Advanced Query Processing

CSEP544

Nov. 2015

Lecture 9: Advanced Query Processing

- Optimal Sequential Algorithms.
- Semijoin Reduction
- Optimal Parallel Algorithms.

Bibliography

- E Friedgut, Hypergraphs, entropy, and inequalities, American Mathematical Monthly, 749-760, 2004.
- Albert Atserias, Martin Grohe, Dniel Marx: Size Bounds and Query Plans for Relational Joins. SIAM J. Comput. 42(4): 1737-1767 (2013)
- Hung Q. Ngo, Christopher Ré, Atri Rudra: Skew strikes back: new developments in the theory of join algorithms. SIGMOD Record 42(4): 5-16 (2013)
- Paul Beame, Paraschos Koutris, Dan Suciu: Skew in parallel query processing. PODS 2014: 212-223
- Paul Beame, Paraschos Koutris, Dan Suciu: Communication steps for parallel query processing. PODS 2013: 273-284

Natural Join

$R \bowtie S$

Joins R, S on all common attributes, removes duplicate attributes

R

A	B
a_1	b_1
a_1	b_2
a_2	b_3
a_3	b_4

S

A	C
a_1	c_1
a_1	c_2
a_3	c_3
a_4	c_4

$T = R \bowtie S$

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2
a_1	b_2	c_1
a_1	b_2	c_2
a_3	b_4	c_3

Input schemas: $R(A, B)$, $S(A, C)$
Output schema: $T(A, B, C)$

Very Quick Review of Basic Join Algorithms

Compute $R \bowtie_{A=B} S$

- Nested-loop join
- Hash-join
- Merge-join

(To describe in class.)

Complexity: $O((|R| + |S| + |R \bowtie_{A=B} S|) \log(|R| + |S|))$

Ignoring log factors, Complexity: $\boxed{O(|\text{Input}| + |\text{Output}|)}$

Conjunctive Queries

Example

$Q_1(x, y, z, u) = R(x, y), S(y, z), T(z, u)$

- Relational Algebra: $(R(x, y) \bowtie S(y, z)) \bowtie T(z, u)$
- First Order Logic:
  $Q_1 = \{(x, y, z, u) \mid (x, y) \in R \wedge (y, z) \in S \wedge (z, u) \in T\}$
- SQL: `select * from` $\cdots$ `where` $\cdots$

Conjunctive Queries

Example

$Q_1(x, y, z, u) = R(x, y), S(y, z), T(z, u)$

- Relational Algebra: $(R(x, y) \bowtie S(y, z)) \bowtie T(z, u)$
- First Order Logic:
  $Q_1 = \{(x, y, z, u) \mid (x, y) \in R \wedge (y, z) \in S \wedge (z, u) \in T\}$
- SQL: `select * from` $\cdots$ `where` $\cdots$

Example

$Q_2(x, u) = R(x, y), S(y, z), T(z, u)$

- Relational Algebra: $\Pi_{x,u}((R(x, y) \bowtie S(y, z)) \bowtie T(z, u))$
- First Order Logic:
  $Q_1 = \{(x, u) \mid \exists y \exists z((x, y) \in R \wedge (y, z) \in S \wedge (z, u) \in T)\}$
- SQL: `select` $\cdots$ `from` $\cdots$ `where` $\cdots$

Traditional Approach to Computing Conjunctive Queries

$$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$$

Optimizer generates a *query plan*:

$$\begin{aligned} \text{Temp}(x,y,z) =& R(x,y) \bowtie S(y,z) \\ Q(x,y,z) =& \text{Temp}(x,y,z) \bowtie T(z,x) \end{aligned}$$

Optimizers examines many possible plans, evaluates the cheapest plan.

Problem: intermediate results may be large, and very hard to estimate.

Upper Bound on the Size of the Answer

Consider the join of two relations:

$$Q(x, y, z) = R(x, y), S(y, z)$$

Question

If $|R| = m_1, |S| = m_2$, how large can $|Q|$ be?

Upper Bound on the Size of the Answer

Consider the join of two relations:

$$Q(x, y, z) = R(x, y), S(y, z)$$

Question

If $|R| = m_1, |S| = m_2$, how large can $|Q|$ be?

- Can be 0

Upper Bound on the Size of the Answer

Consider the join of two relations:

$$Q(x,y,z) = R(x,y), S(y,z)$$

Question

If $|R| = m_1, |S| = m_2$, how large can $|Q|$ be?

