Query Optimization & Evaluation via Information Theory

Mahmoud Abo Khamis

University of Oxford & RelationalAI

PODS 2026 - Tutorial

Conjunctive Queries (CQs)

$$Q(\boldsymbol{F}) \quad :- \quad \bigwedge_{R(\boldsymbol{X}) \in \text{atoms}(Q)} R(\boldsymbol{X})$$

$$Q^{\text{full}}_{\square}(X, Y, Z, W) \quad :- \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

$$Q_{\square}(X, Y) \quad :- \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

R

X	Y
1	p
1	q
2	p

S

Y	Z
p	3
q	4
q	5

T

Z	W
3	i
5	i
5	j

U

W	X
i	1
j	1
k	2

$Q^{\text{full}}_{\square}$

X	Y	Z	W
1	p	3	i
1	q	5	i
1	q	5	j

$Q_{\square}$

X	Y
1	p
1	q

Three Communities Involved

- **Graph Algorithms**
  - Results are tailored to *specific queries*
  - Complexity is measured in graph size, N
- **CSP**
  - Every atom is a constraint
  - *Tree decompositions (TDs)*
  - Complexity is measured in dimensionality, $|\mathrm{Dom}|$
- **Database Theory**

Database Theory: **Query Optimizer = Meta-Algorithm**

- Input
  - **CQ** Q
  - **Statistics** $\mathcal{S}$ about the input database $\mathcal{D}$
    - Total size, N
    - Individual sizes $|R|, |S|, \ldots$
    - FDs, e.g., $X \to Y$ in $R(X, Y)$
    - $\ldots$
- Ideal output
  - **Query Plan (Algorithm)** for Q that is *optimal* given $\mathcal{S}$
    - that minimizes the worst-case runtime overall $\mathcal{D}$ that *satisfy* $\mathcal{S}$, i.e. $\mathcal{D} \models \mathcal{S}$

PANDA

A Unified Framework for Query Optimization and Evaluation

- Rooted in **information theory**
- Builds upon and unifies several lines of work
  - in graph algorithms, CSP, and DB theory
- Subsumes many prior results
- For any CQ Q and statistics $\mathcal{S}$
  - achieves the *best known combinatorial complexity*
  - provably optimal for *some* classes
- Extends to
  - non-combinatorial algorithms (FMM)
  - queries beyond CQs (#CQs, FAQs)

Statistics $\mathcal{S}$

A collection of **Degree Constraints (DCs)**:

- Given $R(\boldsymbol{Z})$ and $\boldsymbol{X}, \boldsymbol{Y} \subseteq \boldsymbol{Z}$
  - $\deg_R(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x})$ is # distinct $\boldsymbol{Y}$-values for a given $\boldsymbol{X}$-value $\boldsymbol{x}$
  - $\deg_R(\boldsymbol{Y} \mid \boldsymbol{X}) = \max_{\boldsymbol{x}} \deg_R(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x})$
  - DC: $\deg_R(\boldsymbol{Y} \mid \boldsymbol{X}) \leq N_{\boldsymbol{Y}\mid\boldsymbol{X}}$

Example

$\deg_R(Y \mid X = 1) = 2$
$\deg_R(Y \mid X = 2) = 1$
$\deg_R(Y \mid X) = 2$
R satisfies the DC: $\deg_R(Y \mid X) \leq 2$

R

X	Y
1	p
1	q
2	p

- DCs subsume
  - *Cardinality Constraints (CCs)*: $|R(\boldsymbol{Z})| = \deg_R(\boldsymbol{Z} \mid \emptyset)$
  - *Functional Dependencies (FDs)*: $X \to Y$ means $\deg_R(Y \mid X) \leq 1$

Roadmap

- **Cardinality Estimation**
  - *Output Size*
    - Given a CQ Q and statistics $\mathcal{S}$,

      what is the worst-case output size of Q given $\mathcal{S}$?
  - *Query Plan Cost*
    - Given Q, $\mathcal{S}$, and a query plan P,

      what is the cost of evaluating Q using P?
- **Query Evaluation**
  - Given Q and $\mathcal{S}$,

    find the best query plan for evaluating Q given $\mathcal{S}$?

Outline

1. Cardinality Estimation
   - Output Size
   - Query Plan Cost
2. Query Evaluation
3. Conclusion

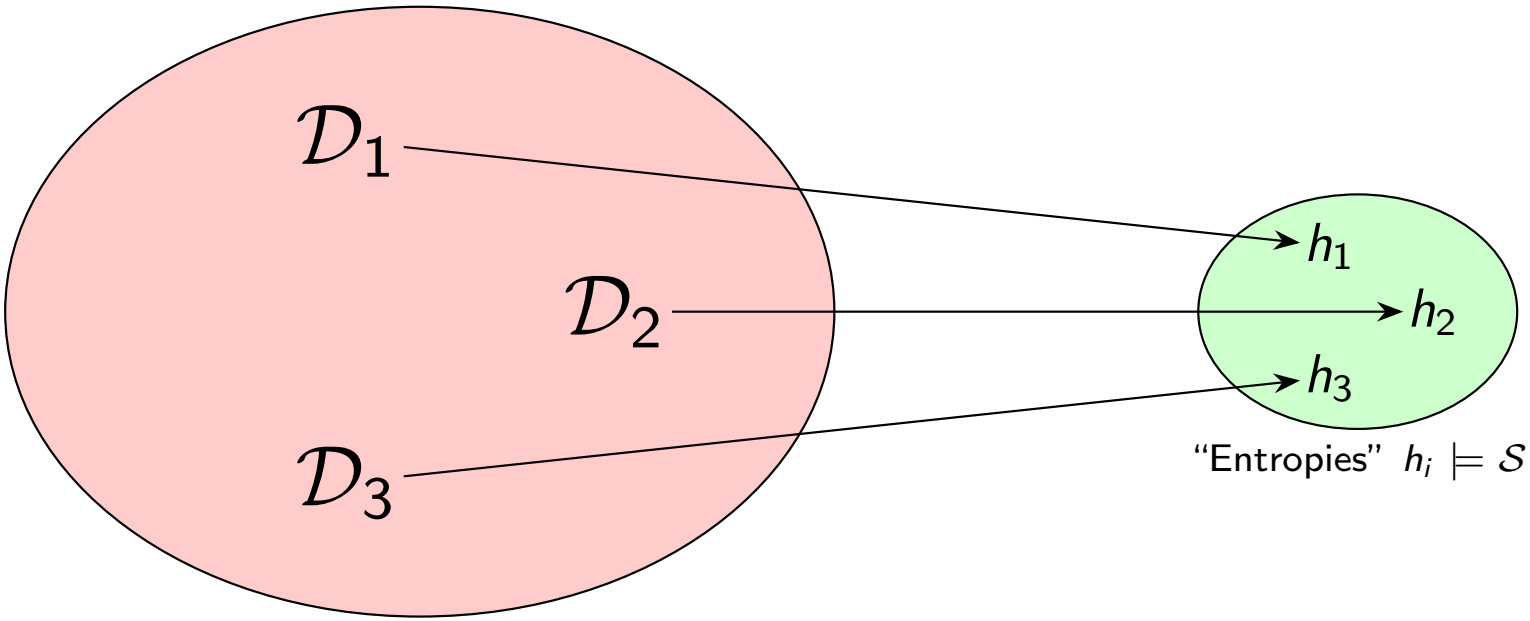
Outline

1. Cardinality Estimation
   - Output Size
   - Query Plan Cost
2. Query Evaluation
3. Conclusion

Cardinality Estimation: Output size of Q given $\mathcal{S}$

- Information theory to bound #occurrences of a pattern in graph [Itai et al. 1977; Alon 1981; Chung et al. 1986; Friedgut and Kahn 1998; Friedgut 2004; Grohe et al. 2006; Grohe et al. 2014]
- Atserias, Grohe,& Marx proved the **AGM-bound** under *CCs* [Atserias et al. 2013]
  - Shearer's inequality [Chung et al. 1986]
- Argument extended to
  - FDs [Gottlob, Lee, et al. 2012]
  - DCs [Abo Khamis, Ngo, et al. 2017]
  - ℓ_p-norms on degree vectors [Jayaraman et al. 2021; Abo Khamis, Nakos, et al. 2024]
  - Correlations among degree vectors [Mayer et al. 2026]

Why Information Theory?



Databases $\mathcal{D}_i \models \mathcal{S}$

Find $\mathcal{D}_i$ that maximizes $|Q(\mathcal{D}_i)|$

Find h_i that maximizes $h_i(\boldsymbol{F})$

Background: Entropic Functions & Polymatroids

- Given a prob. dist. $p : \mathrm{Dom}^{\boldsymbol{V}} \to \mathbb{R}_+$, let $p_{\boldsymbol{X}}$ be the marginal over $\boldsymbol{X} \subseteq \boldsymbol{V}$

$$h(\boldsymbol{X}) \stackrel{\text{def}}{=} -\sum_{\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}} p_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \log p_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

Background: Entropic Functions & Polymatroids

- Given a prob. dist. $p : \mathrm{Dom}^{\boldsymbol{V}} \to \mathbb{R}_+$, let $p_{\boldsymbol{X}}$ be the marginal over $\boldsymbol{X} \subseteq \boldsymbol{V}$

$$h(\boldsymbol{X}) \quad \stackrel{\text{def}}{=} \quad - \sum_{\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}} p_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \log p_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is **entropic** iff there exists p satisfying (1) $\forall \boldsymbol{X} \subseteq \boldsymbol{V}$

Background: Entropic Functions & Polymatroids

- Given a prob. dist. $p : \mathrm{Dom}^{\boldsymbol{V}} \to \mathbb{R}_+$, let $p_{\boldsymbol{X}}$ be the marginal over $\boldsymbol{X} \subseteq \boldsymbol{V}$

$$h(\boldsymbol{X}) \quad \stackrel{\text{def}}{=} \quad - \sum_{\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}} p_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \log p_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is **entropic** iff there exists p satisfying (1) $\forall \boldsymbol{X} \subseteq \boldsymbol{V}$
- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is a **polymatroid** iff it satisfies the **basic Shannon inequalities**:

$$\begin{aligned} h(\emptyset) &= 0 \\ h(\boldsymbol{X}) &\leq h(\boldsymbol{X} \cup \boldsymbol{Y}) && \textit{(monotonicity)} \\ \underbrace{h(\boldsymbol{X}) + h(\boldsymbol{Y}) &\geq h(\boldsymbol{X} \cup \boldsymbol{Y}) + h(\boldsymbol{X} \cap \boldsymbol{Y})}_{} && \textit{(submodularity)} \end{aligned}$$

Background: Entropic Functions & Polymatroids

- Given a prob. dist. $p : \mathrm{Dom}^{\boldsymbol{V}} \to \mathbb{R}_+$, let $p_{\boldsymbol{X}}$ be the marginal over $\boldsymbol{X} \subseteq \boldsymbol{V}$

$$h(\boldsymbol{X}) \quad \stackrel{\text{def}}{=} \quad -\sum_{\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}} p_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \log p_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is **entropic** iff there exists p satisfying (1) $\forall \boldsymbol{X} \subseteq \boldsymbol{V}$
- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is a **polymatroid** iff it satisfies the **basic Shannon inequalities**:

$$\begin{aligned}
h(\emptyset) &= 0 \\
h(\boldsymbol{X}) &\leq h(\boldsymbol{X} \cup \boldsymbol{Y}) && \textit{(monotonicity)} \\
\underbrace{h(\boldsymbol{X}) + h(\boldsymbol{Y}) &\geq h(\boldsymbol{X} \cup \boldsymbol{Y}) + h(\boldsymbol{X} \cap \boldsymbol{Y})}_{\substack{h(\boldsymbol{A}|\boldsymbol{B}) \geq h(\boldsymbol{A}|\boldsymbol{B} \cup \boldsymbol{C}) \\ h(\boldsymbol{A}|\boldsymbol{B}) \stackrel{\text{def}}{=} h(\boldsymbol{A} \cup \boldsymbol{B}) - h(\boldsymbol{B})}} && \textit{(submodularity)}
\end{aligned}$$

Background: Entropic Functions & Polymatroids

- Given a prob. dist. $p : \mathrm{Dom}^{\boldsymbol{V}} \to \mathbb{R}_+$, let $p_{\boldsymbol{X}}$ be the marginal over $\boldsymbol{X} \subseteq \boldsymbol{V}$

$$h(\boldsymbol{X}) \quad \stackrel{\text{def}}{=} \quad -\sum_{\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}} p_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \log p_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is **entropic** iff there exists p satisfying (1) $\forall \boldsymbol{X} \subseteq \boldsymbol{V}$
- $h : 2^{\boldsymbol{V}} \to \mathbb{R}_+$ is a **polymatroid** iff it satisfies the **basic Shannon inequalities**:

$$\begin{aligned}
h(\emptyset) &= 0 \\
h(\boldsymbol{X}) &\leq h(\boldsymbol{X} \cup \boldsymbol{Y}) && \textit{(monotonicity)} \\
\underbrace{h(\boldsymbol{X}) + h(\boldsymbol{Y}) &\geq h(\boldsymbol{X} \cup \boldsymbol{Y}) + h(\boldsymbol{X} \cap \boldsymbol{Y})}_{\substack{h(\boldsymbol{A}|\boldsymbol{B}) \geq h(\boldsymbol{A}|\boldsymbol{B} \cup \boldsymbol{C}) \\ h(\boldsymbol{A}|\boldsymbol{B}) \stackrel{\text{def}}{=} h(\boldsymbol{A} \cup \boldsymbol{B}) - h(\boldsymbol{B})}} && \textit{(submodularity)}
\end{aligned}$$

- Any inequality that can be "inferred" from the above is a **Shannon inequality**

Background: Entropic Functions & Polymatroids (cont.)

Entropic functions satisfy:

- (Basic) Shannon inequalities, *and*
- **Non-Shannon inequalities** [Zhang et al. 1997]

$$\underbrace{\Gamma_n^*}_{\text{Entropic functions}} \subsetneq \underbrace{\Gamma_n}_{\text{Polymatroids}}$$

Mapping from $\mathcal{D}$ to h

R

X	Y
1	p
1	q
2	p

S

Y	Z
p	3
q	4
q	5

T

Z	W
3	i
5	i
5	j

U

W	X
i	1
j	1
k	2

Mapping from $\mathcal{D}$ to h

R

X	Y
1	p
1	q
2	p

S

Y	Z
p	3
q	4
q	5

T

Z	W
3	i
5	i
5	j

U

W	X
i	1
j	1
k	2

$Q_{\square}^{\text{full}}$

X	Y	Z	W
1	p	3	i
1	q	5	i
1	q	5	j

Mapping from $\mathcal{D}$ to h

R

X	Y
1	p
1	q
2	p

S

Y	Z
p	3
q	4
q	5

T

Z	W
3	i
5	i
5	j

U

W	X
i	1
j	1
k	2

$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

Mapping from $\mathcal{D}$ to h

R

X	Y	
1	p	1/3
1	q	2/3
2	p	0

S

Y	Z	
p	3	1/3
q	4	0
q	5	2/3

T

Z	W	
3	i	1/3
5	i	1/3
5	j	1/3

U

W	X	
i	1	2/3
j	1	1/3
k	2	0

$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

Mapping from $\mathcal{D}$ to h

R

X	Y	
1	p	1/3
1	q	2/3
2	p	0

S

Y	Z	
p	3	1/3
q	4	0
q	5	2/3

T

Z	W	
3	i	1/3
5	i	1/3
5	j	1/3

U

W	X	
i	1	2/3
j	1	1/3
k	2	0

$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

$$\log_N |Q(\mathcal{D})| = h(XYZW)$$

Mapping from $\mathcal{D}$ to h

R

X	Y	
1	p	1/3
1	q	2/3
2	p	0

S

Y	Z	
p	3	1/3
q	4	0
q	5	2/3

T

Z	W	
3	i	1/3
5	i	1/3
5	j	1/3

U

W	X	
i	1	2/3
j	1	1/3
k	2	0

$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

$$\log_N |Q(\mathcal{D})| = h(XYZW)$$

$\mathcal{D} \models \mathcal{S}$	$h \models \mathcal{S}$
$\lvert R(X,Y)\rvert \le N$	
Similarly $\lvert S\rvert, \lvert T\rvert, \lvert U\rvert \le N$	
$W \to X$	
$\deg_U(W \mid X) \le C$	

Mapping from $\mathcal{D}$ to h

R

X	Y	
1	p	1/3
1	q	2/3
2	p	0

S

Y	Z	
p	3	1/3
q	4	0
q	5	2/3

T

Z	W	
3	i	1/3
5	i	1/3
5	j	1/3

U

W	X	
i	1	2/3
j	1	1/3
k	2	0

$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

$$\log_N |Q(\mathcal{D})| = h(XYZW)$$

$\mathcal{D} \models \mathcal{S}$	$h \models \mathcal{S}$
$\lvert R(X,Y)\rvert \le N$	$h(XY) \le \log_N N$
Similarly $\lvert S\rvert, \lvert T\rvert, \lvert U\rvert \le N$	...
$W \to X$	
$\deg_U(W \mid X) \le C$	

Mapping from $\mathcal{D}$ to h

R

X	Y	
1	p	1/3
1	q	2/3
2	p	0

S

Y	Z	
p	3	1/3
q	4	0
q	5	2/3

T

Z	W	
3	i	1/3
5	i	1/3
5	j	1/3

U

W	X	
i	1	2/3
j	1	1/3
k	2	0

$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

$$\log_N |Q(\mathcal{D})| = h(XYZW)$$

$\mathcal{D} \models \mathcal{S}$	$h \models \mathcal{S}$
$\lvert R(X,Y)\rvert \le N$	$h(XY) \le \log_N N$
Similarly $\lvert S\rvert, \lvert T\rvert, \lvert U\rvert \le N$	...
$W \to X$	$h(X\lvert W) \le 0$
$\deg_U(W\lvert X) \le C$	

Mapping from $\mathcal{D}$ to h

R

X	Y	
1	p	1/3
1	q	2/3
2	p	0

S

Y	Z	
p	3	1/3
q	4	0
q	5	2/3

T

Z	W	
3	i	1/3
5	i	1/3
5	j	1/3

U

W	X	
i	1	2/3
j	1	1/3
k	2	0

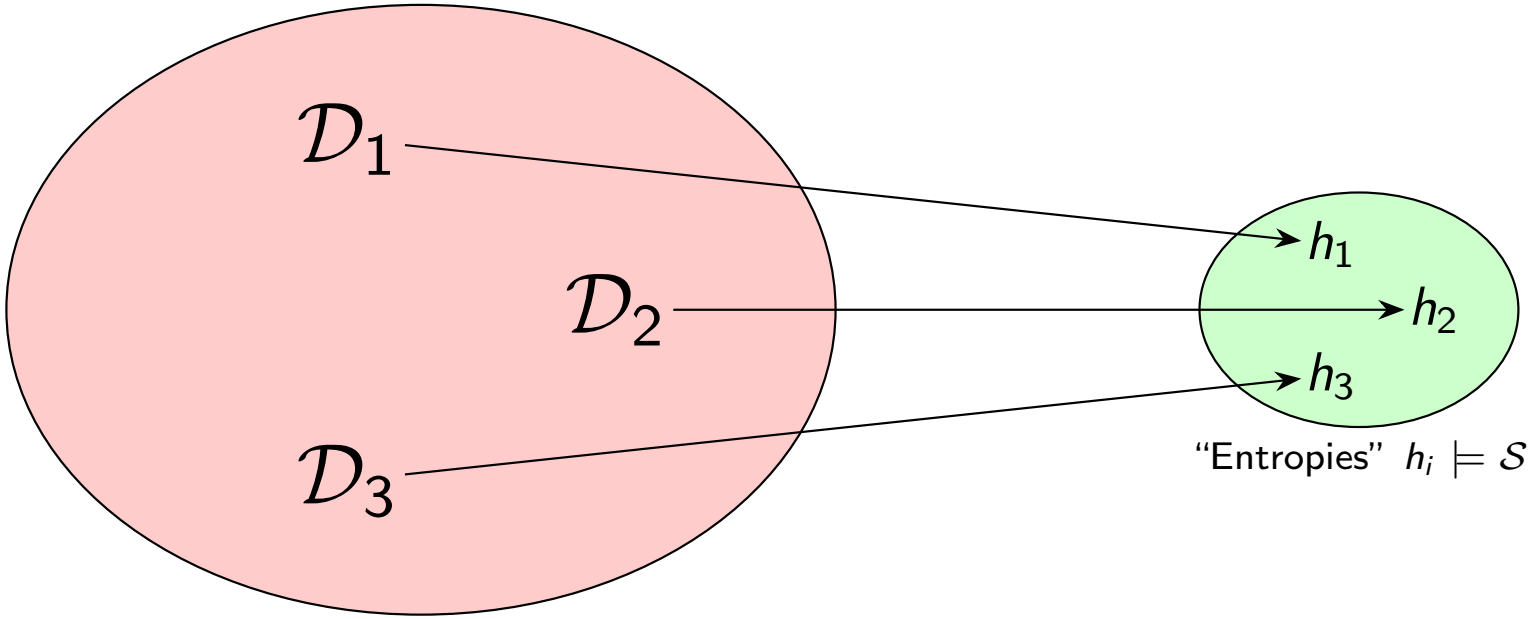
$Q_{\square}^{\text{full}}$

X	Y	Z	W	
1	p	3	i	1/3
1	q	5	i	1/3
1	q	5	j	1/3

$$\log_N |Q(\mathcal{D})| = h(XYZW)$$

$\mathcal{D} \models \mathcal{S}$	$h \models \mathcal{S}$
$\lvert R(X,Y)\rvert \le N$	$h(XY) \le \log_N N$
Similarly $\lvert S\rvert, \lvert T\rvert, \lvert U\rvert \le N$	...
$W \to X$	$h(X\lvert W) \le 0$
$\deg_U(W\lvert X) \le C$	$h(W\lvert X) \le \log_N C$

