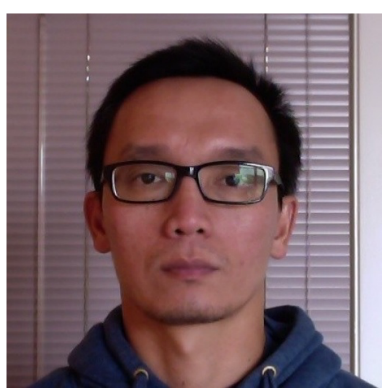
Optimal Query Processing Meets Information Theory

Dan Suciu – University of Washington

Mahmoud Abo-Khamis   Hung Ngo   – RelationalAI Inc.



[PODS'2016]
[PODS'2017]

Basic Question

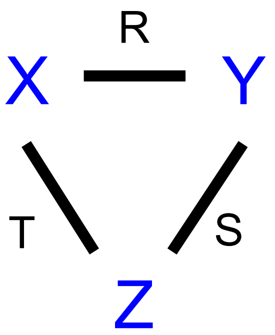
What is the optimal runtime to compute a query Q on a database D?

- Q, D are labeled hypergraphs
- Problem 1: list all occurrences in Q in D
  Problem 2: check if there exists Q in D
- Data complexity: Q is fixed, runtime = $f(D)$

Example Queries

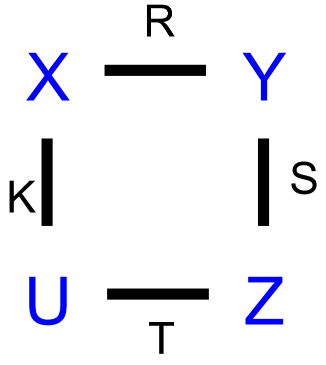
Enumerate all labeled triangles:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$



Check if there exists a labeled 4-cycle

$\exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



Main Results

Fix statistics for D
(cardinalities, functional dependencies, max degrees)
Fix the query Q

Problem 1: enumeration problem

Tight, but open if computable

Computable, but not tight

Thm ∀ D,
(1) |Q(D)| ≤ Entropic-bound ≤ Polymatroid-bound
(2) Q(D) computable in time Õ(Polymatroid-bound)

Problem 2: decision problem

Optimal?

Thm ∀D, Q(D) is computable in time Õ($2^{\text{submodular-width}}$)

Main Principle

- Find information-theoretic *proof* of the upper bound, or the submodular width
- Convert *proof* to *algorithm*

Outline

- Enumeration problem
- Decision problem
- Conclusions

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$
- S.Y is a key:

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$
- S.Y is a key: $|Q(D)| \leq N$

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$
- S.Y is a key: $|Q(D)| \leq N$
- S.Y has degree $\leq d$:

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$
- S.Y is a key: $|Q(D)| \leq N$
- S.Y has degree $\leq d$: $|Q(D)| \leq d \times N$

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$
- S.Y is a key: $|Q(D)| \leq N$
- S.Y has degree $\leq d$: $|Q(D)| \leq d \times N$

E.g. $R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Maximum Output Size

$\max_{D \text{ satisfies stats}} (|Q(D)|)$

E.g. $R(X,Y) \wedge S(Y,Z)$, $|R|, |S| \leq N$

- No other info: $|Q(D)| \leq N^2$
- S.Y is a key: $|Q(D)| \leq N$
- S.Y has degree $\leq d$: $|Q(D)| \leq d \times N$

E.g. $R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

No other info: $|Q(D)| \leq N^{3/2}$

Background: Entropy, Polymatroid

Fix a set $\mathbf{X}=\{X_1,\ldots,X_k\}$ and a function $H: 2^{\mathbf{X}} \rightarrow R_+$

Def H is called _entropic_ if there exists random variables $\mathbf{X}$ s.t. $H(\mathbf{U})$ = entropy of $\mathbf{U}$, for $\mathbf{U} \subseteq \mathbf{X}$

Def H is a _polymatroid_ if

$H(\emptyset) = 0$

$H(\mathbf{V}) \geq H(\mathbf{U})$ for $\mathbf{U} \subseteq \mathbf{V}$

$H(\mathbf{U}) + H(\mathbf{V}) \geq H(\mathbf{U} \cap \mathbf{V}) + H(\mathbf{U} \cup \mathbf{V})$

Shannon inequalities

Every entropic function is a polymatroid

Converse fails for k≥4 [Zhang&Yeung'98]

Enumeration Problem

Fix a set of statistics for D (cardinalities, FDs, degrees)

Fix a query Q with variables $\mathbf{X}=\{X_1,\ldots,X_k\}$

Asymptotically tight, but open if computable

Theorem ∀D that satisfies the statistics

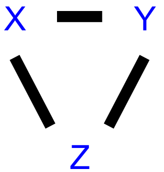
$$\log |Q(D)| \leq \max_{H \text{ entropic satisfying stats}} H(\mathbf{X})$$

$$\leq \max_{H \text{ polymatroid satisfying stats}} H(\mathbf{X})$$

Computable in EXPTIME, but not tight

Thm ∀D, Q(D) computable in time Õ(Polymatroid-bound)

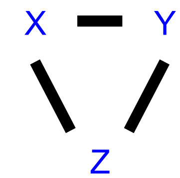
Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Database D → entropic function H

Proof of Upper Bound



Q(X,Y,Z) = R(X,Y) ∧ S(Y,Z) ∧ T(Z,X)

Database D → entropic function H

Database D

R(X,Y)

X	Y
a	3
a	2
b	2
d	3

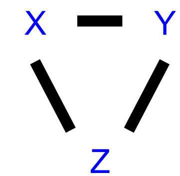
S(Y,Z)

Y	Z
3	m
2	q
3	q
2	m

T(Z,X)

Z	X
m	a
q	a
q	b
m	d

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Database D → entropic function H

Output Q(D)

X	Y	Z
a	3	m
a	2	q
b	2	q
d	3	m
a	3	q

Database D

R(X,Y)

X	Y
a	3
a	2
b	2
d	3

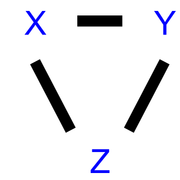
S(Y,Z)

Y	Z
3	m
2	q
3	q
2	m

T(Z,X)

Z	X
m	a
q	a
q	b
m	d

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Database D → entropic function H

Output Q(D)

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

Database D

R(X,Y)

X	Y
a	3
a	2
b	2
d	3

S(Y,Z)

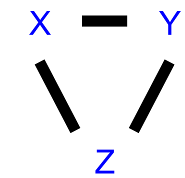
Y	Z
3	m
2	q
3	q
2	m

T(Z,X)

