Paper Presentation:
On the Enumeration Complexity of Unions of Conjunctive Queries

Yuxin Kang

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Scope of Study

tutorials:

Person	Title
Alan Fekete	Making Consistency...
Suresh Venkatasu...	Algorithmic Fairness...

schedule:

Title	Day
Making Consistency...	Tue
Algorithmic Fairness...	Wed
Regularizing Conjunct...	Mon

Person	Day
Alan Fekete	Tue
Suresh Venkatasu...	Wed

research talks:

Person	Title
Pablo Barceló	Regularizing Conjunct...
Peter Lindner	Probabilistic Database...
Muhammad Tibi	Query Evaluation in...

Person	Day
Pablo Barceló	Mon
Martin Grohe	Mon
Muhammad Tibi	Mon

$$Q_1(\text{Person}, \text{Day}) \leftarrow \text{tutorials}(\text{Person}, \text{Title}), \text{schedule}(\text{Title}, \text{Day})$$
$$Q_2(\text{Person}, \text{Day}) \leftarrow \text{research}(\text{Person}, \text{Title}), \text{schedule}(\text{Title}, \text{Day})$$

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Scope of Study

$$Q_3 = Q_1 \cup Q_2$$

Person	Day
Alan Fekete	Tue
Suresh Venkatasu…	Wed
Pablo Barceló	Mon
Martin Grohe	Mon
Muhammad Tibi	Mon

How efficiently can we solve such queries?

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Definitions & Terminologies

Enumeration Complexity: complexity of outputting all answers individually, measured by preprocessing time + per-answer delay, rather than just total runtime.

Linear Preprocessing Time: the setup phase of the enumeration algorithm runs in time proportional to the size of the input (O(|Input|)).

Constant Delay: after the preprocessing step finishes and the first answer is produced, the time to generate each next answer is bounded by a fixed constant (O(1)).

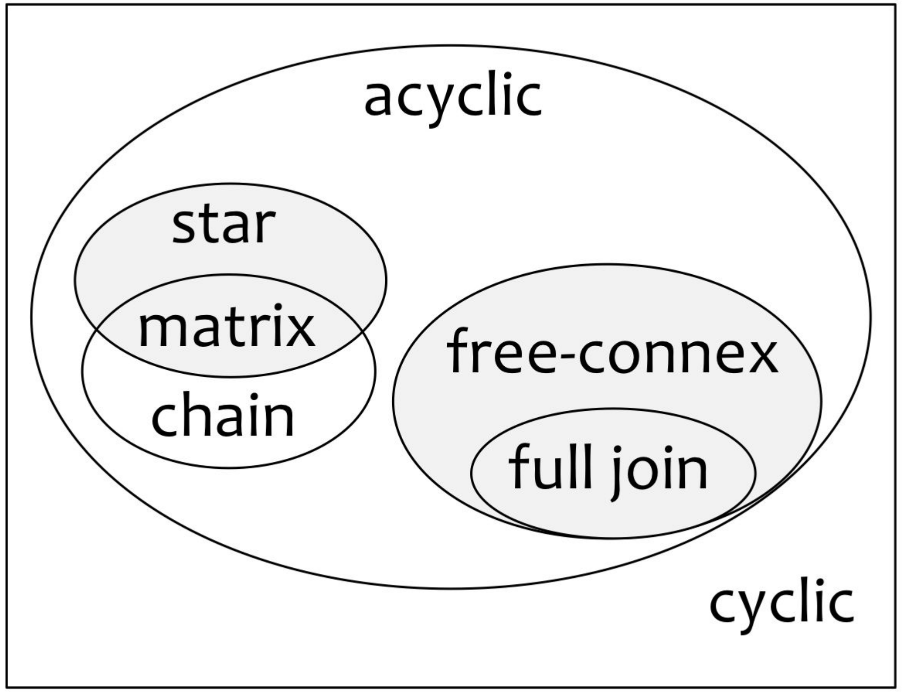
DelayClin: a class of enumeration algorithms with the Linear Preprocessing Time and Constant Delay.

Tractable: we refer to queries in DelayClin as tractable and queries outside of this class as intractable.

Background & Previous Results

1) Free-connex queries are tractable (Bagan et al. & Brault-Baron).

