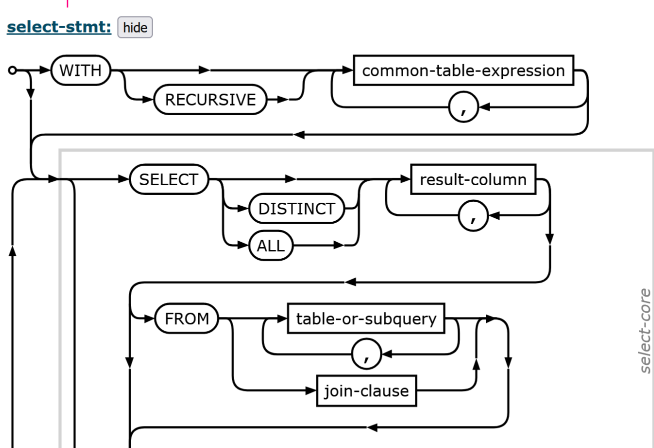
Select	Project	Join	Union	Intersection	Differences
σ	π	$*$	$\cup$	$\cap$	$\setminus$
		$\bowtie$			
		⟕			



Select Stmt ::= RecQual . SelCore
| SelCore

RecQual ::= With CTEs
| With Recursive CTEs

CTEs ::= CTE | CTE , CTEs

Select	Project	Join	Union	Intersection	Differences
σ	π	$*$	$\cup$	$\cap$	$\setminus$
		$\bowtie$			
		$⟕$			

$\sigma_P\ Q$: Finish evaluating Q

Say it produces a table T

Look at all rows of T

Find those which satisfy P

Output those rows

$\pi_{c_1 c_2 \ldots c_K}\ Q$: Finish evaluating Q

Say it produces T

For each row of T,

create a new row with columns $c_1\ c_2 \cdots c_K$

$Q_1 \cap Q_2$: Finish evaluating $Q_1 \longrightarrow T_1$

Finish evaluating $Q_2 \longrightarrow T_2$

Alternative algorithms which evaluate Q_2 "lazily"

Magic Sets Algorithm

for each row $r \in T_1$:

if $r \in T_2$:

then emit r

$Q_1 * Q_2$: $Q_1 \longrightarrow T_1$

$Q_2 \longrightarrow T_2$

for each row $r_1 \in T_1$

for each row $r_2 \in T_2$

emit $r_1 ++ r_2$

$Q_1 \bowtie_c Q_2$: $Q_1 \longrightarrow T_1$

$Q_2 \longrightarrow T_2$

$\forall r_1 \in T_1$

$\forall r_2 \in T_2$ s.t $r_2[c] = r_1[c]$

emit $r_1 ++ r_2$

$\forall r_1 \in T_1$

$\forall r_2 \in T_2$

if $r_1[c] = r_2[c]$

then ______

Loop Nest Join

Query Q

$$\text{Database}, T_1, T_2 \cdots \longrightarrow \boxed{} \longrightarrow \text{Output } T$$

How quickly can you produce T?

Exponential — [Combined complexity] time $= f(D, Q)$

Polynomial time — [Data complexity] time $= f_Q(D)$

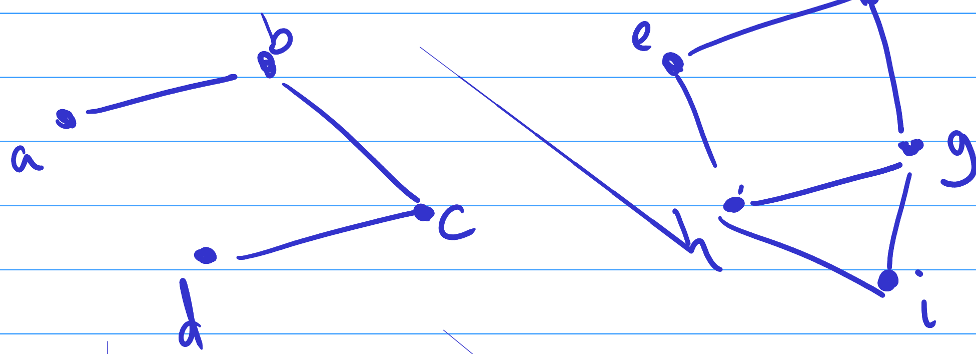
Circuit complexity

$Q_1 \bowtie Q_2 \bowtie Q_3 \cdots \bowtie Q_n$:

for $r_1 \in T_1$:
  for $r_2 \in T_2$:
    $\vdots$
      for $r_n \in T_n$: ___

$|T_1| \cdot |T_2| \cdots |T_n|$

The Triangle Query

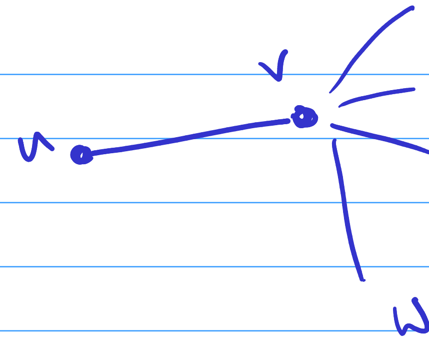


E	
a	b
b	c
c	d

E	
e	f
e	h
f	g
h	g
h	i
g	i

Q: Is there a triangle in the input graph?

$$\Pi_{124}\left(\left(E \bowtie_{c_2=c_1} E\right) \bowtie_{\substack{c_4=c_1 \\ c_1=c_2}} E\right)$$



for $(u,v) \in E$:

for $(v,w) \in E$:

$O(n^2)$ if $(w\ u) \in E$:

then emit (u,v,w)

On the other side:

One concrete family of graphs which have $O(n^{1.76})$ triangles.

Two Practical Optimizations

Two tables : Employee   Zip

Name	Salary	Zipcode

Zipcode	Area name
90089	Univ Park
90015	DTLA
90017	DTLA
⋮	⋮

$$\sigma_{\substack{c_2 > 100K \\ c_5 = DTLA}} \left(\underbrace{\underbrace{E}_{\text{2000 tuples}} \bowtie_{c_3 = c_1} \underbrace{Z}_{\text{2000 tuples}}}_{\text{Say, 30000 tuples}} \right)$$

$\boxed{\text{Selection Pushdown}}$

Say, 30000 tuples ← Requires 2000 log 2000 units of time

Say, 500 tuples ← Requires 30000 units of time

$$\sigma_{\substack{c_2 > 100K \\ c_5 = DTLA}} \left(\overbrace{E \bowtie_{c_3 = c_1} Z}^{2000} \right)$$

