$

$$\begin{aligned}
\underline{h(XY) + h(YZ)} + h(XZ) & \\
&\geq h(XYZ) + \underline{h(Y) + h(XZ)} \\
&\geq h(XYZ) + h(XYZ) + h(\emptyset) \\
&= 2h(XYZ)
\end{aligned}$$

Example

$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$

$$\begin{aligned}
\underline{h(XY) + h(YZ)} + h(XZ) & \\
&\geq h(XYZ) + \underline{h(Y) + h(XZ)} \\
&\geq h(XYZ) + h(XYZ) + h(\emptyset) \\
&= 2h(XYZ)
\end{aligned}$$

We use Shannon inequalities to derive upper bounds

Example

$$Q(X,Y,Z) = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$$

$$\boxed{|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}}$$

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Example

$Q(X,Y,Z) = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$

$$|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}$$

Define the uniform distribution on the output Q:

Q

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Example

$Q(X, Y, Z) = R(X, Y) \bowtie S(Y, Z) \bowtie T(Z, X)$

$|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}$

Define the uniform distribution on the output Q:

Q

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

R

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S

Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5

T

X	Z	
a	m	1/5
a	q	2/5
b	q	1/5
d	m	1/5

$\log |R|$

$\geq h(XY)$

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Example

$Q(X,Y,Z) = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$

$|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}$

Define the uniform distribution on the output Q:

Q

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

R

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S

Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5

T

X	Z	
a	m	1/5
a	q	2/5
b	q	1/5
d	m	1/5

$$\log|R| + \log|S| + \log|T|$$
$$\geq h(XY) + h(YZ) + h(XZ)$$

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Example

$Q(X,Y,Z) = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$

$|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}$

Define the uniform distribution on the output Q:

Q

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

R

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S

Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5

T

X	Z	
a	m	1/5
a	q	2/5
b	q	1/5
d	m	1/5

$$\log|R| + \log|S| + \log|T|$$
$$\geq h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$$

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Example

$Q(X, Y, Z) = R(X, Y) \bowtie S(Y, Z) \bowtie T(Z, X)$

$|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}$

Define the uniform distribution on the output Q:

Q

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

R

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S

Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5

T

X	Z	
a	m	1/5
a	q	2/5
b	q	1/5
d	m	1/5

$$\begin{aligned}\log|R| + \log|S| + \log|T| \\ &\geq h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ) \\ &= 2\log|Q|\end{aligned}$$

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Example

$Q(X,Y,Z) = R(X,Y) \bowtie S(Y,Z) \bowtie T(Z,X)$

$$|Q| \leq (|R| \cdot |S| \cdot |T|)^{1/2}$$

Define the uniform distribution on the output Q:

Q

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

R

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S

Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5

T

X	Z	
a	m	1/5
a	q	2/5
b	q	1/5
d	m	1/5

$$\begin{aligned}\log|R| + \log|S| + \log|T| \\ &\geq h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ) \\ &= 2\log|Q|\end{aligned}$$

$$|R| \cdot |S| \cdot |T| \geq |Q|^2$$

[Atserias et al., 2013, Friedgut and Kahn, 1998]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

[2] [Balister and Bollobás, 2012]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

Proof[2] Equivalently:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + \cdots + k_m h(\boldsymbol{Y}_m) \geq k_0 h(\boldsymbol{X})}$$

where each variable is "covered $\geq k_0$ times".

[2] [Balister and Bollobás, 2012]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

Proof[2] Equivalently:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + \cdots + k_m h(\boldsymbol{Y}_m) \geq k_0 h(\boldsymbol{X})}$$

where each variable is "covered $\geq k_0$ times".

Repeatedly rewrite $h(\boldsymbol{Y}_i) + h(\boldsymbol{Y}_j) \to h(\boldsymbol{Y}_i \cup \boldsymbol{Y}_j) + h(\boldsymbol{Y}_i \cap \boldsymbol{Y}_j)$

[2] [Balister and Bollobás, 2012]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

Proof[2] Equivalently:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + \cdots + k_m h(\boldsymbol{Y}_m) \geq k_0 h(\boldsymbol{X})}$$

where each variable is "covered $\geq k_0$ times".

Repeatedly rewrite $h(\boldsymbol{Y}_i) + h(\boldsymbol{Y}_j) \rightarrow h(\boldsymbol{Y}_i \cup \boldsymbol{Y}_j) + h(\boldsymbol{Y}_i \cap \boldsymbol{Y}_j)$

- Every variable remains covered $\geq k_0$ times.

[2] [Balister and Bollobás, 2012]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

Proof[2] Equivalently:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + \cdots + k_m h(\boldsymbol{Y}_m) \geq k_0 h(\boldsymbol{X})}$$

where each variable is "covered $\geq k_0$ times".

Repeatedly rewrite $h(\boldsymbol{Y}_i) + h(\boldsymbol{Y}_j) \to h(\boldsymbol{Y}_i \cup \boldsymbol{Y}_j) + h(\boldsymbol{Y}_i \cap \boldsymbol{Y}_j)$

- Every variable remains covered $\geq k_0$ times.
- $\sum_\ell |\boldsymbol{Y}_\ell|^2$ strictly increases (homework!).

[2] [Balister and Bollobás, 2012]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

Proof[2] Equivalently:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + \cdots + k_m h(\boldsymbol{Y}_m) \geq k_0 h(\boldsymbol{X})}$$

where each variable is "covered $\geq k_0$ times".

Repeatedly rewrite $h(\boldsymbol{Y}_i) + h(\boldsymbol{Y}_j) \rightarrow h(\boldsymbol{Y}_i \cup \boldsymbol{Y}_j) + h(\boldsymbol{Y}_i \cap \boldsymbol{Y}_j)$

- Every variable remains covered $\geq k_0$ times.
- $\sum_\ell |\boldsymbol{Y}_\ell|^2$ strictly increases (homework!).

When do we stop?