Entropic Bound vs. Polymatroid Bound

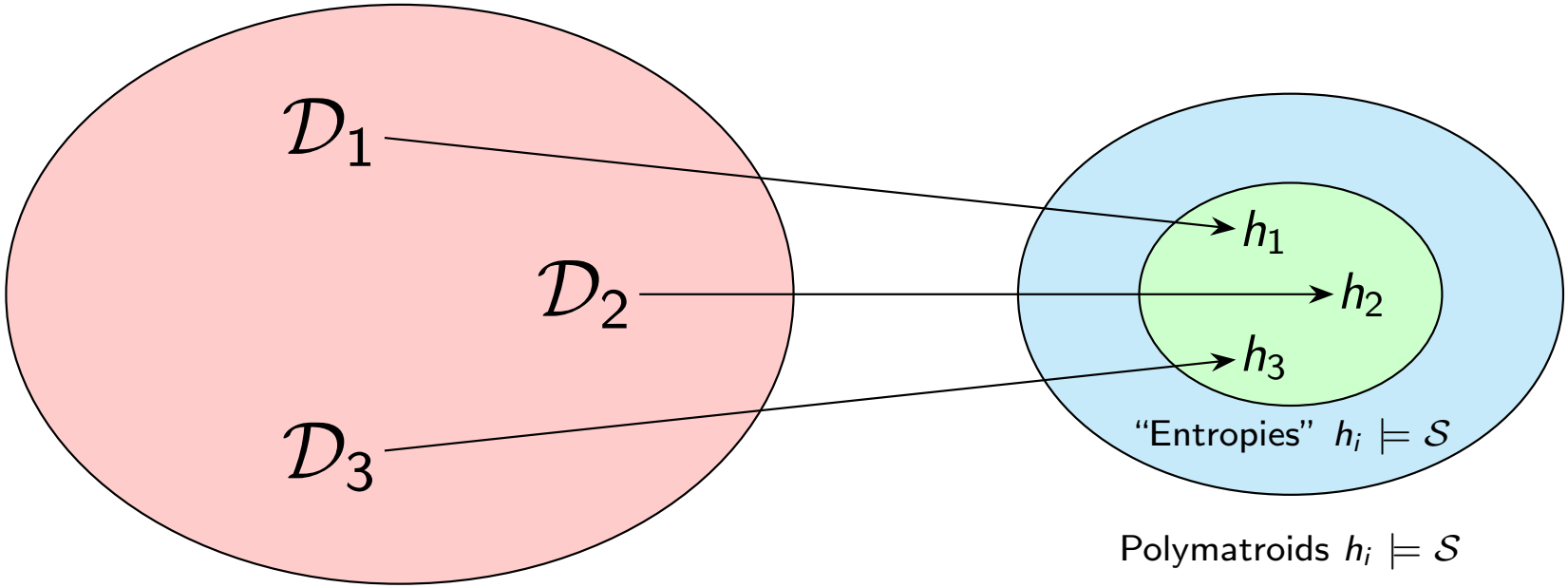


$$\log_N \underbrace{\sup_{\mathcal{D} \models \mathcal{S}} |Q(\mathcal{D})|}_{\textbf{worst-case output size}} \leq \underbrace{\sup_{h \models \mathcal{S}, \Gamma_n^*} h(\boldsymbol{F})}_{\textbf{Entropic-bound}(Q, \mathcal{S})}$$

Tight

Not computable

Entropic Bound vs. Polymatroid Bound

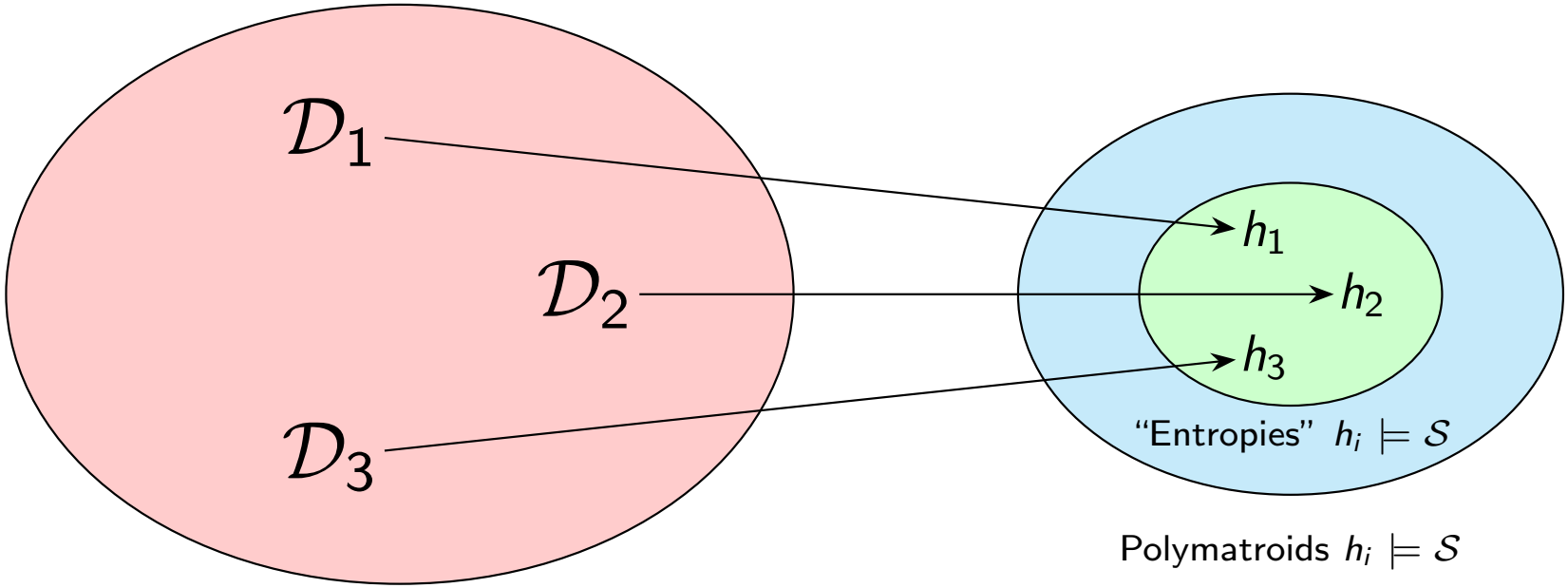


$$\log_N \underbrace{\sup_{\mathcal{D} \models \mathcal{S}} |Q(\mathcal{D})|}_{\textbf{worst-case output size}} \leq \underbrace{\sup_{h \models \mathcal{S}, \Gamma_n^*} h(\boldsymbol{F})}_{\textbf{Entropic-bound}(Q, \mathcal{S})}$$

Tight

Not computable

Entropic Bound vs. Polymatroid Bound



$$\log_N \underbrace{\sup_{\mathcal{D} \models \mathcal{S}} |Q(\mathcal{D})|}_{\textbf{worst-case output size}} \leq \underbrace{\sup_{h \models \mathcal{S}, \Gamma_n^*} h(\boldsymbol{F})}_{\textbf{Entropic-bound}(Q,\mathcal{S})} \leq \underbrace{\max_{h \models \mathcal{S}, \Gamma_n} h(\boldsymbol{F})}_{\textbf{Polymatroid-bound}(Q,\mathcal{S})}$$

Entropic-bound$(Q, \mathcal{S})$
Tight
Not computable

Polymatroid-bound$(Q, \mathcal{S})$
Computable
Tight under CCs
Not tight under FDs

Example on the Polymatroid Bound

$$Q_{\square}^{\text{full}}(X, Y, Z, W) \quad :- \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

$\mathcal{D} \models \mathcal{S}$	$h \models \mathcal{S}$
$\lvert R(X, Y)\rvert \leq N$	$h(XY) \leq \log_N N$
Similarly $\lvert S\rvert, \lvert T\rvert, \lvert U\rvert \leq N$	$\ldots$
$W \to X$	$h(X \mid W) \leq 0$
$\deg_U(W \mid X) \leq C$	$h(W \mid X) \leq \log_N C$

$$\lvert Q_{\square}^{\text{full}}(\mathcal{D})\rvert \quad \leq \quad N^{3/2} \cdot \sqrt{C}$$

Outline

1. Cardinality Estimation
   - Output Size
   - Query Plan Cost
     - Static Plans
     - Adaptive Plans
2. Query Evaluation
3. Conclusion

Outline

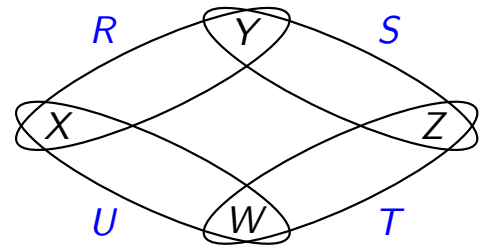
1. Cardinality Estimation
   - Output Size
   - Query Plan Cost
     - Static Plans
     - Adaptive Plans
2. Query Evaluation
3. Conclusion

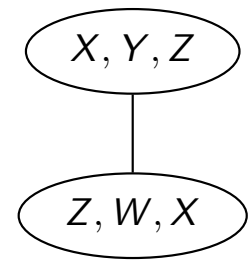
Static Query Plans: Tree Decompositions (TDs)

- Grew out of the CSP and PGM communities
- Widely adopted in DB Theory
  - (Generalized) hypertree decompositions [Gottlob, Leone, et al. 2002; Gottlob, Greco, et al. 2016]
  - Fractional hypertree decompositions [Grohe et al. 2014]
  - *Free-connex* tree decompositions [Bagan et al. 2007]

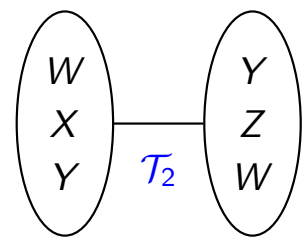
Static Query Plans: Tree Decompositions (TDs)

$$Q_{\square}(X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$





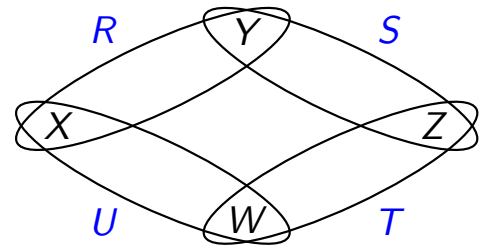
$\mathcal{T}_1$



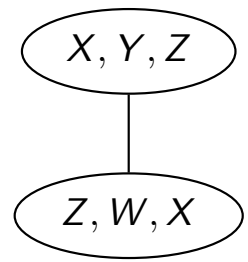
$\mathcal{T}_2$

Static Query Plans: Tree Decompositions (TDs)

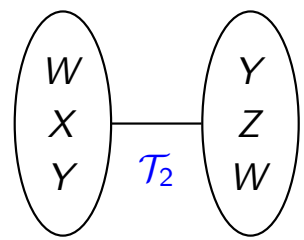
$$Q_{\square}(X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$



- Take a free-connex TD
- Apply the polymatroid bound to every bag
- Result: **Fractional hypertree width**



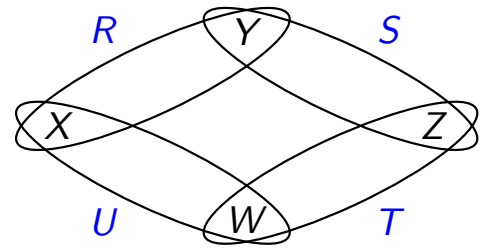
$\mathcal{T}_1$



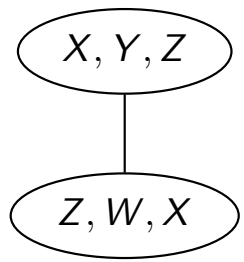
$\mathcal{T}_2$

Static Query Plans: Tree Decompositions (TDs)

$$Q_\square(X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

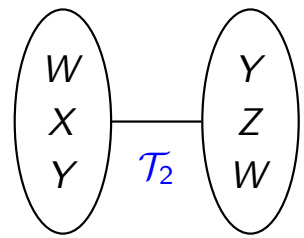


- Take a free-connex TD
- Apply the polymatroid bound to every bag
- Result: **Fractional hypertree width**



$\mathcal{T}_1$

$$\text{fhtw}(Q, \mathcal{S}) \stackrel{\text{def}}{=} \underbrace{\min_{\mathcal{T} \in \text{TD}(Q)}}_{\substack{\text{best static} \\ \text{query plan}}} \underbrace{\max_{\boldsymbol{B} \in \text{bags}(\mathcal{T})}}_{\substack{\text{worst subquery} \\ \text{in the plan}}} \underbrace{\max_{h \models \mathcal{S}, \Gamma_n} h(\boldsymbol{B})}_{\substack{\text{polymatroid bound} \\ \text{of the subquery}}}$$

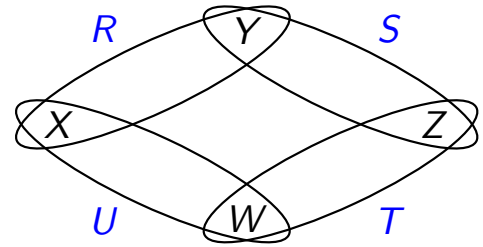


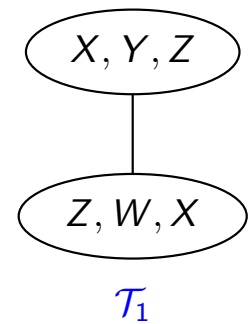
$\mathcal{T}_2$

- Generalizes [Grohe et al. 2014]

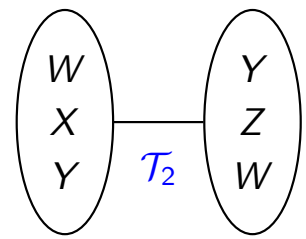
Static Query Plans: Tree Decompositions (TDs)

$$\mathsf{fhtw}(Q_\square, \mathcal{S}_\square) = \min(\max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(XYZ), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(ZWX))$$
$$\max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(YZW), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(WXY)))$$





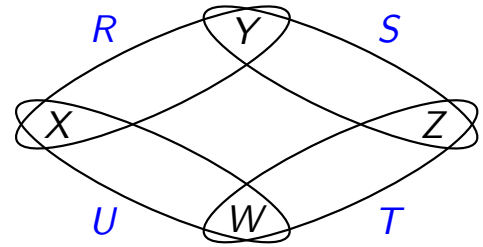
$\mathcal{T}_1$



$\mathcal{T}_2$

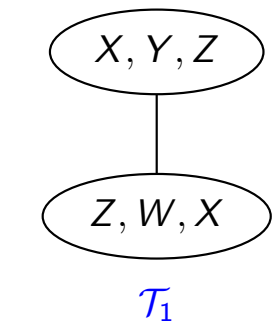
Static Query Plans: Tree Decompositions (TDs)

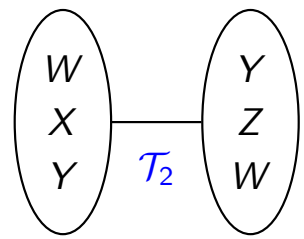
$$\text{fhtw}(Q_\square, \mathcal{S}_\square) = \min(\max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(XYZ), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(ZWX))$$
$$\max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(YZW), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(WXY)))$$



If $\mathcal{S}_\square$ contains only $|R|, |S|, |T|, |U| \leq N$,

$$\text{fhtw}(Q_\square, \mathcal{S}_\square) = 2$$



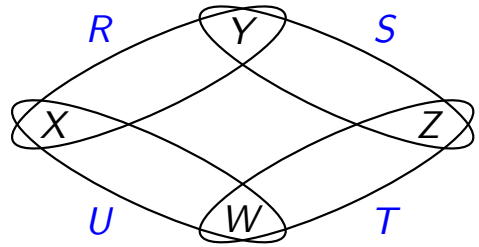


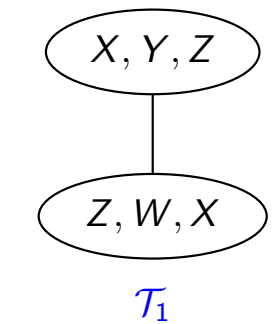
Outline

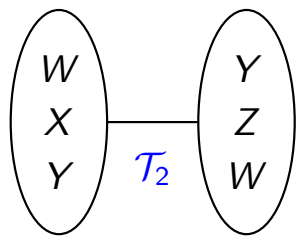
1. Cardinality Estimation
   - Output Size
   - Query Plan Cost
     - Static Plans
     - Adaptive Plans
2. Query Evaluation
3. Conclusion

Adaptive Query Plans

- $\text{fhtw}(Q_\square, \mathcal{S}_\square) = 2$
- But $Q_\square$ admits an $O(N^{3/2})$ algorithm [Alon et al. 1997]
  - *Partition* relations, e.g. $\deg_R(Y|X = x) \leq \sqrt{N}$
  - Pick a *different TD* for each part

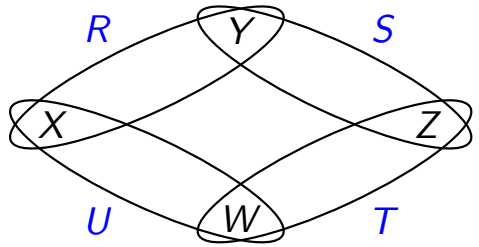


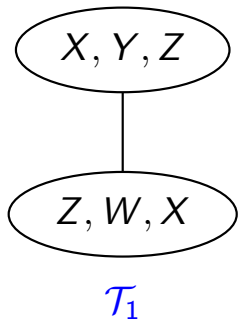


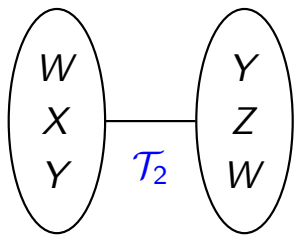


Adaptive Query Plans

- $\text{fhtw}(Q_\square, \mathcal{S}_\square) = 2$
- But $Q_\square$ admits an $O(N^{3/2})$ algorithm [Alon et al. 1997]
  - *Partition* relations, e.g. $\deg_R(Y|X = x) \leq \sqrt{N}$
  - Pick a *different TD* for each part
- We call such algorithms **adaptive** query plans
- But how can we *capture their best cost*?

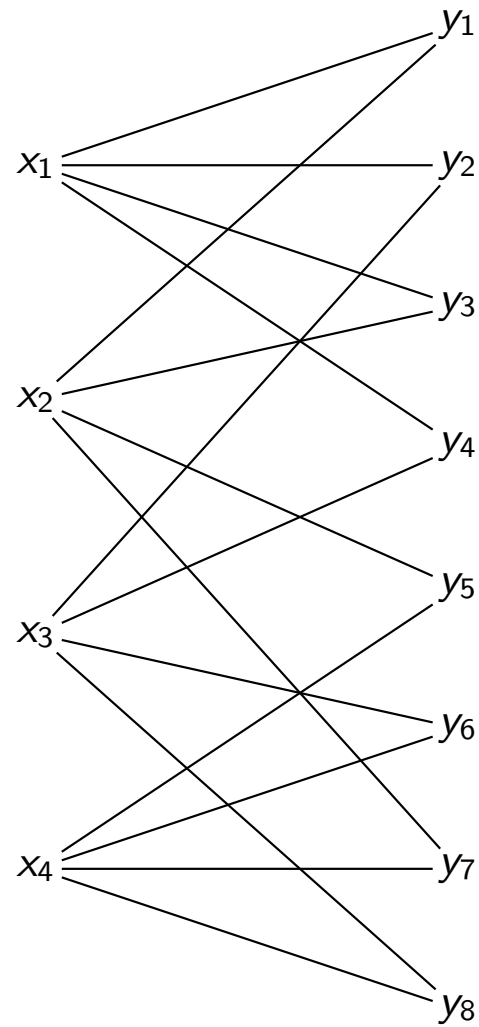






Cardinality Estimation of Adaptive Query Plans

- Partition the input instance into **uniform** instances

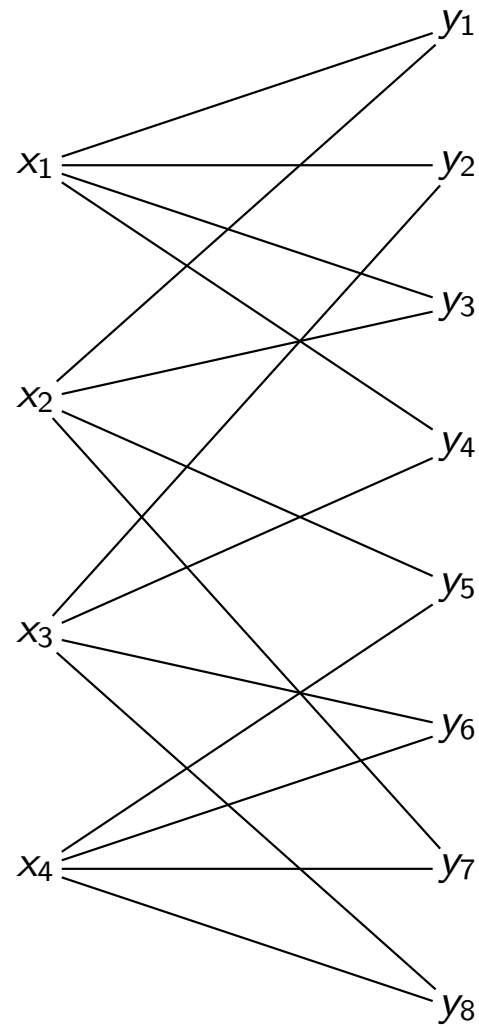


A **uniform** $R(X, Y)$
with 16 tuples

Cardinality Estimation of Adaptive Query Plans

- Partition the input instance into **uniform** instances
- *Uniform instance* $R \Leftrightarrow$ *Entropic function* h

$R(X, Y)$	h
$\lvert R\rvert = 16$	$h(XY) = 4$ bits
$\lvert\pi_X R\rvert = 4$	$h(X) = 2$ bits
$\lvert\pi_Y R\rvert = 8$	$h(Y) = 3$ bits
$\deg_R(Y\lvert X) = 4$	$h(Y\lvert X) = 2$ bits
$\deg_R(X\lvert Y) = 2$	$h(X\lvert Y) = 1$ bit



A **uniform** $R(X, Y)$ with 16 tuples

Cardinality Estimation of Adaptive Query Plans

- *Lower bound*: generalizes [Marx 2013]

$$\mathsf{subw}(Q,\mathcal{S}) \stackrel{\text{def}}{=} \underbrace{\max_{h \models \mathcal{S},\Gamma_n}}_{\text{worst } \textit{uniform} \text{ instance}} \underbrace{\min_{\mathcal{T}\in\mathsf{TD}(Q)}}_{\text{best static query plan}} \underbrace{\max_{\boldsymbol{B}\in\mathsf{bags}(\mathcal{T})}}_{\text{worst subquery in the plan}} \underbrace{h(\boldsymbol{B})}_{\text{subquery cost}}$$