- Can be 0
- Can be $m_1 m_2$

Upper Bound on the Size of the Answer

Consider the join of two relations:

$$Q(x, y, z) = R(x, y), S(y, z)$$

Question

If $|R| = m_1, |S| = m_2$, how large can $|Q|$ be?

- Can be 0
- Can be $m_1 m_2$
- Answer: $0 \leq |Q| \leq m_1 m_2$.

Upper Bound on the Size of the Answer

$$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$$

Question

If $|R| = m_1$, $|S| = m_2$, $|T| = m_3$, how large can the result be?

Upper Bound on the Size of the Answer

$$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$$

Question

If $|R| = m_1$, $|S| = m_2$, $|T| = m_3$, how large can the result be?

- Naive answer: $\leq m_1 m_2 m_3$ (why?)

Upper Bound on the Size of the Answer

$$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$$

Question

If $|R| = m_1$, $|S| = m_2$, $|T| = m_3$, how large can the result be?

- Naive answer: $\leq m_1 m_2 m_3$ (why?)
- Better answer: $\leq m_1 m_2$ (why?)

Upper Bound on the Size of the Answer

$$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$$

Question

If $|R| = m_1$, $|S| = m_2$, $|T| = m_3$, how large can the result be?

- Naive answer: $\leq m_1 m_2 m_3$ (why?)
- Better answer: $\leq m_1 m_2$ (why?)
- But also: $\leq m_1 m_3$, $\leq m_2 m_3$

Upper Bound on the Size of the Answer

$$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$$

Question

If $|R| = m_1$, $|S| = m_2$, $|T| = m_3$, how large can the result be?

- Naive answer: $\leq m_1 m_2 m_3$ (why?)
- Better answer: $\leq m_1 m_2$ (why?)
- But also: $\leq m_1 m_3$, $\leq m_2 m_3$

We will show:

- Also (and better!): $|Q| \leq \sqrt{m_1 m_2 m_3}$
- There exists an algorithm that computes Q in time min(all the above)
- How this generalizes to *any* conjunctive query.

The Hypergraph of a Query

Definition

Let Q be a full conjunctive query without self-joins. The hypergraph G of Q consists of:

- Nodes(G) = Vars(Q) the set of variables of Q
- HyperEdges(G) = Atoms(Q) the set of atoms of Q.

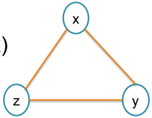
The Hypergraph of a Query

Definition

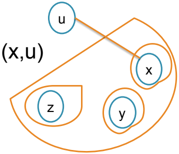
Let Q be a full conjunctive query without self-joins. The hypergraph G of Q consists of:

- Nodes(G) = Vars(Q) the set of variables of Q
- HyperEdges(G) = Atoms(Q) the set of atoms of Q.

Q(x,y,z) = R(x,y),S(y,z),T(z,x)



Q(x,y,z) = R(x,y,z),S(x),T(y),K(z),M(x,u)



Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

An *edge cover* = subset of edges that contain all nodes.

Full conjunctive query: $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$
Relation sizes: $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Proposition (Simple!)

Let $R_{i_1}, \ldots, R_{i_u}$ be any edge cover. Then $\boxed{|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}}$

(proof in class)

Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

An *edge cover* = subset of edges that contain all nodes.

Full conjunctive query: $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$
Relation sizes: $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Proposition (Simple!)

Let $R_{i_1}, \ldots, R_{i_u}$ be any edge cover. Then $\boxed{|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}}$

(proof in class)

Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

An *edge cover* = subset of edges that contain all nodes.

Full conjunctive query: $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$
Relation sizes: $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Proposition (Simple!)

Let $R_{i_1}, \ldots, R_{i_u}$ be any edge cover. Then $\boxed{|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}}$

(proof in class)

Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

An *edge cover* = subset of edges that contain all nodes.

Full conjunctive query: $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$
Relation sizes: $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Proposition (Simple!)

Let $R_{i_1}, \ldots, R_{i_u}$ be any edge cover. Then $\boxed{|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}}$

(proof in class)

Fractional Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

A *fractional edge cover* = sequence of positive numbers $u_1, \ldots, u_\ell$ s.t.:

$$\forall i : \sum_{j : x_i \in R_j} u_j \geq 1$$

Theorem (AGM'13)

Let $u_1, \ldots, u_\ell$ be any fractional edge cover. Then $\boxed{|Q| \leq m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$

Fractional Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

A *fractional edge cover* = sequence of positive numbers $u_1, \ldots, u_\ell$ s.t.:

$$\forall i : \sum_{j : x_i \in R_j} u_j \geq 1$$

Theorem (AGM'13)

Let $u_1, \ldots, u_\ell$ be any fractional edge cover. Then $\boxed{|Q| \leq m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$

Fractional Edge Cover of a Hypergraph G

G = nodes $x_1, \ldots, x_k$ and hyperedges $R_1, \ldots, R_\ell$.