[2] [Balister and Bollobás, 2012]

Generalized Shearer's Inequality

Theorem If $\boldsymbol{w}$ is a fractional edge cover of the hyperedges $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$, then

$$\boxed{w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \geq h(\boldsymbol{X})}$$

Proof[2] Equivalently:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + \cdots + k_m h(\boldsymbol{Y}_m) \geq k_0 h(\boldsymbol{X})}$$

where each variable is "covered $\geq k_0$ times".

Repeatedly rewrite $h(\boldsymbol{Y}_i) + h(\boldsymbol{Y}_j) \longrightarrow h(\boldsymbol{Y}_i \cup \boldsymbol{Y}_j) + h(\boldsymbol{Y}_i \cap \boldsymbol{Y}_j)$

- Every variable remains covered $\geq k_0$ times.
- $\sum_\ell |\boldsymbol{Y}_\ell|^2$ strictly increases (homework!). When do we stop?
- Stop when $\boldsymbol{Y}_1 \supset \boldsymbol{Y}_2 \supset \cdots$ Then $\boldsymbol{Y}_1 = \boldsymbol{X}$ and $k_1 \geq k_0$:

$$\boxed{k_1 h(\boldsymbol{Y}_1) + k_2 h(\boldsymbol{Y}_2) + \cdots \geq k_1 h(\boldsymbol{Y}_1) \geq k_0 h(\boldsymbol{X})}$$

[2] [Balister and Bollobás, 2012]

Alternate Proof of the AGM Bound

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m).$

For every fractional edge cover $\boldsymbol{w}$: $\boxed{|Q(\boldsymbol{D})| \leq |R_1^D|^{w_1} \cdots |R_m^D|^{w_m}}$

Alternate Proof of the AGM Bound

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m).$

For every fractional edge cover $\boldsymbol{w}$: $\boxed{|Q(\boldsymbol{D})| \leq |R_1^D|^{w_1} \cdots |R_m^D|^{w_m}}$

Let h be the entropic vector of the uniform distribution on $Q(\boldsymbol{D})$.

$$\begin{aligned} w_1 \log |R_1^D| + \cdots + w_m \log |R_m^D| \geq \\ \geq w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \end{aligned}$$

Alternate Proof of the AGM Bound

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$.

For every fractional edge cover $\boldsymbol{w}$: $\boxed{|Q(\boldsymbol{D})| \leq |R_1^D|^{w_1} \cdots |R_m^D|^{w_m}}$

Let h be the entropic vector of the uniform distribution on $Q(\boldsymbol{D})$.

$$
\begin{aligned}
w_1 \log |R_1^D| + \cdots + w_m \log |R_m^D| \geq& \\
\geq & w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \\
\geq & h(\boldsymbol{X}) \qquad \text{Shearer's inequality}
\end{aligned}
$$

Alternate Proof of the AGM Bound

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$.

For every fractional edge cover $\boldsymbol{w}$: $\boxed{|Q(\boldsymbol{D})| \leq |R_1^D|^{w_1} \cdots |R_m^D|^{w_m}}$

Let h be the entropic vector of the uniform distribution on $Q(\boldsymbol{D})$.

$$
\begin{aligned}
w_1 \log |R_1^D| + \cdots + w_m \log |R_m^D| \geq & \\
\geq & w_1 h(\boldsymbol{Y}_1) + \cdots + w_m h(\boldsymbol{Y}_m) \\
\geq & h(\boldsymbol{X}) && \text{Shearer's inequality} \\
= & \log Q(\boldsymbol{D}) && \text{uniform distribution on } \boldsymbol{X}
\end{aligned}
$$

Discussion

- AGM/Shearer limited to cardinality statistics.
- More general statistics require Shannon inequalities.
- Everything gets harder: fractional edge cover no longer sufficient, order of the submodularity steps matters, tightness no longer guaranteed.
- Yet, lots of theory insights can make it work very well: next.

The LpBound

The Conditional Entropy

$$\boxed{h(\boldsymbol{V}|\boldsymbol{U}) \stackrel{\text{def}}{=} h(\boldsymbol{UV}) - h(\boldsymbol{U})}$$

May add $\boldsymbol{U}$ to $\boldsymbol{V}$: $h(\boldsymbol{UV}|\boldsymbol{U}) = h(\boldsymbol{V}|\boldsymbol{U})$

Semantics: $h(\boldsymbol{V}|\boldsymbol{U}) = \mathbb{E}_{\boldsymbol{u}}[h(\boldsymbol{V}|\boldsymbol{U} = \boldsymbol{u})]$

The Chain Rule: $\boxed{h(\boldsymbol{U}) + h(\boldsymbol{V}|\boldsymbol{U}) = h(\boldsymbol{UV})}$

Submodularity becomes: $\boxed{h(\boldsymbol{V}|\boldsymbol{U}) \geq h(\boldsymbol{V}|\boldsymbol{UW})}$

Connecting Statistics to the Entropy Vector

Relation $R(X, Y, Z)$:

$$h(XYZ) \leq \log |R|$$

Connecting Statistics to the Entropy Vector

Relation $R(X, Y, Z)$:

$$h(XYZ) \leq \log |R| \qquad h(X) \leq \log |R.X| \qquad h(XY) \leq \log |R.XY|$$

Connecting Statistics to the Entropy Vector

Relation $R(X, Y, Z)$:

$$h(XYZ) \leq \log|R| \qquad h(X) \leq \log|R.X| \qquad h(XY) \leq \log|R.XY|$$

$$h(Y|X) \leq \log||\deg_R(Y|X)||_\infty$$

$$\mathbb{E}_x[h(Y|X=x)] \leq \mathbb{E}_x[\log(\deg(Y|X=x))]$$
$$\leq \max_x(\log(\deg(Y|X=x)))$$