Z	X
m	a
q	a
q	b
m	d

$H(XYZ) = \log |Q(D)|$

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Database D → entropic function H

Output Q(D)

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

Database D

R(X,Y)

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S(Y,Z)

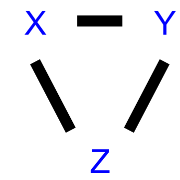
Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5
2	m	0

T(Z,X)

Z	X	
m	a	1/5
q	a	2/5
q	b	1/5
m	d	1/5

H(XYZ) = log |Q(D)|

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Database D → entropic function H

Output Q(D)

X	Y	Z	
a	3	m	1/5
a	2	q	1/5
b	2	q	1/5
d	3	m	1/5
a	3	q	1/5

$H(XYZ) = \log |Q(D)|$

Database D

R(X,Y)

X	Y	
a	3	2/5
a	2	1/5
b	2	1/5
d	3	1/5

S(Y,Z)

Y	Z	
3	m	2/5
2	q	2/5
3	q	1/5
2	m	0

T(Z,X)

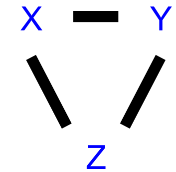
Z	X	
m	a	1/5
q	a	2/5
q	b	1/5
m	d	1/5

$H(XY) \leq \log N_R$ $H(YZ) \leq \log N_S$ $H(XZ) \leq \log N_T$

$H(Z|Y) \leq \log \mathrm{deg}_S(z|y)$

Cardinalitites, functional dependences, max degrees

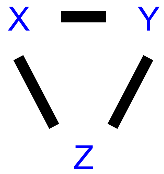
<u>Proof</u> of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$|R|,|S|,|T| \leq N \quad \rightarrow \quad |Q(D)| \leq N^{3/2}$

Proof of Upper Bound

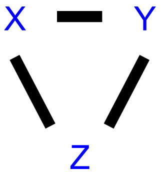


$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$|R|,|S|,|T| \leq N \quad \rightarrow \quad |Q(D)| \leq N^{3/2}$

$3 \log N \geq h(XY) + h(YZ) + h(XZ)$

Proof of Upper Bound



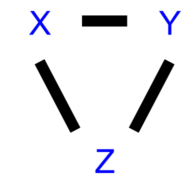
$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$|R|,|S|,|T| \leq N \quad \rightarrow \quad |Q(D)| \leq N^{3/2}$

submodularity

$$3 \log N \geq h(XY) + h(YZ) + h(XZ)$$

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

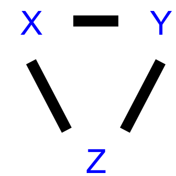
$|R|,|S|,|T| \leq N \quad \rightarrow \quad |Q(D)| \leq N^{3/2}$

submodularity

$3 \log N \geq h(XY) + h(YZ) + h(XZ)$

$\geq h(XYZ) + h(Y) + h(XZ)$

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$|R|,|S|,|T| \leq N \rightarrow |Q(D)| \leq N^{3/2}$

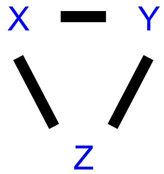
submodularity

$3 \log N \geq h(XY) + h(YZ) + h(XZ)$

submodularity

$\geq h(XYZ) + h(Y) + h(XZ)$

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$|R|,|S|,|T| \leq N \rightarrow |Q(D)| \leq N^{3/2}$

submodularity

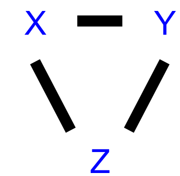
$3 \log N \geq h(XY) + h(YZ) + h(XZ)$

submodularity

$\geq h(XYZ) + h(Y) + h(XZ)$

$\geq h(XYZ) + h(XYZ) + h(\emptyset)$

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \land S(Y,Z) \land T(Z,X)$

$|R|,|S|,|T| \le N \quad \rightarrow \quad |Q(D)| \le N^{3/2}$

submodularity

$3 \log N \ge h(XY) + h(YZ) + h(XZ)$

submodularity

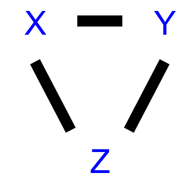
$\ge h(XYZ) + h(Y) + h(XZ)$

$\ge h(XYZ) + h(XYZ) + h(\emptyset)$

$= 2\, h(XYZ)$

$= 2 \log |Q(D)|$

Proof of Upper Bound



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$|R|,|S|,|T| \leq N \quad \rightarrow \quad |Q(D)| \leq N^{3/2}$

submodularity

$3 \log N \geq h(XY) + h(YZ) + h(XZ)$

submodularity

$\geq h(XYZ) + h(Y) + h(XZ)$

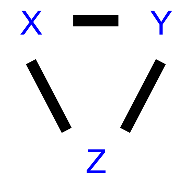
$\geq h(XYZ) + h(XYZ) + h(\emptyset)$

$= 2\, h(XYZ)$

Shearer's inequality
$h(XY) + h(YZ) + h(XZ) \geq 2\, h(XYZ)$

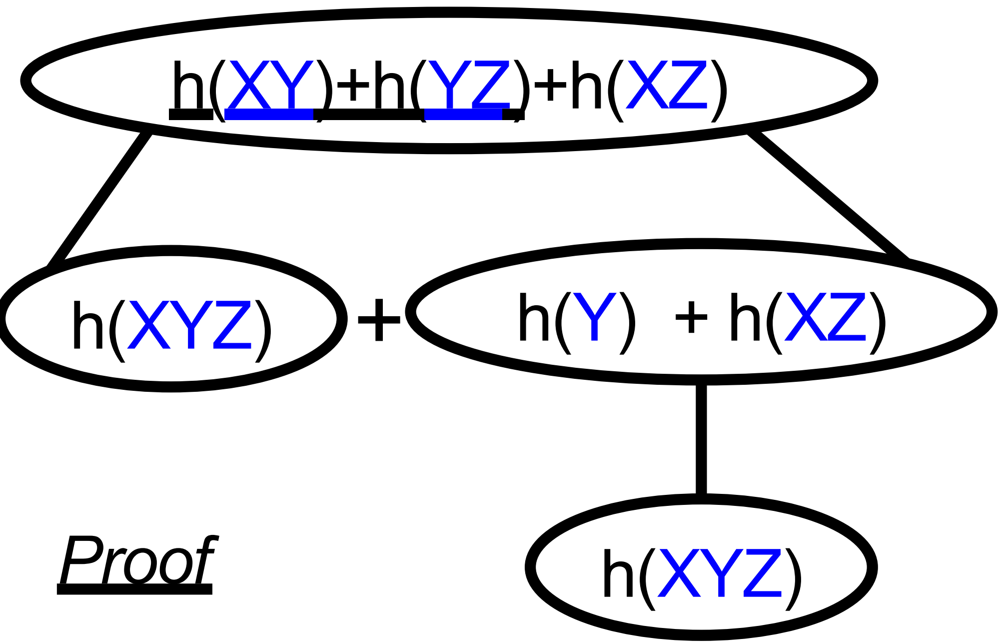
$= 2 \log |Q(D)|$

Proof to *Algorithm*



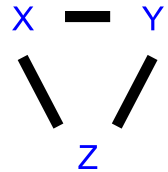
$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$h(XY) + h(YZ) + h(XZ) \geq 2\, h(XYZ)$



