Advanced Query Processing

CSEP544

Nov. 2015
Lecture 9: Advanced Query Processing

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

- 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

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: \( O(|\text{Input}| + |\text{Output}|) \)
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:
  \[ Q_1 = \{(x, y, z, u) \mid (x, y) \in R \land (y, z) \in S \land (z, u) \in T\} \]
- SQL: `select * from ... where ...`
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:
  \[ Q_1 = \{ (x, y, z, u) \mid (x, y) \in R \land (y, z) \in S \land (z, u) \in T \} \]
- SQL: `select * from ... where ...`

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:
  \[ Q_1 = \{ (x, u) \mid \exists y \exists z ((x, y) \in R \land (y, z) \in S \land (z, u) \in T) \} \]
- SQL: `select ... from ... where ...`
Traditional Approach to Computing Conjunctive Queries

\[ Q(x, y, z) = R(x, y), S(y, z), T(z, x) \]

Optimizer generates a \textit{query plan}:

\[ \text{Temp}(x, y, z) = R(x, y) \bowtie S(y, z) \]
\[ Q(x, y, z) = \text{Temp}(x, y, z) \bowtie T(z, x) \]

Optimizers examines many possible plans, evaluates the cheapest plan.

\textbf{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?

Answer: \(0 \leq |Q| \leq m_1 m_2\).
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
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 \( \min(\text{all the above}) \)
- 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:

- $\text{Nodes}(G) = \text{Vars}(Q)$ the set of variables of $Q$
- $\text{HyperEdges}(G) = \text{Atoms}(Q)$ 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:

- $\text{Nodes}(G) = \text{Vars}(Q)$ the set of variables of $Q$
- $\text{HyperEdges}(G) = \text{Atoms}(Q)$ 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 $G$

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

**Definition**

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

**Full conjunctive query:** $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$

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 $|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}$

(proof in class)
Edge Cover of a Hypergraph $G$

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

**Definition**

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

Full conjunctive query: $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$

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 $|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}$

(proof in class)
Edge Cover of a Hypergraph $G$

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

**Definition**

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

Full conjunctive query: $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$

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 $|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}$

(proof in class)
Edge Cover of a Hypergraph $G$

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

**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 $|Q| \leq m_{i_1} \cdot m_{i_2} \cdots m_{i_u}$

(proof in class)
Fractional Edge Cover of a Hypergraph $G$

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

**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

\[ |Q| \leq m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]
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

$$|Q| \leq m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}$$
Fractional Edge Cover of a Hypergraph $G$

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

**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

$$|Q| \leq m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}$$
Examples

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

\[ Q(x, y, z) = R(x, y), S(y, z), T(z, x) \]

\[ |R| = |S| = |T| = m \]
Examples

$\text{AGM}_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}$

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

A fractional edge: $u = (1/2, 1/2, 1/2)$
Examples

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

\[ Q(x, y, z) = R(x, y), S(y, z), T(z, x) \]
\[ |R| = |S| = |T| = m \]

A fractional edge: \( u = (1/2, 1/2, 1/2) \)

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**

\[
AGM(Q) = \min_u m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell}
\]

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

**Example**

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

\[
\begin{array}{cccc}
(1, 1, 0) & (1, 0, 1) & (0, 1, 1) & (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \\
1 & 1 & 1 & 1 \\
\end{array}
\]

**Example**

\[
Q(x, y, z, v, w) = R(x, y), S(y, z), T(z, v), K(v, w)
\]

\[
\begin{array}{cccc}
(1, 0, 1, 1) & (1, 1, 0, 1) \\
1 & 1 & 1 & 1 \\
\end{array}
\]
AGM Bound

Definition

\[ AGM(Q) = \min_u m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

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

Example

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

\[ u = \begin{array}{c}
(1, 1, 0) \\
m_1 m_2 \\
(1, 0, 1) \\
m_1 m_3 \\
(0, 1, 1) \\
m_2 m_3 \\
(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \\
(m_1 m_2 m_3)^{1/2}
\end{array} \]

