Solutions for Exercise Sheet 9

Richter, Yannick
MTK 03741982
ge78tup@mytum.de
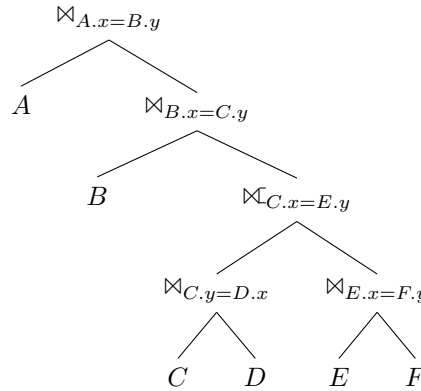
Rodrigues, Diogo
MTK 03770446
diogo.rodrigues@tum.de

TUM – Query Optimization 2022/23
23rd December 2022

Our solutions for Exercise Sheet 9.
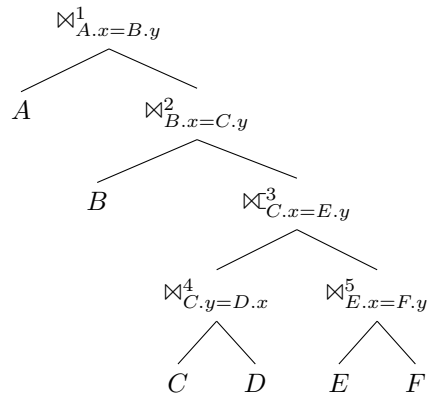
Exercise 1

Given the following join tree, give the reordering restrictions for the individual joins and show the resulting query graph for DPhyp (attention: right outer join!):

Figure 1: $(R \circ_1 S) \circ_2 T \equiv^? R \circ_1 (S \circ_2 T)$. Stated otherwise: is there a conflict if $\circ_2$ is above $\circ_1$? To check if this table's rows/columns are correct (i.e., if I haven't swapped $\circ_1$ with $\circ_2$ by mistake), consider the following statements: $(R \rhd_p S) \bowtie_q T \not\equiv R \rhd_p (S \bowtie_q T)$, $(R \bowtie_p S) \rhd_q T \equiv R \bowtie_p (S \rhd_q T)$

We will consider the following joins with these IDs (just to make it easier to write the table):

$$\bowtie^1_{A.x=B.y}$$

$$A \qquad \bowtie^2_{B.x=C.y}$$

$$B \qquad \bowtie^3_{C.x=E.y}$$

$$\bowtie^4_{C.y=D.x} \qquad \bowtie^5_{E.x=F.y}$$

$$C \quad D \qquad E \quad F$$

	SES	TES
1	$\{A,B\}$	$\{A,B\}$
2	$\{B,C\}$	$\{B,C\}$
3	$\{C,E\}$	$\{C,D,E\}$
4	$\{C,D\}$	$\{C,D\}$
5	$\{E,F\}$	$\{E,F\}$

The reordering restrictions that cannot be violated are:

- 3 after 4

Original query graph:

$$A \xrightarrow{\quad 1 \quad} B \xrightarrow{\quad 2 \quad} C$$

with edge 3 from E to C, edge 4 from C to D, edge 5 from F to E.

Hypergraph:

$$A \xrightarrow{\quad 1 \quad} B \xrightarrow{\quad 2 \quad} C$$

with edge 3, edge 4, edge 5.
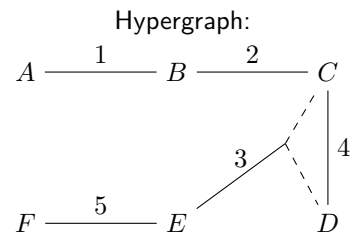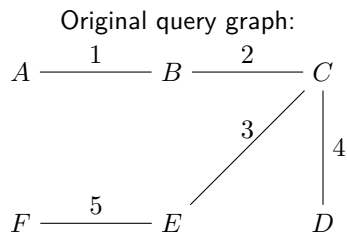
Exercise 2

