Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog

Hung Q. Ngo
(Stealth Mode)

With Mahmoud Abo Khamis and Dan Suciu @ PODS 2017

HIGHLIGHTS
of Logic, Games and Automata

Table of Contents

Table of Contents

Exciting "Recent" Results on Query Evaluation

conjunctive query
information theory
AGM bound
fractional hypertree width
tree width
proof sequence
disjunctive datalog rule
functional dependency
Shannon flow inequality
entropic bound
factorized database
tree decomposition
worst case optimal join
generalized hypertree width
functional aggregate query
hypertree decomposition
Shannon type inequality
algorithm
degree constraints
polymatroid bound
InsideOut
PANDA algorithm
submodular width

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$
- Attribute $A_i, i \in [n]$, $\quad \mathbf{A}_J := \{A_j \mid j \in J\}, J \subseteq [n]$

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$
- Attribute $A_i, i \in [n]$, $\mathbf{A}_J := \{A_j \mid j \in J\}, J \subseteq [n]$
- Relation $R_F(\mathbf{A}_F)$ for each $F \in \mathcal{E}$, R_F for short
  - e.g. R_{123} for $R_{123}(A_1, A_2, A_3)$
  - e.g. T_{25} for $T_{25}(A_2, A_5)$

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$
- Attribute $A_i, i \in [n]$, $\quad \mathbf{A}_J := \{A_j \mid j \in J\}, J \subseteq [n]$
- Relation $R_F(\mathbf{A}_F)$ for each $F \in \mathcal{E}$, $\qquad$ R_F for short
  - e.g. R_{123} for $R_{123}(A_1, A_2, A_3)$
  - e.g. T_{25} for $T_{25}(A_2, A_5)$

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$
- Attribute $A_i, i \in [n]$, $\mathbf{A}_J := \{A_j \mid j \in J\}, J \subseteq [n]$
- Relation $R_F(\mathbf{A}_F)$ for each $F \in \mathcal{E}$, R_F for short
  - e.g. R_{123} for $R_{123}(A_1, A_2, A_3)$
  - e.g. T_{25} for $T_{25}(A_2, A_5)$

Problem (Boolean Conjunctive Query (BCQ))

$$Q : \quad S() \leftarrow \bigwedge_{F \in \mathcal{E}} R_F \quad ,$$

// can all R_F be satisfied at once?

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$
- Attribute $A_i, i \in [n]$, $\mathbf{A}_J := \{A_j \mid j \in J\}, J \subseteq [n]$
- Relation $R_F(\mathbf{A}_F)$ for each $F \in \mathcal{E}$, R_F for short
  - e.g. R_{123} for $R_{123}(A_1, A_2, A_3)$
  - e.g. T_{25} for $T_{25}(A_2, A_5)$

Problem (Boolean Conjunctive Query (BCQ))

$$Q : S() \leftarrow \bigwedge_{F \in \mathcal{E}} R_F ,$$ *// can all R_F be satisfied at once?*

Question

How do we evaluate Q efficiently?

A Query Evaluation Problem

- Hypergraph $\mathcal{H} = ([n], \mathcal{E}), \mathcal{E} \subseteq 2^{[n]}$
- Attribute $A_i, i \in [n]$, $\mathbf{A}_J := \{A_j \mid j \in J\}, J \subseteq [n]$
- Relation $R_F(\mathbf{A}_F)$ for each $F \in \mathcal{E}$, R_F for short
  - e.g. R_{123} for $R_{123}(A_1, A_2, A_3)$
  - e.g. T_{25} for $T_{25}(A_2, A_5)$

Problem (Boolean Conjunctive Query (BCQ))

$$Q : S() \leftarrow \bigwedge_{F \in \mathcal{E}} R_F ,$$ *// can all R_F be satisfied at once?*

Question

How do we evaluate Q efficiently?

Conjunctive, count, aggregate queries are fine too.

What do we mean by "efficiency"?

Database centric complexity framework

- Assumption 1: query size $\ll$ data size
  - Data complexity
  - Fixed parameter tractability (e.g. parameter = some function of query size)

$$\tilde{O}(\text{something}) = O\left(f(|\text{query}|) \cdot \text{polylog}(|\text{data}|) \cdot \text{something}\right)$$

What do we mean by "efficiency"?

Database centric complexity framework

- Assumption 1: query size $\ll$ data size
  - Data complexity
  - Fixed parameter tractability (e.g. parameter = some function of query size)

$$\tilde{O}(\text{something}) = O\left(f(|\text{query}|) \cdot \text{polylog}(|\text{data}|) \cdot \text{something}\right)$$

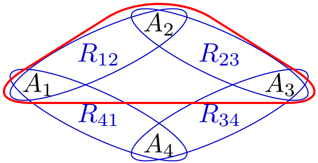
- Assumption 2: known constraints on input relations
  - Cardinalities of materialized relations, *or upper bounds*
    - cardinality constraints (CC)
  - Functional dependencies, *the more the merrier*
    - FD constraints (FDC)
  - Degree bounds, *the more the merrier*
    - Degree constraints (DC)

Example

Q : $S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

$$\wedge A_1 + A_2 = A_3.$$

$$\mathcal{H} = ([4], \{12, 23, 34, 14, 123\})$$



A_1	A_2
2	1
4	1
4	2
5	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5
e	6
e	7

A_4	A_1
3	2
4	4
4	5

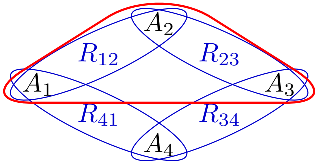
- Cardinalities: $|R_{12}| = 4$, $|R_{23}| = 3$, ...

Example

Q : $S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

$$\wedge A_1 + A_2 = A_3.$$

$$\mathcal{H} = ([4], \{12, 23, 34, 14, 123\})$$



A_1	A_2
2	1
4	1
4	2
5	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5
e	6
e	7

A_4	A_1
3	2
4	4
4	5

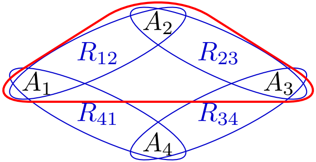
- Cardinalities: $|R_{12}| = 4$, $|R_{23}| = 3, \ldots$
- FD: $\{A_1, A_2\} \to A_3$, $\{A_3, A_2\} \to A_1, \ldots$

Example

Q : $S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

$$\wedge A_1 + A_2 = A_3.$$

$$\mathcal{H} = ([4], \{12, 23, 34, 14, 123\})$$



A_1	A_2
2	1
4	1
4	2
5	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5
e	6
e	7

A_4	A_1
3	2
4	4
4	5

- Cardinalities: $|R_{12}| = 4, |R_{23}| = 3, \ldots$
- FD: $\{A_1, A_2\} \to A_3, \{A_3, A_2\} \to A_1, \ldots$
- Degree Bounds: $\deg_{34}(A_4|A_3 = x) \stackrel{\text{def}}{=} |\sigma_{A_3=x}(R_{34})| \leq 2, \forall x, \ldots$

Degree- generalize cardinality- and FD-constraints

DC ⊇ CC ∪ FDC

- Degree Constraints (DC):

$$\deg_F(\mathbf{A}_Y | \mathbf{A}_X) \le N_{Y|X}, \qquad X \subset Y \subseteq F \in \mathcal{E}$$

Degree- generalize cardinality- and FD-constraints

$$\mathsf{DC} \supseteq \mathsf{CC} \cup \mathsf{FDC}$$

- Degree Constraints (DC):

$$\deg_F(\mathbf{A}_Y | \mathbf{A}_X) \leq N_{Y|X}, \qquad X \subset Y \subseteq F \in \mathcal{E}$$

  - Cardinality Constraints (CC):

$$|R_F| \leq N \Leftrightarrow \deg_F(\mathbf{A}_F | \mathbf{A}_\emptyset) \leq N_{F|\emptyset} \stackrel{\text{def}}{=} N.$$

Degree- generalize cardinality- and FD-constraints

DC ⊇ CC ∪ FDC

- Degree Constraints (DC):

$$\deg_F(\mathbf{A}_Y | \mathbf{A}_X) \leq N_{Y|X}, \qquad X \subset Y \subseteq F \in \mathcal{E}$$

  - Cardinality Constraints (CC):

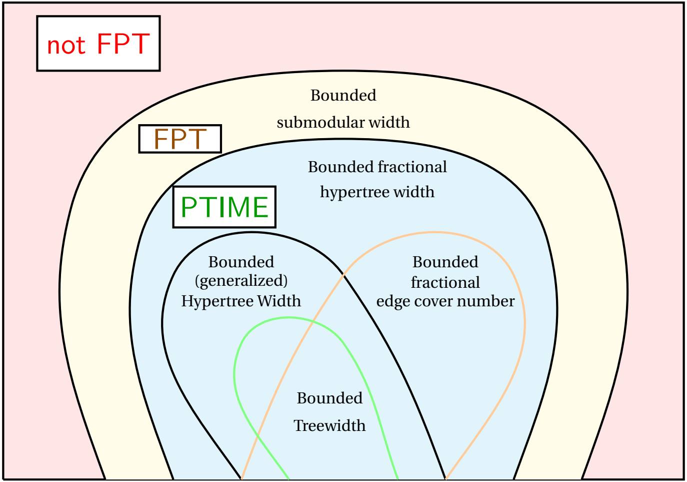
$$|R_F| \leq N \Leftrightarrow \deg_F(\mathbf{A}_F | \mathbf{A}_\emptyset) \leq N_{F|\emptyset} \stackrel{\text{def}}{=} N.$$

  - Functional Dependencies (FDC):

$$\mathbf{A}_X \to \mathbf{A}_Y \Leftrightarrow \deg_F(\mathbf{A}_Y | \mathbf{A}_X) \leq N_{Y|X} \stackrel{\text{def}}{=} 1.$$

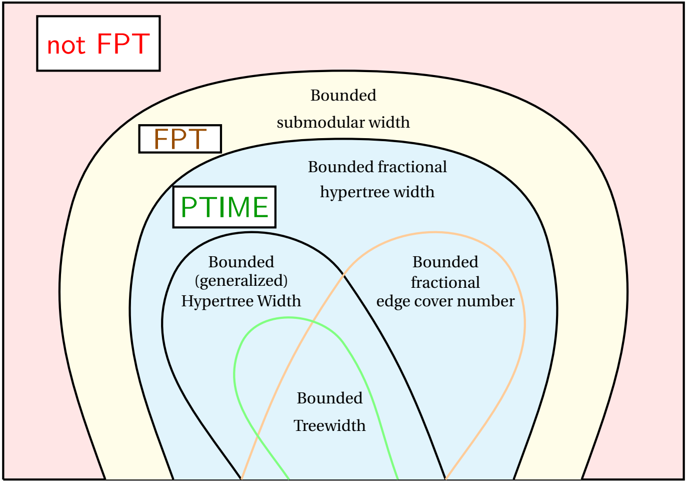
Islands of tractability (redrawn from D. Marx's slides)

Prior results with cardinality constraints



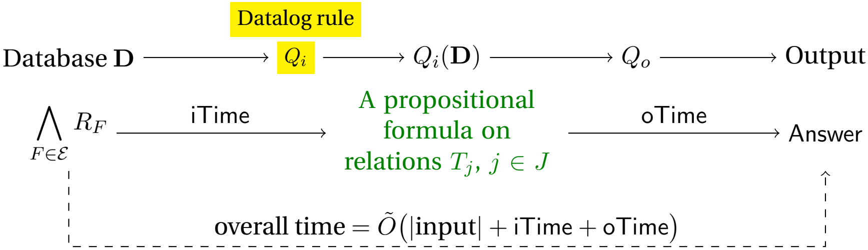
Islands of tractability (redrawn from D. Marx's slides)

Prior results with cardinality constraints

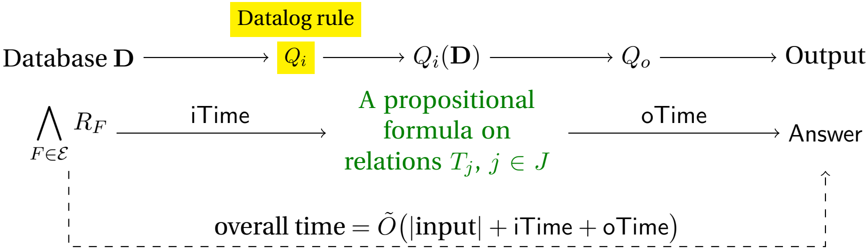


We want the same map with degree constraints.