Proof

Proof to *Algorithm*

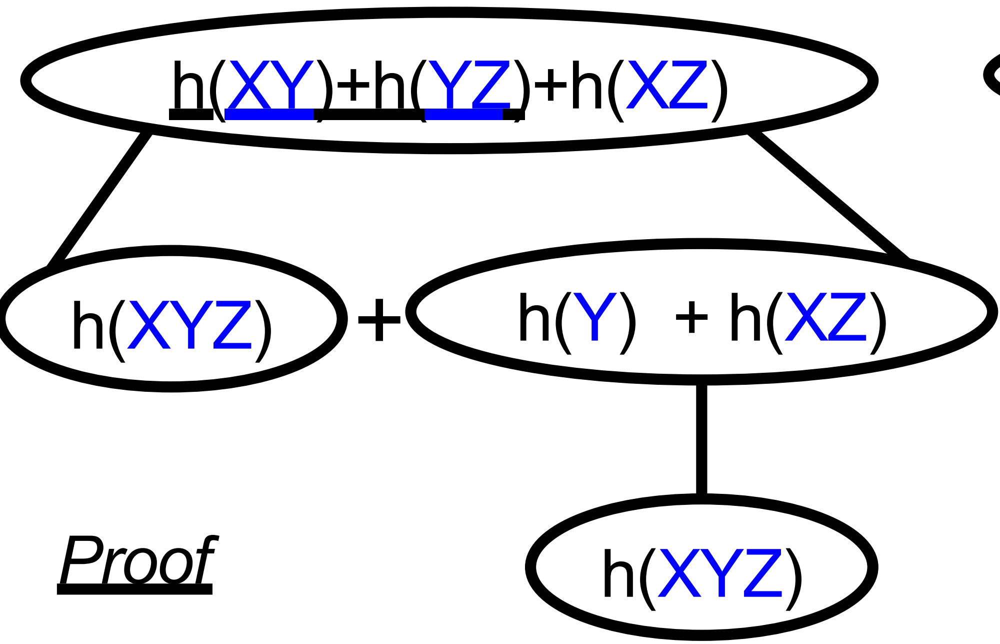


$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$h(XY) + h(YZ) + h(XZ) \geq 2\, h(XYZ)$

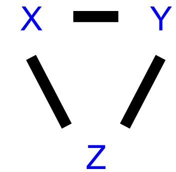
Algorithm

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$



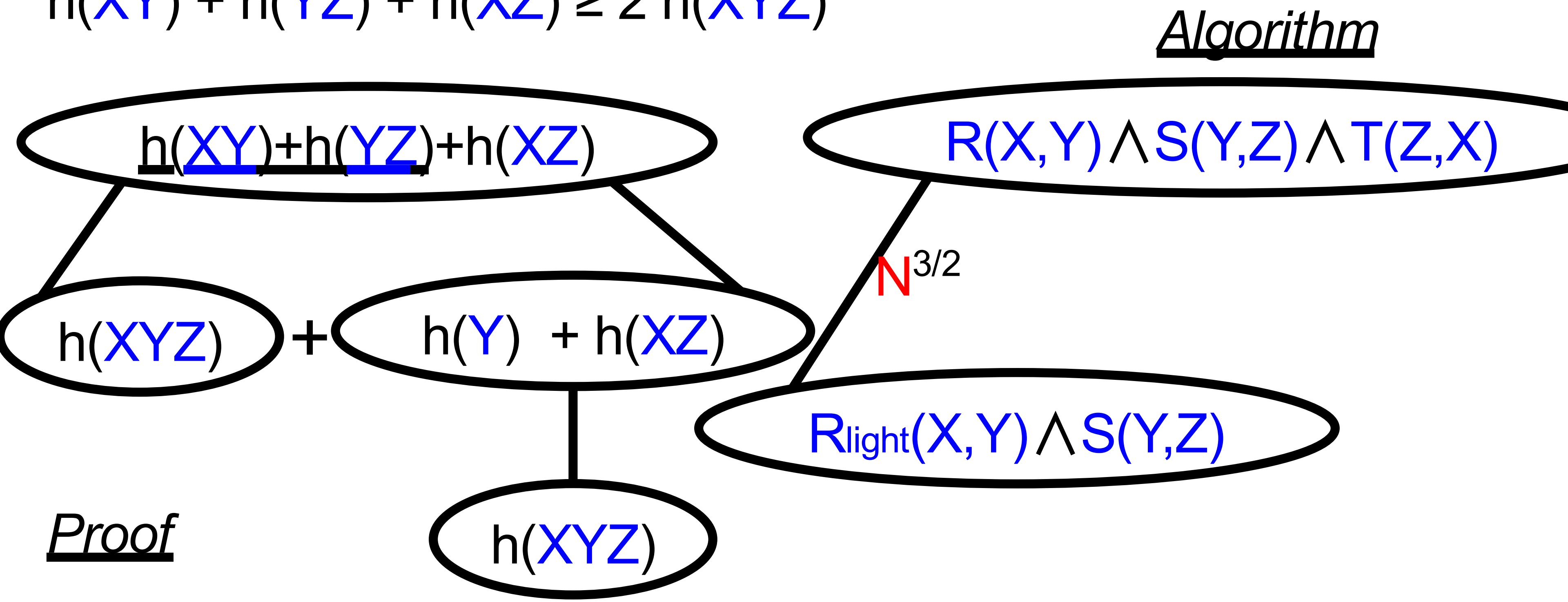
Proof

Proof to *Algorithm*



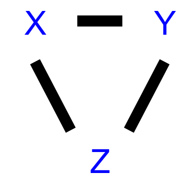
$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$h(XY) + h(YZ) + h(XZ) \geq 2\, h(XYZ)$