Example

\[ Q(x, y, z, v, w) = R(x, y), S(y, z), T(z, v), K(v, w) \]

\[ u = \begin{array}{c}
(1, 0, 1, 1) \\
m_1 m_3 m_4 \\
(1, 1, 0, 1) \\
m_1 m_2 m_4
\end{array} \]
AGM Bound

Definition

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

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

Example

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

$$u = (1, 1, 0) \quad (1, 0, 1) \quad (0, 1, 1) \quad \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$$

$$AGM(Q) = \min$$

$$\begin{array}{cccc}
 m_1 m_2 & m_1 m_3 & m_2 m_3 & (m_1 m_2 m_3)^{1/2} \\
\end{array}$$

Example

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

$$u = (1, 0, 1, 1) \quad (1, 1, 0, 1)$$

$$AGM(Q) = \min$$

$$\begin{array}{cc}
 m_1 m_3 m_4 & m_1 m_2 m_4 \\
\end{array}$$
The Worst-Case Query Output

**Question**

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

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]
The Worst-Case Query Output

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

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_{\ell}^{u_{\ell}} \]

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?

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

**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 \( R = S = T = \lfloor n \rfloor \times \lfloor n \rfloor \). Then \( |Q| = n^3 = m^{3/2} \).
Background in Algebra: Duality of Linear Programs

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

\[ AGM(Q) = \min_u AGM_u(Q) \]
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

\[ \text{AGM}_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

\[ \text{AGM}(Q) = \min_u \text{AGM}_u(Q) \]

is the optimal solution to:

**minimize** \( \sum_j u_j \log m_j \)

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

**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

\[ \text{AGM}_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

\[ \text{AGM}(Q) = \min_u \text{AGM}_u(Q) \] is the optimal solution to:

\[
\begin{align*}
\text{minimize} & \sum_j u_j \log m_j \\
\text{subject to} & \sum_{j: x_i \in R_j} u_j \geq 1 \\
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \sum_i v_i \\
\text{subject to} & \sum_{i: x_i \in R_j} v_i \leq \log m_j \\
\end{align*}
\]

Fractional Edge Cover

Fractional Vertex Packing

**Theorem (Strong Duality of Linear Programs)**

\[
\min_u \sum_j u_j \log m_j = \max_v \sum_i v_i
\]
Background in Algebra: Duality of Linear Programs

\[ Q(x, y, z) = R(x, y), S(y, z), T(z, x) \]

\[
\begin{align*}
\min & \left( u_R \log |R| + u_S \log |S| + u_T \log |T| \right) \\
x : & \quad u_R + u_T \geq 1 \\
y : & \quad u_R + u_S \geq 1 \\
z : & \quad u_S + u_T \geq 1
\end{align*}
\]
Background in Algebra: Duality of Linear Programs

\[ Q(x, y, z) = R(x, y), S(y, z), T(z, x) \]

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

\[
\begin{align*}
\max(v_x + v_y + v_z) \\
R: & \quad v_x + v_y \leq \log |R| \\
S: & \quad v_y + v_z \leq \log |S| \\
T: & \quad v_x + v_z \leq \log |T|
\end{align*}
\]

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

\[ Q(x, y, z) = R(x, y), S(y, z), T(z, x) \]

\[
\begin{align*}
\text{min}(u_R \log |R| + u_S \log |S| + u_T \log |T|) & \quad \text{max}(v_x + v_y + v_z) \\
\text{x: } u_R + u_T & \geq 1 \\
\text{y: } u_R + u_S & \geq 1 \\
\text{z: } u_S + u_T & \geq 1 \\
\end{align*}
\]

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

\[ AGM_u(Q) = m_1^{u_1} \cdot m_2^{u_2} \cdots m_\ell^{u_\ell} \]

\[ AGM(Q) = \min_u AGM_u(Q) \]

is the optimal solution to:

\[
\begin{align*}
\text{minimize} & \sum_j u_j \log m_j \\
\forall i : & \sum_{j : x_i \in R_j} u_j \geq 1
\end{align*}
\]

\[ \begin{align*}
\text{maximize} & \sum_i v_i \\
\forall j : & \sum_{i : x_i \in R_j} v_i \leq \log m_j
\end{align*} \]

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 \( |Q| = 2^{v_1 + v_2 + \ldots} = 2^{u_1 \log m_1 + u_2 \log m_2 + \ldots} = m_1^{u_1} m_2^{u_2} \ldots \)
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 an algorithm that runs in time $AGM(Q)$.