A Meta Algorithm (i.e. Meta Query Plan)

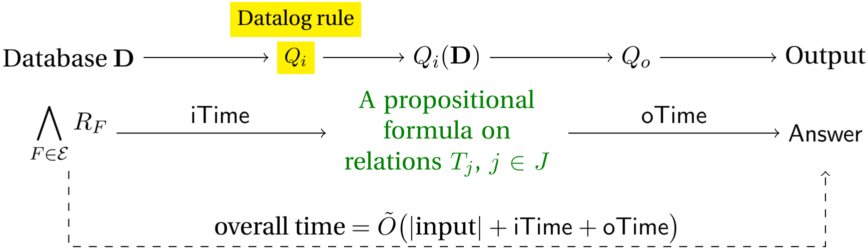


A Meta Algorithm (i.e. Meta Query Plan)



(a) $\text{oTime} = |Q_i(\mathbf{D})| + |\text{answer}|$

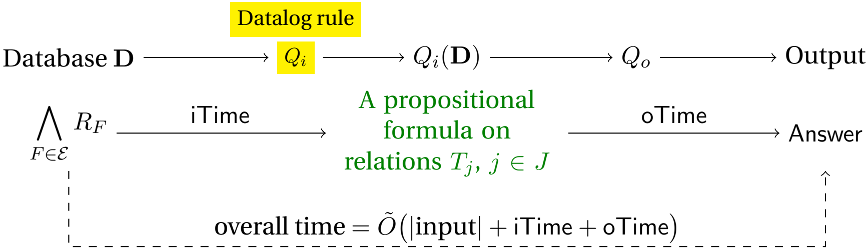
A Meta Algorithm (i.e. Meta Query Plan)



(a) $\text{oTime} = |Q_i(\mathbf{D})| + |\text{answer}|$

- i.e. Q_o evaluatable in linear time

A Meta Algorithm (i.e. Meta Query Plan)

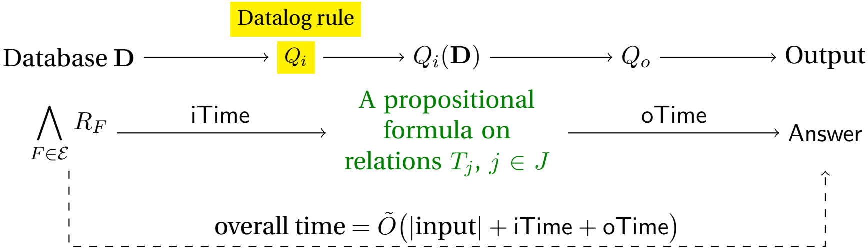


(a) $\mathsf{oTime} = |Q_i(\mathbf{D})| + |\mathsf{answer}|$

- i.e. Q_o evaluatable in linear time

(b) $\mathsf{iTime} = \max_{\mathbf{D} \models \mathsf{DC}} |Q_i(\mathbf{D})|$

A Meta Algorithm (i.e. Meta Query Plan)



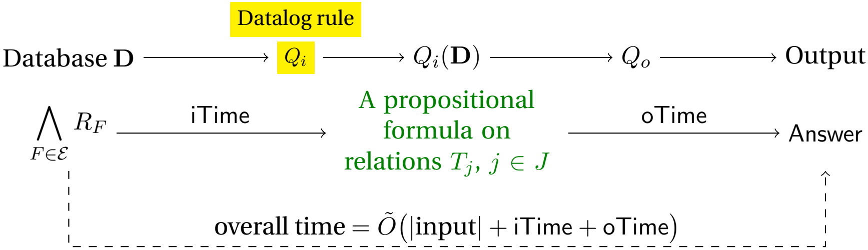
(a) $\text{oTime} = |Q_i(\mathbf{D})| + |\text{answer}|$

- i.e. Q_o evaluatable in linear time

(b) $\text{iTime} = \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})|$

- i.e. Q_i evaluatable within its worst-case output size

A Meta Algorithm (i.e. Meta Query Plan)



(a) $\text{oTime} = |Q_i(\mathbf{D})| + |\text{answer}|$
- i.e. Q_o evaluatable in linear time

(b) $\text{iTime} = \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})|$
- i.e. Q_i evaluatable within its worst-case output size

- Design Q_i s.t. (a) holds and (b) as small as possible

Option 1: Full Conjunctive Query

Full conjunctive query

Database $\mathbf{D}$ $\longrightarrow$ Q_i $\longrightarrow$ $Q_i(\mathbf{D})$ $\longrightarrow$ Q_o $\longrightarrow$ Output

$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} T_{[n]} \xrightarrow{\text{oTime}} \text{Answer}$

Option 1: Full Conjunctive Query

Full conjunctive query

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} T_{[n]} \xrightarrow{\text{oTime}} \text{Answer}$$

- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$ trivially true

Option 1: Full Conjunctive Query

Full conjunctive query

Database $\mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow$ Output

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} T_{[n]} \xrightarrow{\text{oTime}} \text{Answer}$$

- $\text{oTime} = |Q_i(\mathbf{D})| + |\text{answer}|$ trivially true
- $\text{iTime} = \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})|$ worst-case optimal algorithm
  - Known if all DC are cardinality constraints
    - NPRR [Ngo, Porat, Ré, Rudra PODS'12]
    - Leapfrog-Triejoin [Veldhuizen ICDT'14]
    - Generic Join [Ngo, Ré, Rudra SIGMOD Records 2013]
  - Unknown for general DC until our work

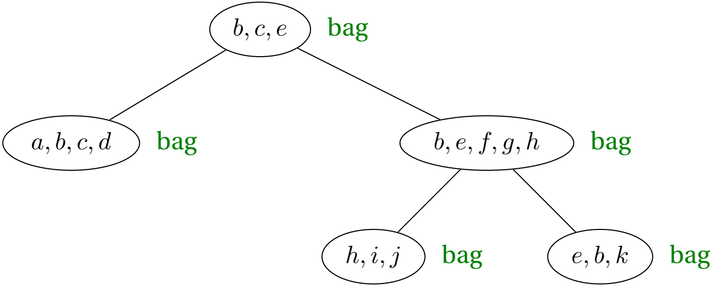
Detour: Tree Decompositions, Informally

Detour: Tree Decompositions, Informally

$$S() \leftarrow R(a,b,d) \wedge c < d \wedge T(c,b,d) \wedge U(b,e) \wedge V(c,e)$$
$$\wedge\, b + e = f \wedge W(b,e,g) \wedge \;\wedge X(i,j,h) \wedge e - b = k.$$

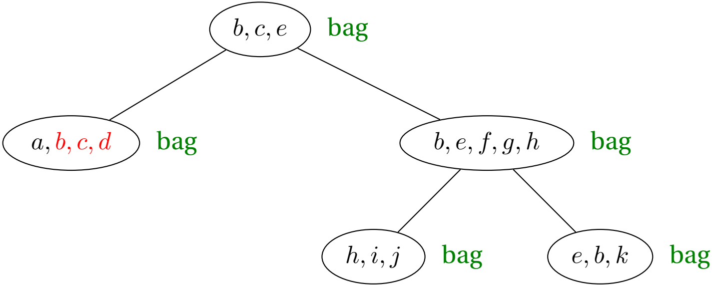
Detour: Tree Decompositions, Informally

$$S() \leftarrow R(a,b,d) \wedge c < d \wedge T(c,b,d) \wedge U(b,e) \wedge V(c,e)$$
$$\wedge\, b + e = f \wedge W(b,e,g) \wedge\ \wedge X(i,j,h) \wedge e - b = k.$$



Detour: Tree Decompositions, Informally

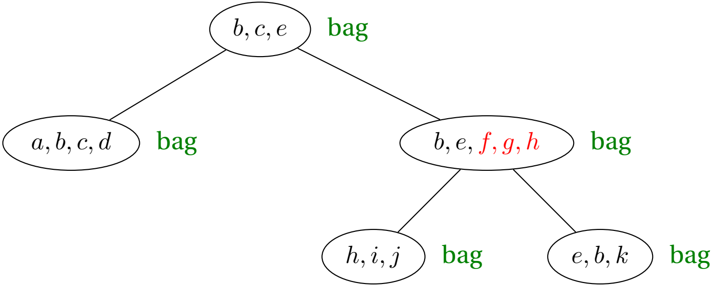
$$S() \leftarrow R(a,b,d) \wedge c < d \wedge T(c,b,d) \wedge U(b,e) \wedge V(c,e)$$
$$\wedge\, b+e=f \wedge W(b,e,g) \wedge \;\wedge X(i,j,h) \wedge e-b=k.$$



- Every relation is covered by some bag

Detour: Tree Decompositions, Informally

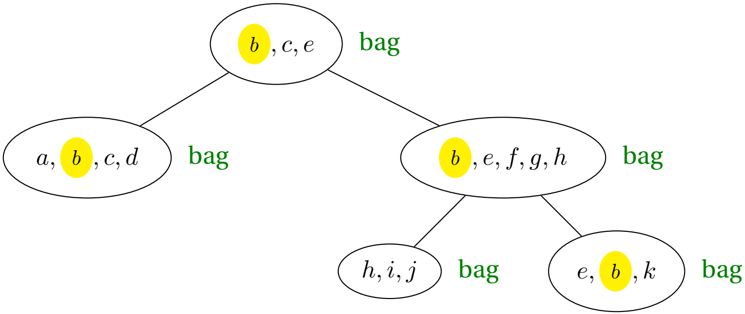
$$S() \leftarrow R(a,b,d) \wedge c < d \wedge T(c,b,d) \wedge U(b,e) \wedge V(c,e)$$
$$\wedge\, b+e=f \wedge W(b,e,g) \wedge g/f=h \wedge X(i,j,h) \wedge e-b=k.$$



- Every relation is covered by some bag

Detour: Tree Decompositions, Informally

$$S() \leftarrow R(a,b,d) \wedge c < d \wedge T(c,b,d) \wedge U(b,e) \wedge V(c,e)$$
$$\wedge\, b + e = f \wedge W(b,e,g) \wedge \;\wedge X(i,j,h) \wedge e - b = k.$$



- Every relation is covered by some bag
- Bags conntaining a given variable are connected

Detour: Tree Decompositions, Formally

- Hypergraph $\mathcal{H} = ([n], \mathcal{E})$
- A Tree Decomposition of $\mathcal{H}$ is a pair $(\mathcal{T}, \chi)$ where
  - $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ is a tree

See [Gottlob et al 2016], Gems of PODS.

Detour: Tree Decompositions, Formally

- Hypergraph $\mathcal{H} = ([n], \mathcal{E})$
- A Tree Decomposition of $\mathcal{H}$ is a pair $(\mathcal{T}, \chi)$ where
  - $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ is a tree
  - $\chi : V(\mathcal{T}) \to 2^{[n]}$ assigns a bag $\chi(v)$ to each tree-node v
    - Every hyperedge $F \in \mathcal{E}$ is covered by some bag $(F \subseteq \chi(v))$
    - Bags containing $\forall i \in [n]$ forms a subtree

See [Gottlob et al 2016], Gems of PODS.