[Lecture 6]



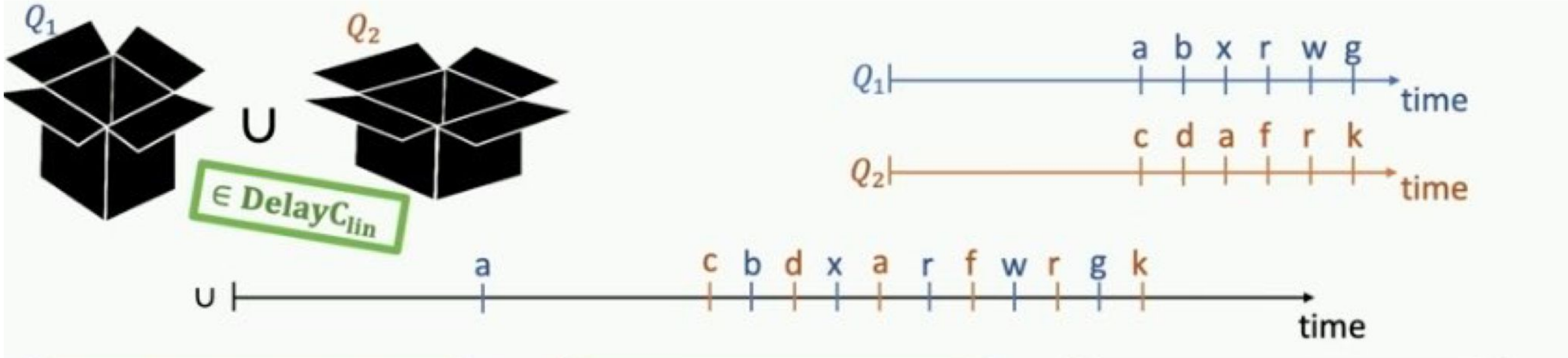
[Lecture 3: A Practical version of Yannakakis's Algorithm]

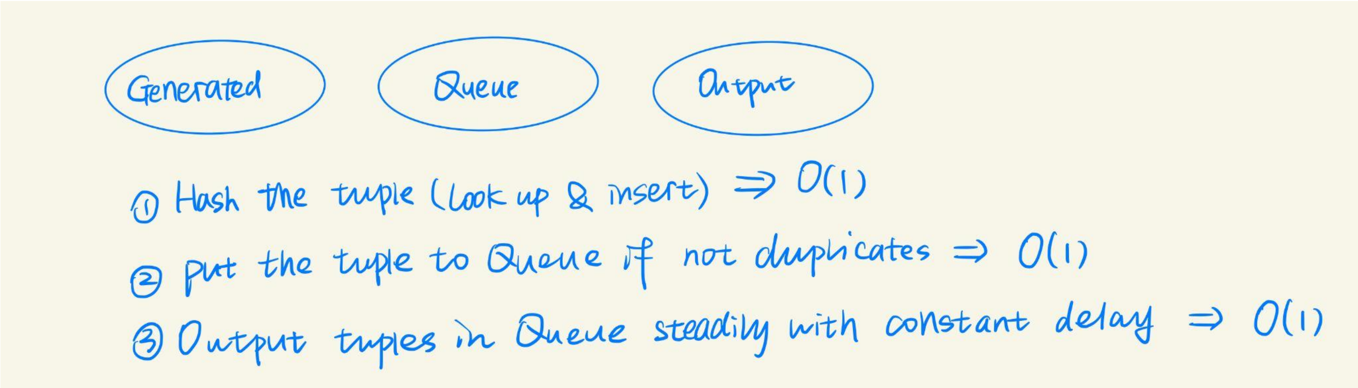
- After $O(N)$ linear preprocessing time, an index of linear size can be built such that every join result can be enumerated with $O(1)$ delay

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Background & Previous Results

2) A union of tractable problems is tractable (Durand & Grandjean).