Definition

A *fractional edge cover* = sequence of positive numbers $u_1, \ldots, u_\ell$ s.t.:

$$\forall i : \sum_{j : x_i \in R_j} u_j \geq 1$$

Theorem (AGM'13)

Let $u_1, \ldots, u_\ell$ be any fractional edge cover. Then $\boxed{|Q| \leq m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$

Examples

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

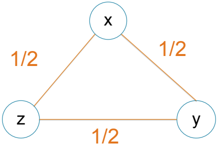
$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$
$|R| = |S| = |T| = m$

Examples

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$
$|R| = |S| = |T| = m$

A fractional edge: $\mathbf{u} = (1/2, 1/2, 1/2)$

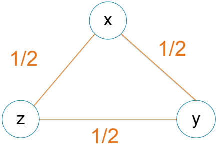


Examples

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$
$|R| = |S| = |T| = m$

A fractional edge: $\mathbf{u} = (1/2, 1/2, 1/2)$



It follows that $|Q| \leq m^{1/2} m^{1/2} m^{1/2} = m^{3/2}$

With m edges you can built at most $m^{3/2}$ triangles!

AGM Bound

Definition

$AGM(Q) = \min_{\mathbf{u}} m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}$

Thus: $|Q| \leq AGM(Q)$.

Example

$Q(x, y, z) = R(x, y), S(y, z), T(z, x),$ $|R| = m_1, |S| = m_2, |T| = m_3$

$\mathbf{u} =$	$(1, 1, 0)$	$(1, 0, 1)$	$(0, 1, 1)$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$
$AGM(Q) =$ min of	$m_1 m_2$	$m_1 m_3$	$m_2 m_3$	$(m_1 m_2 m_3)^{1/2}$

Example

$Q(x, y, z, v, w) = R(x, y), S(y, z), T(z, v), K(v, w)$

$\mathbf{u} =$	$(1, 0, 1, 1)$	$(1, 1, 0, 1)$
$AGM(Q) =$ min of	$m_1 m_3 m_4$	$m_1 m_2 m_4$

AGM Bound

Definition

$AGM(Q) = \min_{\mathbf{u}} m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}$

Thus: $|Q| \leq AGM(Q)$.

Example

$Q(x, y, z) = R(x, y), S(y, z), T(z, x),$ $\quad |R| = m_1, |S| = m_2, |T| = m_3$

$\mathbf{u} =$	$(1,1,0)$	$(1,0,1)$	$(0,1,1)$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$
$AGM(Q) =$ min of	$m_1 m_2$	$m_1 m_3$	$m_2 m_3$	$(m_1 m_2 m_3)^{1/2}$

Example

$Q(x, y, z, v, w) = R(x, y), S(y, z), T(z, v), K(v, w)$

$\mathbf{u} =$	$(1,0,1,1)$	$(1,1,0,1)$
$AGM(Q) =$ min of	$m_1 m_3 m_4$	$m_1 m_2 m_4$

AGM Bound

Definition

$AGM(Q) = \min_{\mathbf{u}} m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}$

Thus: $|Q| \le AGM(Q)$.

Example

$Q(x,y,z) = R(x,y), S(y,z), T(z,x),$ $|R| = m_1, |S| = m_2, |T| = m_3$

$\mathbf{u} =$	$(1,1,0)$	$(1,0,1)$	$(0,1,1)$	$(\frac{1}{2},\frac{1}{2},\frac{1}{2})$
$AGM(Q)$ = min of	$m_1 m_2$	$m_1 m_3$	$m_2 m_3$	$(m_1 m_2 m_3)^{1/2}$

Example

$Q(x,y,z,v,w) = R(x,y), S(y,z), T(z,v), K(v,w)$

$\mathbf{u} =$	$(1,0,1,1)$	$(1,1,0,1)$
$AGM(Q)$ = min of	$m_1 m_3 m_4$	$m_1 m_2 m_4$

The Worst-Case Query Output

Question

Can the query output ever get as large as the AGM bound?

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

The Worst-Case Query Output

Question

Can the query output ever get as large as the AGM bound?