R_{light} or R_{heavy}: degree(Y) ≤ or > $N^{1/2}$

Proof to *Algorithm*



$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$h(XY) + h(YZ) + h(XZ) \geq 2\, h(XYZ)$

Proof

$h(XY)+h(YZ)+h(XZ)$

$h(XYZ)$ + $h(Y) + h(XZ)$

$h(XYZ)$

Algorithm

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$N^{3/2}$

$R_{light}(X,Y) \wedge S(Y,Z)$

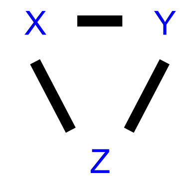
$\cup$

$N^{3/2}$

$R_{heavy}(Y) \wedge T(X,Z)$

R_{light} or R_{heavy}: degree(Y) $\leq$ or $>$ $N^{1/2}$

Proof to *Algorithm*

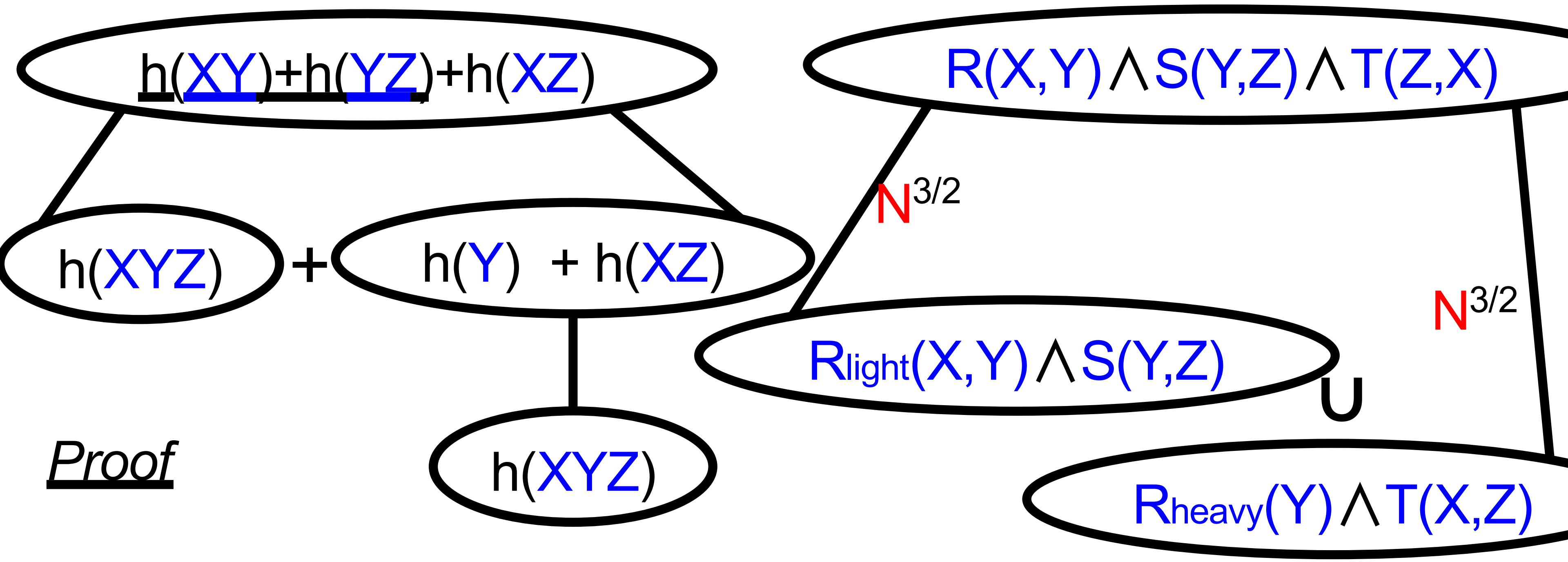


$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

Runtime $\tilde{O}(N^{3/2})$

$h(XY) + h(YZ) + h(XZ) \geq 2\, h(XYZ)$

Algorithm



R_{light} or R_{heavy}: degree$(Y) \leq$ or $> N^{1/2}$

Enumeration Problem: Discussion

Cardinalities: [Atserias,Grohe,Marx'08, Ngo,Re,Rudra'13]

- Entropic bound = polymatroid bound
- Algorithm for Q(D) has single log factor

Cardinalities + FDs + max degrees:

- Entropic bound $\lneqq$ polymatroid bound
- Algorithm for Q(D) has polylog factor

Outline

- Enumeration problem
- Decision problem
- Conclusions

Decision Problem

Fix Q, fix statistics on D
Problem: does Q occur in D?

"submodular width"

Theorem One can check if Q is in D in time $\tilde{O}(2^{\text{subw}(Q)})$

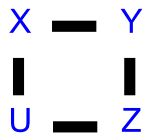
Optimal? (fine grained lower bound is open!)

Background: Tree Decomposition

Informally: TD = a tree where each node t represents an enumeration problem

- Fractional hypetree width [Grohe,Marx'14]
  $\min_{\text{tree}} \max_{\text{node } t} \max_{D}$
- Submodular width [Marx'2013]
  $\max_{D} \min_{\text{tree}} \max_{\text{node } t}$

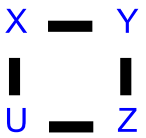
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

$\boxed{\min_{\text{tree}} \max_{\text{node } t} \max_{D}}$

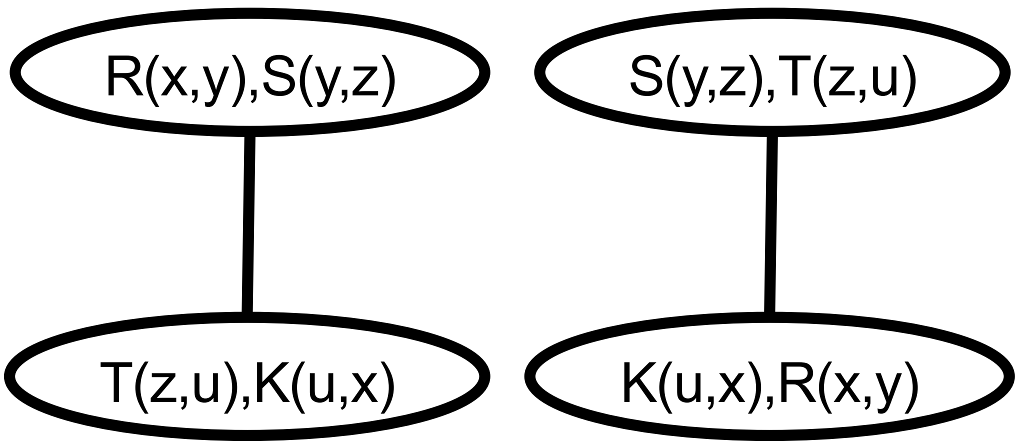
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