Connecting Statistics to the Entropy Vector

Relation $R(X, Y, Z)$:

$$h(XYZ) \leq \log|R| \qquad h(X) \leq \log|R.X| \qquad h(XY) \leq \log|R.XY|$$

$$h(Y|X) \leq \log||\text{deg}_R(Y|X)||_\infty$$

$$\mathbb{E}_x[h(Y|X=x)] \leq \mathbb{E}_x[\log(\text{deg}(Y|X=x))] \leq \max_x(\log(\text{deg}(Y|X=x)))$$

$$\boxed{h(X) + p \cdot h(Y|X) \leq \log||\text{deg}(Y|X)||_p^p}$$

[Zhang et al., 2025]

Equivalently: $\frac{1}{p}h(X) + h(Y|X) \leq \log||\text{deg}(Y|X)||_p$

The LpBound[3]

$Q(\mathbf{X}) = R_1(\mathbf{Y}_1) \wedge \cdots \wedge R_m(\mathbf{Y}_m).$

Collect various statistics of the form $||\text{deg}_{R_j^D}(\mathbf{Y}|\mathbf{X})||_p$, for $\mathbf{X}, \mathbf{Y} \subseteq \texttt{Vars}(R_j)$.

This includes cardinalities, domain sizes, max degrees.

[3] [Zhang et al., 2025]

The LpBound[3]

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$.

Collect various statistics of the form $||\text{deg}_{R_j^D}(\boldsymbol{Y}|\boldsymbol{X})||_p$, for $\boldsymbol{X}, \boldsymbol{Y} \subseteq \texttt{Vars}(R_j)$.
This includes cardinalities, domain sizes, max degrees.

Prove a Shannon inequality of the form:

$$\sum_i w_i\Big(h(X_i) + p \cdot h(Y_i|X_i)\Big) \geq h(\boldsymbol{X})$$

[3] [Zhang et al., 2025]

The LpBound[3]

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$.

Collect various statistics of the form $||\text{deg}_{R_j^D}(\boldsymbol{Y}|\boldsymbol{X})||_p$, for $\boldsymbol{X}, \boldsymbol{Y} \subseteq \texttt{Vars}(R_j)$.
This includes cardinalities, domain sizes, max degrees.

Prove a Shannon inequality of the form:

$$\sum_i w_i \Big(h(X_i) + p \cdot h(Y_i|X_i) \Big) \geq h(\boldsymbol{X})$$

It follows:

$$\boxed{\prod_i ||\text{deg}_{R_{j_i}^D}(\boldsymbol{Y}_i|\boldsymbol{X}_i)||_{p_i}^{w_i p_i} \geq Q(\boldsymbol{D})}$$

[3] [Zhang et al., 2025]

Example 1

$Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$

Then $|Q| \leq \left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$.

Example 1

$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$
Then $|Q| \leq \left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$.

Proof: let h be the entropy vector of the uniform distribution on $Q(\boldsymbol{D})$:

$$\log ||\text{deg}_R(Y|X)||_2^2 + \log ||\text{deg}_S(Z|Y)||_2^2 + \log ||\text{deg}_T(X|Z)||_2^2 \geq$$

$$= 3 \log |Q|$$

Example 1

$Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$
Then $|Q| \leq \left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$.

Proof: let h be the entropy vector of the uniform distribution on $Q(\boldsymbol{D})$:

$$\log ||\text{deg}_R(Y|X)||_2^2 + \log ||\text{deg}_S(Z|Y)||_2^2 + \log ||\text{deg}_T(X|Z)||_2^2 \geq$$
$$\geq h(X) + 2h(Y|X) + h(Y) + 2h(Z|Y) + h(Z) + 2h(X|Z)$$

$$= 3 \log |Q|$$

Example 1

$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$
Then $|Q| \leq \left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$.

Proof: let h be the entropy vector of the uniform distribution on $Q(\boldsymbol{D})$:

$$\log||\text{deg}_R(Y|X)||_2^2 + \log||\text{deg}_S(Z|Y)||_2^2 + \log||\text{deg}_T(X|Z)||_2^2 \geq$$
$$\geq h(X) + 2h(Y|X) + h(Y) + 2h(Z|Y) + h(Z) + 2h(X|Z)$$
$$= h(XY) + h(Y|X) + h(YZ) + h(Z|Y) + h(XZ) + h(X|Z)$$

$$= 3\log|Q|$$

Example 1

$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Then $|Q| \leq \left(||\deg_R(Y|X)||_2^2 \cdot ||\deg_S(Z|Y)||_2^2 \cdot ||\deg_T(X|Z)||_2^2\right)^{1/3}$.

Proof: let h be the entropy vector of the uniform distribution on $Q(\boldsymbol{D})$:

$$\begin{aligned}
&\log ||\deg_R(Y|X)||_2^2 + \log ||\deg_S(Z|Y)||_2^2 + \log ||\deg_T(X|Z)||_2^2 \geq \\
&\geq h(X) + 2h(Y|X) + h(Y) + 2h(Z|Y) + h(Z) + 2h(X|Z) \\
&= h(XY) + \underline{h(Y|X)} + h(YZ) + \underline{h(Z|Y)} + h(XZ) + \underline{h(X|Z)} \\
&\geq h(XY) + h(Y|XZ) + h(YZ) + h(Z|XY) + h(XZ) + h(X|YZ) \\
&= 3 \log |Q|
\end{aligned}$$

Example 1

$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Then $|Q| \leq \left(||\text{deg}_R(Y|X)||_2^2 \cdot ||\text{deg}_S(Z|Y)||_2^2 \cdot ||\text{deg}_T(X|Z)||_2^2\right)^{1/3}$.