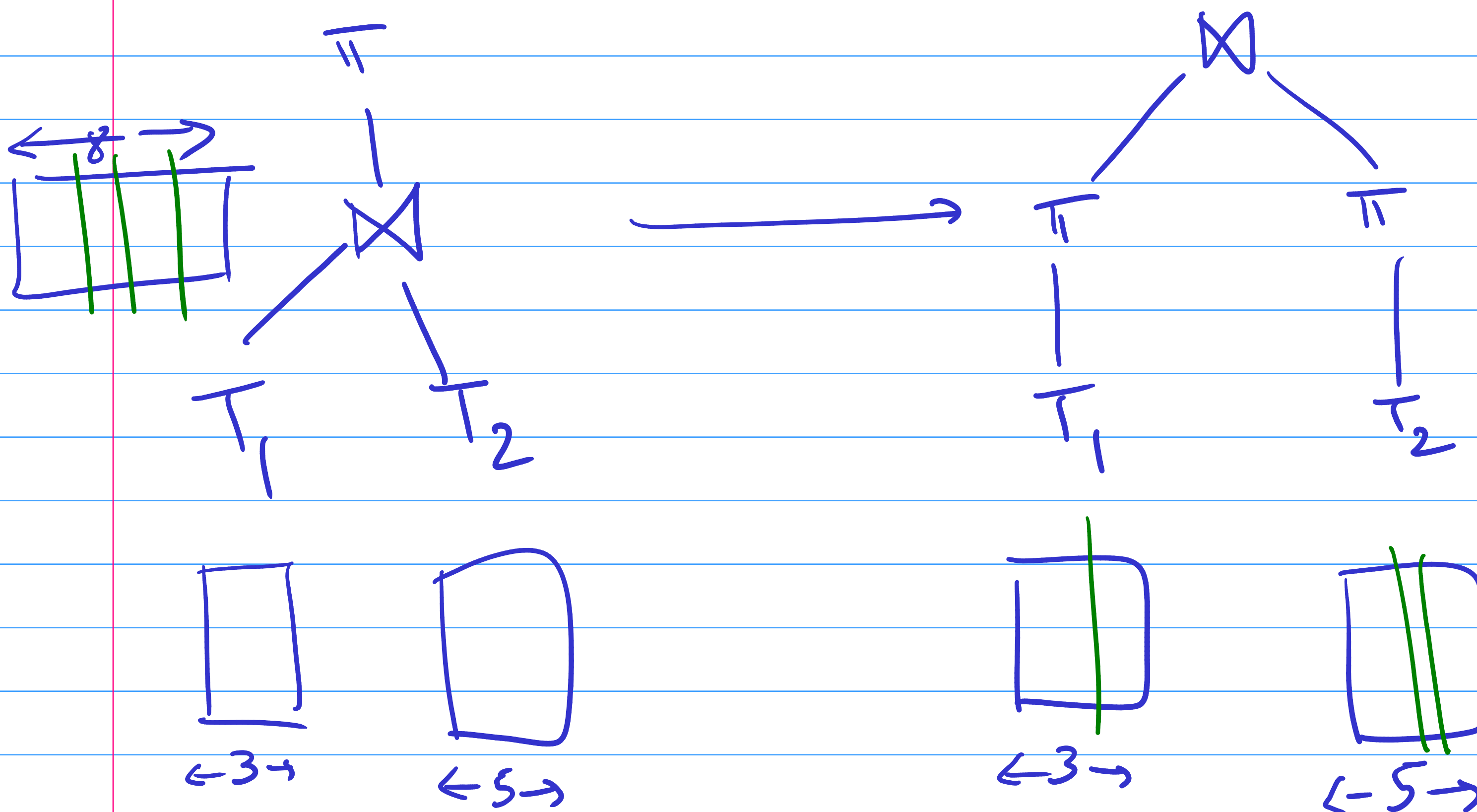
$$\underbrace{\underbrace{\left(\sigma_{c_2 > 100K} E\right)}_{1000} \bowtie_{c_3 = c_1} \underbrace{\left(\sigma_{c_2 = DTLA} Z\right)}_{1/2/3}}_{\text{Total} = 7000 \text{ units of time}}$$

Selection Pushdown = Distributivity of σ over $\bowtie$

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

$$\sigma_f (T_1 \bowtie T_2) = (\sigma_{f_1} T_1) \bowtie (\sigma_{f_2} T_2)$$

Optimization 2 : Projection pushdowns



All states with the HQ of a social network

Ex

Companies	states	city
Apple	CA	Cupertino
Microsoft	WA	Redmond
Oracle	TX	Austin
Twitter	CA	Palo Alto

π_{state}

$\uparrow$

$\bowtie_{Comp}$

Company	Soc Nets
Microsoft	LinkedIn
Twitter	Twitter
Microsoft	GitHub

(4 × 3)

(4 × 1)

$CSC \longrightarrow \pi_{12} \longrightarrow \bowtie_{Company} \longleftarrow \pi_{1} \longleftarrow CS$

$\downarrow$

π_{state}