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

Example

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$ Find three relations $|R| = |S| = |T| = m$ such that $|Q| = m^{3/2}$

The Worst-Case Query Output

Question

Can the query output ever get as large as the AGM bound?

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

Example

$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$ Find three relations $|R| = |S| = |T| = m$ such that $|Q| = m^{3/2}$

Answer: let $n = m^{1/2}$, and $R = S = T = [n] \times [n]$. Then $|Q| = n^3 = m^{3/2}$.

Background in Algebra: Duality of Linear Programs

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

$AGM(Q) = \min_{\mathbf{u}} AGM_{\mathbf{u}}(Q)$ is the optimal solution to:

$$\text{minimize} \sum_j u_j \log m_j$$

$$\forall i : \sum_{j : x_i \in R_j} u_j \geq 1$$

Background in Algebra: Duality of Linear Programs

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

$AGM(Q) = \min_{\mathbf{u}} AGM_{\mathbf{u}}(Q)$ is the optimal solution to:

$$\text{minimize} \sum_j u_j \log m_j$$
$$\forall i : \sum_{j : x_i \in R_j} u_j \geq 1$$

$$\text{maximize} \sum_i v_i$$
$$\forall j : \sum_{i : x_i \in R_j} v_i \leq \log m_j$$

Fractional Edge Cover

Fractional Vertex Packing

Background in Algebra: Duality of Linear Programs

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

$AGM(Q) = \min_{\mathbf{u}} AGM_{\mathbf{u}}(Q)$ is the optimal solution to:

$$\begin{aligned} &\text{minimize} \sum_j u_j \log m_j \\ &\forall i : \sum_{j : x_i \in R_j} u_j \geq 1 \end{aligned}$$

$$\begin{aligned} &\text{maximize} \sum_i v_i \\ &\forall j : \sum_{i : x_i \in R_j} v_i \leq \log m_j \end{aligned}$$

Fractional Edge Cover

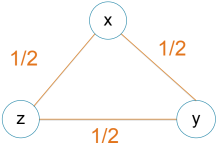
Fractional Vertex Packing

Theorem (Strong Duality of Linear Programs)

$$\min_{\mathbf{u}} \sum_j u_j \log m_j = \max_{\mathbf{v}} \sum_i v_i$$

Background in Algebra: Duality of Linear Programs

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$



$$\min(u_R \log|R| + u_S \log|S| + u_T \log|T|)$$

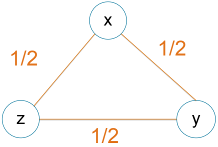
$x: \quad u_R + u_T \geq 1$

$y: \quad u_R + u_S \geq 1$

$z: \quad u_S + u_T \geq 1$

Background in Algebra: Duality of Linear Programs

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$



$$\begin{aligned} &\min(u_R \log|R| + u_S \log|S| + u_T \log|T|) \\ x: &\quad u_R + u_T \geq 1 \\ y: &\quad u_R + u_S \geq 1 \\ z: &\quad u_S + u_T \geq 1 \end{aligned}$$

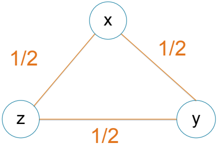
$$\begin{aligned} &\max(v_x + v_y + v_z) \\ R: &\quad v_x + v_y \leq \log|R| \\ S: &\quad v_y + v_z \leq \log|S| \\ T: &\quad v_x + v_z \leq \log|T| \end{aligned}$$

Strong duality theorem:

$\min u_R \log|R| + u_S \log|S| + u_T \log|T| = \max v_x + v_y + v_z.$

Background in Algebra: Duality of Linear Programs

$$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$$



$$\begin{array}{ll} \min(u_R \log|R| + u_S \log|S| + u_T \log|T|) & \\ x: & u_R + u_T \geq 1 \\ y: & u_R + u_S \geq 1 \\ z: & u_S + u_T \geq 1 \end{array}$$

$$\begin{array}{ll} \max(v_x + v_y + v_z) & \\ R: & v_x + v_y \leq \log|R| \\ S: & v_y + v_z \leq \log|S| \\ T: & v_x + v_z \leq \log|T| \end{array}$$

Strong duality theorem:

$\min u_R \log|R| + u_S \log|S| + u_T \log|T| = \max v_x + v_y + v_z$.

