What do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog have to do with one another?

Mahmoud Abo Khamis[1] Hung Q. Ngo[1] Dan Suciu[1,2]

[1]LogicBlox Inc. [2]University of Washington

PODS 2017

Contributions

- A Join Algorithm

Contributions

- A Join Algorithm
  - first to meet the Submodular Width bound!

Contributions

- A Join Algorithm
  - first to meet the Submodular Width bound!
  - works for and relies on Disjunctive Datalog.

Contributions

- A Join Algorithm
  - first to meet the Submodular Width bound!
  - works for and relies on Disjunctive Datalog.
  - fully utilizes Functional DEPs and Degree Bounds.

Contributions

- A Join Algorithm
  - first to meet the Submodular Width bound!
  - works for and relies on Disjunctive Datalog.
  - fully utilizes Functional DEPs and Degree Bounds.
- A Unified Framework for Join Bounds

Contributions

- A Join Algorithm
  - first to meet the Submodular Width bound!
  - works for and relies on Disjunctive Datalog.
  - fully utilizes Functional DEPs and Degree Bounds.
- A Unified Framework for Join Bounds
  - subsumes most known bounds.

Contributions

- A Join Algorithm
  - first to meet the Submodular Width bound!
  - works for and relies on Disjunctive Datalog.
  - fully utilizes Functional DEPs and Degree Bounds.
- A Unified Framework for Join Bounds
  - subsumes most known bounds.
  - extends them to Functional DEPs and Degree Bounds.

Contributions

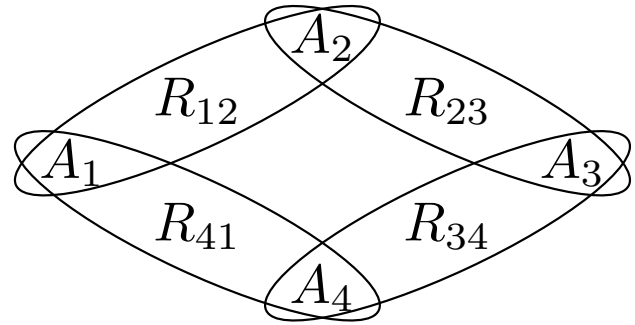
- A Join Algorithm
  - first to meet the Submodular Width bound!
  - works for and relies on Disjunctive Datalog.
  - fully utilizes Functional DEPs and Degree Bounds.
- A Unified Framework for Join Bounds
  - subsumes most known bounds.
  - extends them to Functional DEPs and Degree Bounds.
- Results on Shannon-type Inequalities

Table of Contents

Table of Contents

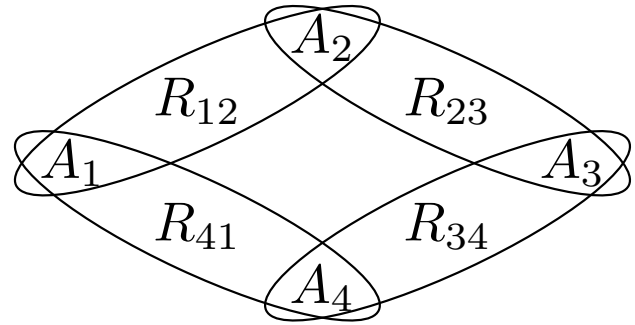
Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



A_1	A_2
a	1
b	1
b	2

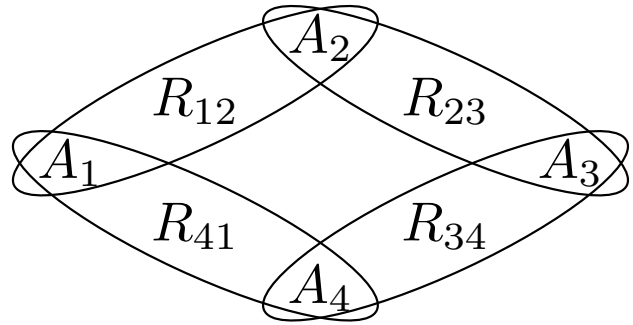
A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

A_1	A_2
a	1
b	1
b	2

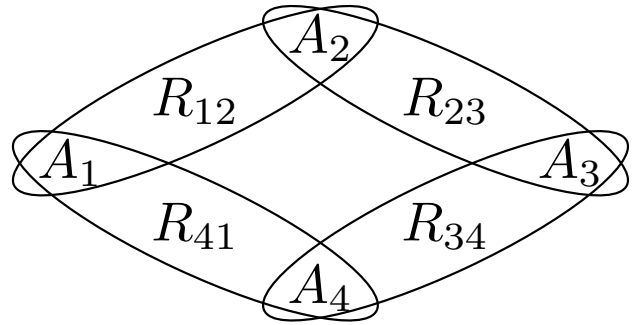
A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

A_1	A_2
a	1
b	1
b	2

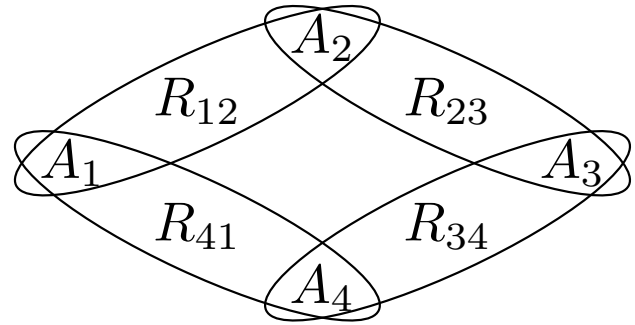
A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

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

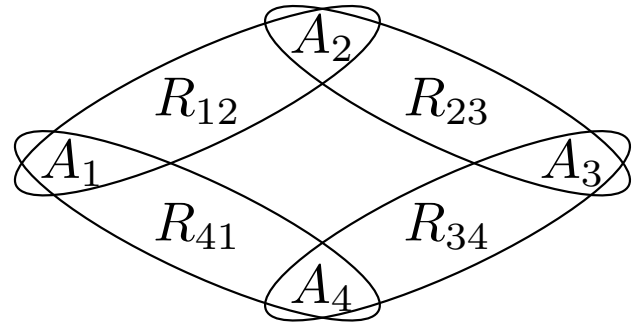
A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

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

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

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_1A_2A_3A_4) = \log |Q|$

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

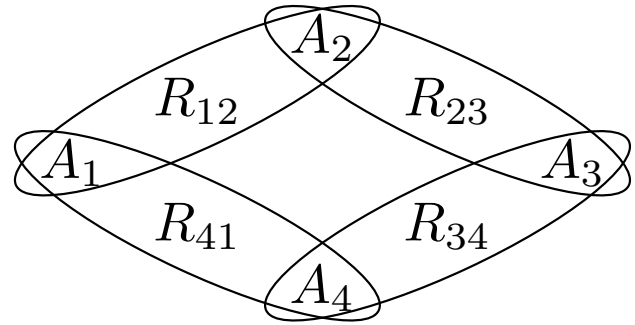
A_2	A_3	
1	c	1/4
1	d	2/4
2	c	1/4

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

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

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_1A_2A_3A_4) = \log|Q|$

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

$h(A_1A_2) \leq \log|R_{12}|,$

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

$h(A_2A_3) \leq \log|R_{23}|,$

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

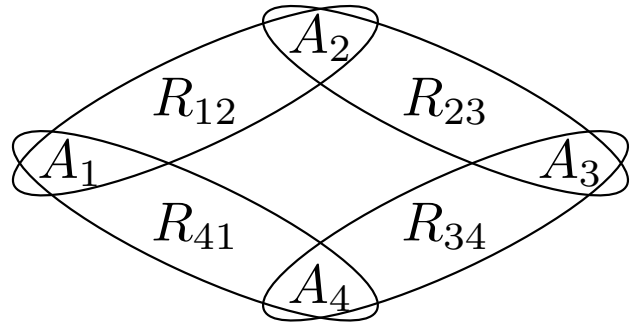
$h(A_3A_4) \leq \log|R_{34}|,$

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

$\ldots$

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_1A_2A_3A_4) = \log|Q|$

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

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

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

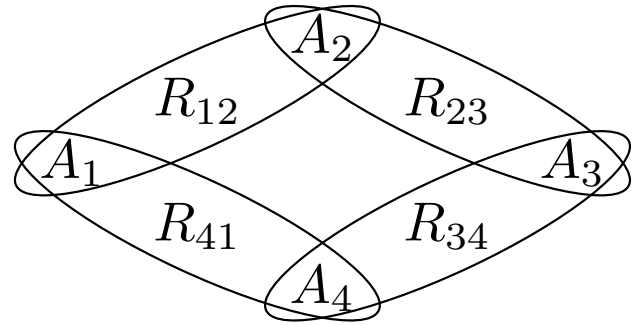
A_4	A_1	
3	b	2/4
4	a	1/4
4	b	1/4

$h(A_1A_2) \le \log|R_{12}|, \quad h(A_2A_3) \le \log|R_{23}|, \quad h(A_3A_4) \le \log|R_{34}|, \quad \ldots$

$h(A_2|A_1 = \text{‘a’}) \le \log|\sigma_{A_1=\text{‘a’}}R_{12}|, \quad h(A_2|A_1 = \text{‘b’}) \le \log|\sigma_{A_1=\text{‘b’}}R_{12}|, \quad \ldots$

