Homework 2

1. $\sigma_{p_1}(R_1 \bowtie_{p_2} R_2) \overset{?}{\equiv} \sigma_{p_1}(R1) \bowtie_{p_2} R_2$ <u>yes</u> ✓

since $\mathcal{F}(p_1) \subseteq A(R_1)$ then the two expressions are the same

2. <u>yes</u> ✓

3. $\sigma_{p_1}(R_1 \bowtie_{p_2} R_2) \overset{?}{\equiv} \sigma_{p_1}(R1) \bowtie_{p_2} R_2$

R_1

a	b
1	2
3	4

R_2

x	y
2	2
5	6

✓

p1 : a = 1
p2 : b = x

1	2	2	2
-	-	5	6

<u>vs</u>

1	2	2	2
-	-	5	6

a = 1 | b = x | 1, 2; 3, 4 | 2, 2; 5, 6

4. $\sigma_{p_1}(R_1 \bowtie_{p_2} R_2) \overset{?}{\equiv} R_1 \bowtie_{p_1 \wedge p_2} R_2$

R_1

a	b
1	2
3	4

R_2

x	y
1	2
5	6

p1 : b = 2
p2 : a = x

1	2	1	2
-	-	5	6

<u>vs</u>

1	2	1	2
-	-	5	6

5. $\sigma_1(R_1 \cup \rho_{x \to a, y \to b}(R_2)) \overset{?}{\equiv} \sigma_{p_1}(R_1) \cup \sigma_{p_1}(\rho_{x \to a, y \to b}(R_2))$

yes $\longrightarrow$ $\sigma_p(A \cup B) = \sigma_p(A) \cup \sigma_p(B)$

6. yes $\rightarrow$ $\sigma_p(A \cap B) = \sigma_p(A) \cap \sigma_p(B)$

Exercise 2 [accesstime]

N_R = 1K pages (memory) each page → 50 tuples
N_S = 100K pages (disk)

av. access time = 10ms
transfer speed = 10K pages/sec.

Nested loops Join : $O(n \times m)$

$$t_{pages_s} = \underbrace{10\text{ ms}}_{\text{access pages in disk}} + \underbrace{N_S / 10K \text{ pages. sec}}_{\text{download pages to memory}}$$

$$t_{NL} = \underbrace{1000}_{\text{each page R}} \times (\underbrace{50}_{\text{each tuple R}} \times \underbrace{t_{pages_s}}_{\text{all pages S}}) = 50\,000 \times \left(10 + \frac{100\,000}{10\,000} \times 10^3\right) =$$

$$= 50000 \times (10 + 10^4) = 500\ 500\ 000$$

Block Nested Loops Join

```
# 10
for b_R in R:                    # 100K
    for each page p_S in S:      # load pages
        for each tuple in p_S:   # 50
            for each page in b_R:
                for each tuple in page:
                    tey-join
```

I/O, CPU (no cost) → comparissions with block of R in RAM

$$t_{pages_s} = 10\text{ ms} + N_S / 10K \text{ pages. sec}$$

$$t_{BNL} = 10 \times (10 + 10^4)$$

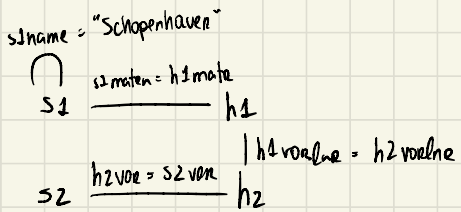
$$= 10 \times (10 + 10^4) = 100\ 100$$

Homework 3

2. [querygraph]

|students| = 8
|hoeren| = 13
|professoren| = 7
|voorlesungen| = 10

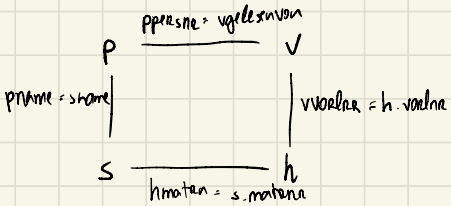
1.



4		
8		
13		
13		
8		

4		
0	0	1
0	1	1
1	2	1
2	3	1

2.



4		
7		
10		
13		
8		

4		
0	1	1
1	2	1
2	3	1
3	0	1

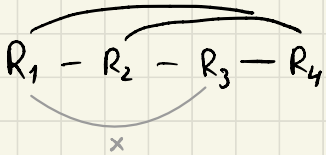
Homework 4

1. [cpoptimal] $|R_1| = 2$ $|R_2| = 100$ $|R_3| = 2$ $|R_4| = 10$

$R_1 \times R_3 = 4$

$f_{1,2} = ?$
$f_{2,3} = ?$
$f_{2,4} = ?$

$f_{1,3} = 1$



$$R_1 \bowtie R_2 = f_{1,2} \times 2 \times 100 + 0 + 0 = 200 f_{1,2}$$
$$R_1 \times R_3 = 4$$
$$R_2 \bowtie R_3 = f_{2,3} \times 100 \times 2 = 200 f_{2,3}$$
$$R_2 \bowtie R_4 = f_{2,4} \times 100 \times 10 = 1000 f_{2,4}$$

$$|R_1 \bowtie R_2| + f_{1,2} \times |R_1 \bowtie R_2| \times |R_3|$$
$$\parallel$$
$$|(R_1 \bowtie R_2) \bowtie R_3|$$

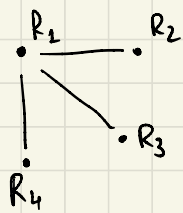
$$((R_1 \bowtie R_2) \bowtie R_3) \bowtie R_4 = 200 f_{1,2} + f_{2,3} \times 200 f_{1,2} \times 2 + (f_{2,4} \times (200 f_{1,2} + f_{2,3} \times 200 f_{1,2} \times 2) \times 10)$$
$$((R_2 \bowtie R_3) \bowtie R_4) \bowtie R_1 = 200 f_{2,3} + f_{2,4} \times 200 f_{2,3} \times 10 + (f_{1,2} \times (200 f_{2,3} + f_{2,4} \times 200 f_{2,3} \times 10) \times 2)$$
$$((R_2 \bowtie R_3) \bowtie R_1) \bowtie R_4 =$$
$$((R_1 \times R_3) \bowtie (R_2 \bowtie R_4)) = 4 + 1000 f_{2,4} + f_{3,4} \times 4 \times 1000 f_{2,4}$$

$$C_{out}(T) = \begin{cases} 0 & \text{if } T \text{ is leaf } R_i \\ |T| + C_{out}(T_1) + C_{out}(T_2) & \text{if } T = T_1 \bowtie T_2 \end{cases}$$

- $f_{2,4}$ – muito pequeno (0.01)

$f_{1,2}$, $f_{2,3}$ – muito grandes (0.5)