In class: what are the optimal solutions in these cases:

- $|R| = |S| = |T|$
- $|R| = |S| \ll |T|$

The AGM Bound is Tight

$$\boxed{AGM_{\mathbf{u}}(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}}$$

$AGM(Q) = \min_{\mathbf{u}} AGM_{\mathbf{u}}(Q)$ is the optimal solution to:

$$\text{minimize} \sum_j u_j \log m_j$$
$$\forall i : \sum_{j : x_i \in R_j} u_j \geq 1$$

$$\text{maximize} \sum_i v_i$$
$$\forall j : \sum_{i : x_i \in R_j} v_i \leq \log m_j$$

Theorem

The AGM bound is tight

Proof: start with an optimal solution v_i.
Define $R(x_1, x_5, x_8) = [2^{v_1}] \times [2^{v_5}] \times [2^{v_8}]$ etc
Then $|Q| = 2^{v_1 + v_2 + \cdots} = 2^{u_1 \log m_1 + u_2 \log m_2 + \cdots} = m_1^{u_1} m_2^{u_2} \cdots$

Computing Full Conjunctive Queries

- Recall: all database systems compute one join at a time
- This may be much larger than the maximum output size, $AGM(Q)$.
- Goal: design a algorithm that runs in time $AGM(Q)$.

Worst-Case-Optimal algorithm: runs in time $AGM(Q)$.

Generic Join

Compute $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$

If $|\mathbf{x}| = 1$ then return $R_1 \cap \cdots \cap R_\ell$.

Otherwise, choose a variable x

Assuem it occurrs in atoms $R_{i_1}, \ldots, R_{i_k}$

- Compute $A = \Pi_x(R_{i_1}) \cap \cdots \cap \Pi_x(R_{i_k})$
- For each $a \in A$, compute $\texttt{Result}_a = Q[a/x]$ using *Generic-Join*
- Return $\bigcup_a \texttt{Result}_a$.

Runtime: $O(AGM(Q))$ (Plus a $\log n$ factor for index lookup)

Generic Join

Compute $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$

If $|\mathbf{x}| = 1$ then return $R_1 \cap \cdots \cap R_\ell$.
Otherwise, choose a variable x
Assuem it occurs in atoms $R_{i_1}, \ldots, R_{i_k}$

- Compute $A = \Pi_x(R_{i_1}) \cap \cdots \cap \Pi_x(R_{i_k})$
- For each $a \in A$, compute $\mathtt{Result}_a = Q[a/x]$ using *Generic-Join*
- Return $\bigcup_a \mathtt{Result}_a$.

Runtime: $O(AGM(Q))$ (Plus a $\log n$ factor for index lookup)

Generic Join – Example

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$

- Compute $A = \Pi_x(R) \cap \Pi_x(T) = \{a_1, \ldots, a_n\}$
- For each $a_i \in A$, denote $R'(y) = R(a_i, y)$, $T'(z) = T(z, a_i)$
  Compute $\texttt{Result}_i(a_i, y, z) = R'(y), S(y, z), T'(z)$
- Return $\bigcup_i \texttt{Result}_i$

Runtime: $O(m^{3/2})$ assuming $|R| = |S| = |T| = m$.

Details of Generic Join

- Fix variable order: $x_1, x_2, \ldots$ (AGM bound always holds, but in practice the order can matter a lot)
- Order/index the relations accordingly. For example, $R(x_3, x_6, x_7)$ has a B+-index on x_3, x_6, x_7.
- Computing $A = \Pi_x(R_{i_1}) \cap \cdots \cap \Pi_x(R_{i_k})$ is similar to multi-way merge join. Must ensure that runtime is $\leq \min(|R_{i_1}|, |R_{i_2}|, \ldots)$.
- When we iterate $a \in A$, we are making one more binding in all indexes.
- LeapFrog Tree Join (by LogicBlox) is based on these ideas.

Details of Generic Join

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$

- Compute $A = \Pi_x(R) \cap \Pi_x(T) = \{a_1, \ldots, a_n\}$
- For each $a_i \in A$, denote $R'(y) = R(a_i, y)$, $T'(z) = T(z, a_i)$
  Compute $\texttt{Result}_i(a_i, y, z) = R'(y), S(y, z), T'(z)$
- Return $\bigcup_i \texttt{Result}_i$

Runtime: $O(m^{3/2})$ assuming $|R| = |S| = |T| = m$.

Variable order: x, y, z. What happens in each case?