Cardinality Estimation of Adaptive Query Plans

- *Lower bound*: generalizes [Marx 2013]

$$\mathsf{subw}(Q,\mathcal{S}) \stackrel{\text{def}}{=} \underbrace{\max_{h\models\mathcal{S},\Gamma_n}}_{\text{worst } \textit{uniform} \text{ instance}} \underbrace{\min_{\mathcal{T}\in\mathsf{TD}(Q)}}_{\text{best static query plan}} \underbrace{\max_{\boldsymbol{B}\in\mathsf{bags}(\mathcal{T})}}_{\text{worst subquery in the plan}} \underbrace{h(\boldsymbol{B})}_{\text{subquery cost}}$$

$$\mathsf{fhtw}(Q,\mathcal{S}) \stackrel{\text{def}}{=} \underbrace{\min_{\mathcal{T}\in\mathsf{TD}(Q)}}_{\text{best static query plan}} \underbrace{\max_{\boldsymbol{B}\in\mathsf{bags}(\mathcal{T})}}_{\text{worst subquery in the plan}} \underbrace{\max_{h\models\mathcal{S},\Gamma_n} h(\boldsymbol{B})}_{\text{polymatroid bound of the subquery}}$$

Cardinality Estimation of Adaptive Query Plans

- *Lower bound*: generalizes [Marx 2013]

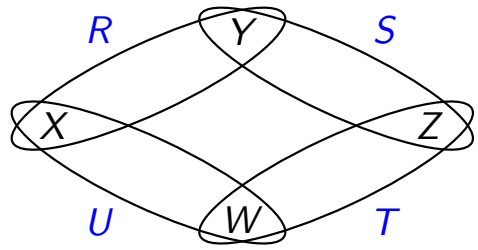
$$\mathsf{subw}(Q,\mathcal{S}) \stackrel{\text{def}}{=} \underbrace{\max_{h\models\mathcal{S},\Gamma_n}}_{\text{worst } \textit{uniform} \text{ instance}} \underbrace{\min_{\mathcal{T}\in\mathsf{TD}(Q)}}_{\text{best static query plan}} \underbrace{\max_{\boldsymbol{B}\in\mathsf{bags}(\mathcal{T})}}_{\text{worst subquery in the plan}} \underbrace{h(\boldsymbol{B})}_{\text{subquery cost}}$$

$$\mathsf{fhtw}(Q,\mathcal{S}) \stackrel{\text{def}}{=} \underbrace{\min_{\mathcal{T}\in\mathsf{TD}(Q)}}_{\text{best static query plan}} \underbrace{\max_{\boldsymbol{B}\in\mathsf{bags}(\mathcal{T})}}_{\text{worst subquery in the plan}} \underbrace{\max_{h\models\mathcal{S},\Gamma_n} h(\boldsymbol{B})}_{\text{polymatroid bound of the subquery}}$$

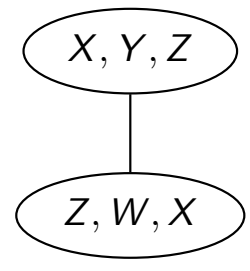
- PANDA runs in $O(N^{\mathsf{subw}(Q,\mathcal{S})} \cdot \log N + \mathrm{OUT})$ for any $Q, \mathcal{S}$

Example

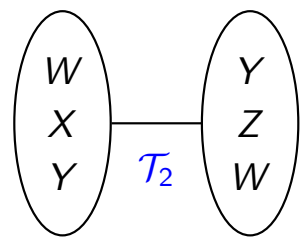
$$\mathsf{subw}(Q_\square, \mathcal{S}_\square) = \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(\max(h(XYZ), h(ZWX)), \max(h(YZW), h(WXY)))$$



$$\mathsf{fhtw}(Q_\square, \mathcal{S}_\square) = \min(\max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(XYZ), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(ZWX)), \max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(YZW), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(WXY)))$$



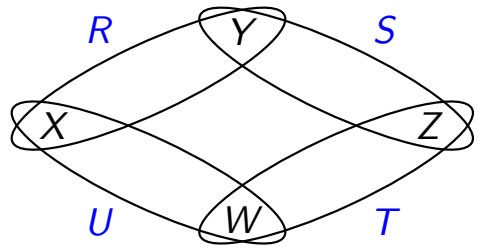
$\mathcal{T}_1$



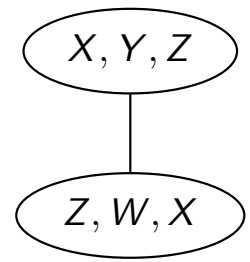
$\mathcal{T}_2$

Example

$$\text{subw}(Q_\square, \mathcal{S}_\square) = \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(\max(h(XYZ), h(ZWX)), \max(h(YZW), h(WXY)))$$



$$\text{fhtw}(Q_\square, \mathcal{S}_\square) = \min(\max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(XYZ), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(ZWX)), \max(\max_{h \models \mathcal{S}_\square, \Gamma_4} h(YZW), \max_{h \models \mathcal{S}_\square, \Gamma_4} h(WXY)))$$

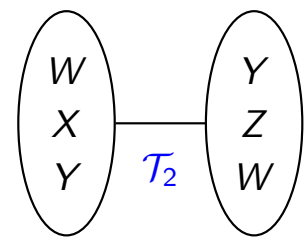


$\mathcal{T}_1$

If $\mathcal{S}_\square$ contains only $|R|, |S|, |T|, |U| \leq N$,

$$\text{subw}(Q_\square, \mathcal{S}_\square) = 3/2$$

$$\text{fhtw}(Q_\square, \mathcal{S}_\square) = 2$$



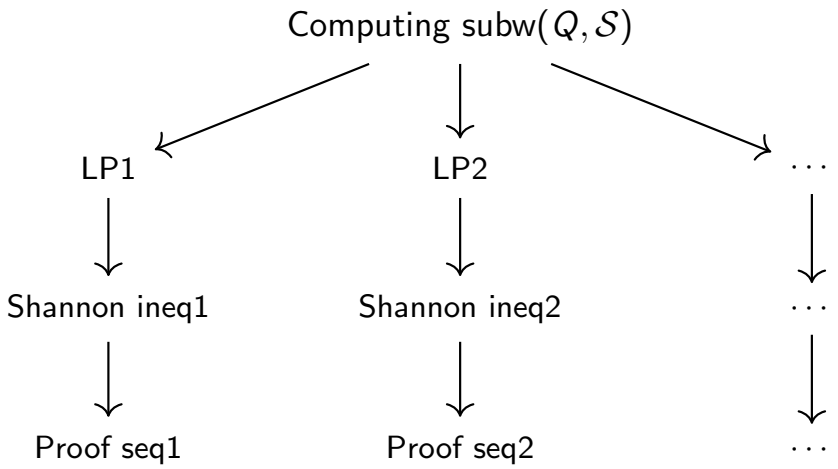
$\mathcal{T}_2$

Outline

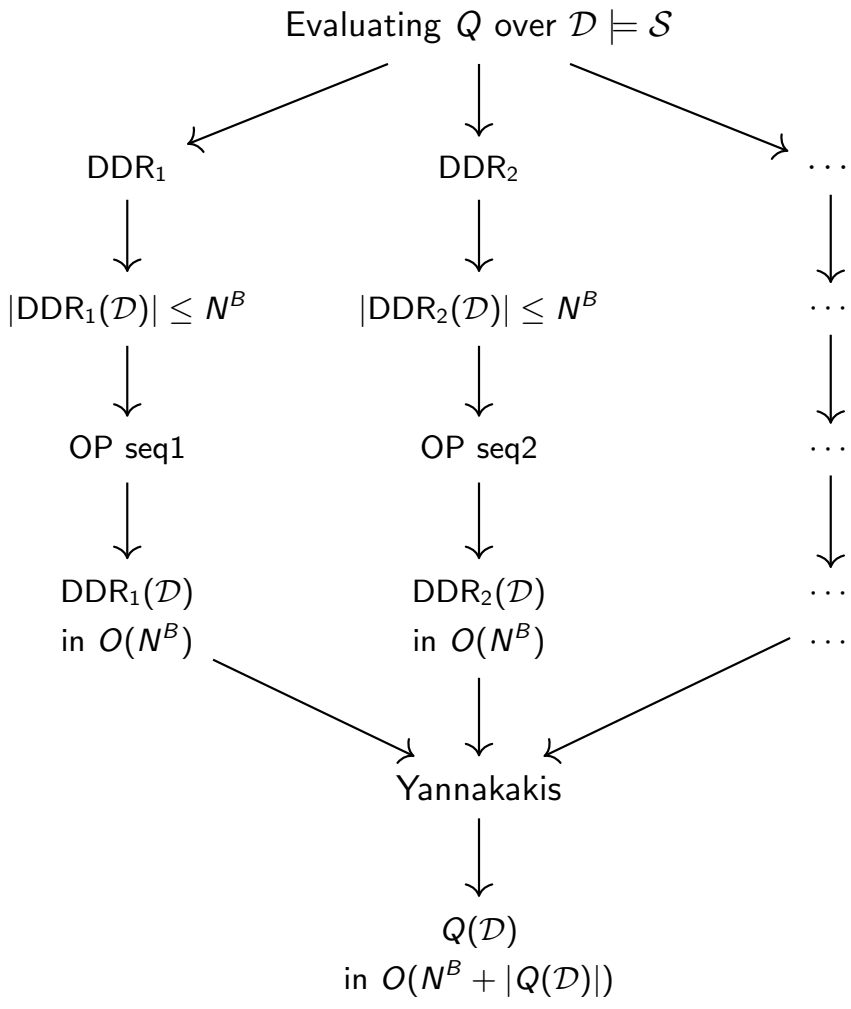
1. Cardinality Estimation
2. Query Evaluation
3. Conclusion

PANDA

Proof that subw$(Q, \mathcal{S}) \leq B$

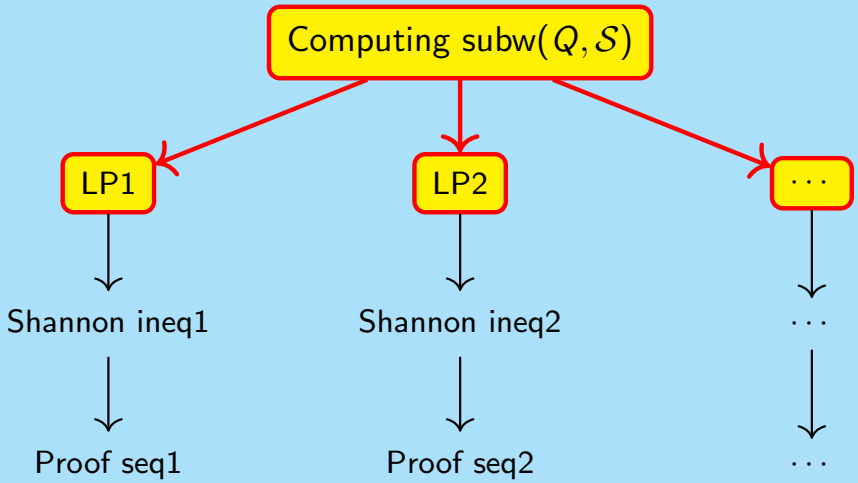


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$

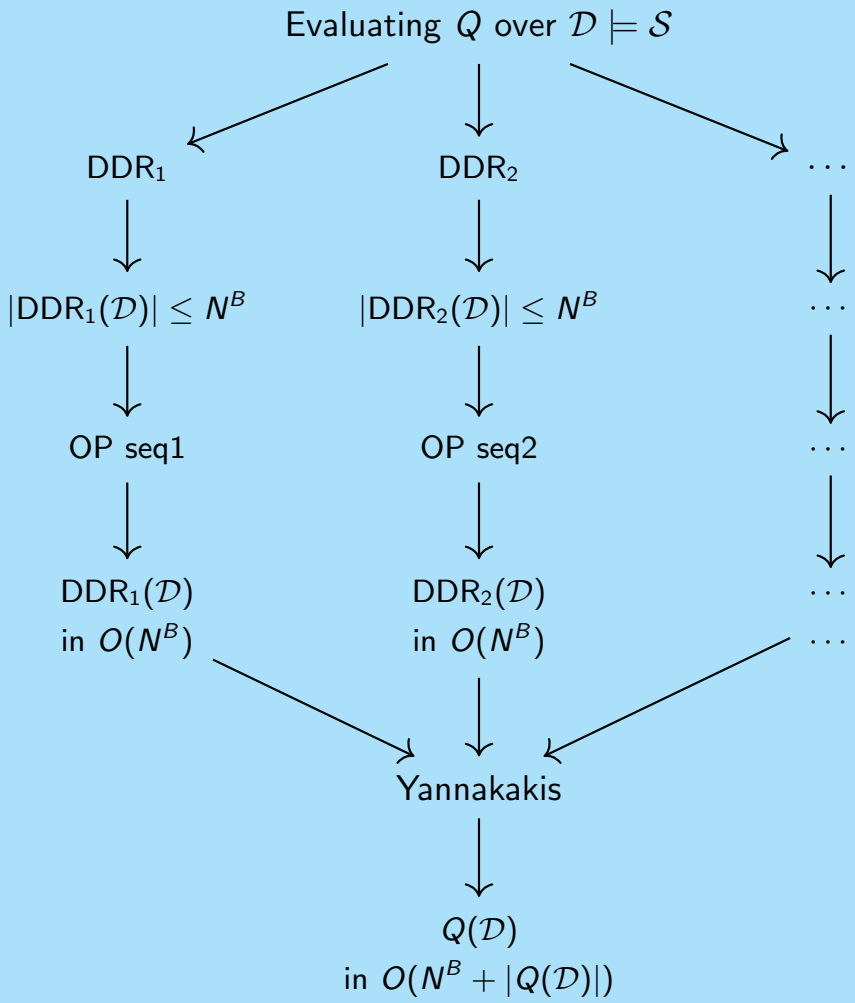


PANDA

Proof that $\text{subw}(Q, \mathcal{S}) \leq B$

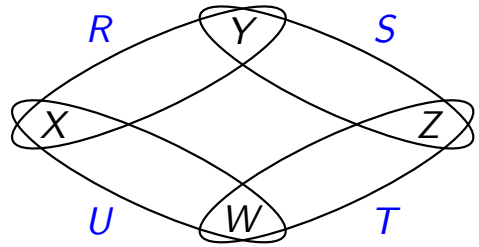


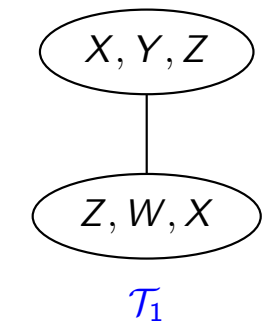
Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



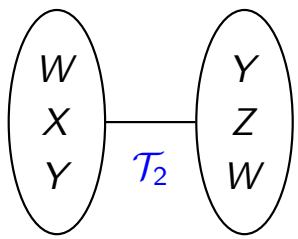
Computing subw$(Q, \mathcal{S})$

$$\mathsf{subw}(Q_\square, \mathcal{S}_\square) = \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(\max(h(XYZ), h(ZWX)), \max(h(YZW), h(WXY)))$$





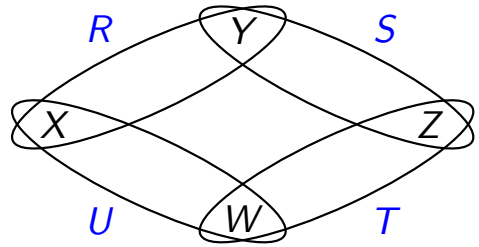
$\mathcal{T}_1$

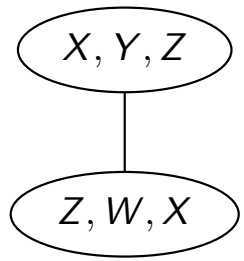


$\mathcal{T}_2$

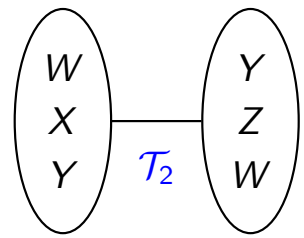
Computing subw$(Q, \mathcal{S})$

$$\mathsf{subw}(Q_\square, \mathcal{S}_\square) = \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(\max(h(XYZ), h(ZWX)),$$
$$\max(h(YZW), h(WXY)))$$
$$= \max\Bigg(\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW)),$$
$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(WXY)),$$
$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(ZWX), h(YZW)),$$
$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(ZWX), h(WXY)) \Bigg)$$





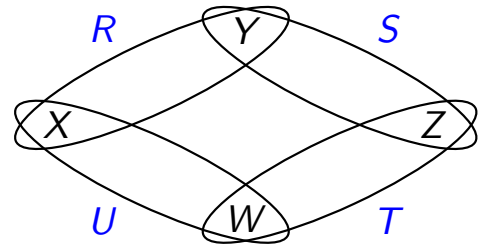
$\mathcal{T}_1$

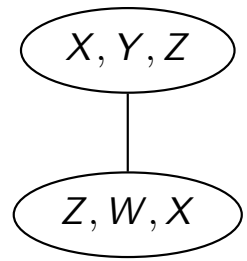


$\mathcal{T}_2$

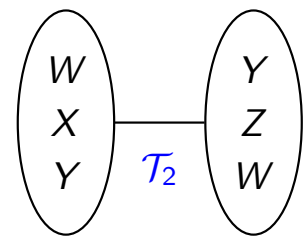
Computing subw$(Q, \mathcal{S})$ $\Rightarrow$ LP1, LP2, ...

$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW))$$





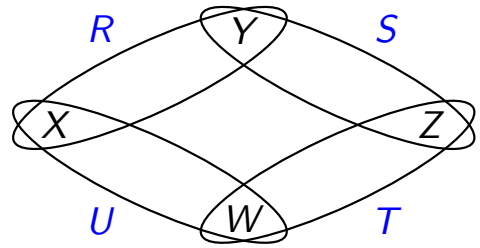
$\mathcal{T}_1$



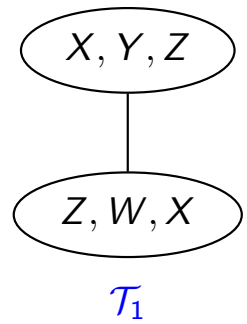
$\mathcal{T}_2$

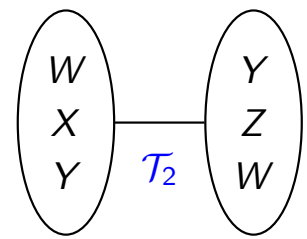
Computing subw$(Q, \mathcal{S})$ $\Rightarrow$ LP1, LP2, ...

$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW))$$



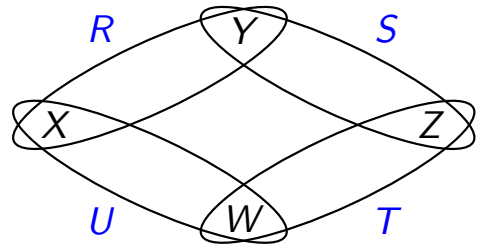
$$\begin{aligned}
\max \quad & t \\
\text{s.t.} \quad & t \le h(XYZ), \quad t \le h(YZW), && (2)\\
& h(XY), h(YZ), h(ZW), h(WX) \le 1, && (3)\\
& \underbrace{h \in \Gamma_4} && (4)
\end{aligned}$$



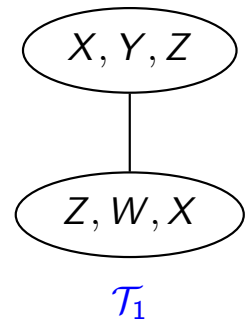


Computing subw$(Q, \mathcal{S})$ $\Rightarrow$ LP1, LP2, ...

$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW))$$



$$\begin{aligned}
\max \quad & t \\
\text{s.t.} \quad & t \le h(XYZ), \quad t \le h(YZW), && (2)\\
& h(XY), h(YZ), h(ZW), h(WX) \le 1, && (3)\\
& \underbrace{h \in \Gamma_4} && (4)
\end{aligned}$$



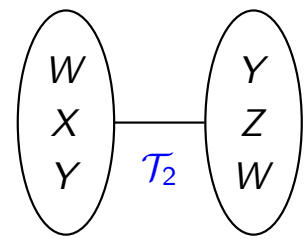
$h(\emptyset) = 0$

$h(Y) \le h(XY)$

...

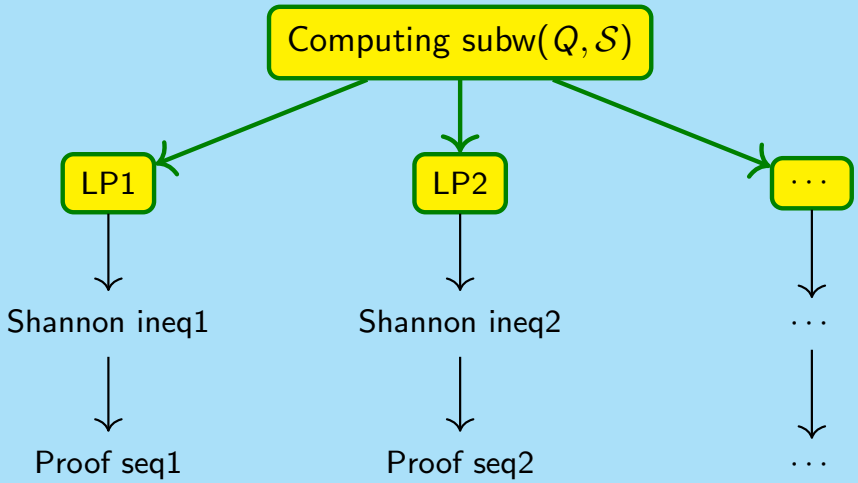
$h(XY) + h(YZ) \ge h(XXYZ) + h(Y)$

...