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_1A_2A_3A_4) = \log|Q|$

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

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

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

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

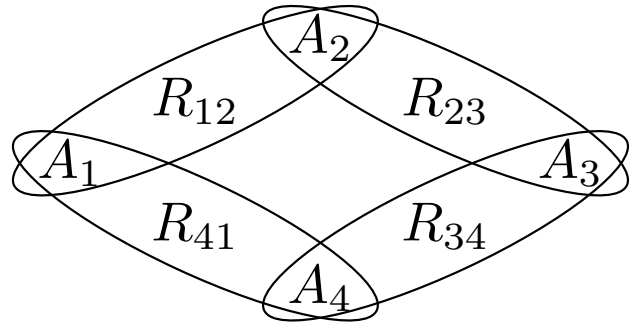
$h(A_1A_2) \le \log|R_{12}|, \quad h(A_2A_3) \le \log|R_{23}|, \quad h(A_3A_4) \le \log|R_{34}|, \quad \ldots$

$h(A_2|A_1 = \text{'a'}) \le \log|\sigma_{A_1=\text{'a'}}R_{12}|, \quad h(A_2|A_1 = \text{'b'}) \le \log|\sigma_{A_1=\text{'b'}}R_{12}|, \quad \ldots$

$h(A_2|A_1) \le \log \max_x |\sigma_{A_1=x}R_{12}|$

Size Bounds for Full Conjunctive Queries

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_1A_2A_3A_4) = \log|Q|$

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

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

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

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

$h(A_1A_2) \le \log|R_{12}|, \quad h(A_2A_3) \le \log|R_{23}|, \quad h(A_3A_4) \le \log|R_{34}|, \quad \ldots$

$h(A_2|A_1 = \text{'a'}) \le \log|\sigma_{A_1=\text{'a'}}R_{12}|, \quad h(A_2|A_1 = \text{'b'}) \le \log|\sigma_{A_1=\text{'b'}}R_{12}|, \quad \ldots$

$$h(A_2|A_1) \le \log \underbrace{\max_x |\sigma_{A_1=x}R_{12}|}_{\deg_{R_{12}}(A_2|A_1)}$$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$
  - Functional Dependencies (FD): $\mathbf{A}_X \to \mathbf{A}_Y$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$
  - Functional Dependencies (FD): $\mathbf{A}_X \to \mathbf{A}_Y$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$
  - Functional Dependencies (FD): $\mathbf{A}_X \to \mathbf{A}_Y$

Bound Idea

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$
  - Functional Dependencies (FD): $\mathbf{A}_X \to \mathbf{A}_Y$

Bound Idea

$\log |Q| \quad \leq \quad$ maximum $\quad h(A_1, \ldots, A_n)$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$
  - Functional Dependencies (FD): $\mathbf{A}_X \to \mathbf{A}_Y$

Bound Idea

$$\log |Q| \quad \leq \quad \text{maximum} \quad h(A_1, \ldots, A_n)$$
$$\text{over all} \quad \textit{entropies } h$$

Size Bounds: Input

- A full conjunctive query

$$Q(\mathbf{A}_{[n]}) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

  - $\mathbf{A}_{[n]} = \{A_1, \ldots, A_n\}$ is the full set of attributes
  - $\mathbf{A}_F = \{A_f \mid f \in F \subseteq [n]\}$ is a subset

- Degree Constraints (DC):

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

  - Cardinality Constraints (CC): $|R_F| \leq N_{F|\emptyset}$
  - Functional Dependencies (FD): $\mathbf{A}_X \to \mathbf{A}_Y$

Bound Idea

$$\log |Q| \quad \leq \quad \text{maximum} \quad h(A_1, \ldots, A_n)$$
$$\text{over all} \quad \textit{entropies } h$$
$$\text{such that} \quad h \textit{ satisfies} \text{ degree constraints of } Q$$

Size Bounds: Preliminaries

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ *satisfying* the degree constraints

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

Size Bounds: Preliminaries

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ *satisfying* the degree constraints

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

- Γ_n^* is the set of *entropic* functions

Size Bounds: Preliminaries

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ *satisfying* the degree constraints

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

- Γ_n^* is the set of *entropic* functions
- $\overline{\Gamma}_n^*$ is the topological closure of Γ_n^*

Size Bounds: Preliminaries

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ *satisfying* the degree constraints

$$\mathsf{HDC} \stackrel{\text{def}}{=} \left\{ h \quad | \quad h(Y|X) \le \log N_{Y|X}, \quad \forall (X, Y, N_{Y|X}) \right\}$$

- Γ_n^* is the set of *entropic* functions
- $\overline{\Gamma}_n^*$ is the topological closure of Γ_n^*
- Γ_n is the set of *polymatroids*, i.e. functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying

$$h(X \cup Y) + h(X \cap Y) \le h(X) + h(Y), \qquad X, Y \subseteq [n] \qquad \text{(submodularity)}$$
$$h(X) \le h(Y), \qquad X \subseteq Y \subseteq [n] \qquad \text{(monotonicity)}$$
$$h(\emptyset) = 0 \qquad \text{(strictness)}$$

Size Bounds: Preliminaries

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ *satisfying* the degree constraints

$$\mathsf{HDC} \stackrel{\text{def}}{=} \left\{ h \quad | \quad h(Y|X) \le \log N_{Y|X}, \quad \forall (X, Y, N_{Y|X}) \right\}$$

- Γ_n^* is the set of *entropic* functions
- $\overline{\Gamma}_n^*$ is the topological closure of Γ_n^*
- Γ_n is the set of *polymatroids*, i.e. functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying

$$\begin{aligned} h(X \cup Y) + h(X \cap Y) \le h(X) + h(Y), \quad & X, Y \subseteq [n] && \text{(submodularity)} \\ h(X) \le h(Y), \quad & X \subseteq Y \subseteq [n] && \text{(monotonicity)} \\ h(\emptyset) = 0 & && \text{(strictness)} \end{aligned}$$

$$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

Size Bounds for Full Conjunctive Queries

$$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

Size Bounds for Full Conjunctive Queries

$$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

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

Size Bounds for Full Conjunctive Queries

$$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

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

Size Bounds for Full Conjunctive Queries

$$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

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

Size Bounds for Full Conjunctive Queries

$$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

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

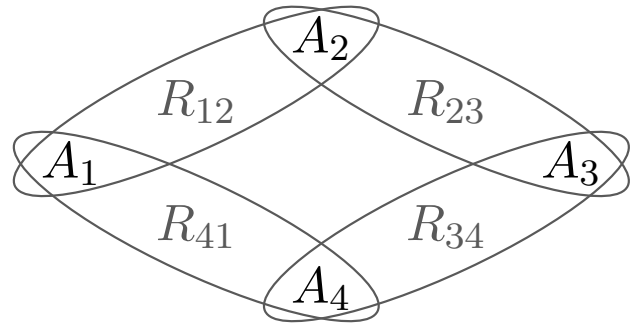
Table of Contents

Disjunctive Datalog

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

Disjunctive Datalog

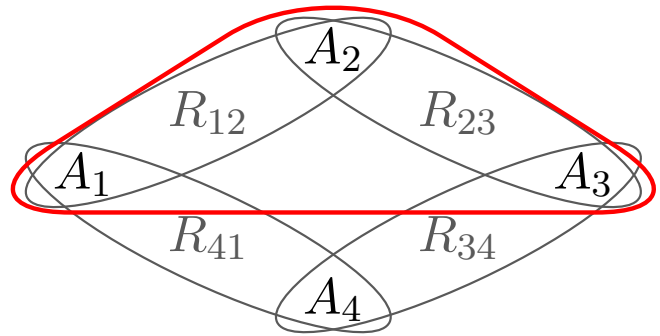
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

Disjunctive Datalog

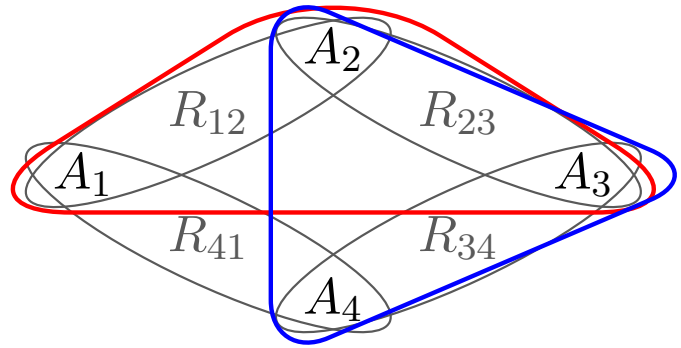
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

Disjunctive Datalog

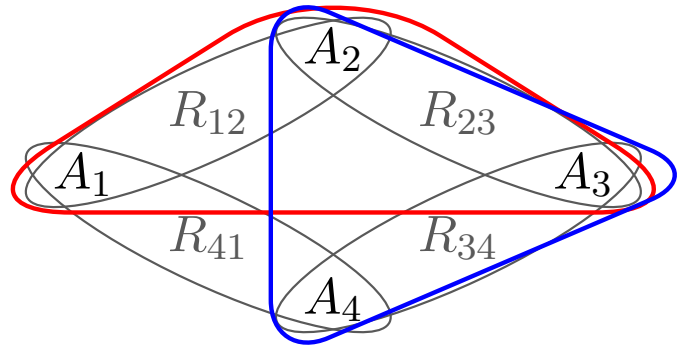
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

