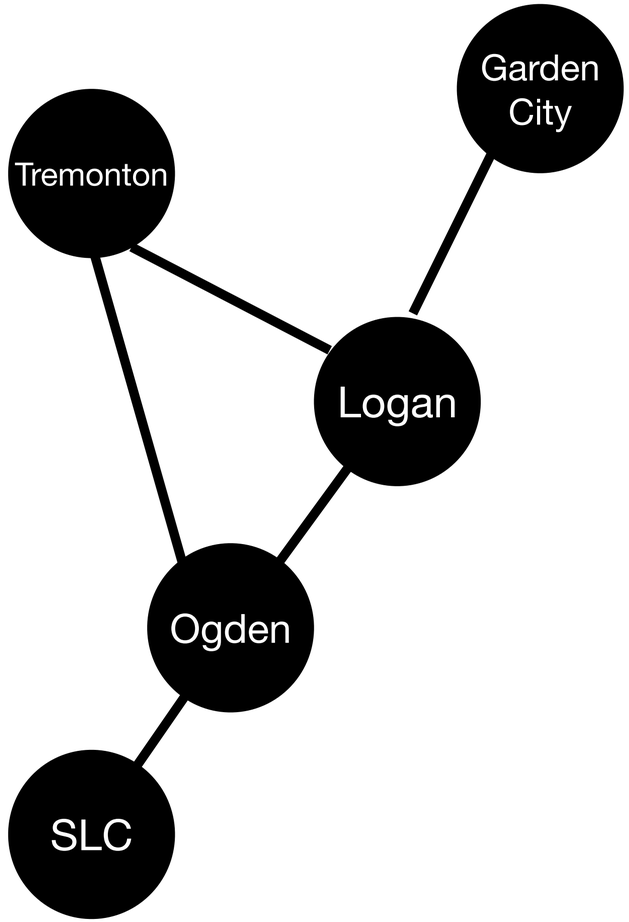
Relational Algebra

Yihao Sun Utah State University Sept. 2026

Two-hop to Bear Lake

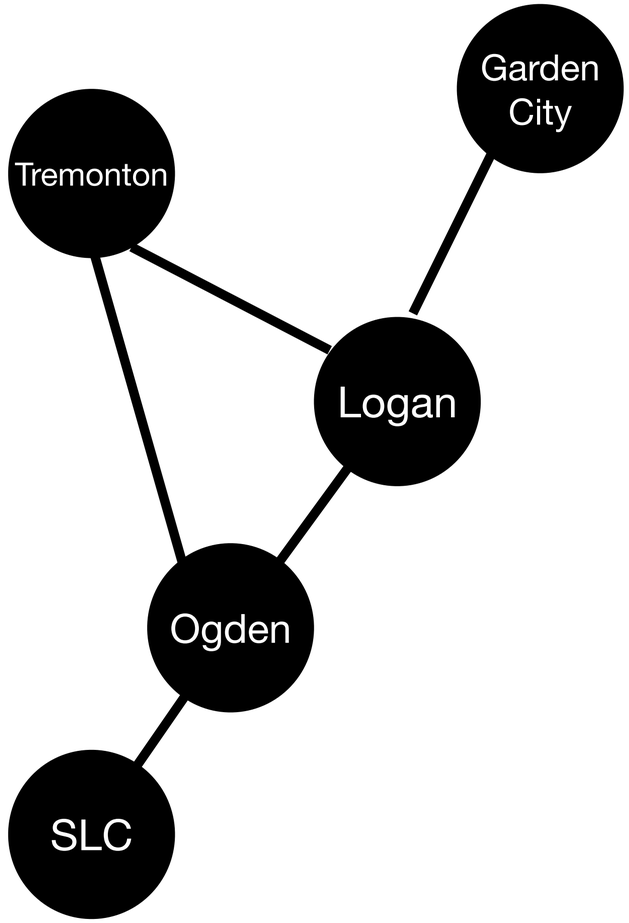


$$
\begin{aligned}
&q(a, b) :- \\
&\quad \text{edge}(a, mid1), \\
&\quad \text{edge}(mid1, b), \\
&\quad \text{lakefront}(b) .
\end{aligned}
$$

$I(\text{lakefront}) = \{(\text{Garden City},)\}$

Each link can be used in either direction.