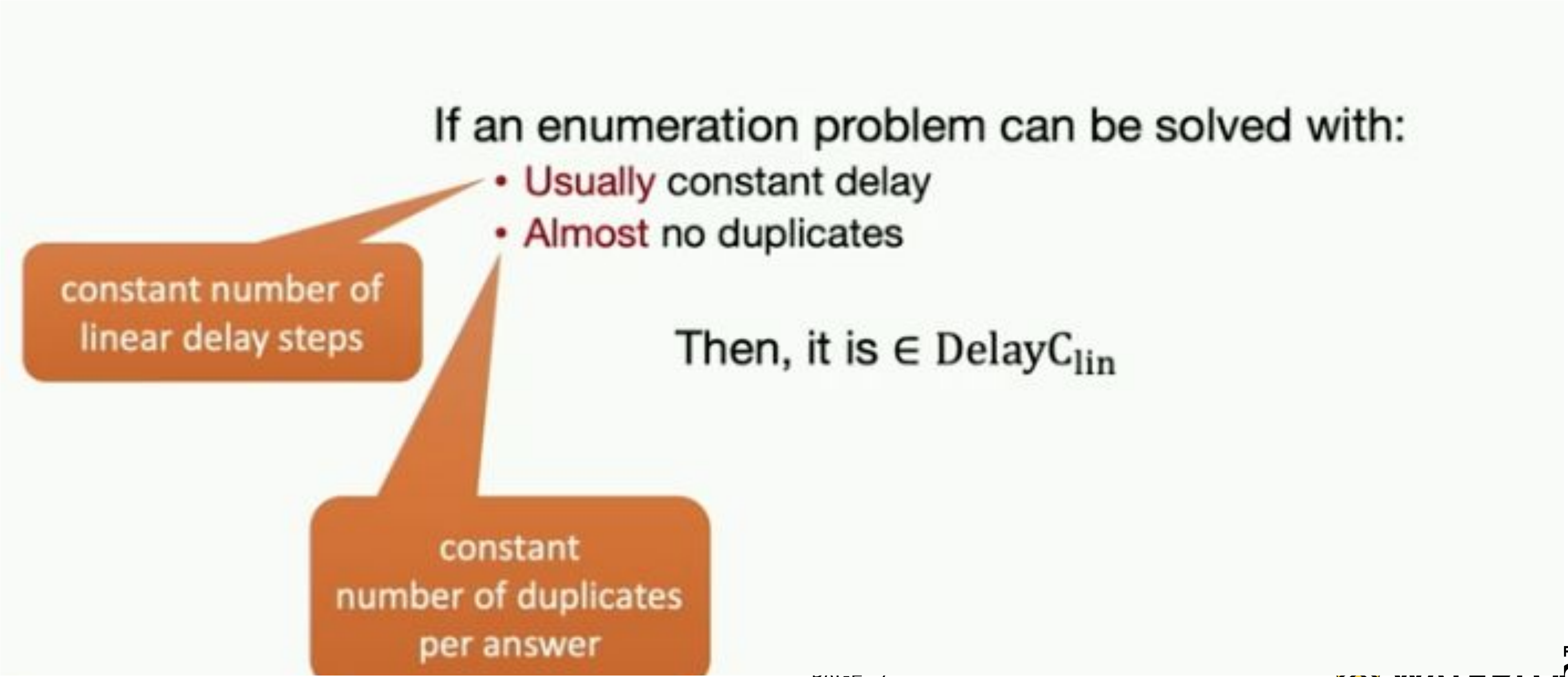


A UCQ of free-connex queries is tractable

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Background & Previous Results

3) Cheater's Lemma (Bagan et al.).



Motivation:

What happens if some CQs of a UCQ are tractable while others are not?

Intuitively, one might be tempted to expect a union of enumeration problems to be harder than a single problem within the union, making such a UCQ intractable as well.

This paper shows that some UCQs containing intractable CQs are, in fact, tractable.

And, some UCQs containing only intractable CQs are tractable!

Upper bound

Definition (Body-homomorphism):

body-homomorphism h: Q2 → Q1: For every atom (relation) in Q2, if we rename its variables using h, it becomes an atom in Q1. ***Q2's head and Q1's head do not have to match.***

Let $Q_1(x, y, w) \leftarrow R_1(x, z), R_2(z, y), R_3(y, w)$;

$Q_2(a, b) \leftarrow R_1(a, c), R_2(c, b)$;

Define $h: Q_2 \rightarrow Q_1$ as $h(a) = x$, $h(c) = z$, $h(b) = y$.