$f_{3,4} = 0.1$



$$|R_1| = |R_2| = |R_3| = |R_4| = c$$

$$f_{1,2} = f_{1,3} = f_{1,4} = 0.5$$

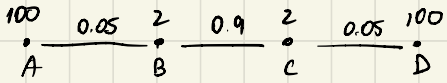
$$R_1 \bowtie R_2 = R_1 \bowtie R_3 = R_1 \bowtie R_4 = 0.5 \times c^2$$

$$(R_1 \bowtie R_2) \bowtie R_3 = 0.5c^2 + 0.5^2 c^2 c = 0.25c^3 + 0.5c^2$$
$$((R_1 \bowtie R_2) \bowtie R_3) \bowtie R_4 \approx 0.5^4 \cdot c^4 \qquad O(c^4 + c^3 + c^2 + c^2)$$

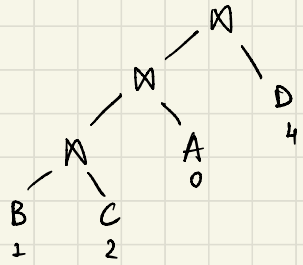
$$R_2 \times R_3 = c^2$$

$$(R_2 \times R_3) \bowtie (R_1 \bowtie R_4) = c^2 + 0.5c^2 + 0.5^3 c^4$$ (no c^3 term)

2. [greedyalgo]



((1 2) 0) 4)



GOO

$A \bowtie B = 10$

$B \bowtie C = 3.6$

$C \bowtie D = 10$

$(B \bowtie C) \bowtie A = 0.05 \times 3.6 \times 100 = 18$

$(B \bowtie C) \bowtie D = \quad 4$

$((B \bowtie C) \bowtie A) \bowtie D = 0.05 \times 18 \times 100 = 90$

$Cout = 90 + 18 + 3.6 = 111.6$

Optimal

$A \bowtie B = 10$

$C \bowtie D = 10$

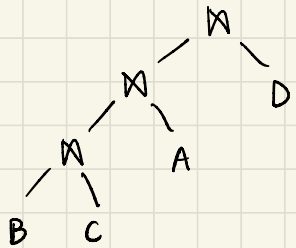
$(A \bowtie B) \bowtie (C \bowtie D) = 0.9 \times 10 \times 10 = 90$

$Cout = 110$

GJO-1

GJO-2

GJO-3



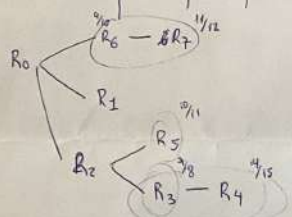
Homework 5.

2. [ikkbz]

R.Name	n	s	C	T	RANK	
R_1	10	1/10	1	1	0	
R_2	50	1/5	10	10	9/10	0.9
R_3	40	1/5	8	8	7/8	0.875
R_4	30	3/10	15	15	14/15	0.933
R_5	55	1/5	11	11	10/11	0.909
R_6	50	1/5	10	10	9/10	0.9
R_7	60	1/5	12	12	11/12	0.9167
R_2R_3	0	0	90	80	79/80	0.988

$$\text{Rank} = \frac{T-1}{C}$$

1.

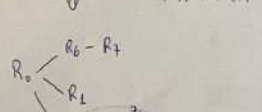


$\Rightarrow \dots R_2 - R_3 - R_5 - R_4$

Rank(4) > Rank(3)

$\Downarrow$

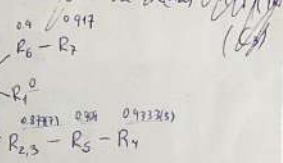
2.



$C(R_2R_3) = C(R_2) + T(R_2)C(R_3)$
$= 10 + 10 \times 8 = 90$

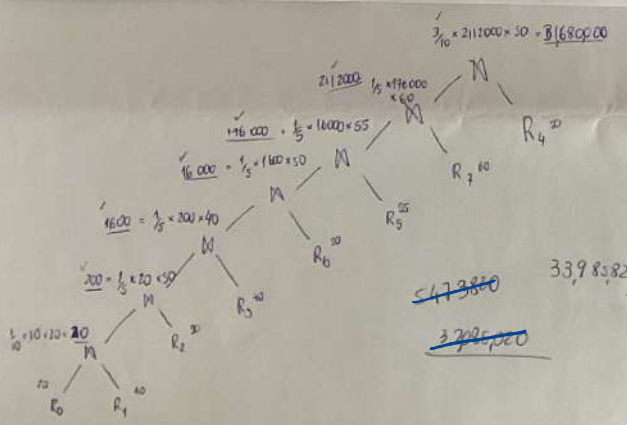
$T(R_2R_3) = \prod_{i=2,3} n_i \cdot s_i$
$= \frac{1}{5} \times \frac{1}{5} \times 50 \times 40$
$= 10 \times 8 = 80$

3.



4.

$R_0 - R_2 - R_{2,3} - R_6 - R_5 - R_1 - R_4$



~~5173820~~

~~3202500~~

33,985,820

Homework 6

1. [ikkbzmvp]

$$\overset{10}{R_0} \;\overset{0.1}{—}\; \overset{20}{R_1} \;\overset{0.2}{—}\; \overset{10}{R_2} \;\overset{0.05}{—}\; \overset{100}{R_3}$$

$$\text{rank} = \frac{T-1}{C}$$

R_0

R. name	n	s	C	T	rank
R_1	20	0.1	2	2	1/2
R_2	10	0.2	2	2	1/2
R_3	100	0.05	5	5	4/5

1. $R_0 \;—\; \overset{1/2}{R_1} \;—\; \overset{1/2}{R_2} \;—\; \overset{4/5}{R_3}$ DONE!

R_1

R. name	n	s	C	T	rank
R_0	10	0.1	1	1	0
R_2	10	0.2	2	2	1/2
R_3	100	0.05	5	5	4/5

1. $R_1 \;/\; \overset{0}{R_0}$, $R_1 \;\backslash\; \overset{1/2}{R_2} \;—\; \overset{4/5}{R_3}$ $\xRightarrow{\text{normalize}}$ $R_1 — R_0 — R_2 — R_3$ DONE!