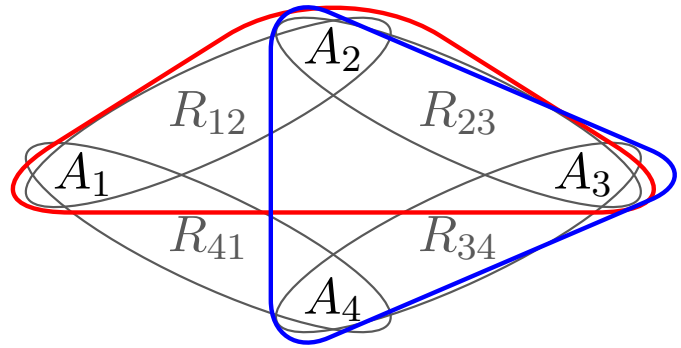
A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

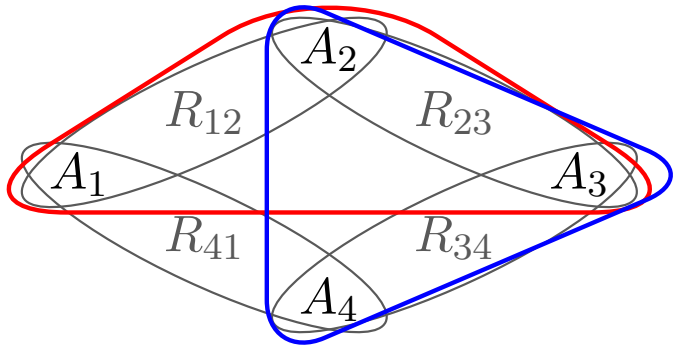
A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

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

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

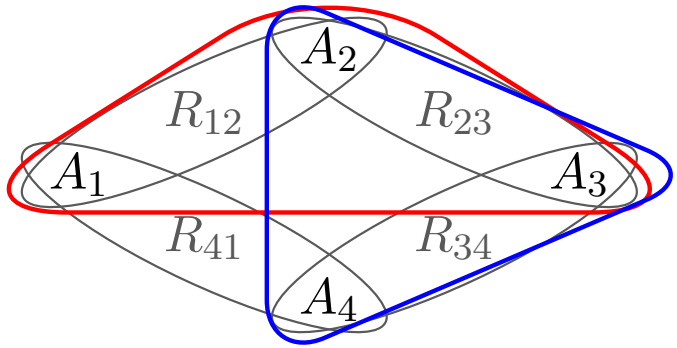
A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

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

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

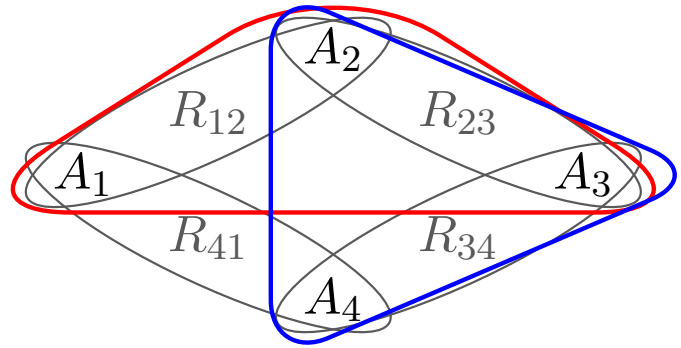
A_1	A_2	A_3	A_4
a	1	d	4
b	1	c	3
b	1	d	4
b	2	c	3

A_1	A_2	A_3
b	1	c
b	2	c

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

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

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

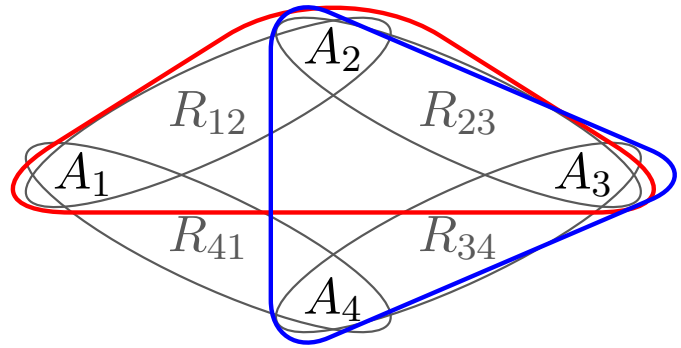
A_1	A_2	A_3
b	1	c
b	2	c

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

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

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

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

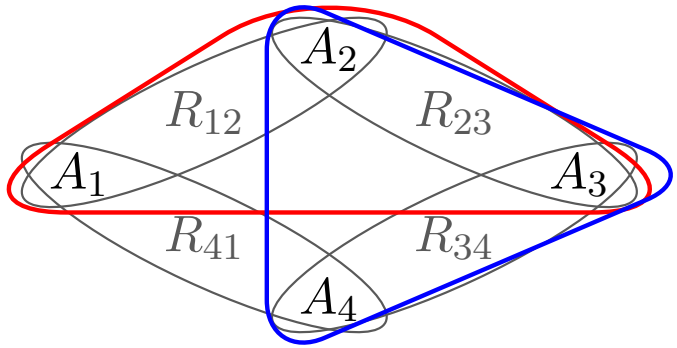
A_1	A_2	A_3
b	1	c
b	2	c

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

Output size is the *minimum* over all models

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

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

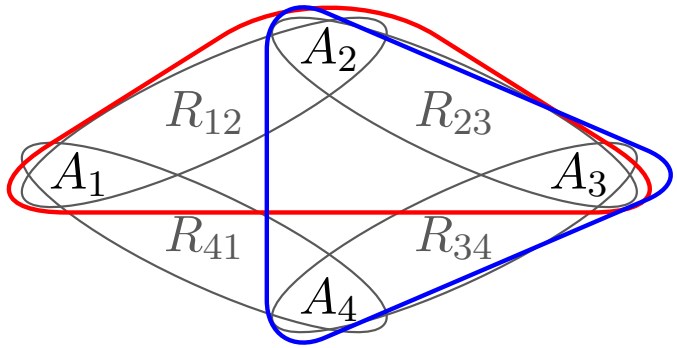
A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4

A minimum-sized model of size 2

Disjunctive Datalog

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



$$P: T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

A_1	A_2
a	1
b	1
b	2

A_2	A_3
1	c
1	d
2	c

A_3	A_4
c	3
d	4
d	5

A_4	A_1
3	b
4	a
4	b

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

A_1	A_2	A_3
b	1	c
b	2	c

A_2	A_3	A_4
1	d	4

A minimum-sized model of size 2

$\Rightarrow$ Output size is 2

Disjunctive Datalog: Output Size

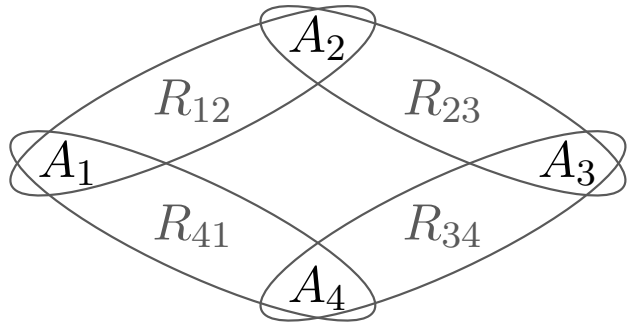
$$P: \quad \bigvee_{B \in \mathcal{B}} T_B(\mathbf{A}_B) \text{ :- } \bigwedge_{F \in \mathcal{E}} R_F(\mathbf{A}_F)$$

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

Disjunctive Datalog: Output Size

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

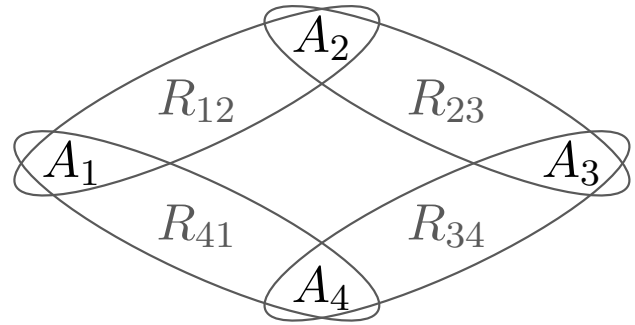
$$|P(\mathbf{D})| \stackrel{\text{def}}{=} \min_{\mathbf{T}:\mathbf{T}\models P} \max_{B\in\mathcal{B}} |T_B|$$



Disjunctive Datalog: Output Size

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

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

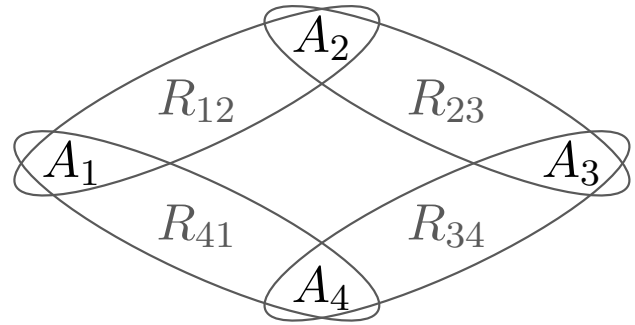


$$\mathbf{D}: \quad |R_{12}| \le N, \quad |R_{23}| \le N, \quad |R_{34}| \le N, \quad |R_{41}| \le N.$$