Option 2: A Single Tree Decomposition

Fix $(\mathcal{T}, \chi)$

Multiple Conjunctive Rules

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\text{oTime}} \text{Answer}$$

Option 2: A Single Tree Decomposition

Fix $(\mathcal{T}, \chi)$

Multiple Conjunctive Rules

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\text{oTime}} \text{Answer}$$

- $Q_i(\mathbf{D})$ is Olteanu's factorized database (FDB)

Option 2: A Single Tree Decomposition

Fix $(\mathcal{T}, \chi)$

Multiple Conjunctive Rules

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\text{oTime}} \text{Answer}$$

- $Q_i(\mathbf{D})$ is Olteanu's factorized database (FDB)
- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$
  - Yannakakis for join, FDB/InsideOut for aggregates

Option 2: A Single Tree Decomposition

Fix $(\mathcal{T}, \chi)$

Multiple Conjunctive Rules

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\text{oTime}} \text{Answer}$$

- $Q_i(\mathbf{D})$ is Olteanu's factorized database (FDB)
- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$
  - Yannakakis for join, FDB/InsideOut for aggregates
- iTime $= \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})| \leq \max_{\mathbf{D} \models \text{DC}} \max_{v \in V(\mathcal{T})} |P_v(\mathbf{D})|$
  - $P_v : T_{\chi(v)} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F \qquad v \in V(\mathcal{T})$

Option 2: A Single Tree Decomposition

Fix $(\mathcal{T}, \chi)$

Multiple Conjunctive Rules

Database $\mathbf{D}$ $\longrightarrow$ Q_i $\longrightarrow$ $Q_i(\mathbf{D})$ $\longrightarrow$ Q_o $\longrightarrow$ Output

$\bigwedge_{F \in \mathcal{E}} R_F$ $\xrightarrow{\text{iTime}}$ $\bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)}$ $\xrightarrow{\text{oTime}}$ Answer

- $Q_i(\mathbf{D})$ is Olteanu's factorized database (FDB)
- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$
  - Yannakakis for join, FDB/InsideOut for aggregates
- iTime $= \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})| \leq \max_{\mathbf{D} \models \text{DC}} \max_{v \in V(\mathcal{T})} |P_v(\mathbf{D})|$
  - $P_v \; : \; T_{\chi(v)} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F \qquad v \in V(\mathcal{T})$

Option 2: A Single Tree Decomposition

Fix $(\mathcal{T}, \chi)$

Multiple Conjunctive Rules

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\quad \mathsf{iTime} \quad} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\quad \mathsf{oTime} \quad} \mathsf{Answer}$$

- $Q_i(\mathbf{D})$ is Olteanu's factorized database (FDB)
- $\mathsf{oTime} = |Q_i(\mathbf{D})| + |\mathsf{answer}|$
  - Yannakakis for join, FDB/InsideOut for aggregates
- $\mathsf{iTime} = \max_{\mathbf{D} \models \mathsf{DC}} |Q_i(\mathbf{D})| \leq \max_{\mathbf{D} \models \mathsf{DC}} \max_{v \in V(\mathcal{T})} |P_v(\mathbf{D})|$
  - $P_v \ : \ T_{\chi(v)} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F \qquad v \in V(\mathcal{T})$

$$\min_{(\mathcal{T},\chi)} \max_{\mathbf{D} \models \mathsf{CC}} \max_{v \in V(\mathcal{T})} |P_v(\mathbf{D})| \leq N^{\mathsf{fhtw}(\mathcal{H})} \leq N^{\mathsf{ghtw}(\mathcal{H})} \leq N^{\mathsf{tw}(\mathcal{H})+1}$$

Option 3: Multiple Tree Decompositions

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\mathsf{iTime}} \bigvee_{(\mathcal{T},\chi)} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\mathsf{oTime}} \mathsf{Answer}$$

- $\bigvee$ ranges over non-redundant TDs $(\mathcal{T}, \chi)$

Option 3: Multiple Tree Decompositions

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\quad \text{iTime} \quad} \bigvee_{(\mathcal{T},\chi)} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\quad \text{oTime} \quad} \text{Answer}$$

- $\bigvee$ ranges over non-redundant TDs $(\mathcal{T}, \chi)$
- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$
  - Union of Yannakakis on all TDs

Option 3: Multiple Tree Decompositions

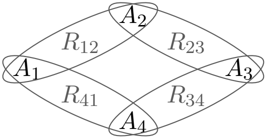
How to evaluate this?

Database $\mathbf{D}$ $\longrightarrow$ Q_i $\longrightarrow$ $Q_i(\mathbf{D})$ $\longrightarrow$ Q_o $\longrightarrow$ Output

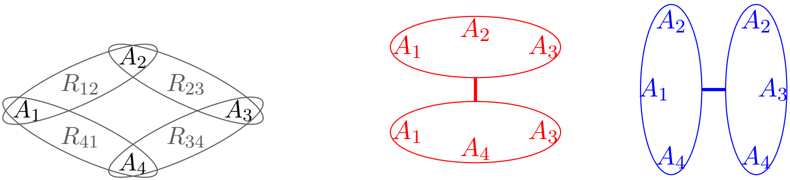
$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\quad \mathsf{iTime} \quad} \bigvee_{(\mathcal{T},\chi)} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\quad \mathsf{oTime} \quad} \mathsf{Answer}$$

- $\bigvee$ ranges over non-redundant TDs $(\mathcal{T}, \chi)$
- $\mathsf{oTime} = |Q_i(\mathbf{D})| + |\mathsf{answer}|$
  - Union of Yannakakis on all TDs
- $\mathsf{iTime} = \max_{\mathbf{D} \models \mathsf{DC}} |Q_i(\mathbf{D})| \leq$?

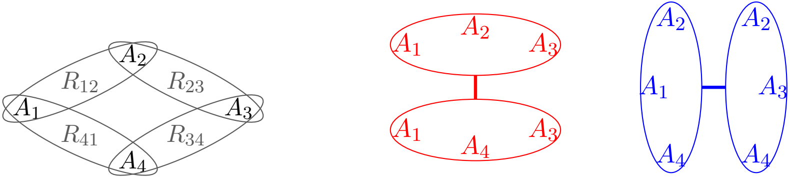
Example: $Q \; : \; S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$



Example: $Q \ : \ S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

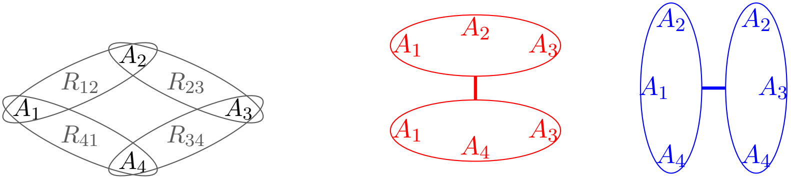


Example: $Q \; : \; S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$



$$(T_{123} \wedge T_{134}) \vee (T_{124} \wedge T_{234}) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Example: $Q \ : \ S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$



$$(T_{123} \wedge T_{134}) \vee (T_{124} \wedge T_{234}) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

By distributivity, rewrite an equivalent the head:

$$(T_{123} \vee T_{124}) \wedge (T_{123} \vee T_{234}) \wedge (T_{134} \vee T_{124}) \wedge (T_{134} \vee T_{234})$$
$$\leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

(Each “clause” has one bag per TD)

Option 3: Multiple Tree Decompositions

Multiple Disjunctive Datalog Rules!

Database $\mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow$ Output

$$\bigwedge_{F\in\mathcal{E}} R_F \xrightarrow{\text{iTime}} \bigwedge_{\mathcal{B}} \bigvee_{B\in\mathcal{B}} T_B \xrightarrow{\text{oTime}} \text{Answer}$$

- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$
  - Union of Yannakakis on all TDs
- iTime $= \max_{\mathbf{D}\models\text{DC}} |Q_i(\mathbf{D})| \leq \max_{\mathbf{D}\models\text{DC}} \max_{\mathcal{B}} |P_{\mathcal{B}}(\mathbf{D})|$
  - $P_{\mathcal{B}} : \bigvee_{B\in\mathcal{B}} T_B \leftarrow \bigwedge_{F\in\mathcal{E}} R_F$

disjunctive datalog rule

Option 3: Multiple Tree Decompositions

Multiple Disjunctive Datalog Rules!

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\text{iTime}} \bigwedge_{\mathcal{B}} \bigvee_{B \in \mathcal{B}} T_B \xrightarrow{\text{oTime}} \text{Answer}$$

- oTime $= |Q_i(\mathbf{D})| + |\text{answer}|$
  - Union of Yannakakis on all TDs
- iTime $= \max_{\mathbf{D} \models \mathsf{DC}} |Q_i(\mathbf{D})| \leq \max_{\mathbf{D} \models \mathsf{DC}} \max_{\mathcal{B}} |P_{\mathcal{B}}(\mathbf{D})|$
  - $P_{\mathcal{B}} : \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$

disjunctive datalog rule

$$\max_{\mathbf{D} \models \mathsf{CC}} \max_{\mathcal{B}} |P_{\mathcal{B}}(\mathbf{D})| \leq N^{\mathsf{subw}(\mathcal{H})} \leq N^{\mathsf{fhtw}(\mathcal{H})}$$

subw = submodular width (Daniel Marx, JACM'2013)

Example: $Q\ : S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 1:

$$Q_i : \qquad T_{1234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Example: $Q\ : S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 1:

$$Q_i : \qquad T_{1234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Option 2:

$$\text{either } Q_i : \qquad T_{123} \wedge T_{134} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

$$\text{or } Q_i : \qquad T_{124} \wedge T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Example: $Q\ : S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 1:

$$Q_i : \qquad T_{1234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Option 2:

$$\text{either } Q_i : \qquad T_{123} \wedge T_{134} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

$$\text{or } Q_i : \qquad T_{124} \wedge T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Option 3:

$$Q_i : \qquad (T_{123} \wedge T_{134}) \vee (T_{124} \wedge T_{234}) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Equivalent to:

$$P_{123,124} : \qquad T_{123} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

$$P_{123,234} : \qquad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

$$P_{134,124} : \qquad T_{134} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

$$P_{134,234} : \qquad T_{134} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Example: $Q\ : S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 1: $\mathsf{iTime} = \max_{\mathbf{D} \models \mathsf{DC}} |Q_i(\mathbf{D})| = N^2$

$$Q_i : \quad T_{1234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Option 2:

either $Q_i : \quad T_{123} \wedge T_{134} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

or $Q_i : \quad T_{124} \wedge T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 3:

$$Q_i : \quad (T_{123} \wedge T_{134}) \vee (T_{124} \wedge T_{234}) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Equivalent to:

$$P_{123,124} : \quad T_{123} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{123,234} : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{134,124} : \quad T_{134} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{134,234} : \quad T_{134} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Example: $Q \; : S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 1: $\textsf{iTime} = \max_{\mathbf{D} \models \textsf{DC}} |Q_i(\mathbf{D})| = N^2$

$$Q_i : \quad T_{1234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Option 2: $\textsf{iTime} = \max_{\mathbf{D} \models \textsf{DC}} |Q_i(\mathbf{D})| = N^2$

either $Q_i : \quad T_{123} \wedge T_{134} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

or $Q_i : \quad T_{124} \wedge T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 3:

$$Q_i : \quad (T_{123} \wedge T_{134}) \vee (T_{124} \wedge T_{234}) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Equivalent to:

$$P_{123,124} : \quad T_{123} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{123,234} : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{134,124} : \quad T_{134} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{134,234} : \quad T_{134} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Example: $Q\ : S() \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 1: $\text{iTime} = \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})| = N^2$

$$Q_i : \quad T_{1234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Option 2: $\text{iTime} = \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})| = N^2$

either $Q_i : \quad T_{123} \wedge T_{134} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

or $Q_i : \quad T_{124} \wedge T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$

Option 3: $\text{iTime} = \max_{\mathbf{D} \models \text{DC}} |Q_i(\mathbf{D})| = N^{3/2}$

$$Q_i : \quad (T_{123} \wedge T_{134}) \vee (T_{124} \wedge T_{234}) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Equivalent to:

$$P_{123,124} : \quad T_{123} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{123,234} : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{134,124} : \quad T_{134} \vee T_{124} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$
$$P_{134,234} : \quad T_{134} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$$

Roadmap

Given degree constraints DC, and a disjunctive datalog rule

$$P \; : \; \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

Roadmap

Given degree constraints DC, and a disjunctive datalog rule

$$P \ : \ \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

Question (Worst-case Output Size Bound)

Find a good upper-bound for $\max_{\mathbf{D} \models DC} |P(\mathbf{D})|$

Roadmap

Given degree constraints DC, and a disjunctive datalog rule

$$P \; : \; \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

Question (Worst-case Output Size Bound)

Find a good upper-bound for $\max_{\mathbf{D} \models DC} |P(\mathbf{D})|$

Question (Algorithm)

Design an algorithm evaluating P *within the bound.*

Roadmap

Given degree constraints DC, and a disjunctive datalog rule

$$P \; : \; \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

Question (Worst-case Output Size Bound)

Find a good upper-bound for $\max_{\mathbf{D} \models DC} |P(\mathbf{D})|$

Question (Algorithm)

Design an algorithm evaluating P *within the bound.*

Question (Gathering fruits)

Plug bound/algorithm into Meta Algorithm, what do we get?