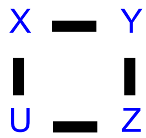
$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon, Yuster, Zwick'97]

$\min_{\text{tree}} \max_{\text{node } t} \max_{D}$

Tree decompositions



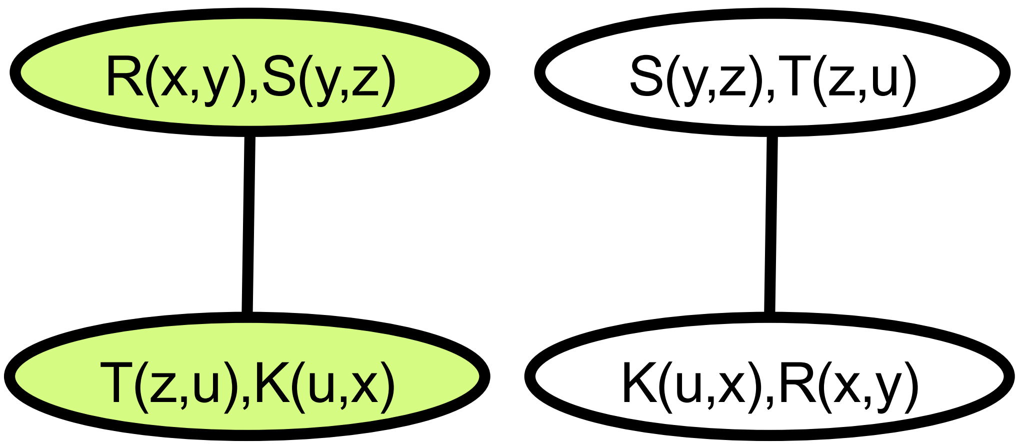
$Q() = \exists x \exists y \exists z \exists u \; R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

$\min_{\text{tree}} \max_{\text{node } t} \max_{D}$

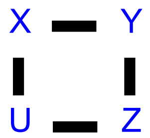
Tree decompositions



Runtime $\tilde{O}(N^2)$

(suboptimal)

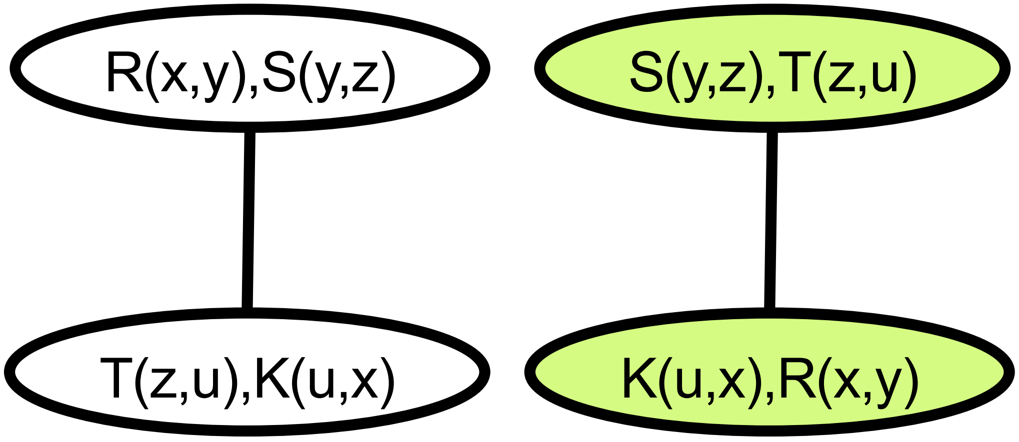
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|, |S|, |T|, |K| \leq N$ $O(N^{3/2})$ algorithm [Alon, Yuster, Zwick'97]

$\min_{tree} \max_{node\ t} \max_D$

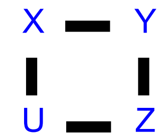
Tree decompositions



Runtime $\tilde{O}(N^2)$

(suboptimal)

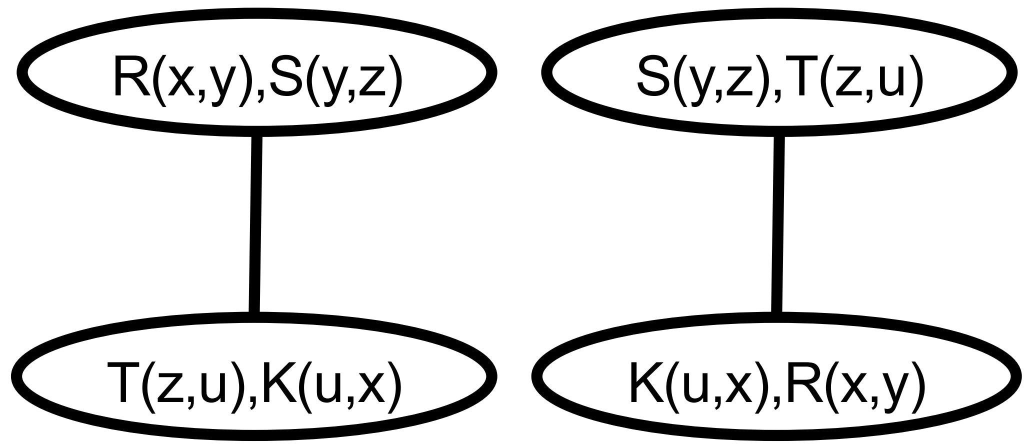
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



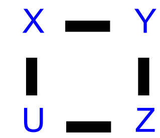
$|R|,|S|,|T|,|K| \le N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

$\boxed{\max_D \min_{tree} \max_{node\ t}}$

Tree decompositions



$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$

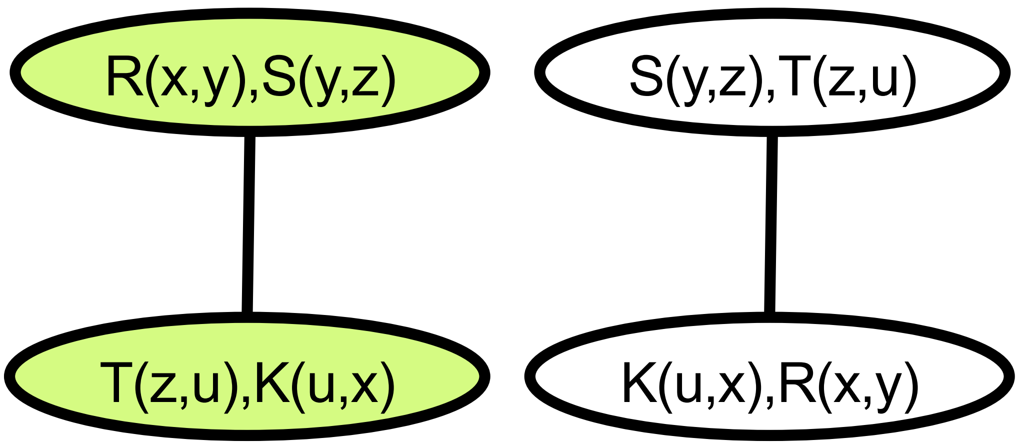


$|R|,|S|,|T|,|K| \leq N$ $\quad O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