Disjunctive Datalog: Output Size

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

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



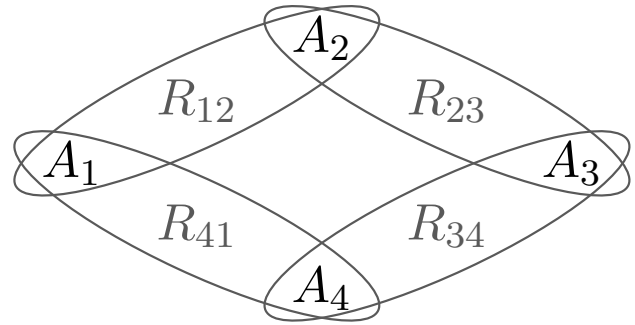
$$\mathbf{D}: \quad |R_{12}| \le N, \quad |R_{23}| \le N, \quad |R_{34}| \le N, \quad |R_{41}| \le N.$$

- $P: \; T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4)$:-
  $R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$

Disjunctive Datalog: Output Size

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

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



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

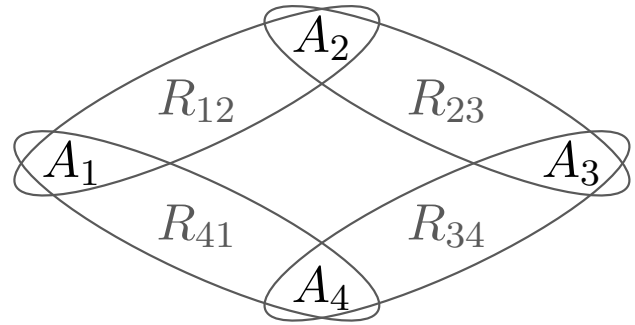
- $P: \; T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4)$:-
  $R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$

$|P(\mathbf{D})| \le N^{3/2}$, for all $\mathbf{D}$

Disjunctive Datalog: Output Size

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

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



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

- $P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4)$:-
  $R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$

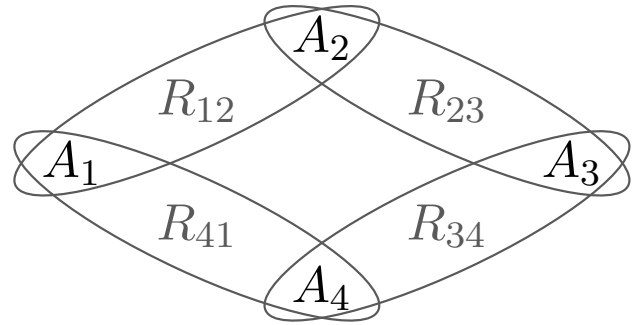
  $|P(\mathbf{D})| \le N^{3/2}$, for all $\mathbf{D}$

- $P': \quad T_{123}(A_1, A_2, A_3)$:-
  $R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$

Disjunctive Datalog: Output Size

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

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



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

- $P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4)$:-
  $R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$

  $|P(\mathbf{D})| \le N^{3/2}$, for all $\mathbf{D}$

- $P': \quad T_{123}(A_1, A_2, A_3)$:-
  $R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$

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

Disjunctive Datalog: Size Bounds

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

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

Disjunctive Datalog: Size Bounds

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

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

Recall that:

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying the degree constraints
- $$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

Disjunctive Datalog: Size Bounds

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

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

Recall that:

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying the degree constraints
- $$\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$$

Theorem

$\log |P(\mathbf{D})|$

Disjunctive Datalog: Size Bounds

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

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

Recall that:

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying the degree constraints
- $\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$

Theorem

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

Disjunctive Datalog: Size Bounds

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

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

Recall that:

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying the degree constraints
- $\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$

Theorem

$$\log|P(\mathbf{D})| \quad \leq \quad \underbrace{\max_{h\in\overline{\Gamma}_n^*\cap HDC} \min_{B\in\mathcal{B}} h(B)}_{\textit{entropic bound}} \quad \leq \quad \underbrace{\max_{h\in\Gamma_n\cap HDC} \min_{B\in\mathcal{B}} h(B)}_{\textit{polymatroid bound}}$$

Disjunctive Datalog: Size Bounds

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

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

Recall that:

- HDC is the set of functions $h : 2^{[n]} \to \mathbb{R}_+$ satisfying the degree constraints
- $\underbrace{\Gamma_n^*}_{\text{entropic functions}} \subset \underbrace{\overline{\Gamma}_n^*}_{\text{topological closure of } \Gamma_n^*} \subset \underbrace{\Gamma_n}_{\text{polymatroids}}$

Theorem

$$\log|P(\mathbf{D})| \quad \leq \quad \underbrace{\max_{h\in\overline{\Gamma}_n^*\cap HDC} \min_{B\in\mathcal{B}} h(B)}_{\substack{\textit{entropic bound} \\ \textit{(asymptotically tight!)}}} \quad \leq \quad \underbrace{\max_{h\in\Gamma_n\cap HDC} \min_{B\in\mathcal{B}} h(B)}_{\textit{polymatroid bound}}$$

Table of Contents

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

- An algorithm for disjunctive datalog

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

- An algorithm for disjunctive datalog
  - computes *a model*

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

- An algorithm for disjunctive datalog
  - computes *a model*
  - within the polymatroid bound:

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

- An algorithm for disjunctive datalog
  - computes *a model*
  - within the polymatroid bound:
    - the worst-case size of the *minimum model*.

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

- An algorithm for disjunctive datalog
  - computes *a model*
  - within the polymatroid bound:
    - the worst-case size of the *minimum model*.
- Outline

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

- An algorithm for disjunctive datalog
  - computes *a model*
  - within the polymatroid bound:
    - the worst-case size of the *minimum model*.
- Outline
  - Construct a **Proof Sequence** for the bound.

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

- An algorithm for disjunctive datalog
  - computes *a model*
  - within the polymatroid bound:
    - the worst-case size of the *minimum model*.
- Outline
  - Construct a **Proof Sequence** for the bound.
  - Interpret each proof step as an algorithmic step.

PANDA

- Polymatroid bound: $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$

PANDA

- Polymatroid bound: $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$
- $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap \mathsf{HDC}} \sum_{B \in \mathcal{B}} \lambda_B h(B)$

PANDA

- Polymatroid bound: $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$
- $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap \mathsf{HDC}} \sum_{B \in \mathcal{B}} \lambda_B h(B)$
- $\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot \underbrace{h(Y|X)}_{\substack{| \wedge \\ \log N_{Y|X}}}$

PANDA

- Polymatroid bound: $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$
- $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap \mathsf{HDC}} \sum_{B \in \mathcal{B}} \lambda_B h(B)$
- $\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot \underbrace{h(Y|X)}_{\substack{| \wedge \\ \log N_{Y|X}}}$
- Proof Sequence

  Given $X \subseteq Y$:

$$h(X) + h(Y|X) \to h(Y)$$

$$h(Y) \to h(X) + h(Y|X)$$

$$h(Y) \to h(X)$$

$$h(Y|X) \to h(Y \cup Z|X \cup Z)$$

PANDA

- Polymatroid bound: $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$
- $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap \mathsf{HDC}} \sum_{B \in \mathcal{B}} \lambda_B h(B)$
- $\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot \underbrace{h(Y|X)}_{\substack{| \wedge \\ \log N_{Y|X}}}$
- Proof Sequence

  Given $X \subseteq Y$:

$$h(X) + h(Y|X) \to h(Y) \qquad \text{(join)}$$
$$h(Y) \to h(X) + h(Y|X) \qquad \text{(data partition)}$$
$$h(Y) \to h(X) \qquad \text{(projection)}$$
$$h(Y|X) \to h(Y \cup Z|X \cup Z) \qquad \text{(nothing)}$$

- Algorithmic Sequence

PANDA

- Polymatroid bound: $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B)$
- $\max_{h \in \Gamma_n \cap \mathsf{HDC}} \min_{B \in \mathcal{B}} h(B) = \max_{h \in \Gamma_n \cap \mathsf{HDC}} \sum_{B \in \mathcal{B}} \lambda_B h(B)$
- $\sum_{B \in \mathcal{B}} \lambda_B \cdot h(B) \leq \sum_{(X,Y,N_{Y|X})} \delta_{Y|X} \cdot \underbrace{h(Y|X)}_{\substack{| \wedge \\ \log N_{Y|X}}}$
- Proof Sequence
  Given $X \subseteq Y$:

$$h(X) + h(Y|X) \to h(Y) \qquad \text{(join)}$$
$$h(Y) \to h(X) + h(Y|X) \qquad \text{(data partition)}$$
$$h(Y) \to h(X) \qquad \text{(projection)}$$
$$h(Y|X) \to h(Y \cup Z|X \cup Z) \qquad \text{(nothing)}$$

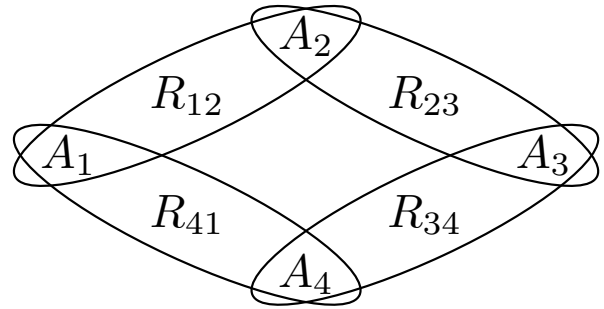
- Algorithmic Sequence

Theorem

PANDA solves any disjunctive datalog rule P in time within the polymatroid bound of P.