Table of Contents

High-level View of the Bound

Given degree constraints DC, a disjunctive datalog rule

$$P \ : \ \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

We shall prove bounds of the form

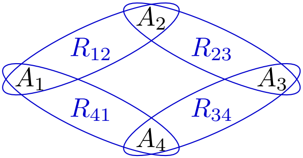
$$\max_{\mathbf{D} \models \mathsf{DC}} \log |P(\mathbf{D})| \leq \text{some function of } h$$

s.t. h is (approximately) entropic

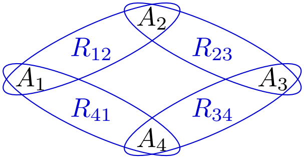
and h satisfies degree constraints

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$



An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

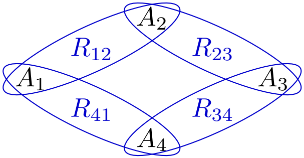
A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

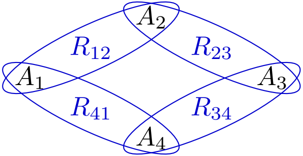
A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

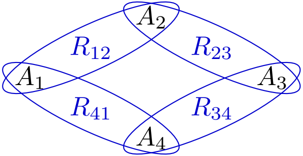
A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

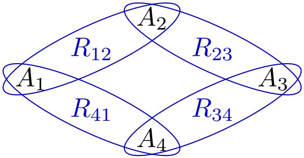
A_1	A_2	
a	1	1/4
b	1	2/4
b	2	1/4

A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

A_3	A_4	
c	3	2/4
d	4	2/4
d	5	0

A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

$H_{\mathsf{unif}}(A_1 A_2 A_3 A_4) = \log |Q|$

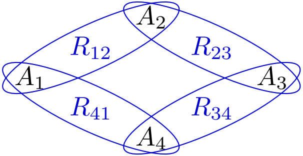
A_1	A_2	
a	1	1/4
b	1	2/4
b	2	1/4

A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

A_3	A_4	
c	3	2/4
d	4	2/4
d	5	0

A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

$H_{\mathsf{unif}}(A_1 A_2 A_3 A_4) = \log |Q|$

A_1	A_2	
a	1	1/4
b	1	2/4
b	2	1/4

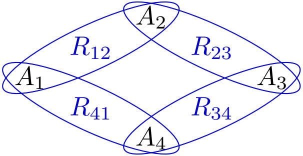
A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

A_3	A_4	
c	3	2/4
d	4	2/4
d	5	0

A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

$H_{\mathsf{unif}}(A_1 A_2) \leq \log |R_{12}|$, $H_{\mathsf{unif}}(A_2 A_3) \leq \log |R_{23}|$, $H_{\mathsf{unif}}(A_3 A_4) \leq \log |R_{34}|$, $\ldots$

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

$H_{\mathsf{unif}}(A_1 A_2 A_3 A_4) = \log |Q|$

A_1	A_2	
a	1	1/4
b	1	2/4
b	2	1/4

A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

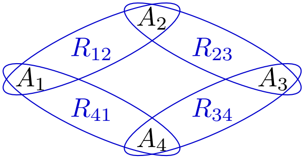
A_3	A_4	
c	3	2/4
d	4	2/4
d	5	0

A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

$H_{\mathsf{unif}}(A_1 A_2) \leq \log |R_{12}|, \quad H_{\mathsf{unif}}(A_2 A_3) \leq \log |R_{23}|, \quad H_{\mathsf{unif}}(A_3 A_4) \leq \log |R_{34}|, \quad \ldots$

$H_{\mathsf{unif}}(A_2 | A_1 = \text{‘a’}) \leq \log \left|\sigma_{A_1 = \text{‘a’}} R_{12}\right|, \quad H_{\mathsf{unif}}(A_2 | A_1 = \text{‘b’}) \leq \log \left|\sigma_{A_1 = \text{‘b’}} R_{12}\right|, \ldots$

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12



$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

$H_{\mathsf{unif}}(A_1 A_2 A_3 A_4) = \log |Q|$

A_1	A_2	
a	1	1/4
b	1	2/4
b	2	1/4

A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

A_3	A_4	
c	3	2/4
d	4	2/4
d	5	0

A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

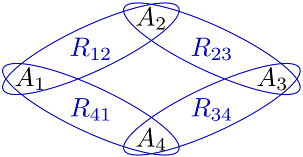
$H_{\mathsf{unif}}(A_1 A_2) \le \log |R_{12}|, \quad H_{\mathsf{unif}}(A_2 A_3) \le \log |R_{23}|, \quad H_{\mathsf{unif}}(A_3 A_4) \le \log |R_{34}|, \quad \ldots$

$H_{\mathsf{unif}}(A_2 | A_1 = \text{'a'}) \le \log \left|\sigma_{A_1=\text{'a'}} R_{12}\right|, \quad H_{\mathsf{unif}}(A_2 | A_1 = \text{'b'}) \le \log \left|\sigma_{A_1=\text{'b'}} R_{12}\right|, \ldots$

$H_{\mathsf{unif}}(A_2 | A_1) \le \log \max_x \left|\sigma_{A_1=x} R_{12}\right|$

An Idea From Gottlob-Lee-Valiant-Valiant, JACM'12

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$



A_1	A_2	A_3	A_4	
a	1	d	4	1/4
b	1	c	3	1/4
b	1	d	4	1/4
b	2	c	3	1/4

$H_{\mathsf{unif}}(A_1 A_2 A_3 A_4) = \log |Q|$

A_1	A_2	
a	1	1/4
b	1	2/4
b	2	1/4

A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

A_3	A_4	
c	3	2/4
d	4	2/4
d	5	0

A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

$H_{\mathsf{unif}}(A_1 A_2) \leq \log |R_{12}|, \quad H_{\mathsf{unif}}(A_2 A_3) \leq \log |R_{23}|, \quad H_{\mathsf{unif}}(A_3 A_4) \leq \log |R_{34}|, \quad \ldots$

$H_{\mathsf{unif}}(A_2 | A_1 = \text{'a'}) \leq \log |\sigma_{A_1=\text{'a'}} R_{12}|, \quad H_{\mathsf{unif}}(A_2 | A_1 = \text{'b'}) \leq \log |\sigma_{A_1=\text{'b'}} R_{12}|, \ldots$

$$H_{\mathsf{unif}}(A_2 | A_1) \leq \log \underbrace{\max_x |\sigma_{A_1=x} R_{12}|}_{\deg_{R_{12}}(A_2|A_1)}$$

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$
- subject to (whatever H_{unif} satisfies):

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$
- subject to (whatever H_{unif} satisfies):
  - h is Entropic

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$
- subject to (whatever H_{unif} satisfies):
  - h is Entropic
    - There is some distribution on $\mathbf{A}_{[n]}$ such that $h(X)$ is the marginal entropy on $\mathbf{A}_X$, for all X

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$
- subject to (whatever H_{unif} satisfies):
  - h is Entropic
    - There is some distribution on $\mathbf{A}_{[n]}$ such that $h(X)$ is the marginal entropy on $\mathbf{A}_X$, for all X
  - h satisfies DC

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$
- subject to (whatever H_{unif} satisfies):
  - h is Entropic
    - There is some distribution on $\mathbf{A}_{[n]}$ such that $h(X)$ is the marginal entropy on $\mathbf{A}_X$, for all X
  - h satisfies DC
    - $h(Y|X) \stackrel{\text{def}}{=} h(Y) - h(X) \leq \log N_{Y|X}, \qquad X \subset Y \subseteq F \in \mathcal{E}$

Entropic Bound for Full Conjunctive Queries

- $Q : T_{[n]} \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$, and degree constraints DC
- $\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| \leq \sup h([n])$
- subject to (whatever H_{unif} satisfies):
  - h is Entropic
    - There is some distribution on $\mathbf{A}_{[n]}$ such that $h(X)$ is the marginal entropy on $\mathbf{A}_X$, for all X
  - h satisfies DC
    - $h(Y|X) \stackrel{\text{def}}{=} h(Y) - h(X) \leq \log N_{Y|X}, \qquad X \subset Y \subseteq F \in \mathcal{E}$
- Good Bound, but not computable!

Hierarchy of Set Functions

$h : 2^{[n]} \to \mathbb{R}_+$, non-negative, monotone, $h(\emptyset) = 0$

$h(X) \leq h(Y)$ if $X \subseteq Y$

$\mathsf{SA}_n := \{h \mid h \text{ is sub-additive}\}$

$$h(X \cup Y) \leq h(X) + h(Y)$$

$\Gamma_n := \{h \mid h \text{ is submodular}\}$

$$h(X \cup Y) + h(X \cap Y) \leq h(X) + h(Y)$$

$\overline{\Gamma}^*_n$: topological closure of Γ^*_n

$\Gamma^*_n = \{h : h \text{ is entropic}\}$

M_n : Modular

$$h(X) = \sum_{x \in x} h(x)$$

Bounds for Full Conjunctive Query

- HDC $\stackrel{\text{def}}{=} \{h \quad | \quad h(Y|X) \leq \log N_{Y|X}, \quad \forall (X, Y, N_{Y|X})\}$

Bounds for Full Conjunctive Query

- $\mathsf{HDC} \stackrel{\text{def}}{=} \{h \quad | \quad h(Y|X) \leq \log N_{Y|X}, \quad \forall (X, Y, N_{Y|X})\}$
- Then,

$$\begin{aligned}
\max_{\mathbf{D} \models \mathsf{DC}} \log |Q(\mathbf{D})| &\leq \max_{h \in \overline{\Gamma}^*_n \cap \mathsf{HDC}} h([n]) && \text{entropic bound} \\
&\leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} h([n]) && \text{polymatroid bound} \\
&\leq \max_{h \in \mathsf{SA}_n \cap \mathsf{HDC}} h([n]) && \text{sub-additive bound.}
\end{aligned}$$

Size Bounds for Full Conjunctive Queries

Bound	Entropic Bound	Polymatroid Bound
Definition	$\log \lvert Q \rvert \leq \max_{h \in \overline{\Gamma}^*_n \cap \mathsf{HDC}} h([n])$	$\log \lvert Q \rvert \leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} h([n])$