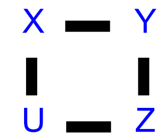
$\max_D \min_{tree} \max_{node\ t}$

$\overbrace{\max(h(xyz),h(zux))}^{T1}$

Tree decompositions



$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$

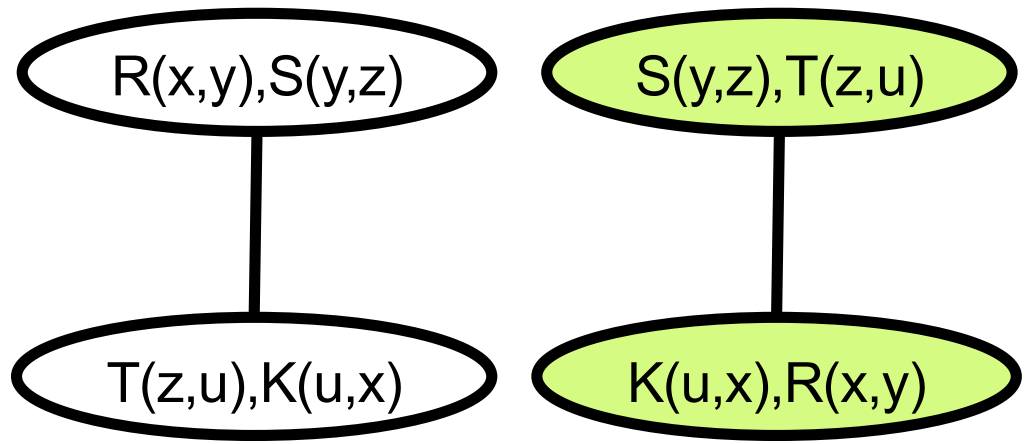


$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

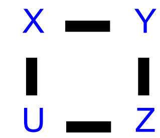
$\max_D \min_{tree} \max_{node\ t}$

$$\overbrace{\max(h(xyz),h(zux))}^{T1},$$
$$\underbrace{\max(h(yzu),h(uxy))}_{T2}$$

Tree decompositions



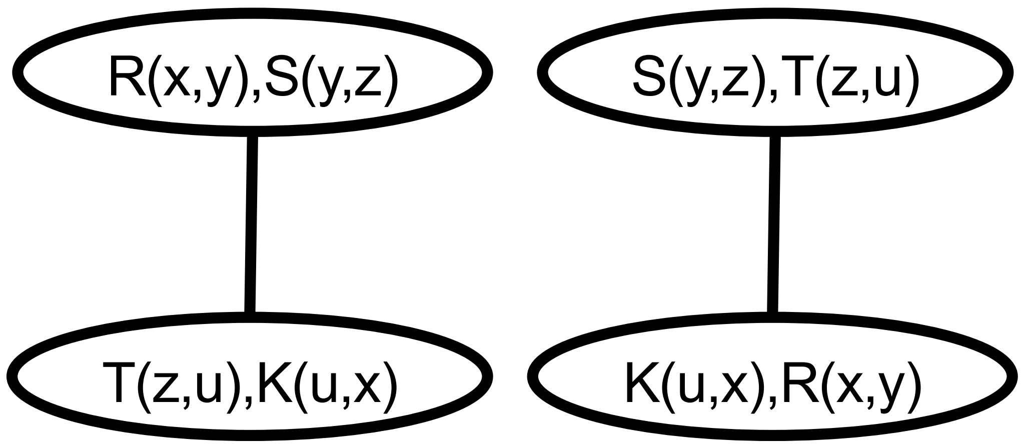
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

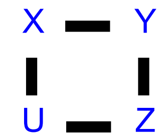
$\max_D \min_{tree} \max_{node\ t}$

Tree decompositions



$$\min(\overbrace{\max(h(xyz),h(zux))}^{T1},\ \underbrace{\max(h(yzu),h(uxy))}_{T2}) =$$

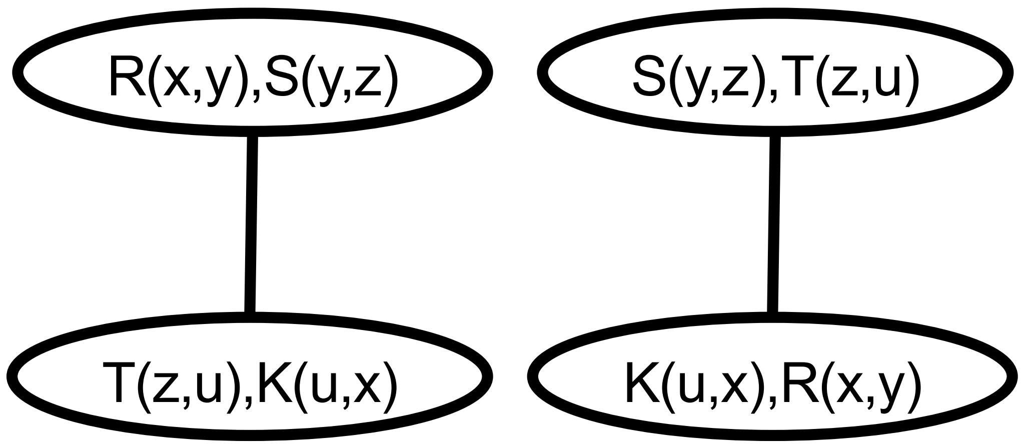
Q() = ∃x∃y∃z∃u R(x,y)∧S(y,z)∧T(z,u)∧K(u,x)



|R|,|S|,|T|,|K| ≤ N O($N^{3/2}$) algorithm [Alon,Yuster,Zwick'97]

$\max_D \min_{tree} \max_{node\ t}$

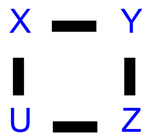
Tree decompositions



$$\min(\overbrace{\max(h(xyz),h(zux))}^{T1}, \underbrace{\max(h(yzu),h(uxy))}_{T2}) =$$

$$= \max(\min(h(xyz),h(yzu)), \min(h(xyz),h(uxy)), \min(h(zux),h(yzu)), \min(h(zux),h(uxy)))$$