The Answer Is a Set



$$E = I(\text{edge}) \subseteq D^2$$

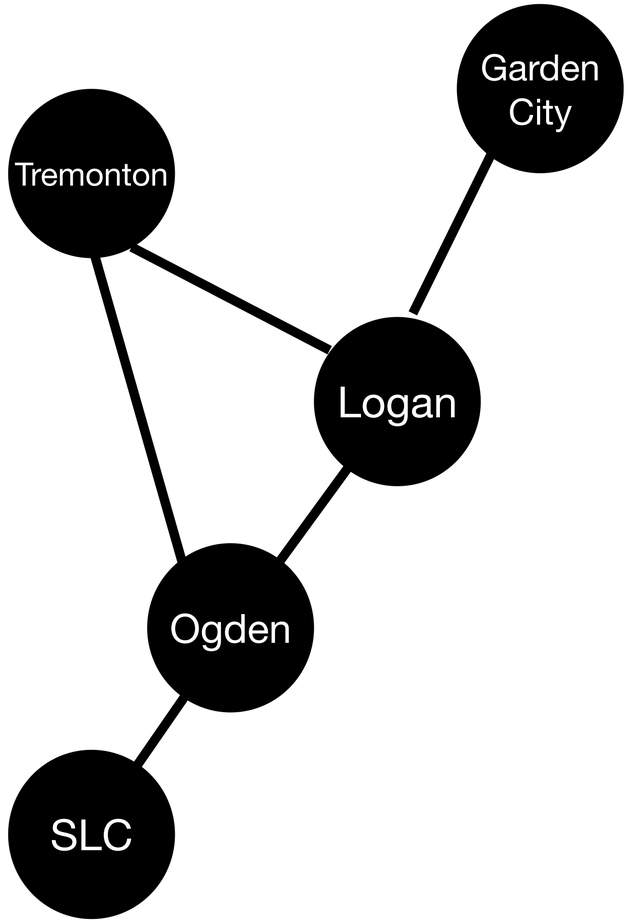
$$B = I(\text{lakefront}) = \{\text{Garden City}\}$$

$$H_2 = \{(u, w) \in D^2 \mid \exists z \in D : \ (u, z) \in E \wedge (z, w) \in E \wedge w \in B\}$$

$$(\text{Tremonton}, \text{Garden City}) \in H_2$$

The intermediate city is a witness.

The Same Answer, by Valuation



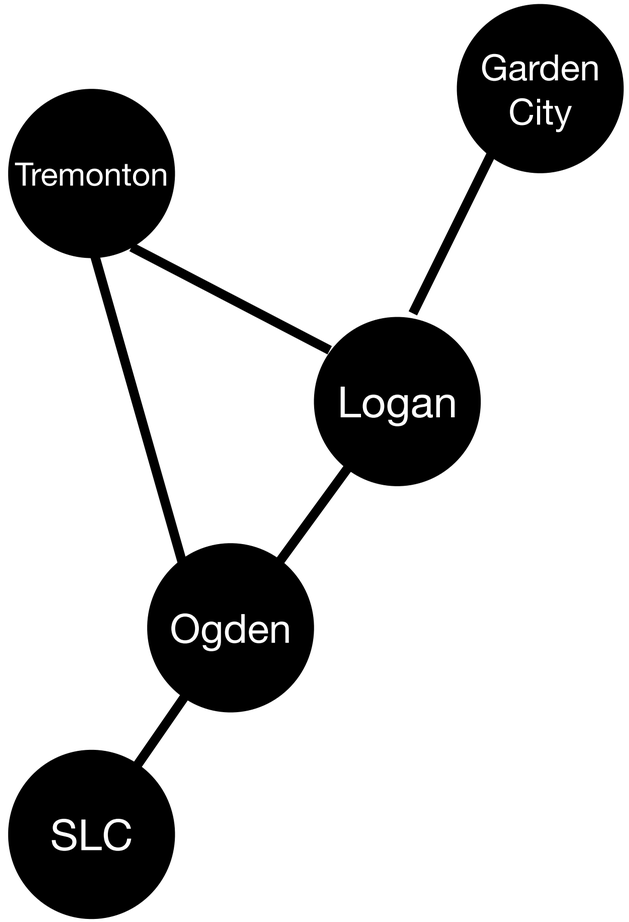
$$C = \text{edge}(a, mid1) \\ \quad \wedge \text{edge}(mid1, b) \\ \quad \wedge \text{lakefront}(b)$$

$$H_2 = \{(v(a), v(b)) \mid \\ \quad v : \{a, mid1, b\} \to D, \\ \quad I \models C[v]\}$$

The definition chooses no assignment order.