R_2

R. name	n	s	C	T	rank
R_0	10	0.1	1	1	0
R_1	20	0.2	4	4	3/4
R_3	100	0.05	5	5	4/5
R_1R_0	0	0	8	4	3/8 = 0.375

$$C(R_1R_0) = C(R_1) + T(R_1)\,C(R_0) = 4 + 4 \times 1 = 8$$

$$T(R_1R_0) = \prod s_i n_i = 0.1 \times 0.2 \times 10 \times 20 = 4$$

1. $R_2 \;/\; \overset{3/4}{R_1} — \overset{0}{R_0}$, $R_2 \;\backslash\; \overset{4/5}{R_3}$

$\text{rank}(R_1) > \text{rank}(R_0)$

2. $R_2 \;/\; \overset{3/8}{R_{10}}$, $R_2 \;\backslash\; \overset{4/5}{R_3}$

3: $R_2 - R_{10} - R_3 \Rightarrow R_2 - R_1 - R_0 - R_3$ DONE

R_3

R.name	n	s	C	T	rank
R_0	10	0.1	1	1	0
R_1	20	0.2	4	4	3/4
R_2	10	0.05	0.5	0.5	-1
$R_1 R_0$	0	0	8	4	0.375

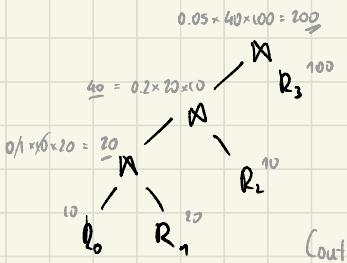
1. $R_3 - R_2^{-1} - R_1^{3/4} - R_0^{0} \Rightarrow R_3 - R_2^{-1} - R_{10}^{0.375}$

$\text{rank}(R_0) < \text{rank}(R_1)$

2. $R_3 - R_2 - R_1 - R_0$

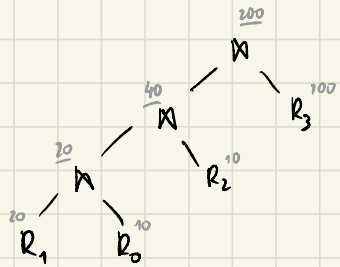
Costs

R_0



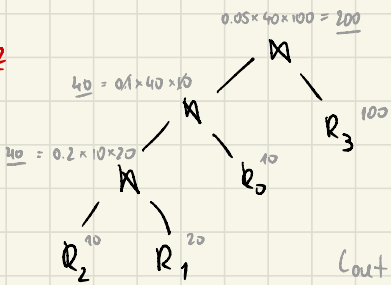
$C_{out} = 260$

R_1



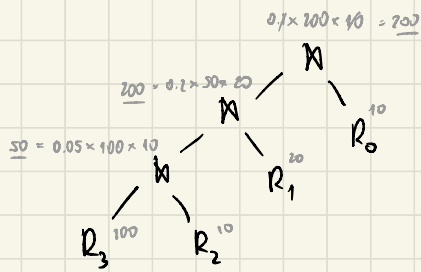
$C_{out} = 260$

R_2



$C_{out} = 280$

R_3



$C_{out} = 450$

2. [mst]

$C_{edge} = |R_0 \bowtie R_1|$

$|R_1 \bowtie R_3| = 0.1 \times 20 \times 30 = 60$ 3

$|R_1 \bowtie R_0| = 0.3 \times 20 \times 50 = 300$

$|R_3 \bowtie R_0| = 0.1 \times 30 \times 50 = 150$ 5

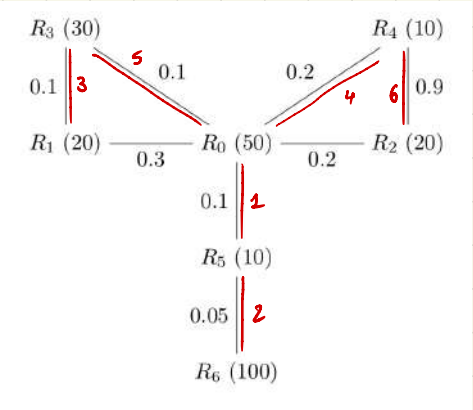
$|R_0 \bowtie R_4| = 0.2 \times 10 \times 50 = 100$ 4

$|R_0 \bowtie R_2| = 0.2 \times 50 \times 20 = 200$

$|R_4 \bowtie R_2| = 0.9 \times 10 \times 20 = 180$ 6

$|R_0 \bowtie R_5| = 0.1 \times 10 \times 50 = 50$ 1

$|R_5 \bowtie R_6| = 0.05 \times 10 \times 100 = 50$ 2



same but join nodes when edge is selected

~~$|R_1 \bowtie R_3| = 0.1 \times 20 \times 30 = 60$~~ · 2

~~$|R_1 \bowtie R_{0,5}| = 0.3 \times 20 \times 50 = 300$~~

~~$|R_3 \bowtie R_{0,5}| = 0.1 \times 30 \times 50 = 150$~~

~~$|R_{0,5} \bowtie R_4| = 0.2 \times 10 \times 50 = 100$~~

~~$|R_{0,5} \bowtie R_2| = 0.2 \times 50 \times 20 = 100$~~

~~$|R_4 \bowtie R_2| = 0.9 \times 10 \times 20 = 180$~~

~~$|R_0 \bowtie R_5| = 0.1 \times 10 \times 50 = 50$~~ 1

~~$|R_5 \bowtie R_0| = 0.05 \times 10 \times 100 = 50$~~

~~$|R_{0,5} \bowtie R_6| = 0.05 \times 50 \times 100 = 250$~~

~~$|R_{13} \bowtie R_{05}| = 0.1 \times 0.3 \times 60 \times 50 = 90$~~ · 3

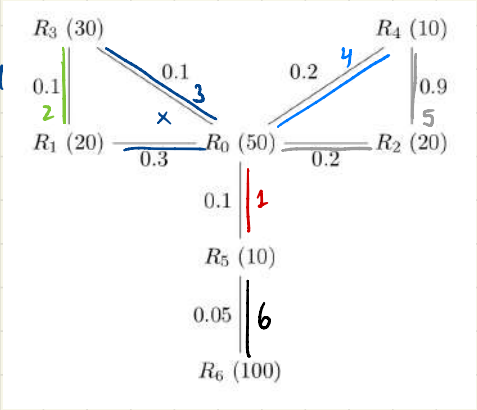
~~$|R_{0135} \bowtie R_4| = 0.2 \times 90 \times 10 = 180$~~ · 4

~~$|R_{0135} \bowtie R_6| = 0.05 \times 90 \times 100 = 450$~~

~~$|R_{01354} \bowtie R_2| = 0.2 \times 0.9 \times 180 \times 20 = 648$~~ · 5

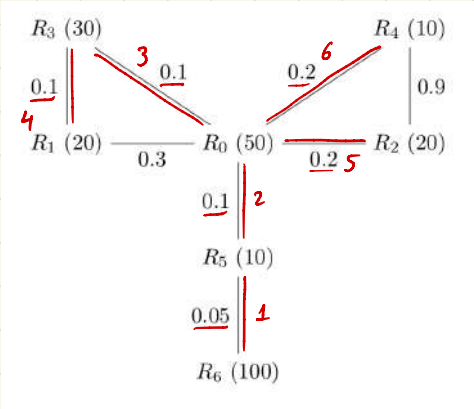
~~$|R_{01354} \bowtie R_6| = 0.05 \times 180 \times 100 = 900$~~

