SQL / relational algebra

SELECT	PROJECT	JOINS	Union	Intersections	Set differences
σ	π	$\bowtie$ $\times$	$\cup$	$\cap$	$\setminus$

$\sigma_P\ T$ $\qquad$ $\pi_{c_1 c_2}\ T$

$Q_1 \cap Q_2$

1. Finish evaluating Q_1. Say T_1
2. Finish evaluating Q_2. Say T_2
3. Calculate intersection.

Alternative approach
Use T_1 to even prune which parts of Q_2 you want to evaluate.

→ Prolog top/down evaluation

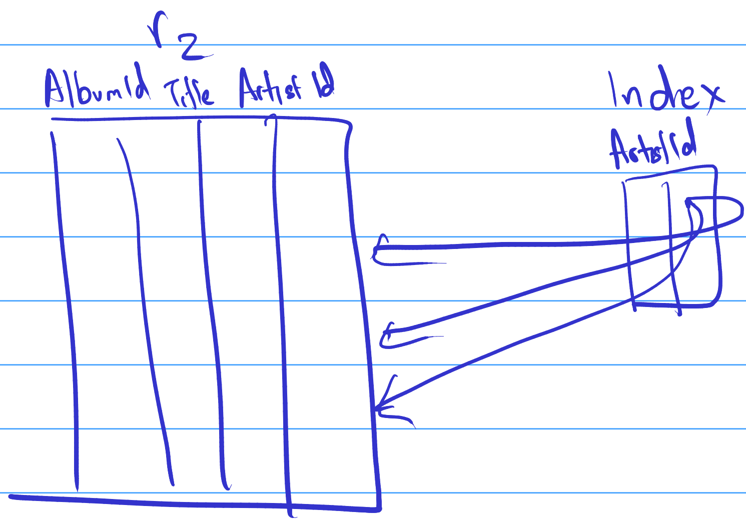
→ "Magic sets" algorithms

$Q_1 \bowtie Q_2$. $Q_1 \longrightarrow T_1$ $Q_2 \longrightarrow T_2$

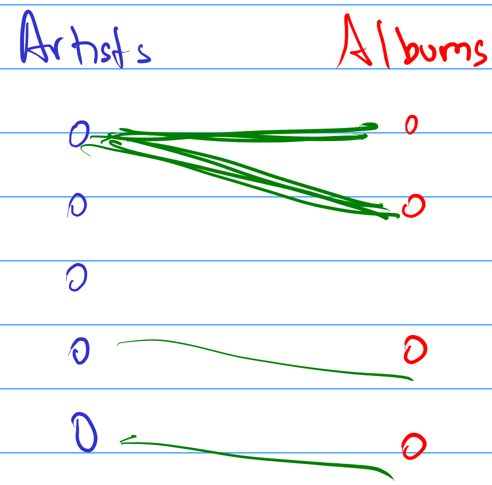
$\forall\, r_1 \in T_1$

$\forall$ matching $r_2 \in T_2$

emit $(r_1 \bowtie r_2)$



- Find all pairs of albums which were recorded by the same artist.

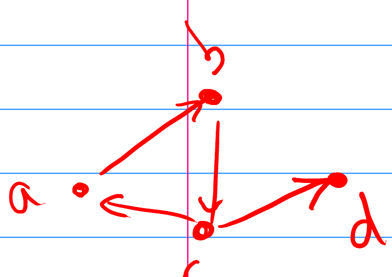


Q: Given graph G write down T_G

adjacency relation written down in table form

Q: Find out if G has a triangle.