One Variable at a Time

Choose a value for a

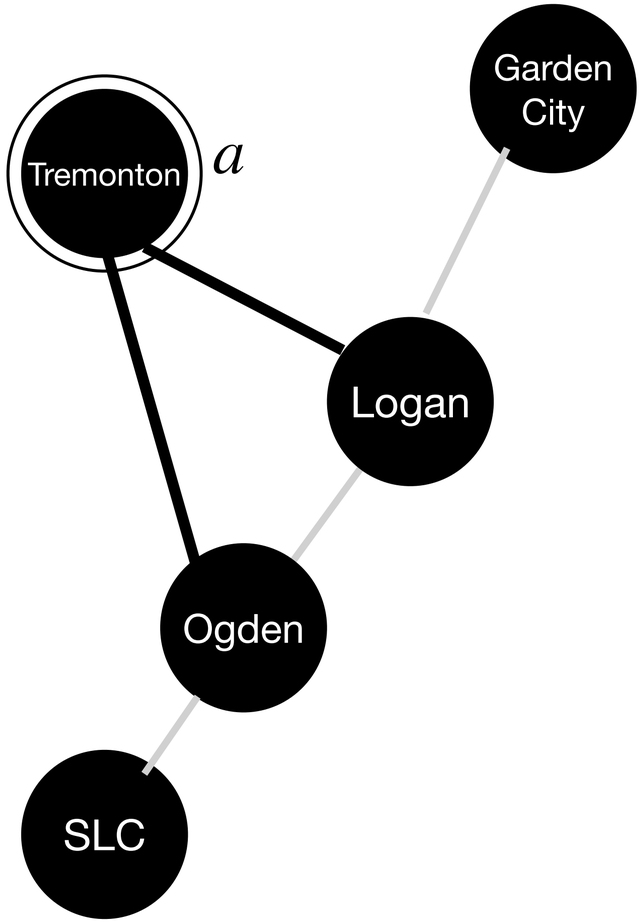


$$V(a) = \{\text{Logan, Garden City, Tremonton, Ogden, SLC}\}$$

One branch: $a = \text{Tremonton}$

Valuate $mid1$

Keep a = Tremonton



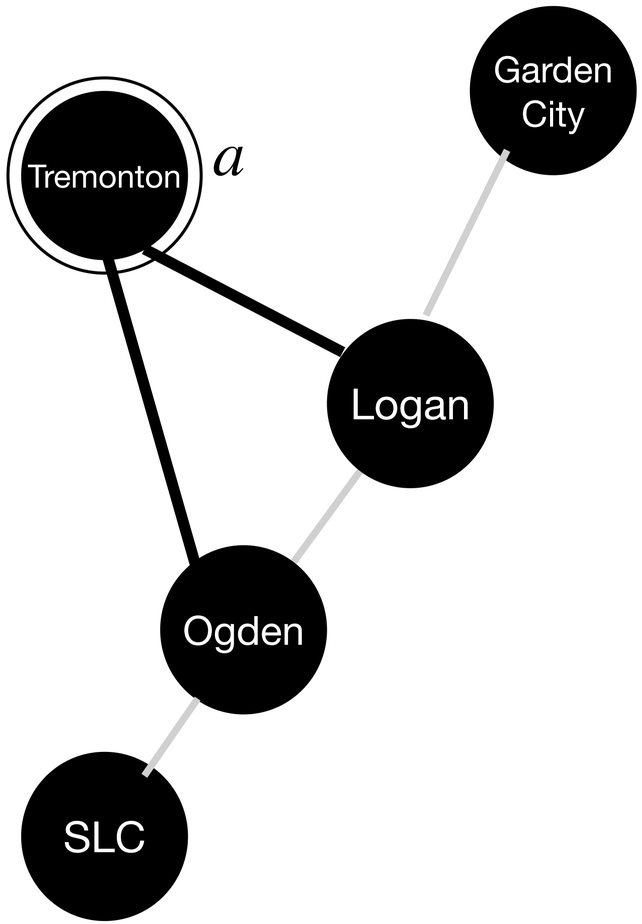
$v = \{a \mapsto \text{Tremonton}\}$

$V(mid1) = \{\text{Logan, Garden City,}$
$\text{Tremonton, Ogden, SLC}\}$

Check $\text{edge}(a, mid1)$

Valuate $mid1$

Keep a = Tremonton



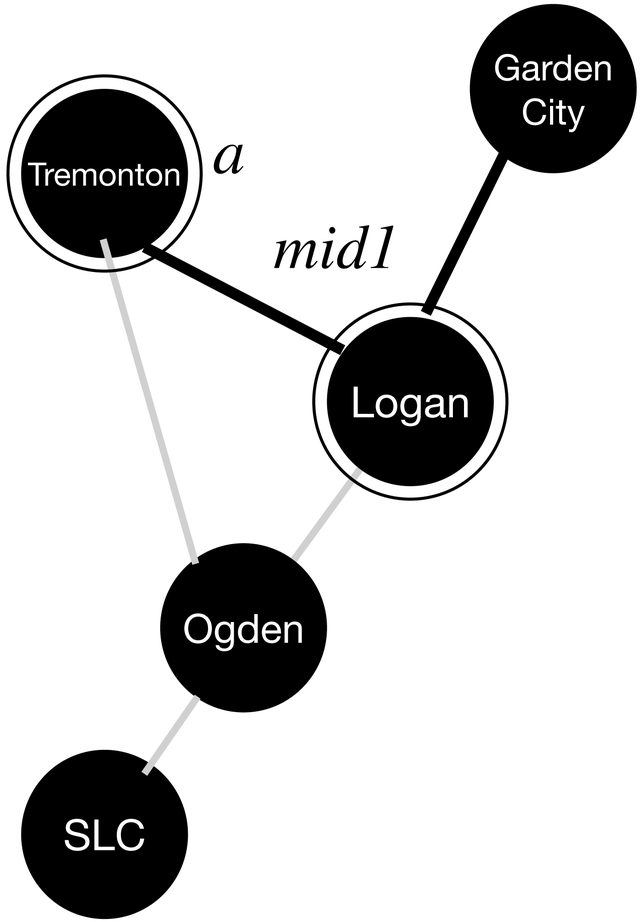
$v = \{a \mapsto \text{Tremonton}\}$

$V(mid1) = \{\text{Logan}, \text{Ogden}\}$

$\text{edge}(\text{Tremonton}, mid1)$

