Updated 3/23/2024

Topic 2: Complexity of Query Evaluation
Unit 4: Reverse Data Management (RDM)
Lecture 18

Wolfgang Gatterbauer

CS7240 Principles of scalable data management (sp24)

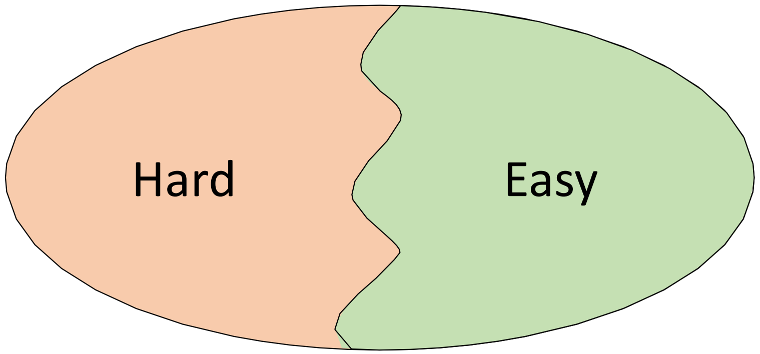
https://northeastern-datalab.github.io/cs7240/sp24/

3/22/2024

Dichotomy theorems

Dichotomy theorem

classifying every member of a family of problems as easy or hard.



In database context

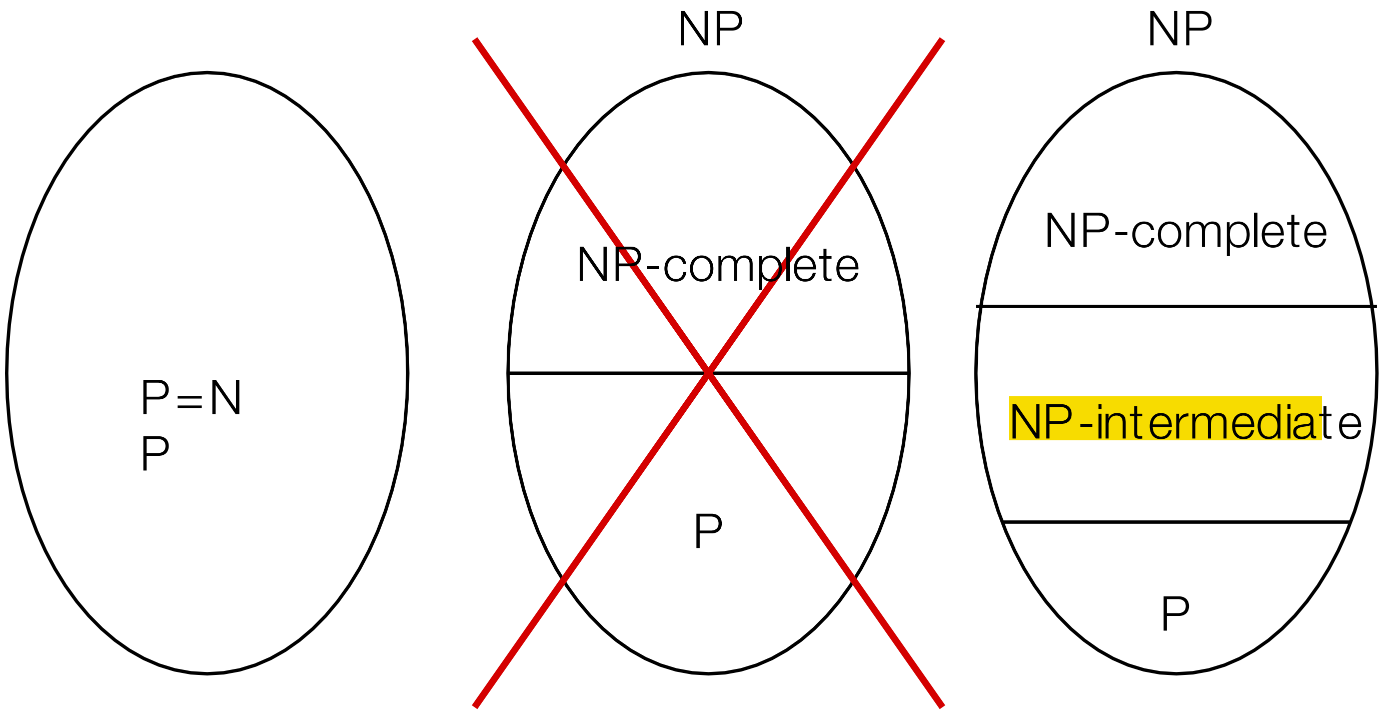
Given a certain problem and a query. Solving this problem for a query is either easy or hard.

Dichotomy theorems

Every problem is either in P or NP-complete.

Why are such theorems surprising?

Theorem [Ladner, 1973]
If P ≠ NP, then there is a language L ∈ NP \ P that is not NP-complete.



Dichotomy theorems

- Dichotomy theorems give goods research programs: easy to formulate, but can be hard to complete.
- The search for dichotomy theorems may uncover algorithmic results that no one has thought of.
- Proving dichotomy theorems requires attacking the problem both from the algorithmic and the complexity side. Requires good command of both algorithmic and hardness proof techniques.
- Possible outcomes:
  - Everything is hard, except some trivial cases.
  - Everything is hard, except the famous known nontrivial positive cases.
  - Some unexpected easy cases are found.

Source: Daniel Marx. Every graph is easy or hard: dichotomy theorems for graph problems, 2015. http://www.cs.bme.hu/~dmarx/papers/marx-icgt2014-slides.pdf

Example dichotomy theorems in DB theory

- Probabilistic databases
  - Self-join (SJ) free: [Dalvi, Suciu, VLDB 2004]
  - SJ: [Dalvi, Suciu, JACM 2012]
- Resilience
  - SJ-free: [Freire+, VLDB 2015]
  - SJ: open (some progress in [Freire+, PODS'20], [Makhija+, SIGMOD'24])
- View-side effect problem
  - SJ free with FDs [Kimelfeld, PODS 2012]
- Consistent query answering
  - SJ-free: [Koutris, Wijsen, PODS 2015]

Source: Dalvi, Suciu. "Efficient query evaluation on probabilistic databases", VLDB 2004. https://dl.acm.org/doi/abs/10.5555/1316689.1316764 , Dalvi, Suciu. "The dichotomy of probabilistic inference for unions of conjunctive queries", JACM 2012. https://doi.org/10.1145/2395116.2395119 , Freire, Gatterbauer, Immerman, Meliou, The complexity of resilience and responsibility for self-join-free conjunctive queries. PVLDB 2015. https://doi.org/10.14778/2850583.2850592 , Freire, Gatterbauer, Immerman, Meliou. New Results for the Complexity of Resilience for Binary Conjunctive Queries with Self-Joins. PODS 2020. https://doi.org/10.1145/3375395.3387647 , Kimelfeld. "A dichotomy in the complexity of deletion propagation with functional dependencies". PODS 2012. https://doi.org/10.1145/2213556.2213584 ,Koutris, Wijsen, "The Data Complexity of Consistent Query Answering for Self-Join-Free Conjunctive Queries Under Primary Key Constraints", PODS 2015, https://doi.org/10.1145/2745754.2745769

The view deletion problem

- *D* a database instance and *V=Q(D)* a view defined over *D*.
  - *Find a set of tuples ΔD to remove from D so that a specific tuple t is removed from the view*
  - *Minimize the number of side-effects in the view* VIEW
    - *View side-effect problem*
      - *Hard: queries with joins and projection or union*
      - *PTIME: the rest*
  - *Minimize the number of tuples deleted from D* SOURCE
    - *Source side-effect problem*
      - *Same dichotomy*

DATALAB
@Northeastern

Slides slightly extended from this presentation in Jan 2024

Solving Reverse DM Problems with ILPs and LP relaxations

Wolfgang Gatterbauer

Based on joint work with **Neha Makhija**

Dagstuhl seminar 24032
Jan 15, 2024

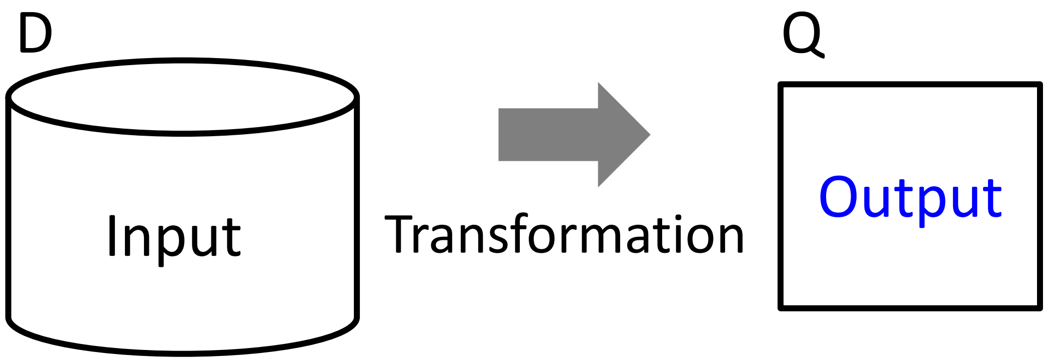
https://northeastern-datalab.github.io/unified-reverse-data-management/

https://www.dagstuhl.de/en/seminars/seminar-calendar/seminar-details/24032

Reverse Data Management

QUERY EVALUATION:

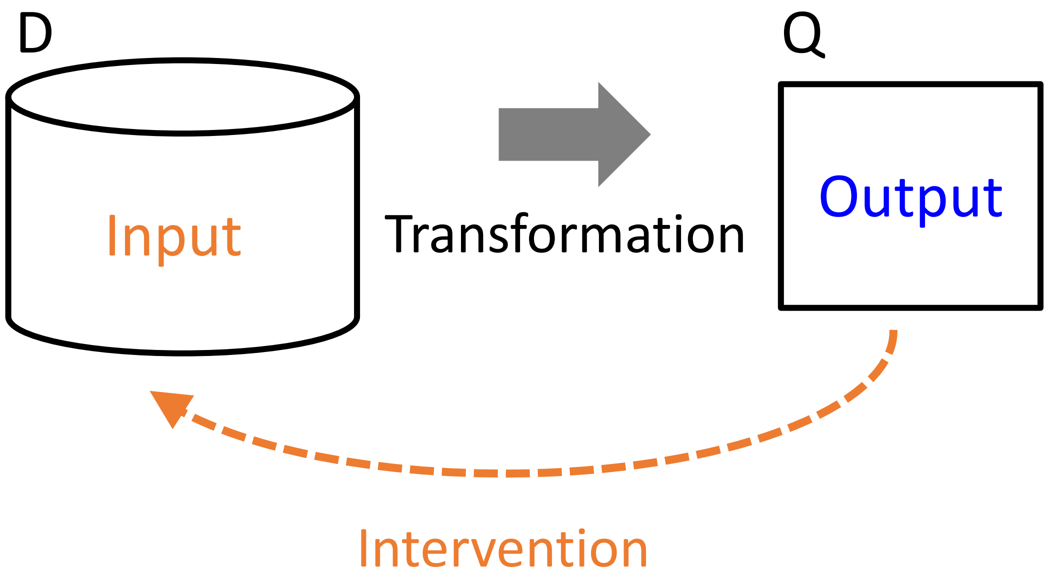
A transformation of the input to the output



Reverse Data Management

QUERY EVALUATION:

A transformation of the input to the output



uses some notion of minimality

REVERSE DATA MANAGEMENT:

What are the required changes to the input, in order to achieve a desired output?

Cp. to "explanation by intervention" (Sudeepa's earlier talk today), and "minimum change required" (Pablo's earlier talk today)

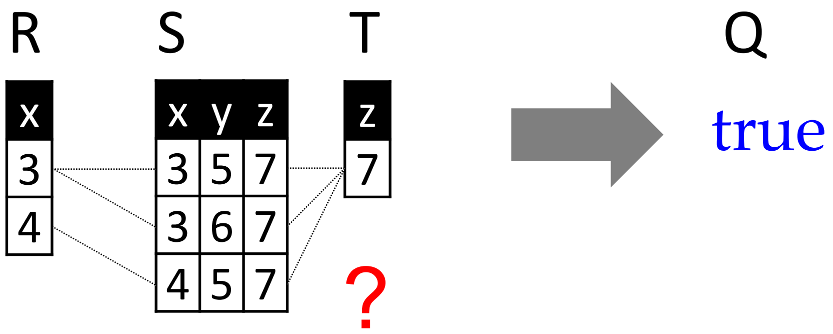
Meliou, Gatterbauer, Suciu. Reverse Data Management. VLDB 2011 Vision track. https://doi.org/10.14778/3402755.3402803

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false



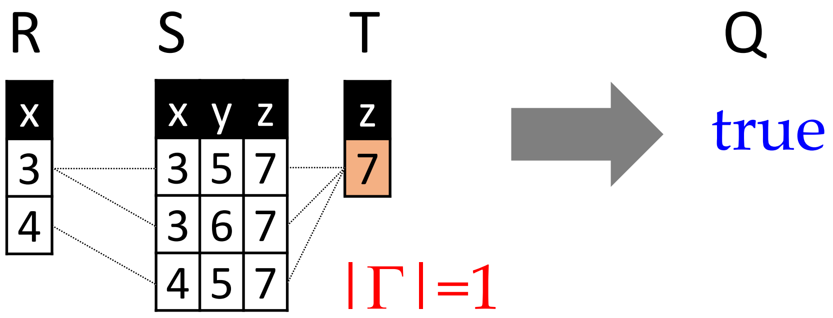
Freire, Gatterbauer, Immerman, Meliou. The complexity of resilience and responsibility for self-join-free conjunctive queries, VLDB 2015. https://doi.org/10.14778/2850583.2850592

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false



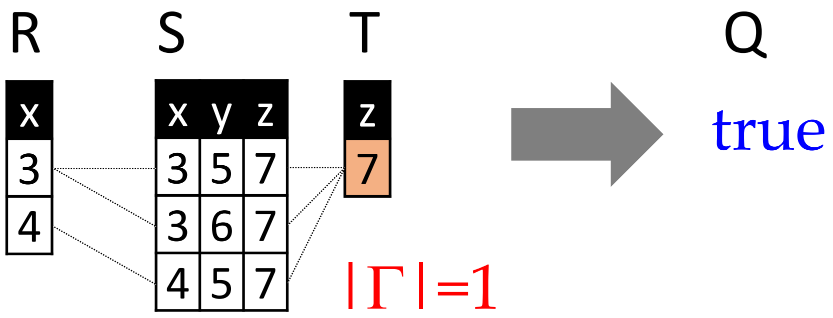
Freire, Gatterbauer, Immerman, Meliou. The complexity of resilience and responsibility for self-join-free conjunctive queries, VLDB 2015. https://doi.org/10.14778/2850583.2850592

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

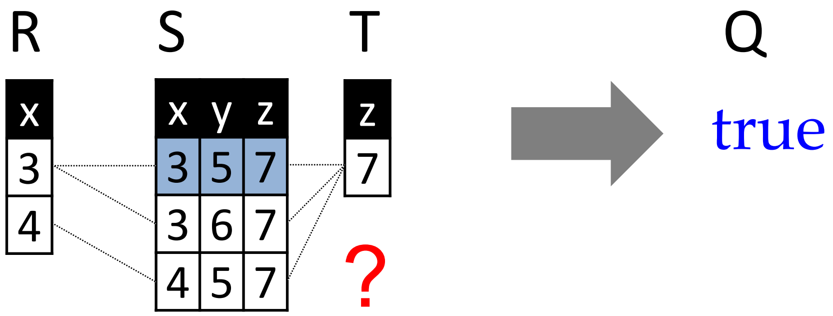
Resilience

Delete min number of tuples to make Q false



Causal responsibility

Delete min number of tuples to make
an input tuple counterfactual



Freire, Gatterbauer, Immerman, Meliou. The complexity of resilience and responsibility for self-join-free conjunctive queries, VLDB 2015. https://doi.org/10.14778/2850583.2850592
Meliou, Gatterbauer, Moore, Suciu. The complexity of causality and responsibility for query answers and non-answers, VLDB 2010. https://doi.org/10.14778/1880172.1880176

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

Q: true

$|\Gamma|=1$

recall Pablo's discussion on "interactions" for actual causes in disjunctions

Causal responsibility

Delete min number of tuples to make *<u>an input tuple</u> counterfactual*

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

Q: true

$|\Gamma|=2$

Freire, Gatterbauer, Immerman, Meliou. The complexity of resilience and responsibility for self-join-free conjunctive queries, VLDB 2015. https://doi.org/10.14778/2850583.2850592
Meliou, Gatterbauer, Moore, Suciu. The complexity of causality and responsibility for query answers and non-answers, VLDB 2010. https://doi.org/10.14778/1880172.1880176

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

$|\Gamma|=1$

Q: true

Q(u,w):-R'(u,x),S(x,y,z),T'(z,w)

Source side-effects (in deletion prop.)

Delete min number of tuples to delete an output tuple

R'

u	x
1	3
1	4
2	3

S

x	y	z
3	5	7
3	6	7
4	5	7

T'

z	w
7	9

?

Q

u	w
1	9
2	9

Dayal, Bernstein. On the correct translation of update operations on relational views, TODS 1982, https://doi.org/10.1145/319732.319740
Buneman, Khanna, Tan. On propagation of deletions and annotations through views, PODS 2002, https://doi.org/10.1145/543613.543633

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

$|\Gamma|=1$

Q

true

Q(u,w):-R'(u,x),S(x,y,z),T'(z,w)

Source side-effects (in deletion prop.)

Delete min number of tuples to delete an output tuple

R'

u	x
1	3
1	4
2	3

S

x	y	z
3	5	7
3	6	7
4	5	7

T'

z	w
7	9

$|\Gamma|=1$

Q

u	w
1	9
~~2~~	~~9~~

$|\Delta|=1$

Dayal, Bernstein. On the correct translation of update operations on relational views, TODS 1982, https://doi.org/10.1145/319732.319740
Buneman, Khanna, Tan. On propagation of deletions and annotations through views, PODS 2002, https://doi.org/10.1145/543613.543633

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

$|\Gamma|=1$

Q: true

Q(u,w):-R'(u,x),S(x,y,z),T'(z,w)

Source side-effects (in deletion prop.)

Delete min number of tuples to delete an output tuple

R'

u	x
1	3
1	4
2	3

S

x	y	z
3	5	7
3	6	7
4	5	7

T'

z	w
7	9

$|\Gamma|=1$

Q

u	w
1	9
~~2~~	~~9~~

$|\Delta|=1$

View side-effects (in deletion propagation)

Delete tuples in order to delete an output tuple,
while minimizing the other output tuples deleted

R'

u	x
1	3
1	4
2	3

S

x	y	z
3	5	7
3	6	7
4	5	7

T'

z	w
7	9

?

Q

u	w
1	9
2	9

?

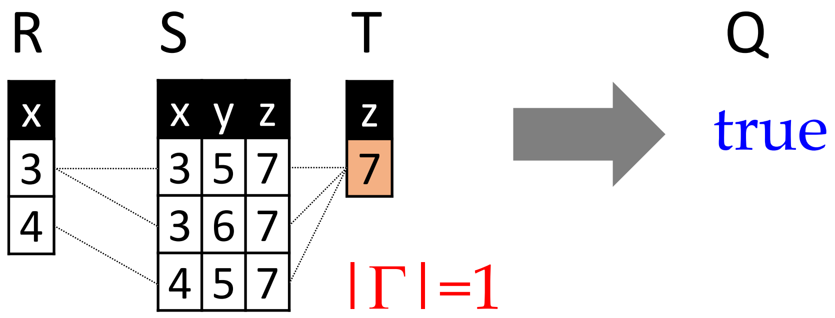
Dayal, Bernstein. On the correct translation of update operations on relational views, TODS 1982, https://doi.org/10.1145/319732.319740
Buneman, Khanna, Tan. On propagation of deletions and annotations through views, PODS 2002, https://doi.org/10.1145/543613.543633
Kimelfeld, Vondrak, Williams. Maximizing conjunctive views in deletion propagation, PODS 2011, https://doi.org/10.1145/1989284.1989308

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

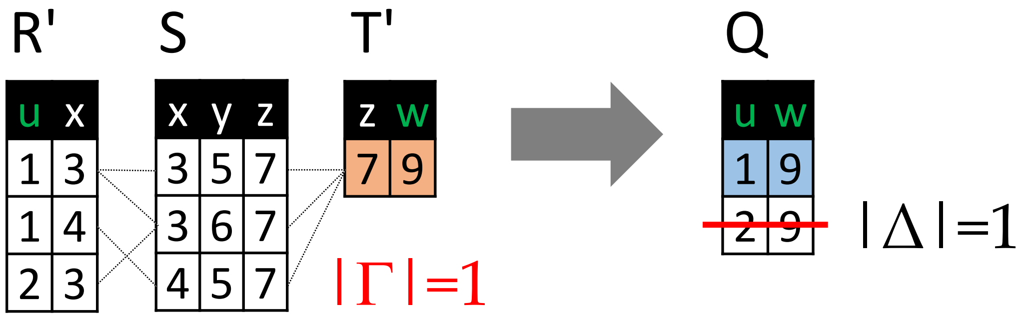
Delete min number of tuples to make Q false



Q(u,w):-R'(u,x),S(x,y,z),T'(z,w)

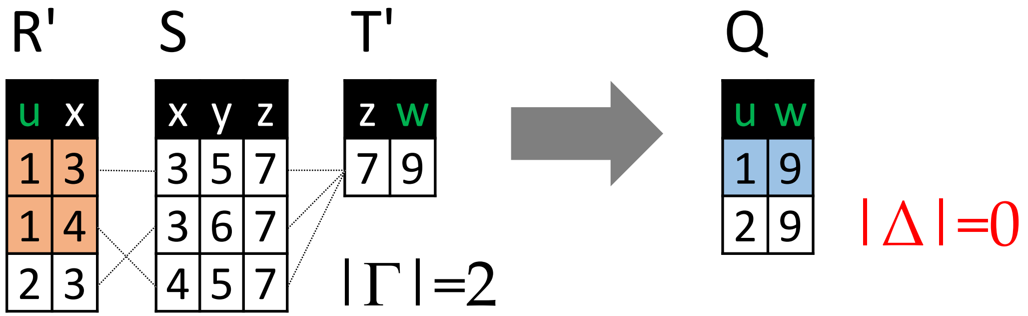
Source side-effects (in deletion prop.)