Proof: let h be the entropy vector of the uniform distribution on $Q(\boldsymbol{D})$:

$$\begin{aligned}
\log ||\text{deg}_R(Y|X)||_2^2 + \log ||\text{deg}_S(Z|Y)||_2^2 + \log ||\text{deg}_T(X|Z)||_2^2 \geq \\
\geq & h(X) + 2h(Y|X) + h(Y) + 2h(Z|Y) + h(Z) + 2h(X|Z) \\
= & h(XY) + \underline{h(Y|X)} + h(YZ) + \underline{h(Z|Y)} + h(XZ) + \underline{h(X|Z)} \\
\geq & h(XY) + h(Y|XZ) + h(YZ) + h(Z|XY) + h(XZ) + h(X|YZ) \\
= & 3h(XYZ) \\
= & 3 \log |Q|
\end{aligned}$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \left(X + Z = U\right) \wedge \left(Y * U = X\right)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Example 2

$$Q = R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Using FDs $XZ \rightarrow U$ and $YU \rightarrow X$: $\qquad |Q| \leq N^{3/2}$.

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\quad AGM(Q) = N^2$.
Using FDs $XZ \to U$ and $YU \to X$: $\quad |Q| \leq N^{3/2}$.

$$\log|R| + \log|S| + \log|T| \geq h(XY) + h(YZ) + h(ZU)$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.
Using FDs $XZ \rightarrow U$ and $YU \rightarrow X$: $\qquad |Q| \leq N^{3/2}$.

$$\begin{aligned}\log|R| + \log|S| + \log|T| &\geq h(XY) + h(YZ) + h(ZU)\\ &= h(XY)+h(YZ) + h(ZU) + h(U|XZ) + h(X|YU)\end{aligned}$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Using FDs $XZ \rightarrow U$ and $YU \rightarrow X$: $\qquad |Q| \leq N^{3/2}$.

$$\log|R| + \log|S| + \log|T| \geq h(XY) + h(YZ) + h(ZU)$$
$$= \underline{h(XY) + h(YZ)} + h(ZU) + h(U|XZ) + h(X|YU)$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Using FDs $XZ \rightarrow U$ and $YU \rightarrow X$: $\qquad |Q| \leq N^{3/2}$.

$$\begin{aligned}
\log|R| + \log|S| + \log|T| &\geq h(XY) + h(YZ) + h(ZU) \\
&= \underline{h(XY) + h(YZ)} + h(ZU) + h(U|XZ) + h(X|YU) \\
&\geq h(XYZ) + h(Y) + h(ZU) + h(U|XZ) + h(X|YU)
\end{aligned}$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Using FDs $XZ \to U$ and $YU \to X$: $\qquad |Q| \leq N^{3/2}$.

$$\begin{aligned}
\log|R| + \log|S| + \log|T| &\geq h(XY) + h(YZ) + h(ZU) \\
&= h(XY) + h(YZ) + h(ZU) + h(U|XZ) + h(X|YU) \\
&\geq h(XYZ) + \underline{h(Y) + h(ZU)} + h(U|XZ) + h(X|YU)
\end{aligned}$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Using FDs $XZ \rightarrow U$ and $YU \rightarrow X$: $\qquad |Q| \leq N^{3/2}$.

$$\begin{aligned}
\log|R| + \log|S| + \log|T| &\geq h(XY) + h(YZ) + h(ZU) \\
&= h(XY) + h(YZ) + h(ZU) + h(U|XZ) + h(X|YU) \\
&\geq h(XYZ) + \underline{h(Y) + h(ZU)} + h(U|XZ) + h(X|YU) \\
&\geq h(XYZ) + h(YZU) + h(U|XZ) + h(X|YU)
\end{aligned}$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \big(X + Z = U\big) \wedge \big(Y * U = X\big)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $\qquad AGM(Q) = N^2$.

Using FDs $XZ \to U$ and $YU \to X$: $\qquad |Q| \le N^{3/2}$.

$$\begin{aligned}
\log|R| + \log|S| + \log|T| &\ge h(XY) + h(YZ) + h(ZU) \\
= h(XY) + h(YZ) + h(ZU) &+ h(U|XZ) + h(X|YU) \\
&\ge h(XYZ) + h(Y) + h(ZU) + h(U|XZ) + h(X|YU) \\
&\ge h(XYZ) + h(YZU) + \underline{h(U|XZ)} + \underline{h(X|YU)}
\end{aligned}$$

Example 2

$$Q = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge \left(X + Z = U\right) \wedge \left(Y * U = X\right)$$

Assume $|R^D| = |S^D| = |T^D| = N$ $AGM(Q) = N^2$.
Using FDs $XZ \rightarrow U$ and $YU \rightarrow X$: $|Q| \leq N^{3/2}$.

$$\begin{aligned}
\log|R| + \log|S| + \log|T| &\geq h(XY) + h(YZ) + h(ZU) \\
&= h(XY) + h(YZ) + h(ZU) + h(U|XZ) + h(X|YU) \\
&\geq h(XYZ) + h(Y) + h(ZU) + h(U|XZ) + h(X|YU) \\
&\geq h(XYZ) + h(YZU) + \underline{h(U|XZ)} + \underline{h(X|YU)} \\
&\geq h(XYZ) + h(YZU) + h(U|XYZ) + h(X|YZU) \\
&= 2h(XYZU) = \boxed{2\log|Q|}
\end{aligned}$$

Discussion: AGM v.s. LpBound

- Any Shannon inequality gives a bound on $Q(\boldsymbol{D})$; similar to AGM.
- Beyond Shearer inequalities: general Shannon inequalities are more difficult (next).
- AGM bound: tight up to a rounding factor $\frac{1}{2^n}$.
  LpBound: not tight in general (!!!)
  LpBound for simple conditionals: tight up to rounding factor $\frac{1}{2^{2^n-1}}$

Computing the Upper Bound

Motivation

n = number of variables in the query.

AGM bound:

- Optimal fractional edge cover: Primal Linear Program.
- Optimal vertex packing: Dual Linear Program, with n unknowns.

LpBound:

- Optimal Shannon inequality: Dual Linear Program.
- Optimal vector h: Primal Linear program, with 2^n unknowns.

Note: my choice of which LP to call "primal" or "dual" is inconsistent between AGM and LpBound.