$|R_{012345} \bowtie R_6| = 0.05 \times 100 \times 648 = 3240$

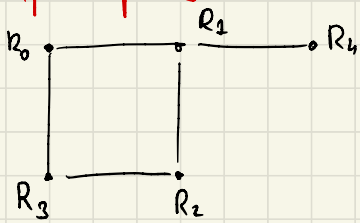


Solution:

$C_{edge} = \text{selectivity}_{edge}$



3. [dpsizedpsub]



DPsize without crossproducts:

1: 0, 1, 2, 3, 4

2: 0|1[1], 0|3[2], 1|2[3], 1|4[4], 2|3[5] ✓

3: 0|12[6], 0|14[7], 0|23[8], 1|03[9], 1|23[10], 2|01[11], 2|14[12], 2|03[13], 3|01[14], 3|12[15], 4|01[16], 4|12[17] ✓

~~0|12~~, ~~01|3~~, ~~01|4~~, ~~03|1~~, ~~03|2~~, ~~12|0~~, ~~12|3~~, ~~12|4~~, ~~14|0~~, ~~14|2~~, ~~23|0~~, ~~23|1~~

4: 0|123[18], 0|124[19], 1|023[20], 2|013[21], 2|014[22], 3|012[23], 3|014[24], 3|124[25], 4|012[26], 4|013[27], 4|123[28], 01|23[29], 03|12[30], ~~12|03~~, 03|14[31], 14|23[32], ~~21|03~~, ~~23|01~~, ~~23|14~~, ~~012|3~~, ~~012|4~~

5: 0|1234[33], 2|0134[34], 3|0124[35], 4|0123[36], 03|124[37], 14|023[38], 23|014[39]

DPsub with crossproducts:

0: -

1: -

10: 0 $1 \bowtie 0$[1]

20: 0 2×0[2]

21: 1 $2 \bowtie 1$[3]

210: 0, 1, 10 $0 \bowtie 21$[4], $1 \bowtie 20$[5], $10 \bowtie 2$[6]

3: -

30: $3 \bowtie 0$[7]

31: 1 3×1[8]

310: 0, 1, 10 $0 \bowtie 31$[9], $1 \bowtie 30$[10], $10 \bowtie 3$[11]

32 : 2 $\quad$ 3 ⋈ 2[12]

320 : 0, 2, 20 $\quad$ 0 ⋈ 32[13], 2 ⋈ 30[14], 20 ⋈ 3[15]

321 : 1, 2, 21 $\quad$ 1 ⋈ 32[16], 2 ⋈ 31[17], 3 ⋈ 21[18]

3210 : 0, 1, 10, 20, 21, 210 : $\quad$ 0 ⋈ 321[19], 1 ⋈ 320[20], 10 ⋈ 32[21], 2 ⋈ 310,
2,
20 ⋈ 13[22], 21 ⋈ 03[23], 210 ⋈ 3[24]

4 : –

40 : 0 $\quad$ 4 ⋈ 0[25]

41 : 1 $\quad$ 4 ⋈ 1[26]

410 : 0, 1, 10 $\quad$ 0 ⋈ 41[27], 1 ⋈ 40[28], 10 ⋈ 4[29]

42 : 2 $\quad$ 4 ⋈ 2[30]

420 : 0, 2, 20 $\quad$ 0 ⋈ 42[31], 2 ⋈ 40[32], 20 ⋈ 4[33]

421 : 1, 2, 21 $\quad$ 1 ⋈ 42[34], 2 ⋈ 41[35], 21 ⋈ 4[36]

4210 : 0, 1, 10, 2, 20, 21, 210 $\quad$ 0 ⋈ 421[37], 1 ⋈ 420[38], 10 ⋈ 42[39], 2 ⋈ 410[40],
20 ⋈ 41[41], 21 ⋈ 40[42], 210 ⋈ 4[43]

43 : 3 $\quad$ 4 ⋈ 3[44]

430 : 0, 3, 30 $\quad$ 0 ⋈ 43[45], 3 ⋈ 40[46], 30 ⋈ 4[47]

431 : 1, 3, 31 $\quad$ 1 ⋈ 43[48], 3 ⋈ 41[49], 31 ⋈ 4[50]

4310 : 0, 1, 10, 3, 30, 31, 310 $\quad$ 0 ⋈ 431[51], 1 ⋈ 430[52], 10 ⋈ 43[53], 3 ⋈ 410[54]
30 ⋈ 41[55], 31 ⋈ 40[56], 310 ⋈ 4[57]

432 : 2, 3, 32 $\quad$ 2 ⋈ 43[58], 3 ⋈ 42[59], 32 ⋈ 4[60]

4320 : 0, 2, 20, 3, 30, 32, 320 $\quad$ 0 ⋈ 432[61], 2 ⋈ 430[62], 20 ⋈ 43[63], 3 ⋈ 420[64]
30 ⋈ 42[65], 32 ⋈ 40[66], 320 ⋈ 4[67]

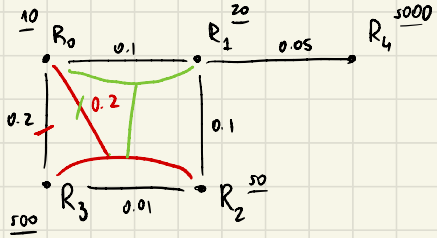
4321 : 1, 2, 21, 3, 31, 32, 321 $\quad$ 1 ⋈ 432[68], 2 ⋈ 431[69], 21 ⋈ 43[70], 3 ⋈ 421[71]
31 ⋈ 42[72], 32 ⋈ 41[73], 231 ⋈ 4[74]

43210 : 0, 1, 10, 2, 20, 21, 210, 3, 30, 31, 310, 32, 320, 321, 3210
0 ⋈ 4321[75], 1 ⋈ 4320[76], 10 ⋈ 432[77], 2 ⋈ 4310[78],
20 ⋈ 431[79], 21 ⋈ 430[80], 210 ⋈ 43[81], 3 ⋈ 4210[82],
30 ⋈ 421[83], 31 ⋈ 420[84], 310 ⋈ 42[85], 32 ⋈ 410[86]
320 ⋈ 41[87], 321 ⋈ 40[88], 3210 ⋈ 4[89]

Total : 89 + 1

Homework 7

4. [simplification]



1 $R_0 \bowtie R_1$	$= 20$	
$R_0 \bowtie R_3$	$= 1000$	
5 $R_1 \bowtie R_2$	$= 100$	
$R_1 \bowtie R_4$	$= 5000$	
$R_2 \bowtie R_3$	$= 250$	

1.

$\bowtie_1$: $R_3 \bowtie R_0$

$\bowtie_2$: $R_3 \bowtie R_2$

$\Downarrow$

$\{R_2 R_3\} \bowtie R_0$

NEW

7. $R_0 \bowtie R_1$	$= 20$
$R_2 R_3 \bowtie R_0$	$= 750$
$R_1 \bowtie R_2$	$= 100$
$R_1 \bowtie R_4$	$= 5000$
$R_2 \bowtie R_3$	$= 250$

2.