Delete min number of tuples to delete an output tuple



View side-effects (in deletion propagation)

Delete tuples in order to delete an output tuple, while minimizing the other output tuples deleted



Dayal, Bernstein. On the correct translation of update operations on relational views, TODS 1982, https://doi.org/10.1145/319732.319740
Buneman, Khanna, Tan. On propagation of deletions and annotations through views, PODS 2002, https://doi.org/10.1145/543613.543633
Kimelfeld, Vondrak, Williams. Maximizing conjunctive views in deletion propagation, PODS 2011, https://doi.org/10.1145/1989284.1989308

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

Q: true

$|\Gamma|=1$

Q:-R(x),S(x,y,z),T(z)

Smallest witness problem

Delete max number of tuples while keeping Q true

R

x
3
4

S

x	y	z
3	5	7
3	6	7
4	5	7

T

z
7

Q: true

?

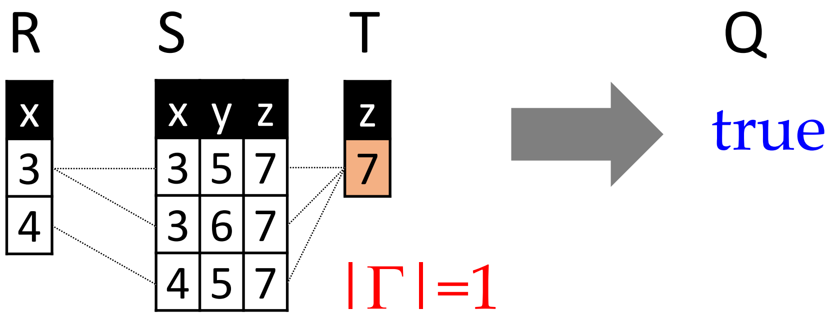
Miao, Roy, Yang. Explaining Wrong Queries Using Small Examples. SIGMOD 2019. https://doi.org/10.1145/3299869.3319866
Hu, Sintos. Finding Smallest Witnesses for Conjunctive Queries, ICDT 2024. https://cs.uwaterloo.ca/~xiaohu/icdt2024.pdf

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

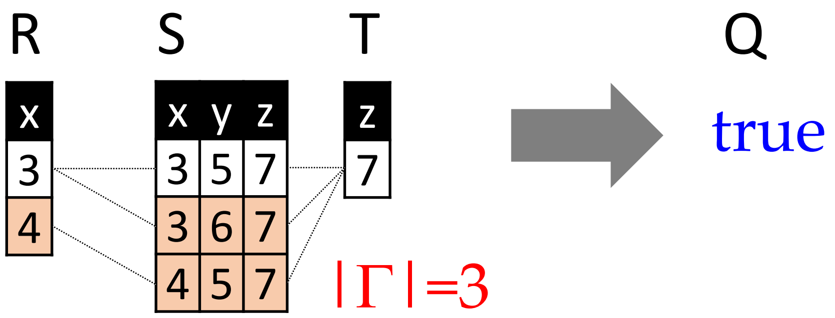
Delete min number of tuples to make Q false



Q:-R(x),S(x,y,z),T(z)

Smallest witness problem

Delete max number of tuples while keeping Q true



becomes more interesting once you add projections

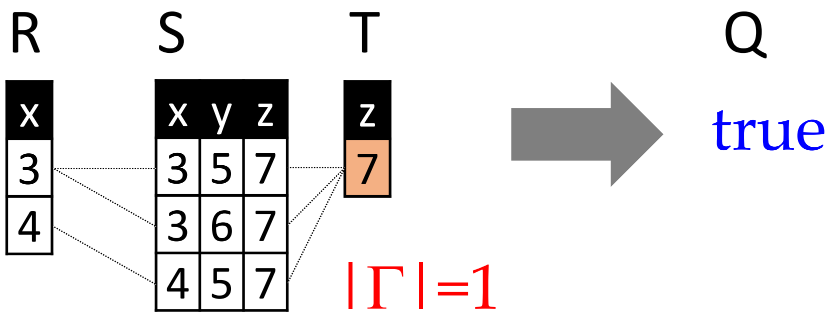
Miao, Roy, Yang. Explaining Wrong Queries Using Small Examples. SIGMOD 2019. https://doi.org/10.1145/3299869.3319866
Hu, Sintos. Finding Smallest Witnesses for Conjunctive Queries, ICDT 2024. https://cs.uwaterloo.ca/~xiaohu/icdt2024.pdf

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

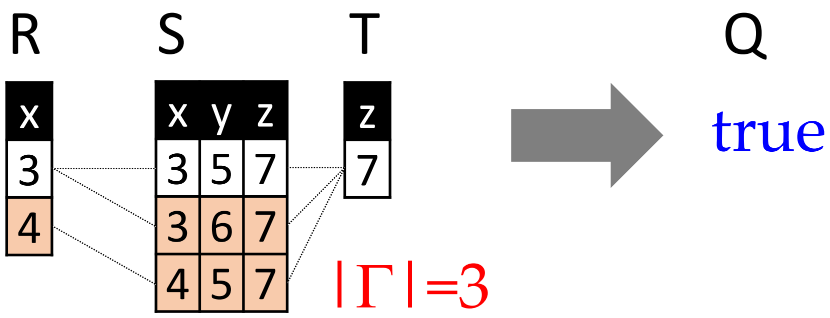
Delete min number of tuples to make Q false



Q:-R(x),S(x,y,z),T(z)

Smallest witness problem

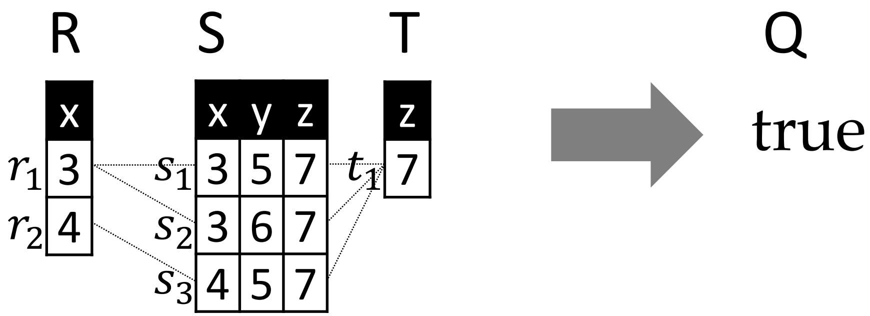
Delete max number of tuples while keeping Q true



Minimal size factorization

Find a provenance factorization of minimum length

$\text{Prov}(Q, D) = \ \dots$?



Miao, Roy, Yang. Explaining Wrong Queries Using Small Examples. SIGMOD 2019. https://doi.org/10.1145/3299869.3319866

Hu, Sintos. Finding Smallest Witnesses for Conjunctive Queries, ICDT 2024. https://cs.uwaterloo.ca/~xiaohu/icdt2024.pdf

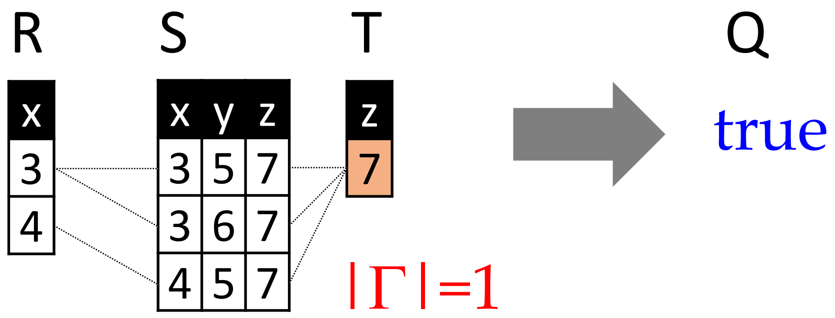
Makhija, Gatterbauer: Towards a Dichotomy for Minimally Factorizing the Provenance of Self-Join Free Conjunctive Queries. https://arxiv.org/pdf/2105.14307

Example Reverse Data Management Problems

Q:-R(x),S(x,y,z),T(z)

Resilience

Delete min number of tuples to make Q false

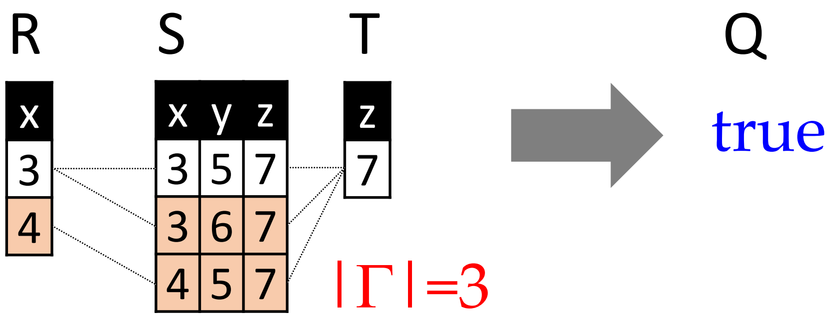


|Γ|=1

Q:-R(x),S(x,y,z),T(z)

Smallest witness problem

Delete max number of tuples while keeping Q true

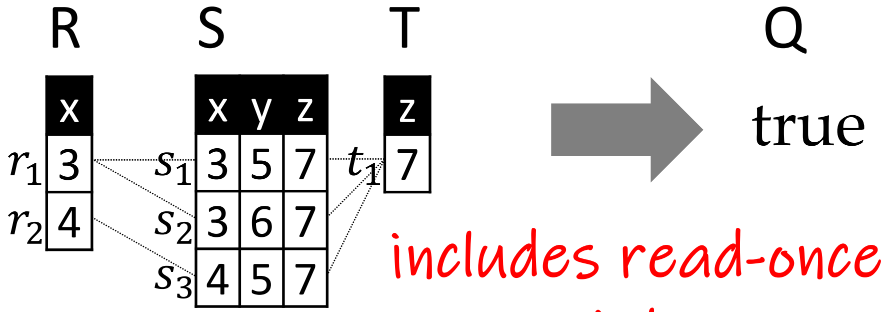


|Γ|=3

Minimal size factorization

Find a provenance factorization of minimum length

$\mathrm{Prov}(Q, D) = t_1(r_1(s_1 + s_2) + r_2 s_3)$ |Γ|=6



includes read-once provenance as special case

Miao, Roy, Yang. Explaining Wrong Queries Using Small Examples. SIGMOD 2019. https://doi.org/10.1145/3299869.3319866
Hu, Sintos. Finding Smallest Witnesses for Conjunctive Queries, ICDT 2024. https://cs.uwaterloo.ca/~xiaohu/icdt2024.pdf
Makhija, Gatterbauer: Towards a Dichotomy for Minimally Factorizing the Provenance of Self-Join Free Conjunctive Queries. https://arxiv.org/pdf/2105.14307

A recipe for solving RDM problems that goes a long way

1. Model problem as "appropriate" Integer Linear Program (ILP)
2. For what queries does the LP relaxation always return the same optimal objective ("ILP = LP")?

3. For queries with ILP ≠ LP: prove hardness

Interesting: hardness proofs can be automated! (with ASP Answer Set Programming)

4. For queries with ILP = LP: find a MFMC (Max-Flow Min-Cut) encoding

Interesting: the encoding can be quite different, e.g. the original constraint matrix is usually not Totally Unimodular

This is the difficult part!

Source: Makhija, Gatterbauer: A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations. SIGMOD'24. https://doi.org/10.1145/3626715

ILP for triangle query $Q^{\Delta}\text{:}-R(x,y),S(y,z),T(z,x)$

R

x	y	
1	1	r_{11}
2	1	r_{21}
2	3	r_{23}
5	3	r_{53}
5	1	r_{51}

$\bowtie$

S

y	z	
1	1	s_{11}
3	1	s_{31}
3	4	s_{34}
1	4	s_{14}

$\bowtie$

T

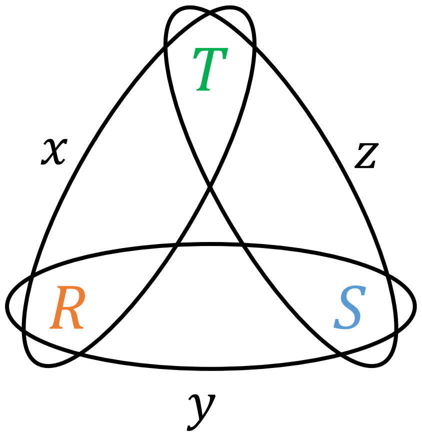
z	x	
1	1	t_{11}
1	2	t_{12}
4	2	t_{42}
4	5	t_{45}
4	1	t_{41}

$Q^{\Delta}\text{:}-R(x,y),S(y,z),T(z,x)$

Boolean query

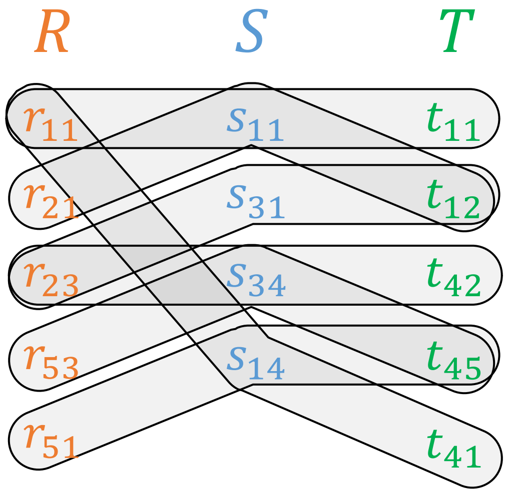
$Q_f^{\Delta}(x,y,z)\text{:}-R(x,y),S(y,z),T(z,x)$

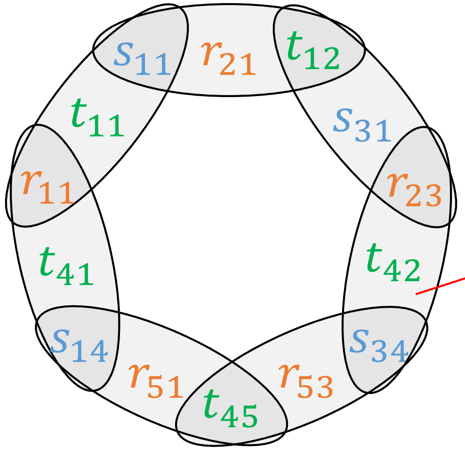
Full query



Query dual hypergraph
- atoms as vertices
- variables as edges

Data instance hypergraph
(tuples as vertices, "witnesses" as edges)





Q_f^{Δ}

	x	y	z	
w_1	1	1	1	$r_{11}s_{11}t_{11}$
w_2	2	1	1	$r_{21}s_{11}t_{12}$
w_3	2	3	1	$r_{23}s_{31}t_{12}$
w_4	2	3	4	$r_{23}s_{34}t_{42}$
w_5	5	3	4	$r_{53}s_{34}t_{45}$
w_6	5	1	4	$r_{51}s_{14}t_{45}$
w_7	1	1	4	$r_{11}s_{14}t_{41}$

ILP for triangle query Q^Δ: $-R(x,y), S(y,z), T(z,x)$

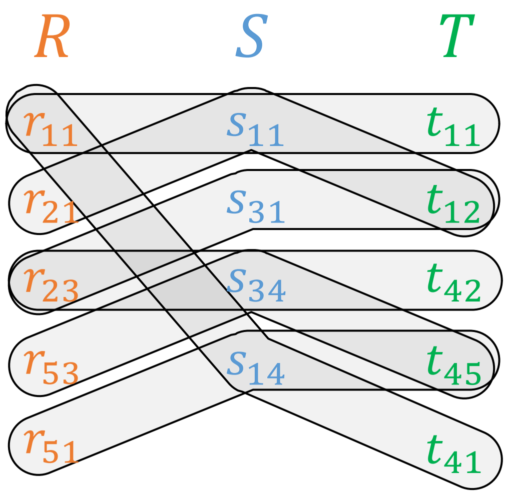
$$\min_{\mathbf{x}} \mathbf{c}^\mathrm{T}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in \{0,1\}^n$$

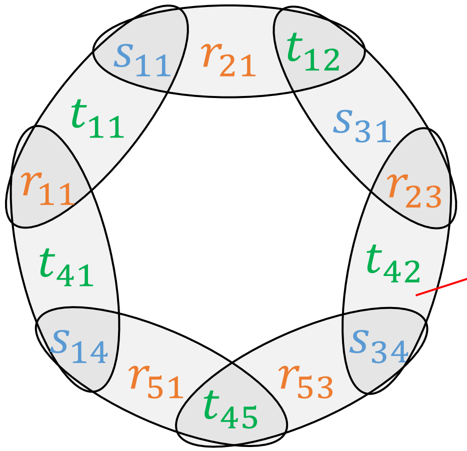
$\mathbf{c}$ objective vector
$\mathbf{A}$ constraint matrix
$\mathbf{b}$ constraint vector
$\mathbf{x}$ decision variables

$\mathbf{A}$ is 0,1
$\mathbf{b} = \mathbf{1}_m$
} set covering problems

$\mathbf{c}^\mathrm{T} = \mathbf{1}_n$ unweighted: $\min \sum \mathbf{x}$

Data instance hypergraph
(tuples as vertices, "witnesses" as edges)





Q_f^Δ

	x	y	z	
w_1	1	1	1	$r_{11}s_{11}t_{11}$
w_2	2	1	1	$r_{21}s_{11}t_{12}$
w_3	2	3	1	$r_{23}s_{31}t_{12}$
w_4	2	3	4	$r_{23}s_{34}t_{42}$
w_5	5	3	4	$r_{53}s_{34}t_{45}$
w_6	5	1	4	$r_{51}s_{14}t_{45}$
w_7	1	1	4	$r_{11}s_{14}t_{41}$

ILP for triangle query $Q^\Delta: -R(x,y), S(y,z), T(z,x)$

$$\min_{\mathbf{x}} \mathbf{c}^\mathrm{T}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in \{0,1\}^n$$

$\mathbf{c}$ objective vector
$\mathbf{A}$ constraint matrix
$\mathbf{b}$ constraint vector
$\mathbf{x}$ decision variables

$\mathbf{A}$ is 0,1
$\mathbf{b} = \mathbf{1}_m$
} set covering problems

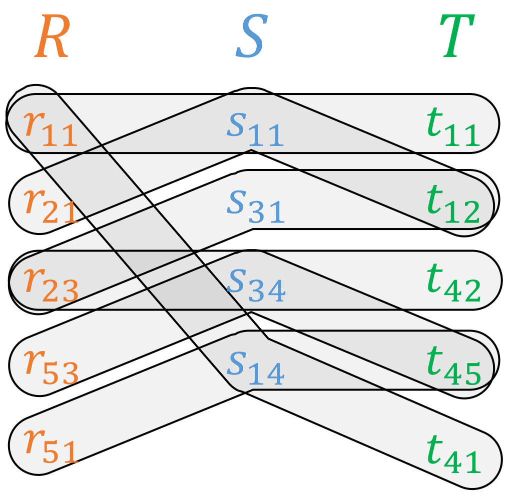
$\mathbf{c}^\mathrm{T} = \mathbf{1}_n$ unweighted: $\min \sum \mathbf{x}$

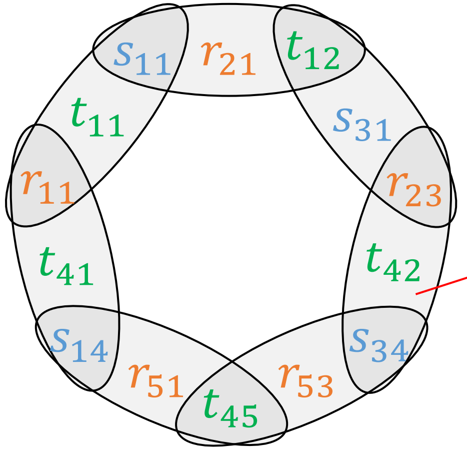
Objective

$$\min \quad x[r_{11}] + x[r_{21}] + x[r_{23}] + x[r_{53}] + x[r_{51}] + x[s_{11}] + x[s_{31}] + x[s_{34}] + x[s_{14}] + x[t_{11}] + x[t_{12}] + x[t_{42}] + x[t_{45}] + x[t_{42}]$$

optimal value: ?