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

Compute $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$

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

Otherwise, choose a variable $x$.

Assume 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 $\text{Result}_a = Q[a/x]$ using Generic-Join
- Return $\bigcup_a \text{Result}_a$.

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

Compute $Q(x) = R_1(x_1), \ldots, R_{\ell}(x_{\ell})$

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

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

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 \( \text{Result}_i(a_i, y, z) = R'(y), S(y, z), T'(z) \)
- Return \( \bigcup_i \text{Result}_i \)

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

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 \( x \): \( R = [1] \times [m], S = [m] \times [m], T = [m] \times [1] \).
- Skewed at \( y \): \( R = [m] \times [1], S = [1] \times [m], T = [m] \times [m] \).
- Skewed at \( z \): \( R = [m] \times [m], S = [m] \times [1], T = [1] \times [m] \).
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: \[ \sum_i a_i b_i \leq \left( \sum_i a_i^2 \right)^{\frac{1}{2}} \left( \sum_i b_i^2 \right)^{\frac{1}{2}} \]
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: \[ \sum_i a_i b_i \leq \left( \sum_i a_i^2 \right)^{\frac{1}{2}} \left( \sum_i b_i^2 \right)^{\frac{1}{2}} \]

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

Hölder \((u + v + w \geq 1)\): \[ \sum_i a_i b_i c_i \leq \left( \sum_i a_i^u \right)^{\frac{1}{u}} \left( \sum_i b_i^v \right)^{\frac{1}{v}} \left( \sum_i c_i^w \right)^{\frac{1}{w}} \]
Friedgut’s Inequality

Cauchy-Schwartz:  \[ \sum_i a_i b_i \leq \left( \sum_i a_i^2 \right)^{\frac{1}{2}} \left( \sum_i b_i^2 \right)^{\frac{1}{2}} \]

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

Hölder \((u + v + w \geq 1)\):  \[ \sum_i a_i b_i c_i \leq \left( \sum_i a_i^u \right)^u \left( \sum_i b_i^v \right)^v \left( \sum_i c_i^w \right)^w \]

**Theorem (Friedgut’04)**

Let \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \) be a query and \( u_1, \ldots, u_\ell \) be a fractional edge cover. Then:

\[
\sum_x a_{1,x_1} \cdots a_{\ell,x_\ell} \leq \left( \sum_{x_1} a_{1,x_1}^{\frac{1}{u_1}} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell}^{\frac{1}{u_\ell}} \right)^{u_\ell}
\]

What are the queries in the examples above?
Friedgut’s Inequality

Cauchy-Schwartz: \[ \sum_i a_i b_i \leq \left( \sum_i a_i^2 \right)^{\frac{1}{2}} \left( \sum_i b_i^2 \right)^{\frac{1}{2}} \]

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

Hölder \((u + v + w \geq 1)\): \[ \sum_i a_i b_i c_i \leq \left( \sum_i a_i^u \right)^u \left( \sum_i b_i^v \right)^v \left( \sum_i c_i^w \right)^w \]

**Theorem (Friedgut’04)**

Let \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \) be a query and \( u_1, \ldots, u_\ell \) be a fractional edge cover. Then:

\[
\sum_x a_{1,x_1} \cdots a_{\ell,x_\ell} \leq \left( \sum_{x_1} a_{1,x_1}^{u_1} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell}^{u_\ell} \right)^{u_\ell}
\]

What are the queries in the examples above?