$\bowtie_1$: $R_0 \bowtie R_2 R_3$

$\bowtie_2$: $R_0 \bowtie R_1$

$\Downarrow$

$\{R_0, R_1\} \bowtie \{R_2 R_3\}$

6 options:

1. $$oB(R_0 \bowtie \underline{R_1}, R_0 \bowtie \underline{R_3}) = \frac{C((R_0 \bowtie R_1) \bowtie R_3)}{C((R_0 \bowtie R_3) \bowtie R_1)} = \frac{20 + 0.2 \times 20 \times 500}{1000 + 0.1 \times 1000 \times 20} = \frac{2020}{3000}$$

$$oB(R_0 \bowtie \underline{R_3}, R_0 \bowtie \underline{R_1}) = \frac{300}{202}$$

2. $$oB(R_1 \bowtie \underline{R_0}, R_1 \bowtie \underline{R_2}) = \frac{20 + 0.1 \times 20 \times 50}{100 + 0.1 \times 100 \times 10} = \frac{120}{200}$$

$$oB(R_1 \bowtie \underline{R_2}, R_1 \bowtie \underline{R_0}) = \frac{20}{12}$$

3. $$oB(R_1 \bowtie \underline{R_0}, R_1 \bowtie \underline{R_4}) = \frac{20 + 0.05 \times 20 \times 5000}{5000 + 0.1 \times 5000 \times 10} = \frac{5020}{10.000} \quad / \quad \frac{1000}{502}$$

4. $$oB(R_1 \bowtie \underline{R_2}, R_1 \bowtie \underline{R_4}) = \frac{100 + 0.05 \times 100 \times 5000}{5000 + 0.1 \times 5000 \times 50} = \frac{25.100}{30.000} \quad / \quad \frac{300}{251}$$

5. $$oB(R_2 \bowtie \underline{R_1}, R_2 \bowtie \underline{R_3}) = \frac{100 + 0.01 \times 100 \times 500}{250 + 0.1 \times 250 \times 20} = \frac{600}{750} \quad / \quad \frac{750}{600}$$

6. $$oB(R_3 \bowtie \underline{R_0}, R_3 \bowtie \underline{R_2}) = \frac{1000 + 0.01 \times 1000 \times 50}{250 + 0.2 \times 250 \times 10} = \frac{1500}{750} = 2 \quad / \quad \frac{1}{2}$$

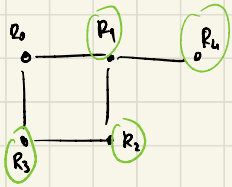
Best

NEW:

7. $$oB(R_0 \bowtie (\underline{R_2 R_3}), R_0 \bowtie \underline{R_1}) = \frac{750 + 0.1 \times 0.1 \times 750 \times 20}{20 + 0.1 \times 0.2 \times 20 \times 250}$$

Best $\longrightarrow$ $$= \frac{900}{120} \quad / \quad \frac{120}{900}$$

3. [dpccp]



$\{4\}$: —

$\{3\}$: —

$\{2\}$: $\{3\}$

valid neighbors: $\{2\}, \{4\}$

$\{2,3\}$: —

$\{1\}$: $\{4\}, \{2\}, \{2,3\}$

$\{1,2\}$: $\{4\}, \{3\}$

$\{1,2,3\}$: $\{4\}$

$\{1,4\}$: $\{2\}, \{2,3\}$

$\{1,2,4\}$: $\{3\}$

$\{1,2,3,4\}$: — ✓

here we consider $2>0$, not $2<3$ so this pair is valid.

$\{0\}$: $\{3\}, \{2,3\}, \{1\}, \{1,2\}, \{1,2,3\}, \{1,4\}, \{1,2,4\}, \{1,2,3,4\}$ — ✓

$\{0,1\}$: $\{4\}, \{3\}, \{2\}, \{2,3\}$

$\{0,1,2\}$: $\{4\}, \{3\}$

$\{0,1,4\}$: $\{3\}, \{2\}, \{2,3\}$

$\{0,1,2,4\}$: $\{3\}$

$\{0,3\}$: $\{2\}, \{1\}, \{1,2\}, \{1,4\}, \{1,2,4\}$

$\{0,2,3\}$: $\{1\}, \{1,4\}$ — ✓

$\{0,1,3\}$: $\{4\}, \{2\}$

$\{0,1,2,3\}$: $\{4\}$

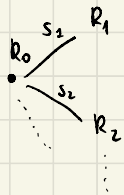
$\{0,1,3,4\}$: $\{2\}$

$\{0,1,2,3,4\}$: —

— ✓

1. Star QO, Greedy star Cout

↳ sort relations by $|R_i| \cdot s_i$



Proof by induction: (on the tree)

Basis: R_0 is optimal for size 1

step: Assume T_K is optimal for size K

T_{K+1}: $T_K \bowtie R_5$, $T_K \bowtie R_3$, $T_K \bowtie R_2$, ...

pick relation $R_i \in R \setminus R(T)$ that minimizes $|T_{K+1}| = |T_K| \cdot s_i \cdot |R_i|$ s.t. $|R_i| \cdot s_i$ is minimal

Principle of optimality:

~~If subtrees are optimal, full tree is optimal~~ ✗

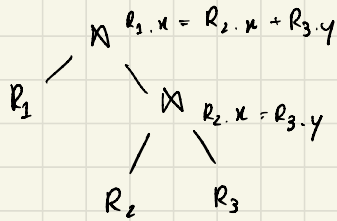
If full tree is optimal, then subtrees <u>must</u> be optimal.

1. GS = IKKBZ on star
2. on star, all trees are Left Deep

↳ <u>GS is optimal on star</u>

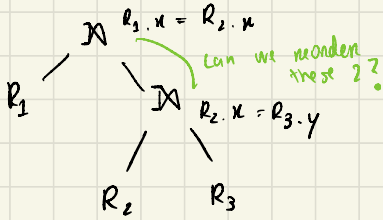
- DP algorithms would have taken too much time => not good for star queries

Homework 8



joins <u>cannot</u> be reordered

SES	TES
R_1 R_2 R_3	R_1, R_2, R_3
R_2 R_3	R_2, R_3 (?)



join <u>can</u> be reordered

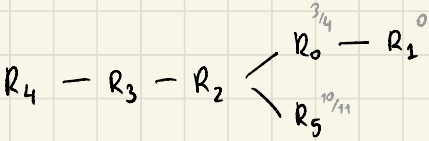
SES	TES
R_1 R_2	R_1, R_2, R_3
R_2 R_3	R_2, R_3

$\Rightarrow$ $(\{R_1\}, \{R_2, R_3\})$
$(\{R_2\}, \{R_3\})$

1. [lindp]

R. name	n	s	C	T	rank
R_0	20	1/5	4	4	3/4
R_1	10	1/10	1	1	0
R_2	50	1/5	10	10	9/10
R_3	40	3/10	12	12	11/12
R_5	55	1/5	11	11	10/11
$R_0 R_1$	–	–	8	4	3/8

1.