Data instance hypergraph
(tuples as vertices, "witnesses" as edges)





Q_f^Δ

	x	y	z	
w_1	1	1	1	$r_{11}s_{11}t_{11}$
w_2	2	1	1	$r_{21}s_{11}t_{12}$
w_3	2	3	1	$r_{23}s_{31}t_{12}$
w_4	2	3	4	$r_{23}s_{34}t_{42}$
w_5	5	3	4	$r_{53}s_{34}t_{45}$
w_6	5	1	4	$r_{51}s_{14}t_{45}$
w_7	1	1	4	$r_{11}s_{14}t_{41}$

Constraints (one per witness)

$$x[r_{11}] + x[s_{11}] + x[t_{11}] \geq 1$$
$$x[r_{21}] + x[s_{11}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{31}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{34}] + x[t_{42}] \geq 1$$
$$x[r_{53}] + x[s_{34}] + x[t_{45}] \geq 1$$
$$x[r_{51}] + x[s_{14}] + x[t_{45}] \geq 1$$
$$x[r_{11}] + x[s_{14}] + x[t_{42}] \geq 1$$

$$\mathbf{x} \in \{0,1\}^{11}$$

ILP for triangle query Q^Δ: $-R(x,y), S(y,z), T(z,x)$

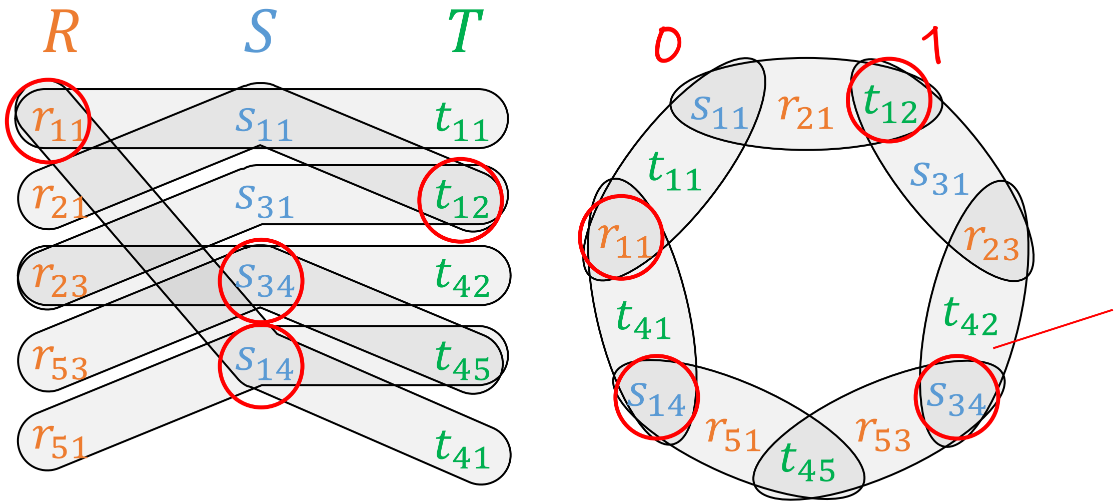
$$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$$
$$A\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in \{0,1\}^n$$

- $\mathbf{c}$ objective vector
- A constraint matrix
- $\mathbf{b}$ constraint vector
- $\mathbf{x}$ decision variables

A is 0,1; $\mathbf{b} = \mathbf{1}_m$ } set covering problems

$\mathbf{c}^T = \mathbf{1}_n$ unweighted: $\min \sum \mathbf{x}$

Data instance hypergraph
(tuples as vertices, "witnesses" as edges)



Q_f^Δ

	x	y	z	
w_1	1	1	1	$r_{11}s_{11}t_{11}$
w_2	2	1	1	$r_{21}s_{11}t_{12}$
w_3	2	3	1	$r_{23}s_{31}t_{12}$
w_4	2	3	4	$r_{23}s_{34}t_{42}$
w_5	5	3	4	$r_{53}s_{34}t_{45}$
w_6	5	1	4	$r_{51}s_{14}t_{45}$
w_7	1	1	4	$r_{11}s_{14}t_{41}$

Objective

$$\min\ x[r_{11}] + x[r_{21}] + x[r_{23}] + x[r_{53}] + x[r_{51}] + x[s_{11}] + x[s_{31}] + x[s_{34}] + x[s_{14}] + x[t_{11}] + x[t_{12}] + x[t_{42}] + x[t_{45}] + x[t_{42}]$$

=1 =0

optimal value: 4

Constraints (one per witness)

$$x[r_{11}] + x[s_{11}] + x[t_{11}] \geq 1$$
$$x[r_{21}] + x[s_{11}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{31}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{34}] + x[t_{42}] \geq 1$$
$$x[r_{53}] + x[s_{34}] + x[t_{45}] \geq 1$$
$$x[r_{51}] + x[s_{14}] + x[t_{45}] \geq 1$$
$$x[r_{11}] + x[s_{14}] + x[t_{42}] \geq 1$$

$$\mathbf{x} \in \{0,1\}^{11}$$

ILP for triangle query $Q^{\Delta}\text{:}-R(x,y), S(y,z), T(z,x)$

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in \{0,1\}^n$$

- $\mathbf{c}$ objective vector
- $\mathbf{A}$ constraint matrix
- $\mathbf{b}$ constraint vector
- $\mathbf{x}$ decision variables

$\mathbf{A}$ is 0,1; $\mathbf{b} = \mathbf{1}_m$ } set covering problems

$\mathbf{c}^{\mathrm{T}} = \mathbf{1}_n$ unweighted: $\min \sum \mathbf{x}$

Objective

$$\min\ x[r_{11}] + x[r_{21}] + x[r_{23}] + x[r_{53}] + x[r_{51}] + x[s_{11}] + x[s_{31}] + x[s_{34}] + x[s_{14}] + x[t_{11}] + x[t_{12}] + x[t_{42}] + x[t_{45}] + x[t_{42}]$$

=1 =0

optimal value: 4

A

	r_{11}	r_{21}	r_{23}	r_{53}	r_{51}	s_{11}	s_{31}	s_{34}	s_{14}	t_{11}	t_{12}	t_{42}	t_{45}	t_{42}
w_1	1					1				1				
w_2		1				1					1			
w_3			1				1				1			
w_4			1					1				1		
w_5				1				1					1	
w_6					1				1				1	
w_7	1								1				1	1

$\cdot\ \mathbf{x} = \begin{bmatrix} x[r_{11}] \\ x[r_{21}] \\ \dots \\ x[s_{11}] \\ x[s_{31}] \\ \dots \\ x[t_{45}] \\ x[t_{42}] \end{bmatrix} \geq \mathbf{1}_7 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Constraints (one per witness)

$$x[r_{11}] + x[s_{11}] + x[t_{11}] \geq 1$$
$$x[r_{21}] + x[s_{11}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{31}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{34}] + x[t_{42}] \geq 1$$
$$x[r_{53}] + x[s_{34}] + x[t_{45}] \geq 1$$
$$x[r_{51}] + x[s_{14}] + x[t_{45}] \geq 1$$
$$x[r_{11}] + x[s_{14}] + x[t_{42}] \geq 1$$

$$\mathbf{x} \in \{0,1\}^{11}$$

LP relaxation for triangle query Q^Δ: $-R(x,y), S(y,z), T(z,x)$

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in [0,1]^n$$
~~$\mathbf{x} \in \{0,1\}^n$~~

- $\mathbf{c}$ objective vector
- $\mathbf{A}$ constraint matrix
- $\mathbf{b}$ constraint vector
- $\mathbf{x}$ decision variables

$\mathbf{A}$ is 0,1
$\mathbf{b} = \mathbf{1}_m$ } fractional set covering problems
$\mathbf{c}^{\mathrm{T}} = \mathbf{1}_n$ unweighted: $\min \sum \mathbf{x}$

Objective

$$\min x[r_{11}] + x[r_{21}] + x[r_{23}] + x[r_{53}] + x[r_{51}] + x[s_{11}] + x[s_{31}] + x[s_{34}] + x[s_{14}] + x[t_{11}] + x[t_{12}] + x[t_{42}] + x[t_{45}] + x[t_{42}]$$

=1 =0

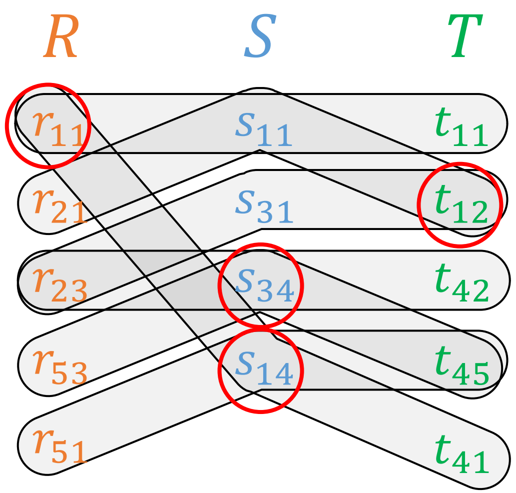
optimal value: ?

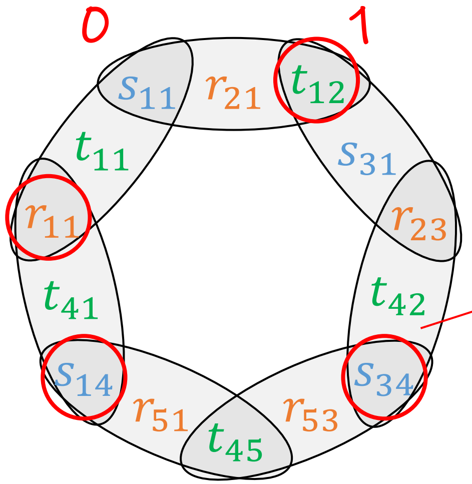
Constraints (one per witness)

$$x[r_{11}] + x[s_{11}] + x[t_{11}] \geq 1$$
$$x[r_{21}] + x[s_{11}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{31}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{34}] + x[t_{42}] \geq 1$$
$$x[r_{53}] + x[s_{34}] + x[t_{45}] \geq 1$$
$$x[r_{51}] + x[s_{14}] + x[t_{45}] \geq 1$$
$$x[r_{11}] + x[s_{14}] + x[t_{42}] \geq 1$$

~~$\mathbf{x} \in \{0,1\}^{11}$~~
$\mathbf{x} \in [0,1]^{11}$

Data instance hypergraph
(tuples as vertices, "witnesses" as edges)





Q_f^Δ

	x	y	z	
w_1	1	1	1	$r_{11}s_{11}t_{11}$
w_2	2	1	1	$r_{21}s_{11}t_{12}$
w_3	2	3	1	$r_{23}s_{31}t_{12}$
w_4	2	3	4	$r_{23}s_{34}t_{42}$
w_5	5	3	4	$r_{53}s_{34}t_{45}$
w_6	5	1	4	$r_{51}s_{14}t_{45}$
w_7	1	1	4	$r_{11}s_{14}t_{41}$

LP relaxation for triangle query Q^Δ: $-R(x,y), S(y,z), T(z,x)$

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in [0,1]^n$$
~~$\mathbf{x} \in \{0,1\}^n$~~

c objective vector
A constraint matrix
b constraint vector
x decision variables

A is 0,1
$\mathbf{b} = \mathbf{1}_m$
} fractional set covering problems

$\mathbf{c}^{\mathrm{T}} = \mathbf{1}_n$ unweighted: $\min \sum \mathbf{x}$

Objective

$$\min\ x[r_{11}] + x[r_{21}] + x[r_{23}] + x[r_{53}] + x[r_{51}] + x[s_{11}] + x[s_{31}] + x[s_{34}] + x[s_{14}] + x[t_{11}] + x[t_{12}] + x[t_{42}] + x[t_{45}] + x[t_{42}]$$

=½ =0

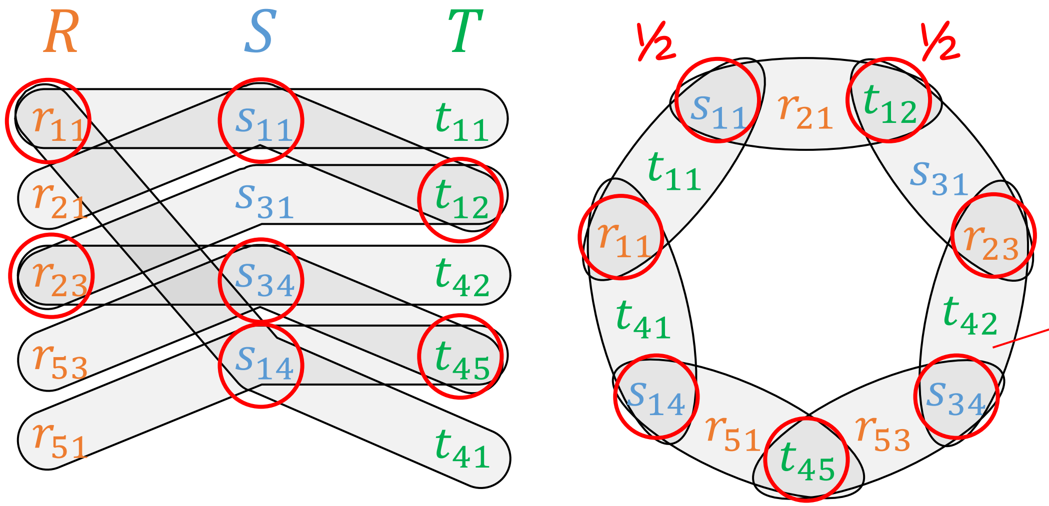
optimal value: 3.5

Constraints (one per witness)

$$x[r_{11}] + x[s_{11}] + x[t_{11}] \geq 1$$
$$x[r_{21}] + x[s_{11}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{31}] + x[t_{12}] \geq 1$$
$$x[r_{23}] + x[s_{34}] + x[t_{42}] \geq 1$$
$$x[r_{53}] + x[s_{34}] + x[t_{45}] \geq 1$$
$$x[r_{51}] + x[s_{14}] + x[t_{45}] \geq 1$$
$$x[r_{11}] + x[s_{14}] + x[t_{42}] \geq 1$$

~~$\mathbf{x} \in \{0,1\}^{11}$~~
$$\mathbf{x} \in [0,1]^{11}$$

Data instance hypergraph
(tuples as vertices, "witnesses" as edges)



Q^Δ_f

	x	y	z	
w_1	1	1	1	$r_{11}s_{11}t_{11}$
w_2	2	1	1	$r_{21}s_{11}t_{12}$
w_3	2	3	1	$r_{23}s_{31}t_{12}$
w_4	2	3	4	$r_{23}s_{34}t_{42}$
w_5	5	3	4	$r_{53}s_{34}t_{45}$
w_6	5	1	4	$r_{51}s_{14}t_{45}$
w_7	1	1	4	$r_{11}s_{14}t_{41}$

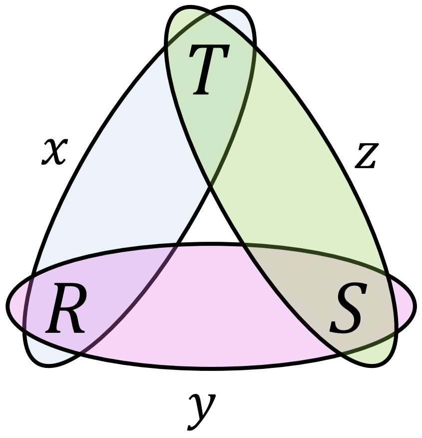
Outline

- Resilience for self-join free CQs
  - hard & easy cases
- Resilience for general CQs
  - hard & easy cases

"Active triads" as hardness criterion for SJ-free CQs

Triangle query

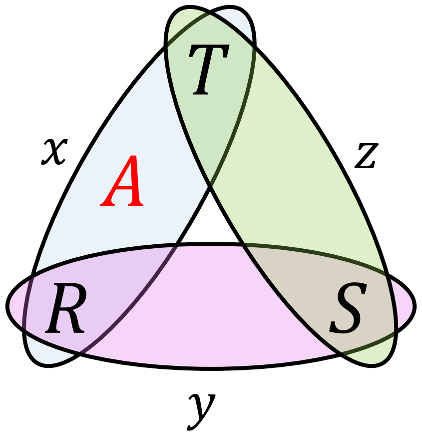
$Q^{\Delta}\text{:-}R(x,y), S(y,z), T(x,z)$



Query dual hypergraph
- atoms as vertices
- variables as edges

Triangle unary

$Q^{\Delta}_{A}\text{:-}R(x,y), S(y,z), T(x,z), A(x)$

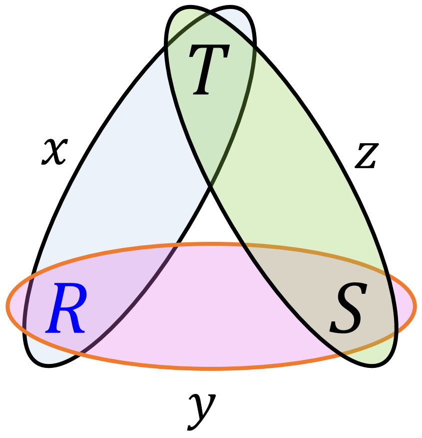


Originally just called "triad" in [Freire+ VLDB'15]. Renamed to "active triad" in [Makhija+. SIGMOD'24] as new notion allows more fine-grained analysis under both set and bag semantics

"Active triads" as hardness criterion for SJ-free CQs

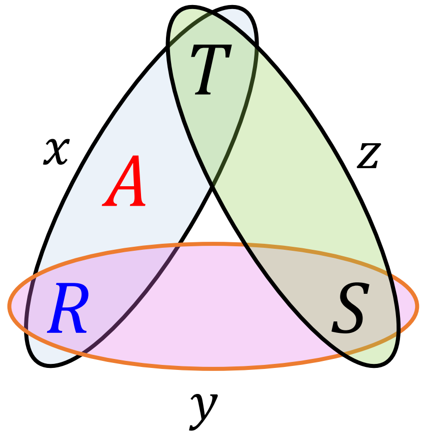
Triangle query

$Q^{\Delta}\text{:} -R(x, y), S(y, z), T(x, z)$



Triangle unary

$Q^{\Delta}_{A}\text{:} -R(x, y), S(y, z), T(x, z), A(x)$



The path $R - y - S$ from R to S is INDEPENDENT of $T(x, z)$, as it uses no variable from $T\{x, z\}$

Atom $R(x, y)$ is UNDOMINATED since there is no other atom S s.t. $\text{var}(S) \subset \text{var}(R)$

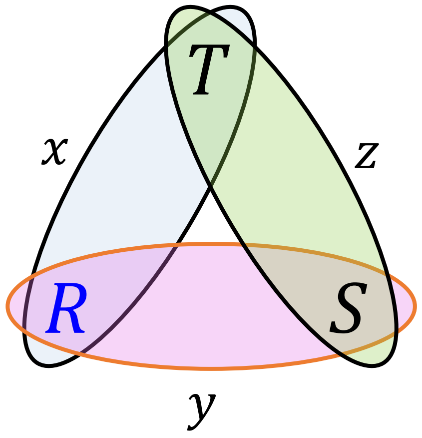
Atom R is DOMINATED since $\text{var}(A) = \{x\} \subset \{\text{x}, \text{y}\} = \text{var}(R)$

Originally just called "triad" in [Freire+ VLDB'15]. Renamed to "active triad" in [Makhija+. SIGMOD'24] as new notion allows more fine-grained analysis under both set and bag semantics

"Active triads" as hardness criterion for SJ-free CQs

Triangle query

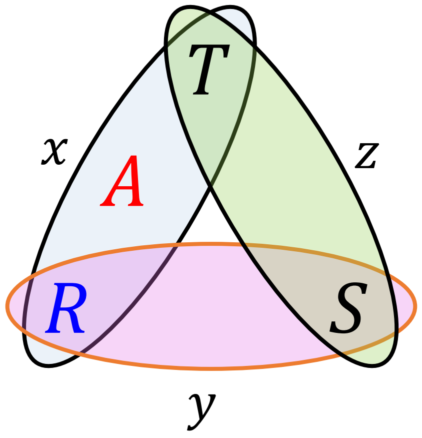
Q^{Δ}:$-R(x,y), S(y,z), T(x,z)$



NPC

Triangle unary

Q^{Δ}_{A}:$-R(x,y), S(y,z), T(x,z), A(x)$



PTIME

The path $R - y - S$ from R to S is INDEPENDENT of $T(x,z)$, as it uses no variable from $T\{x,z\}$

Atom $R(x,y)$ is UNDOMINATED since there is no other atom S s.t. $\text{var}(S) \subset \text{var}(R)$

Atom R is DOMINATED since $\text{var}(A) = \{x\} \subset \{\text{x}, \text{y}\} = \text{var}(R)$

ACTIVE TRIAD: 3 undominated atoms R,S,T that have independent paths among each other

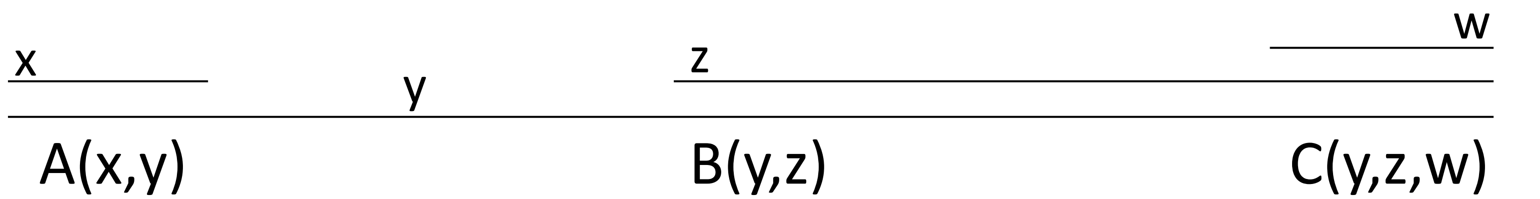
TRIAD: 3 atoms R,S,T that have independent paths among each other

does not yet make a query hard (under set semantics)