\( Q_{\text{Cauchy-Schwartz}}(x) = R(x), S(x); \)
Friedgut’s Inequality

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

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

Hölder \((u + v + w \geq 1)\): \( \sum_i a_i b_i c_i \leq \left( \sum_i a_i^u \right)^u \left( \sum_i b_i^v \right)^v \left( \sum_i c_i^w \right)^w \)

Theorem (Friedgut’04)

Let \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \) be a query and \( u_1, \ldots, u_\ell \) be a fractional edge cover. Then:

\[
\sum_x a_{1,x_1} \cdots a_{\ell,x_\ell} \leq \left( \sum_{x_1} a_{1,x_1}^{u_1} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell}^{u_\ell} \right)^{u_\ell}
\]

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:**
\[ \sum_i a_i b_i \leq \left( \sum_i a_i^2 \right)^{\frac{1}{2}} \left( \sum_i b_i^2 \right)^{\frac{1}{2}} \]

**Triangle:**
\[ \sum_{i,j,k} a_{ij} b_{jk} c_{ki} \leq \left( \sum_{i,j} a_{ij}^2 \right)^{\frac{1}{2}} \left( \sum_{j,k} b_{jk}^2 \right)^{\frac{1}{2}} \left( \sum_{k,i} c_{ki}^2 \right)^{\frac{1}{2}} \]

**Hölder \((u + v + w \geq 1)\):**
\[ \sum_i a_i b_i c_i \leq \left( \sum_i a_i^u \right)^u \left( \sum_i b_i^v \right)^v \left( \sum_i c_i^w \right)^w \]

**Theorem (Friedgut’04)**

Let \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \) be a query and \( u_1, \ldots, u_\ell \) be a fractional edge cover. Then:
\[ \sum_x a_{1,x_1} \cdots a_{\ell,x_\ell} \leq \left( \sum_{x_1} a_{1,x_1}^{\frac{1}{u_1}} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell}^{\frac{1}{u_\ell}} \right)^{u_\ell} \]

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 \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell, x_\ell}^{u_\ell} \leq \left( \sum_{x_1} a_{1,x_1} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell} \right)^{u_\ell}
\]
Friedgut’s Inequality – Proof

Query $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq \left( \sum_{x_1} a_{1,x_1} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell} \right)^{u_\ell}$$

**Proof:** by induction on $|x|$
Friedgut’s Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_{x_1} a_{1,x_1})^{u_1} \cdots (\sum_{x_\ell} a_{\ell,x_\ell})^{u_\ell}
\]

**Proof:** by induction on \(|x|\)

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

Prove: \( \sum_x a_{1,x}^{u_1} \cdots a_{\ell,x}^{u_\ell} \leq (\sum_{x_1} a_{1,x})^{u_1} \cdots (\sum_{x_\ell} a_{\ell,x})^{u_\ell} \) This is Hölder.
Friedgut’s Inequality – Proof

Query $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$, fractional cover $u_1, \ldots, u_\ell$

$$\sum_x a_1^{u_1} \cdots a_\ell^{u_\ell} \leq (\sum_{x_1} a_1^{x_1})^{u_1} \cdots (\sum_{x_\ell} a_\ell^{x_\ell})^{u_\ell}$$

Proof: by induction on $|x|$

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

Prove: $\sum_x a_1^{u_1} \cdots a_\ell^{u_\ell} \leq (\sum_x a_1^{x_1})^{u_1} \cdots (\sum_x a_\ell^{x_\ell})^{u_\ell}$ 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 \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq \left( \sum x_1 a_{1,x_1} \right)^{u_1} \cdots \left( \sum x_\ell a_{\ell,x_\ell} \right)^{u_\ell}
\]

Proof: by induction on \( |x| \)

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

Prove: \( \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq \left( \sum x_1 a_{1,x_1} \right)^{u_1} \cdots \left( \sum x_\ell a_{\ell,x_\ell} \right)^{u_\ell} \) 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 \( Q'(y, z) = R'(y), S(y, z), T'(z) \)

\[
\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_x a_{xy}^{u_1} c_{zx}^{u_3} \quad \text{group by } \sum_x
\]
Friedgut’s Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_x a_{1,x_1})^{u_1} \cdots (\sum_x a_{\ell,x_\ell})^{u_\ell}
\]