$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



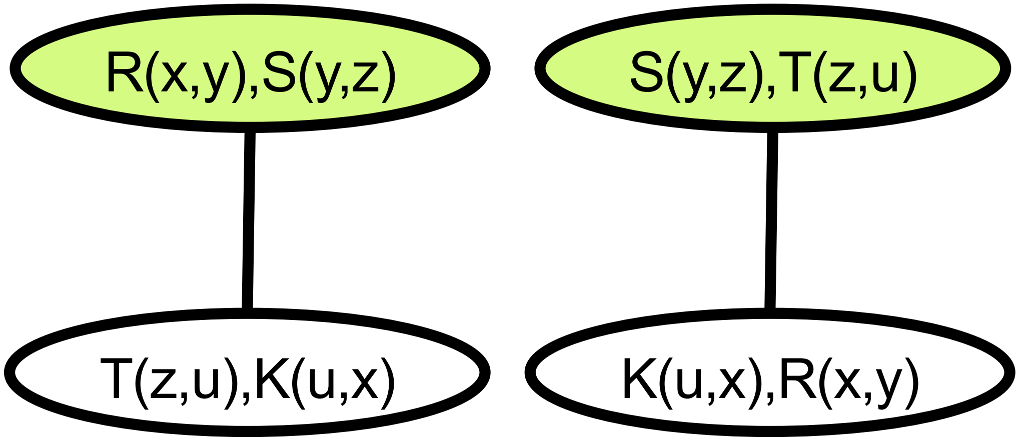
$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon, Yuster, Zwick'97]

$\max_D \min_{tree} \max_{node\ t}$

$$\min\Big(\overbrace{\max(h(xyz),h(zux))}^{T1},\ \underbrace{\max(h(yzu),h(uxy))}_{T2}\Big) =$$

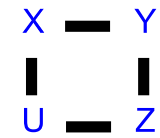
$$= \max(\mathbf{\min(h(xyz),h(yzu))}, \min(h(xyz),h(uxy)), \min(h(zux),h(yzu)), \min(h(zux),h(uxy)))$$

Tree decompositions



$3 \log N \geq h(xy) + h(yz) + h(zu)$

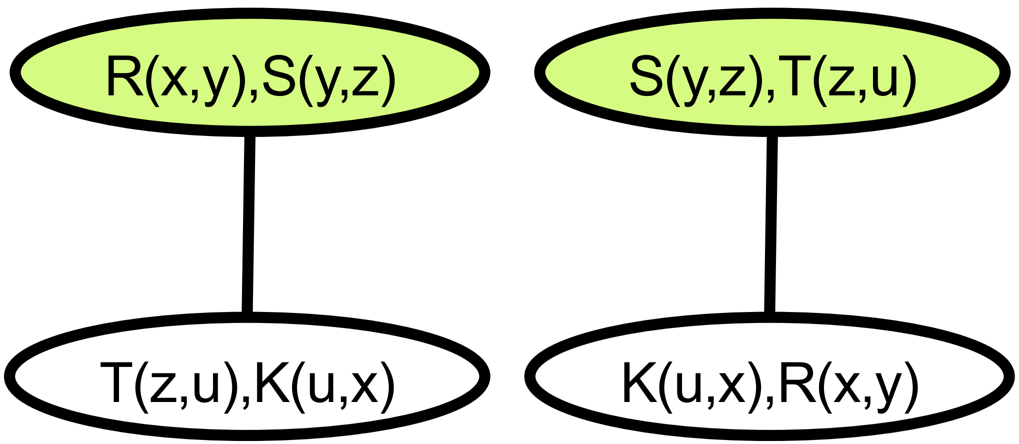
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

$\max_D \min_{tree} \max_{node\ t}$

Tree decompositions

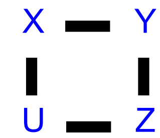


$$\min(\overbrace{\max(h(xyz),h(zux))}^{T1}, \underbrace{\max(h(yzu),h(uxy))}_{T2}) =$$

$$= \max(\mathbf{\min(h(xyz),h(yzu))}, \min(h(xyz),h(uxy)), \min(h(zux),h(yzu)), \min(h(zux),h(uxy)))$$

$$3 \log N \geq \underline{h(xy) + h(yz)} + h(zu)$$
$$\geq h(xyz) + h(y) + h(zu)$$

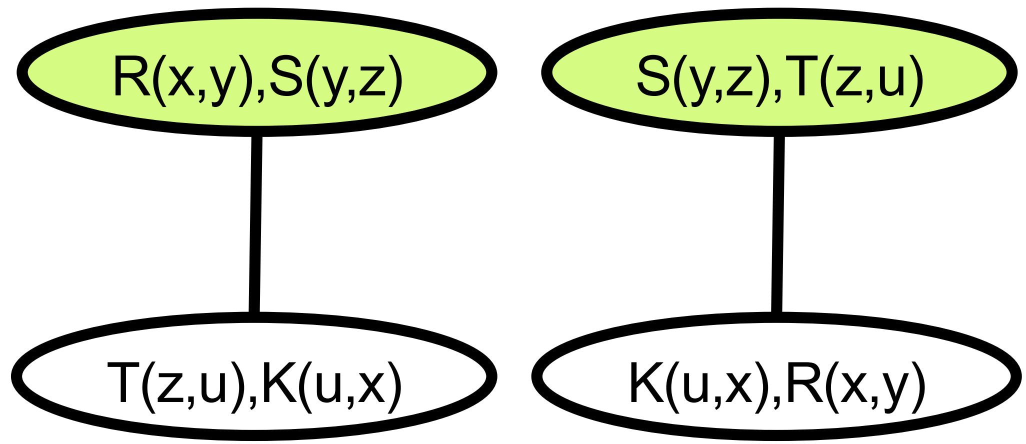
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon, Yuster, Zwick'97]

$\max_D \min_{tree} \max_{node\ t}$

Tree decompositions

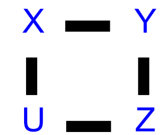


$$\overbrace{\min(\max(h(xyz),h(zux)),}^{T1}$$
$$\underbrace{\max(h(yzu),h(uxy)))}_{T2} =$$

$$= \max(\mathbf{\min(h(xyz),h(yzu))}, \min(h(xyz),h(uxy)), \min(h(zux),h(yzu)), \min(h(zux),h(uxy)))$$

$$3 \log N \geq \underline{h(xy) + h(yz)} + h(zu)$$
$$\geq h(xyz) + \underline{h(y) + h(zu)}$$
$$\geq h(xyz) + h(yzu) + h(\emptyset)$$