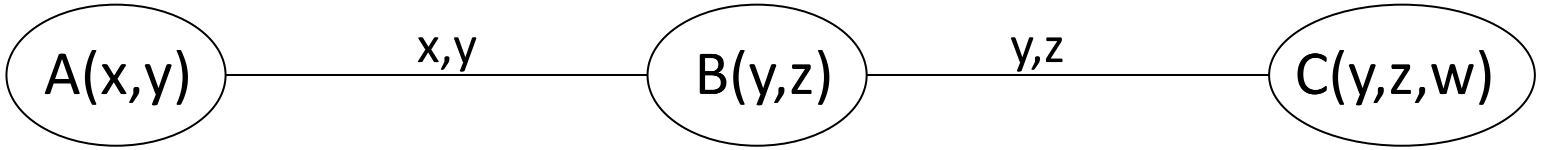
Originally just called "triad" in [Freire+ VLDB'15]. Renamed to "active triad" in [Makhija+. SIGMOD'24] as new notion allows more fine-grained analysis under both set and bag semantics

Example linear query $Q{:-}A(x,y), B(y,z), C(y,z,w)$

Interval graph:



Join path:

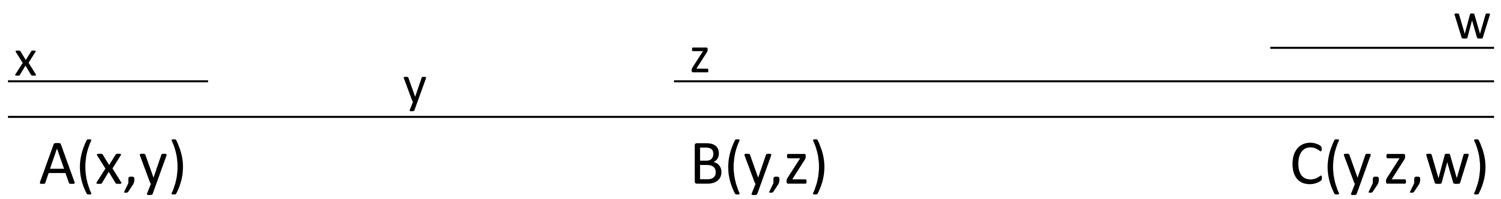


DEFINITION: A query q is LINEAR if it has a join path (i.e. a join tree that forms a path)

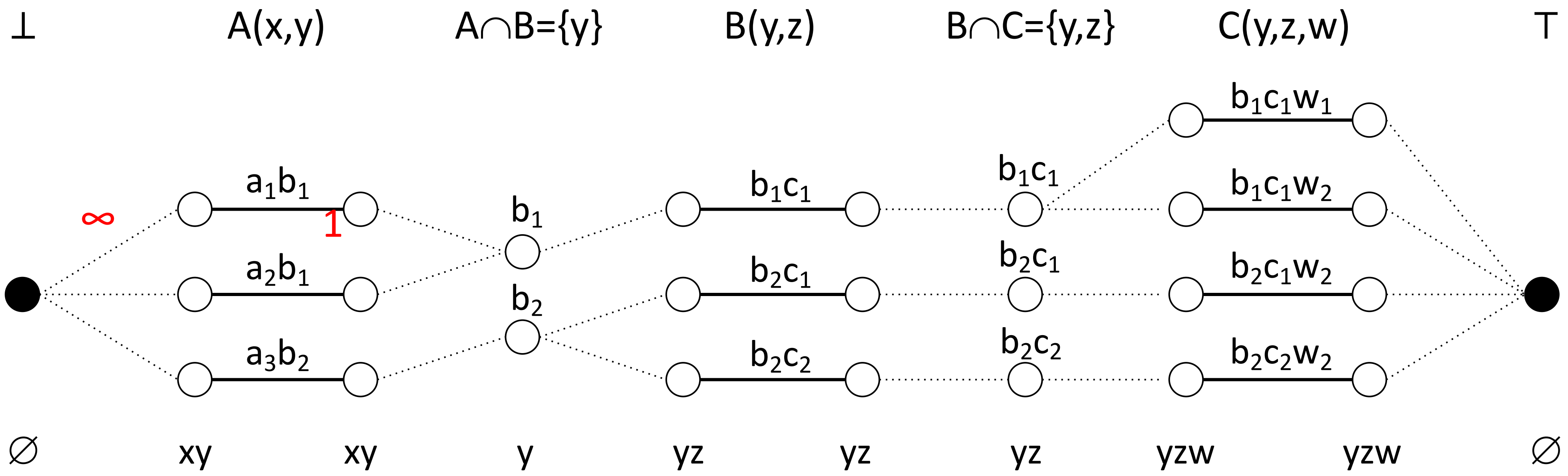
Linear queries were originally defined via "linear orders" in [Meliou+ VLDB'10] and [Freire+ VLDB'15]

Example linear query $Q\!:\!-A(x,y),B(y,z),C(y,z,w)$

Interval graph:



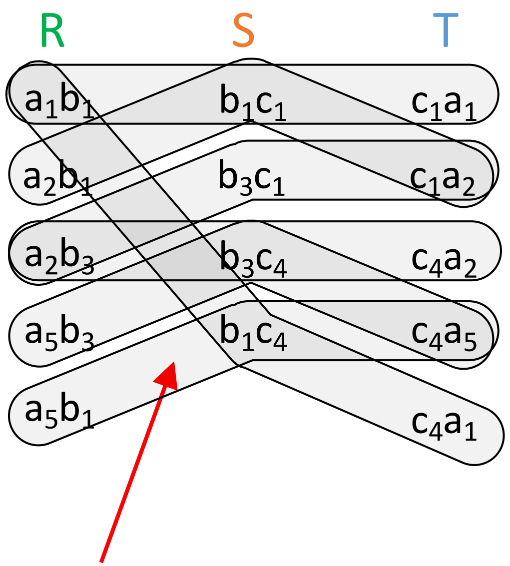
Source-Target flow graph encoding (MFMC)

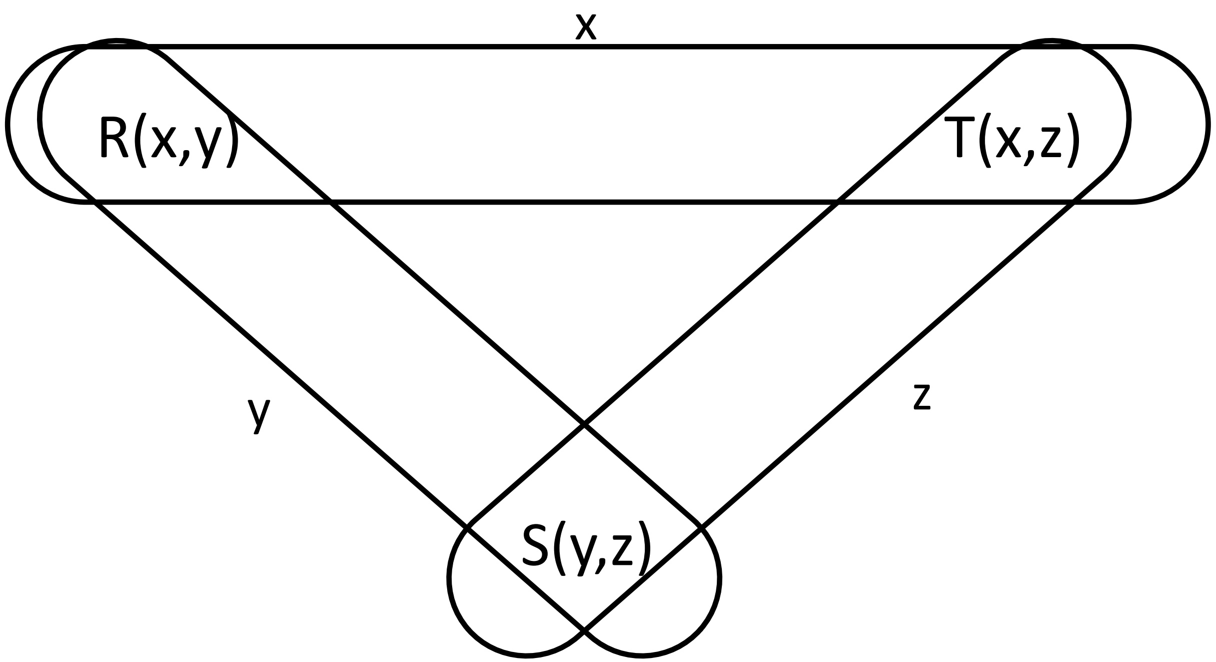


Non-Linearizable Queries: those with active triads

$Q^{\Delta}\text{:}- R(x,y), S(y,z), T(x,z)$

Triangle query





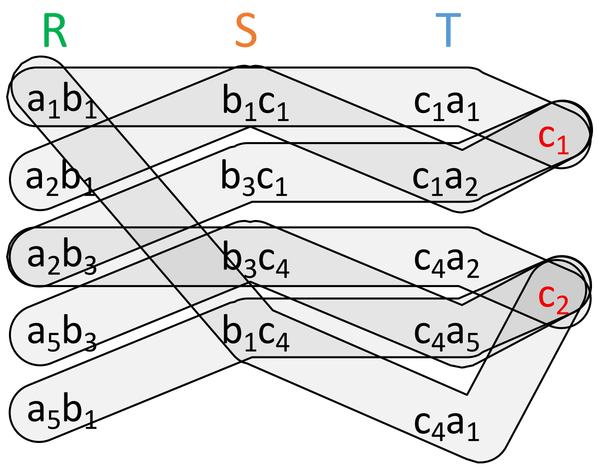
each hyperedge corresponds to one witness,
each tuple of domain values to one database tuple,
if Q is evaluated over the database over all shown tuples

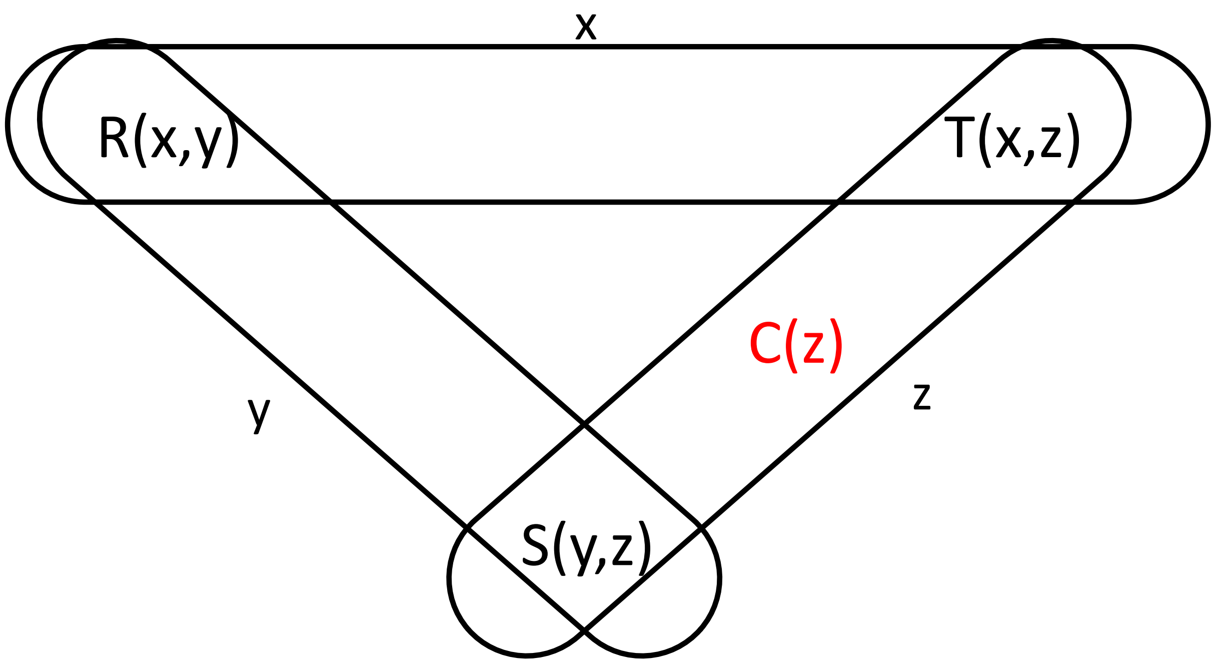
"Hard instance" (more complicated than it needs to be). A 6-hole in the constraint matrix of the associated ILP. We can also make a 3-hole

Linearizable Queries: "domination" to the rescue

$Q_C^\Delta{:}-R(x,y),S(y,z),T(x,z),C(z)$

Triangle unary



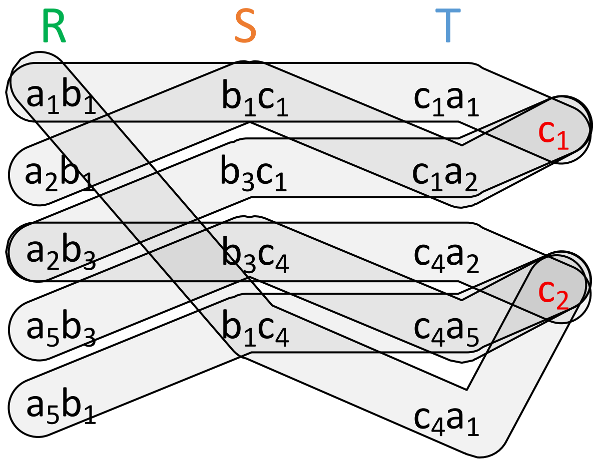


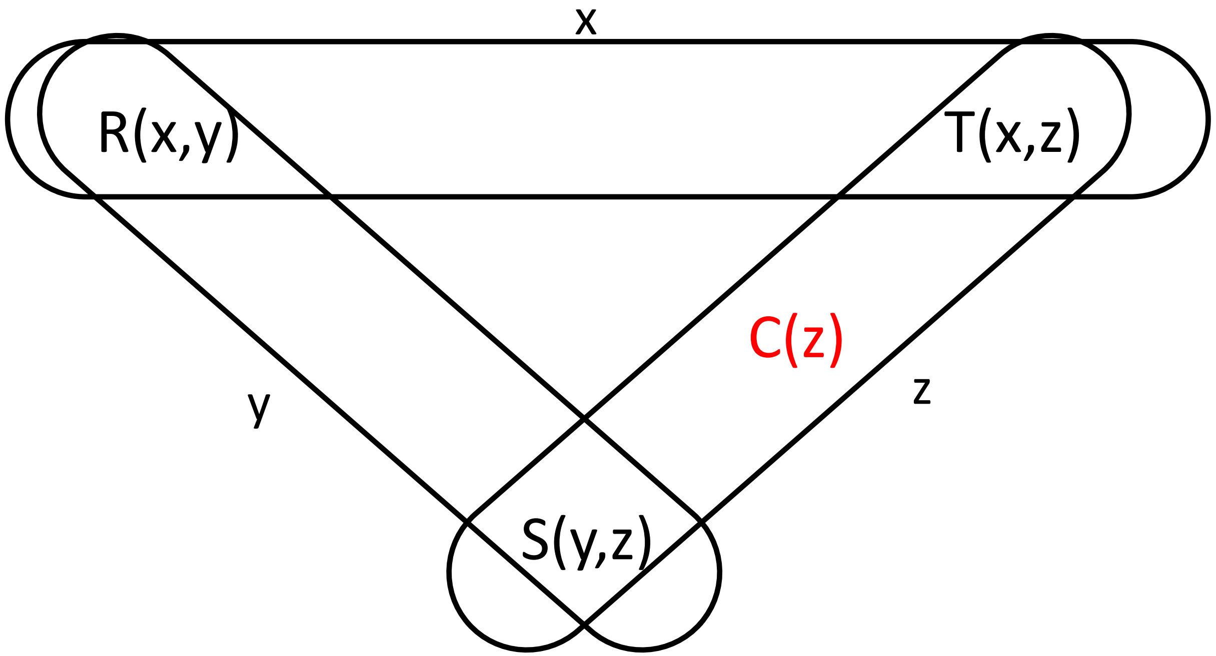
This instance (similar for smaller 3-hole) can be solved with flow. Key insight: The resulting flow graph does not "correspond exactly" to the original instance, but a *minimal* cut of the resulting implementation still corresponds to a *minimal* VC of the original problem

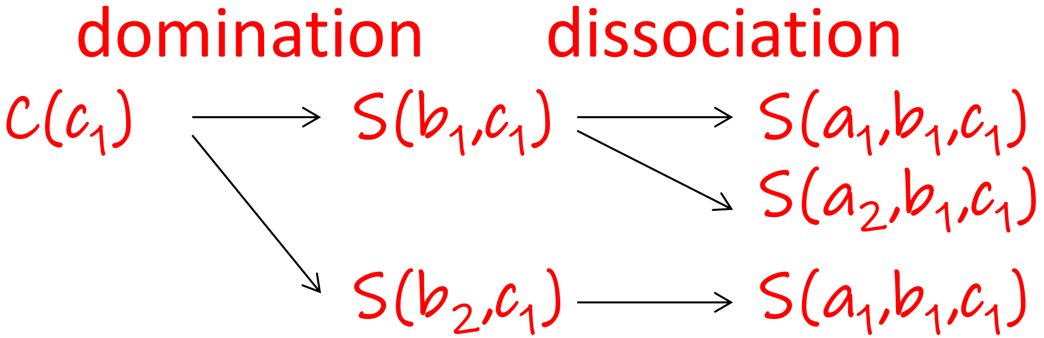
Linearizable Queries: "domination" to the rescue

$Q_C^{\Delta}\!: -R(x,y), S(y,z), T(x,z), C(z)$

Triangle unary





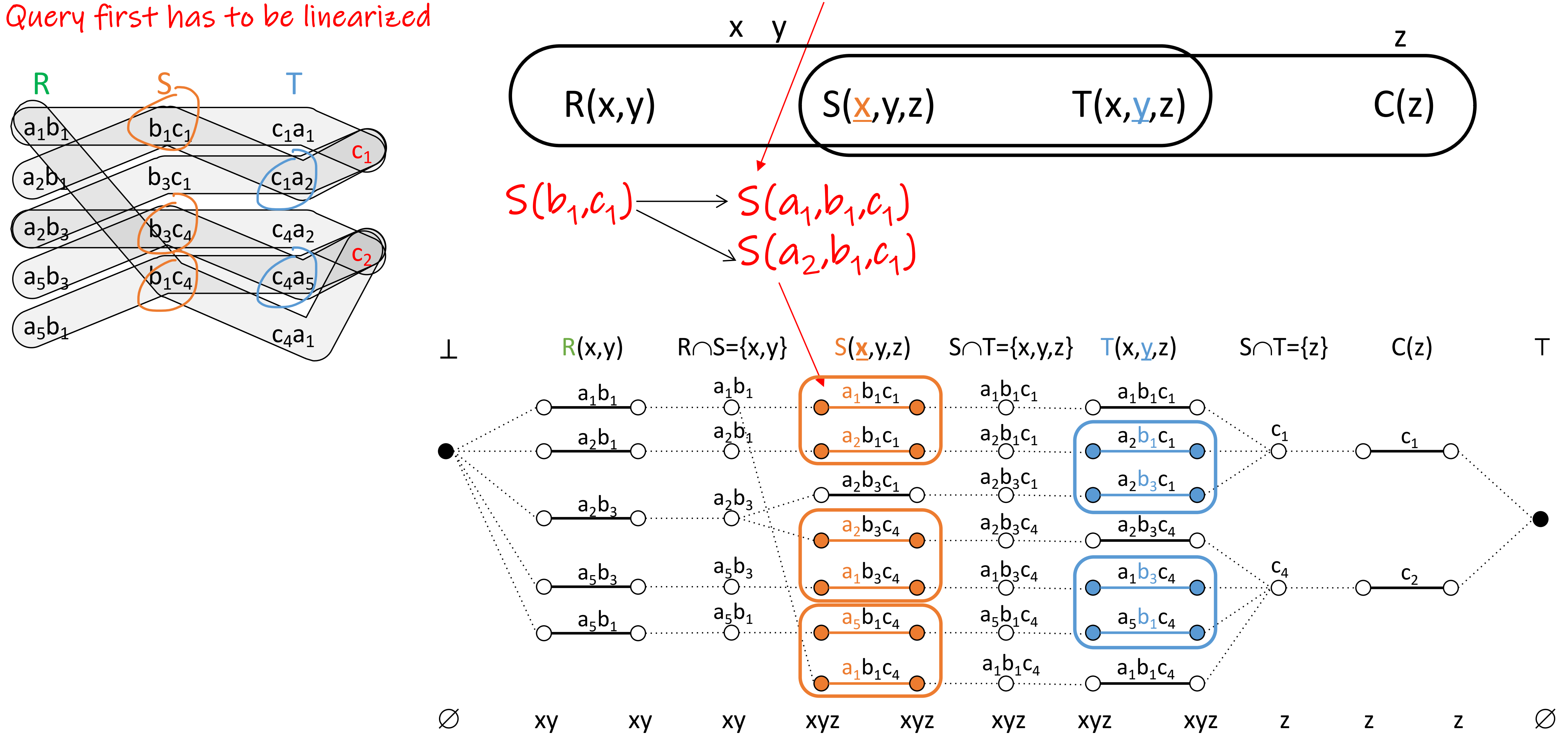


Linearizable Queries

$Q_C^{\Delta'}$:− $R(x,y), S(x,y,z), T(x,y,z), C(z)$

Query first has to be linearized

Notice that these are now "dissociated" tuples, they don't correspond anymore to real tuples. But we would never need to pick them anyway ☺