The Primal Linear Program

$Q(\boldsymbol{X}) = \bigwedge_j R_j(\boldsymbol{X}_j)$, m atoms, n variables.

$\boxed{\text{Maximize } h(\boldsymbol{X})}$ where:

- There are 2^n unknowns: $h(\boldsymbol{U})$ for all $\boldsymbol{U} \subseteq \boldsymbol{X}$.

The Primal Linear Program

$Q(\boldsymbol{X}) = \bigwedge_j R_j(\boldsymbol{X}_j)$, m atoms, n variables.

$\boxed{\text{Maximize } h(\boldsymbol{X})}$ where:

- There are 2^n unknowns: $h(\boldsymbol{U})$ for all $\boldsymbol{U} \subseteq \boldsymbol{X}$.
- For each stats add the corresponding constraint:

$$h(\boldsymbol{X}_j) \leq \log |R_j|$$

$$\cdots$$

$$h(\boldsymbol{U}) + p h(\boldsymbol{V}|\boldsymbol{U}) \leq \log ||\text{deg}_{R_k}(\boldsymbol{V}|\boldsymbol{U})||_p^p$$

The Primal Linear Program

$Q(\boldsymbol{X}) = \bigwedge_j R_j(\boldsymbol{X}_j)$, m atoms, n variables.

Maximize $h(\boldsymbol{X})$ where:

- There are 2^n unknowns: $h(\boldsymbol{U})$ for all $\boldsymbol{U} \subseteq \boldsymbol{X}$.
- For each stats add the corresponding constraint:

$$h(\boldsymbol{X}_j) \leq \log |R_j|$$

$$\ldots$$

$$h(\boldsymbol{U}) + ph(\boldsymbol{V}|\boldsymbol{U}) \leq \log ||\text{deg}_{R_k}(\boldsymbol{V}|\boldsymbol{U})||_p^p$$

- Add all Shannon inequalities as constraints:

$$-h(XY) - h(YZ) + h(XYZ) + h(Y) \leq 0$$

$$\ldots$$

Example

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$

Example

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$

Primal:

Maximize $h(XYZ)$, where:

c_1 : $h(XY) \leq \log |R|$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$c_1 : \quad h(XY) \leq \log |R|$$
$$c_2 : \quad h(YZ) \leq \log |S|$$

Example

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$

Primal:

Maximize $h(XYZ)$, where:

$$c_1 : \quad h(XY) \leq \log |R|$$
$$c_2 : \quad h(YZ) \leq \log |S|$$
$$c_3 : \quad h(XZ) \leq \log |T|$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 &: & h(XY) &\leq \log|R| \\
c_2 &: & h(YZ) &\leq \log|S| \\
c_3 &: & h(XZ) &\leq \log|T| \\
\sigma_1 &: & -h(XY) - h(YZ) & \\
& & +h(XYZ) + h(Y) \leq 0 &
\end{aligned}
$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 :&\quad h(XY) \leq \log|R| \\
c_2 :&\quad h(YZ) \leq \log|S| \\
c_3 :&\quad h(XZ) \leq \log|T| \\
\sigma_1 :&\quad -h(XY) - h(YZ) \\
&\quad +h(XYZ) + h(Y) \leq 0 \\
\sigma_2 :&\quad -h(Y) - h(XZ) \\
&\quad +h(XYZ) \leq 0 \\
&\quad \ldots
\end{aligned}
$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
{\color{red}c_1}: &\quad h(XY) \leq \log|R| \\
{\color{red}c_2}: &\quad h(YZ) \leq \log|S| \\
{\color{red}c_3}: &\quad h(XZ) \leq \log|T| \\
\sigma_1: &\quad -h(XY) - h(YZ) \\
&\quad +h(XYZ) + h(Y) \leq 0 \\
\sigma_2: &\quad -h(Y) - h(XZ) \\
&\quad +h(XYZ) \leq 0 \\
&\quad \ldots \\
\sigma_{18}: &\quad \cdots \leq 0
\end{aligned}
$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 &: & h(XY) &\leq \log|R| \\
c_2 &: & h(YZ) &\leq \log|S| \\
c_3 &: & h(XZ) &\leq \log|T| \\
\sigma_1 &: & -h(XY) - h(YZ) & \\
& & +h(XYZ) + h(Y) &\leq 0 \\
\sigma_2 &: & -h(Y) - h(XZ) & \\
& & +h(XYZ) &\leq 0 \\
& & &\cdots \\
\sigma_{18} &: & \cdots &\leq 0
\end{aligned}
$$

Dual:

Minimize $c_1 \log|R| + c_2 \log|S| + c_3 \log|T|$

where:

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 &: & h(XY) &\leq \log |R| \\
c_2 &: & h(YZ) &\leq \log |S| \\
c_3 &: & h(XZ) &\leq \log |T| \\
\sigma_1 &: & -h(XY) - h(YZ) & \\
& & +h(XYZ) + h(Y) &\leq 0 \\
\sigma_2 &: & -h(Y) - h(XZ) & \\
& & +h(XYZ) &\leq 0 \\
& & &\ldots \\
\sigma_{18} &: & \cdots &\leq 0
\end{aligned}
$$

Dual:

Minimize $c_1 \log |R| + c_2 \log |S| + c_3 \log |T|$ where:

$$h(XYZ): \quad \sigma_1 + \sigma_2 + \cdots \geq 1$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 &: & h(XY) &\leq \log|R| \\
c_2 &: & h(YZ) &\leq \log|S| \\
c_3 &: & h(XZ) &\leq \log|T| \\
\sigma_1 &: & -h(XY) - h(YZ) & \\
& & +h(XYZ) + h(Y) &\leq 0 \\
\sigma_2 &: & -h(Y) - h(XZ) & \\
& & +h(XYZ) &\leq 0 \\
& & &\cdots \\
\sigma_{18} &: & \cdots &\leq 0
\end{aligned}
$$