Size Bounds for Full Conjunctive Queries

Bound	Entropic Bound	Polymatroid Bound
Definition	$\log \lvert Q\rvert \leq \max_{h \in \overline{\Gamma}^*_n \cap \mathsf{HDC}} h([n])$	$\log \lvert Q\rvert \leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} h([n])$
CC only	AGM bound (Tight) [Atserias et al. FOCS'08]	AGM bound (Tight) [Atserias et al. FOCS'08]

Size Bounds for Full Conjunctive Queries

Bound	Entropic Bound	Polymatroid Bound
Definition	$\log\|Q\| \leq \max_{h \in \overline{\Gamma}^*_n \cap \mathsf{HDC}} h([n])$	$\log\|Q\| \leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} h([n])$
CC only	AGM bound (Tight) [Atserias et al. FOCS'08]	AGM bound (Tight) [Atserias et al. FOCS'08]
CC + FD only	Entropic Bound for FD **[Gottlob et al. JACM'12]** (Tight [Gogacz et al. ICDT'17])	Polymatroid Bound for FD **[Gottlob et al. JACM'12]** (Not tight [our work])

Size Bounds for Full Conjunctive Queries

Bound	Entropic Bound	Polymatroid Bound
Definition	$\log \lvert Q \rvert \leq \max_{h \in \overline{\Gamma}^*_n \cap \mathsf{HDC}} h([n])$	$\log \lvert Q \rvert \leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} h([n])$
CC only	AGM bound (Tight) [Atserias et al. FOCS'08]	AGM bound (Tight) [Atserias et al. FOCS'08]
CC + FD only	Entropic Bound for FD **[Gottlob et al. JACM'12]** (Tight [Gogacz et al. ICDT'17])	Polymatroid Bound for FD **[Gottlob et al. JACM'12]** (Not tight [our work])
DC	Entropic Bound for DC (Tight [our work])	Polymatroid Bound for DC (Not tight [our work])

Disjunctive Datalog: Size Bounds

$$P: \bigvee_{B\in\mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F\in\mathcal{E}} R_F(\mathbf{A}_F) \qquad |P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T}\models P} \max_{B\in\mathcal{B}} |T_B|$$

Theorem (our work)

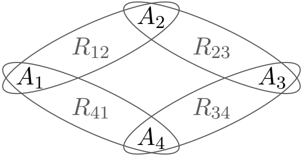
$$\max_{\mathbf{D}\models DC} \log|P(\mathbf{D})| \leq \underbrace{\max_{h\in\overline{\Gamma}^*_n\cap HDC} \min_{B\in\mathcal{B}} h(B)}_{\textit{Entropic bound}} \qquad \textit{Tight}$$

$$\leq \underbrace{\max_{h\in\Gamma_n\cap HDC} \min_{B\in\mathcal{B}} h(B)}_{\textit{Polymatroid bound}} \qquad \textit{Not Tight}$$

Imply all known bounds for (Full) Conjunctive Queries!

Earlier Example

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



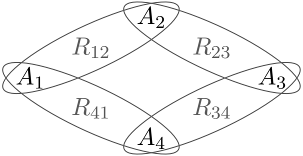
$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$

$$\mathsf{CC}: \quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$$

$$P_{123,234} \quad : \quad T_{123} \vee T_{234} \quad \leftarrow \quad R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

Earlier Example

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$

$$\mathsf{CC}: \quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$$

$$P_{123,234} \quad : \quad T_{123} \vee T_{234} \quad \leftarrow \quad R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

$$\begin{aligned}
\max_{\mathbf{D} \models \mathsf{CC}} \log |P_{123,234}(\mathbf{D})| &\leq \max_{h \in \Gamma_n \cap \mathsf{CC}} \min\{h(A_1A_2A_3), h(A_2A_3A_4)\} \\
&\leq \max_{h \in \Gamma_n \cap \mathsf{CC}} \frac{1}{2}[h(A_1A_2A_3), h(A_2A_3A_4)] \\
&\leq \max_{h \in \Gamma_n \cap \mathsf{CC}} \frac{1}{2}[h(A_1A_2) + h(A_2A_3) + h(A_3A_4)] \\
&\leq \frac{3}{2} \log N.
\end{aligned}$$

Table of Contents

Roadmap

Given degree constraints DC and a disjunctive datalog rule

$$P : \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

Answer (Worst-case Output Size Bound)

$$\max_{\mathbf{D} \models \mathsf{DC}} \log |P(\mathbf{D})| \leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B) = \text{polymatroid bound.}$$

Question (Algorithm)

Compute a model for P within $\tilde{O}(2^{\text{polymatroid bound}})$

Question (Gathering fruits)

Plug bound/algorithm into Meta Algorithm, what do we get?

Connection to Shannon-flow Inequalities

Lemma (Linearize it)

There exists non-negative $\boldsymbol{\lambda} = (\lambda_B)_{B \in \mathcal{B}}$*, with* $\|\boldsymbol{\lambda}\|_1 = 1$*, s.t.*

$$\max_{h \in \Gamma_n \cap HDC} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap HDC} \sum_{B \in \mathcal{B}} \lambda_B \, h(B) \tag{1}$$

Connection to Shannon-flow Inequalities

Lemma (Linearize it)

There exists non-negative $\boldsymbol{\lambda} = (\lambda_B)_{B \in \mathcal{B}}$*, with* $\|\boldsymbol{\lambda}\|_1 = 1$*, s.t.*

$$\max_{h \in \Gamma_n \cap \mathit{HDC}} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap \mathit{HDC}} \sum_{B \in \mathcal{B}} \lambda_B \, h(B) \tag{1}$$

Lemma (Shannon-flow inequality)

There exists $\boldsymbol{\delta} \geq \mathbf{0}$ *s.t.* $2^{\textit{polymatroid bound}} = \prod_{(X,Y,N_{Y|X})} N_{Y|X}^{\delta_{Y|X}}$*, and*

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X), \quad \forall h \in \Gamma_n \tag{2}$$

Connection to Shannon-flow Inequalities

Lemma (Linearize it)

There exists non-negative $\boldsymbol{\lambda} = (\lambda_B)_{B \in \mathcal{B}}$, *with* $\|\boldsymbol{\lambda}\|_1 = 1$, *s.t.*

$$\max_{h \in \Gamma_n \cap HDC} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap HDC} \sum_{B \in \mathcal{B}} \lambda_B \, h(B) \tag{1}$$

Lemma (Shannon-flow inequality)

There exists $\boldsymbol{\delta} \geq \mathbf{0}$ *s.t.* $2^{\textit{polymatroid bound}} = \prod_{(X,Y,N_{Y|X})} N_{Y|X}^{\delta_{Y|X}}$, *and*

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X), \; \forall h \in \Gamma_n \tag{2}$$

(2) is a (vast) generalization of Shearer's lemma

PANDA (**P**roof-**A**ssisted e**N**tropic **D**egree-**A**ware)

- What?
  - Compute a model for our disjunctive datalog rule
  - Run within $\tilde{O}\left(2^{\text{polymatroid bound}}\right) = \tilde{O}\left(\prod_{(X,Y,N_{Y|X})} N_{Y|X}^{\delta_{Y|X}}\right)$:

PANDA (**P**roof-**A**ssisted e**N**tropic **D**egree-**A**ware)

- What?
  - Compute a model for our disjunctive datalog rule
  - Run within $\tilde{O}\left(2^{\text{polymatroid bound}}\right) = \tilde{O}\left(\prod_{(X,Y,N_{Y|X})} N_{Y|X}^{\delta_{Y|X}}\right)$:
- How? Proof as symbolic instructions
  - Construct a Proof Sequence for the corresponding Shannon-flow inequality
  - Proof steps $\rightarrow$ relational operators.

Proof sequence

Shannon-flow inequality: $h(Y|X) \stackrel{\text{def}}{=} h(Y) - h(X)$, $X \subseteq Y$

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence, convert RHS to LHS using following steps

(In)equality	Steps ($X \subseteq Y$)
$h(X) + h(Y\|X) = h(Y)$	$h(X) + h(Y\|X) \to h(Y)$
$h(Y) = h(X) + h(Y\|X)$	$h(Y) \to h(X) + h(Y\|X)$
$h(Y) \geq h(X)$	$h(Y) \to h(X)$
$h(Y\|X) \geq h(Y \cup Z\|X \cup Z)$	$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$

Proof sequence

Shannon-flow inequality: $h(Y|X) \stackrel{\text{def}}{=} h(Y) - h(X),\, X \subseteq Y$

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence, convert RHS to LHS using following steps

(In)equality	Steps ($X \subseteq Y$)
$h(X) + h(Y|X) = h(Y)$	$h(X) + h(Y|X) \to h(Y)$
$h(Y) = h(X) + h(Y|X)$	$h(Y) \to h(X) + h(Y|X)$
$h(Y) \geq h(X)$	$h(Y) \to h(X)$
$h(Y|X) \geq h(Y \cup Z|X \cup Z)$	$h(Y|X) \to h(Y \cup Z|X \cup Z)$

Theorem

There is a proof sequence for every Shannon-flow inequality. The length is data-independent.

Proof Steps As Relational Operators

Shannon-flow inequality:

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence,

Steps ($X \subseteq Y$)	Relational Operator
$h(X) + h(Y\|X) \to h(Y)$	
$h(Y) \to h(X) + h(Y\|X)$	
$h(Y) \to h(X)$	
$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$	

Proof Steps As Relational Operators

Shannon-flow inequality:

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence,

Steps ($X \subseteq Y$)	Relational Operator
$h(X) + h(Y\|X) \to h(Y)$	(join)
$h(Y) \to h(X) + h(Y\|X)$	
$h(Y) \to h(X)$	
$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$	

Proof Steps As Relational Operators

Shannon-flow inequality:

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence,

Steps ($X \subseteq Y$)	Relational Operator
$h(X) + h(Y\|X) \to h(Y)$	(join)
$h(Y) \to h(X) + h(Y\|X)$	(data partition)
$h(Y) \to h(X)$	
$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$	

Proof Steps As Relational Operators

Shannon-flow inequality:

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence,

Steps ($X \subseteq Y$)	Relational Operator
$h(X) + h(Y\|X) \to h(Y)$	(join)
$h(Y) \to h(X) + h(Y\|X)$	(data partition)
$h(Y) \to h(X)$	(projection)
$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$	

Proof Steps As Relational Operators

Shannon-flow inequality:

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence,

Steps ($X \subseteq Y$)	Relational Operator
$h(X) + h(Y\|X) \to h(Y)$	(join)
$h(Y) \to h(X) + h(Y\|X)$	(data partition)
$h(Y) \to h(X)$	(projection)
$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$	(NOP)

Proof Steps As Relational Operators

Shannon-flow inequality:

$$\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot h(Y|X)$$

Proof sequence,

Steps ($X \subseteq Y$)	Relational Operator
$h(X) + h(Y\|X) \to h(Y)$	(join)
$h(Y) \to h(X) + h(Y\|X)$	(data partition)
$h(Y) \to h(X)$	(projection)
$h(Y\|X) \to h(Y \cup Z\|X \cup Z)$	(NOP)

Theorem

PANDA solves any disjunctive datalog rule P in time

$$\tilde{O}(N + \textbf{poly}(\log N) \cdot 2^{\textit{polymatroid bound for } P})$$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\begin{aligned}
\log |P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)} \\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)} \\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)} \\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Example: $P: \ T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$

$$\begin{aligned}
\log|P| &\le \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)}\\
&\le \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)}\\
&\le \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)}\\
&\le \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\begin{aligned}
\log|P| &\le \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)}\\
&\le \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)}\\
&\le \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)}\\
&\le \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4\vert A_3) + h(A_3)\big)$

Example: $P: \ T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$
\begin{aligned}
\log|P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)}\\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)}\\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)}\\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}
$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4\|A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4\|A_3) + h(A_3)$	