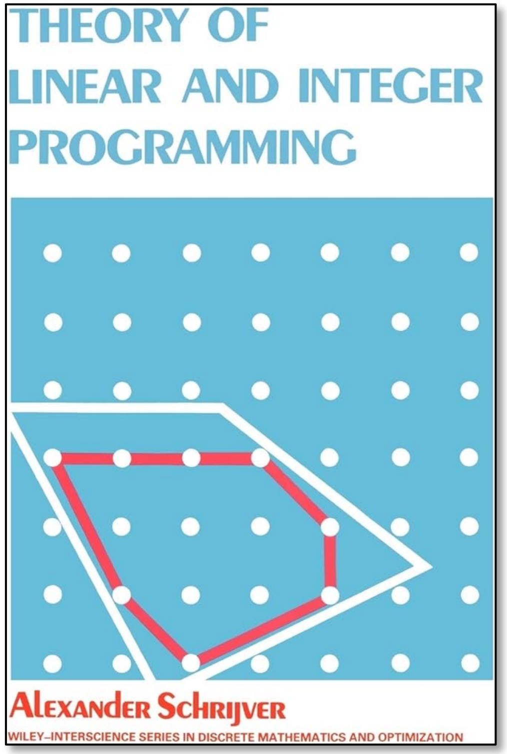
Polyhedral theory on ILP for triangle unary query

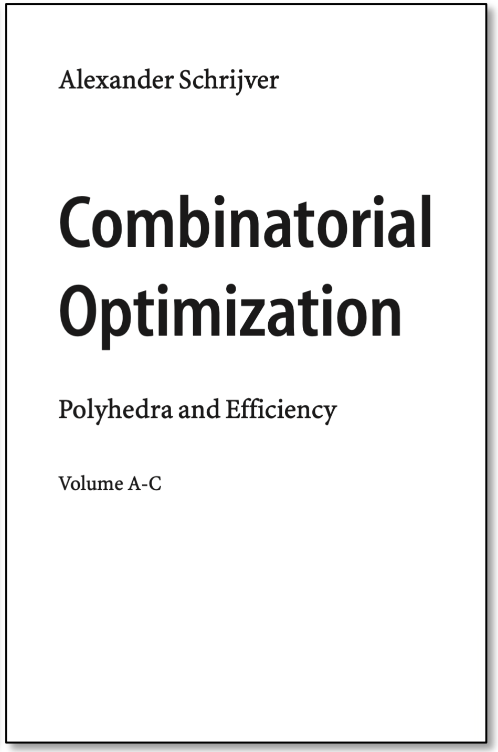
$$\min \ \mathbf{c}^{\mathrm{T}}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in \{0,1\}^n$$

objective vector

constraint vector

constraint matrix





Polyhedral theory on ILP for triangle unary query: not useful ☹

$$\min \ \mathbf{c}^{\mathrm{T}}\mathbf{x}$$

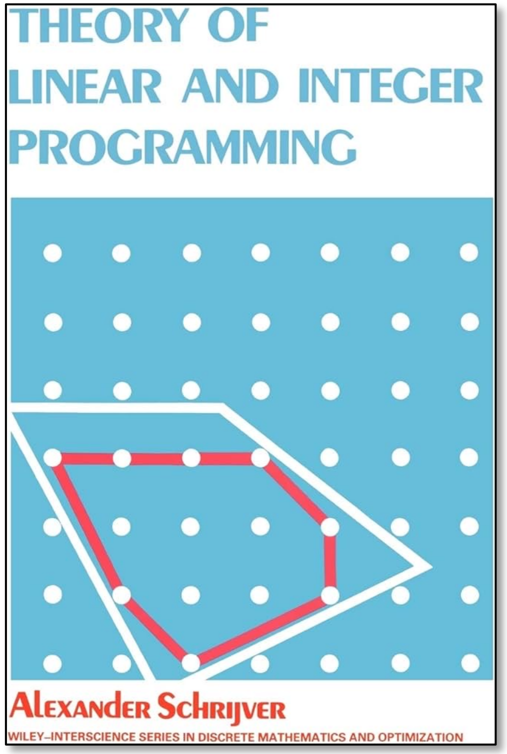
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$

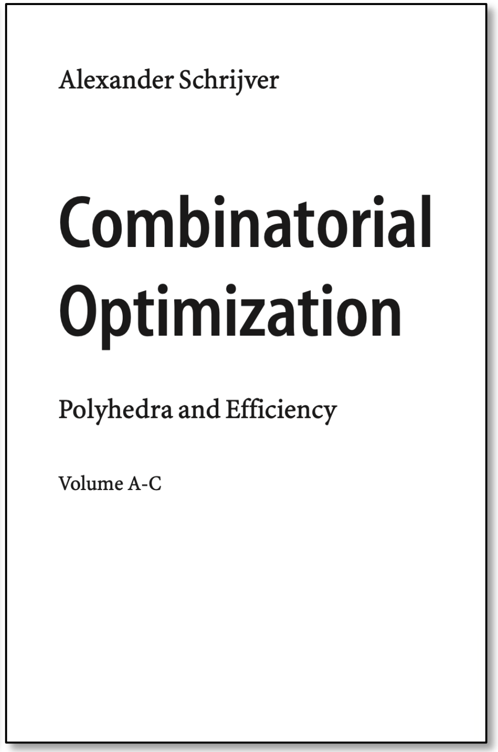
$$\mathbf{x} \in \{0,1\}^n$$

objective vector → $\mathbf{c}$

constraint vector → $\mathbf{b}$

constraint matrix → $\mathbf{A}$





- Focus of polyhedral theory mainly on constraint matrix $\mathbf{A}$. But our PTIME constraint matrixes need not be balanced, nor Totally Unimodular, etc.

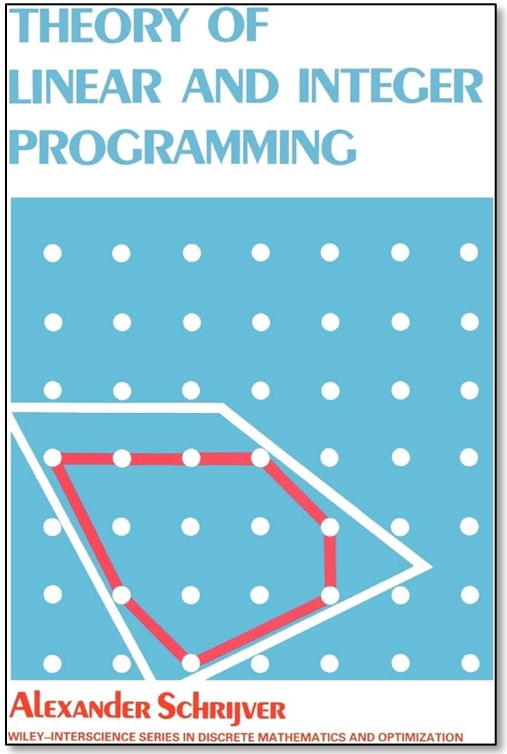
Polyhedral theory on ILP for triangle unary query: not useful ☹

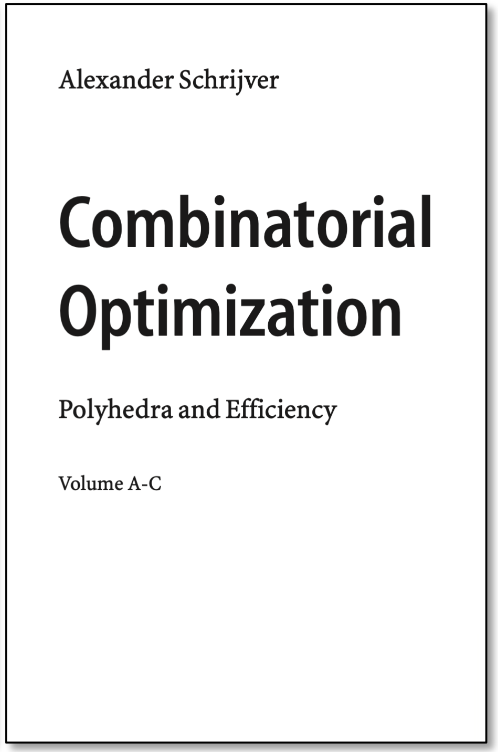
$$\min \ \mathbf{c}^{\mathrm{T}}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \in \{0,1\}^n$$

objective vector → $\mathbf{c}$

constraint matrix → $\mathbf{A}$

constraint vector → $\mathbf{b}$





- Focus of polyhedral theory mainly on constraint matrix $\mathbf{A}$. But our PTIME constraint matrixes need not be balanced, nor Totally Unimodular, etc.

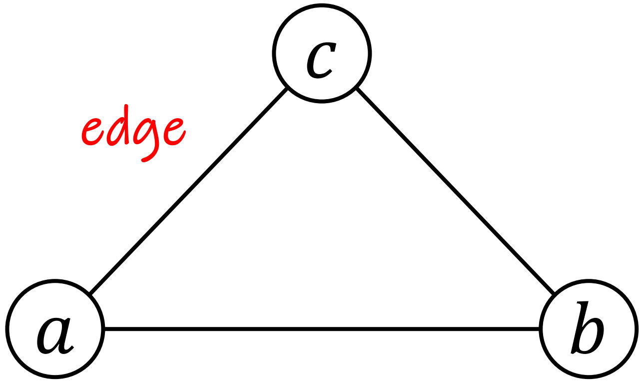
1. Our complexity results take into account the objective vector $\mathbf{c}$!
2. This gives us a separation between the problem under set vs. bag semantics!
3. We use an indirect proof via the earlier-described MFMC encoding ☺

Source: Makhija, Gatterbauer: A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations. SIGMOD'24. https://doi.org/10.1145/3626715

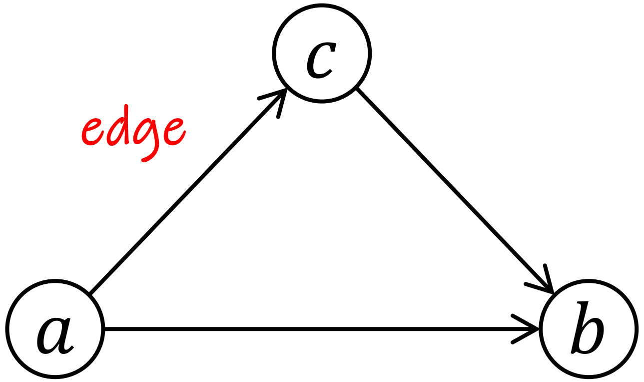
Outline

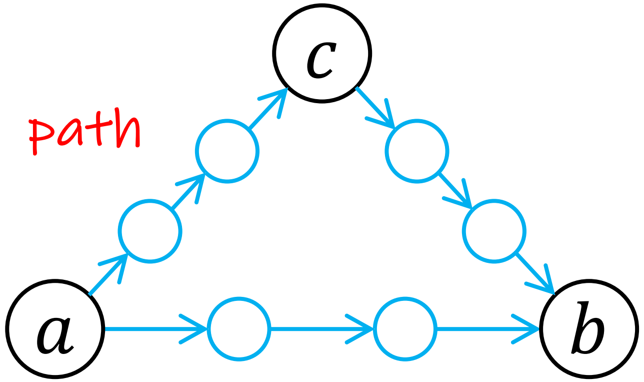
- Resilience for self-join free CQs
  - hard & easy cases
- Resilience for general CQs
  - hard & easy cases

A template of hardness reduction from minimal VC



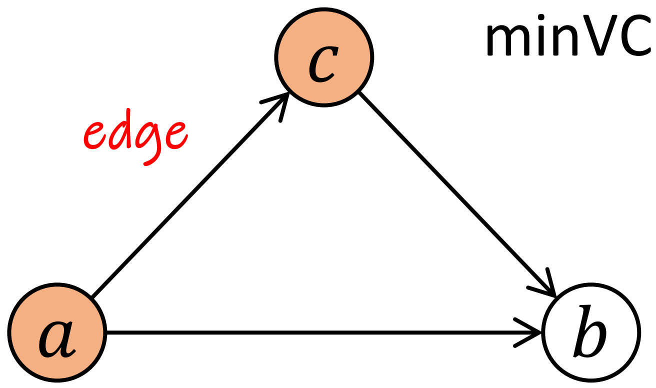
A template of hardness reduction from minimal VC



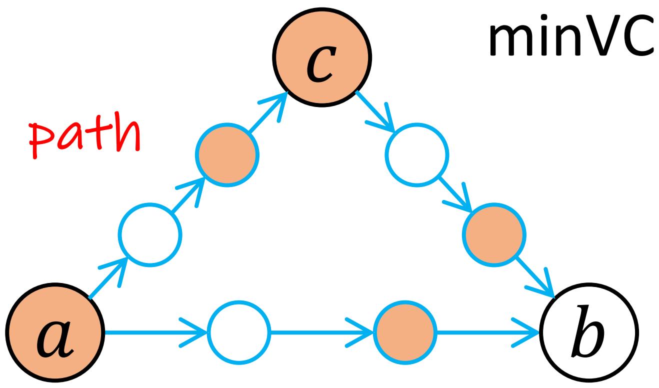


What is the minimum VC for either graph ?

A template of hardness reduction from minimal VC



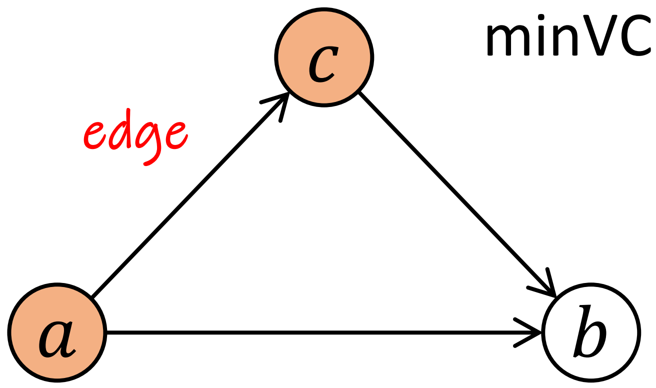
minVC = 2



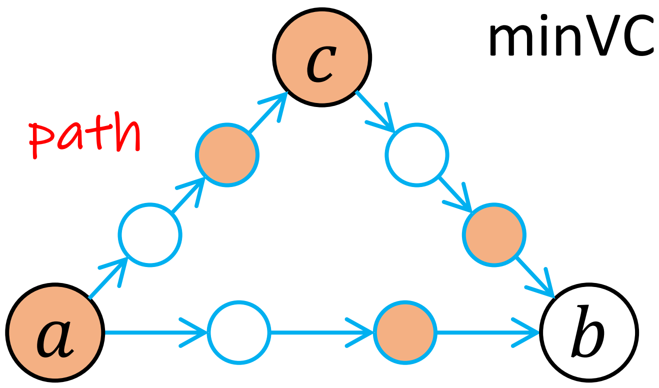
minVC = 5

Side-topic:
What is the minimum fractional VC for either graph ?
(Assign weights in [0,1] to vertices s.t. the sum of weights of endpoints for each edge is ≥ 1)

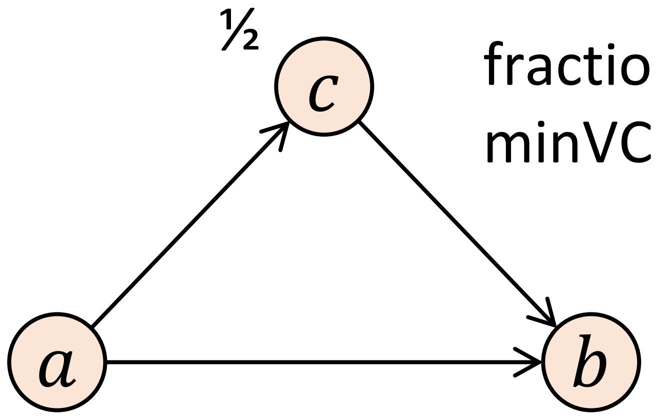
A template of hardness reduction from minimal VC



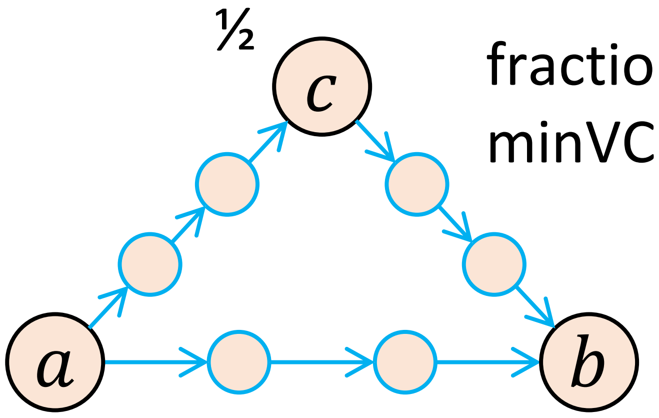
minVC = 2



minVC = 5

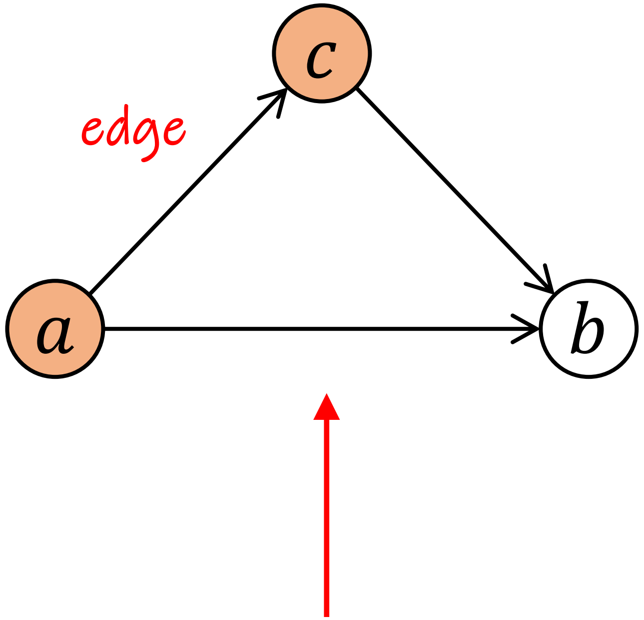


fractional
minVC = 1.5



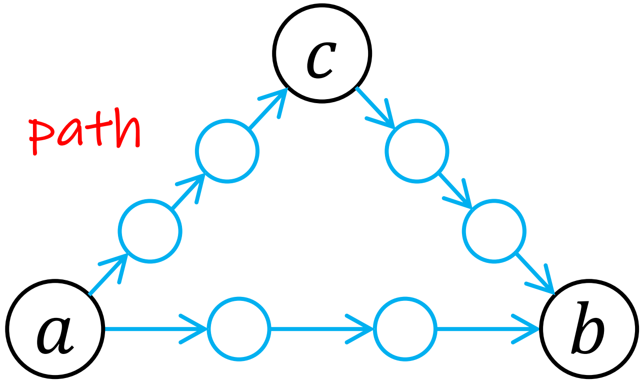
fractional
minVC = 4.5

A template of hardness reduction from minimal VC



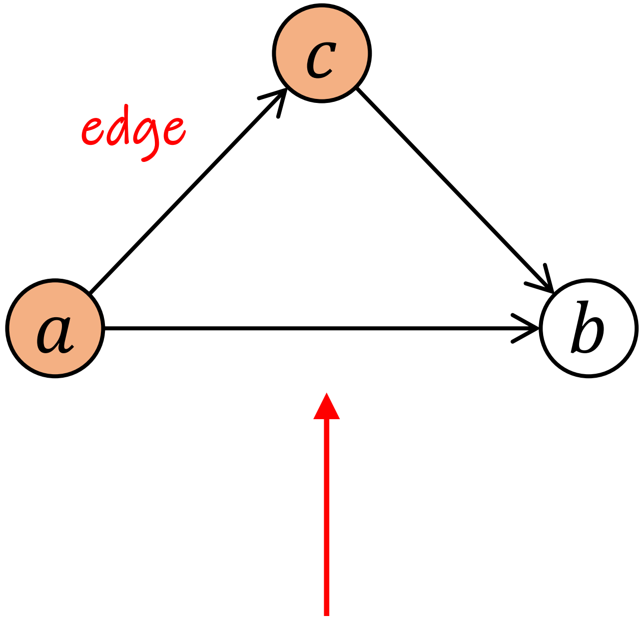
VC of size k

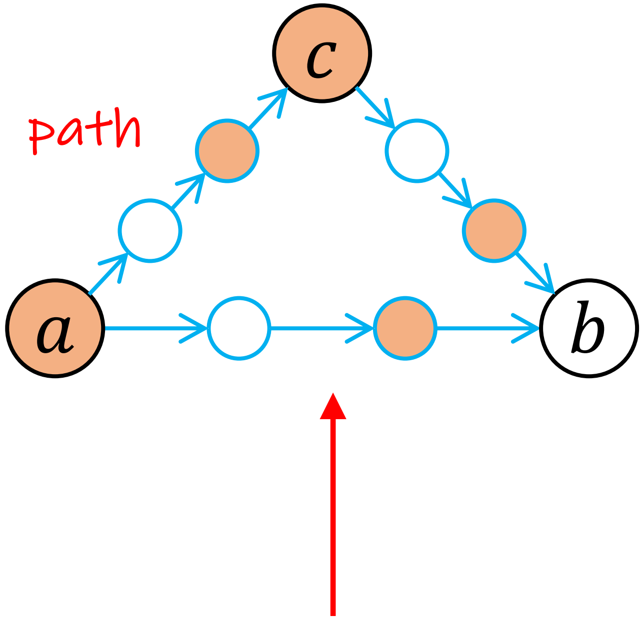
⇒



VC of size ?