Dual:

Minimize $c_1 \log|R| + c_2 \log|S| + c_3 \log|T|$ where:

$$
\begin{aligned}
h(XYZ) &: & \sigma_1 + \sigma_2 + \cdots &\geq 1 \\
h(XY) &: & c_1 - \sigma_1 + \cdots &\geq 0 \\
h(YZ) &: & c_2 - \sigma_1 + \cdots &\geq 0 \\
h(XZ) &: & c_3 - \sigma_2 + \cdots &\geq 0
\end{aligned}
$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 &: & h(XY) &\leq \log|R| \\
c_2 &: & h(YZ) &\leq \log|S| \\
c_3 &: & h(XZ) &\leq \log|T| \\
\sigma_1 &: & -h(XY) - h(YZ) & \\
& & +h(XYZ) + h(Y) &\leq 0 \\
\sigma_2 &: & -h(Y) - h(XZ) & \\
& & +h(XYZ) &\leq 0 \\
& & & \cdots \\
\sigma_{18} &: & \cdots &\leq 0
\end{aligned}
$$

Dual:

Minimize $c_1 \log|R| + c_2 \log|S| + c_3 \log|T|$ where:

$$
\begin{aligned}
h(XYZ) &: & \sigma_1 + \sigma_2 + \cdots &\geq 1 \\
h(XY) &: & c_1 - \sigma_1 + \cdots &\geq 0 \\
h(YZ) &: & c_2 - \sigma_1 + \cdots &\geq 0 \\
h(XZ) &: & c_3 - \sigma_2 + \cdots &\geq 0 \\
h(X) &: & \cdots &\geq 0 \\
h(Y) &: & \sigma_1 - \sigma_2 + \cdots &\geq 0 \\
h(Z) &: & \cdots &\geq 0
\end{aligned}
$$

Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Primal:

Maximize $h(XYZ)$, where:

$$
\begin{aligned}
c_1 &: & h(XY) &\leq \log|R| \\
c_2 &: & h(YZ) &\leq \log|S| \\
c_3 &: & h(XZ) &\leq \log|T| \\
\sigma_1 &: & -h(XY) - h(YZ) & \\
& & +h(XYZ) + h(Y) &\leq 0 \\
\sigma_2 &: & -h(Y) - h(XZ) & \\
& & +h(XYZ) &\leq 0 \\
& & & \cdots \\
\sigma_{18} &: & \cdots &\leq 0
\end{aligned}
$$

Dual:

Minimize $c_1 \log|R| + c_2 \log|S| + c_3 \log|T|$ where:

$$
\begin{aligned}
h(XYZ) &: & \sigma_1 + \sigma_2 + \cdots &\geq 1 \\
h(XY) &: & c_1 - \sigma_1 + \cdots &\geq 0 \\
h(YZ) &: & c_2 - \sigma_1 + \cdots &\geq 0 \\
h(XZ) &: & c_3 - \sigma_2 + \cdots &\geq 0 \\
h(X) &: & \cdots &\geq 0 \\
h(Y) &: & \sigma_1 - \sigma_2 + \cdots &\geq 0 \\
h(Z) &: & \cdots &\geq 0
\end{aligned}
$$

Correctness: any feasible solution $c_1, c_2, c_3, \sigma_1, \ldots, \sigma_{18}$ of the dual defines a Shannon inequality $c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$.

Correctness Proof

Theorem

Any feasible solution $c_1, c_2, c_3, \sigma_1, \ldots, \sigma_{18}$ *of the primal defines a Shannon inequality* $c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$.

Proof: Multiply each inequality in the dual with its h-term and add them:

$$h(XYZ)(\sigma_1 + \cdots) + h(XY)(c_1 - \sigma_1 + \cdots) + \cdots \geq h(XYZ)$$

Correctness Proof

Theorem

Any feasible solution $c_1, c_2, c_3, \sigma_1, \ldots, \sigma_{18}$ *of the primal defines a Shannon inequality* $c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$.

Proof: Multiply each inequality in the dual with its h-term and add them:

$$h(XYZ)(\sigma_1 + \cdots) + h(XY)(c_1 - \sigma_1 + \cdots) + \cdots \geq h(XYZ)$$

Group by the coefficients $c_1, c_2, c_3, \sigma_1, \sigma_2, \ldots$

Correctness Proof

Theorem

Any feasible solution $c_1, c_2, c_3, \sigma_1, \ldots, \sigma_{18}$ *of the primal defines a Shannon inequality* $c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$.

Proof: Multiply each inequality in the dual with its h-term and add them:

$$h(XYZ)(\sigma_1 + \cdots) + h(XY)(c_1 - \sigma_1 + \cdots) + \cdots \geq h(XYZ)$$

Group by the coefficients $c_1, c_2, c_3, \sigma_1, \sigma_2, \ldots$

$$c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) + \sigma_1(\cdots) + \cdots \geq h(XYZ)$$

Correctness Proof

Theorem

Any feasible solution $c_1, c_2, c_3, \sigma_1, \ldots, \sigma_{18}$ of the primal defines a Shannon inequality $c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$.

Proof: Multiply each inequality in the dual with its h-term and add them:

$$h(XYZ)(\sigma_1 + \cdots) + h(XY)(c_1 - \sigma_1 + \cdots) + \cdots \geq h(XYZ)$$

Group by the coefficients $c_1, c_2, c_3, \sigma_1, \sigma_2, \ldots$

$$c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) + \sigma_1(\cdots) + \cdots \geq h(XYZ)$$

By design, the co-factor of σ_i is the LHS of a Shannon inequality,

e.g. $\sigma_1(-h(XY) - h(YZ) + h(XYZ) + h(Y))$

Correctness Proof

Theorem

Any feasible solution $c_1, c_2, c_3, \sigma_1, \ldots, \sigma_{18}$ of the primal defines a Shannon inequality $c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$.