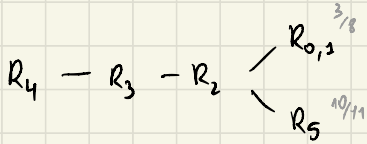


$C(R_0 R_1) = C(R_0) + T(R_0)\, C(R_1)$
$= 4 + 4 \times 1 = 8$

$T(R_0 R_1) = \prod s_i c_i$
$= \frac{1}{5} \times \frac{1}{10} \times 20 \times 10 = 4$

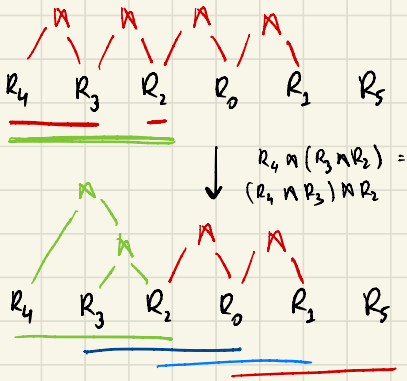
$\text{rank}(R_1) < \text{rank}(R_0)$

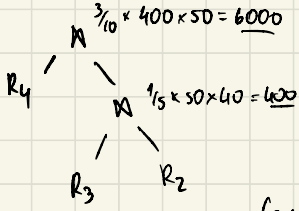
2.



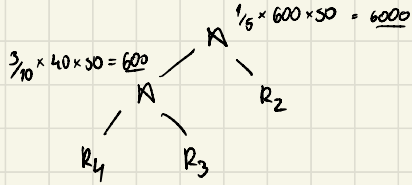
$\Rightarrow$ Lin : $R_4 - R_3 - R_2 - R_{0,1} - R_5$

	R_4	R_3	R_2	R_0	R_1	R_5
R_4	0	600	6400	24800	24820	267.020
R_3		0	400	1800	1820	19120
R_2			0	200	220	2420
R_0				0	20	1120
R_1					0	550
R_5						0



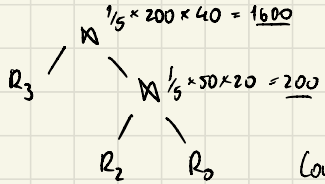


Cout: 6400



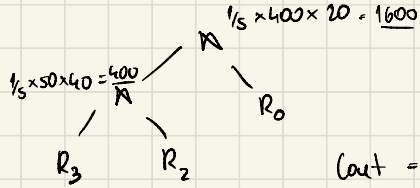
Cout = 6600

$R_3 \bowtie (R_2 \bowtie R_0)$



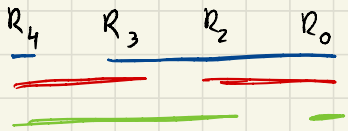
Cout = 1800

$(R_3 \bowtie R_2) \bowtie R_0$



Cout = 2000

...

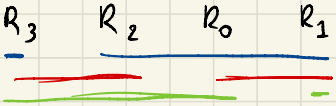


1. $\text{Cout} = \overbrace{1800}^{200+1600} + \frac{3}{10} \times 1600 \times 50 = \times\ 25.800$ ✓

2. $\text{Cout} = \overset{600}{\cancel{50}} + 200 + \frac{1}{5} \times \overset{600}{\cancel{30}} \times 200 = 24.800$ ✓

ver acima! ← $\overbrace{6400}^{400+6000} + \frac{1}{5} \times 6000 \times 20 = 30.400$

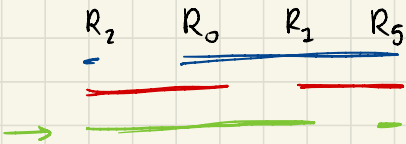
3.



1. $\text{Cout} = \overbrace{220}^{20+200} + \frac{1}{5} \times 200 \times 40 = 1820$ ✓

2. $\text{Cout} = 400 + 20 + \frac{1}{5} \times 400 \times 20 = \overset{2020}{\cancel{2200}}$

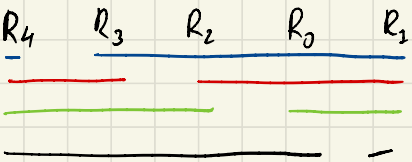
3. $\text{Cout} = \overbrace{1800}^{200+1600} \times \frac{1}{\cancel{10}} \times 1600 \times \cancel{10} = 3400$ ✓



1. $\text{Cout} = 1120 + \frac{1}{5} \times \frac{1}{5} \times 1100 \times 50 = 3320$

2. $\text{Cout} = 200 + 550 + \frac{1}{10} \times \frac{1}{5} \times 200 \times 550 = 2950$

3. $\text{Cout} = 220 + \frac{1}{5} \times 200 \times 55 = 2420$ ✓



1. $\text{Cout} = 1820 + \frac{3}{10} \times 1600 \times 50 = 25.820$ ✗

2. $\text{Cout} = 600 + \overbrace{220}^{20+200} + \frac{1}{5} \times 600 \times 200 = 24.820$ ✓ ✗

3. $\text{Cout} = 6400 + 20 + \frac{1}{5} \times 6000 \times 20 = 30.420$ ✓ ✗

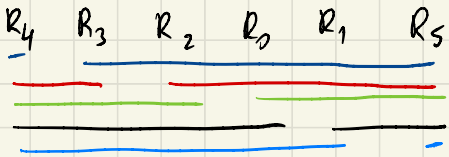
4. $\text{Cout} = \overset{24.800}{\cancel{2250}} + \frac{1}{10} \times 24.000 \times \cancel{10} = 48.800$

1. $\overbrace{2420}^{220+2200} + \frac{1}{5} \times 2200 \times 40 = 20.020$

2. $400 + 1120 + \frac{1}{5} \times \frac{1}{5} \times 400 \times 1100 = 19.120$

3. $\overbrace{1800}^{200+1600} + 550 + \frac{1}{10} \times \frac{1}{5} \times 1600 \times 550 = 19.950$

4. $1820 + \frac{1}{5} \times 1600 \times 55 = 19.420$



1. $19.120 + \frac{3}{10} \times 17.600 \times 50 = 283.120$

2. $600 + 2420 + \frac{1}{5} \times 600 \times 2200 = 267.020$

3. $6400 + 1120 + \frac{1}{5} \times \frac{1}{5} \times 6000 \times 1100 = 271.520$

4. $24800 + 550 + \frac{1}{10} \times \frac{1}{5} \times 24.000 \times 550 = 289.350$