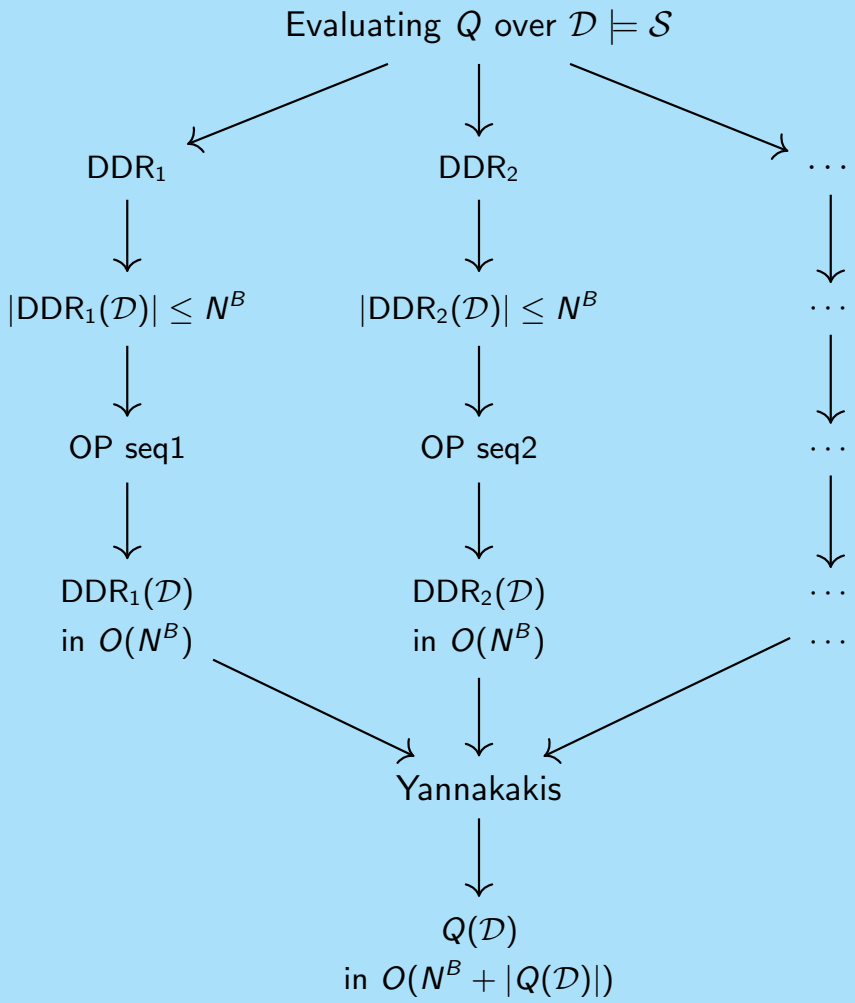


PANDA

Proof that subw$(Q, \mathcal{S}) \leq B$

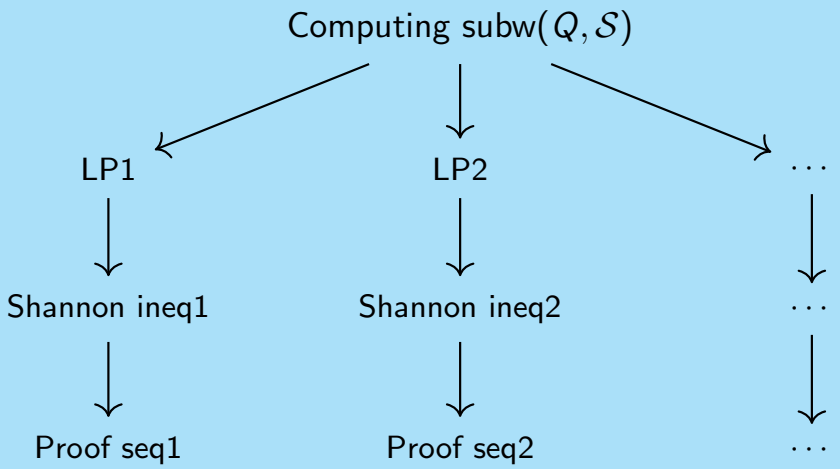


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$

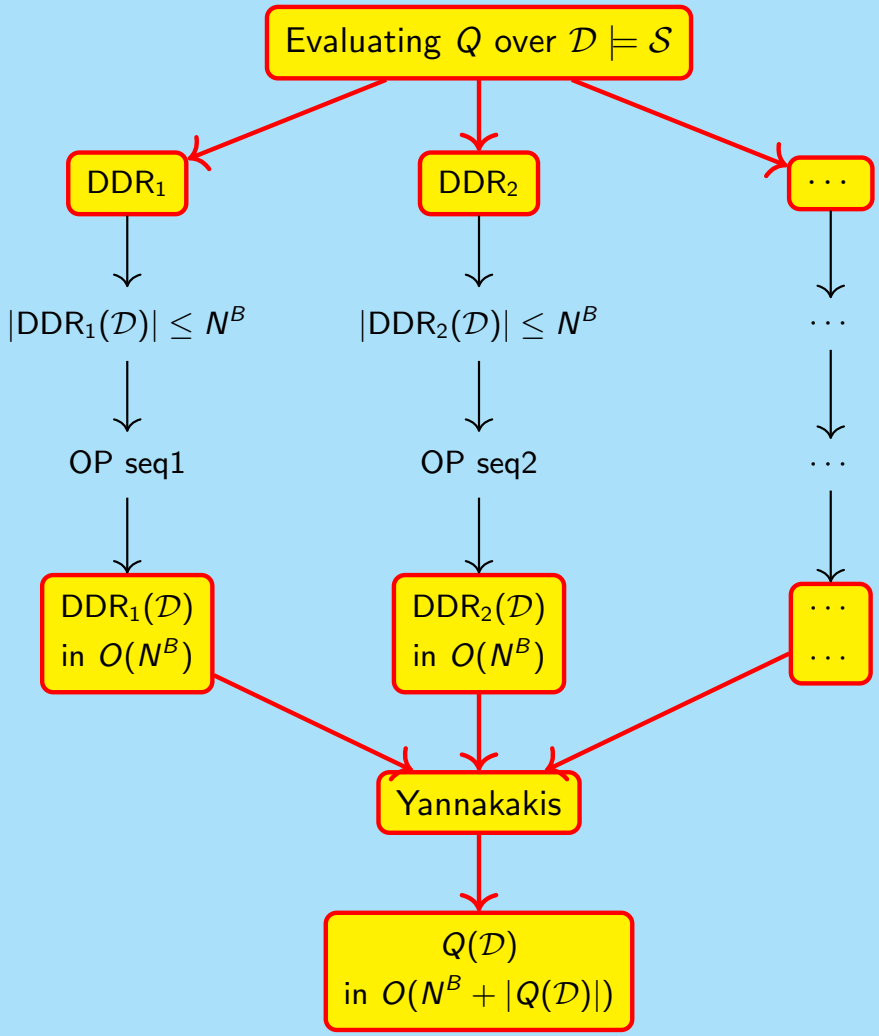


PANDA

Proof that subw$(Q, \mathcal{S}) \leq B$

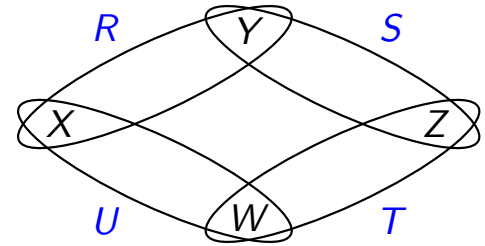


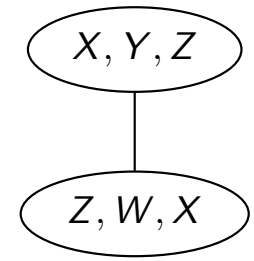
Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



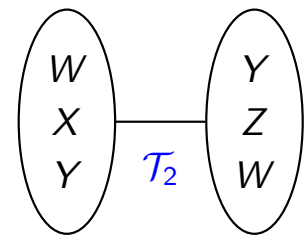
Evaluating Q over $\mathcal{D} \models \mathcal{S}$

$$Q_{\square}(X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$





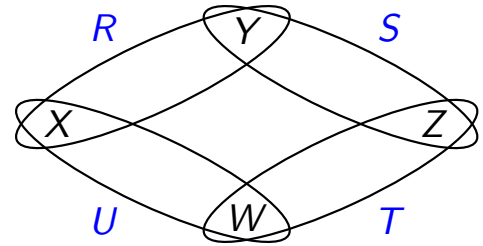
$\mathcal{T}_1$



$\mathcal{T}_2$

Evaluating Q over $\mathcal{D} \models \mathcal{S}$

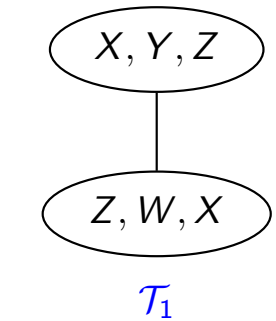
$$Q_{\square}(X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

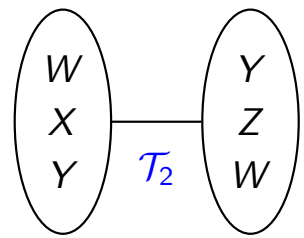


1. **Solve**

$$\big(A_{11}(X, Y, Z) \wedge A_{12}(Z, W, X)\big) \vee \big(A_{21}(Y, Z, W) \wedge A_{22}(W, X, Y)\big)$$

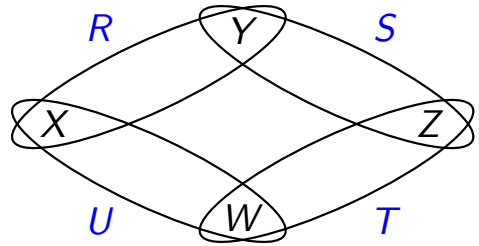
$$\text{:-} \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$





Evaluating Q over $\mathcal{D} \models \mathcal{S}$

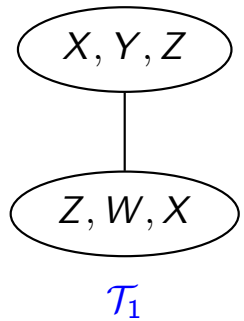
$$Q_{\square}(X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$



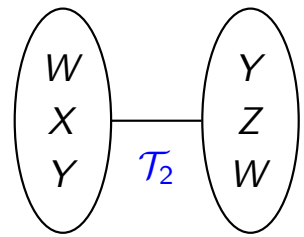
1. **Solve**

$$\big(A_{11}(X, Y, Z) \wedge A_{12}(Z, W, X)\big) \vee \big(A_{21}(Y, Z, W) \wedge A_{22}(W, X, Y)\big)$$

$$\text{:-} \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$



2. **Semi-join reduce** $A_{11}, A_{12}, A_{21}, A_{22}$ with R, S, T, U
3. **Run Yannakakis**



$$Q^{(1)}_{\square}(X, Y) \stackrel{\text{def}}{=} A_{11}(X, Y, Z) \wedge A_{12}(Z, W, X)$$

$$Q^{(2)}_{\square}(X, Y) \stackrel{\text{def}}{=} A_{21}(Y, Z, W) \wedge A_{22}(W, X, Y)$$

Evaluating Q over $\mathcal{D} \models \mathcal{S}$

$$(A_{11}(X, Y, Z) \wedge A_{12}(Z, W, X)) \vee (A_{21}(Y, Z, W) \wedge A_{22}(W, X, Y))$$

$$\text{:-} \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

Evaluating Q over $\mathcal{D} \models \mathcal{S}$

$$(A_{11}(X, Y, Z) \wedge A_{12}(Z, W, X)) \vee (A_{21}(Y, Z, W) \wedge A_{22}(W, X, Y))$$

$$\text{:-} \quad R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

$$A_{11}(X, Y, Z) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

$$A_{11}(X, Y, Z) \vee A_{22}(W, X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

$$A_{12}(Z, W, X) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

$$A_{12}(Z, W, X) \vee A_{22}(W, X, Y) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

Evaluating Q over $\mathcal{D} \models \mathcal{S}$

$$(A_{11}(X,Y,Z) \wedge A_{12}(Z,W,X)) \vee (A_{21}(Y,Z,W) \wedge A_{22}(W,X,Y))$$

$$\text{:-} \quad R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$A_{11}(X,Y,Z) \vee A_{22}(W,X,Y) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

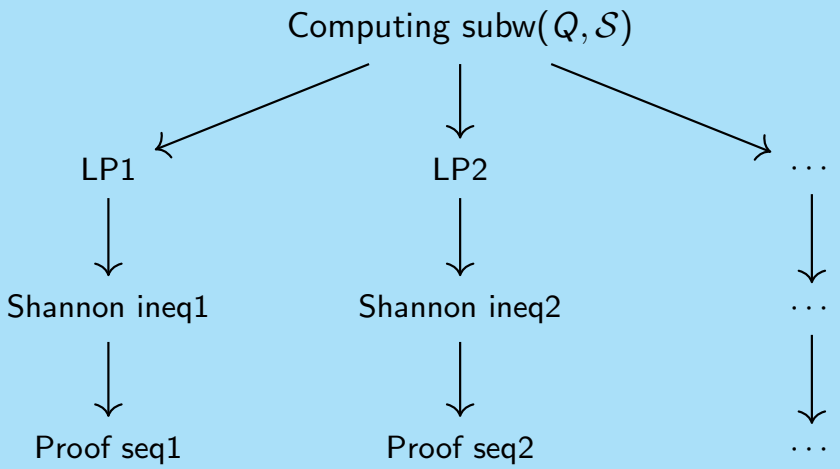
$$A_{12}(Z,W,X) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$A_{12}(Z,W,X) \vee A_{22}(W,X,Y) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

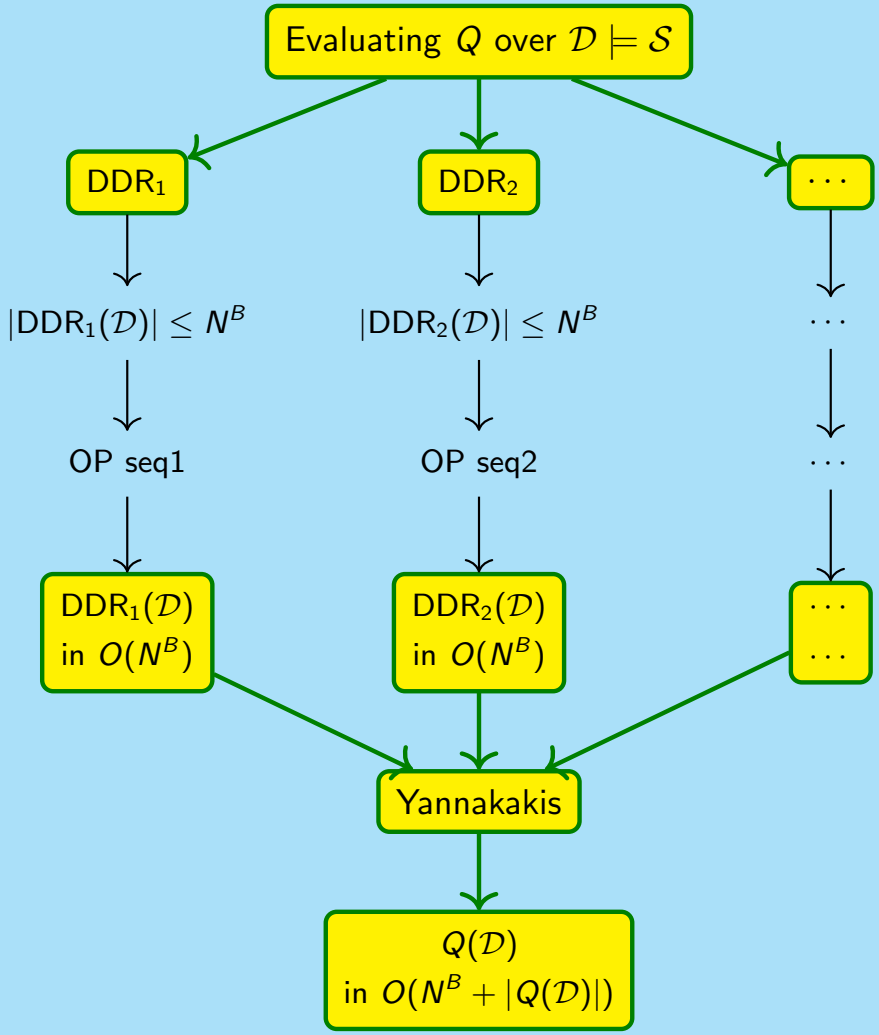
$$\begin{aligned} \text{subw}(Q_\square, \mathcal{S}_\square) = \max\Bigg(& \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW)), \\ & \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(WXY)), \\ & \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(ZWX), h(YZW)), \\ & \max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(ZWX), h(WXY)) \Bigg) \end{aligned}$$

PANDA

Proof that subw$(Q, \mathcal{S}) \leq B$

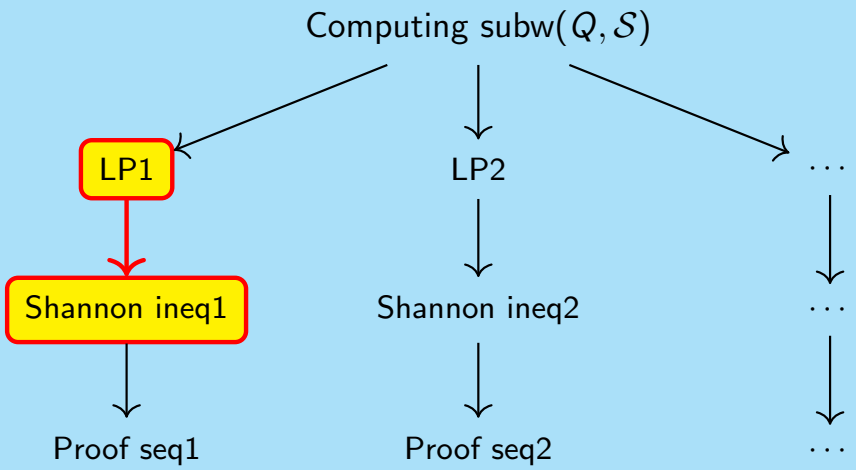


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$

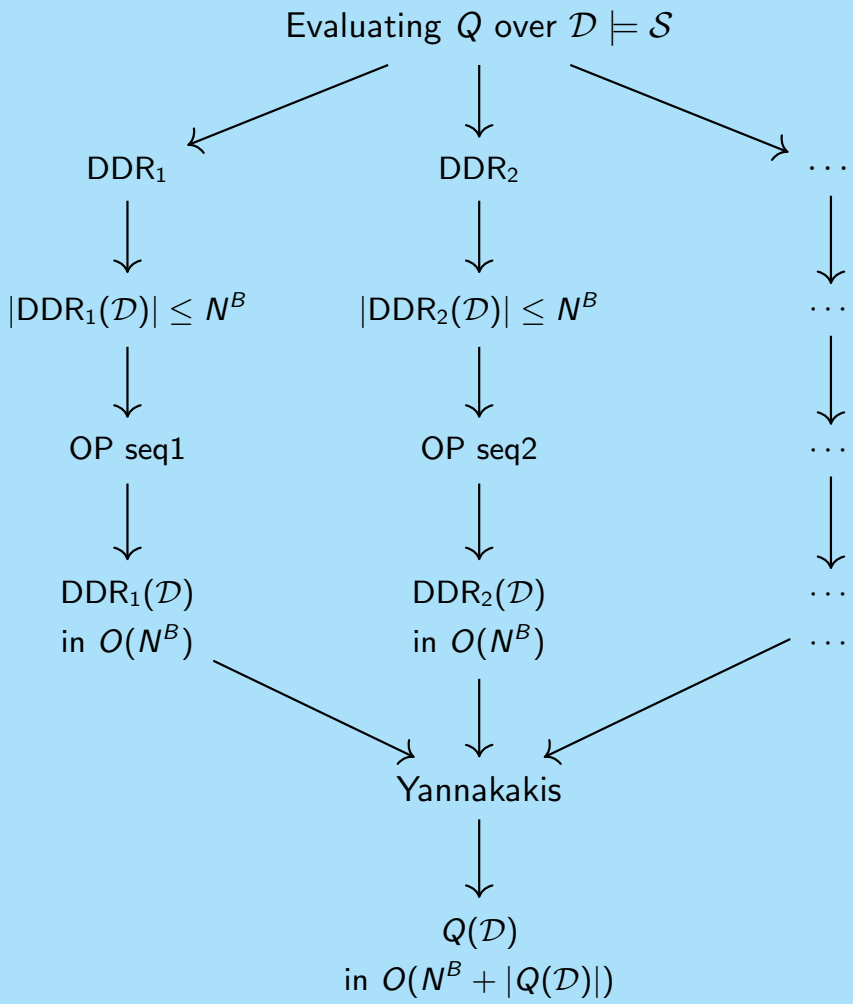


PANDA

Proof that $\text{subw}(Q, \mathcal{S}) \leq B$



Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



LP $\Rightarrow$ Shannon inequality

$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW))$$

Primal LP

$$\begin{aligned}
\max \quad & t \\
\text{s.t.} \quad & t \le h(XYZ), \quad t \le h(YZW), && (2)\\
& h(XY), h(YZ), h(ZW), h(WX) \le 1, && (3)\\
& h \in \Gamma_4 && (4)
\end{aligned}$$

LP $\Rightarrow$ Shannon inequality

$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW))$$

Primal LP

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & t \le h(XYZ), \quad t \le h(YZW), && (2) \\ & h(XY), h(YZ), h(ZW), h(WX) \le 1, && (3) \\ & h \in \Gamma_4 && (4) \end{aligned}$$

Dual LP

- Feasible Space: **Inequalities** $t \le C$ that can be proved from (2)–(4)
- Goal: **Minimize** C

LP $\Rightarrow$ Shannon inequality

$$\max_{h \models \mathcal{S}_\square, \Gamma_4} \min(h(XYZ), h(YZW))$$

Primal LP

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & t \le h(XYZ), \quad t \le h(YZW), && (2) \\ & h(XY), h(YZ), h(ZW), h(WX) \le 1, && (3) \\ & h \in \Gamma_4 && (4) \end{aligned}$$

Dual LP

- Feasible Space: **Inequalities** $t \le C$ that can be proved from (2)–(4)
- Goal: **Minimize** C

But how can we prove such an inequality?