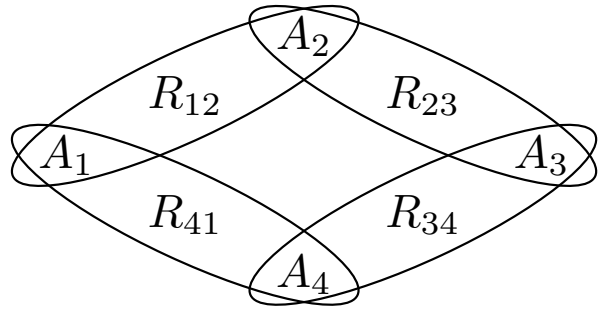
PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



PANDA: Example

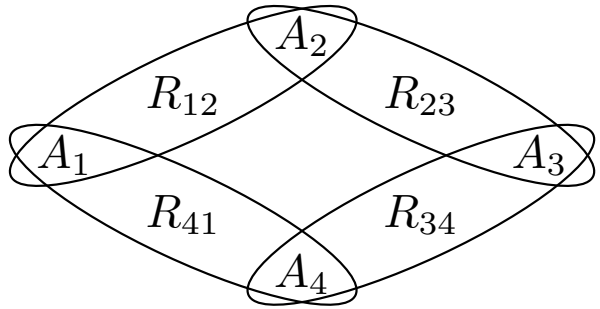
$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

PANDA: Example

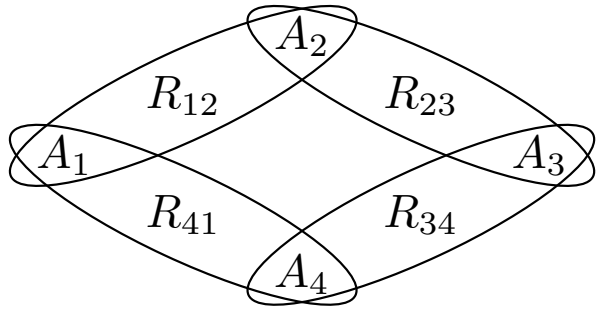
$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

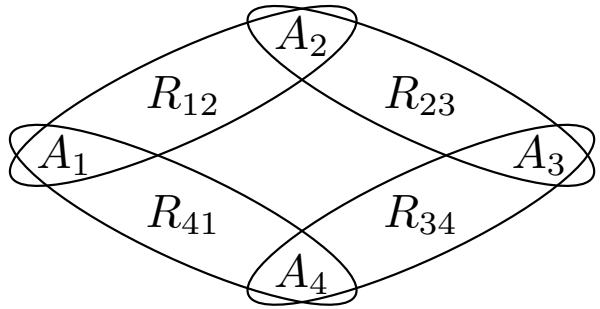


$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad |P| \leq N^{3/2}$

$\log |P| \quad \leq \quad \min(h(A_1A_2A_3), h(A_2A_3A_4))$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

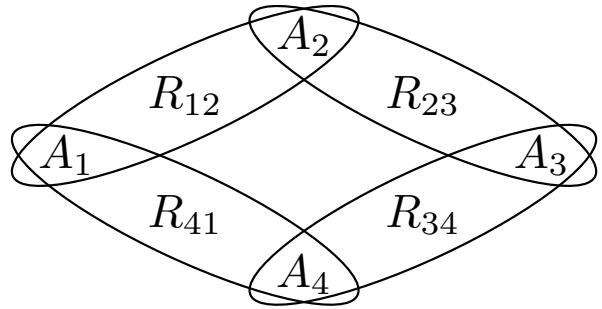


$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad \boxed{|P| \leq N^{3/2}}$$

$$\log|P| \quad \leq \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \leq \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



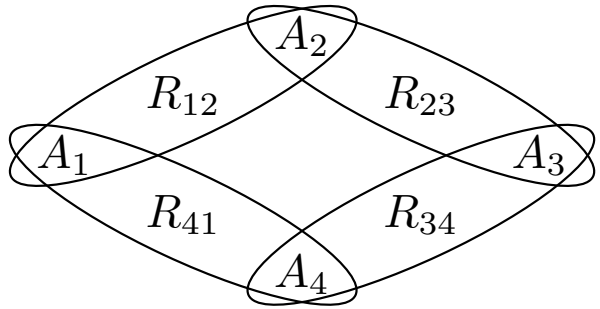
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \leq N \quad \Rightarrow \quad \boxed{|P| \leq N^{3/2}}$$

$$\log|P| \quad \leq \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \leq \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$

$$\boxed{\leq} \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

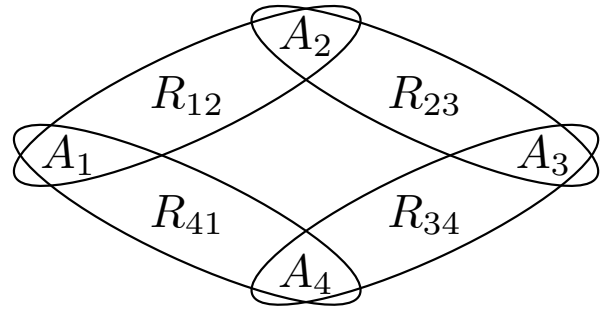


$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad \boxed{|P| \le N^{3/2}}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\boxed{\le} \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



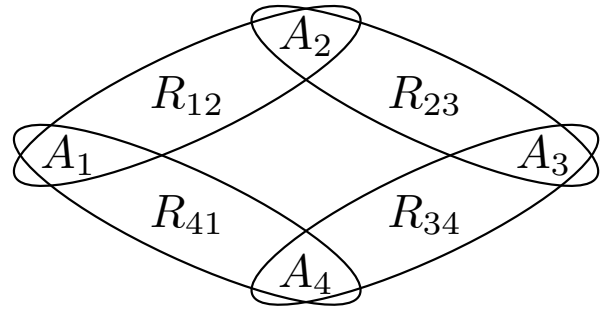
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$h(A_1A_2) + h(A_2A_3) + h(A_3A_4)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



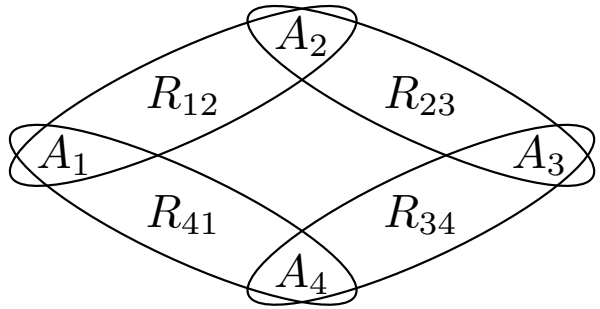
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad \to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



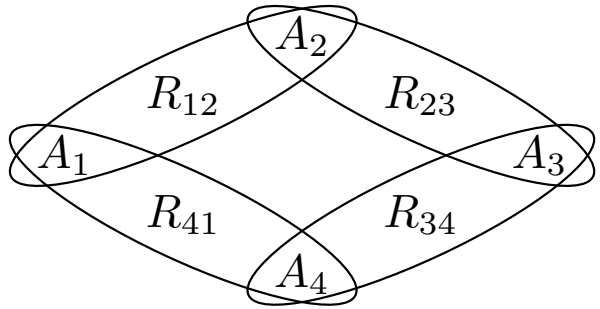
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad \to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)$$
$$h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



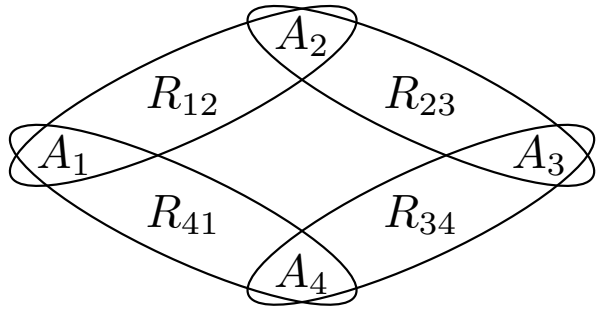
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad \boxed{|P| \le N^{3/2}}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\boxed{\le} \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad \to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)$$
$$h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) \quad \to \quad \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



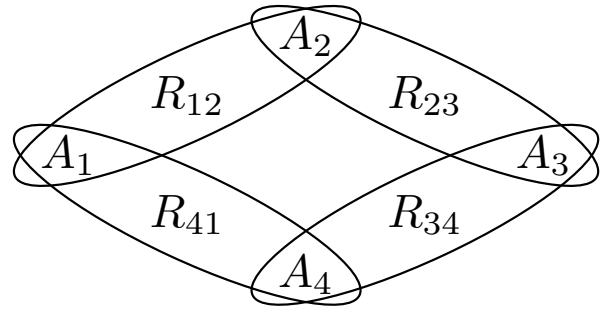
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad \to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)$$
$$h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) \quad \to \quad \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big)$$
$$h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3)$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