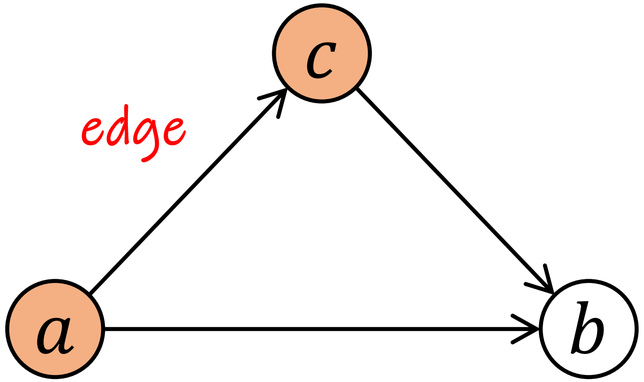
A template of hardness reduction from minimal VC

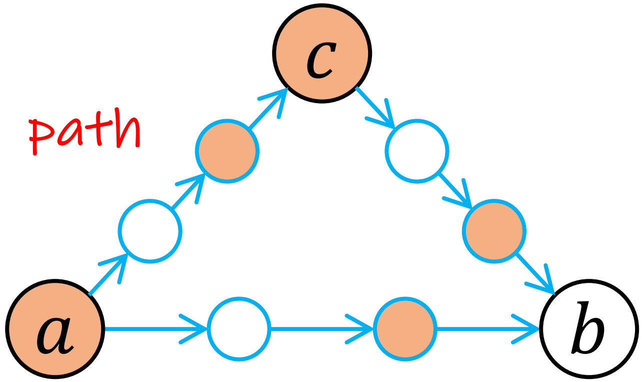




VC of size k ⇒ VC of size k + # edges

A template of hardness reduction from minimal VC

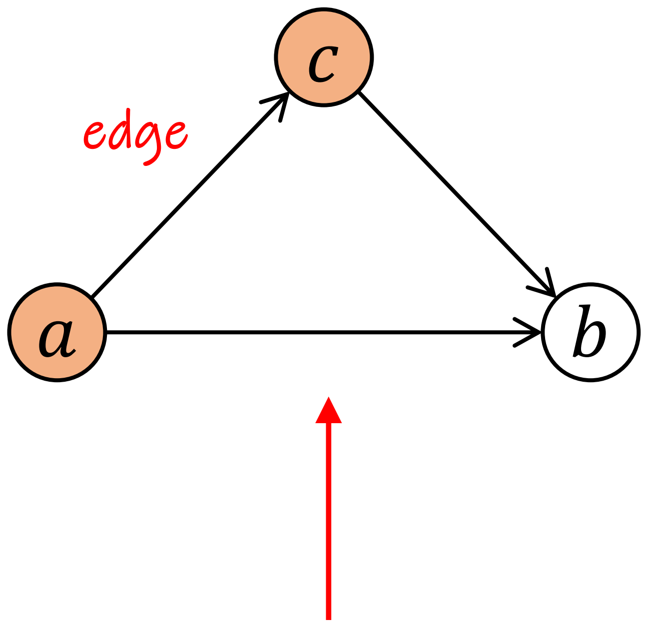


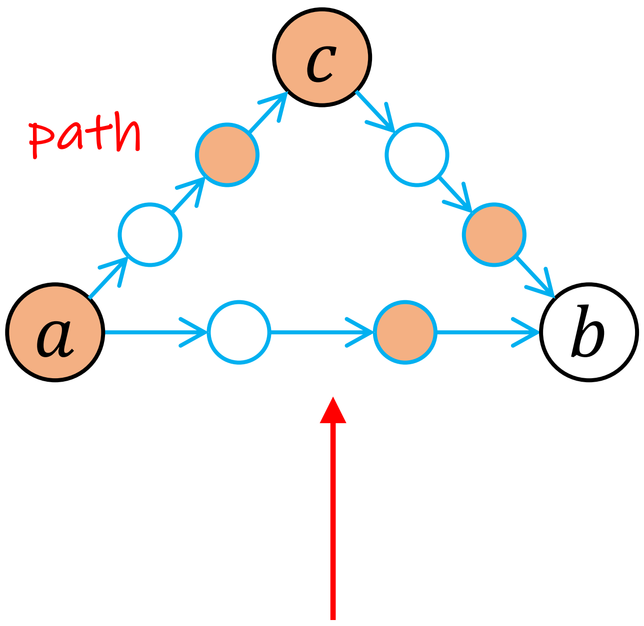


VC of size k ⇐ VC of size k + # edges

?

A template of hardness reduction from minimal VC

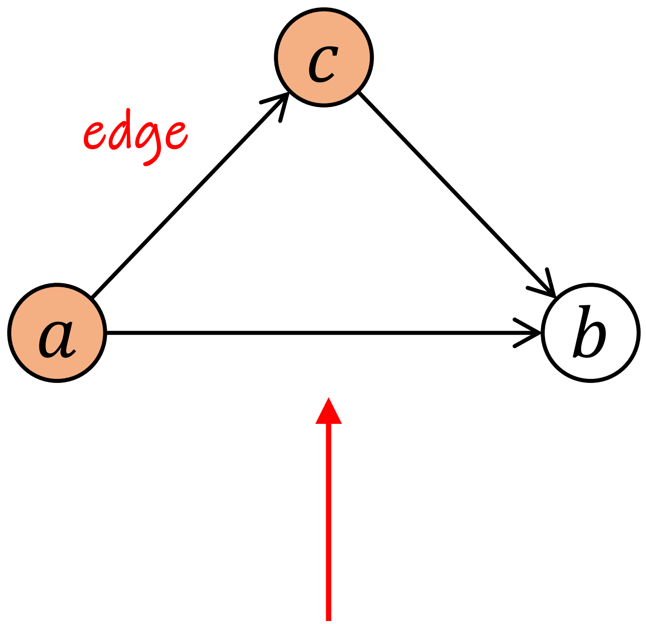


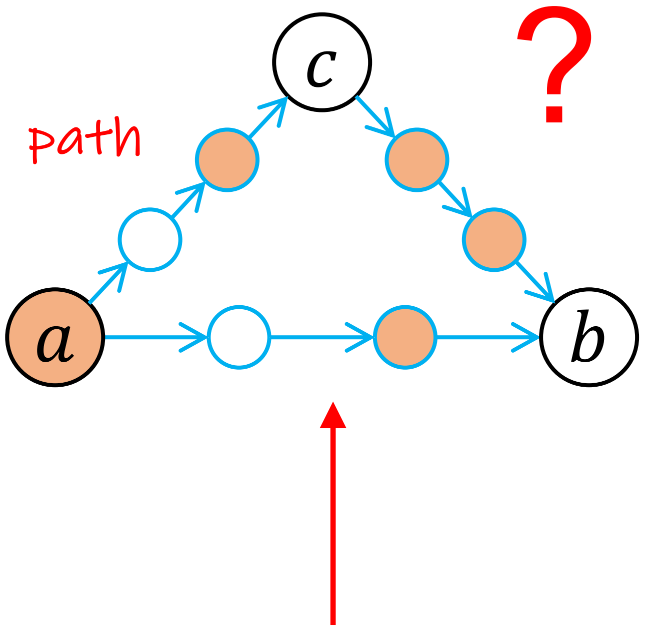


VC of size k ⟺ VC of size k + # edges

(as long as k ≤ # vertices)

A template of hardness reduction from minimal VC

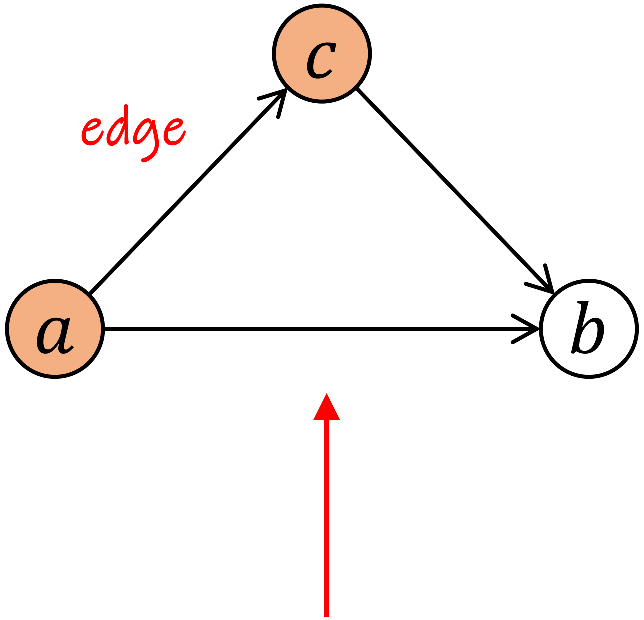


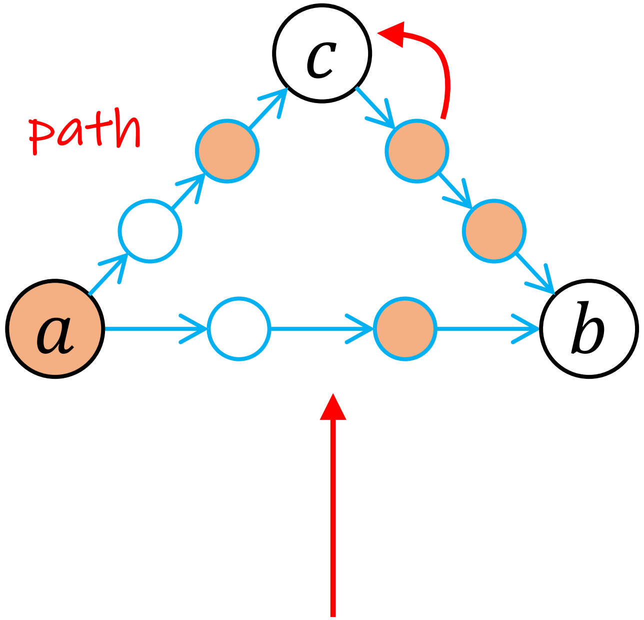


VC of size k ⟺ VC of size k + # edges

(as long as k ≤ # vertices)

A template of hardness reduction from minimal VC

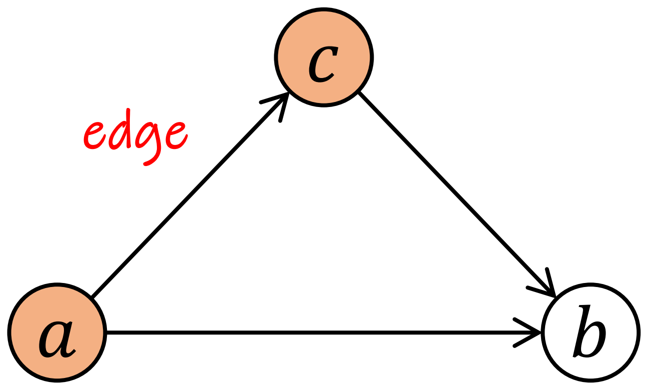


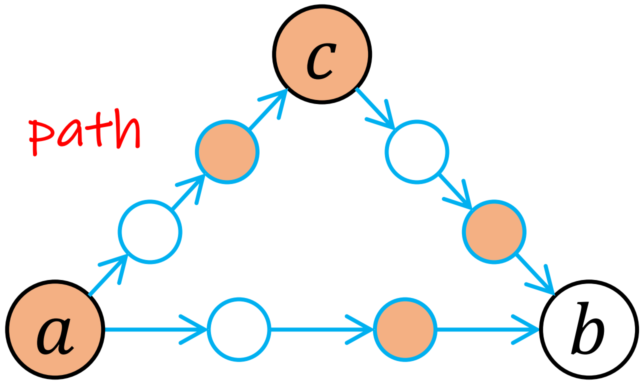


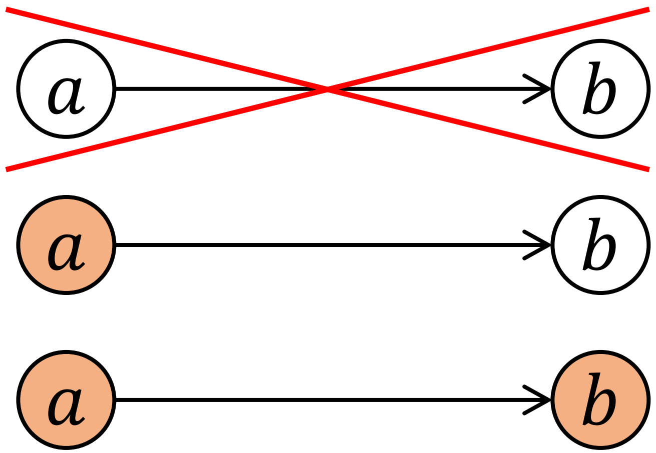
VC of size k ⟺ VC of size k + # edges

(as long as k ≤ # vertices)

A template of hardness reduction from minimal VC

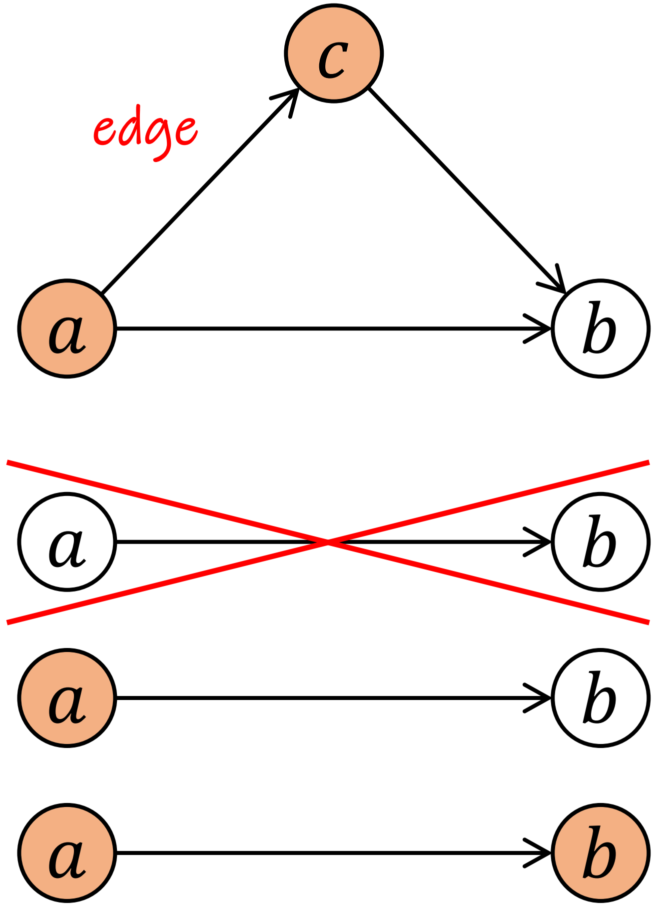


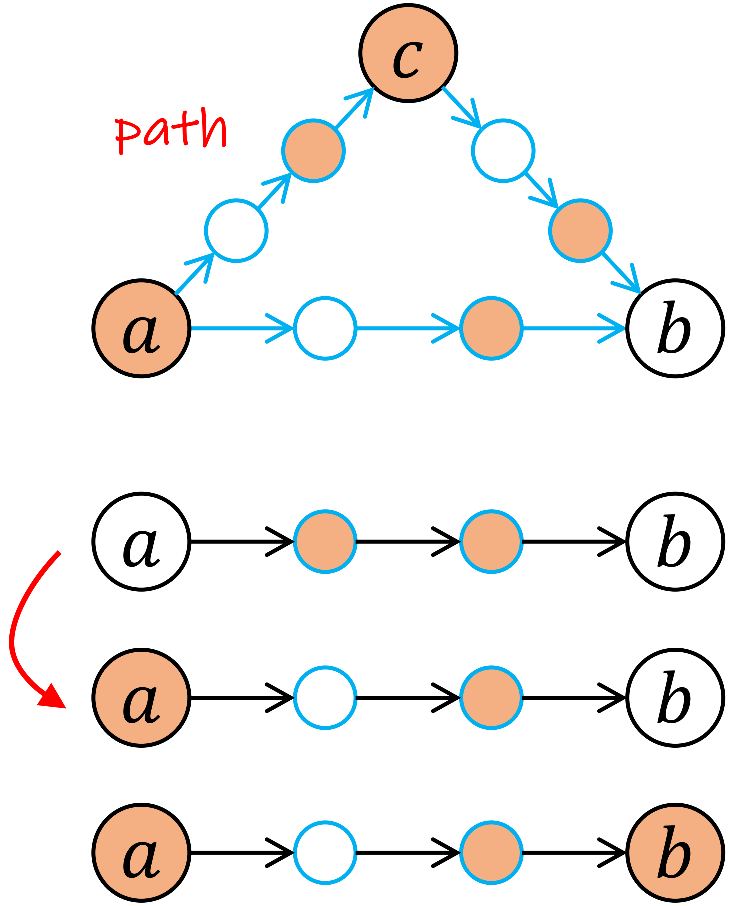




OR-property of an edge: we need to choose one (or both) of the end points

A template of hardness reduction from minimal VC



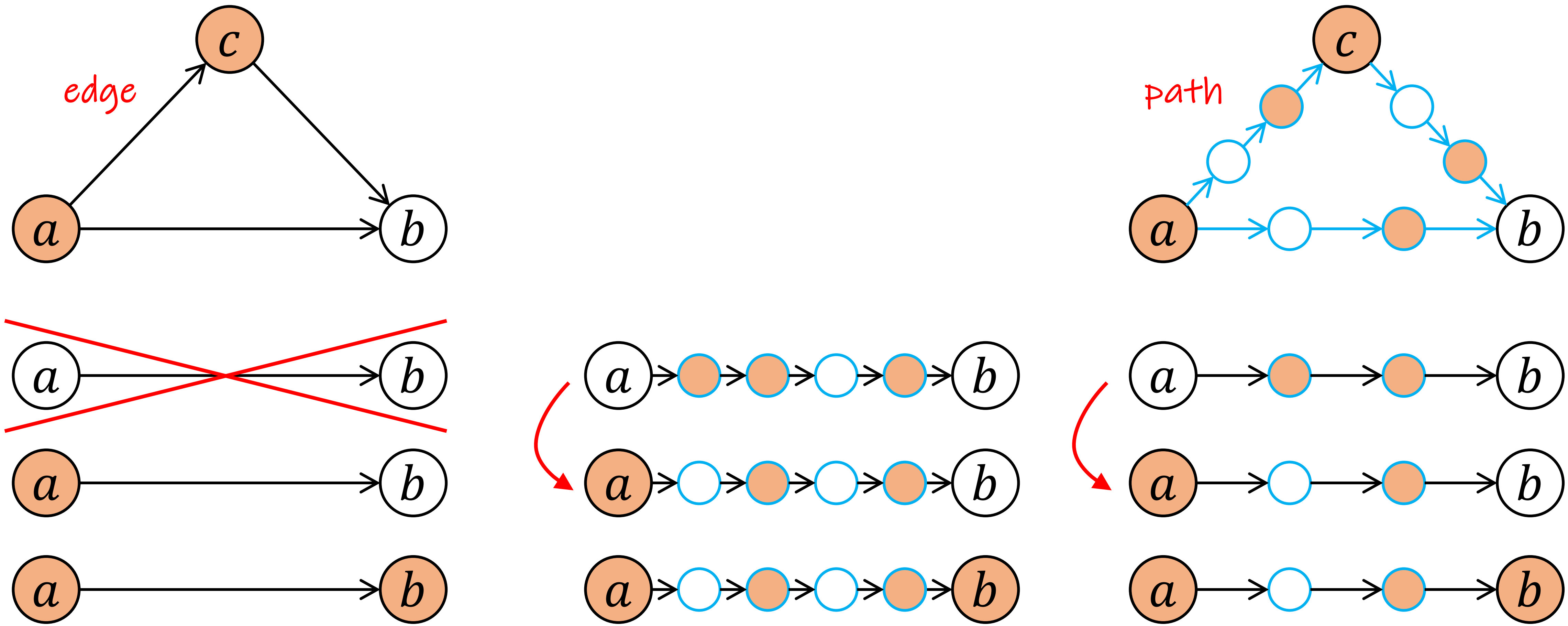


It is not immediately obvious why this is the "right" way to think about "edge gadgets", but we found it is ☺

OR-property of an edge: we need to choose one (or both) of the end points

OR-property of a path gadget: choosing one (or both) of the end points of a path reduces the Minimum VC of the remaining path gadget by 1 (we need to pick ≥ 2 vertices per path anyway)

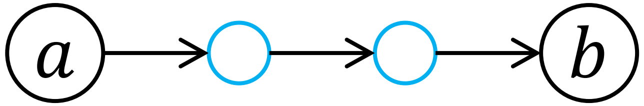
A template of hardness reduction from minimal VC

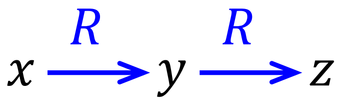


OR-property of an edge: we need to choose one (or both) of the end points

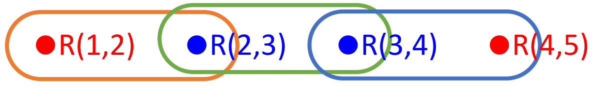
OR-property of a path gadget: choosing one (or both) of the end points of a path reduces the Minimum VC of the remaining path gadget by 1 (we need to pick ≥ 2 vertices per path anyway)