Valuate b

Choose $mid1 =$ Logan

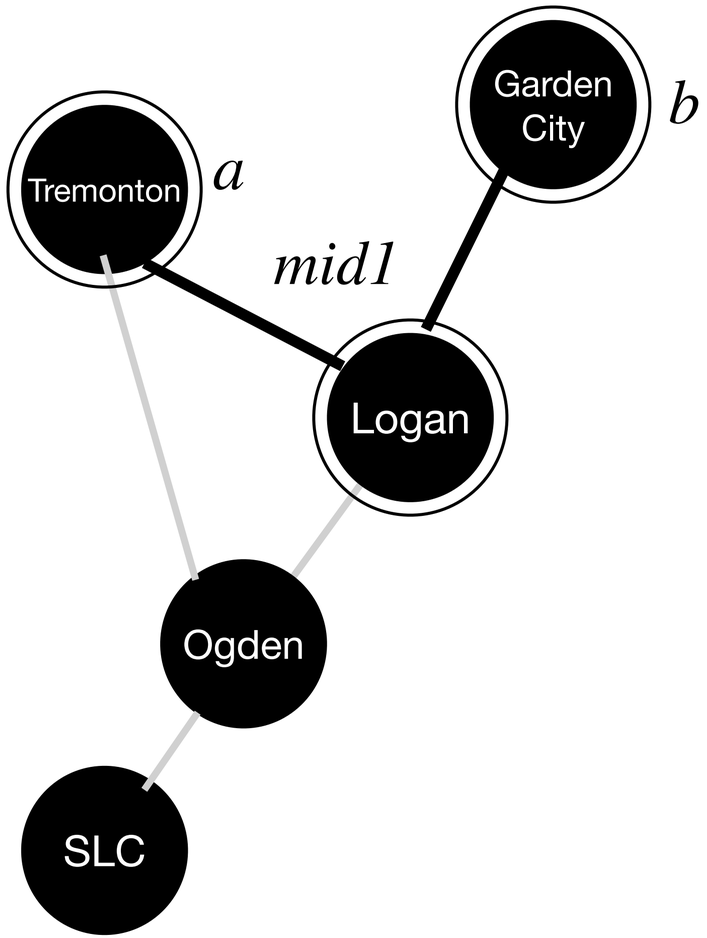


$$v = \{a \mapsto \text{Tremonton},\ mid1 \mapsto \text{Logan}\}$$

$$V(b) = \{\text{Garden City}, \text{Tremonton}, \text{Ogden}\} \cap B$$
$$= \{\text{Garden City}\}$$

Choose $b =$ Garden City

A Satisfying Valuation

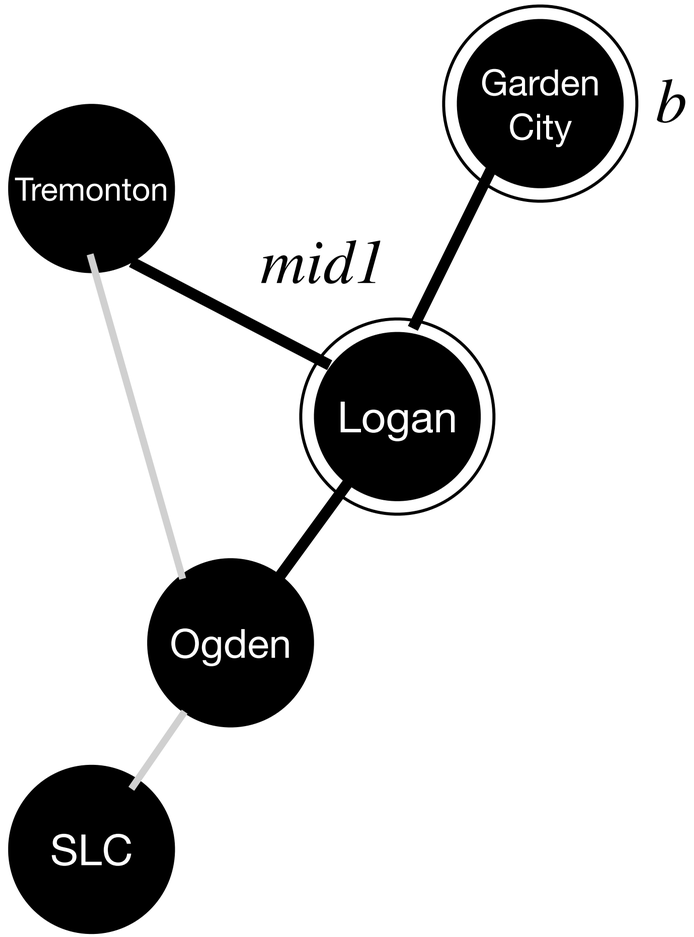


$$v = \{a \mapsto \text{Tremonton},\ mid1 \mapsto \text{Logan},\ b \mapsto \text{Garden City}\}$$

$$\frac{I \vDash \text{edge}(a, mid1)[v] \quad I \vDash \text{edge}(mid1, b)[v] \quad I \vDash \text{lakefront}(b)[v]}{I \vDash C[v]}$$