**Proof:** by induction on \( |x| \)

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

Prove: \( \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_x a_{1,x_1})^{u_1} \cdots (\sum_x a_{\ell,x_\ell})^{u_\ell} \) 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 \( Q'(y,z) = R'(y), S(y,z), T'(z) \)

\[
\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_x a_{xy}^{u_1} c_{zx}^{u_3} \quad \text{group by } \sum_x
\]

\[
\leq \sum_{yz} b_{yz}^{u_2} (\sum_x a_{xy})^{u_1} (\sum_x c_{zx})^{u_3} \quad \text{Hölder } u_1 + u_3 \geq 1
\]
Friedgut’s Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum x \ a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq \left( \sum x_1 \ a_{1,x_1} \right)^{u_1} \cdots \left( \sum x_\ell \ a_{\ell,x_\ell} \right)^{u_\ell}
\]

**Proof:** by induction on \( |x| \)

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

Prove: \( \sum x \ a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq \left( \sum x_1 \ a_{1,x_1} \right)^{u_1} \cdots \left( \sum x_\ell \ a_{\ell,x_\ell} \right)^{u_\ell} \) 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 \( Q'(y,z) = R'(y), S(y,z), T'(z) \)

\[
\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_x a_{xy}^{u_1} c_{zx}^{u_3} \leq \sum_{yz} b_{yz}^{u_2} \left( \sum_x a_{xy} \right)^{u_1} \left( \sum_x c_{zx} \right)^{u_3} \leq \sum_{yz} b_{yz}^{u_2} A_y^{u_1} C_z^{u_3} \leq \left( \sum_{yz} b_{yz} \right)^{u_2} \left( \sum_y A_y \right)^{u_1} \left( \sum_z C_z \right)^{u_3}
\]

Hölder \( u_1 + u_3 \geq 1 \)

Group by \( \sum_x \)

Induction for \( Q' \)
Friedgut’s Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_x a_{1,x_1})^{u_1} \cdots (\sum_x a_{\ell,x_\ell})^{u_\ell}
\]

**Proof:** by induction on \( |x| \)

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

Prove: \( \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_x a_{1,x_1})^{u_1} \cdots (\sum_x a_{\ell,x_\ell})^{u_\ell} \) 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 \( Q'(y, z) = R'(y), S(y, z), T'(z) \)

\[
\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_{x} a_{xy}^{u_1} c_{zx}^{u_3}
\]

\[
\leq \sum_{yz} b_{yz}^{u_2} (\sum_{x} a_{xy})^{u_1} (\sum_{x} c_{zx})^{u_3}
\]

\[
= \sum_{yz} b_{yz}^{u_2} A_y^{u_1} C_z^{u_3} \leq (\sum_{yz} b_{yz})^{u_2} (\sum_{y} A_y)^{u_1} (\sum_{z} C_z)^{u_3}
\]

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

Hölder \( u_1 + u_3 \geq 1 \)  

Induction for \( Q' \)
Friedgut’s Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_{\ell}) \), fractional cover \( u_1, \ldots, u_\ell \)

\[
\sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_{\ell}}^{u_\ell} \leq \left( \sum_x a_{1,x_1} \right)^{u_1} \cdots \left( \sum_x a_{\ell,x_{\ell}} \right)^{u_\ell}
\]

Proof: by induction on \( |x| \)

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

Prove: \( \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_{\ell}}^{u_\ell} \leq \left( \sum_x a_{1,x_1} \right)^{u_1} \cdots \left( \sum_x a_{\ell,x_{\ell}} \right)^{u_\ell} \) 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 \( Q'(y, z) = R'(y), S(y, z), T'(z) \)

\[
\sum_{xyz} a_{xy}^{u_1} b_{yz}^{u_2} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \sum_{x} a_{xy}^{u_1} c_{zx}^{u_3} = \sum_{yz} b_{yz}^{u_2} \left( \sum_{x} a_{xy} \right)^{u_1} \left( \sum_{x} c_{zx} \right)^{u_3}
\]

group by \( \sum_x \)