5. $24820 + \frac{1}{5} \times 24000 \times 55 = 288.820$

Optimal order : ((4 3)((2 (0 1)) 5))

2. [Restrictions]

SES	TES
{A, B}	{A, B, C, D, E}
{B, C}	{B, C, D, E}
{C, E}	{C, E, D}
{C, D}	{C, D}
{E, F}	{E, F}

hyper-edges:

A	BCDE
B	CDE
CD	E
C	D
E	F

Homework 9

1. rank = 64 [random permut]

$\langle \underline{1}, 2, 3, 4, 5, 6, 7, \underline{8} \rangle$	$i = 8$	$R \bmod i = 0$	$R = 8$
$\langle 8, \underline{2}, 3, 4, 5, 6, \underline{7}, 1 \rangle$	$i = 7$	$R \bmod i = 1$	$R = 1$
$\langle 8, \underline{7}, 3, 4, 5, \underline{6}, 2, 1 \rangle$	$i = 6$	$R \bmod i = 1$	$R = 0$
$\langle \underline{8}, 6, 3, 4, \underline{5}, 7, 2, 1 \rangle$	$i = 5$	$R \bmod i = 0$	$R = 0$
$\langle \underline{5}, 6, 3, \underline{4}, 8, 7, 2, 1 \rangle$	$i = 4$	$R \bmod i = 0$	$R = 0$
$\langle \underline{4}, 6, \underline{3}, 5, 8, 7, 2, 1 \rangle$	$i = 3$	$R \bmod i = 0$	$R = 0$
$\langle \underline{3}, \underline{6}, 4, 5, 8, 7, 2, 1 \rangle$	$i = 2$	$R \bmod i = 0$	$R = 0$
$\langle \underline{\underline{6}}, 3, 4, 5, 8, 7, 2, 1 \rangle$	$i = 1$	$R \bmod i = 0$	$R = 0$
$\langle 6, 3, 4, 5, 8, 7, 2, 1 \rangle$ ✓			

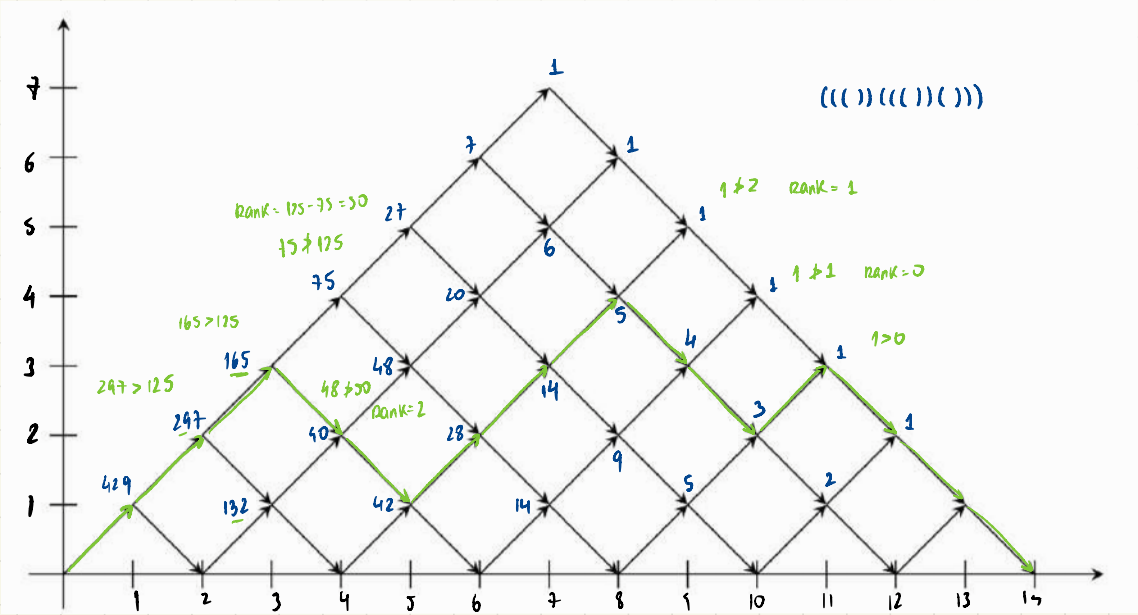
2. [random tree]

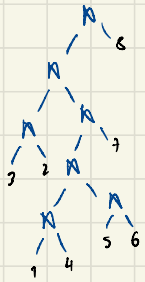
rank = 125

8 relations → 7 inner nodes

$$q(i,j) = p(2n-i, j) = \frac{j+1}{2n-i+1}\binom{2n-i+1}{\frac{1}{2}(2n-i+j)+1}$$

$$= p(7-i, j) = \frac{j+1}{15-i}\binom{15-i}{\frac{1}{2}(14-i+j)+1}$$





permutation : 3 2 1 4 5 6 7 8

(((3 2)(((1 4) (5 6)) 7)) 8)

Homework 10

1. [uniform random]

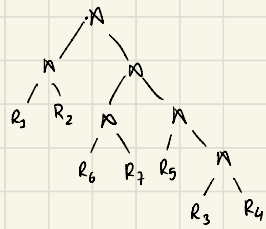
$T_1 : (\langle (R_1 \bowtie R_2), R_3 \rangle, 4) = (1\ 2), 3, 4$

$T_2 : (\langle (R_6 \bowtie R_7), R_5 \rangle, 4) = (6\ 7), 5, 4$

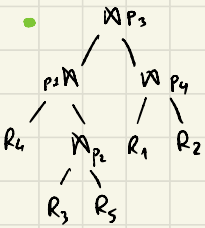
$\alpha = [1, 0, 1]$

Merged: $(\langle (R_1 \bowtie R_2), (R_6 \bowtie R_7), R_5, R_3 \rangle, R_4)$

$\rightarrow (1\ 2), (6\ 7), 5, 3, 4$



2. [genetic encoding]

- 

ordered list : 2 1 4 3

ordered number : 35 41 23 12

$\langle R_3, R_2, \overset{3}{R_3}, R_4, \overset{5}{R_5} \rangle \rightarrow \langle \overset{1}{R_3 \bowtie R_5}, R_1, R_2, \overset{4}{R_4} \rangle \rightarrow \langle (R_4 \bowtie (R_3 \bowtie R_5)), \overset{2}{R_2}, \overset{3}{R_2} \rangle \downarrow$

$\langle \underbrace{R_1 \bowtie R_2}_{1}, \underbrace{R_4 \bowtie (R_3 \bowtie R_5)}_{2} \rangle$