Destination First

Choose b from lakefront first



$V(b) = B = \{\text{Garden City}\}$

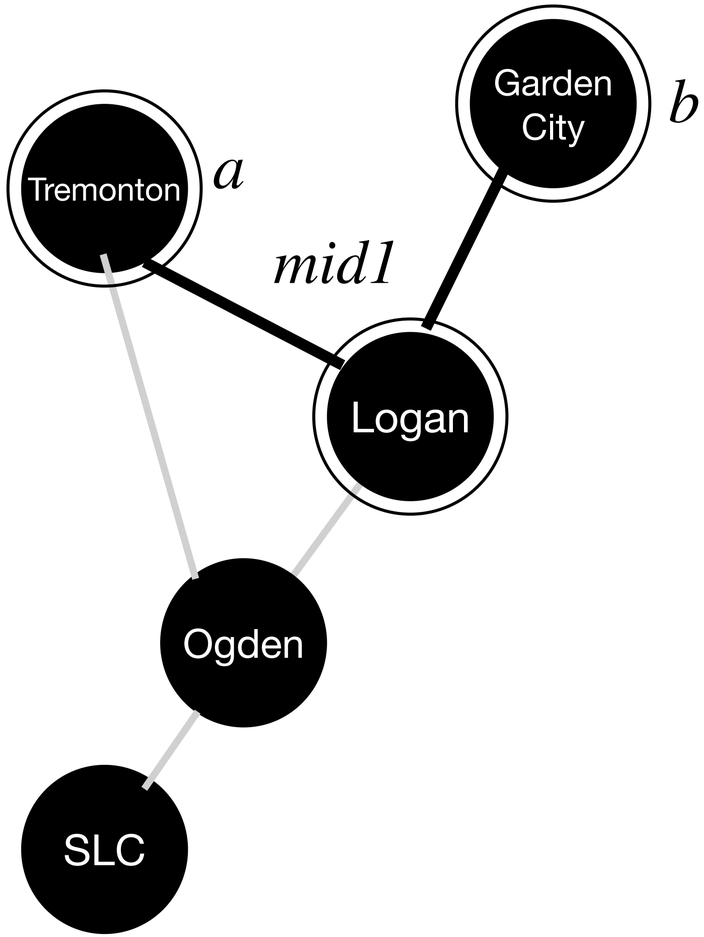
$V(mid1) = \{\text{Logan}\}$

$\text{edge}(\text{Logan}, \text{Garden City})$

$V(a) = \{\text{Garden City}, \text{Tremonton}, \text{Ogden}\}$

Set Intersection

Keep a = Tremonton and b = Garden City



$$V(mid1) =$$

$$\{\text{Logan, Ogden}\}$$

$$\cap$$

$$\{\text{Logan}\}$$

$$= \{\text{Logan}\}$$

Same Answers, Different Work

Surviving partial valuations for the whole query

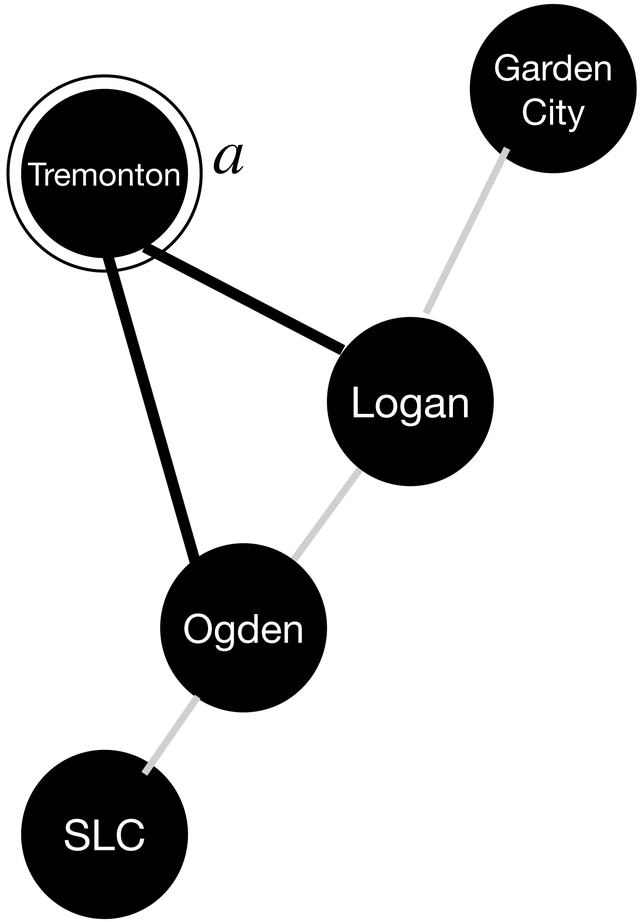
	After one variable	After two	Complete
$a \to mid1 \to b$	5	10	3
$b \to mid1 \to a$	1	1	3

The same answers. A different amount of intermediate work.

Check each constraint as soon as its variables are bound.

Valuation Relations

First atom with a = Tremonton



$\text{edge}(a, mid1)$

R

a	$mid1$
Tremonton	Logan
Tremonton	Ogden

Constraints on a Valuation

$$X = \text{vars}(C_1), \quad Y = \text{vars}(C_2), \quad t : X \cup Y \to D$$

$$t \in \text{Val}_I(C_1 \wedge C_2)$$

$$\iff I \vDash C_1[t] \wedge I \vDash C_2[t]$$

$$\iff t\big|_X \in E_1 \wedge t\big|_Y \in E_2$$

One valuation must pass both local tests.