Hölder \( u_1 + u_3 \geq 1 \)

\[
= \sum_{yz} b_{yz}^{u_2} A_y^{u_1} C_z^{u_3} \leq \left( \sum_{yz} b_{yz} \right)^{u_2} \left( \sum_{y} A_y \right)^{u_1} \left( \sum_{z} C_z \right)^{u_3}
\]

Induction for \( Q' \)

\[
= \left( \sum_{yz} b_{yz} \right)^{u_2} \left( \sum_{xy} a_{xy} \right)^{u_1} \left( \sum_{zx} c_{zx} \right)^{u_3}
\]

QED
The AGM Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

Sizes \( |R_1| = m_1, \ldots, |R_\ell| = m_\ell \)

Prove \( |Q| \leq m_1^{u_1} \cdots m_\ell^{u_\ell} \) 

Let \( \text{Dom} = \) the domain of all constants in the relations \( R_1, \ldots, R_\ell \).

For every \( j = 1, \ldots, \ell \), and every tuple \( x_j \in \text{Dom} \backslash x_j \), define:

\[
 a_{j,x_j} = \begin{cases} 1 & \text{if the tuple } x_j \text{ belongs to } R_j \\ 0 & \text{otherwise} \end{cases}
\]

Then: \( m_j = |R_j| = \sum_{x_j \in \text{Dom} \backslash x_j} a_{j,x_j} \), \( |Q| = \sum_{x \in \text{Dom} \backslash x} a_{1,x_1} \cdots a_{\ell,x_\ell} \)

Now use Friedgut’s inequality:

\[
 |Q| = \sum_{x} a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq \left( \sum_{x_1} a_{1,x_1} \right)^{u_1} \cdots \left( \sum_{x_\ell} a_{\ell,x_\ell} \right)^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}
\]

QED
The AGM Inequality – Proof

Query $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$, fractional cover $u_1, \ldots, u_\ell$

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

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

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

For every $j = 1, \ldots, \ell$, and every tuple $x_j \in \text{Dom} \setminus x_j$, define:

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

Then: $m_j = |R_j| = \sum_{x_j \in \text{Dom} \setminus x_j} a_{j, x_j}$, $|Q| = \sum_{x \in \text{Dom} \setminus x} a_{1, x_1} \cdots a_{\ell, x_\ell}$

Now use Friedgut’s inequality:

$$|Q| = \sum_x a_{1, x_1}^{u_1} \cdots a_{\ell, x_\ell}^{u_\ell} \leq \left( \sum_{x_1} a_{1, x_1}^{u_1} \right) \cdots \left( \sum_{x_\ell} a_{\ell, x_\ell}^{u_\ell} \right) = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED
The AGM Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

Sizes \( |R_1| = m_1, \ldots, |R_\ell| = m_\ell \)

Prove \( |Q| \leq m_1^{u_1} \cdots m_\ell^{u_\ell} \)

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

For every \( j = 1, \ldots, \ell \), and every tuple \( x_j \in \text{Dom} \mid x_j \mid \), define:

\[
a_{j, x_j} = \begin{cases} 
1 & \text{if the tuple } x_j \text{ belongs to } R_j \\
0 & \text{otherwise}
\end{cases}
\]

Then: \( m_j = |R_j| = \sum_{x_j \in \text{Dom} \mid x_j \mid} a_{j, x_j} \), \( |Q| = \sum_{x \in \text{Dom} \mid x \mid} a_{1, x_1} \cdots a_{\ell, x_\ell} \)

Now use Friedgut’s inequality:

\[
|Q| = \sum_{x} a_{1, x_1}^{u_1} \cdots a_{\ell, x_\ell}^{u_\ell} \leq (\sum_{x_1} a_{1, x_1}^{u_1}) \cdots (\sum_{x_\ell} a_{\ell, x_\ell}^{u_\ell}) = m_1^{u_1} \cdots m_\ell^{u_\ell}
\]

QED
The AGM Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

Sizes \( |R_1| = m_1, \ldots, |R_\ell| = m_\ell \)

Prove \( |Q| \leq m_1^{u_1} \cdots m_\ell^{u_\ell} \)

Let \( \text{Dom} = \) the domain of all constants in the relations \( R_1, \ldots, R_\ell \).