- $(((R_4 \bowtie R_1) \bowtie R_3) \bowtie R_2)$

ordered list : 4 1 3 2

ordered number : 4 1 2 1

$\langle R_1, R_2, R_3, \overset{4}{R_4} \rangle$

$\langle \underline{R_1}, R_2, R_3 \rangle$

$\langle \underline{R_2}, R_3 \rangle$

$\langle \underline{R_2} \rangle$

Encoding : 35 13 23 12

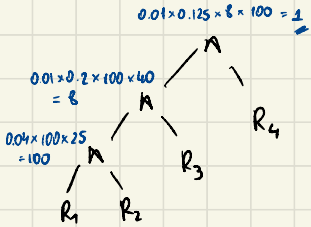
$\langle \underline{3 \bowtie 5}, 1, \underline{2}, 4 \rangle$

$\langle (3 \bowtie 5) \bowtie 2, 1, 4 \rangle$

$\langle 1 \bowtie 4, (3 \bowtie 5) \bowtie 2 \rangle$

3. [order preserving]

$((R_1 \bowtie R_2) \bowtie R_3) \bowtie R_4$ vs. $(R_1 \bowtie R_2) \bowtie (R_3 \bowtie R_4)$



$R_1 \bowtie R_2 = 100$

$R_2 \bowtie R_3 = 200$

$R_3 \bowtie R_4 = 500$

Cout = 109

- $(R_1 \bowtie R_2) \bowtie R_3 < R_1 \bowtie (R_2 \bowtie R_3)$

0.01×0.2×100×40

cardinalities statistics

	1	2	3	4
1	100	100	8	1
2		25	200	25
3			40	500
4				100

costs

	1	2	3	4
1	0	100	108	109
2		0	200	225
3			0	500
4				0

split points t

	1	2	3	4
1		1	2	3
2			2	3
3				3
4				

Homework 11.

3. [yao]

$m = \#pages = 3$ $\quad n = \#tuples = 2$

$$\bar{Y}_n^{N,m} = \bar{Y}_2^{6,3} = 3 \times \bar{Y}_2^{6}(2) = 3 \times (1-p) = 3 \times \frac{3}{5} = 1.8$$

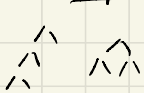
$N-n = 4 \quad K = 2 \Rightarrow K \leq N-n$

- $$[1-p] = 1 - \frac{\binom{N-n}{K}}{\binom{N}{K}} = 1 - \prod_{i=0}^{n-1} \frac{N-K-i}{N-i} = 1 - \prod_{i=0}^{n-1} \frac{4-i}{6-i} = 1 - \left(\frac{4}{6} \times \frac{3}{5}\right) = \frac{3}{5}$$

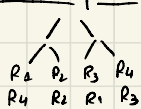
1. [tree count]

shape × order of rec.

ex. trans clique ⇒



C_{n-1}



$n!$ combinations

Horrible!

DPsub

principle of optimality for $S_1 \subseteq R$

R_1, R_2, R_3 for $S_2 \subseteq R \setminus S_1$

S_1	S_2	$R \setminus (S_1 \cup S_2)$	join (S_1, S_2)
1	1	3	
4	1	0	

subtract cases where S_1 is $\{\}$ and S_2 is $\{\}$

$O(3^n)$

	S_1 $\{\}$; S_2 not $\{\}$		S_2 $\{\}$, S_1 not $\{\}$		Both $\{\}$
3^n -	$(2^n - 1)$	-	$(2^n - 1)$	-	1

Number of times join will be called: $3^n - 2^{n+1} + 1$

2. • $\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!(n-(n-k))!} \Leftrightarrow \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!\,k!}$

• $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k} \Leftrightarrow \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!}$

$= \frac{(n-1)!}{(k-1)!(n-1-k)!} \times \left(\frac{n}{k(n-1)}\right) = \frac{n!}{k!(n-k)!}$

Homework 12

1. [cheung]

$m =$ 3 pages, $n =$ 2 tuples per page $k =$ 4 not necessarily dist. tuple

$\overline{\text{Cheung}}_n^{N,m}(k) = m \times \text{Cheung}_n^N(k)$

$\text{Cheung}_n^N(k) = [1 - \tilde{p}]$

$$\tilde{p} = \frac{\binom{N-n+K-1}{K}}{\binom{N+K-1}{K}} = \prod_{i=0}^{K-1} \frac{N-n+i}{N+i} = \prod_{i=0}^{3} \frac{6-2+i}{6+i} = \frac{4}{6} \times \frac{5}{7} \times \frac{6}{8} \times \frac{7}{9}$$

$$= 0.27(7)$$

$$\overline{\text{Cheung}}_{n}^{N,m}(K) = 3 \times [1 - 0.27(7)] = 2.16(6)$$

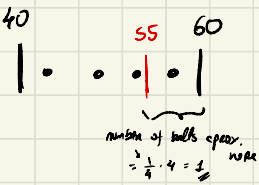
Homework 13

1. [histogram]

$A > c$ $\quad \frac{\sum_{b \in B: c \in b} \frac{\max(b)-c}{\max(b)-\min(b)} H_A(b) + \sum_{b \in B: \min(b) > c} H_A(b)}{\sum_{b \in B} H_A(b)}$

$R.a \geq 55$

$$\frac{60-55}{60-40} \cdot 4 + 2 = \underline{3} \text{ elements}$$



2. [join_histogram]

$A_1 = A_2$ $\quad \frac{\sum_{b_1 \in B_1, b_2 \in B_2, b' = b_1 \cap b_2, b' \neq \emptyset} \frac{\max(b')-\min(b')}{\max(b_1)-\min(b_1)} H_{A_1}(b_1) \frac{\max(b')-\min(b')}{\max(b_2)-\min(b_2)} H_{A_2}(b_2)}{\sum_{b_1 \in B_1} H_{A_1}(b_1) \sum_{b_2 \in B_2} H_{A_2}(b_2)}$

$R.a = S.b$

$$\overbrace{\frac{10-0}{20-0} \cdot 1 \cdot \frac{10-0}{10-0} \cdot 2}^{[0,10) \cap [0,10) \; = 1} + \overbrace{\frac{20-10}{20-0} \cdot 1 \cdot \frac{20-10}{20-10} \cdot 4}^{[0,20] \cap [10,20) \; = 2} + \overbrace{\frac{40-20}{40-20} \cdot 3 \cdot \frac{40-20}{40-20} \cdot 1}^{[20,40) \cap [20,40) \; = 3} +$$

$$\boxed{\frac{\text{intersection interval}}{\text{partition interval}} \times \text{elem}}$$

$$+ \frac{50-40}{60-40} \cdot 4 \cdot \frac{50-40}{50-40} \cdot \overset{=12}{6} + \frac{60-50}{60-40} \cdot 4 \cdot \frac{60-50}{100-50} \cdot \overset{=1.6}{4} +$$

$$+ \frac{80-60}{80-60} \cdot 2 \cdot \frac{80-60}{100-50} \cdot \overset{=3.2}{4} + \frac{100-80}{100-80} \cdot 0 \cdot \frac{100-80}{100-50} \cdot \overset{=0}{4} = 22.8$$