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\begin{aligned}
\log|P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)} \\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)} \\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)} \\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4\|A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4\|A_3) + h(A_3)$	$\big(h(A_4\|A_3) \to h(A_4\|A_2A_3)\big)$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\begin{aligned}
\log|P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)}\\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)}\\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)}\\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4 \mid A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_3) + h(A_3)$	$\big(h(A_4 \mid A_3) \to h(A_4 \mid A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_2A_3) + h(A_3)$	

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\begin{aligned}
\log|P| &\le \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)}\\
&\le \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)}\\
&\le \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)}\\
&\le \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4 \mid A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_3) + h(A_3)$	$\big(h(A_4 \mid A_3) \to h(A_4 \mid A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_2A_3) + h(A_3)$	$\big(h(A_2A_3) + h(A_4 \mid A_2A_3) \to h(A_2A_3A_4)\big)$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\log|P| \leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \text{(polymatroid bound)}$$

$$\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) \quad \text{(linearize)}$$

$$\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \text{(Shannon-flow)}$$

$$\leq \frac{3}{2}\log N \quad \text{(Cardinality constraints)}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4 \vert A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \vert A_3) + h(A_3)$	$\big(h(A_4 \vert A_3) \to h(A_4 \vert A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \vert A_2A_3) + h(A_3)$	$\big(h(A_2A_3) + h(A_4 \vert A_2A_3) \to h(A_2A_3A_4)\big)$
$h(A_1A_2) + h(A_2A_3A_4) + h(A_3)$	

Example: $P: \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\begin{aligned}
\log |P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)} \\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)} \\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)} \\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4 \vert A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \vert A_3) + h(A_3)$	$\big(h(A_4 \vert A_3) \to h(A_4 \vert A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \vert A_2A_3) + h(A_3)$	$\big(h(A_2A_3) + h(A_4 \vert A_2A_3) \to h(A_2A_3A_4)\big)$
$h(A_1A_2) + h(A_2A_3A_4) + h(A_3)$	$\big(h(A_1A_2) \to h(A_1A_2 \vert A_3)\big)$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$
\begin{aligned}
\log |P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)} \\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)} \\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)} \\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}
$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4 \mid A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_3) + h(A_3)$	$\big(h(A_4 \mid A_3) \to h(A_4 \mid A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_2A_3) + h(A_3)$	$\big(h(A_2A_3) + h(A_4 \mid A_2A_3) \to h(A_2A_3A_4)\big)$
$h(A_1A_2) + h(A_2A_3A_4) + h(A_3)$	$\big(h(A_1A_2) \to h(A_1A_2 \mid A_3)\big)$
$h(A_1A_2 \mid A_3) + h(A_2A_3A_4) + h(A_3)$	

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\begin{aligned}
\log|P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)} \\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)} \\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)} \\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4 \mid A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_3) + h(A_3)$	$\big(h(A_4 \mid A_3) \to h(A_4 \mid A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4 \mid A_2A_3) + h(A_3)$	$\big(h(A_2A_3) + h(A_4 \mid A_2A_3) \to h(A_2A_3A_4)\big)$
$h(A_1A_2) + h(A_2A_3A_4) + h(A_3)$	$\big(h(A_1A_2) \to h(A_1A_2 \mid A_3)\big)$
$h(A_1A_2 \mid A_3) + h(A_2A_3A_4) + h(A_3)$	$\big(h(A_1A_2 \mid A_3) + h(A_3) \to h(A_1A_2A_3)\big)$

Example: $P : T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$$\begin{aligned}
\log |P| &\leq \min(h(A_1A_2A_3), h(A_2A_3A_4)) && \text{(polymatroid bound)} \\
&\leq \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big) && \text{(linearize)} \\
&\leq \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) && \text{(Shannon-flow)} \\
&\leq \frac{3}{2}\log N && \text{(Cardinality constraints)}
\end{aligned}$$

Proof sequence	Proof Step
$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$	$\big(h(A_3A_4) \to h(A_4\|A_3) + h(A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4\|A_3) + h(A_3)$	$\big(h(A_4\|A_3) \to h(A_4\|A_2A_3)\big)$
$h(A_1A_2) + h(A_2A_3) + h(A_4\|A_2A_3) + h(A_3)$	$\big(h(A_2A_3) + h(A_4\|A_2A_3) \to h(A_2A_3A_4)\big)$
$h(A_1A_2) + h(A_2A_3A_4) + h(A_3)$	$\big(h(A_1A_2) \to h(A_1A_2\|A_3)\big)$
$h(A_1A_2\|A_3) + h(A_2A_3A_4) + h(A_3)$	$\big(h(A_1A_2\|A_3) + h(A_3) \to h(A_1A_2A_3)\big)$
$h(A_1A_2A_3) + h(A_2A_3A_4)$	

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$h(A_3A_4) \to h(A_4|A_3) + h(A_3)$

$h(A_4|A_3) \to h(A_4|A_2A_3)$

$h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)$

$h(A_1A_2) \to h(A_1A_2|A_3)$

$h(A_1A_2|A_3) + h(A_3) \to h(A_1A_2A_3)$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$h(A_3 A_4) \rightarrow h(A_4|A_3) + h(A_3)$

$R_{34}(A_3, A_4) \rightarrow R_{34}^{(\ell)}(A_3, A_4), R_3^{(h)}(A_3)$

$h(A_4|A_3) \rightarrow h(A_4|A_2 A_3)$

$h(A_2 A_3) + h(A_4|A_2 A_3) \rightarrow h(A_2 A_3 A_4)$

$h(A_1 A_2) \rightarrow h(A_1 A_2|A_3)$

$h(A_1 A_2|A_3) + h(A_3) \rightarrow h(A_1 A_2 A_3)$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$h(A_3 A_4) \to h(A_4|A_3) + h(A_3)$

$h(A_4|A_3) \to h(A_4|A_2 A_3)$

$h(A_2 A_3) + h(A_4|A_2 A_3) \to h(A_2 A_3 A_4)$

$h(A_1 A_2) \to h(A_1 A_2|A_3)$

$h(A_1 A_2|A_3) + h(A_3) \to h(A_1 A_2 A_3)$

$R_{34}(A_3, A_4) \to R_{34}^{(\ell)}(A_3, A_4), R_3^{(h)}(A_3)$

$R_{34}^{(\ell)}(A_3, A_4) \to R_{34}^{(\ell)}(A_3, A_4)$

Example: $P : T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$h(A_3 A_4) \rightarrow h(A_4|A_3) + h(A_3)$

$h(A_4|A_3) \rightarrow h(A_4|A_2 A_3)$

$h(A_2 A_3) + h(A_4|A_2 A_3) \rightarrow h(A_2 A_3 A_4)$

$h(A_1 A_2) \rightarrow h(A_1 A_2|A_3)$

$h(A_1 A_2|A_3) + h(A_3) \rightarrow h(A_1 A_2 A_3)$

$R_{34}(A_3, A_4) \rightarrow R_{34}^{(\ell)}(A_3, A_4), R_3^{(h)}(A_3)$

$R_{34}^{(\ell)}(A_3, A_4) \rightarrow R_{34}^{(\ell)}(A_3, A_4)$

$R_{23}(A_2, A_3) \bowtie R_{34}^{(\ell)}(A_3, A_4) \rightarrow T_{234}(A_2, A_3, A_4)$

Example: $P : \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$h(A_3 A_4) \to h(A_4|A_3) + h(A_3)$

$h(A_4|A_3) \to h(A_4|A_2 A_3)$

$h(A_2 A_3) + h(A_4|A_2 A_3) \to h(A_2 A_3 A_4)$

$h(A_1 A_2) \to h(A_1 A_2|A_3)$

$h(A_1 A_2|A_3) + h(A_3) \to h(A_1 A_2 A_3)$

$R_{34}(A_3, A_4) \to R_{34}^{(\ell)}(A_3, A_4), R_3^{(h)}(A_3)$

$R_{34}^{(\ell)}(A_3, A_4) \to R_{34}^{(\ell)}(A_3, A_4)$

$R_{23}(A_2, A_3) \bowtie R_{34}^{(\ell)}(A_3, A_4) \to T_{234}(A_2, A_3, A_4)$

$R_{12}(A_1, A_2) \to R_{12}(A_1, A_2)$

Example: $P: \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}$.

$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$$

$h(A_3A_4) \rightarrow h(A_4\mid A_3) + h(A_3)$	$R_{34}(A_3, A_4) \rightarrow R^{(\ell)}_{34}(A_3, A_4), R^{(h)}_3(A_3)$
$h(A_4\mid A_3) \rightarrow h(A_4\mid A_2A_3)$	$R^{(\ell)}_{34}(A_3, A_4) \rightarrow R^{(\ell)}_{34}(A_3, A_4)$
$h(A_2A_3) + h(A_4\mid A_2A_3) \rightarrow h(A_2A_3A_4)$	$R_{23}(A_2, A_3) \bowtie R^{(\ell)}_{34}(A_3, A_4) \rightarrow T_{234}(A_2, A_3, A_4)$
$h(A_1A_2) \rightarrow h(A_1A_2\mid A_3)$	$R_{12}(A_1, A_2) \rightarrow R_{12}(A_1, A_2)$
$h(A_1A_2\mid A_3) + h(A_3) \rightarrow h(A_1A_2A_3)$	$R_{12}(A_1, A_2) \bowtie R^{(h)}_3(A_3) \rightarrow T_{123}(A_1, A_2, A_3)$

Table of Contents

Roadmap

Given degree constraints DC and a disjunctive datalog rule

$$P : \bigvee_{B \in \mathcal{B}} T_B \leftarrow \bigwedge_{F \in \mathcal{E}} R_F$$

Answer (Worst-case Output Size Bound)

Polymatroid bound $\max_{\mathbf{D} \models \mathsf{DC}} \log |P(\mathbf{D})| \leq \max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$

Answer (Algorithm)

PANDA computes a model for P within $\tilde{O}(2^{\text{polymatroid bound}})$

Question (Gathering fruits)

Plug bound/algorithm into Meta Algorithm, what do we get?

Option 1: Full Conjunctive Query

Full conjunctive query

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\quad \textsf{iTime} \quad} T_{[n]} \xrightarrow{\quad \textsf{oTime} \quad} \textsf{Answer}$$

$$\textsf{iTime} = \max_{\mathbf{D} \models \textsf{DC}} |Q_i(\mathbf{D})| \leq \max_{h \in \Gamma_n \cap \textsf{HDC}} 2^{h([n])} = \textsf{PANDA's} \text{ runtime}$$

PANDA is worst-case optimal whenever the polymatroid bound is tight!

Option 2: A Single Tree Decomposition

Let $(\mathcal{T}, \chi)$ be a tree decomposition of $\mathcal{H}$ to be chosen later

Multiple Conjunctive Rules

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F \in \mathcal{E}} R_F \xrightarrow{\quad \mathsf{iTime} \quad} \bigwedge_{v \in V(\mathcal{T})} T_{\chi(v)} \xrightarrow{\quad \mathsf{oTime} \quad} \text{Answer}$$

$$\mathsf{iTime} = \max_{\mathbf{D} \models \mathsf{DC}} |Q_i(\mathbf{D})| \leq \max_{\mathbf{D} \models \mathsf{DC}} \max_{v \in V(\mathcal{T})} |P_v(\mathbf{D})|$$

$$\leq 2^{\max_{v \in V(\mathcal{T})} \max_{h \in \Gamma_n \cap \mathsf{DC}} h(\chi(v))} = \text{PANDA's runtime}$$

Pick the best $(\mathcal{T}, \chi)$ before running PANDA:

$$\min_{(\mathcal{T}, \chi)} \max_{v \in V(\mathcal{T})} \max_{h \in \Gamma_n \cap \mathsf{CC}} h(\chi(v)) \leq \log N \cdot \mathsf{fhtw}(Q)$$

PANDA evaluates Q within $\tilde{O}(N^{\mathsf{fhtw}(Q)})$-time.