a	b
b	c
c	a
c	d



Alg 1: For each vertex u

For each neighbor v s.t. $u \to v$

For each neighbor w of v

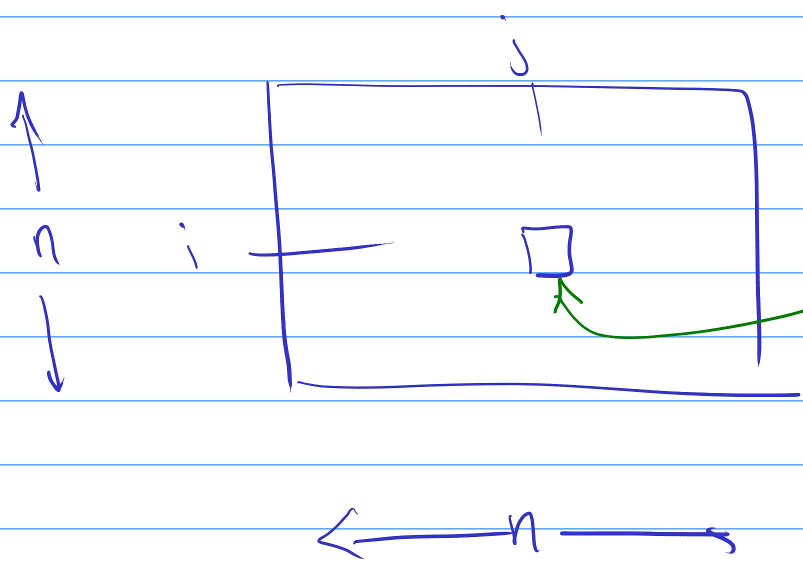
if $w \to u$, return true.

Return false.

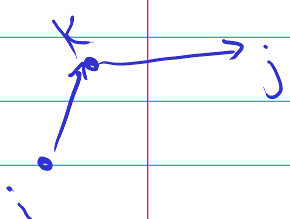
$O(n^3)$

number of vertices

Alg 2: Assume G provided as its adjacency matrix E.



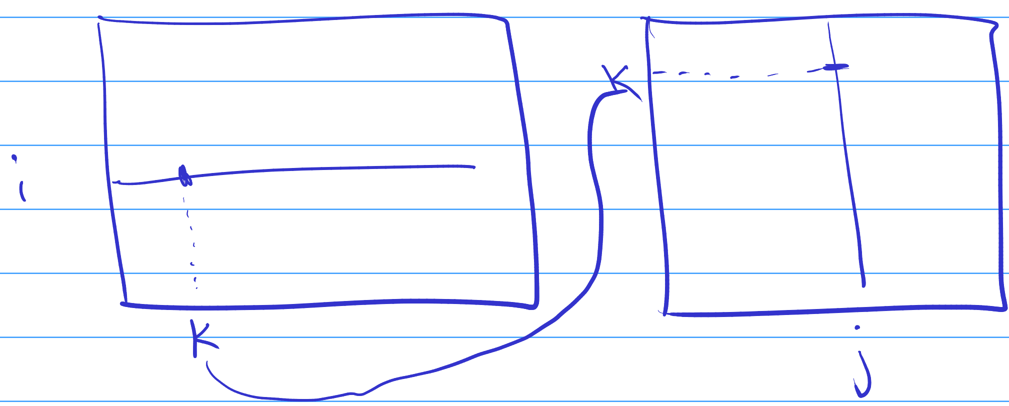
Is there an edge from $v_i \to v_j$.



$E \times E$

$F = E \times E$

$$f_{ij} = \sum_{k} e_{ik} \cdot e_{kj} = \bigvee_{k} e_{ik} \wedge e_{kj}$$



F = all pairs of vertices that are two apart

$H = E \times E \times E \qquad O(n^{\omega})$ matrix multiplication exponent.

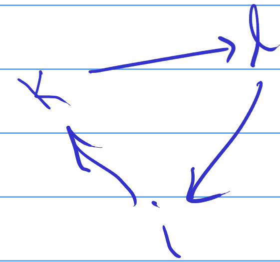
$2 \leq \omega \lesssim 2.$

$h_{ij} = \text{true}$ iff $\exists\, k \quad l$

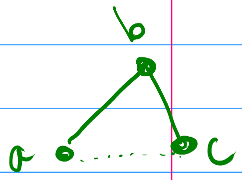
$i \rightarrow k \rightarrow l \rightarrow j$

$\Rightarrow h_{ii} = \text{true}$ iff $\exists\, k \quad l$

$i \rightarrow k \rightarrow l \rightarrow i$



In <u>other words</u>: Diagonal entry $h_{ii} = \text{true}$ iff i is the corner of a triangle.