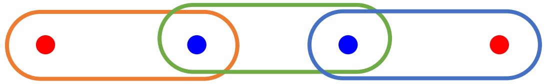
A template of hardness reduction from minimal VC



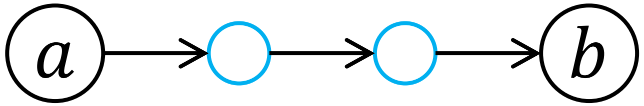


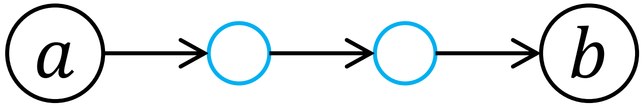
q_{2chain}:−$R(x,y),R(y,z)$

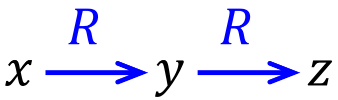


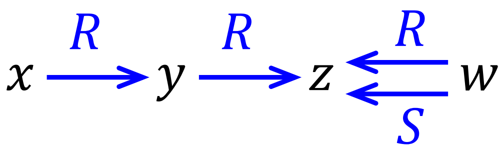


A template of hardness reduction from minimal VC



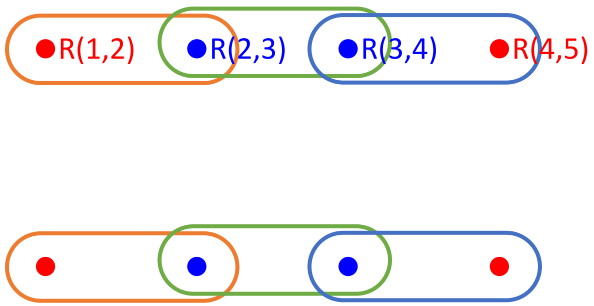


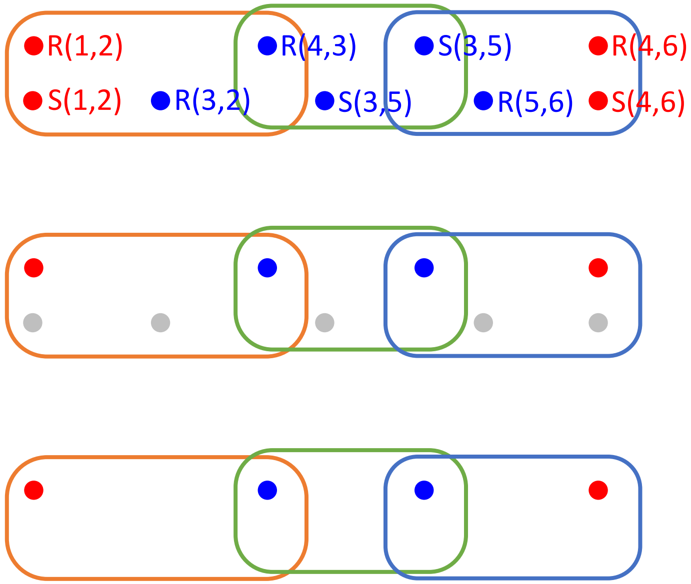




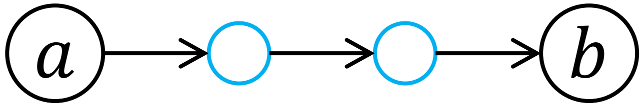
q_{2chain}:−$R(x,y)$,$R(y,z)$

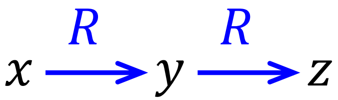
q_{3cc}^{S}:−$R(x,y)$,$R(y,z)$,$R(w,z)$,$S(w,z)$



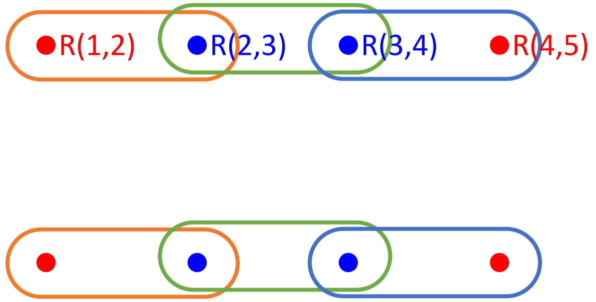


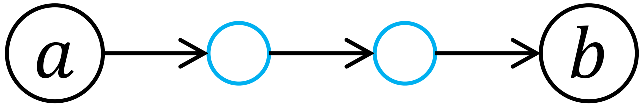
A template of hardness reduction from minimal VC

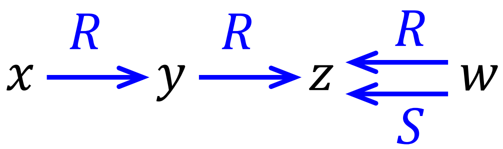




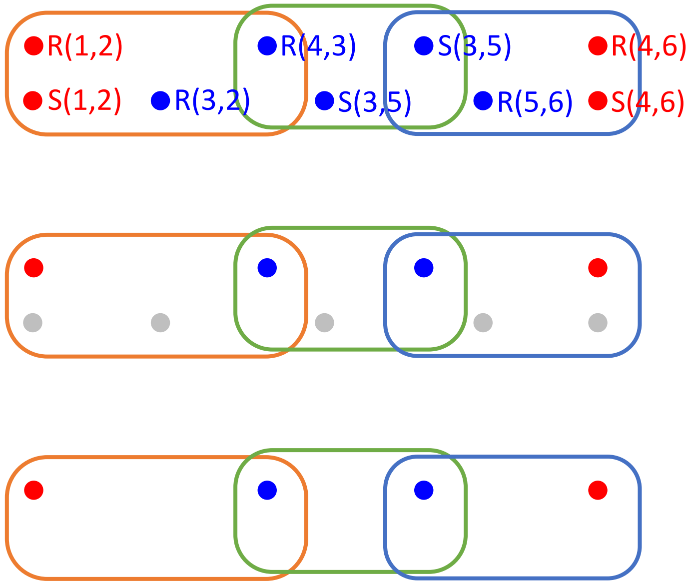
q_{2chain}:$-R(x,y),R(y,z)$

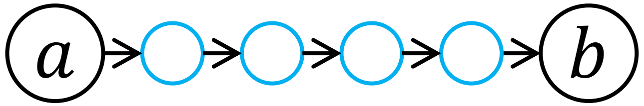


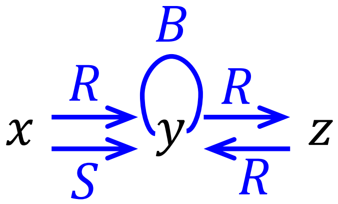




q_{3cc}^{S}:$-R(x,y),R(y,z),R(w,z),S(w,z)$



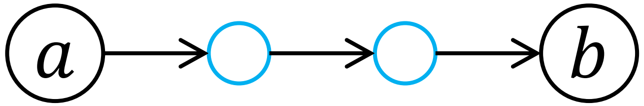




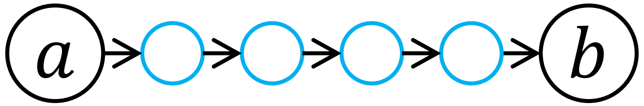
$q_{3perm-R}^{SxyB}$:$-\ S(x,y),R(x,y),B(y),R(y,z),R(z,y)$

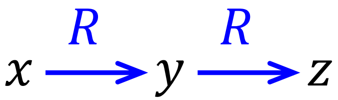
?

A template of hardness reduction from minimal VC

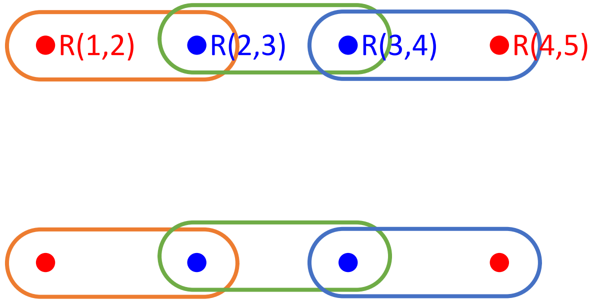


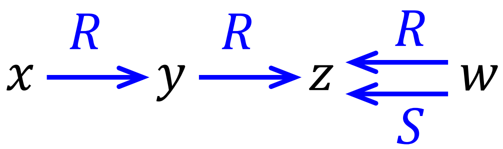
computer-generated "IJP"
with disjunctive logic program
(answer set programming ASP)



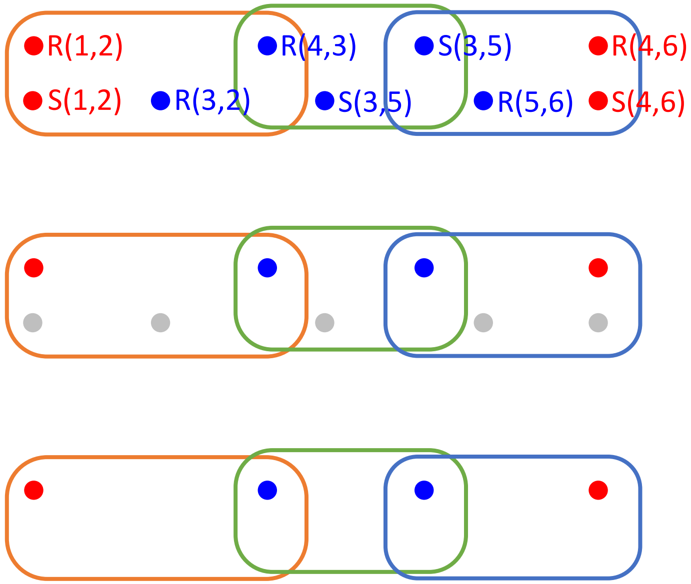


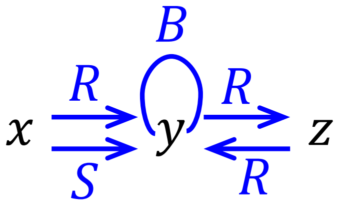
q_{2chain}:–$R(x,y), R(y,z)$



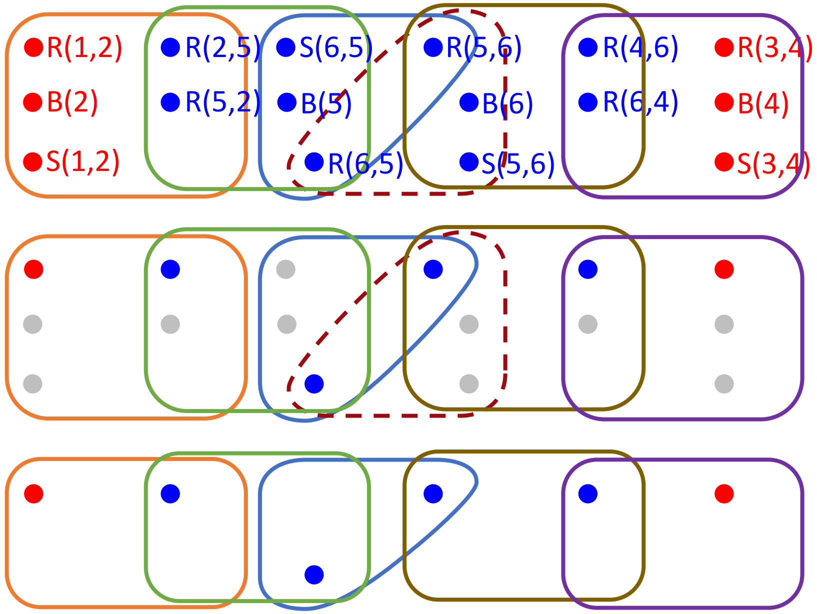


q_{3cc}^{S}:–$R(x,y), R(y,z), R(w,z), S(w,z)$





$q_{3perm-R}^{SxyB}$:– $S(x,y), R(x,y), B(y), R(y,z), R(z,y)$



Source: Makhija, Gatterbauer: A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations. SIGMOD'24. https://doi.org/10.1145/3626715

Formal definitions of IJPs

1. Join paths with end points (possible set of tuples)

DETAILS

Definition 7.1 (Join Path (JP)). A database D (under set or bag semantics) forms a Join Path from a set of tuples $\mathcal{S}$ (start) to a set of tuples $\mathcal{T}$ (terminal), for query Q if

(1) Each tuple in D participates in some witness (i.e. D is reduced).
(2) The witness hypergraph is connected.
(3) $\mathcal{S}$ and $\mathcal{T}$ form a *valid endpoint pair*, i.e.:
  (i) $\mathcal{S}$ and $\mathcal{T}$ are isomorphic and non-identical.
  (ii) There is no endogenous tuple $t \in D$, $t \notin \mathcal{S} \cup \mathcal{T}$ whose constants are a subset of the constants of tuples in $\mathcal{S} \cup \mathcal{T}$.

We also call two join paths *isomorphic* if there is a bijective mapping between the shared constants across the witnesses. Given a fixed query, we usually leave away the implied qualifier "isomorphic" when discussing join paths. We talk about the "*composition*" of two join paths if one endpoint of the first is identical to an endpoint of the second, and all other constants are different. We call a composition of join paths "non-leaking" if the composition adds no additional witnesses that were not already present in any of the non-composed join paths.

3. Composability defined semantically: composition cannot create additional witnesses = does not "leak"

Definition 7.3 (Independent Join Path). A Join Path D forms an *Independent Join Path (IJP)* if it fulfills two additional conditions:

(4) "OR-property": Let c be the resilience of Q on D. Then resilience is $c-1$ in all 3 cases of removing either $\mathcal{S}$ or $\mathcal{T}$ or both.
(5) Any composition of two or more isomorphic JPs is non-leaking.

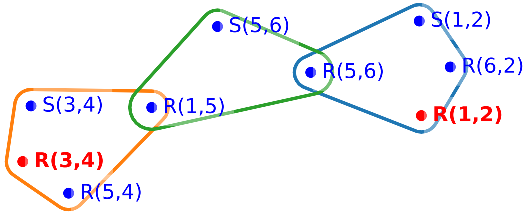
2. OR-property

Source: Makhija, Gatterbauer: A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations. SIGMOD'24. https://doi.org/10.1145/3626715

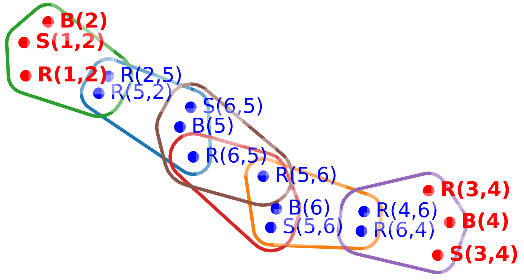
Finding IJPs with ASP: 5 New Hardness Gadgets

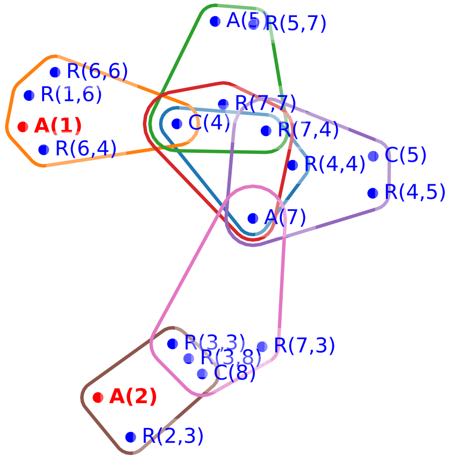
- Using the **Automatic IJP Generator**, able to prove 5 queries hard (out of 7 previously open from [Freire+ PODS'20])

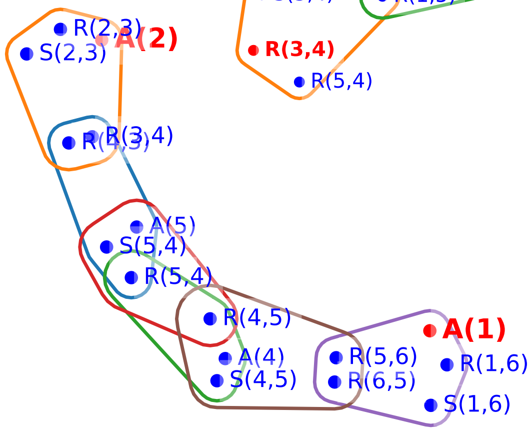
q^{S}_{3cc}:-R(x,y),R(y,z),R(w,z),S(w,z)



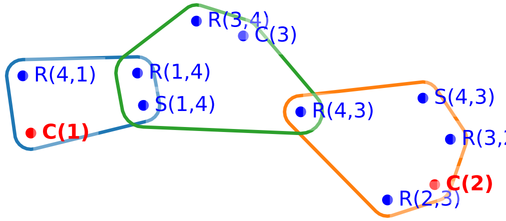
$q^{S_{xy}B}_{3perm-R}$:- S(x,y),R(x,y),B(y),R(y,z),R(z,y)







$q^{AS_{xy}}_{3perm-R}$:- A(x),S(x,y),R(x,y),R(y,z),R(z,y)



$q^{S_{xy}C}_{3perm-R}$:- S(x,y),R(x,y),R(y,z),R(z,y),C(z)

z_6:− A(x),R(x,y),R(y,y),R(y,z),C(z)

- Can recover all previous hardness results + find new ones!

Dichotomy Conjectures for Resilience

Theorem.

IJP → **NPC**

Conjecture. [Hardness]

IJP ↔ **NPC**

Conjecture. [Hardness]

IJP → IJP of domain size $\leq 7 * var(Q)$

Conjecture. [PTIME]

∄IJP → LP = ILP

Theorem.

IJP ↔ **NPC** for SJ-Free Queries

Conjecture. [Hardness] Corollary

NPC → DLP finds a hardness proof

Conjecture. [PTIME]

∄IJP → There is a flow graph that encodes resilience

Slide by Neha Makhija. https://northeastern-datalab.github.io/unified-reverse-data-management/download/Unified-Reverse-Data-Management-Simons-Institute-Nov-2023.pdf

Outlook

- Flow becomes complicated. Goal: can we automate as well?
- Resilience Binary CQs: is the ASP really "all we need"?
- Resilience Union of Binary CQs: the current ASP abstraction is not enough (but can likely be fixed)
- Resilience Non-binary CQs: gets more hairy
- Minimum size factorization of provenance
  - constraint matrices contain {-1,0,1}; minimal vertex cover analogy breaks
  - flow encodings still possible but become factored in nontrivial ways
  - leverages a connection between factorizations, query plans, and VEOs

Thank you ☺

BACKUP

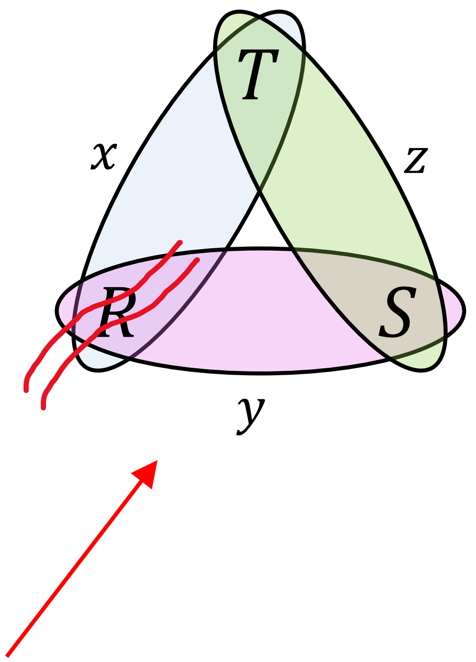
IJPs generalize triads

"IJPs" generalize "active triads"

Triangle query

$Q^{\Delta}: -R(x,y), S(y,z), T(x,z)$

Triad in the Triangle query



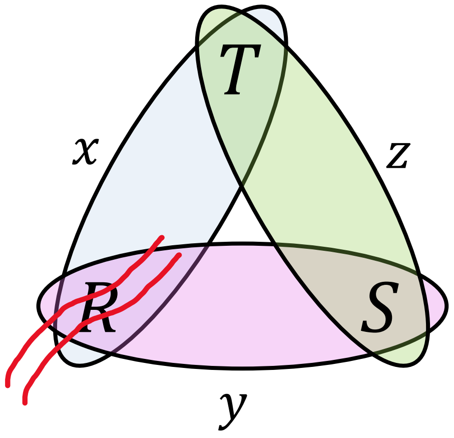
This this the query dual hypergraph.
Each node corresponds to one tuple, each hyperedge to one variable in the query

"IJPs" generalize "active triads"