- Complete bipartite graph: $R = S = T = [m^{1/2}] \times [m^{1/2}]$.
- Skewed at x: $R = [1] \times [m]$, $S \subseteq [m] \times [m]$, $T = [m] \times [1]$.
- Skewed at y: $R = [m] \times [1]$, $S = [1] \times [m]$, $T \subseteq [m] \times [m]$
- Skewed at z: $R \subseteq [m] \times [m]$, $S = [m] \times [1]$, $T \subseteq [1] \times [m]$

Discussion

- Major advantage over one-join-at-at-time algorithms *for cyclic queries!*
- For acyclic queries, the story is more complex (discussed later)
- We haven't proven its optimality yet: will do this next.

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \leq (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \leq (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Triangle: $\sum_{i,j,k} a_{ij} b_{jk} c_{ki} \leq (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}} (\sum_{j,k} b_{jk}^2)^{\frac{1}{2}} (\sum_{k,i} c_{ki}^2)^{\frac{1}{2}}$

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \leq (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Triangle: $\sum_{i,j,k} a_{ij} b_{jk} c_{ki} \leq (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}} (\sum_{j,k} b_{jk}^2)^{\frac{1}{2}} (\sum_{k,i} c_{ki}^2)^{\frac{1}{2}}$

Hölder $(u + v + w \geq 1)$: $\sum_i a_i b_i c_i \leq (\sum_i a_i^{\frac{1}{u}})^u (\sum_i b_i^{\frac{1}{v}})^v (\sum_i c_i^{\frac{1}{w}})^w$

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \leq (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Triangle: $\sum_{i,j,k} a_{ij} b_{jk} c_{ki} \leq (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}} (\sum_{j,k} b_{jk}^2)^{\frac{1}{2}} (\sum_{k,i} c_{ki}^2)^{\frac{1}{2}}$

Hölder ($u + v + w \geq 1$): $\sum_i a_i b_i c_i \leq (\sum_i a_i^{\frac{1}{u}})^u (\sum_i b_i^{\frac{1}{v}})^v (\sum_i c_i^{\frac{1}{w}})^w$

Theorem (Friedgut'04)

Let $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$ be a query and $u_1, \ldots, u_\ell$ be a fractional edge cover. Then:

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}^{\frac{1}{u_1}}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}^{\frac{1}{u_\ell}}\right)^{u_\ell}}$$

What are the queries in the examples above?

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \le (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Triangle: $\sum_{i,j,k} a_{ij} b_{jk} c_{ki} \le (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}} (\sum_{j,k} b_{jk}^2)^{\frac{1}{2}} (\sum_{k,i} c_{ki}^2)^{\frac{1}{2}}$

Hölder ($u + v + w \ge 1$): $\sum_i a_i b_i c_i \le (\sum_i a_i^{\frac{1}{u}})^u (\sum_i b_i^{\frac{1}{v}})^v (\sum_i c_i^{\frac{1}{w}})^w$

Theorem (Friedgut'04)

Let $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$ be a query and $u_1, \ldots, u_\ell$ be a fractional edge cover. Then:

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell} \le \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}^{\frac{1}{u_1}}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}^{\frac{1}{u_\ell}}\right)^{u_\ell}}$$

What are the queries in the examples above?

$Q_{\text{Cauchy-Schwartz}}(x) = R(x), S(x);$

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \leq (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Triangle: $\sum_{i,j,k} a_{ij} b_{jk} c_{ki} \leq (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}} (\sum_{j,k} b_{jk}^2)^{\frac{1}{2}} (\sum_{k,i} c_{ki}^2)^{\frac{1}{2}}$

Hölder ($u + v + w \geq 1$): $\sum_i a_i b_i c_i \leq (\sum_i a_i^{\frac{1}{u}})^u (\sum_i b_i^{\frac{1}{v}})^v (\sum_i c_i^{\frac{1}{w}})^w$

Theorem (Friedgut'04)

Let $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$ be a query and $u_1, \ldots, u_\ell$ be a fractional edge cover. Then:

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}^{\frac{1}{u_1}}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}^{\frac{1}{u_\ell}}\right)^{u_\ell}}$$

What are the queries in the examples above?