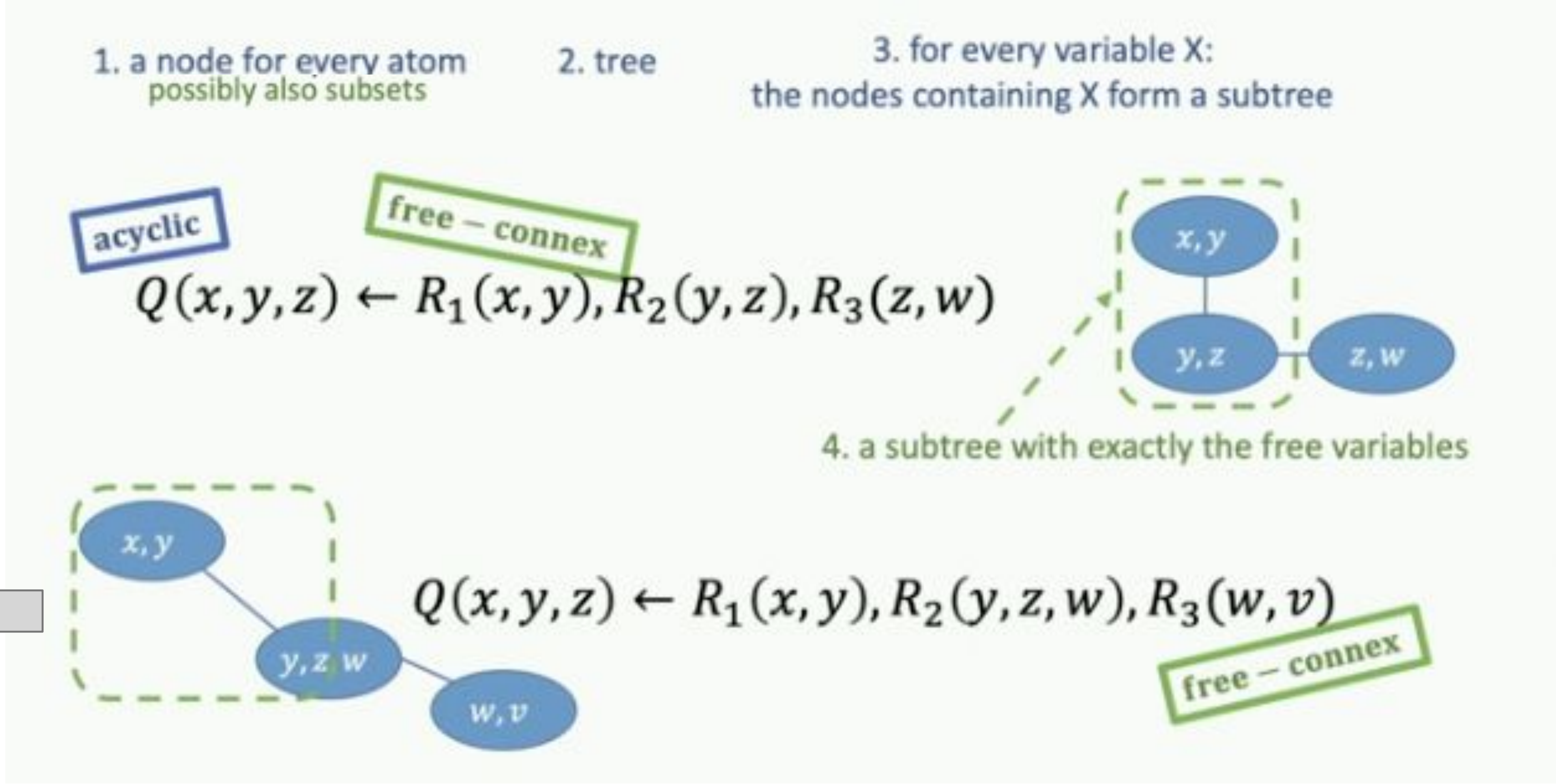
So, $R_1(a, c)$ under h becomes $R_1(x, z)$;

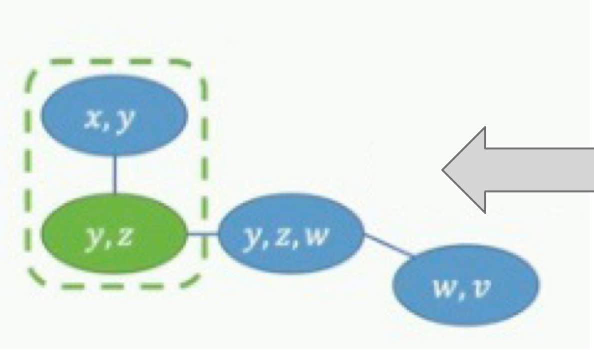
$R_2(c, b)$ under h becomes $R_2(z, y)$.

Definition (Body-isomorphism): a body-homomorphism that's also bijective.

Upper bound

Free-Connex Revisit:





UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Upper bound

Some UCQs containing intractable CQs are tractable.