Proving an inequality $t \leq C$

$$
\begin{aligned}
\max \quad & t \\
\text{s.t.} \quad & t \leq h(XYZ), \quad t \leq h(YZW), && (2)\\
& h(XY), h(YZ), h(ZW), h(WX) \leq 1, && (3)\\
& h \in \Gamma_4 && (4)
\end{aligned}
$$

Proving an inequality $t \leq C$

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & t \leq h(XYZ), \quad t \leq h(YZW), && (2) \\ & h(XY), h(YZ), h(ZW), h(WX) \leq 1, && (3) \\ & h \in \Gamma_4 && (4) \end{aligned}$$

- $t \leq \cdots$ using (2)

$$t \leq \lambda_1 h(XYZ) + \lambda_2 h(YZW), \qquad \lambda_1 + \lambda_2 \geq 1$$

Proving an inequality $t \leq C$

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & t \leq h(XYZ), \quad t \leq h(YZW), && (2) \\ & h(XY), h(YZ), h(ZW), h(WX) \leq 1, && (3) \\ & h \in \Gamma_4 && (4) \end{aligned}$$

- $t \leq \cdots$ using (2)

$$t \leq \lambda_1 h(XYZ) + \lambda_2 h(YZW), \qquad \lambda_1 + \lambda_2 \geq 1$$

- $\cdots \leq C$ using (3)

$$w_1 h(XY) + w_2 h(YZ) + w_3 h(ZW) + w_4 h(WX) \leq w_1 + w_2 + w_3 + w_4$$

Proving an inequality $t \le C$

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & t \le h(XYZ), \quad t \le h(YZW), && (2) \\ & h(XY), h(YZ), h(ZW), h(WX) \le 1, && (3) \\ & h \in \Gamma_4 && (4) \end{aligned}$$

- $t \le \cdots$ using (2)

$$t \le \lambda_1 h(XYZ) + \lambda_2 h(YZW), \qquad \lambda_1 + \lambda_2 \ge 1$$

- $\cdots \le C$ using (3)

$$w_1 h(XY) + w_2 h(YZ) + w_3 h(ZW) + w_4 h(WX) \le w_1 + w_2 + w_3 + w_4$$

- Stitch using (4)

$$\underbrace{t \le}_{(2)} \underbrace{\underbrace{\lambda_1 h(XYZ) + \lambda_2 h(YZW) \le w_1 h(XY) + w_2 h(YZ) + w_3 h(ZW) + w_4 h(WX)}_{\textbf{Shannon inequality } (4)} \le w_1 + w_2 + w_3 + w_4}_{(3)}$$

Optimal inequality $t \leq 3/2$

$$
\begin{aligned}
\max \quad & t \\
\text{s.t.} \quad & t \leq h(XYZ), \quad t \leq h(YZW), && (2)\\
& h(XY), h(YZ), h(ZW), h(WX) \leq 1, && (3)\\
& h \in \Gamma_4 && (4)
\end{aligned}
$$

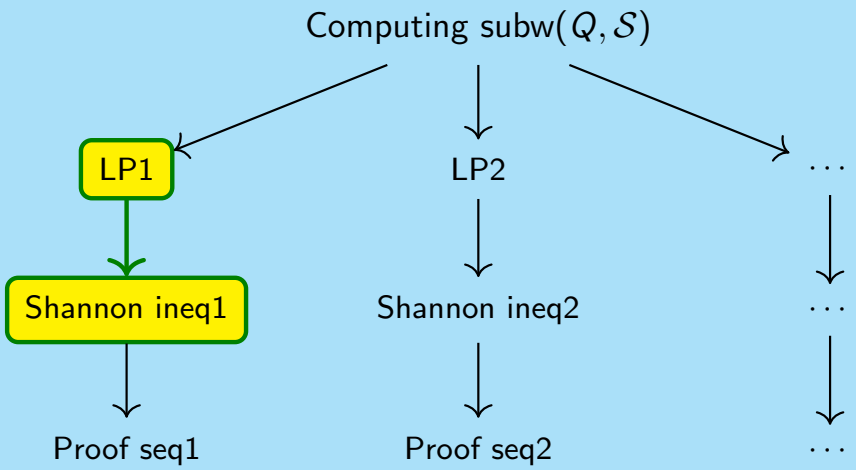
$$
t \leq \underbrace{\frac{1}{2}h(XYZ) + \frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY) + \frac{1}{2}h(YZ) + \frac{1}{2}h(ZW)}_{\textbf{Shannon inequality}} \leq \frac{3}{2}
$$

$$
\tfrac{1}{2}[h(XY) + h(YZ) - h(Y) - h(XYZ)] \geq 0
$$

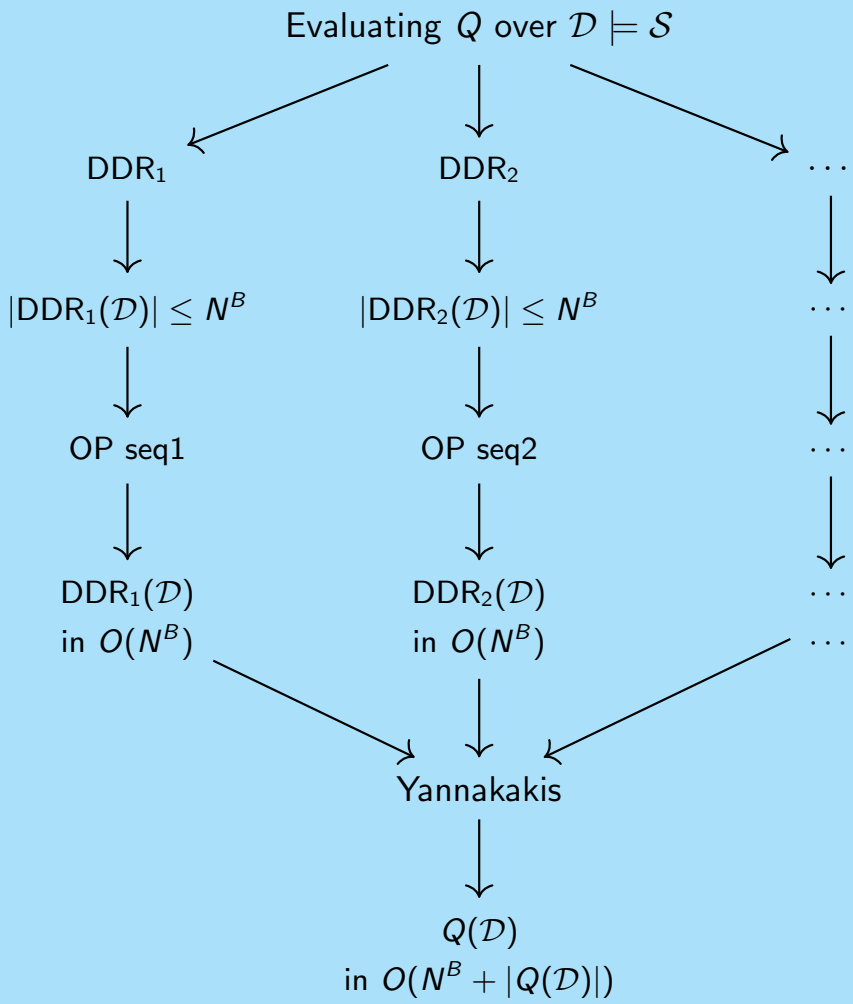
$$
\tfrac{1}{2}[h(Y) + h(ZW) - h(YZW)] \geq 0
$$

PANDA

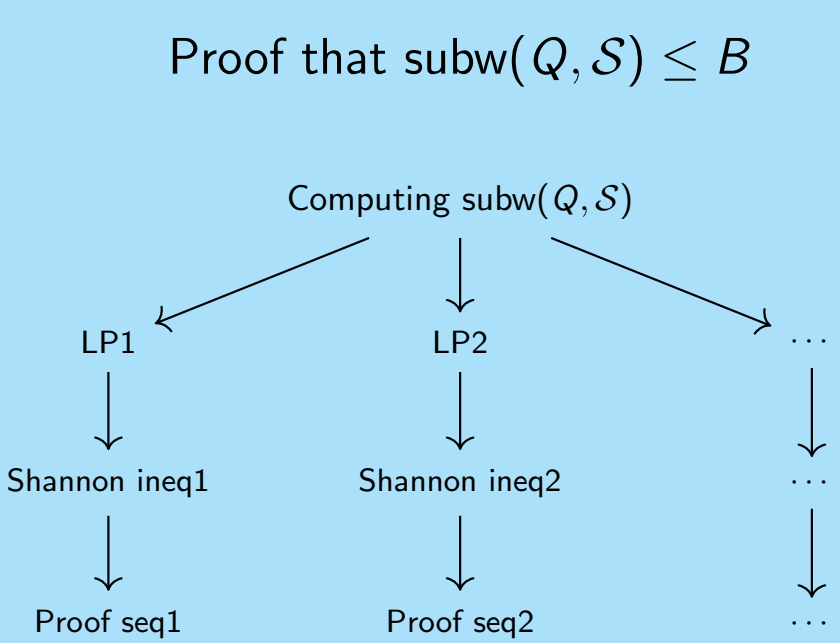
Proof that $\text{subw}(Q, \mathcal{S}) \leq B$

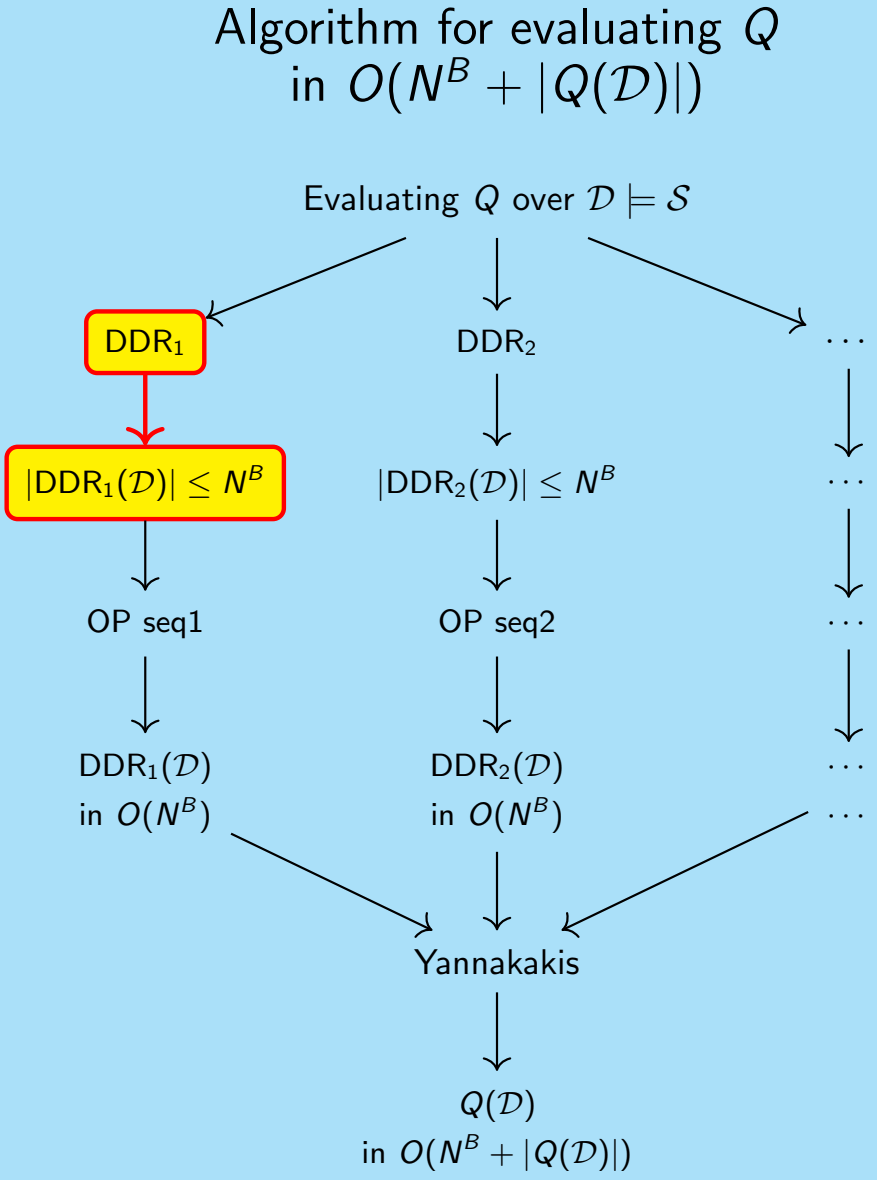


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



PANDA


Proof that subw$(Q, \mathcal{S}) \leq B$
Computing subw$(Q, \mathcal{S})$
LP1
LP2
$\cdots$
Shannon ineq1
Shannon ineq2
$\cdots$
Proof seq1
Proof seq2
$\cdots$


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$
Evaluating Q over $\mathcal{D} \models \mathcal{S}$
DDR_1
DDR_2
$\cdots$
$|\mathrm{DDR}_1(\mathcal{D})| \leq N^B$
$|\mathrm{DDR}_2(\mathcal{D})| \leq N^B$
$\cdots$
OP seq1
OP seq2
$\cdots$
$\mathrm{DDR}_1(\mathcal{D})$ in $O(N^B)$
$\mathrm{DDR}_2(\mathcal{D})$ in $O(N^B)$
$\cdots$
$\cdots$
Yannakakis
$Q(\mathcal{D})$ in $O(N^B + |Q(\mathcal{D})|)$

Output Size Bound of a DDR

$$A_{11}(X, Y, Z) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

Output Size Bound of a DDR

$$A_{11}(X, Y, Z) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

- *Many* valid outputs (A_{11}, A_{21})
- Each has *size* $\max(|A_{11}|, |A_{21}|)$
- The **output size** is their *min*

Output Size Bound of a DDR

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

- *Many* valid outputs (A_{11}, A_{21})
- Each has *size* $\max(|A_{11}|, |A_{21}|)$
- The **output size** is their *min*

$$t \leq \frac{1}{2}h(XYZ) + \frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY) + \frac{1}{2}h(YZ) + \frac{1}{2}h(ZW) \leq \frac{3}{2}$$

Output Size Bound of a DDR

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

- *Many* valid outputs (A_{11}, A_{21})
- Each has *size* $\max(|A_{11}|, |A_{21}|)$
- The **output size** is their *min*

$$t \leq \frac{1}{2}h(XYZ) + \frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY) + \frac{1}{2}h(YZ) + \frac{1}{2}h(ZW) \leq \frac{3}{2}$$

$$\Downarrow$$

The DDR has output size $\leq N^{3/2}$

Output Size Bound of a DDR (cont.)

$$A_{11}(X, Y, Z) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

Proving the DDR has output size $\leq N^{3/2}$

Output Size Bound of a DDR (cont.)

$$A_{11}(X, Y, Z) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

Proving the DDR has output size $\leq N^{3/2}$

$A_{11}, A_{21}, A \leftarrow \emptyset$
for $(x, y, z, w) \in R \bowtie S \bowtie T \bowtie U$ **do**
    if $(x, y, z) \notin A_{11} \quad \wedge \quad (y, z, w) \notin A_{21}$ **then**
        Insert (x, y, z) into A_{11}, (y, z, w) into A_{21}, and (x, y, z, w) into A

Output Size Bound of a DDR (cont.)

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

Proving the DDR has output size $\leq N^{3/2}$

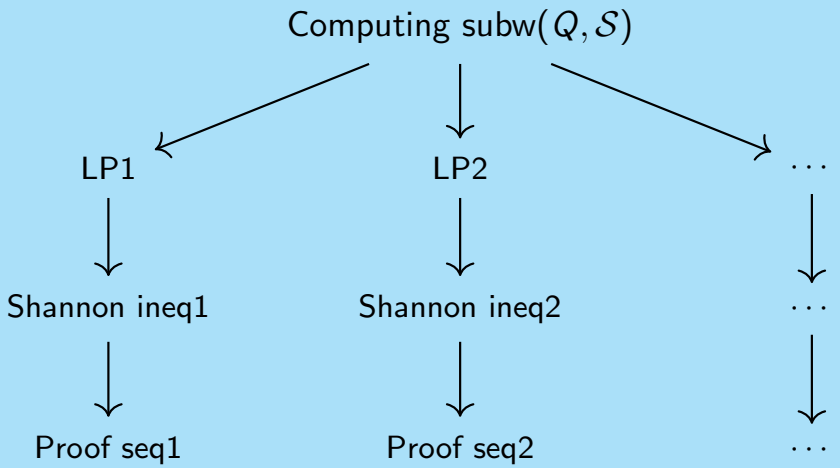
$A_{11}, A_{21}, A \leftarrow \emptyset$
for $(x,y,z,w) \in R \bowtie S \bowtie T \bowtie U$ **do**
  if $(x,y,z) \notin A_{11} \quad \wedge \quad (y,z,w) \notin A_{21}$ **then**
    Insert (x,y,z) into A_{11}, (y,z,w) into A_{21}, and (x,y,z,w) into A

$$\log_N |A_{11}| = \log_N |A_{21}| = \log_N |A| = h(XYZW) =$$

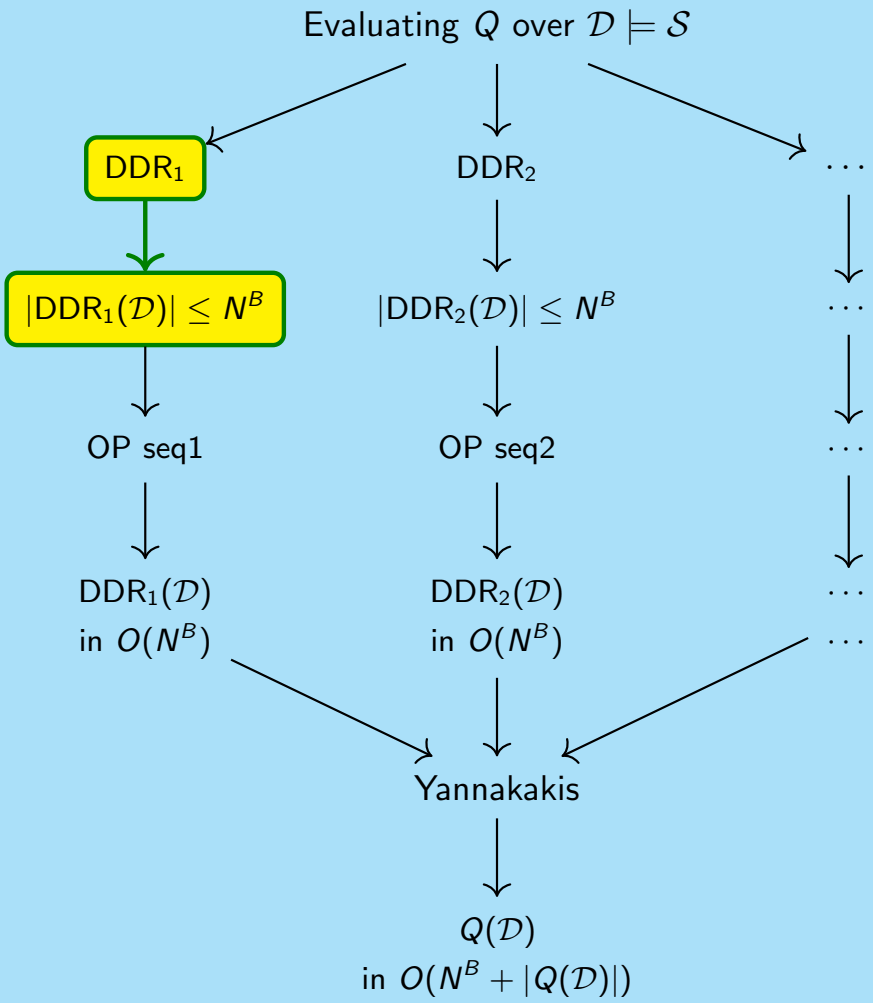
$$\frac{1}{2}h(XYZ) + \frac{1}{2}h(YZW) \quad \leq \quad \frac{1}{2}h(XY) + \frac{1}{2}h(YZ) + \frac{1}{2}h(ZW) \quad \leq \quad \frac{3}{2}$$

PANDA

Proof that $\mathrm{subw}(Q, \mathcal{S}) \leq B$

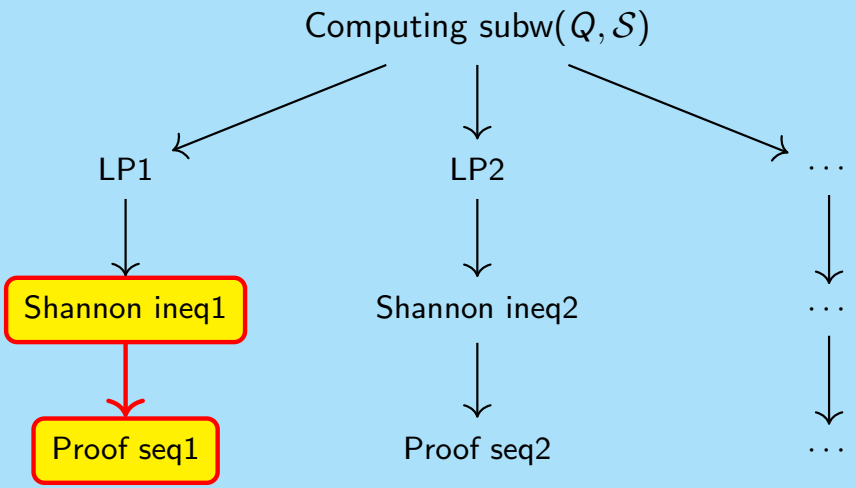


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$

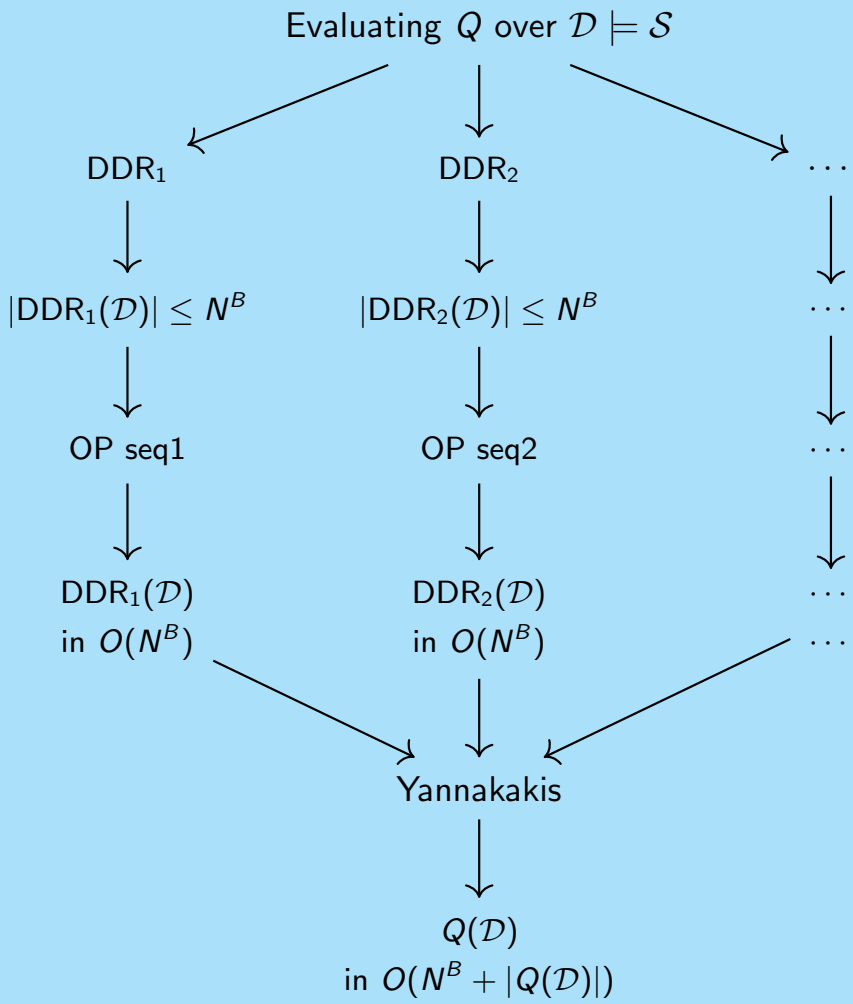


PANDA

Proof that $\text{subw}(Q, \mathcal{S}) \leq B$



Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ) + \frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY) + \frac{1}{2}h(YZ) + \frac{1}{2}h(ZW)$$

$$h(XYZ) + h(YZW) \leq h(XY) + h(YZ) + h(ZW)$$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ) + \frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY) + \frac{1}{2}h(YZ) + \frac{1}{2}h(ZW)$$

$$h(XYZ) + h(YZW) \leq h(XY) + h(YZ) + h(ZW)$$

$h(XY) + h(YZ) + h(ZW)$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$h(XY)+h(YZ)+h(ZW)$