$Q_{\text{Cauchy-Schwartz}}(x) = R(x), S(x)$;

$Q_{\text{triangle}}(x, y, z) = R(x, y), S(y, z), T(z, x)$;

Friedgut's Inequality

Cauchy-Schwartz: $\sum_i a_i b_i \leq (\sum_i a_i^2)^{\frac{1}{2}} (\sum_i b_i^2)^{\frac{1}{2}}$

Triangle: $\sum_{i,j,k} a_{ij} b_{jk} c_{ki} \leq (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}} (\sum_{j,k} b_{jk}^2)^{\frac{1}{2}} (\sum_{k,i} c_{ki}^2)^{\frac{1}{2}}$

Hölder ($u + v + w \geq 1$): $\sum_i a_i b_i c_i \leq (\sum_i a_i^{\frac{1}{u}})^u (\sum_i b_i^{\frac{1}{v}})^v (\sum_i c_i^{\frac{1}{w}})^w$

Theorem (Friedgut'04)

Let $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$ be a query and $u_1, \ldots, u_\ell$ be a fractional edge cover. Then:

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}^{\frac{1}{u_1}}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}^{\frac{1}{u_\ell}}\right)^{u_\ell}}$$

What are the queries in the examples above?

$Q_{\text{Cauchy-Schwartz}}(x) = R(x), S(x)$;

$Q_{\text{triangle}}(x, y, z) = R(x, y), S(y, z), T(z, x)$;

$Q_{\text{Hölder}}(x) = R(x), S(x), T(x)$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \ge 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \le (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \geq 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \leq (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Induction Step. Pick a variable x, and remove it. For example,

$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$ becomes $Q'(y, z) = R'(y), S(y, z), T'(z)$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \ge 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \le (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Induction Step. Pick a variable x, and remove it. For example,

$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$ becomes $Q'(y,z) = R'(y), S(y,z), T'(z)$

$$\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_{x} a_{xy}^{u_1} c_{zx}^{u_3} \qquad \text{group by } \sum_x$$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \ge 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \le (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Induction Step. Pick a variable x, and remove it. For example,
$Q(x, y, z) = R(x, y), S(y, z), T(z, x)$ becomes $Q'(y, z) = R'(y), S(y, z), T'(z)$

$$\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_{x} a_{xy}^{u_1} c_{zx}^{u_3} \qquad \text{group by } \textstyle\sum_x$$

$$\le \sum_{yz} b_{yz}^{u_2} \Big(\sum_{x} a_{xy}\Big)^{u_1} \Big(\sum_{x} c_{zx}\Big)^{u_3} \qquad \text{Hölder } u_1 + u_3 \ge 1$$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \geq 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \leq (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Induction Step. Pick a variable x, and remove it. For example,

$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$ becomes $Q'(y,z) = R'(y), S(y,z), T'(z)$

$$\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_{x} a_{xy}^{u_1} c_{zx}^{u_3} \qquad \text{group by } \textstyle\sum_x$$

$$\leq \sum_{yz} b_{yz}^{u_2} (\sum_{x} a_{xy})^{u_1} (\sum_{x} c_{zx})^{u_3} \qquad \text{Hölder } u_1 + u_3 \geq 1$$

$$= \sum_{yz} b_{yz}^{u_2} A_y^{u_1} C_z^{u_3} \leq (\sum_{yz} b_{yz})^{u_2} (\sum_{y} A_y)^{u_1} (\sum_{z} C_z)^{u_3} \qquad \text{Induction for } Q'$$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \ge 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \le (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Induction Step. Pick a variable x, and remove it. For example,

$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$ becomes $Q'(y,z) = R'(y), S(y,z), T'(z)$

$$\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_x a_{xy}^{u_1} c_{zx}^{u_3} \qquad \text{group by } \textstyle\sum_x$$

$$\le \sum_{yz} b_{yz}^{u_2} \Big(\sum_x a_{xy}\Big)^{u_1} \Big(\sum_x c_{zx}\Big)^{u_3} \qquad \text{Hölder } u_1 + u_3 \ge 1$$

$$= \sum_{yz} b_{yz}^{u_2} A_y^{u_1} C_z^{u_3} \le \Big(\sum_{yz} b_{yz}\Big)^{u_2} \Big(\sum_y A_y\Big)^{u_1} \Big(\sum_z C_z\Big)^{u_3} \qquad \text{Induction for } Q'$$

$$= \Big(\sum_{yz} b_{yz}\Big)^{u_2} \Big(\sum_{xy} a_{xy}\Big)^{u_1} \Big(\sum_{zx} c_{zx}\Big)^{u_3}$$

Friedgut's Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\boxed{\sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell}}$$

Proof: by induction on $|\mathbf{x}|$

Base Case. $|\mathbf{x}| = 1$: $Q(x) = R_1(x), \ldots, R_\ell(x)$, $u_1 + \ldots + u_\ell \geq 1$

Prove: $\sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \leq (\sum_x a_{1,x})^{u_1} \cdots (\sum_x a_{\ell,x})^{u_\ell}$ This is Hölder.