Triangle query

$Q^{\Delta}: -R(x,y), S(y,z), T(x,z)$

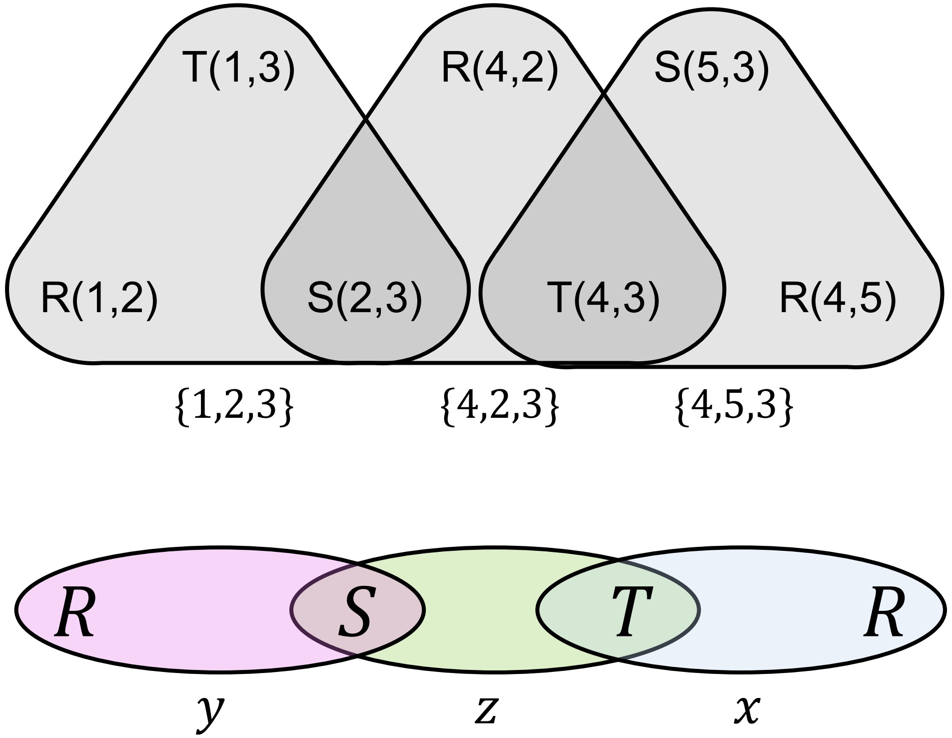
Triad in the Triangle query



This this the query dual hypergraph. Each node corresponds to one tuple, each hyperedge to one variable in the query

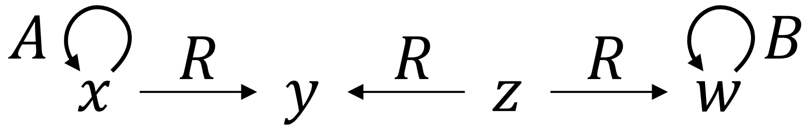
This this the data instance hypergraph. Each node corresponds to one tuple, each hyperedge to one witness

Implied IJP



IJPs unify past proofs for SJs

Evidence for IJPs as unifying criterion



Original hardness proof for via reduction from Max 2SAT

$Q_{3\text{conf}}^{AC}$: $-A(x), R(x,y), R(z,y), R(z,w), C(w)$

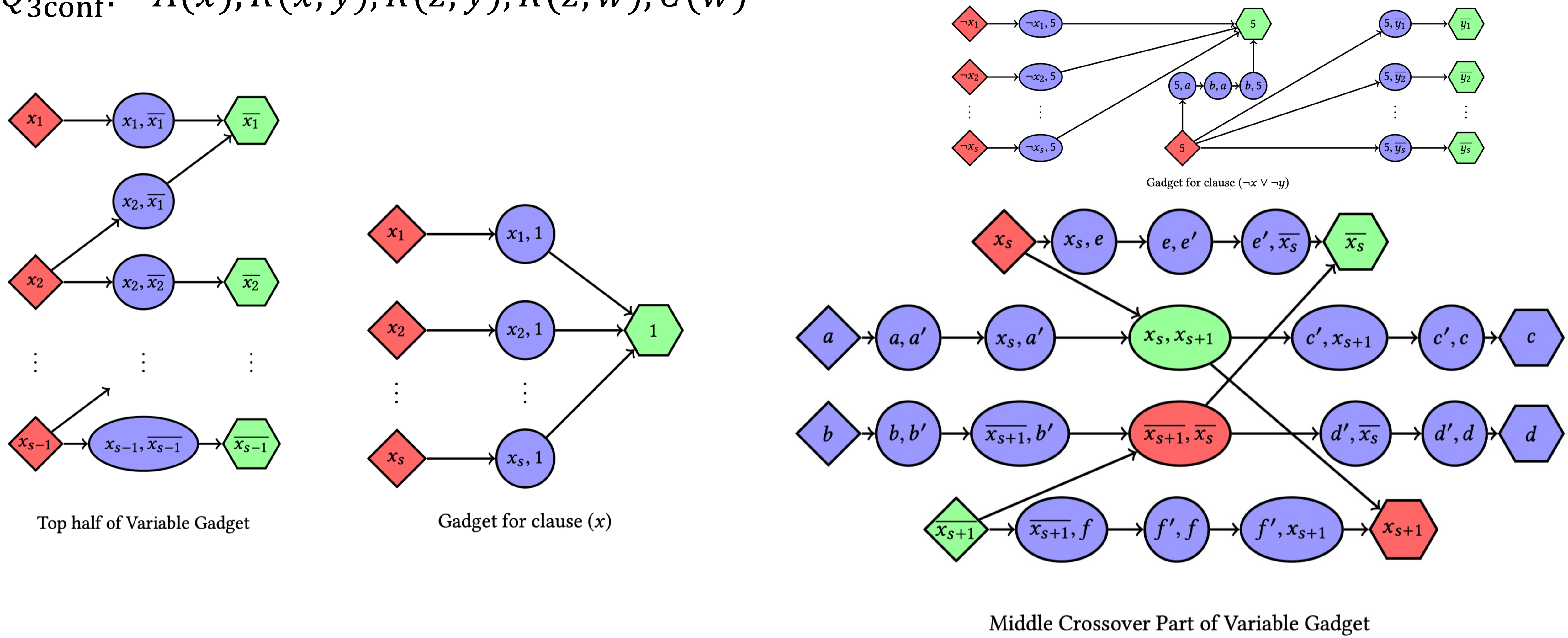
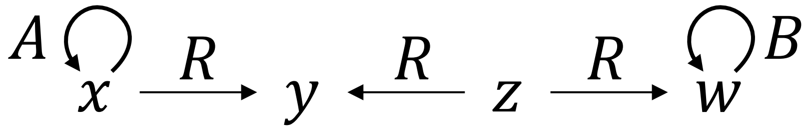


Figure 15: Reduction Gadgets for proof of Proposition 39: diamonds represent A, ellipses, R, and hexagons, C. In the variable gadgets, the minimum contingency sets choose all red vertices and no green, or all green vertices and no red.

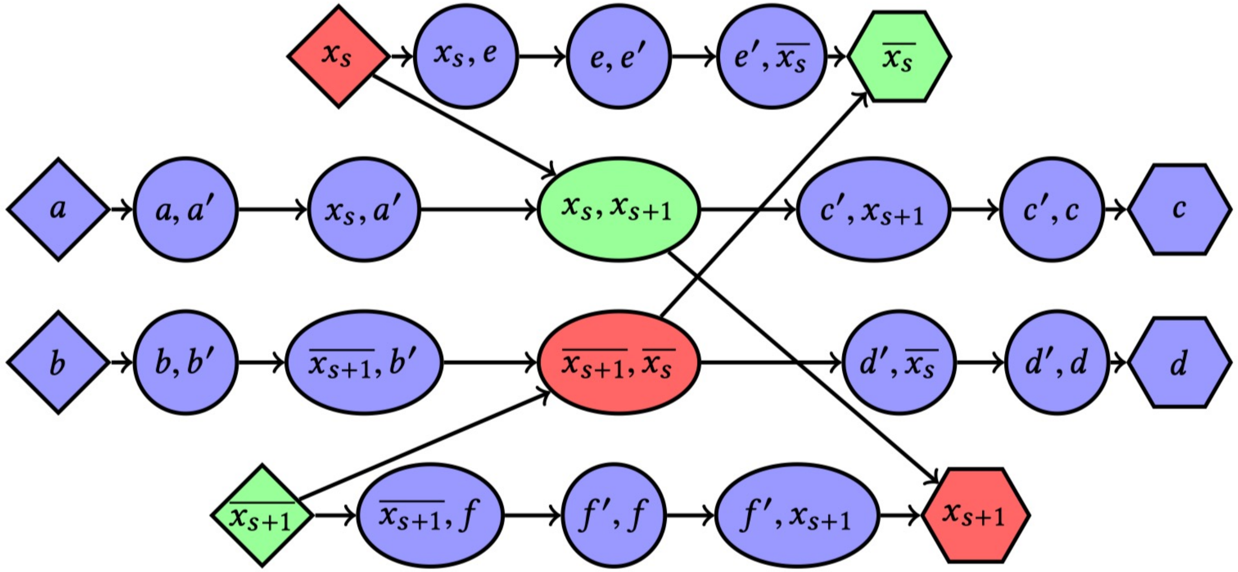
Source: Freire+. New Results for the Complexity of Resilience for Binary Conjunctive Queries with Self-Joins. PODS 2020. https://arxiv.org/pdf/1907.01129

Evidence for IJPs as unifying criterion



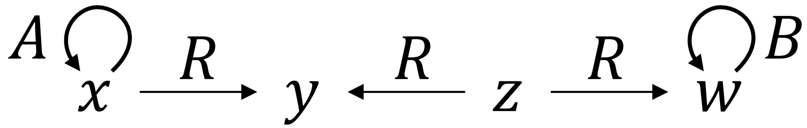
Original hardness proof for via reduction from Max 2SAT

$Q^{AC}_{3\text{conf}}$: $-A(x), R(x,y), R(z,y), R(z,w), C(w)$



Middle Crossover Part of Variable Gadget

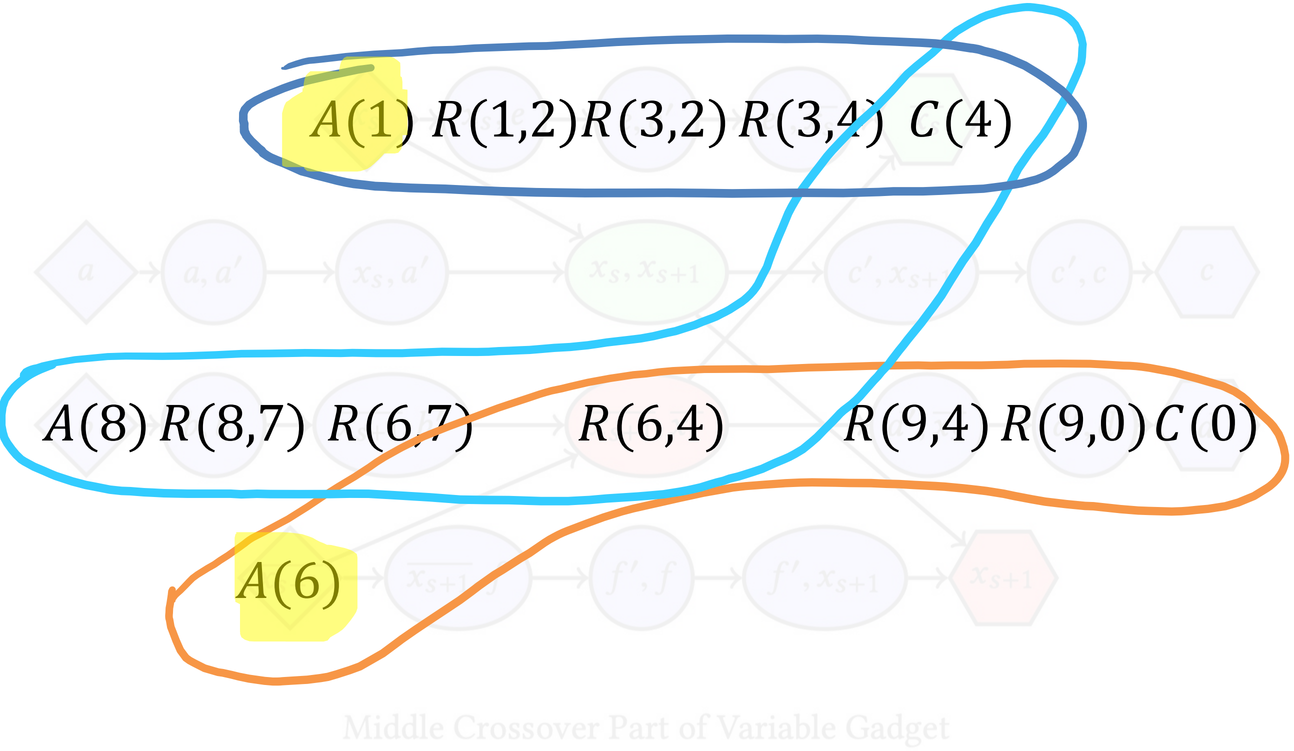
Evidence for IJPs as unifying criterion

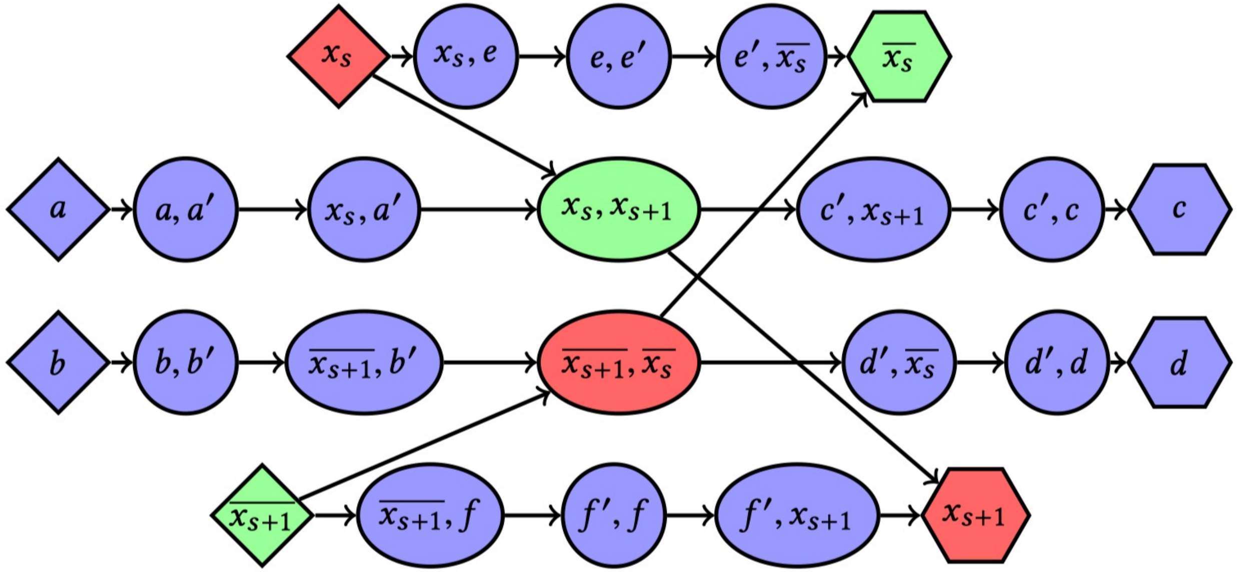


Original hardness proof for via reduction from Max 2SAT

Q^{AC}_{3conf}: $-A(x), R(x,y), R(z,y), R(z,w), C(w)$

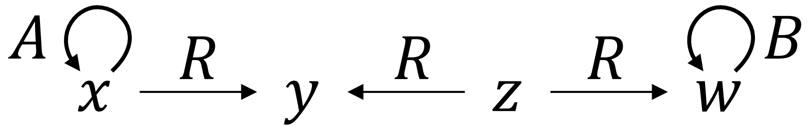
IJP with pivots A(1), A(6) was already hiding in the original gadgets
[not shown are "extensions" one still needs to add to the end points]





Middle Crossover Part of Variable Gadget

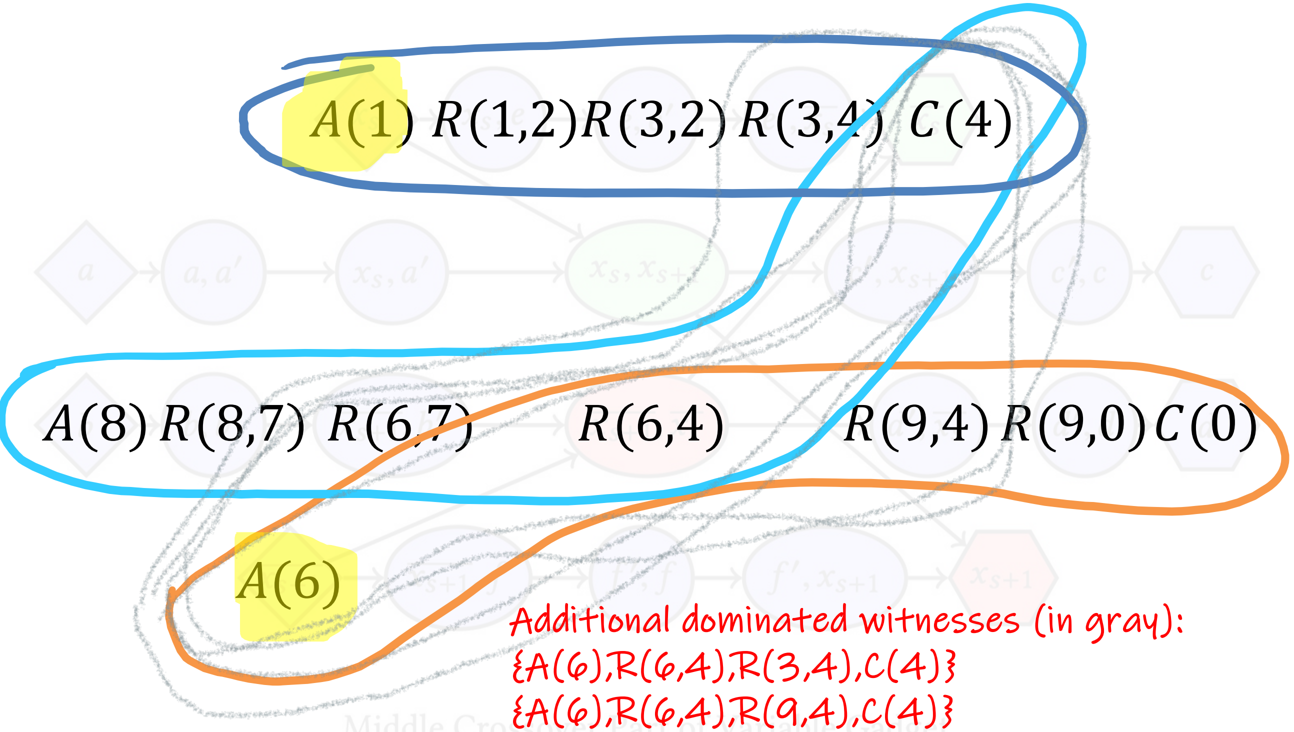
Evidence for IJPs as unifying criterion



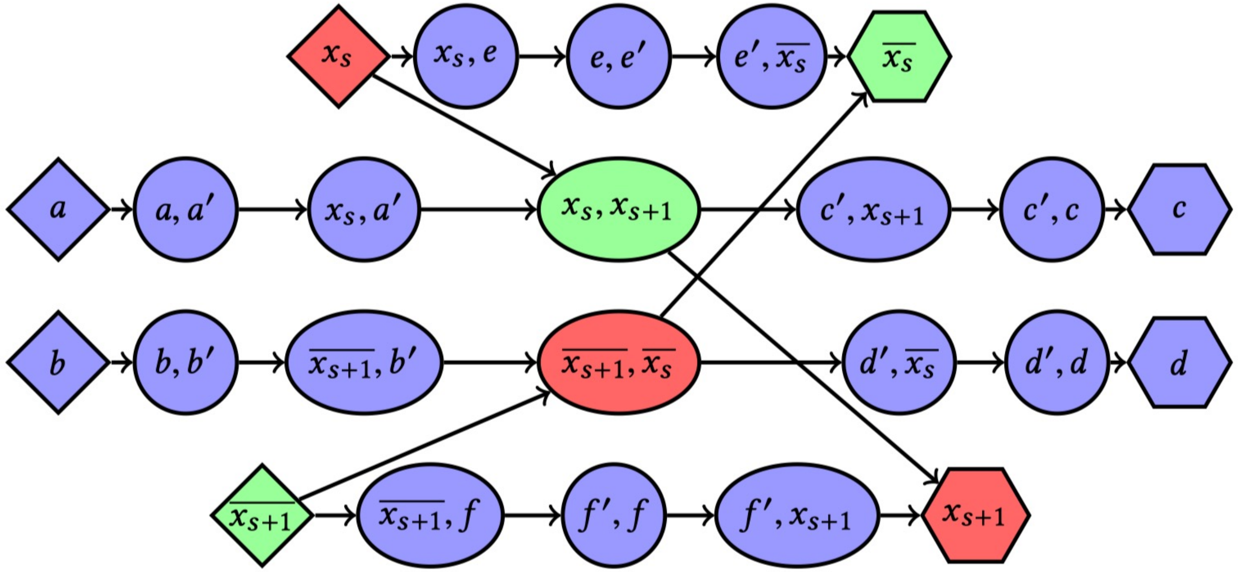
Original hardness proof for via reduction from Max 2SAT

$Q^{AC}_{3\text{conf}}$: $-A(x), R(x,y), R(z,y), R(z,w), C(w)$

IJP with pivots A(1), A(6) was already hiding in the original gadgets
[not shown are "extensions" one still needs to add to the end points]



Additional dominated witnesses (in gray):
{A(6),R(6,4),R(3,4),C(4)}
{A(6),R(6,4),R(9,4),C(4)}
{A(6),R(6,7),R(6,4),C(4)}
{A(6),R(6,7),R(6,4),R(3,4),C(4)}
{A(6),R(6,7),R(6,4),R(9,4),C(4)}



Middle Crossover Part of Variable Gadget