$h(YZ) \to h(Y)+h(Z|Y)$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$h(XY)+h(YZ)+h(ZW)$

$=h(XY)+h(Y)+h(Z|Y)+h(ZW)$

$h(YZ) \rightarrow h(Y)+h(Z|Y)$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$h(XY)+h(YZ)+h(ZW)$$

$$= h(XY)+h(Y)+h(Z|Y)+h(ZW)$$

$$h(YZ) \to h(Y)+h(Z|Y)$$

$$h(Z|Y) \to h(Z|XY)$$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$h(XY)+h(YZ)+h(ZW)$$
$$= h(XY)+h(Y)+h(Z|Y)+h(ZW)$$
$$\geq h(XY)+h(Y)+h(Z|XY)+h(ZW)$$

$h(YZ) \rightarrow h(Y)+h(Z|Y)$

$h(Z|Y) \rightarrow h(Z|XY)$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$\begin{aligned} & h(XY)+h(YZ)+h(ZW) \\ =\ & h(XY)+h(Y)+h(Z|Y)+h(ZW) \\ \geq\ & h(XY)+h(Y)+h(Z|XY)+h(ZW) \end{aligned}$$

$$\begin{aligned} & h(YZ) \to h(Y)+h(Z|Y) \\ & h(Z|Y) \to h(Z|XY) \\ & h(XY)+h(Z|XY) \to h(XYZ) \end{aligned}$$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$\begin{aligned}
& h(XY)+h(YZ)+h(ZW) \\
=\ & h(XY)+h(Y)+h(Z|Y)+h(ZW) \\
\geq\ & h(XY)+h(Y)+h(Z|XY)+h(ZW) \\
=\ & h(XYZ)+h(Y)+h(ZW)
\end{aligned}$$

$$\begin{aligned}
& h(YZ) \to h(Y)+h(Z|Y) \\
& h(Z|Y) \to h(Z|XY) \\
& h(XY)+h(Z|XY) \to h(XYZ)
\end{aligned}$$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$\begin{aligned}
& h(XY)+h(YZ)+h(ZW) \\
=\ & h(XY)+h(Y)+h(Z|Y)+h(ZW) \\
\geq\ & h(XY)+h(Y)+h(Z|XY)+h(ZW) \\
=\ & h(XYZ)+h(Y)+h(ZW)
\end{aligned}$$

$$\begin{aligned}
& h(YZ) \to h(Y)+h(Z|Y) \\
& h(Z|Y) \to h(Z|XY) \\
& h(XY)+h(Z|XY) \to h(XYZ) \\
& h(Y) \to h(Y|ZW)
\end{aligned}$$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$\begin{aligned}
& h(XY)+h(YZ)+h(ZW) \\
=\ & h(XY)+h(Y)+h(Z|Y)+h(ZW) \\
\geq\ & h(XY)+h(Y)+h(Z|XY)+h(ZW) \\
=\ & h(XYZ)+h(Y)+h(ZW) \\
\geq\ & h(XYZ)+h(Y|ZW)+h(ZW)
\end{aligned}$$

$$\begin{aligned}
& h(YZ) \to h(Y)+h(Z|Y) \\
& h(Z|Y) \to h(Z|XY) \\
& h(XY)+h(Z|XY) \to h(XYZ) \\
& h(Y) \to h(Y|ZW)
\end{aligned}$$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$h(XY)+h(YZ)+h(ZW)$	$h(YZ) \to h(Y)+h(Z\|Y)$
$=h(XY)+h(Y)+h(Z\|Y)+h(ZW)$	$h(Z\|Y) \to h(Z\|XY)$
$\geq h(XY)+h(Y)+h(Z\|XY)+h(ZW)$	$h(XY)+h(Z\|XY) \to h(XYZ)$
$=h(XYZ)+h(Y)+h(ZW)$	$h(Y) \to h(Y\|ZW)$
$\geq h(XYZ)+h(Y\|ZW)+h(ZW)$	$h(ZW)+h(Y\|ZW) \to h(YZW)$

Shannon Inequality $\Rightarrow$ Proof Sequence

$$\frac{1}{2}h(XYZ)+\frac{1}{2}h(YZW) \leq \frac{1}{2}h(XY)+\frac{1}{2}h(YZ)+\frac{1}{2}h(ZW)$$

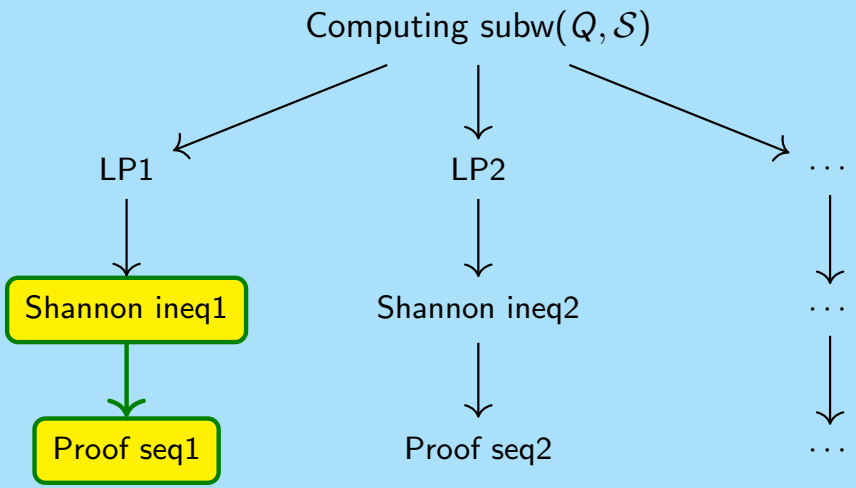
$$h(XYZ)+h(YZW) \leq h(XY)+h(YZ)+h(ZW)$$

$$\begin{aligned}
& h(XY)+h(YZ)+h(ZW) \\
=\ & h(XY)+h(Y)+h(Z|Y)+h(ZW) \\
\geq\ & h(XY)+h(Y)+h(Z|XY)+h(ZW) \\
=\ & h(XYZ)+h(Y)+h(ZW) \\
\geq\ & h(XYZ)+h(Y|ZW)+h(ZW) \\
=\ & h(XYZ)+h(YZW)
\end{aligned}$$

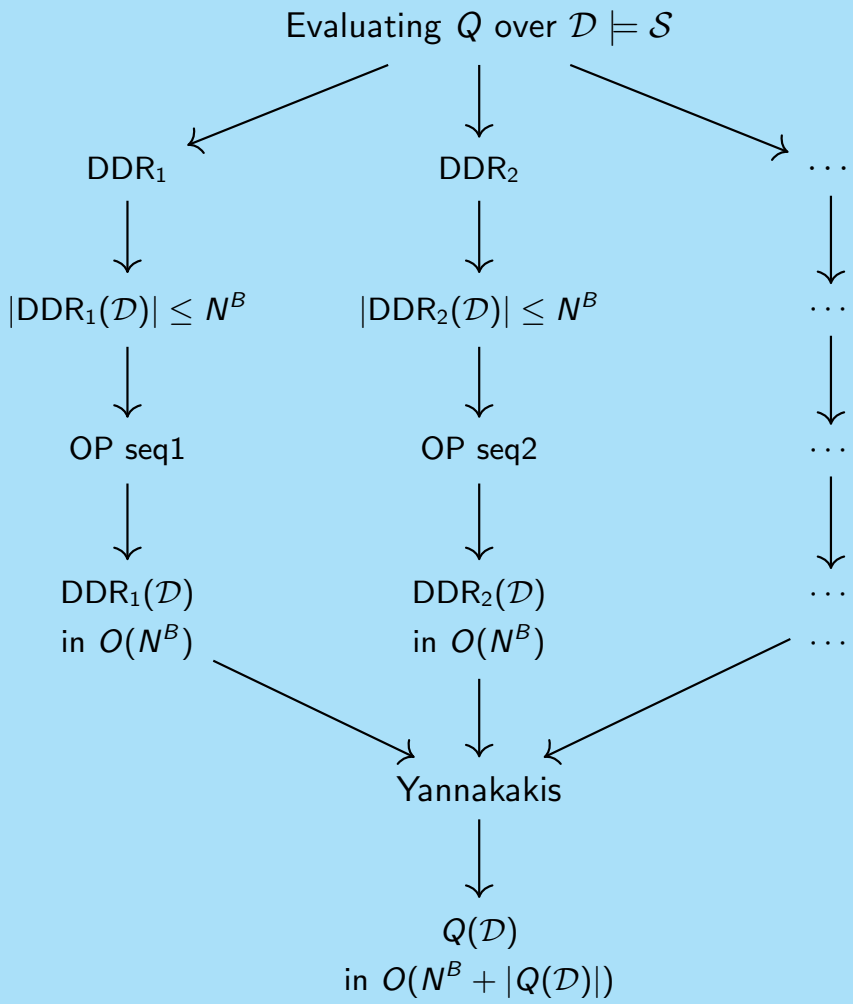
$$\begin{aligned}
& h(YZ) \to h(Y)+h(Z|Y) \\
& h(Z|Y) \to h(Z|XY) \\
& h(XY)+h(Z|XY) \to h(XYZ) \\
& h(Y) \to h(Y|ZW) \\
& h(ZW)+h(Y|ZW) \to h(YZW)
\end{aligned}$$

PANDA

Proof that $\text{subw}(Q, \mathcal{S}) \leq B$

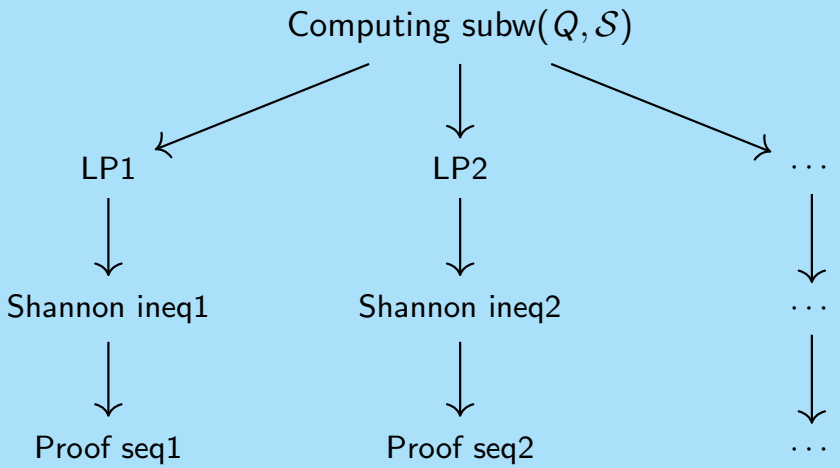


Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$

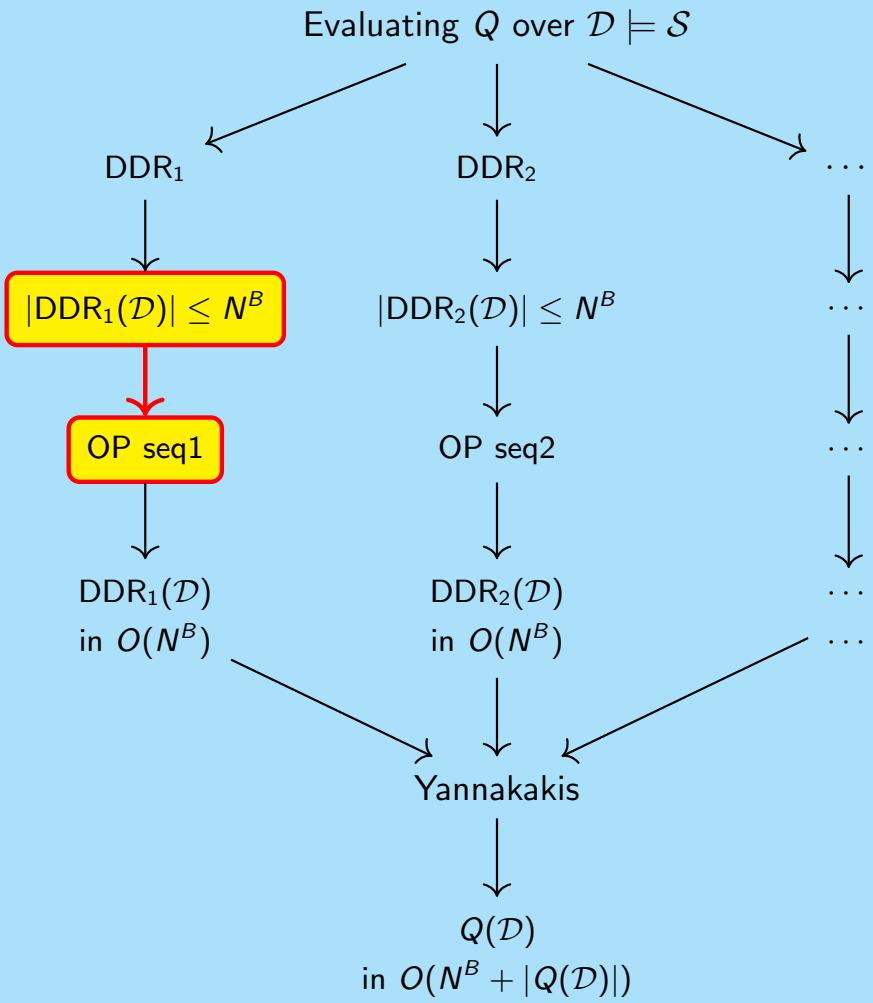


PANDA

Proof that subw$(Q, \mathcal{S}) \leq B$



Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



Evaluating a DDR in its output size bound

$$A_{11}(X, Y, Z) \vee A_{21}(Y, Z, W) \text{ :- } R(X, Y) \wedge S(Y, Z) \wedge T(Z, W) \wedge U(W, X)$$

has an output size bound of $N^{3/2}$, based on

$$h(XYZ) + h(YZW) \quad \leq \quad h(XY) + h(YZ) + h(ZW)$$

Evaluating a DDR in its output size bound

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

has an output size bound of $N^{3/2}$, based on

$$h(XYZ) + h(YZW) \quad \leq \quad h(XY) + h(YZ) + h(ZW)$$

$h(YZ) \rightarrow h(Y) + h(Z|Y)$

$h(Z|Y) \rightarrow h(Z|XY)$

$h(XY) + h(Z|XY) \rightarrow h(XYZ)$

$h(Y) \rightarrow h(Y|ZW)$

$h(ZW) + h(Y|ZW) \rightarrow h(YZW)$

Evaluating a DDR in its output size bound

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \text{ :- } R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

has an output size bound of $N^{3/2}$, based on

$$h(XYZ) + h(YZW) \quad \leq \quad h(XY) + h(YZ) + h(ZW)$$

$h(YZ) \to h(Y) + h(Z|Y)$

$h(Z|Y) \to h(Z|XY)$

$h(XY) + h(Z|XY) \to h(XYZ)$

$h(Y) \to h(Y|ZW)$

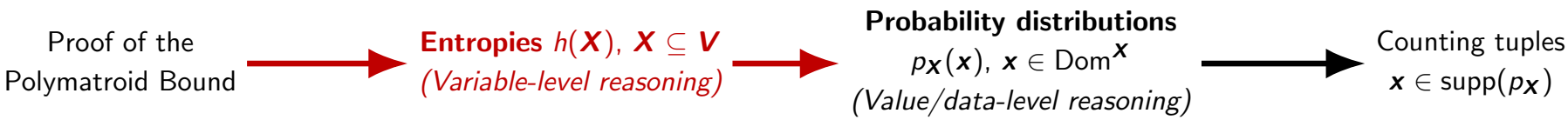
$h(ZW) + h(Y|ZW) \to h(YZW)$

Two Algorithmic Translations:

- **Classic PANDA** [Abo Khamis, Ngo, et al. 2017]
  - Extra $O(\text{polylog}(N))$
- **PANDAExpress** [Abo Khamis, Ngo, et al. 2026]
  - $\text{polylog}(N)$ removed!

Intuition

Classic PANDA route



Intuition

Classic PANDA route

Proof of the Polymatroid Bound → **Entropies** $h(\boldsymbol{X})$, $\boldsymbol{X} \subseteq \boldsymbol{V}$ *(Variable-level reasoning)* → **Probability distributions** $p_{\boldsymbol{X}}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}$ *(Value/data-level reasoning)* → Counting tuples $\boldsymbol{x} \in \mathrm{supp}(p_{\boldsymbol{X}})$

NEW Proof of the Polymatroid Bound → **Sub-probability measures** $p_{\boldsymbol{X}}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathrm{Dom}^{\boldsymbol{X}}$ *(Value/data-level reasoning)* → Counting tuples $\boldsymbol{x} \in \mathrm{supp}(p_{\boldsymbol{X}})$

PANDAExpress route

Background: **Sub-probability Measures**

- *"Relaxed"* probability distribution, $p_{\boldsymbol{X}}(\boldsymbol{x})$

$$\sum_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x}) \leq 1$$

Background: **Sub-probability Measures**

- *"Relaxed"* probability distribution, $p_{\boldsymbol{X}}(\boldsymbol{x})$

$$\sum_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x}) \leq 1$$

- *Conditional* measures: $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$

Background: **Sub-probability Measures**

- *"Relaxed"* probability distribution, $p_{\boldsymbol{X}}(\boldsymbol{x})$

$$\sum_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x}) \leq 1$$

- *Conditional* measures: $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$
- $p_{\boldsymbol{X}}(\boldsymbol{x}) \geq \frac{1}{B}$, for all $\boldsymbol{x} \in supp(p_{\boldsymbol{X}}) \quad \Rightarrow \quad |\text{supp}(p_{\boldsymbol{X}})| \leq B$

Background: **Sub-probability Measures**

- *"Relaxed"* probability distribution, $p_{\boldsymbol{X}}(\boldsymbol{x})$

$$\sum_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x}) \leq 1$$

- *Conditional* measures: $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$
- $p_{\boldsymbol{X}}(\boldsymbol{x}) \geq \frac{1}{B}$, for all $\boldsymbol{x} \in supp(p_{\boldsymbol{X}}) \quad \Rightarrow \quad |\text{supp}(p_{\boldsymbol{X}})| \leq B$
- **Closed** under

Marginals: $$p_{\boldsymbol{X}}(\boldsymbol{x}) \stackrel{\text{def}}{=} \sum_{\boldsymbol{y}} p_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})$$

Conditionals: $$p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{p_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})}{p_{\boldsymbol{X}}(\boldsymbol{x})}$$

Geometric means: $$p_{\boldsymbol{X}}(\boldsymbol{x}) \stackrel{\text{def}}{=} \left[p_{\boldsymbol{X}}^{(1)}(\boldsymbol{x}) \cdot \ldots \cdot p_{\boldsymbol{X}}^{(k)}(\boldsymbol{x})\right]^{1/k}$$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3 \log_N N$	
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3 \log_N N$	

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$ $p_{Z\|Y}(z\|y) \stackrel{\text{def}}{=} \frac{p_{YZ}(y,z)}{p_Y(y)} = \frac{1}{\deg_S(Z\|Y=y)}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$ $p_{Z\|Y}(z\|y) \stackrel{\text{def}}{=} \frac{p_{YZ}(y,z)}{p_Y(y)} = \frac{1}{\deg_S(Z\|Y=y)}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$ $p_{Z\|XY}(z\|xy) \stackrel{\text{def}}{=} p_{Z\|Y}(z\|y) = \frac{1}{\deg_S(Z\|Y=y)}$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3\log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$ $p_{Z\|Y}(z\|y) \stackrel{\text{def}}{=} \frac{p_{YZ}(y,z)}{p_Y(y)} = \frac{1}{\deg_S(Z\|Y=y)}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$ $p_{Z\|XY}(z\|xy) \stackrel{\text{def}}{=} p_{Z\|Y}(z\|y) = \frac{1}{\deg_S(Z\|Y=y)}$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$ $p_{XYZ}(x,y,z) \stackrel{\text{def}}{=} p_{XY}(x,y) \cdot p_{Z\|XY}(z\|xy) = \frac{1}{N \cdot \deg_S(Z\|Y=y)}$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3\log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3 \log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$ $p_{Z\|Y}(z\|y) \stackrel{\text{def}}{=} \frac{p_{YZ}(y,z)}{p_Y(y)} = \frac{1}{\deg_S(Z\|Y=y)}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$ $p_{Z\|XY}(z\|xy) \stackrel{\text{def}}{=} p_{Z\|Y}(z\|y) = \frac{1}{\deg_S(Z\|Y=y)}$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$ $p_{XYZ}(x,y,z) \stackrel{\text{def}}{=} p_{XY}(x,y) \cdot p_{Z\|XY}(z\|xy) = \frac{1}{N \cdot \deg_S(Z\|Y=y)}$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$ $p_{Y\|ZW}(y\|zw) \stackrel{\text{def}}{=} p_Y(y) = \frac{\deg_S(Z\|Y=y)}{N}$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3 \log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound

$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3 \log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$ $p_{Z\|Y}(z\|y) \stackrel{\text{def}}{=} \frac{p_{YZ}(y,z)}{p_Y(y)} = \frac{1}{\deg_S(Z\|Y=y)}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$ $p_{Z\|XY}(z\|xy) \stackrel{\text{def}}{=} p_{Z\|Y}(z\|y) = \frac{1}{\deg_S(Z\|Y=y)}$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$ $p_{XYZ}(x,y,z) \stackrel{\text{def}}{=} p_{XY}(x,y) \cdot p_{Z\|XY}(z\|xy) = \frac{1}{N \cdot \deg_S(Z\|Y=y)}$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$ $p_{Y\|ZW}(y\|zw) \stackrel{\text{def}}{=} p_Y(y) = \frac{\deg_S(Z\|Y=y)}{N}$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$ $p_{YZW}(y,z,w) \stackrel{\text{def}}{=} p_{ZW}(z,w) \cdot p_{Y\|ZW}(y\|zw) = \frac{\deg_S(Z\|Y=y)}{N^2}$
$h(XYZ) + h(YZW) \leq$ $h(XYZ) + h(YZW) \leq 3 \log_N N$	$p_{XYZ}(x,y,z) \cdot p_{YZW}(y,z,w) \geq 1/N^3$

NEW Proof of DDR Output Size Bound $B \stackrel{\text{def}}{=} N^{3/2}$

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \quad :- \quad R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$\underbrace{p_{XYZ}(x,y,z)}_{} \cdot \underbrace{p_{YZW}(y,z,w)}_{} \geq \frac{1}{N^3} = \frac{1}{B^2} \qquad \forall (x,y,z,w) \in \bowtie$$

NEW Proof of DDR Output Size Bound $B \stackrel{\text{def}}{=} N^{3/2}$

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \quad :- \quad R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$\underbrace{p_{XYZ}(x,y,z)} \quad \cdot \quad \underbrace{p_{YZW}(y,z,w)} \quad \geq \quad \frac{1}{N^3} = \frac{1}{B^2} \qquad \forall (x,y,z,w) \in \bowtie$$

For every $(x,y,z,w) \in \bowtie$

either $p_{XYZ}(x,y,z) \geq \frac{1}{B}$

or $p_{YZW}(y,z,w) \geq \frac{1}{B}$

NEW Proof of DDR Output Size Bound $B \stackrel{\text{def}}{=} N^{3/2}$

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \quad :- \quad R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$\underbrace{p_{XYZ}(x,y,z)}_{\frac{1}{N \cdot \deg_S(Z|Y=y)}} \cdot \underbrace{p_{YZW}(y,z,w)}_{\frac{\deg_S(Z|Y=y)}{N^2}} \geq \frac{1}{N^3} = \frac{1}{B^2} \qquad \forall (x,y,z,w) \in \bowtie$$

For every $(x,y,z,w) \in \bowtie$

either $p_{XYZ}(x,y,z) \geq \frac{1}{B}$

or $p_{YZW}(y,z,w) \geq \frac{1}{B}$

NEW Proof of DDR Output Size Bound $B \stackrel{\text{def}}{=} N^{3/2}$

$$A_{11}(X,Y,Z) \vee A_{21}(Y,Z,W) \quad :- \quad R(X,Y) \wedge S(Y,Z) \wedge T(Z,W) \wedge U(W,X)$$

$$\underbrace{p_{XYZ}(x,y,z)}_{\frac{1}{N \cdot \deg_S(Z|Y=y)}} \cdot \underbrace{p_{YZW}(y,z,w)}_{\frac{\deg_S(Z|Y=y)}{N^2}} \geq \frac{1}{N^3} = \frac{1}{B^2} \qquad \forall (x,y,z,w) \in \bowtie$$

For every $(x,y,z,w) \in \bowtie$

either	$p_{XYZ}(x,y,z) \geq \frac{1}{B}$	$\Leftrightarrow$	$\deg_S(Z\|Y=y) \leq \sqrt{N}$
or	$p_{YZW}(y,z,w) \geq \frac{1}{B}$	$\Leftrightarrow$	$\deg_S(Z\|Y=y) \geq \sqrt{N}$

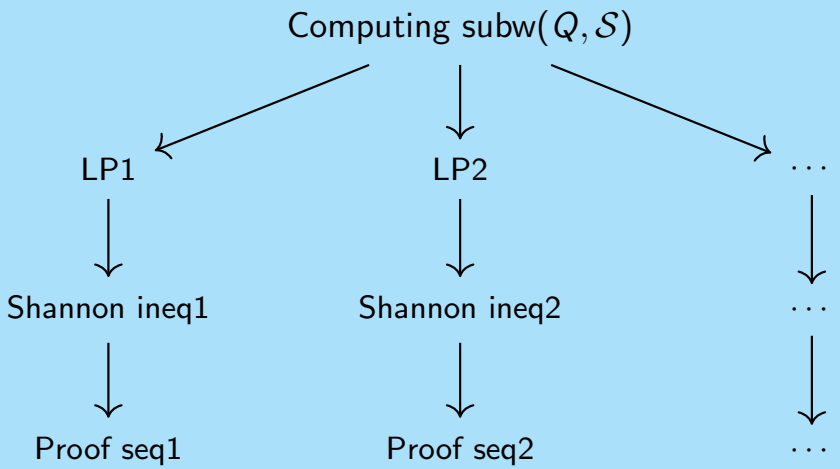
NEW Algorithm for Evaluating the DDR

This **IS** the algorithm, **BUT** compute *truncated* p_{XYZ}, p_{YZW} by $\geq 1/B$

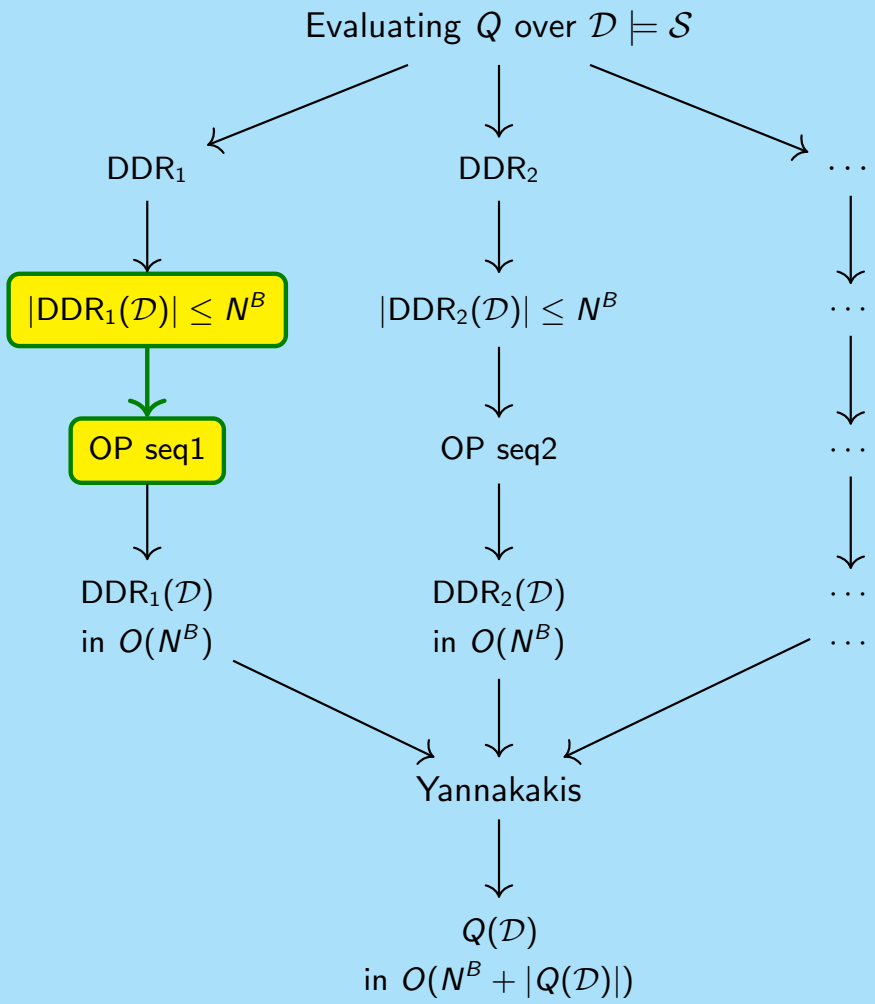
$h(XYZ) + h(YZW) \leq$ $h(XY) + h(YZ) + h(ZW) \leq 3 \log_N N$	$p_{XY}(x,y) \cdot p_{YZ}(y,z) \cdot p_{ZW}(z,w) \geq 1/N^3$ $p_{XY}(x,y) \stackrel{\text{def}}{=} 1/N, \quad p_{YZ}(y,z) \stackrel{\text{def}}{=} 1/N, \quad p_{ZW}(z,w) \stackrel{\text{def}}{=} 1/N$
$h(YZ) \quad \rightarrow \quad h(Y) + h(Z\|Y)$	$p_{YZ}(y,z) \quad \rightarrow \quad p_Y(y) \cdot p_{Z\|Y}(z\|y)$ $p_Y(y) \stackrel{\text{def}}{=} \sum_z p_{YZ}(y,z) = \frac{\deg_S(Z\|Y=y)}{N}$ $p_{Z\|Y}(z\|y) \stackrel{\text{def}}{=} \frac{p_{YZ}(y,z)}{p_Y(y)} = \frac{1}{\deg_S(Z\|Y=y)}$
$h(Z\|Y) \quad \rightarrow \quad h(Z\|XY)$	$p_{Z\|Y}(z\|y) \quad \rightarrow \quad p_{Z\|XY}(z\|xy)$ $p_{Z\|XY}(z\|xy) \stackrel{\text{def}}{=} p_{Z\|Y}(z\|y) = \frac{1}{\deg_S(Z\|Y=y)}$
$h(XY) + h(Z\|XY) \rightarrow h(XYZ)$	$p_{XY}(x,y) \quad \cdot p_{Z\|XY}(z\|xy) \rightarrow \quad p_{XYZ}(x,y,z)$ $p_{XYZ}(x,y,z) \stackrel{\text{def}}{=} p_{XY}(x,y) \cdot p_{Z\|XY}(z\|xy) = \frac{1}{N \cdot \deg_S(Z\|Y=y)}$
$h(Y) \quad \rightarrow \quad h(Y\|ZW)$	$p_Y(y) \quad \rightarrow \quad p_{Y\|ZW}(y\|zw)$ $p_{Y\|ZW}(y\|zw) \stackrel{\text{def}}{=} p_Y(y) = \frac{\deg_S(Z\|Y=y)}{N}$
$h(ZW) + h(Y\|ZW) \rightarrow h(Y\|ZW)$	$p_{ZW}(z,w) \quad \cdot p_{Y\|ZW}(y\|zw) \rightarrow \quad p_{YZW}(y,z,w)$ $p_{YZW}(y,z,w) \stackrel{\text{def}}{=} p_{ZW}(z,w) \cdot p_{Y\|ZW}(y\|zw) = \frac{\deg_S(Z\|Y=y)}{N^2}$

PANDA

Proof that $\mathrm{subw}(Q, \mathcal{S}) \leq B$



Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



Outline

1. Cardinality Estimation
2. Query Evaluation
3. Conclusion

Extensions

- **Other Queries:**
- **Other Statistics:**
- **Other Operators:**

Extensions

- **Other Queries:** FAQs, #CQs
  - FAQs: Aggregate queries over a semiring $(K, \oplus, \otimes)$

    Idempotent e.g. $(\mathbb{R}, \min, +)$ **Works**

    Non-idempotent e.g. $(\mathbb{N}, +, \times)$ or #CQs **Open**
    - $\#\mathsf{subw}(Q) \geq \mathsf{subw}(Q)$ [Abo Khamis, Curtin, et al. 2020]
    - Not tight [Bringmann et al. 2025]

- **Other Statistics:**

- **Other Operators:**

Extensions

- **Other Queries:** FAQs, #CQs
  - FAQs: Aggregate queries over a semiring $(K, \oplus, \otimes)$

    Idempotent e.g. $(\mathbb{R}, \min, +)$ **Works**

    Non-idempotent e.g. $(\mathbb{N}, +, \times)$ or #CQs **Open**
    - $\#\mathsf{subw}(Q) \geq \mathsf{subw}(Q)$ [Abo Khamis, Curtin, et al. 2020]
    - Not tight [Bringmann et al. 2025]

- **Other Statistics:** ℓ_p-norms of Degree Vectors
  - [Jayaraman et al. 2021; Abo Khamis, Nakos, et al. 2024]

- **Other Operators:**

Extensions

- **Other Queries:** FAQs, #CQs
  - FAQs: Aggregate queries over a semiring $(K, \oplus, \otimes)$

    Idempotent e.g. $(\mathbb{R}, \min, +)$ **Works**

    Non-idempotent e.g. $(\mathbb{N}, +, \times)$ or #CQs **Open**
    - $\#\mathsf{subw}(Q) \geq \mathsf{subw}(Q)$ [Abo Khamis, Curtin, et al. 2020]
    - Not tight [Bringmann et al. 2025]

- **Other Statistics:** ℓ_p-norms of Degree Vectors
  - [Jayaraman et al. 2021; Abo Khamis, Nakos, et al. 2024]

- **Other Operators:** Fast Matrix Multiplication (FMM)
  - $\omega\text{-}\mathsf{subw}(Q) \leq \mathsf{subw}(Q)$ [Abo Khamis, Hu, et al. 2025]
    - $\omega\text{-}\mathsf{subw}(Q_\triangle) = \frac{2\omega}{\omega+1}$ [Alon et al. 1997]

Open Problems

- #CQs **in submodular width time?**
- **Entropic version of the submodular width, entw$(Q) \leq$ subw(Q)**
  - Runtime $O(N^{\text{entw}(Q)} + \text{OUT})$?
- **Tractabilty of computing the polymatroid bound**
  - Is it in NP (in $(Q, \mathcal{S})$ size)?
  - Tractable for *"simple"* DCs [Im et al. 2025]
- **Length of the proof sequence**
  - Exponential in general [Abo Khamis, Ngo, et al. 2017]
  - Polynomial for *"simple"* DCs [Im et al. 2025]

Open Problems (cont.)

- **Lower bounds**
  - *Classic lower bound*
    - For self-join-free BCQs: Unbounded subw $\Leftrightarrow$ No FPT [Marx 2013]
  - *Fine-grained lower bounds*
    - $\Omega(N^{\mathsf{subw}(Q)} + \mathsf{OUT})$ for $\mathsf{subw}(Q) < 2$, binary relations [Bringmann et al. 2025]
    - $\Omega(N^{\mathsf{clemb}(Q)} + \mathsf{OUT})$ for $(\min, +)$-aggregate queries [Fan et al. 2023]
    - $\Omega(N^{(1-\epsilon)\cdot \mathsf{entw}(Q)})$ on the semiring $(\oplus, \otimes)$-circuit size [Fan et al. 2024]
    - $N^{\Omega(\mathsf{subw}(Q)^{1/4})}$ on the $(\cup, \times)$-circuit size [Berkholz et al. 2026]

Open Problems (cont.)

- **Lower bounds**

  Best known *combinatorial* upper bound: $O(N^{\text{subw}(Q)} + \text{OUT})$

  - **[Q1]** Does $\Omega(N^{\text{subw}(Q)} + \text{OUT})$ follow from existing conjectures?

  - **[Q2]** What is a *"combinatorial algorithm"*?

    Two ways to rule out FMM

    - Use $(\min, +)$-semiring [Fan et al. 2023; Bringmann et al. 2025]
    - Use *full* queries [Bringmann et al. 2025]

  - **[Q3]** A *non-combinatorial* $\Omega(N^{\omega\text{-subw}(Q)} + \text{OUT})$?

Meet **Meng Meng!** (SIGMOD'25 - Berlin)





Meet **Meng Meng!** (SIGMOD'25 - Berlin)





Thank You!

References

References I


Abo Khamis, Mahmoud, Ryan R. Curtin, Benjamin Moseley, Hung Q. Ngo, Xuanlong Nguyen, Dan Olteanu, and Maximilian Schleich (Dec. 2020). "Functional Aggregate Queries with Additive Inequalities". In: *ACM Trans. Database Syst.* 45.4. DOI: 10.1145/3426865.


Abo Khamis, Mahmoud, Xiao Hu, and Dan Suciu (2025). "Fast Matrix Multiplication meets the Submodular Width". In: *Proc. ACM Manag. Data* 3.2, 98:1–98:26. DOI: 10.1145/3725235.


Abo Khamis, Mahmoud, Vasileios Nakos, Dan Olteanu, and Dan Suciu (May 2024). "Join Size Bounds using lp-Norms on Degree Sequences". In: *Proc. ACM Manag. Data* 2.2. DOI: 10.1145/3651597.


Abo Khamis, Mahmoud, Hung Q. Ngo, and Dan Suciu (2017). "What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?" In: *Proceedings of the 36th ACM PODS 2017, Chicago, IL, USA, May 14-19, 2017*. Extended version available at http://arxiv.org/abs/1612.02503. ACM, pp. 429–444. DOI: 10.1145/3034786.3056105.


— (2026). "PANDAExpress: a Simpler and Faster PANDA Algorithm". In: *Proc. ACM Manag. Data*. DOI: 10.1145/3801910.


Alon, Noga (1981). "On the number of subgraphs of prescribed type of graphs with a given number of edges". In: *Israel J. Math.* 38.1-2, pp. 116–130. DOI: 10.1007/BF02761855.

References II


Alon, Noga, Raphael Yuster, and Uri Zwick (1997). "Finding and Counting Given Length Cycles". In: *Algorithmica* 17.3, pp. 209–223. DOI: 10.1007/BF02523189.


Atserias, Albert, Martin Grohe, and Dániel Marx (2013). "Size Bounds and Query Plans for Relational Joins". In: *SIAM J. Comput.* 42.4, pp. 1737–1767. DOI: 10.1137/110859440.


Bagan, Guillaume, Arnaud Durand, and Etienne Grandjean (2007). "On Acyclic Conjunctive Queries and Constant Delay Enumeration". In: *Computer Science Logic, 21st International Workshop, CSL 2007, 16th Annual Conference of the EACSL, Lausanne, Switzerland, September 11-15, 2007, Proceedings*. Vol. 4646. Lecture Notes in Computer Science. Springer, pp. 208–222. DOI: 10.1007/978-3-540-74915-8_18.


Berkholz, Christoph and Harry Vinall-Smeeth (2026). "Factorised Representations of Join Queries: Tight Bounds and a New Dichotomy". In: *29th International Conference on Database Theory (ICDT 2026)*. Ed. by Balder ten Cate and Maurice Funk. Vol. 365. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 11:1–11:20. DOI: 10.4230/LIPIcs.ICDT.2026.11.


Bringmann, Karl and Egor Gorbachev (2025). "A Fine-Grained Classification of Subquadratic Patterns for Subgraph Listing and Friends". In: *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025*. ACM, pp. 2145–2156. DOI: 10.1145/3717823.3718141.

References III


Chung, F. R. K., R. L. Graham, P. Frankl, and J. B. Shearer (1986). "Some intersection theorems for ordered sets and graphs". In: *J. Combin. Theory Ser. A* 43.1, pp. 23–37. DOI: 10.1016/0097-3165(86)90019-1.


Fan, Austen Z., Paraschos Koutris, and Hangdong Zhao (2023). "The Fine-Grained Complexity of Boolean Conjunctive Queries and Sum-Product Problems". In: *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*. Vol. 261. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 127:1–127:20. DOI: 10.4230/LIPIcs.ICALP.2023.127.


— (May 2024). "Tight Bounds of Circuits for Sum-Product Queries". In: *Proc. ACM Manag. Data* 2.2. DOI: 10.1145/3651588.


Friedgut, Ehud (2004). "Hypergraphs, entropy, and inequalities". In: *Amer. Math. Monthly* 111.9, pp. 749–760. DOI: 10.2307/4145187.


Friedgut, Ehud and Jeff Kahn (1998). "On the number of copies of one hypergraph in another". In: *Israel J. Math.* 105, pp. 251–256. DOI: 10.1007/BF02780332.

References IV


Gottlob, Georg, Gianluigi Greco, Nicola Leone, and Francesco Scarcello (2016). "Hypertree Decompositions: Questions and Answers". In: *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. ACM, pp. 57–74. DOI: 10.1145/2902251.2902309.


Gottlob, Georg, Stephanie Tien Lee, Gregory Valiant, and Paul Valiant (2012). "Size and Treewidth Bounds for Conjunctive Queries". In: *J. ACM* 59.3, 16:1–16:35. DOI: 10.1145/2220357.2220363.


Gottlob, Georg, Nicola Leone, and Francesco Scarcello (2002). "Hypertree Decompositions and Tractable Queries". In: *Journal of Computer and System Sciences* 64.3, pp. 579–627. DOI: https://doi.org/10.1006/jcss.2001.1809.


Grohe, Martin and Dániel Marx (2006). "Constraint solving via fractional edge covers". In: *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006*. ACM Press, pp. 289–298.


— (2014). "Constraint Solving via Fractional Edge Covers". In: *ACM Trans. Algorithms* 11.1, 4:1–4:20. DOI: 10.1145/2636918.

References V


Im, Sungjin, Benjamin Moseley, Hung Q. Ngo, and Kirk Pruhs (2025). "Efficient Algorithms for Cardinality Estimation and Conjunctive Query Evaluation With Simple Degree Constraints". In: *Proc. ACM Manag. Data* 3.2, 96:1–96:26. DOI: 10.1145/3725233.


Itai, Alon and Michael Rodeh (1977). "Finding a Minimum Circuit in a Graph". In: *Proceedings of the 9th Annual ACM Symposium on Theory of Computing, May 4-6, 1977, Boulder, Colorado, USA*. ACM, pp. 1–10. DOI: 10.1145/800105.803390.


Jayaraman, Sai Vikneshwar Mani, Corey Ropell, and Atri Rudra (Dec. 2021). "Worst-case Optimal Binary Join Algorithms under General ℓ_p Constraints". In: *arXiv e-prints*, arXiv:2112.01003, arXiv:2112.01003. DOI: 10.48550/arXiv.2112.01003.


Marx, Dániel (2013). "Tractable Hypergraph Properties for Constraint Satisfaction and Conjunctive Queries". In: *J. ACM* 60.6, 42:1–42:51. DOI: 10.1145/2535926.


Mayer, Christoph, Haozhe Zhang, Mahmoud Abo Khamis, Kyle Deeds, Dan Olteanu, and Dan Suciu (Apr. 2026). "CorrBound: Cardinality Estimation Accounting for Inter- and Intra-relation Correlations". In: *Proc. ACM Manag. Data* 4.1. DOI: 10.1145/3786633.