Proof: Multiply each inequality in the dual with its h-term and add them:

$$h(XYZ)(\sigma_1 + \cdots) + h(XY)(c_1 - \sigma_1 + \cdots) + \cdots \geq h(XYZ)$$

Group by the coefficients $c_1, c_2, c_3, \sigma_1, \sigma_2, \ldots$

$$c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) + \sigma_1(\cdots) + \cdots \geq h(XYZ)$$

By design, the co-factor of σ_i is the LHS of a Shannon inequality,

e.g. $\sigma_1(-h(XY) - h(YZ) + h(XYZ) + h(Y))$

Shannon inequalities $-h(XY) - h(YZ) + h(XYZ) + h(Y) \leq 0$ imply:

$$c_1 h(XY) + c_2 h(YZ) + c_3 h(XZ) \geq h(XYZ)$$

Discussion of the LpBound

- Primal: maximize $h(\boldsymbol{X})$, where h is a polymatroid[4]
  2^n unknowns $\boldsymbol{h} \in \mathbb{R}_+^{2^n}$.
  $\boldsymbol{h}$ is not a database instance in general.

- Special case: all stats are cardinalities (AGM bound)
  Shearer inequality; $\boldsymbol{h}$ is modular, defines product database.
  LP has n unknowns $h(X_1), \ldots, h(X_n)$: fractional vertex packing.

- Special case: all degree stats are simple.
  $\boldsymbol{h}$ is normal, defines normal database [Suciu, 2023].
  Solvable by an LP with n^2 unknowns [Im et al., 2025].

- Special case: Q is Berge acyclic, degree stats are simple.
  Solvable by an LP with $m + n$ unknowns.

[4] Satisfies all basic Shannon inequalities

Final Thoughts: Entropic Inequalities in Databases

- LpBound, (plus successor CorrBound): a practical cardinality estimator.
- Generic Join: matches the AGM bound. Practical, but incompatible with query engines.
- PandaExpress: matches the LpBound. Not yet practical. [Khamis et al., 2025].

Atserias, A., Grohe, M., and Marx, D. (2013).
Size bounds and query plans for relational joins.
SIAM J. Comput., 42(4):1737–1767.

Balister, P. and Bollobás, B. (2012).
Projections, entropy and sumsets.
Comb., 32(2):125–141.

Cai, W., Balazinska, M., and Suciu, D. (2019).
Pessimistic cardinality estimation: Tighter upper bounds for intermediate join cardinalities.
In Boncz, P. A., Manegold, S., Ailamaki, A., Deshpande, A., and Kraska, T., editors, *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*, pages 18–35. ACM.

Friedgut, E. and Kahn, J. (1998).
On the number of copies of one hypergraph in another.
Israel Journal of Mathematics, 105(1):251–256.

Im, S., Moseley, B., Ngo, H. Q., and Pruhs, K. (2025).
Efficient algorithms for cardinality estimation and conjunctive query evaluation with simple degree constraints.
Proc. ACM Manag. Data, 3(2):96:1–96:26.

Khamis, M. A., Ngo, H. Q., and Suciu, D. (2025).
Pandaexpress: a simpler and faster PANDA algorithm.
CoRR, abs/2512.10217.