Consider the following query graph with selectivities and cardinalities and the C_{out} cost function:

$$R_1(10)$$

$1/10$

$$R_0(20)$$

$1/5$

$$R_2(50)$$

$1/5$ $1/5$

$$R_5(55) \qquad\qquad R_3(40) \xrightarrow{3/10} R_4(50)$$
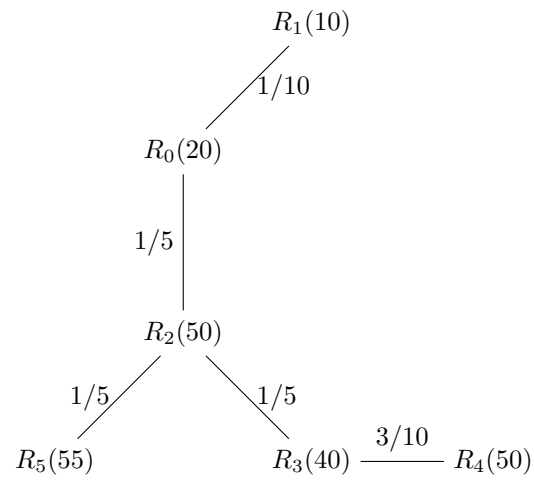
- linearize the search space for the given query graph using R_4 as the start relation
- perform linearized DP on the resulting linearized search space
- perform a full DP (any bushy variant) on the original query graph
- compare the result of linearized DP with the result of the full DP

Linearize search space:

-
```
        R_4
         |
        R_3
         |
        R_2
        / \
      R_0   R_5
       |
      R_1
```
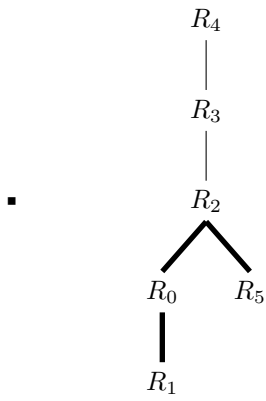
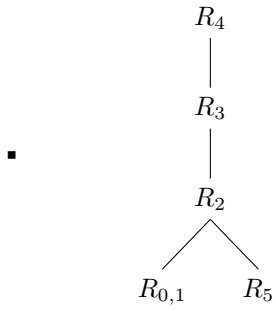|	n	s	C	T	$rank$
R_0	20	1/5	4	4	$3/4 = 0.75$
R_1	10	1/10	1	1	$0/1 = 0$
R_2	50	1/5	10	10	$9/10 = 0.9$
R_3	40	3/10	12	12	$11/12 \approx 0.917$
R_5	55	1/5	11	11	$10/11 \approx 0.909$

-
```
        R_4
         |
        R_3
         |
        R_2
        / \
    R_0,1   R_5
```

	n	s	C	T	$rank$
$R_{0,1}$	20	1/5	8	4	$3/8 = 0.375$
R_2	50	1/5	10	10	$9/10 = 0.9$
R_3	40	3/10	12	12	$11/12 \approx 0.917$
R_5	55	1/5	11	11	$10/11 \approx 0.909$

-
```
        R_4
         |
        R_3
         |
        R_2
         |
        R_0
         |
        R_1
         |
        R_5
```
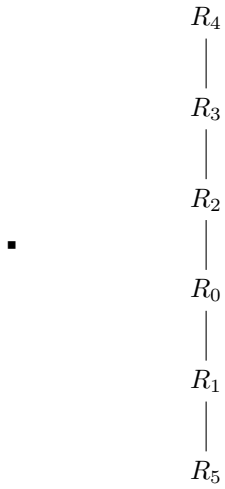
Linearized DP

$$R_4 \text{———} R_3 \text{———} R_2 \text{———} R_0 \text{———} R_1 \text{———} R_5$$