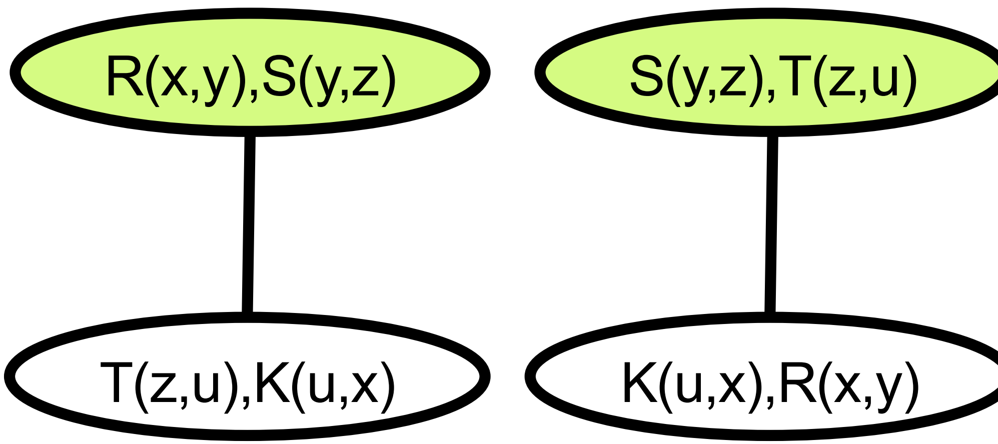
$Q() = \exists x \exists y \exists z \exists u\ R(x,y) \wedge S(y,z) \wedge T(z,u) \wedge K(u,x)$



$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

$\max_D \min_{tree} \max_{node\ t}$

Tree decompositions

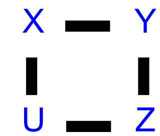


$$\min(\overbrace{\max(h(xyz),h(zux))}^{T1},\ \underbrace{\max(h(yzu),h(uxy))}_{T2}) =$$

$$= \max(\mathbf{\min(h(xyz),h(yzu))}, \min(h(xyz),h(uxy)), \min(h(zux),h(yzu)), \min(h(zux),h(uxy)))$$

$$3 \log N \geq \underline{h(xy) + h(yz)} + h(zu)$$
$$\geq h(xyz) + \underline{h(y) + h(zu)}$$
$$\geq h(xyz) + h(yzu) + h(\emptyset)$$
$$\geq 2\ \mathbf{\min(h(xyz),h(yzu))}$$

$Q() = \exists x \exists y \exists z \exists u \; R(x,y) \land S(y,z) \land T(z,u) \land K(u,x)$

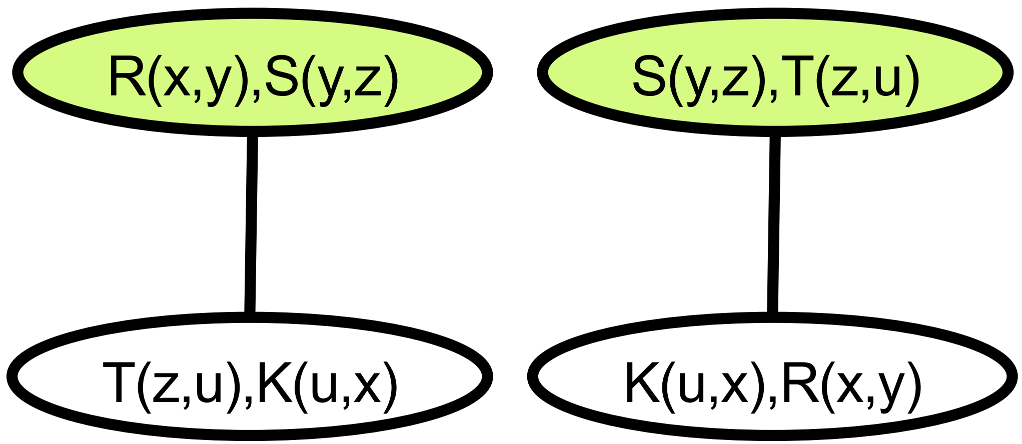


$|R|,|S|,|T|,|K| \leq N$ $O(N^{3/2})$ algorithm [Alon,Yuster,Zwick'97]

$\max_D \min_{tree} \max_{node\ t}$

$\boxed{subw(Q) = 3/2 \log N}$

Tree decompositions



$$\overbrace{\min(\max(h(xyz),h(zux)),}^{T1}\ \underbrace{\max(h(yzu),h(uxy)))}_{T2} =$$

$$= \max(\min(h(xyz),h(yzu)),\ \min(h(xyz),h(uxy)),\ \min(h(zux),h(yzu)),\ \min(h(zux),h(uxy))) \leq \mathbf{3/2 \log N}$$

$$3 \log N \geq \underline{h(xy) + h(yz)} + h(zu)$$
$$\geq h(xyz) + \underline{h(y) + h(zu)}$$
$$\geq h(xyz) + h(yzu) + h(\emptyset)$$
$$\geq 2 \min(h(xyz),h(yzu))$$

Proof to *Algorithm*

Use the proof of:

$$\boxed{h(xyz)+h(yzu) \leq h(xy) + h(yz) + h(zu)}$$

to compute the *disjunctive datalog* rule:

$$\boxed{A(x,y,z) \vee B(y,z,u) \leftarrow R(x,y) \wedge S(y,z) \wedge T(z,u)}$$

(details omitted)

Runtime $\tilde{O}(N^{3/2})$

Outline

- Enumeration problem
- Decision problem
- Conclusions

Conclusions

Query evaluation summary:

- Information theory → *Proof*
- *Proof* → *Algorithm*

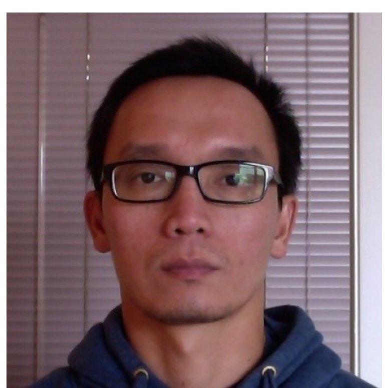
Open problems:

- Better "Proof → Algorithm"
- Fine-grained lower bounds

Thank You!

Questions?

Mahmoud Abo-Khamis   Hung Ngo   – RelationalAI Inc.



[PODS’2016]
[PODS’2017]