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A. W., Safaya, A., Tazarv, A., Xiang, A., Parrish, A., Nie, A., Hussain, A., Askell, A., Dsouza, A., Slone, A., Rahane, A., Iyer, A. S., Andreassen, A., Madotto, A., Santilli, A., Stuhlmüller, A., Dai, A. M., La, A., Lampinen, A. K., Zou, A., Jiang, A., Chen, A., Vuong, A., Gupta, A., Gottardi, A., Norelli, A., Venkatesh, A., Gholamidavoodi, A., Tabassum, A., Menezes, A., Kirubarajan, A., Mullokandov, A., Sabharwal, A., Herrick, A., Efrat, A., Erdem, A., Karakas, A., Roberts, B. R., Loe, B. S., Zoph, B., Bojanowski, B., Özyurt, B., Hedayatnia, B., Neyshabur, B., Inden, B., Stein, B., Ekmekci, B., Lin, B. Y., Howald, B., Orinion, B., Diao, C., Dour, C., Stinson, C., Argueta, C., Ramírez, C. F., Singh, C., Rathkopf, C., Meng, C., Baral, C., Wu, C., Callison-Burch, C., Waites, C., Voigt, C., Manning, C. D., Potts, C., Ramirez, C., Rivera, C. E., Siro, C., Raffel, C., Ashcraft, C., Garbacea, C., Sileo, D., Garrette, D., Hendrycks, D., Kilman, D., Roth, D., Freeman, D., Khashabi, D., Levy, D., González, D. M., Perszyk, D., Hernandez, D., Chen, D., Ippolito, D., Gilboa, D., Dohan, D., Drakard, D., Jurgens, D., Datta, D., Ganguli, D., Emelin, D., Kleyko, D., Yuret, D., Chen, D., Tam, D., Hupkes, D., Misra, D., Buzan, D., Mollo, D. C., Yang, D., Lee, D., Schrader, D., Shutova, E., Cubuk, E. D., Segal, E., Hagerman, E., Barnes, E., Donoway, E., Pavlick, E., Rodolà, E., Lam, E., Chu, E., Tang, E., Erdem, E., Chang, E., Chi, E. A., Dyer, E., Jerzak, E. J., Kim, E., Manyasi, E. E., Zheltonozhskii, E., Xia, F., Siar, F., Martínez-Plumed, F., Happé, F., Chollet, F., Rong, F., Mishra, G., Winata, G. I., de Melo, G., Kruszewski, G., Parascandolo, G., Mariani, G., Wang, G., Jaimovitch-López, G., Betz, G., Gur-Ari, G., Galijasevic, H., Kim, H., Rashkin, H., Hajishirzi, H., Mehta, H., Bogar, H., Shevlin, H., Schütze, H., Yakura, H., Zhang, H., Wong, H. M., Ng, I., Noble, I., Jumelet, J., Geissinger, J., Kernion, J., Hilton, J., Lee, J., Fisac, J. F., Simon, J. B., Koppel, J., Zheng, J., Zou, J., Kocon, J., Thompson, J., Wingfield, J., Kaplan, J., Radom, J., Sohl-Dickstein, J., Phang, J., Wei, J., Yosinski, J., Novikova, J., Bosscher, J., Marsh, J., Kim, J., Taal, J., Engel, J. H., Alabi, J., Xu, J., Song, J., Tang, J., Waweru, J., Burden, J., Miller, J., Balis, J. U., Batchelder, J., Berant, J., Frohberg, J., Rozen, J., Hernández-Orallo, J., Boudeman, J., Guerr, J., Jones, J., Tenenbaum, J. B., Rule, J. S., Chua, J., Kanclerz, K., Livescu, K., Krauth, K., Gopalakrishnan, K., Ignatyeva, K., Markert, K., Dhole, K. D., Gimpel, K., Omondi, K., Mathewson, K., Chiafullo, K., Shkaruta, K., Shridhar, K., McDonell, K., Richardson, K., Reynolds, L., Gao, L., Zhang, L., Dugan, L., Qin, L., Ochando, L. C., Morency, L., Moschella, L., Lam, L., Noble, L., Schmidt, L., He, L., Colón, L. O., Metz, L., Senel, L. K., Bosma, M., Sap, M., ter Hoeve, M., Farooqi, M., Faruqui, M., Mazeika, M., Baturan, M., Marelli, M., Maru, M., Ramírez-Quintana, M. J., Tolkiehn, M., Giulianelli, M., Lewis, M., Potthast, M., Leavitt, M. L., Hagen, M., Schubert, M., Baitemirova, M., Arnaud, M., McElrath, M., Yee, M. A., Cohen, M., Gu, M., Ivanitskiy, M. I., Starritt, M., Strube, M., Swedrowski, M., Bevilacqua, M., Yasunaga, M., Kale, M., Cain, M., Xu, M., Suzgun, M., Walker, M., Tiwari, M., Bansal, M., Aminnaseri, M., Geva, M., Gheini, M., T., M. V., Peng, N., Chi, N. A., Lee, N., Krakover, N. G., Cameron, N., Roberts, N., Doiron, N., Martinez, N., Nangia, N., Deckers, N., Muennighoff, N., Keskar, N. S., Iyer, N., Constant, N., Fiedel, N., Wen, N., Zhang, O., Agha, O., Elbaghdadi, O., Levy, O., Evans, O., Casares, P. A. M., Doshi, P., Fung, P., Liang, P. P., Vicol, P., Alipoormolabashi, P., Liao, P., Liang, P., Chang, P., Eckersley, P., Htut, P. M., Hwang, P., Milkowski, P., Patil, P., Pezeshkpour, P., Oli, P., Mei, Q., Lyu, Q., Chen, Q., Banjade, R., Rudolph, R. E., Gabriel, R., Habacker, R., Risco, R., Millière, R., Garg, R., Barnes, R., Saurous, R. A., Arakawa, R., Raymaekers, R., Frank, R., Sikand, R., Novak, R., Sitelew, R., LeBras, R., Liu, R., Jacobs, R., Zhang, R., Salakhutdinov, R., Chi, R., Lee, R., Stovall, R., Teehan, R., Yang, R., Singh, S., Mohammad, S. M., Anand, S., Dillavou, S., Shleifer, S., Wiseman, S., Gruetter, S., Bowman, S. R., Schoenholz, S. S., Han, S., Kwatra, S., Rous, S. A., Ghazarian, S., Ghosh, S., Casey, S., Bischoff, S., Gehrmann, S., Schuster, S., Sadeghi, S., Hamdan, S., Zhou, S., Srivastava, S., Shi, S., Singh, S., Asaadi, S., Gu, S. S., Pachchigar, S., Toshniwal, S., Upadhyay, S., Debnath, S. S., Shakeri, S., Thormeyer, S., Melzi, S., Reddy, S., Makini, S. P., Lee, S., Torene, S., Hatwar, S., Dehaene, S., Divic, S., Ermon, S., Biderman, S., Lin, S., Prasad, S., Piantadosi, S. T., Shieber, S. M., Misherghi, S., Kiritchenko, S., Mishra, S., Linzen, T., Schuster, T., Li, T., Yu, T., Ali, T., Hashimoto, T., Wu, T., Desbordes, T., Rothschild, T., Phan, T., Wang, T., Nkinyili, T., Schick, T., Kornev, T., Tunduny, T., Gerstenberg, T., Chang, T., Neeraj, T., Khot, T., Shultz, T., Shaham, U., Misra, V., Demberg, V., Nyamai, V., Raunak, V., Ramasesh, V. V., Prabhu, V. U., Padmakumar, V., Srikumar, V., Fedus, W., Saunders, W., Zhang, W., Vossen, W., Ren, X., Tong, X., Zhao, X., Wu, X., Shen, X., Yaghoobzadeh, Y., Lakretz, Y., Song, Y., Bahri, Y., Choi, Y., Yang, Y., Hao, Y., Chen, Y., Belinkov, Y., Hou, Y., Hou, Y., Bai, Y., Seid, Z., Zhao, Z., Wang, Z., Wang, Z. J., Wang, Z., and Wu, Z. (2023).

Beyond the imitation game: Quantifying and extrapolating the capabilities of language models.
Trans. Mach. Learn. Res., 2023.

Suciu, D. (2023).
Applications of information inequalities to database theory problems.
In *LICS*, pages 1–30.

Zhang, H., Mayer, C., Khamis, M. A., Olteanu, D., and Suciu, D. (2025).
Lpbound: Pessimistic cardinality estimation using ℓ_p-norms of degree sequences.
Proc. ACM Manag. Data, 3(3):184:1–184:27.