"Proto" triangles

$$\underbrace{\overset{a\,b}{E} \bowtie \overset{b\,c}{E}}_{\uparrow\, m \quad \uparrow\, m} \bowtie \overset{c\,a}{E}$$

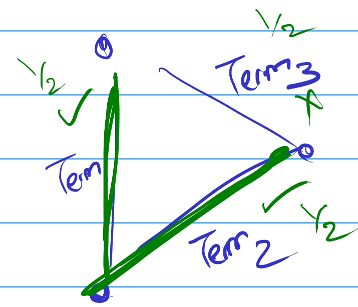
Prunes the proto triangles

Claim: $E \bowtie E \bowtie E$ has at most $O(m^2)$ entries

# of edges.

Question: How many triangles might a graph even have?

$$\underbrace{\underset{\text{Term 1}}{\overset{a\,b}{E}} \bowtie \underset{\text{Term}_2}{\overset{b\,c}{E}}} \bowtie \underbrace{\overset{c\,a}{E}}_{\text{Filter}}\ \text{Term}_3$$



$$(E)^1 \cdot (E^1) \cdot (E^0)$$

$$E^1 E^0 E^1 \qquad E^0 E^1 E^1 \qquad m^2$$

$$E^{\frac{1}{2}} \quad E^{1/2} \quad E^{\frac{1}{2}}$$

NPRR: There is an algorithm which can achieve this bound

AGM: There is a family of graphs with $O(m^{3/2})$ triangles.

This bound is tight.

Two practical optimizations

Selection pushdown

$$\sigma_p\left(\overbrace{\underbrace{Q_1}_{m} \bowtie \underbrace{Q_2}_{n}}^{mn}\right)$$

$$\Downarrow$$

When p only depends on columns in Q_1

$$(\sigma_p Q_1) \bowtie Q_2$$

Projection pushdown.

$$\pi_{C_1}\left(Q_1 \bowtie Q_2\right)$$

$$\Downarrow$$

$$\pi_{C_1}\left(\left(\pi_{C_1,\, col(Q_1) \cap col(Q_2)}\, Q_1\right) \bowtie Q_2\right)$$

Invoice

Customer Id	Amount

Salary

Person Id	Salary

Ex: T_1

Company	State	City
Apple	CA	Cupertino
Microsoft	WA	Redmond
Oracle	TX	Austin
Twitter	CA	SF

T_2

Company	Social Netwo[rk]
Microsoft	LinkedIn
Microsoft	GitHub
Twitter	X

π_{state}

$\bowtie$

$1 = 4$

Useless

Useless.

All states that have a social network HQ.

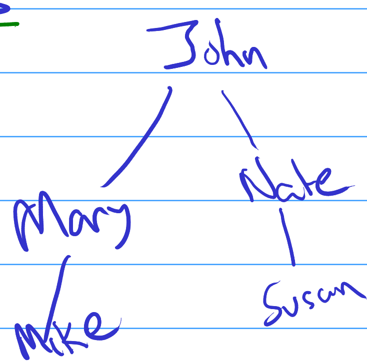
$$\pi_S \left(\pi_{CS} T_1 \bowtie \pi_C T_2 \right)$$

Datalog Relation algebra with recursion.

Ex: Non-recursive

All pairs of cousins

People with the same grandparent.



Parent

Parent (P)	Child (C)
John	Mary
Mary	Mike
John	Nate
Nate	Susan

$$\pi_{ad}\left(\underset{a\ \ b}{\text{Parent}} \underset{b=c}{\bowtie} \underset{c\ \ d}{\text{Parent}}\right) = GP$$

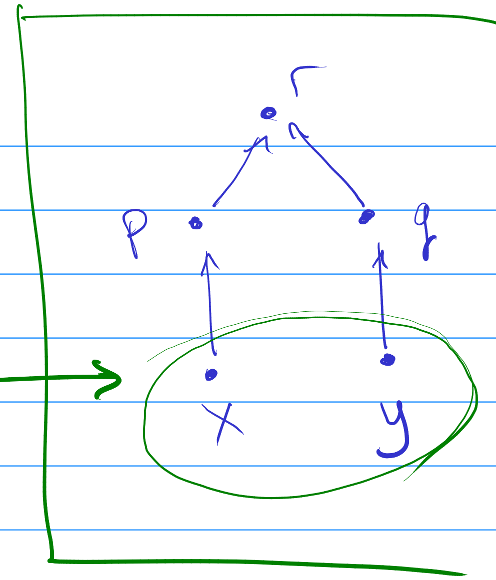
"Views"

Common Table Expression

$$\pi_{bd}\left(\underset{a\ \ b}{GP} \underset{a=c}{\bowtie} \underset{c\ \ d}{GP}\right) = \text{Cousins}.$$

Datalog (The big idea)

Why don't we just write down this picture?



Cousins(x y) :- Parent(p x), Parent(r p),
Parent(q y), Parent(r q).

① Whenever I can find five people (p q x y r) related in this way...

② x & y must be cousins.