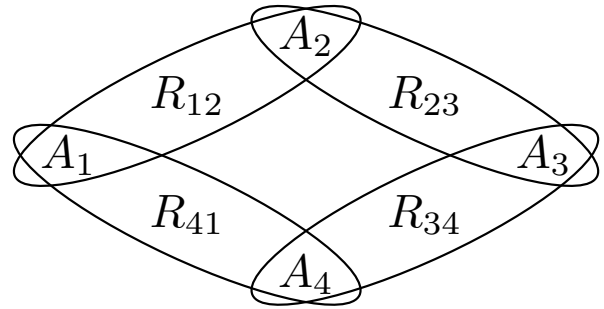
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad \boxed{|P| \le N^{3/2}}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\boxed{\le} \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$\begin{aligned}
h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad &\to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big) \\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) \quad &\to \quad \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big) \\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3) \quad &\to \quad \big(h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)\big)
\end{aligned}$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



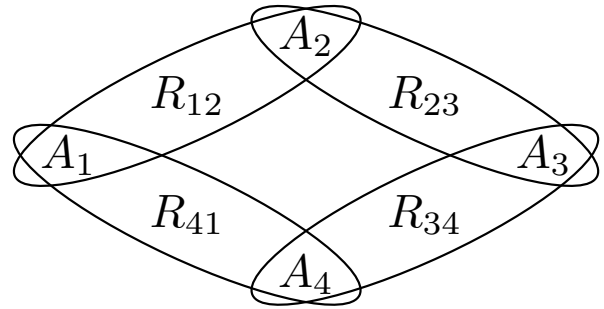
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad \boxed{|P| \le N^{3/2}}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$\begin{aligned}
h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad &\to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big) \\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) \quad &\to \quad \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big) \\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3) \quad &\to \quad \big(h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)\big) \\
h(A_1A_2) + h(A_2A_3A_4) + h(A_3)
\end{aligned}$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



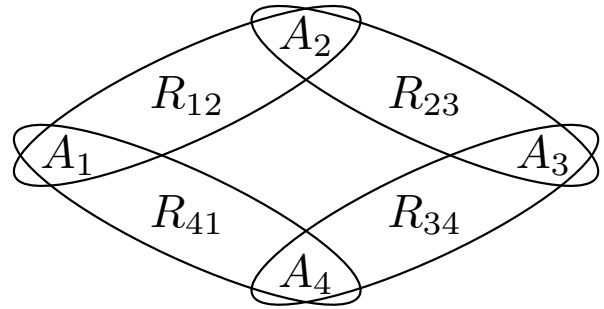
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad \boxed{|P| \le N^{3/2}}$$

$$\log|P| \quad \le \quad \min(h(A_1A_2A_3), h(A_2A_3A_4)) \quad \le \quad \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\boxed{\le} \quad \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \quad \le \quad \frac{3}{2}\log N$$

$$\begin{aligned}
h(A_1A_2) + h(A_2A_3) + h(A_3A_4) \quad &\to \quad \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big) \\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) \quad &\to \quad \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big) \\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3) \quad &\to \quad \big(h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)\big) \\
h(A_1A_2) + h(A_2A_3A_4) + h(A_3) \quad &\to \quad \big(h(A_1A_2) \to h(A_1A_2|A_3)\big)
\end{aligned}$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



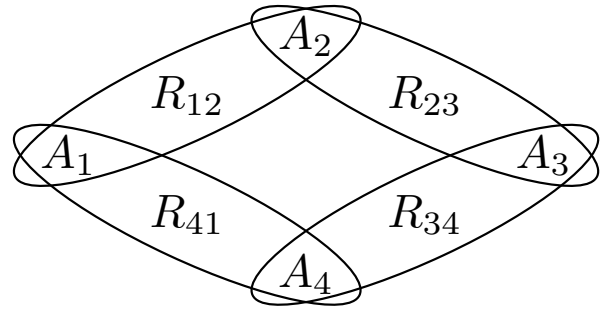
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \le \min(h(A_1A_2A_3), h(A_2A_3A_4)) \le \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \le \frac{3}{2}\log N$$

$$\begin{aligned}
h(A_1A_2) + h(A_2A_3) + h(A_3A_4) &\to \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)\\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) &\to \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big)\\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3) &\to \big(h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)\big)\\
h(A_1A_2) + h(A_2A_3A_4) + h(A_3) &\to \big(h(A_1A_2) \to h(A_1A_2|A_3)\big)\\
h(A_1A_2|A_3) + h(A_2A_3A_4) + h(A_3)
\end{aligned}$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



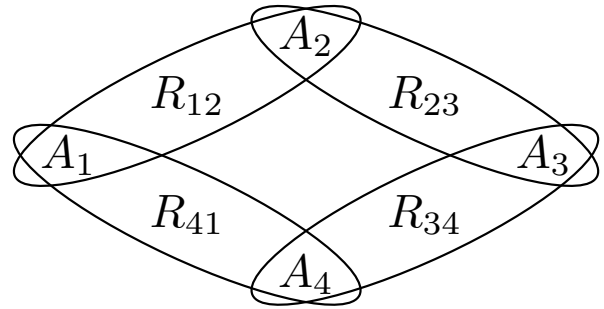
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \le \min(h(A_1A_2A_3), h(A_2A_3A_4)) \le \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \le \frac{3}{2}\log N$$

$$\begin{aligned}
h(A_1A_2) + h(A_2A_3) + h(A_3A_4) &\to \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)\\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) &\to \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big)\\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3) &\to \big(h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)\big)\\
h(A_1A_2) + h(A_2A_3A_4) + h(A_3) &\to \big(h(A_1A_2) \to h(A_1A_2|A_3)\big)\\
h(A_1A_2|A_3) + h(A_2A_3A_4) + h(A_3) &\to \big(h(A_1A_2|A_3) + h(A_3) \to h(A_1A_2A_3)\big)
\end{aligned}$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



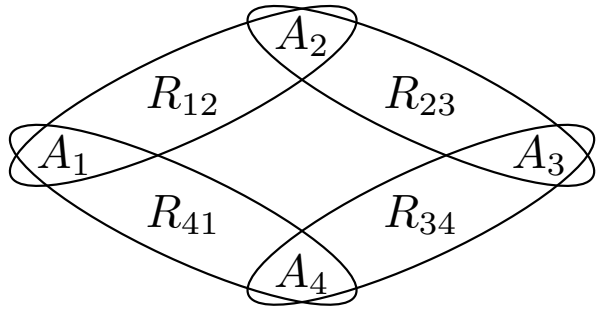
$$|R_{12}|, |R_{23}|, |R_{34}|, |R_{41}| \le N \quad \Rightarrow \quad |P| \le N^{3/2}$$

$$\log|P| \le \min(h(A_1A_2A_3), h(A_2A_3A_4)) \le \frac{1}{2}\big(h(A_1A_2A_3) + h(A_2A_3A_4)\big)$$
$$\le \frac{1}{2}\big(h(A_1A_2) + h(A_2A_3) + h(A_3A_4)\big) \le \frac{3}{2}\log N$$

$$\begin{aligned}
h(A_1A_2) + h(A_2A_3) + h(A_3A_4) &\to \big(h(A_3A_4) \to h(A_4|A_3) + h(A_3)\big)\\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_3) + h(A_3) &\to \big(h(A_4|A_3) \to h(A_4|A_2A_3)\big)\\
h(A_1A_2) + h(A_2A_3) + h(A_4|A_2A_3) + h(A_3) &\to \big(h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)\big)\\
h(A_1A_2) + h(A_2A_3A_4) + h(A_3) &\to \big(h(A_1A_2) \to h(A_1A_2|A_3)\big)\\
h(A_1A_2|A_3) + h(A_2A_3A_4) + h(A_3) &\to \big(h(A_1A_2|A_3) + h(A_3) \to h(A_1A_2A_3)\big)\\
h(A_1A_2A_3) + h(A_2A_3A_4)
\end{aligned}$$

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

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

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

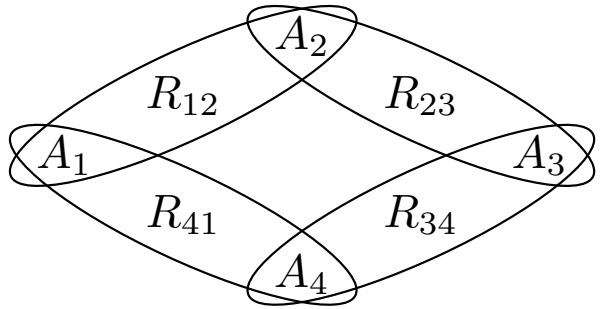
$h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)$

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

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

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_3A_4) \to h(A_4|A_3) + h(A_3)$ $\qquad$ $R_{34}(A_3, A_4) \to R_{34}^{(l)}(A_3, A_4), R_3^{(h)}(A_3)$