Natural Join

A over X, $\quad B$ over Y

$$A \bowtie B = \{t : X \cup Y \to D \mid t\big|_X \in A \wedge t\big|_Y \in B\}$$

Closure: the result is another relation.

$$\mathrm{Val}_I(C_1 \wedge C_2) = E_1 \bowtie E_2$$

Compatible local assignments unite into one output valuation.

Join

Combine rows that agree on *mid1*

R

a	*mid1*
Tremonton	Logan
Tremonton	Ogden

Q

mid1	*b*
Logan	Garden City

$$R \bowtie Q$$

Join

Combine rows that agree on $mid1$

R

a	$mid1$
Tremonton	Logan
Tremonton	Ogden

Q

$mid1$	b
Logan	Garden City

$R \bowtie Q$

a	$mid1$	b
Tremonton	Logan	Garden City

From Valuations to Answers

$$E_1 = \mathrm{Val}_I(C_1), \qquad E_2 = \mathrm{Val}_I(C_2)$$

$$H_2 = \{(t(a), t(b)) \mid t \in E_1 \bowtie E_2\}$$

$$\pi_{\{a,b\}}(E_1 \bowtie E_2) = \{t|_{\{a,b\}} \mid t \in E_1 \bowtie E_2\}$$

(Tremonton, Logan, Garden City) $\longmapsto$ (Tremonton, Garden City)

Keep the head variables. Equal head tuples collapse.

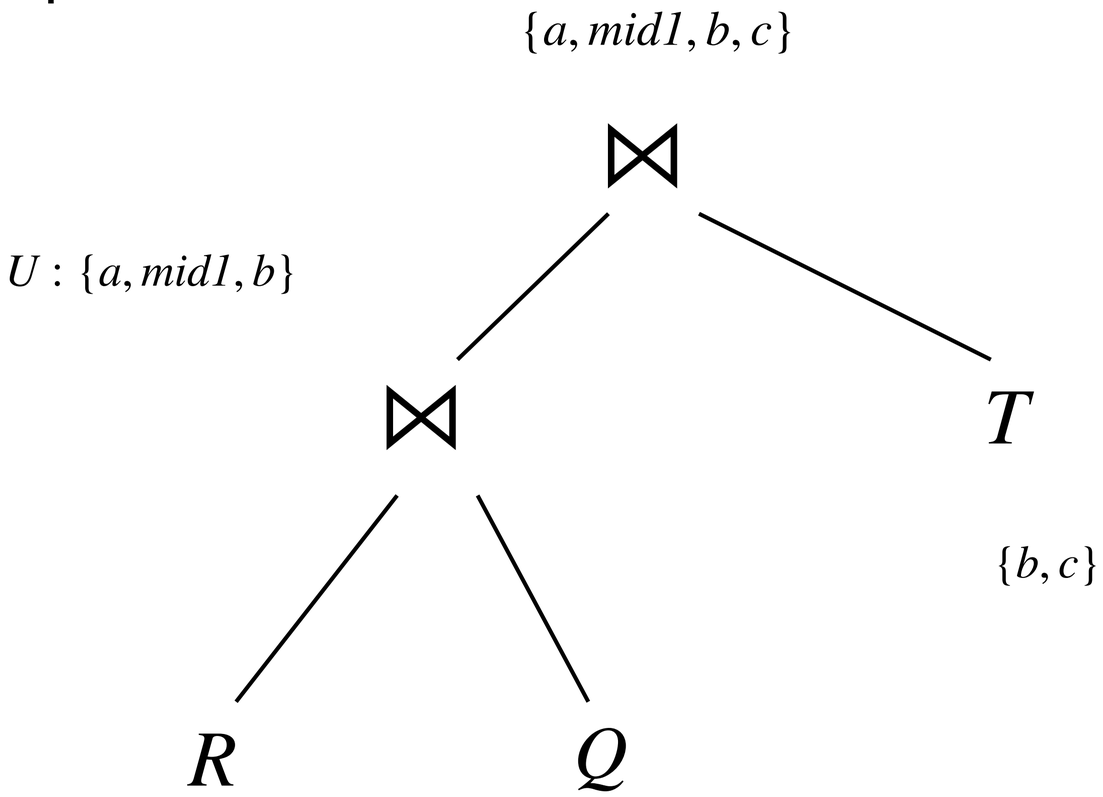
Composition and Associativity

The output relation becomes the next input

$$U = R \bowtie Q$$

$$\begin{aligned} U \bowtie T &= (R \bowtie Q) \bowtie T \\ &= R \bowtie (Q \bowtie T) \end{aligned}$$

T contains the matches for $\text{edge}(b, c)$.