For every \( j = 1, \ldots, \ell \), and every tuple \( x_j \in \text{Dom}^{x_j} \), define:

\[
a_{j,x_j} = \begin{cases} 
1 & \text{if the tuple } x_j \text{ belongs to } R_j \\
0 & \text{otherwise}
\end{cases}
\]

Then: \( m_j = |R_j| = \sum_{x_j \in \text{Dom}^{x_j}} a_{j,x_j}, \quad |Q| = \sum_{x \in \text{Dom}^x} a_{1,x_1} \cdots a_{\ell,x_\ell} \)

Now use Friedgut’s inequality:

\[
|Q| = \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_{x_1} a_{1,x_1})^{u_1} \cdots (\sum_{x_\ell} a_{\ell,x_\ell})^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}
\]

QED
The AGM Inequality – Proof

Query \( Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell) \), fractional cover \( u_1, \ldots, u_\ell \)

Sizes \( |R_1| = m_1, \ldots, |R_\ell| = m_\ell \)

Prove \( |Q| \leq \prod_{j=1}^{\ell} m_j^{u_j} \)

Let \( \text{Dom} = \) the domain of all constants in the relations \( R_1, \ldots, R_\ell \).

For every \( j = 1, \ldots, \ell \), and every tuple \( x_j \in \text{Dom} |x_j| \), define:

\[
a_{j,x_j} = \begin{cases} 
1 & \text{if the tuple } x_j \text{ belongs to } R_j \\
0 & \text{otherwise}
\end{cases}
\]

Then: \( m_j = |R_j| = \sum_{x_j \in \text{Dom} |x_j|} a_{j,x_j} \), \( |Q| = \sum_{x \in \text{Dom} |x|} a_{1,x_1} \cdots a_{\ell,x_\ell} \)

Now use Friedgut’s inequality:

\[
|Q| = \sum_{x} a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_{x_1} a_{1,x_1}^{u_1})^{u_1} \cdots (\sum_{x_\ell} a_{\ell,x_\ell}^{u_\ell})^{u_\ell} = \prod_{j=1}^{\ell} m_j^{u_j}
\]

QED
The AGM Inequality – Proof

Query $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$, fractional cover $u_1, \ldots, u_\ell$

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

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

Let $\text{Dom} = \text{the domain of all constants in the relations } R_1, \ldots, R_\ell$.
For every $j = 1, \ldots, \ell$, and every tuple $x_j \in \text{Dom}^{|x_j|}$, define:

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

Then: $m_j = |R_j| = \sum_{x_j \in \text{Dom}^{|x_j|}} a_{j,x_j}$, $|Q| = \sum_{x \in \text{Dom}^{|x|}} a_{1,x_1} \cdots a_{\ell,x_\ell}$

Now use Friedgut’s inequality:

$$|Q| = \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_{x_1} a_{1,x_1}^{u_1}) \cdots (\sum_{x_\ell} a_{\ell,x_\ell}^{u_\ell}) = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED
The AGM Inequality – Proof

Query $Q(x) = R_1(x_1), \ldots, R_\ell(x_\ell)$, fractional cover $u_1, \ldots, u_\ell$

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

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

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

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

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

Then: $m_j = |R_j| = \sum_{x_j \in \text{Dom}^{|x_j|}} a_{j,x_j}$, $|Q| = \sum_{x \in \text{Dom}^{|x|}} a_{1,x_1} \cdots a_{\ell,x_\ell}$

Now use Friedgut’s inequality:

$$|Q| = \sum_x a_{1,x_1}^{u_1} \cdots a_{\ell,x_\ell}^{u_\ell} \leq (\sum_{x_1} a_{1,x_1}^{u_1})^{u_1} \cdots (\sum_{x_\ell} a_{\ell,x_\ell}^{u_\ell})^{u_\ell} = m_1^{u_1} \cdots m_\ell^{u_\ell}$$

QED
Proof of Optimality for Generic Join

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