Example:

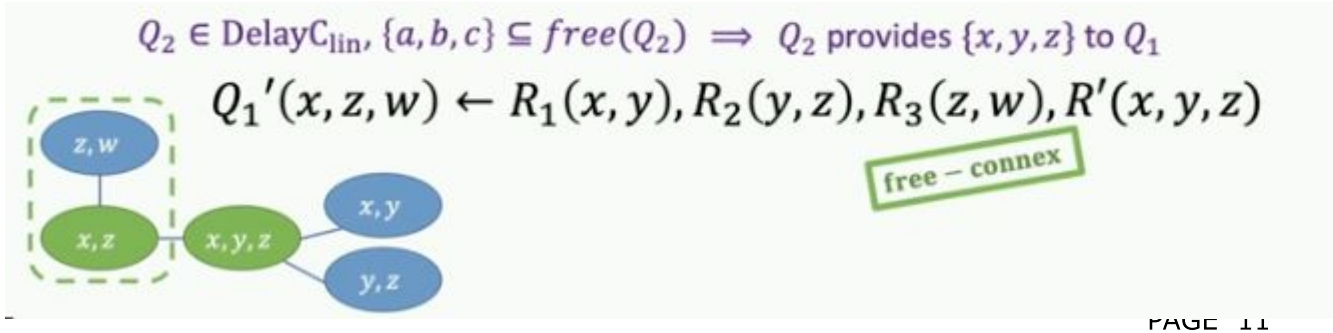
Q1(x, z, w) ← R1(x, y), R2(y, z), R3(z, w) (Non Free-Connex)

Q2(a, b, c) ← R1(a, b), R2(b, c) (Free-Connex)

Observations:

We have a body-homomorphism here: **h(a) = x, h(b) = y, h(c) = z**

So, Q2(x, y, z) ← R1(x, y), R2(y, z)



A **union extension** takes an intractable query Q1 and step by step adds “virtual atoms,” which are new relations formed from the answers of other tractable queries in the same union, until the extended query becomes free-connex.

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Upper bound

Some UCQs containing only intractable CQs are tractable.

Example:

$$Q_1(x,z,w,u) \leftarrow R_1(x,y), R_2(y,z), R_3(z,w), R_4(w,u)$$
$$Q_2(a,c,d,e) \leftarrow R_1(e,d), R_2(d,c), R_3(c,b), R_4(b,a)$$

R_1	
x1	y1
x1	y2

R_2	
y1	z1
y2	z1

R_3	
z1	w1
z1	w2

R_4	
w1	u1
w2	u2

Some Observations:

We have a body-isomorphism here:

h(e) = x, h(d) = y, h(c) = z, h(b) = w, h(a) = u;

h^{-1}(x) = e, h^{-1}(y) = d, h^{-1}(z) = c, h^{-1}(w) = b, h^{-1}(u) = a.

So, Q2(u, z, y, x) ← R1(x, y), R2(y, z), R3(z, w), R4(w, u)

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Upper bound

$$Q_1(x,z,w,u), Q_2(u,z,y,x) \leftarrow R_1(x,y), R_2(y,z), R_3(z,w), R_4(w,u)$$

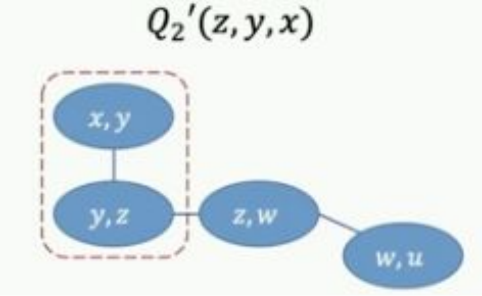
R_1	
x1	y1
x1	y2

R_2	
y1	z1
y2	z1

R_3	
z1	w1
z1	w2

R_4	
w1	u1
w2	u2

Step 1: Solve Q2', output a subset of the answers of Q2

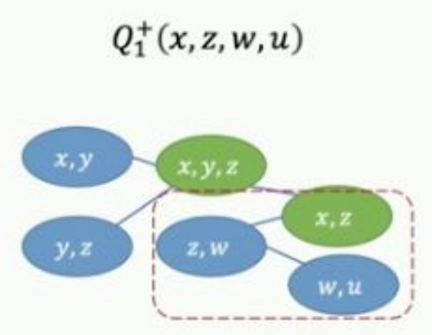


Q_2'		
x1	y1	z1
x1	y2	z1

$\subseteq Q_2$			
x1	y1	z1	u1
x1	y2	z1	u1

Step 2: Solve Q1†, output Q1 (union extension of Q1 with respect Q2)

Q1†(x, z, w, u) ← R1(x, y), R2(y, z), R3(z, w), R4(w, u), RQ2′(z, y, x)