Induction Step. Pick a variable x, and remove it. For example,

$Q(x,y,z) = R(x,y), S(y,z), T(z,x)$ becomes $Q'(y,z) = R'(y), S(y,z), T'(z)$

$$\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_x a_{xy}^{u_1} c_{zx}^{u_3} \qquad \text{group by } \textstyle\sum_x$$

$$\leq \sum_{yz} b_{yz}^{u_2} \Big(\sum_x a_{xy}\Big)^{u_1} \Big(\sum_x c_{zx}\Big)^{u_3} \qquad \text{Hölder } u_1 + u_3 \geq 1$$

$$= \sum_{yz} b_{yz}^{u_2} A_y^{u_1} C_z^{u_3} \leq \Big(\sum_{yz} b_{yz}\Big)^{u_2} \Big(\sum_y A_y\Big)^{u_1} \Big(\sum_z C_z\Big)^{u_3} \qquad \text{Induction for } Q'$$

$$= \Big(\sum_{yz} b_{yz}\Big)^{u_2} \Big(\sum_{xy} a_{xy}\Big)^{u_1} \Big(\sum_{zx} c_{zx}\Big)^{u_3}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

Sizes $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \le m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \ldots, R_\ell$.

For every $j = 1, \ldots, \ell$, and every tuple $\mathbf{x}_j \in \mathrm{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \mathrm{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \mathrm{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \Big(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\Big)^{u_1} \cdots \Big(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\Big)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

Sizes $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \le m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \ldots, R_\ell$.

For every $j = 1, \ldots, \ell$, and every tuple $\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \text{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \Big(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\Big)^{u_1} \cdots \Big(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\Big)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

Sizes $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \leq m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \ldots, R_\ell$.

For every $j = 1, \ldots, \ell$, and every tuple $\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \text{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \Big(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\Big)^{u_1} \cdots \Big(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\Big)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

Sizes $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \leq m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \ldots, R_\ell$.

For every $j = 1, \ldots, \ell$, and every tuple $\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \text{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \Big(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\Big)^{u_1} \cdots \Big(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\Big)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

Sizes $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \le m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \ldots, R_\ell$.

For every $j = 1, \ldots, \ell$, and every tuple $\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \text{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \left(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\right)^{u_1} \cdots \left(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\right)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \ldots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \ldots, u_\ell$

Sizes $|R_1| = m_1, \ldots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \leq m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \ldots, R_\ell$.

For every $j = 1, \ldots, \ell$, and every tuple $\mathbf{x}_j \in \mathrm{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \mathrm{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \mathrm{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \leq \Big(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\Big)^{u_1} \cdots \Big(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\Big)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

The AGM Inequality – Proof

Query $Q(\mathbf{x}) = R_1(\mathbf{x}_1), \dots, R_\ell(\mathbf{x}_\ell)$, fractional cover $u_1, \dots, u_\ell$

Sizes $|R_1| = m_1, \dots, |R_\ell| = m_\ell$

Prove $\boxed{|Q| \le m_1^{u_1} \cdots m_\ell^{u_\ell}}$

Let Dom = the domain of all constants in the relations $R_1, \dots, R_\ell$.

For every $j = 1, \dots, \ell$, and every tuple $\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}$, define:

$$a_{j,\mathbf{x}_j} = \begin{cases} 1 & \text{if the tuple } \mathbf{x}_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}$$

Then: $m_j = |R_j| = \sum_{\mathbf{x}_j \in \text{Dom}^{|\mathbf{x}_j|}} a_{j,\mathbf{x}_j}$, $|Q| = \sum_{\mathbf{x} \in \text{Dom}^{|\mathbf{x}|}} a_{1,\mathbf{x}_1} \cdots a_{\ell,\mathbf{x}_\ell}$

Now use Friedgut's inequality:

$$|Q| = \sum_{\mathbf{x}} a_{1,\mathbf{x}_1}^{u_1} \cdots a_{\ell,\mathbf{x}_\ell}^{u_\ell} \le \Big(\sum_{\mathbf{x}_1} a_{1,\mathbf{x}_1}\Big)^{u_1} \cdots \Big(\sum_{\mathbf{x}_\ell} a_{\ell,\mathbf{x}_\ell}\Big)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED

Proof of Optimality for Generic Join

Use *exactly* the same induction step as in Friedgut's inequality.