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

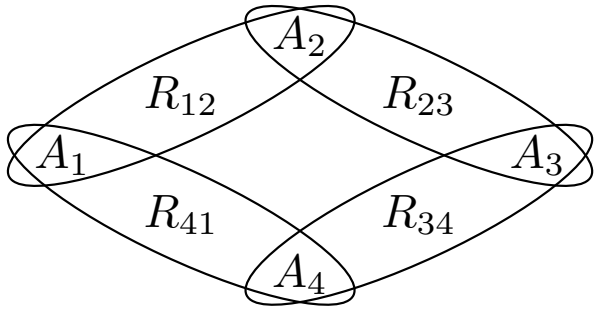
$h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)$

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

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

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

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

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

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

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

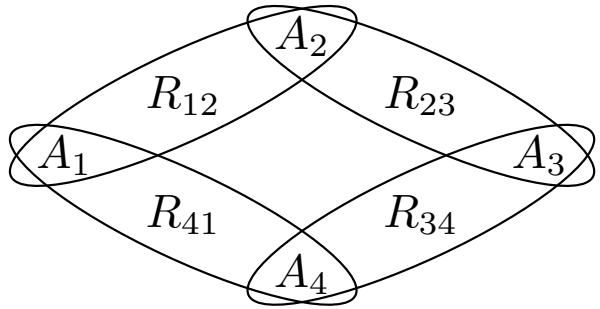
$h(A_1A_2|A_3) + h(A_3) \to h(A_1A_2A_3)$

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

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

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

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

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

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

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

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

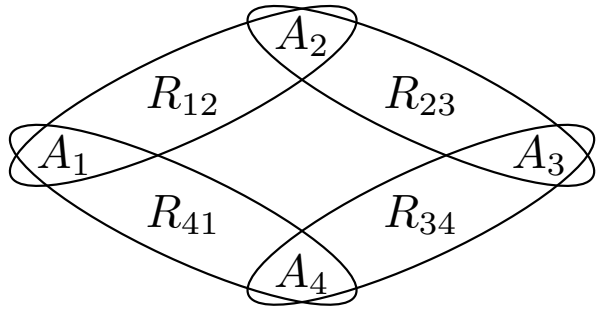
$R_{34}(A_3, A_4) \to R^{(l)}_{34}(A_3, A_4), R^{(h)}_{3}(A_3)$

$R^{(l)}_{34}(A_3, A_4) \to R^{(l)}_{34}(A_3, A_4)$

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

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

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

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

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

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

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

$R_{34}(A_3, A_4) \to R^{(l)}_{34}(A_3, A_4), R^{(h)}_{3}(A_3)$

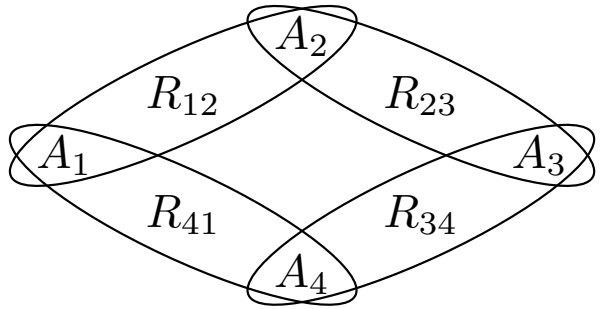
$R^{(l)}_{34}(A_3, A_4) \to R^{(l)}_{34}(A_3, A_4)$

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

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

PANDA: Example

$$P: \quad T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3),$$
$$R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



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

$h(A_3A_4) \to h(A_4|A_3) + h(A_3)$ $\qquad$ $R_{34}(A_3, A_4) \to R_{34}^{(l)}(A_3, A_4), R_3^{(h)}(A_3)$

$h(A_4|A_3) \to h(A_4|A_2A_3)$ $\qquad$ $R_{34}^{(l)}(A_3, A_4) \to R_{34}^{(l)}(A_3, A_4)$

$h(A_2A_3) + h(A_4|A_2A_3) \to h(A_2A_3A_4)$ $\qquad$ $R_{23}(A_2, A_3) \bowtie R_{34}^{(l)}(A_3, A_4) \to T_{234}(A_2, A_3, A_4)$

$h(A_1A_2) \to h(A_1A_2|A_3)$ $\qquad$ $R_{12}(A_1, A_2) \to R_{12}(A_1, A_2)$

$h(A_1A_2|A_3) + h(A_3) \to h(A_1A_2A_3)$ $\qquad$ $R_{12}(A_1, A_2) \bowtie R_3^{(h)}(A_3) \to T_{123}(A_1, A_2, A_3)$

Table of Contents

Beyond Worst-case Optimality

- Output-sensitive algorithms

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

- Submodular width as a candidate for d

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

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

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

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

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

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

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

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

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

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

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

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

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

- Our goals

Beyond Worst-case Optimality

- Output-sensitive algorithms

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

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

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

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

- Our goals

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

Submodular Width

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

Submodular Width

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

Submodular Width

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

Submodular Width

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

Submodular Width

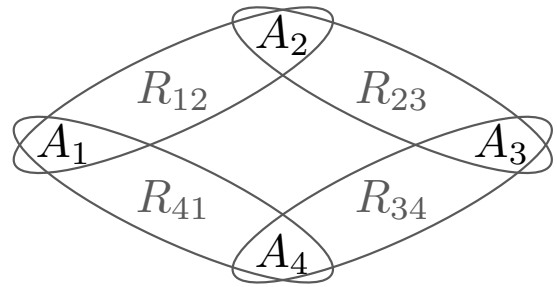
$$
\begin{aligned}
\mathsf{fhtw}(Q) &\stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)) \\
&= \min_{(T,\chi)} \max_{t \in V(T)} \max_{h \in \mathsf{ED} \cap \Gamma_n} h(\chi(t)) \\
&= \min_{(T,\chi)} \max_{h \in \mathsf{ED} \cap \Gamma_n} \max_{t \in V(T)} h(\chi(t)) \\
\mathsf{subw}(Q)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

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

Submodular Width: Example

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

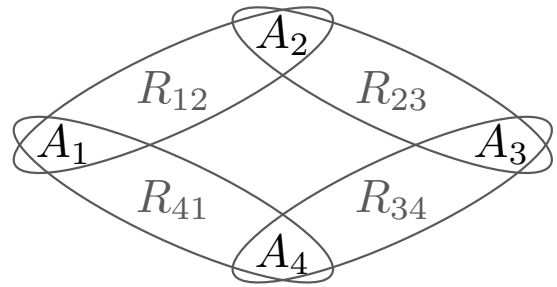
$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

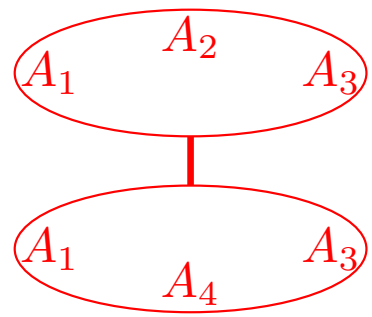


Submodular Width: Example

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

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

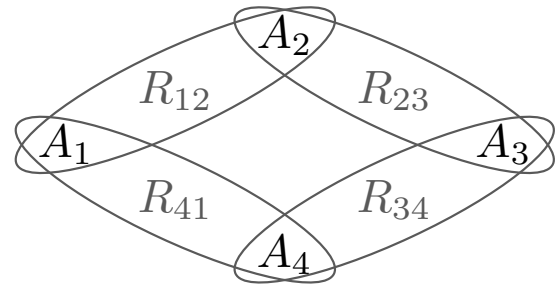


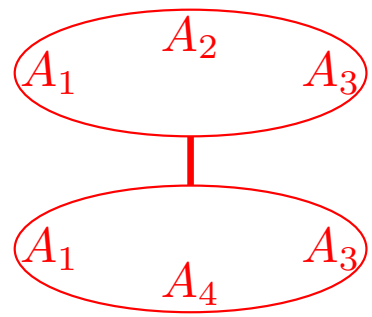


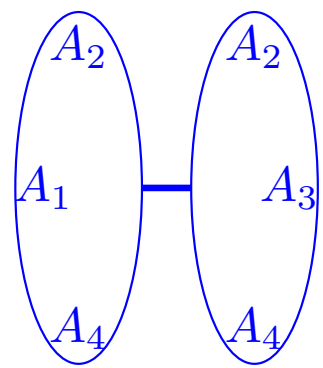
Submodular Width: Example

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

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$





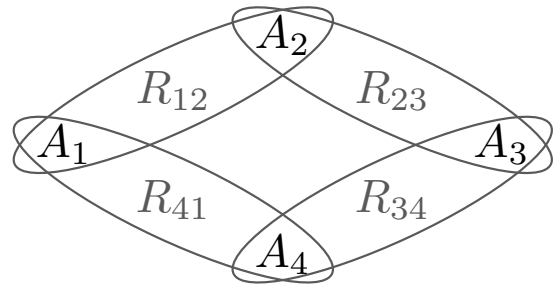


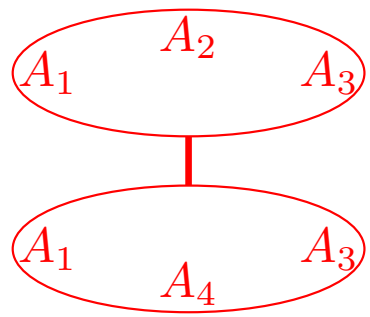
Submodular Width: Example

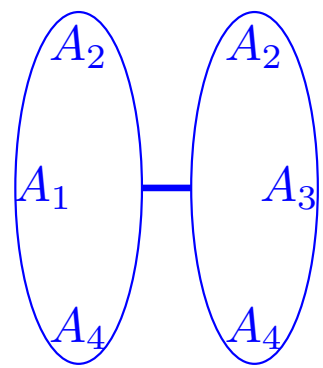
$$\mathsf{fhtw}(Q) \stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)), \qquad \mathsf{subw}(Q) \stackrel{\text{def}}{=} \max_{h \in \mathsf{ED} \cap \Gamma_n} \min_{(T,\chi)} \max_{t \in V(T)} h(\chi(t))$$