Q_1			
x1	z1	w1	u1
x1	z1	w2	u1

ERSITY OF
ERLOO | FACULTY OF MATHEMATICS

Upper bound

$$Q_1(x,z,w,u), Q_2(u,z,y,x) \leftarrow R_1(x,y), R_2(y,z), R_3(z,w), R_4(w,u)$$

R_1	
x1	y1
x1	y2

R_2	
y1	z1
y2	z1

R_3	
z1	w1
z1	w2

R_4	
w1	u1
w2	u2

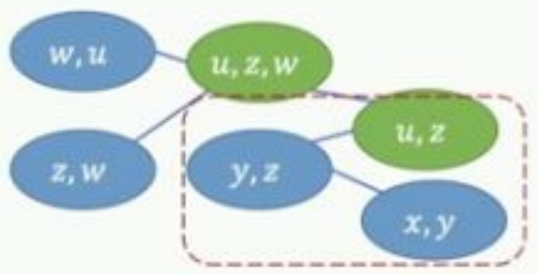
Step 3: Get Q1’(z,w,u), the projection of Q1.

Q1’		
z1	w1	u1
z1	w2	u1

Step 4 Solve Q2†, output Q2 (union extension of Q2 with respect Q1)

Q2†(u, z, y, x) ← R1(x, y), R2(y, z), R3(z, w), R4(w, u), RQ1′(z, w, u)

$Q_2^{+}(u,z,y,x)$



$Q_2^{+}(u,z,y,x)$

Q_2			
x1	y1	z1	u1
x1	y2	z1	u1
x1	y1	z1	u2
x1	y2	z1	u2

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Upper Bound

Let Q be a UCQ.

If <u>each</u> CQ in Q can become free-connex by adding provided atoms $\Longrightarrow$ $Q \in DelayC_{lin}$

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Lower bound

Definition (Boolean Matrix Multiplication (BMM) conjecture):

Boolean $n \times n$ matrices cannot be multiplied in time $O(n^2)$

$$\begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}\begin{pmatrix}0 & 1\\0 & 1\end{pmatrix} = \begin{pmatrix}0 & 1\\0 & 1\end{pmatrix}$$

Acyclic non-free-connex:

$$Q(x,z) \leftarrow R_1(x,y), R_2(y,z)$$

Q

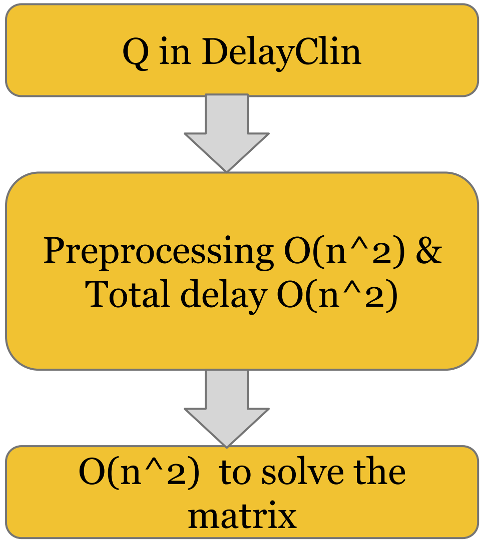
R	C
1	2
2	2

A

R	C
1	1
1	2
2	2

B

R	C
1	2
2	2



contradiction!

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Conditional Lower Bound

Let Q be a UCQ. If there exists intractable CQ that cannot be fixed to tractable CQ. → $Q \notin DelayC_{lin}$

(Under the assumption of BMM)

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Conclusions

1. Extending enumeration theory from CQs to UCQs.

2. The Definition of Union Extensions.

3. Upper Bound: a UCQ is tractable if each CQ in it can become free-connex by union extension. (sufficient but not necessary).

4. Conditional Lower Bound: if a UCQ contains an intractable CQ that cannot be fixed via union extensions, then enumerating it in DelayClin would break known hardness assumptions (BMM).

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Future Directions

1. Find a complete characterization (a dichotomy): a criterion that is necessary and sufficient for UCQ DelayClin enumeration.

2. Space vs. Time trade-offs. The DelayClin for UCQs via union extensions is space demanding.

References

https://dl.acm.org/doi/10.1145/3450263

https://av.tib.eu/media/42887

https://cs.uwaterloo.ca/~xiaohu/courses/CS848/yannakakis-part2.pdf

https://cs.uwaterloo.ca/~xiaohu/courses/CS848/aggregation-part1.pdf