Semijoin

Keep W rows whose b occurs in B

$W : a =$ Tremonton

a	$mid1$	b
Tremonton	Logan	Garden City
Tremonton	Logan	Tremonton
Tremonton	Logan	Ogden
Tremonton	Ogden	Logan
Tremonton	Ogden	Tremonton
Tremonton	Ogden	SLC

B : Bear Lake-front cities

b
Garden City

Here,

$W \ltimes B = W \bowtie B$

B adds no new column.

$W \ltimes B$

a	$mid1$	b
Tremonton	Logan	Garden City

Join and Semijoin

The right relation now has a $lake$ column

Use the same W.

F

b	$lake$
Garden City	Bear Lake

$W \bowtie F$

a	$mid1$	b	$lake$
Tremonton	Logan	Garden City	Bear Lake

$W \ltimes F$

a	$mid1$	b
Tremonton	Logan	Garden City

Keeps W's columns

Both keep one matching valuation. The join also carries $lake$.

Cartesian Join

Ordinary three-hop: a = Logan, c = Ogden . No lakefront condition.

R

a	$mid1$
Logan	Garden City
Logan	Tremonton
Logan	Ogden

T

b	c
Logan	Ogden
Tremonton	Ogden
SLC	Ogden

No shared variable

Cartesian Join

$R \times T$ $3 \times 3 = 9$ candidate valuations

a	$mid1$	b	c
Logan	Garden City	Logan	Ogden
Logan	Garden City	Tremonton	Ogden
Logan	Garden City	SLC	Ogden
Logan	Tremonton	Logan	Ogden
Logan	Tremonton	Tremonton	Ogden
Logan	Tremonton	SLC	Ogden
Logan	Ogden	Logan	Ogden
Logan	Ogden	Tremonton	Ogden
Logan	Ogden	SLC	Ogden

The Middle Constraint

Keep valuations that also satisfy $\text{edge}(mid1, b)$

a	$mid1$	b	c
Logan	Garden City	Logan	Ogden
Logan	Tremonton	Logan	Ogden
Logan	Ogden	Logan	Ogden
Logan	Ogden	Tremonton	Ogden
Logan	Ogden	SLC	Ogden

5 valuations support the answer $(\text{Logan}, \text{Ogden})$.

Expression Syntax

A grammar fragment

$E ::= r$ — Stored relation

$\mid \sigma_{A=d}(E)$ — One relation input

$\mid (E_L \bowtie E_R)$

$\mid (E_L \ltimes E_R)$ — Two relation inputs

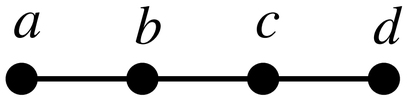
$\sigma_{src=\text{Tremonton}^{(\text{edge})}}$

Closure permits nested expressions.

CQ Zoo: Trees

Vertices are variables. Edges are binary atoms.

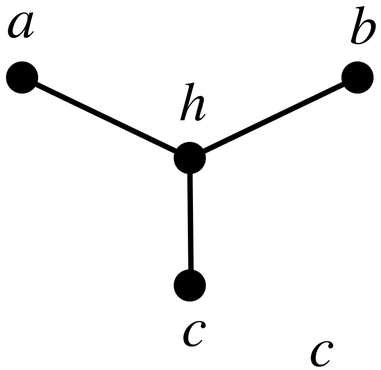
Path / Chain



$$q(a,d) :- R(a,b), S(b,c), T(c,d).$$

Route lookup
Connected legs

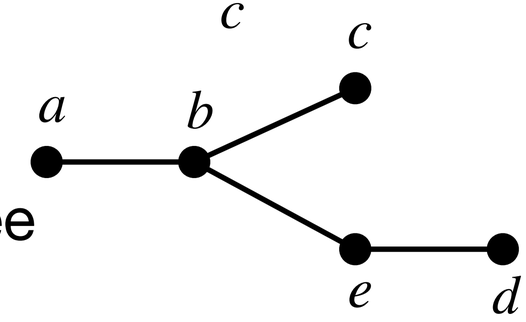
Star



$$q(a,b,c) :- R(h,a), S(h,b), T(h,c).$$

Entity profiles
Name, email, job

Branched Tree



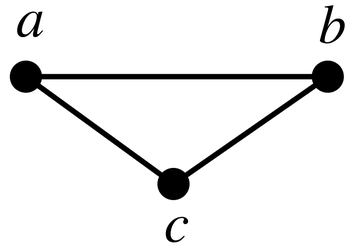
$$q(a,c,d) :- R(a,b), S(b,c), T(b,e), U(e,d).$$

Order reporting
Buyer, date, item

A two-edge path is also called a wedge.