Option 3: Multiple Tree Decompositions

Multiple Disjunctive Datalog Rules!

$$\text{Database } \mathbf{D} \longrightarrow Q_i \longrightarrow Q_i(\mathbf{D}) \longrightarrow Q_o \longrightarrow \text{Output}$$

$$\bigwedge_{F\in\mathcal{E}} R_F \xrightarrow{\quad \mathsf{iTime} \quad} \bigwedge_{\mathcal{B}} \bigvee_{B\in\mathcal{B}} T_B \xrightarrow{\quad \mathsf{oTime} \quad} \mathsf{Answer}$$

$$\log \mathsf{iTime} = \max_{\mathbf{D}\models\mathsf{DC}} \log |Q_i(\mathbf{D})| \leq \max_{\mathbf{D}\models\mathsf{DC}} \max_{\mathcal{B}} \log |P_{\mathcal{B}}(\mathbf{D})|$$

$$\leq \max_{\mathcal{B}} \max_{h\in\Gamma_n\cap\mathsf{HDC}} \min_{B\in\mathcal{B}} h(B) = \max_{h\in\Gamma_n\cap\mathsf{HDC}} \max_{\mathcal{B}} \min_{B\in\mathcal{B}} h(B)$$

$$= \max_{h\in\Gamma_n\cap\mathsf{HDC}} \max_{(\mathcal{T},\chi)} \min_{v\in V(\mathcal{T})} h(\chi(v)) = \log(\text{PANDA's runtime})$$

$$\max_{h\in\Gamma_n\cap\mathsf{HDC}} \max_{(\mathcal{T},\chi)} \min_{v\in V(\mathcal{T})} h(\chi(v)) \leq \log N \cdot \mathsf{subw}(Q)$$

PANDA evaluates Q within $\tilde{O}(N^{\mathsf{subw}(Q)})$-time.

Quantities of Interests

- $X \in \{\overline{\Gamma}^*_n, \Gamma_n, \mathsf{SA}_n\}$
- $Y \in \{\mathsf{HDC}, \mathsf{HFD}, \mathsf{HCC}, \log N \cdot \mathsf{ED}, \log N \cdot \mathsf{VD}\}$

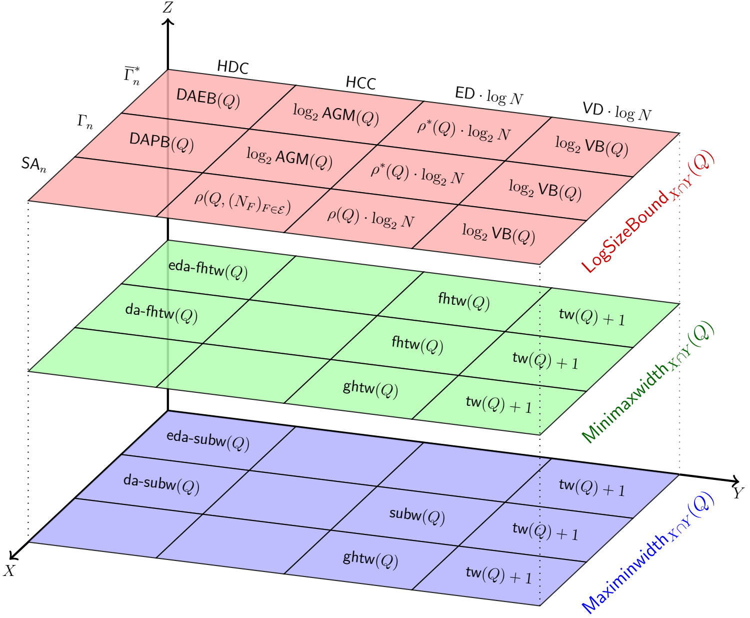
Define

$$\mathsf{LogSizeBound}_{X \cap Y}(P) \stackrel{\text{def}}{=} \max_{h \in X \cap Y} \min_{B \in \mathcal{B}} h(B)$$

$$\mathsf{Minimaxwidth}_{X \cap Y}(Q) \stackrel{\text{def}}{=} \min_{(\mathcal{T}, \chi)} \max_{v \in V(\mathcal{T})} \max_{h \in X \cap Y} h(\chi(v)),$$

$$\mathsf{Maximinwidth}_{X \cap Y}(Q) \stackrel{\text{def}}{=} \max_{h \in X \cap Y} \min_{(\mathcal{T}, \chi)} \max_{v \in V(\mathcal{T})} h(\chi(v)).$$

Summary of Bounds



Many Open Questions

- Is the entropic bound computable under $\mathsf{CC} \cup \mathsf{HDC}$ or DC?

Many Open Questions

- Is the entropic bound computable under $\mathsf{CC} \cup \mathsf{HDC}$ or DC?
- Worst-case optimal algorithm for full conjunctive queries under $\mathsf{CC} \cup \mathsf{HDC}$ or DC

Many Open Questions

- Is the entropic bound computable under $\mathsf{CC} \cup \mathsf{HDC}$ or DC?
- Worst-case optimal algorithm for full conjunctive queries under $\mathsf{CC} \cup \mathsf{HDC}$ or DC
- Worst-case optimal algorithm for disjunctive datalog rules under $\mathsf{CC} \cup \mathsf{HDC}$ or DC

Many Open Questions

- Is the entropic bound computable under $\mathsf{CC} \cup \mathsf{HDC}$ or DC?
- Worst-case optimal algorithm for full conjunctive queries under $\mathsf{CC} \cup \mathsf{HDC}$ or DC
- Worst-case optimal algorithm for disjunctive datalog rules under $\mathsf{CC} \cup \mathsf{HDC}$ or DC
- Remove the annoying poly-log factor from PANDA

Many Open Questions

- Is the entropic bound computable under CC ∪ HDC or DC?
- Worst-case optimal algorithm for full conjunctive queries under CC ∪ HDC or DC
- Worst-case optimal algorithm for disjunctive datalog rules under CC ∪ HDC or DC
- Remove the annoying poly-log factor from PANDA
- Other choices for the propositional formula in the Meta Algorithm, perhaps tradding off iTime and $\otimes$?

Many Open Questions

- Is the entropic bound computable under $\mathsf{CC} \cup \mathsf{HDC}$ or DC?
- Worst-case optimal algorithm for full conjunctive queries under $\mathsf{CC} \cup \mathsf{HDC}$ or DC
- Worst-case optimal algorithm for disjunctive datalog rules under $\mathsf{CC} \cup \mathsf{HDC}$ or DC
- Remove the annoying poly-log factor from PANDA
- Other choices for the propositional formula in the Meta Algorithm, perhaps tradding off iTime and $\otimes$?
- What about negations?

Many Thanks!
Questions?

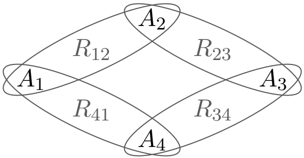
Table of Contents

Semantics of Our Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

Semantics of Our Disjunctive Datalog Rule

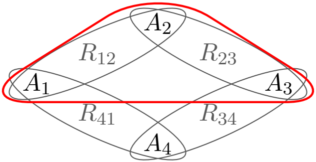
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

Semantics of Our Disjunctive Datalog Rule

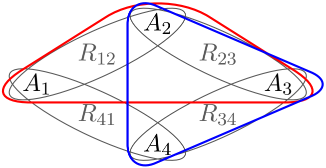
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

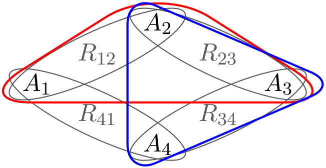
Semantics of Our Disjunctive Datalog Rule

$$P : \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P_{123,234} : T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

Semantics of Our Disjunctive Datalog Rule



$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

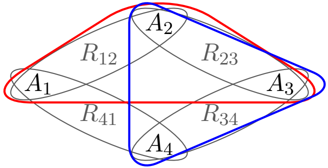
A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

Semantics of Our Disjunctive Datalog Rule



$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234}: \color{red}{T_{123}} \vee \color{blue}{T_{234}} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

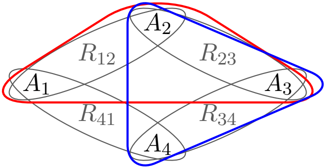
A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

Semantics of Our Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

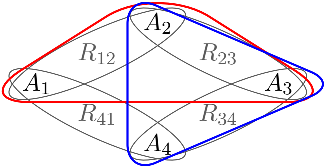
A_4	A_1
3	b
4	a
4	b

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4
1	c	3
2	d	4

Semantics of Our Disjunctive Datalog Rule



$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234} : T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

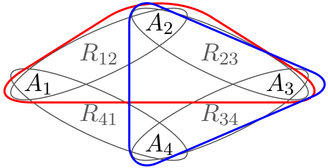
A_4	A_1
3	b
4	a
4	b

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4
1	c	3
2	d	4

Semantics of Our Disjunctive Datalog Rule



$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

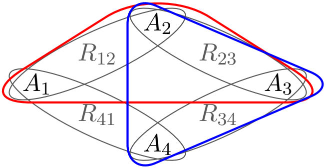
A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4
1	c	3
2	d	4

Inclusive Disjunction

Semantics of Our Disjunctive Datalog Rule



$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

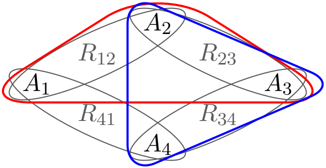
A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4
1	c	3
2	d	4

No *minimal model* requirement

Semantics of Our Disjunctive Datalog Rule



$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

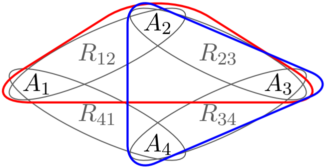
A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4
1	c	3
2	d	4

Model size is $\max(|T_{123}|, |T_{234}|) = 3$

Semantics of Our Disjunctive Datalog Rule



$$P: \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

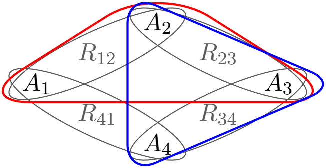
A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4
1	c	3
2	d	4

Output size is the minimum over all models

Semantics of Our Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P_{123,234}: T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

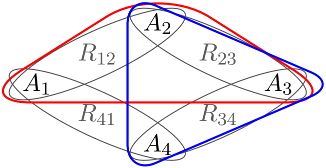
A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4

A minimum-sized model of size 2

Semantics of Our Disjunctive Datalog Rule

$$P : \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P_{123,234} : T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4

A minimum-sized model of size 2
Hence, the output size is 2

Output Size of a Disjunctive Datalog Rule

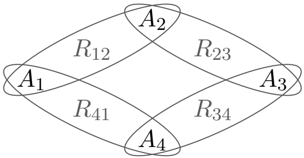
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$

Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

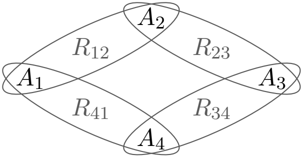
$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$



Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$

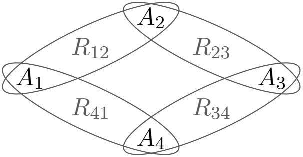


CC : $\quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$

Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$



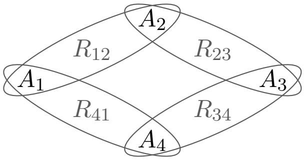
CC : $\quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$

- $P_{123,234} \quad : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$



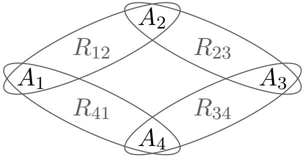
CC : $\quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$

- $P_{123,234} \quad : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$\max_{\mathbf{D} \models \mathsf{CC}} |P_{123,234}(\mathbf{D})| \leq N^{3/2}$$

Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$

$$\mathsf{CC}: \quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$$

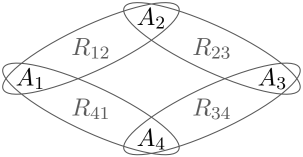
- $P_{123,234} \; : \; T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$\max_{\mathbf{D} \models \mathsf{CC}} |P_{123,234}(\mathbf{D})| \leq N^{3/2}$$

- $P' \; : \; T_{123} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$

$$\mathsf{CC}: \quad |R_{12}| \le N, \quad |R_{23}| \le N, \quad |R_{34}| \le N, \quad |R_{41}| \le N.$$

- $P_{123,234} \quad : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$\max_{\mathbf{D} \models \mathsf{CC}} |P_{123,234}(\mathbf{D})| \le N^{3/2}$$

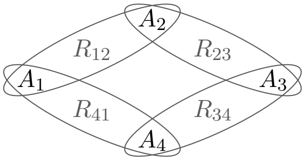
- $P' \quad : \quad T_{123} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$|P'(\mathbf{D})| = N^2$, for some $\mathbf{D}$

Output Size of a Disjunctive Datalog Rule

$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \leftarrow \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T} \models P} \max_{B \in \mathcal{B}} |T_B|$$



CC : $\quad |R_{12}| \leq N, \quad |R_{23}| \leq N, \quad |R_{34}| \leq N, \quad |R_{41}| \leq N.$

- $P_{123,234} \quad : \quad T_{123} \vee T_{234} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$$\max_{\mathbf{D} \models \mathsf{CC}} |P_{123,234}(\mathbf{D})| \leq N^{3/2}$$

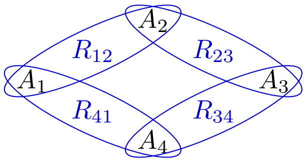
- $P' \quad : \quad T_{123} \leftarrow R_{12} \wedge R_{23} \wedge R_{34} \wedge R_{41}.$

$|P'(\mathbf{D})| = N^2$, for some $\mathbf{D}$

Using Option 3, 4-cycle query answerable in $\tilde{O}(N^{3/2})$-time, matching [Alon, Yuster, Zwick'97]

Polymatroid Bound: Examples

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

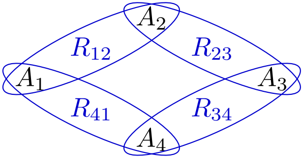


- $|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N$

$|Q| \leq N^2$

Polymatroid Bound: Examples

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



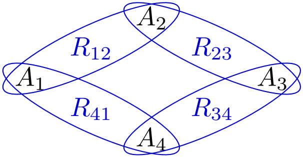
- $|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N$

$|Q| \leq N^2$

$\log |Q| = h(A_1A_2A_3A_4) \leq h(A_1A_2) + h(A_3A_4) \leq 2 \log N$

Polymatroid Bound: Examples

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



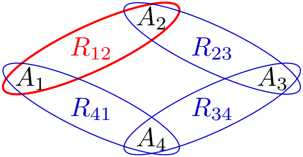
- $|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N$ $\quad$ $|Q| \leq N^2$

  $\log|Q| = h(A_1A_2A_3A_4) \leq h(A_1A_2) + h(A_3A_4) \leq 2\log N$

- $\deg_{12}(A_1A_2|A_1), \deg_{12}(A_1A_2|A_2) \leq D$ $\quad$ $|Q| \leq D \cdot N^{3/2}$

Polymatroid Bound: Examples

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



- $|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N$ $\quad |Q| \leq N^2$

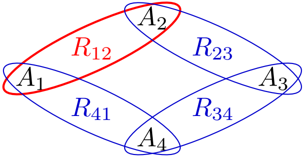
  $\log |Q| = h(A_1 A_2 A_3 A_4) \leq h(A_1 A_2) + h(A_3 A_4) \leq 2 \log N$

- $\deg_{12}(A_1 A_2 | A_1), \deg_{12}(A_1 A_2 | A_2) \leq D$ $\quad |Q| \leq D \cdot N^{3/2}$

$$\begin{aligned} 2 \log |Q| &= 2h(A_1 A_2 A_3 A_4) \\ &\leq h(A_2 A_3) + h(A_3 A_4) + h(A_4 A_1) + h(A_2 | A_1) + h(A_1 | A_2) \\ &\leq 3 \log N + 2 \log D \end{aligned}$$

Polymatroid Bound: Examples

$$Q(A_1, A_2, A_3, A_4) \leftarrow R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



- $|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N$ $\quad$ $|Q| \leq N^2$

  $\log |Q| = h(A_1A_2A_3A_4) \leq h(A_1A_2) + h(A_3A_4) \leq 2 \log N$

- $\deg_{12}(A_1A_2|A_1), \deg_{12}(A_1A_2|A_2) \leq D$ $\quad$ $|Q| \leq D \cdot N^{3/2}$

$$\begin{aligned} 2 \log |Q| &= 2h(A_1A_2A_3A_4) \\ &\leq h(A_2A_3) + h(A_3A_4) + h(A_4A_1) + h(A_2|A_1) + h(A_1|A_2) \\ &\leq 3 \log N + 2 \log D \end{aligned}$$

- $A_1 \rightarrow A_2, \qquad A_2 \rightarrow A_1$ $\quad$ $|Q| \leq N^{3/2}$

Polymatroid Bound is *not* tight!

Polymatroid Bound is *not* tight!

Proof Strategy

Polymatroid Bound is *not* tight!

Proof Strategy

- Take a non-Shannon inequality (modified from [Zhang-Yeung'98])

$$
\begin{aligned}
11h(ABXYC) \leq\ & 3h(XY) + 3h(AX) + 3h(AY) + h(BX) \\
+\ & h(BY) + 5h(C) + h(ABXYC|AB) + h(ABXYC|AC) \\
+\ & 4h(ABXYC|AXY) + h(ABXYC|BXY) \\
+\ & 2h(ABXYC|XC) + 2h(ABXYC|YC) \qquad (3)
\end{aligned}
$$

Polymatroid Bound is *not* tight!

Proof Strategy

- Take a non-Shannon inequality (modified from [Zhang-Yeung'98])

$$\begin{aligned} 11h(ABXYC) &\leq 3h(XY) + 3h(AX) + 3h(AY) + h(BX) \\ &+ h(BY) + 5h(C) + h(ABXYC|AB) + h(ABXYC|AC) \\ &+ 4h(ABXYC|AXY) + h(ABXYC|BXY) \\ &+ 2h(ABXYC|XC) + 2h(ABXYC|YC) \end{aligned} \quad (3)$$

- Construct a query Q where (3) gives the (entropic) bound

$$11 \log |Q| = 11h(ABXYC) \leq 11 \log N^3 + 5 \log N^2 = 43 \log N$$

Polymatroid Bound is *not* tight!

Proof Strategy

- Take a non-Shannon inequality (modified from [Zhang-Yeung'98])

$$
\begin{aligned}
11h(ABXYC) \quad \leq \quad & 3h(XY) + 3h(AX) + 3h(AY) + h(BX) \\
+ \quad & h(BY) + 5h(C) + h(ABXYC|AB) + h(ABXYC|AC) \\
+ \quad & 4h(ABXYC|AXY) + h(ABXYC|BXY) \\
+ \quad & 2h(ABXYC|XC) + 2h(ABXYC|YC)
\end{aligned} \tag{3}
$$

- Construct a query Q where (3) gives the (entropic) bound

$$11 \log |Q| = 11h(ABXYC) \leq 11 \log N^3 + 5 \log N^2 = 43 \log N$$

- Construct a polymatroid h satisfying Q's constraints such that

$$11h(ABXYC) = 44 \log N$$

Beyond Worst-case Optimality

- Output-sensitive algorithms

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\quad N^d \quad + \quad |\mathsf{output}| \quad \Big)$$

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + |\mathsf{output}| \Big)$$

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

- Submodular width as a candidate for d

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

- Submodular width as a candidate for d [Marx JACM'13]

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

- Submodular width as a candidate for d [Marx JACM'13]

$$\mathbf{FPT} \quad \Leftrightarrow \quad \text{Bounded } \mathsf{subw}(Q)$$

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

- Submodular width as a candidate for d [Marx JACM'13]

$$\textbf{FPT} \Leftrightarrow \text{Bounded } \mathsf{subw}(Q)$$

$$\text{Boolean } Q \Rightarrow \tilde{O}\left(N^{\mathsf{subw}(Q) \times c}\right)$$

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

- Submodular width as a candidate for d [Marx JACM'13]

$$\textbf{FPT} \quad \Leftrightarrow \quad \text{Bounded } \mathsf{subw}(Q)$$

$$\text{Boolean } Q \quad \Rightarrow \quad \tilde{O}\left(N^{\mathsf{subw}(Q) \times c}\right)$$

- Our goals

Beyond Worst-case Optimality

- Output-sensitive algorithms

$$\tilde{O}\Big(\underbrace{N^d}_{\text{Intrinsic Cost}} + \underbrace{|\mathsf{output}|}_{\text{Output Cost}} \Big)$$

- Submodular width as a candidate for d [Marx JACM'13]

$$\textbf{FPT} \quad \Leftrightarrow \quad \text{Bounded } \mathsf{subw}(Q)$$

$$\text{Boolean } Q \quad \Rightarrow \quad \tilde{O}\Big(N^{\mathsf{subw}(Q) \times c}\Big)$$

- Our goals

$$\textit{Any } Q \quad \Rightarrow \quad \tilde{O}\Big(N^{\text{da-}\mathsf{subw}(Q) \times 1} + |\mathsf{output}|\Big)$$

Submodular Width

$$\mathsf{fhtw}(Q) \stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t))$$

Submodular Width

$$\begin{aligned}
\mathsf{fhtw}(Q) &\stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)) \\
&= \min_{(T,\chi)} \max_{t \in V(T)} \max_{h \in \mathsf{ED} \cap \Gamma_n} h(\chi(t))
\end{aligned}$$

Submodular Width

$$\mathsf{fhtw}(Q) \stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t))$$

$$= \min_{(T,\chi)} \max_{t \in V(T)} \max_{h \in \mathsf{ED} \cap \Gamma_n} h(\chi(t))$$

$$= \min_{(T,\chi)} \max_{h \in \mathsf{ED} \cap \Gamma_n} \max_{t \in V(T)} h(\chi(t))$$

Submodular Width

$$\begin{aligned}
\mathsf{fhtw}(Q) &\stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)) \\
&= \min_{(T,\chi)} \max_{t \in V(T)} \max_{h \in \mathsf{ED} \cap \Gamma_n} h(\chi(t)) \\
&= \min_{(T,\chi)} \max_{h \in \mathsf{ED} \cap \Gamma_n} \max_{t \in V(T)} h(\chi(t)) \\
\mathsf{subw}(Q) &\stackrel{\text{def}}{=} \max_{h \in \mathsf{ED} \cap \Gamma_n} \min_{(T,\chi)} \max_{t \in V(T)} h(\chi(t))
\end{aligned}$$

Submodular Width

$$\begin{aligned}
\mathsf{fhtw}(Q) &\stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)) \\
&= \min_{(T,\chi)} \max_{t \in V(T)} \max_{h \in \mathsf{ED} \cap \Gamma_n} h(\chi(t)) \\
&= \min_{(T,\chi)} \max_{h \in \mathsf{ED} \cap \Gamma_n} \max_{t \in V(T)} h(\chi(t)) \\
\mathsf{subw}(Q) &\stackrel{\text{def}}{=} \max_{h \in \mathsf{ED} \cap \Gamma_n} \min_{(T,\chi)} \max_{t \in V(T)} h(\chi(t))
\end{aligned}$$

$$\mathsf{subw}(Q) \leq \mathsf{fhtw}(Q)$$