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

$$\mathsf{fhtw}(Q) \quad = \quad 2$$





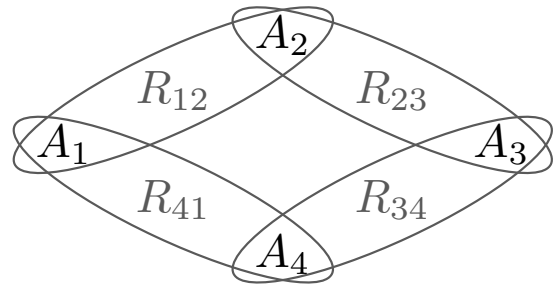


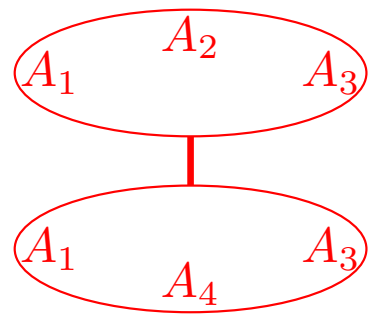
Submodular Width: Example

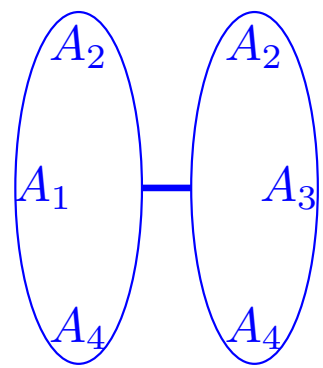
$$\mathsf{fhtw}(Q) \stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)), \qquad \mathsf{subw}(Q) \stackrel{\text{def}}{=} \max_{h \in \mathsf{ED} \cap \Gamma_n} \min_{(T,\chi)} \max_{t \in V(T)} h(\chi(t))$$

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

$$\begin{aligned} \mathsf{fhtw}(Q) &= 2 \\ \mathsf{subw}(Q) &= 3/2 \end{aligned}$$



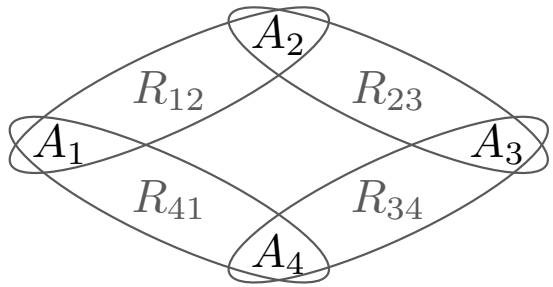




Submodular Width: Example

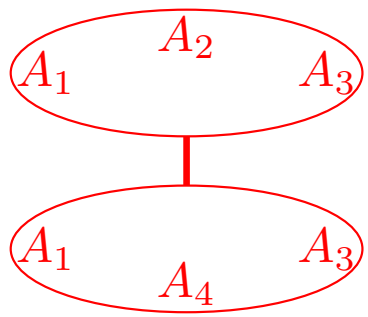
$$\mathsf{fhtw}(Q) \stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)), \qquad \mathsf{subw}(Q) \stackrel{\text{def}}{=} \max_{h \in \mathsf{ED} \cap \Gamma_n} \min_{(T,\chi)} \max_{t \in V(T)} h(\chi(t))$$

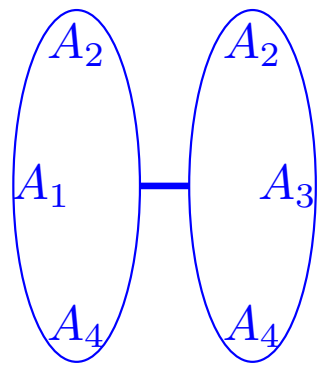
$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



$$\begin{aligned} \mathsf{fhtw}(Q) &= 2 \\ \mathsf{subw}(Q) &= 3/2 \end{aligned}$$

$$\min(\max(h(A_1A_2A_3), h(A_3A_4A_1)), \\ \max(h(A_4A_1A_2), h(A_2A_3A_4))) \leq 3/2$$

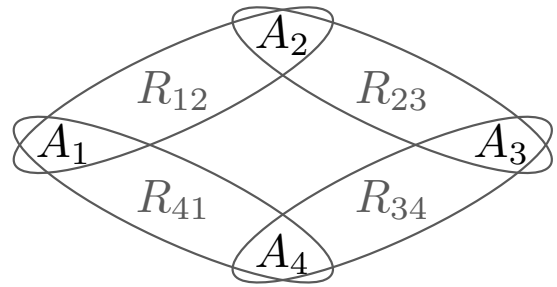




Submodular Width: Example

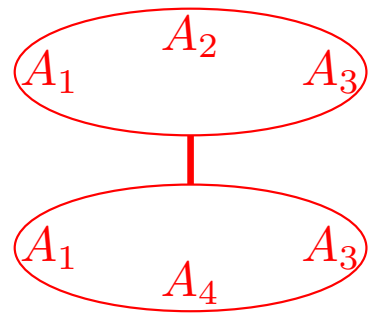
$$\mathsf{fhtw}(Q) \stackrel{\text{def}}{=} \min_{(T,\chi)} \max_{t \in V(T)} \rho^*(\chi(t)), \qquad \mathsf{subw}(Q) \stackrel{\text{def}}{=} \max_{h \in \mathsf{ED} \cap \Gamma_n} \min_{(T,\chi)} \max_{t \in V(T)} h(\chi(t))$$

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), \\ R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

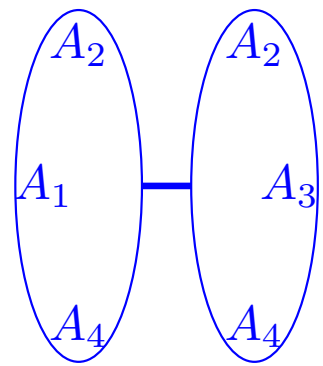


$$\begin{aligned} \mathsf{fhtw}(Q) &= 2 \\ \mathsf{subw}(Q) &= 3/2 \end{aligned}$$

$$\min(\max(h(A_1A_2A_3), h(A_3A_4A_1)), \\ \max(h(A_4A_1A_2), h(A_2A_3A_4))) \le 3/2$$



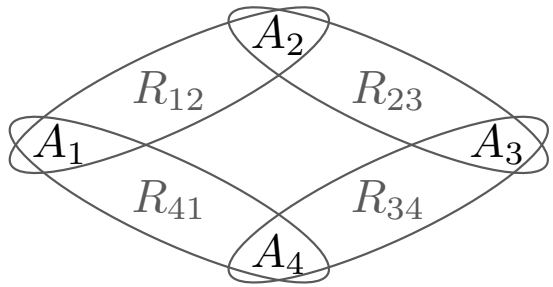
$$\begin{aligned} \min(h(A_1A_2A_3), h(A_4A_1A_2)) &\le 3/2 \\ \min(h(A_1A_2A_3), h(A_2A_3A_4)) &\le 3/2 \\ \min(h(A_3A_4A_1), h(A_4A_1A_2)) &\le 3/2 \\ \min(h(A_3A_4A_1), h(A_2A_3A_4)) &\le 3/2 \end{aligned}$$



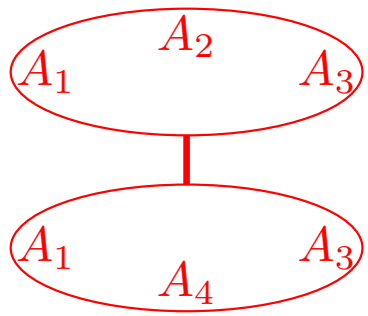
Submodular Width: Example

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

$$Q(A_1, A_2, A_3, A_4) \text{ :- } \quad R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

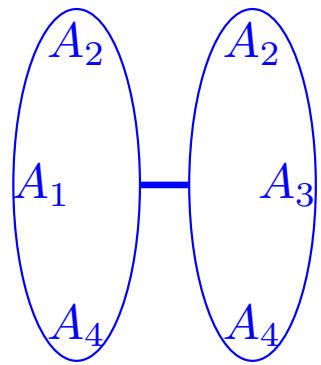


$$T_{123}(A_1, A_2, A_3) \vee T_{412}(A_4, A_1, A_2) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



$$T_{123}(A_1, A_2, A_3) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

$$T_{341}(A_3, A_4, A_1) \vee T_{412}(A_4, A_1, A_2) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$



$$T_{341}(A_3, A_4, A_1) \vee T_{234}(A_2, A_3, A_4) \text{ :-}$$
$$R_{12}(A_1, A_2), R_{23}(A_2, A_3), R_{34}(A_3, A_4), R_{41}(A_4, A_1).$$

Summary of Bounds