Wang, Yilei (2022). *Personal Communication*.

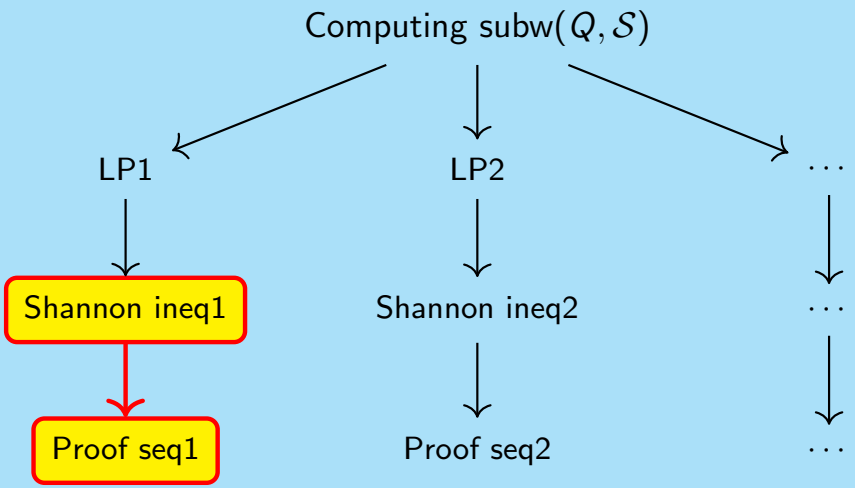
References VI

 Zhang, Zhen and Raymond W. Yeung (1997). "A non-Shannon-type conditional inequality of information quantities". In: *IEEE Trans. Information Theory* 43.6, pp. 1982–1986.

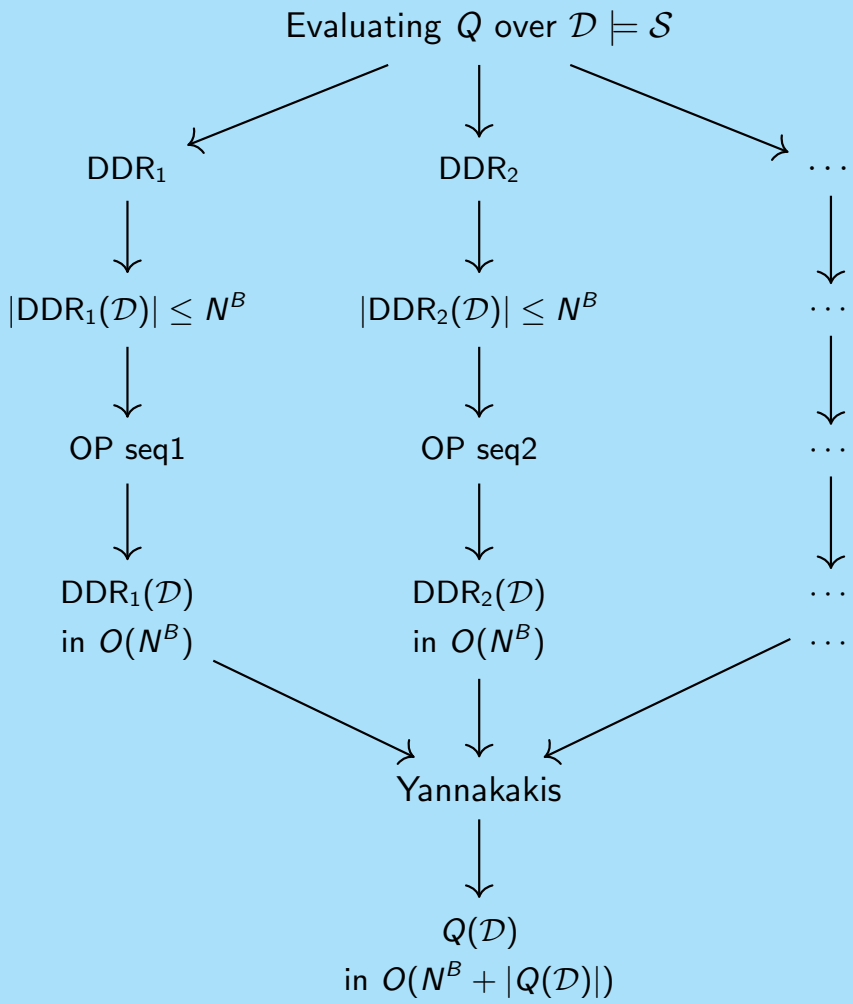
Appendix

PANDA

Proof that $\text{subw}(Q, \mathcal{S}) \leq B$



Algorithm for evaluating Q in $O(N^B + |Q(\mathcal{D})|)$



Shannon Inequality $\Rightarrow$ **Identity**

$$h(XYZ) + h(YZW) \quad \leq \quad h(XY) + h(YZ) + h(ZW)$$

$$-h(XY) - h(YZ) + h(Y) + h(XYZ) \leq 0 \qquad \text{(submodularity)}$$

$$-h(Y) - h(ZW) + h(YZW) \leq 0 \qquad \text{(submodularity)}$$

Shannon Inequality $\Rightarrow$ **Identity**

$$h(XYZ) + h(YZW) \quad \leq \quad h(XY) + h(YZ) + h(ZW)$$

$$-h(XY) - h(YZ) + h(Y) + h(XYZ) \leq 0 \qquad \text{(submodularity)}$$
$$-h(Y) - h(ZW) + h(YZW) \leq 0 \qquad \text{(submodularity)}$$

Identity

$$\overbrace{h(XYZ) + h(YZW)}^{\text{“target terms”}} = \overbrace{h(XY) + h(YZ) + h(ZW)}^{\text{“source terms”}}$$
$$+ (-h(XY) - h(YZ) + h(Y) + h(XYZ)) \quad \} \leq 0 \text{ (submodularity)}$$
$$+ (-h(Y) - h(ZW) + h(YZW)) \quad \} \leq 0 \text{ (submodularity)}$$

Shannon Inequality $\Rightarrow$ **Identity**

$$h(XYZ) + h(YZW) \quad \leq \quad h(XY) + h(YZ) + h(ZW)$$

$$-h(XY) - h(YZ) + h(Y) + h(XYZ) \leq 0 \qquad \text{(submodularity)}$$
$$-h(Y) - h(ZW) + h(YZW) \leq 0 \qquad \text{(submodularity)}$$

Identity

$$\overbrace{h(XYZ) + h(YZW)}^{\text{“target terms”}} = \overbrace{h(XY) + h(YZ) + h(ZW)}^{\text{“source terms”}}$$
$$+ (-h(XY) - h(YZ) + h(Y) + h(XYZ)) \quad \} \leq 0 \text{ (submodularity)}$$
$$+ (-h(Y) - h(ZW) + h(YZW)) \quad \} \leq 0 \text{ (submodularity)}$$

Claim: # of target terms $\leq$ # of $\underbrace{\textit{unconditional}}_{h(\boldsymbol{A}|\emptyset) = h(\boldsymbol{A})}$ source terms

Identity $\Rightarrow$ Proof Sequence

$$
\begin{aligned}
h(XYZ) + h(YZW) = & h(XY) + h(YZ) + h(ZW) \\
& + (-h(XY) - h(YZ) + h(Y) + h(XYZ)) \\
& + (-h(Y) - h(ZW) + h(YZW))
\end{aligned}
$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW)$$
$$+ (-h(XY) \; -h(YZ) \; + h(Y) + h(XYZ))$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$
\begin{aligned}
h(XYZ) + h(YZW) = & h(XY) + h(YZ) + h(ZW) \\
& + (-h(XY) - h(YZ) + h(Y) + h(XYZ)) \\
& + (-h(Y) - h(ZW) + h(YZW))
\end{aligned}
$$

Identity $\Rightarrow$ Proof Sequence

$$
\begin{aligned}
h(XYZ) + h(YZW) = & h(XY) \boxed{+h(YZ)} + h(ZW) \\
& + (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ)) \\
& + (-h(Y) - h(ZW) + h(YZW))
\end{aligned}
$$

$$
\begin{aligned}
\boxed{h(YZ)} &\rightarrow h(Y) + h(Z|Y) \\
h(Z|Y) &\rightarrow h(Z|XY)
\end{aligned}
$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW)$$
$$+ (-h(XY) - h(YZ) + h(Y) + h(XYZ))$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$h(YZ) \rightarrow h(Y) + h(Z|Y)$

$h(Z|Y) \rightarrow h(Z|XY)$

$$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z|XY) + h(ZW)$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW)$$
$$+ (-h(XY) - h(YZ) + h(Y) + h(XYZ))$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$$h(YZ) \rightarrow h(Y) + h(Z|Y)$$
$$h(Z|Y) \rightarrow h(Z|XY)$$

$$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z|XY) + h(ZW)$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW)$$
$$+ (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ))$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(YZ)} \rightarrow h(Y) + h(Z|Y)$$
$$h(Z|Y) \rightarrow h(Z|XY)$$

$$h(XYZ) + h(YZW) = \boxed{h(XY)} + h(Y) + \boxed{h(Z|XY)} + h(ZW)$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW)$$
$$+ (-h(XY) - h(YZ) + h(Y) + h(XYZ))$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$$h(YZ) \rightarrow h(Y) + h(Z|Y)$$
$$h(Z|Y) \rightarrow h(Z|XY)$$

$$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z|XY) + h(ZW)$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$$h(XY) + h(Z|XY) \rightarrow h(XYZ)$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW) + (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ)) + (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(YZ)} \to h(Y) + h(Z|Y)$$

$$h(Z|Y) \to h(Z|XY)$$

$$h(XYZ) + h(YZW) = \boxed{h(XY)} + h(Y) + \boxed{h(Z|XY)} + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(XY)} + \boxed{h(Z|XY)} \to h(XYZ)$$

$$h(XYZ) + h(YZW) = h(XYZ) + h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW)$$
$$+ (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ))$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(YZ)} \rightarrow h(Y) + h(Z|Y)$$
$$h(Z|Y) \rightarrow h(Z|XY)$$

$$h(XYZ) + h(YZW) = \boxed{h(XY)} + h(Y) + \boxed{h(Z|XY)} + h(ZW)$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(XY)} + \boxed{h(Z|XY)} \rightarrow h(XYZ)$$

$$h(XYZ) + h(YZW) = \boxed{h(XYZ)} + h(Y) + h(ZW)$$
$$+ (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW) + (-h(XY) - h(YZ) + h(Y) + h(XYZ)) + (-h(Y) - h(ZW) + h(YZW))$$

$$h(YZ) \to h(Y) + h(Z|Y)$$

$$h(Z|Y) \to h(Z|XY)$$

$$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z|XY) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$$h(XY) + h(Z|XY) \to h(XYZ)$$

$$h(XYZ) + h(YZW) = h(XYZ) + h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW) + (-h(XY) - h(YZ) + h(Y) + h(XYZ)) + (-h(Y) - h(ZW) + h(YZW))$$

$$h(YZ) \to h(Y) + h(Z|Y)$$

$$h(Z|Y) \to h(Z|XY)$$

$$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z|XY) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$$h(XY) + h(Z|XY) \to h(XYZ)$$

$$h(XYZ) + h(YZW) = h(XYZ) + h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

–

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW) + (-h(XY) - h(YZ) + h(Y) + h(XYZ)) + (-h(Y) - h(ZW) + h(YZW))$$

$$h(YZ) \to h(Y) + h(Z|Y)$$
$$h(Z|Y) \to h(Z|XY)$$

$$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z|XY) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$$h(XY) + h(Z|XY) \to h(XYZ)$$

$$h(XYZ) + h(YZW) = h(XYZ) + h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$-$

$$h(YZW) = h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW) + (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ)) + (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(YZ)} \rightarrow h(Y) + h(Z|Y)$$

$$h(Z|Y) \rightarrow h(Z|XY)$$

$$h(XYZ) + h(YZW) = \boxed{h(XY)} + h(Y) + \boxed{h(Z|XY)} + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(XY)} + \boxed{h(Z|XY)} \rightarrow h(XYZ)$$

$$\boxed{h(XYZ)} + h(YZW) = \boxed{h(XYZ)} + h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

–

$$h(YZW) = \boxed{h(Y)} + h(ZW) + (\boxed{-h(Y)} - h(ZW) + h(YZW))$$

$$\boxed{h(Y)} \rightarrow h(Y|ZW)$$

Identity $\Rightarrow$ Proof Sequence

$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW)$ $+ (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ))$ $+ (-h(Y) - h(ZW) + h(YZW))$	$\boxed{h(YZ)} \rightarrow h(Y) + h(Z\|Y)$ $h(Z\|Y) \rightarrow h(Z\|XY)$
$h(XYZ) + h(YZW) = \boxed{h(XY)} + h(Y) + \boxed{h(Z\|XY)} + h(ZW)$ $+ (-h(Y) - h(ZW) + h(YZW))$	$\boxed{h(XY)} + \boxed{h(Z\|XY)} \rightarrow h(XYZ)$
$\boxed{h(XYZ)} + h(YZW) = \boxed{h(XYZ)} + h(Y) + h(ZW)$ $+ (-h(Y) - h(ZW) + h(YZW))$	–
$h(YZW) = \boxed{h(Y)} + h(ZW)$ $+ (\boxed{-h(Y)} - h(ZW) + h(YZW))$	$\boxed{h(Y)} \rightarrow h(Y\|ZW)$
$h(YZW) = \boxed{h(Y\|ZW)} + \boxed{h(ZW)}$	

Identity $\Rightarrow$ Proof Sequence

$$h(XYZ) + h(YZW) = h(XY) \boxed{+h(YZ)} + h(ZW) + (-h(XY) \boxed{-h(YZ)} + h(Y) + h(XYZ)) + (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(YZ)} \to h(Y) + h(Z|Y)$$

$$h(Z|Y) \to h(Z|XY)$$

$$h(XYZ) + h(YZW) = \boxed{h(XY)} + h(Y) + \boxed{h(Z|XY)} + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

$$\boxed{h(XY)} + \boxed{h(Z|XY)} \to h(XYZ)$$

$$\boxed{h(XYZ)} + h(YZW) = \boxed{h(XYZ)} + h(Y) + h(ZW) + (-h(Y) - h(ZW) + h(YZW))$$

–

$$h(YZW) = \boxed{h(Y)} + h(ZW) + (\boxed{-h(Y)} - h(ZW) + h(YZW))$$

$$\boxed{h(Y)} \to h(Y|ZW)$$

$$h(YZW) = \boxed{h(Y|ZW)} + \boxed{h(ZW)}$$

$$\boxed{h(ZW)} + \boxed{h(Y|ZW)} \to h(YZW)$$

Identity $\Rightarrow$ Proof Sequence

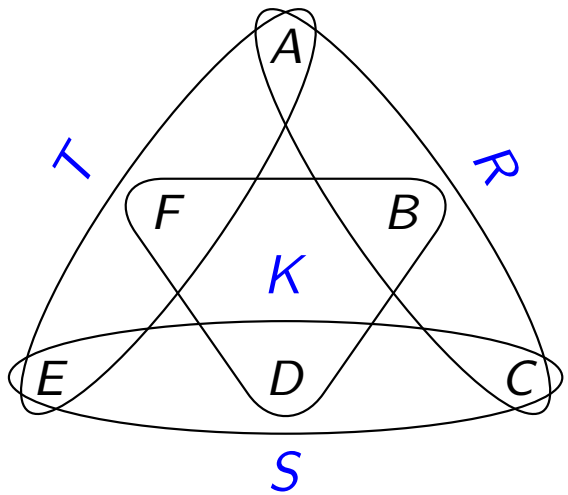
$h(XYZ) + h(YZW) = h(XY) + h(YZ) + h(ZW)$ $+ (-h(XY) - h(YZ) + h(Y) + h(XYZ))$ $+ (-h(Y) - h(ZW) + h(YZW))$	$h(YZ) \to h(Y) + h(Z\|Y)$ $h(Z\|Y) \to h(Z\|XY)$
$h(XYZ) + h(YZW) = h(XY) + h(Y) + h(Z\|XY) + h(ZW)$ $+ (-h(Y) - h(ZW) + h(YZW))$	$h(XY) + h(Z\|XY) \to h(XYZ)$
$h(XYZ) + h(YZW) = h(XYZ) + h(Y) + h(ZW)$ $+ (-h(Y) - h(ZW) + h(YZW))$	–
$h(YZW) = h(Y) + h(ZW)$ $+ (-h(Y) - h(ZW) + h(YZW))$	$h(Y) \to h(Y\|ZW)$
$h(YZW) = h(Y\|ZW) + h(ZW)$	$h(ZW) + h(Y\|ZW) \to h(YZW)$
$h(YZW) = h(YZW)$	–

Appendix: The Hexagon Query

The Hexagon Query [Wang 2022]

$$Q(A, B, C, D, E, F) \text{ :- } R(A, B, C) \wedge S(C, D, E) \wedge T(E, F, A) \wedge K(B, D, F)$$

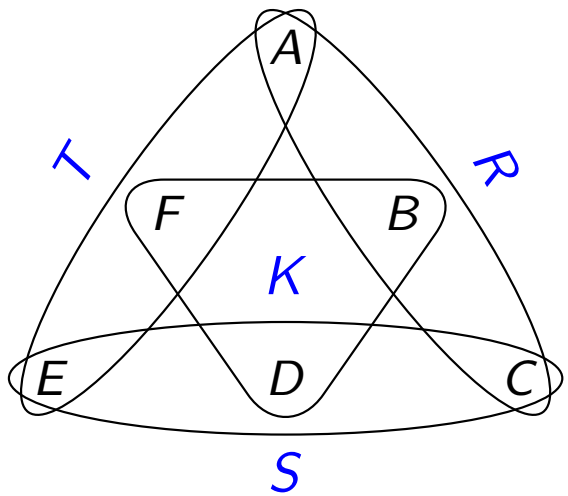
- AGM bound is N^2
- Any WCOJ algorithm can solve it in $O(N^2)$
- But **no heavy-light partitioning** algorithm



The Hexagon Query [Wang 2022]

$$Q(A,B,C,D,E,F) \text{ :- } R(A,B,C) \wedge S(C,D,E) \wedge T(E,F,A) \wedge K(B,D,F)$$

$$2h(ABCDEF) \leq h(ABC) + h(CDE) + h(EFA) + h(BDF)$$



$$h(ABCDEF) \leq h(ABC) + \underbrace{h(DE|C)}_{\text{light } C} + \underbrace{h(F)}_{\text{heavy } F}$$

$$h(ABCDEF) \leq h(EFA) + \underbrace{h(BD|F)}_{\text{light } F} + \underbrace{h(C)}_{\text{heavy } C}$$

But what about light C with light F, or heavy C with heavy F?!

$$Q(A,B,C,D,E,F) \quad :- \quad R(A,B,C) \wedge S(C,D,E) \wedge T(E,F,A) \wedge K(B,D,F)$$

$2h(ABCDEF) \leq$ $h(ABC) + h(CDE) + h(EFA) + h(BDF)$	$p_{ABC}(abc) \cdot p_{CDE}(cde) \cdot p_{EFA}(efa) \cdot p_{BDF}(bdf) \geq \frac{1}{N^4} = \frac{1}{B^2}$
$h(CDE) \to h(C) + h(DE\|C)$	$p_C(c) \stackrel{\text{def}}{=} \sum_{de} p_{CDE}(cde) = \frac{\deg_S(DE\|C=c)}{N}$ $p_{DE\|C}(de\|c) \stackrel{\text{def}}{=} \frac{p_{CDE}(cde)}{p_C(c)} = \frac{1}{\deg_S(DE\|C=c)}$
$h(ABC) + h(DE\|C) \to h(ABCDE)$	$p_{ABCDE}(abcde) \stackrel{\text{def}}{=} p_{ABC}(abc) \cdot p_{DE\|C}(de\|c) = \frac{1}{N \cdot \deg_S(DE\|C=c)}$
$h(BDF) \to h(F) + h(BD\|F)$	$p_F(f) \stackrel{\text{def}}{=} \sum_{bd} p_{BDF}(bdf) = \frac{\deg_K(BD\|F=f)}{N}$ $p_{BD\|F}(bd\|f) \stackrel{\text{def}}{=} \frac{p_{BDF}(bdf)}{p_F(f)} = \frac{1}{\deg_K(BD\|F=f)}$
$h(EFA) + h(BD\|F) \to h(ABDEF)$	$p_{ABDEF}(abdef) \stackrel{\text{def}}{=} p_{EFA}(efa) \cdot p_{BD\|F}(bd\|f) = \frac{1}{N \cdot \deg_K(BD\|F=f)}$
$h(ABCDE) + h(F) \to h(ABCDEF)$	$p'_{ABCDEF}(abcdef) \stackrel{\text{def}}{=} p_{ABCDE}(abcde) \cdot p_F(f) = \frac{\deg_K(BD\|F=f)}{N^2 \cdot \deg_S(DE\|C=c)}$
$h(ABDEF) + h(C) \to h(ABCDEF)$	$p''_{ABCDEF}(abcdef) \stackrel{\text{def}}{=} p_{ABDEF}(abdef) \cdot p_C(c) = \frac{\deg_S(DE\|C=c)}{N^2 \cdot \deg_K(BD\|F=f)}$

PANDAExpress on the Hexagon Query

$$Q(A,B,C,D,E,F) \text{ :- } R(A,B,C) \wedge S(C,D,E) \wedge T(E,F,A) \wedge K(B,D,F)$$

$$\underbrace{p'_{ABCDEF}(abcdef)}_{\frac{\deg_K(BD|F=f)}{N^2 \cdot \deg_S(DE|C=c)}} \cdot \underbrace{p''_{ABCDEF}(abcdef)}_{\frac{\deg_S(DE|C=c)}{N^2 \cdot \deg_K(BD|F=f)}} \geq \frac{1}{N^4} = \frac{1}{B^2}$$

For every $(a,b,c,d,e,f) \in \bowtie$

either $p'_{ABCDEF}(abcdef) \geq \frac{1}{B} \Leftrightarrow \deg_K(BD|F=f) \geq \deg_S(DE|C=c)$

or $p''_{ABCDEF}(abcdef) \geq \frac{1}{B} \Leftrightarrow \underbrace{\deg_K(BD|F=f) \leq \deg_S(DE|C=c)}_{\textbf{Hyperplane Partitioning}}$

Algorithm: Computed *"truncated"* p'_{ABCDEF} and p''_{ABCDEF} by $\geq \frac{1}{B}$