CQ Zoo: Cycles and Dense Patterns

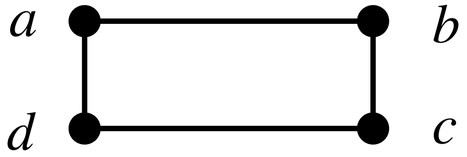
Triangle



$$q(a,b,c) :- R(a,b), S(b,c), T(c,a)\,.$$

Social networks
Mutual friends

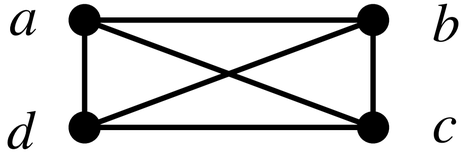
Cycle



$$q(a,b,c,d) :- R(a,b), S(b,c), T(c,d), U(d,a)\,.$$

Shopping patterns
Shared purchases

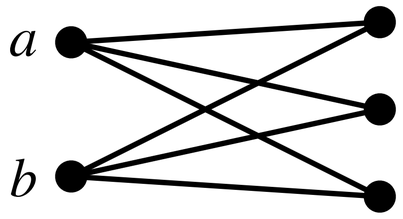
Clique



$$q(a,b,c,d) :- R_{ab}(a,b), R_{ac}(a,c), R_{ad}(a,d), R_{bc}(b,c), R_{bd}(b,d), R_{cd}(c,d)\,.$$

Collaboration
Pairwise coauthors

Biclique



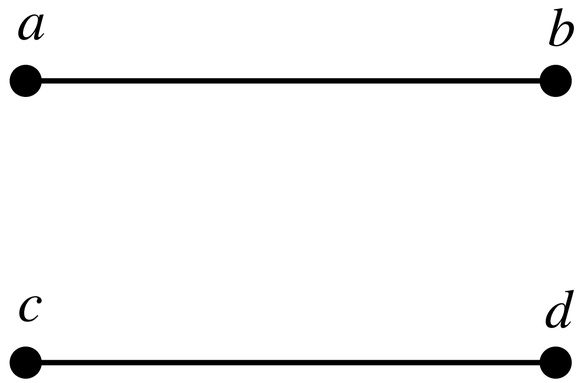
$$q(a,b,x,y,z) :- R_{ax}(a,x), R_{ay}(a,y), R_{az}(a,z), R_{bx}(b,x), R_{by}(b,y), R_{bz}(b,z)\,.$$

Interest groups
Shared products

The clique and biclique shown here have six binary atoms.

CQ Zoo: Atoms and Components

Disconnected

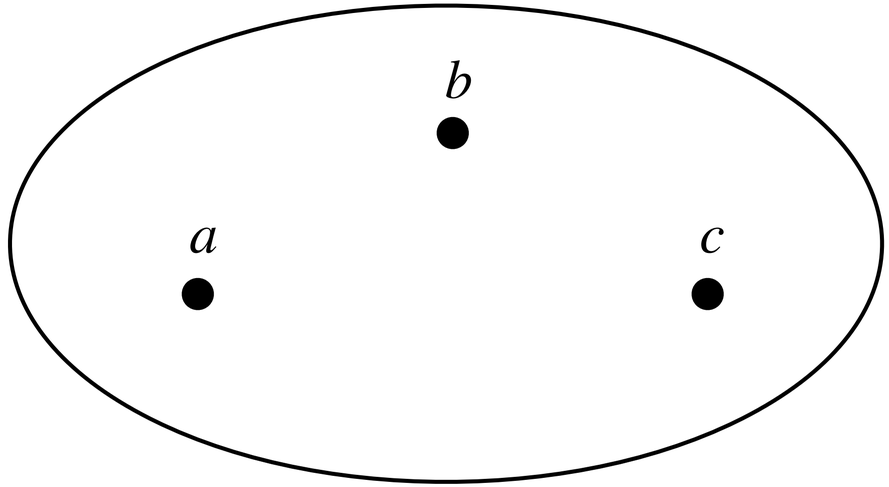


Travel: flight and hotel options

$$q(a, b, c, d) :- R(a, b), S(c, d) .$$

Independent assignments combine by product.

One Ternary Atom



Inventory: depot, item, day

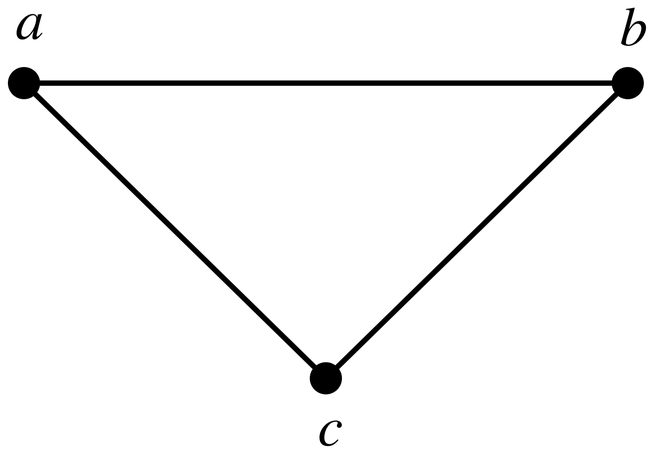
$$q(a, b, c) :- T(a, b, c) .$$

One group means one tuple test.

CQ Zoo: Loomis–Whitney

Each atom omits one variable.

Three variables



Inventory views: depot, item, day

$$q_3(a,b,c) :- R(a,b),\ S(a,c),\ T(b,c)\,.$$

Four variables

- = variable occurs in atom

	a	b	c	d
R	•	•	•	
S	•	•		•
T	•		•	•
U		•	•	•

Inventory views: add supplier

$$q_4(a,b,c,d) :- R(a,b,c),\ S(a,b,d),\ T(a,c,d),\ U(b,c,d)\,.$$