All steps				
N	$Pair_1$	$Pair_2$	$|R|$	C_{out}
R_4R_3	R_4	R_3	600	600
R_3R_2	R_3	R_2	400	400
R_2R_0	R_2	R_0	200	200
R_0R_1	R_0	R_1	20	20
R_1R_5	R_1	R_5	—	—
$R_4R_3R_2$	R_4	R_3R_2	6000	6400
$R_4R_3R_2$	R_4R_3	R_2	6000	6600
$R_3R_2R_0$	R_3	R_2R_0	1600	1800
$R_3R_2R_0$	R_3R_2	R_0	1600	2000
$R_2R_0R_1$	R_2	R_0R_1	200	220
$R_2R_0R_1$	R_2R_0	R_1	200	400
$R_0R_1R_5$	R_0	R_1R_5	—	—
$R_0R_1R_5$	R_0R_1	R_5	—	—
$R_4R_3R_2R_0$	R_4	$R_3R_2R_0$	24000	25800
$R_4R_3R_2R_0$	R_4R_3	R_2R_0	24000	24800
$R_4R_3R_2R_0$	$R_4R_3R_2$	R_0	24000	30400
$R_3R_2R_0R_1$	R_3	$R_2R_0R_1$	1600	1820
$R_3R_2R_0R_1$	R_3R_2	R_0R_1	1600	2200
$R_3R_2R_0R_1$	$R_3R_2R_0$	R_1	1600	3400
$R_2R_0R_1R_5$	R_2	$R_0R_1R_5$	—	—
$R_2R_0R_1R_5$	R_2R_0	R_1R_5	—	—
$R_2R_0R_1R_5$	$R_2R_0R_1$	R_5	2200	2420
$R_4R_3R_2R_0R_1$	R_4	$R_3R_2R_0R_1$	24000	25820
$R_4R_3R_2R_0R_1$	R_4R_3	$R_2R_0R_1$	24000	24820
$R_4R_3R_2R_0R_1$	$R_4R_3R_2$	R_0R_1	24000	30420
$R_4R_3R_2R_0R_1$	$R_4R_3R_2R_0$	R_1	24000	48800
$R_3R_2R_0R_1R_5$	R_3	$R_2R_0R_1R_5$	17600	20020
$R_3R_2R_0R_1R_5$	R_3R_2	$R_0R_1R_5$	—	—
$R_3R_2R_0R_1R_5$	$R_3R_2R_0$	R_1R_5	—	—
$R_3R_2R_0R_1R_5$	$R_3R_2R_0R_1$	R_5	17600	19420
$R_4R_3R_2R_0R_1R_5$	R_4	$R_3R_2R_0R_1R_5$	264000	283420
$R_4R_3R_2R_0R_1R_5$	R_4R_3	$R_2R_0R_1R_5$	264000	267020
$R_4R_3R_2R_0R_1R_5$	$R_4R_3R_2$	$R_0R_1R_5$	—	—
$R_4R_3R_2R_0R_1R_5$	$R_4R_3R_2R_0$	R_1R_5	—	—
$R_4R_3R_2R_0R_1R_5$	$R_4R_3R_2R_0R_1$	R_5	264000	288820

Final Table				
N	$Pair_1$	$Pair_2$	$|R|$	C_{out}
R_4R_3	R_4	R_3	600	600
R_3R_2	R_3	R_2	400	400
R_2R_0	R_2	R_0	200	200
R_0R_1	R_0	R_1	20	20
R_1R_5	R_1	R_5	—	—
$R_4R_3R_2$	R_4	R_3R_2	6000	6400
$R_3R_2R_0$	R_3	R_2R_0	1600	1800
$R_2R_0R_1$	R_2	R_0R_1	200	220
$R_0R_1R_5$	R_0	R_1R_5	—	—
$R_4R_3R_2R_0$	R_4R_3	R_2R_0	24000	24800
$R_3R_2R_0R_1$	R_3	$R_2R_0R_1$	1600	1820
$R_2R_0R_1R_5$	$R_2R_0R_1$	R_5	2200	2420
$R_4R_3R_2R_0R_1$	R_4R_3	$R_2R_0R_1$	24000	24820
$R_3R_2R_0R_1R_5$	$R_3R_2R_0R_1$	R_5	17600	19420
$R_4R_3R_2R_0R_1R_5$	R_4R_3	$R_2R_0R_1R_5$	264000	267020

Full DP:

N	$|R_N|$	Cost	Join
000001	20	0	R_0
000010	10	0	R_1
000011	20	20	$R_0 \bowtie R_1$
000100	50	0	R_2
000101	200	200	$R_0 \bowtie R_2$
000110	—	—	—
000111	200	220	$(R_0 \bowtie R_1) \bowtie R_2$
001000	40	0	R_3
001001	—	—	—
001010	—	—	—
001011	—	—	—
001100	400	400	$R_2 \bowtie R_3$
001101	1600	1800	$(R_0 \bowtie R_2) \bowtie R_3$
001110	—	—	—
001111	1600	1820	$((R_0 \bowtie R_1) \bowtie R_2) \bowtie R_3$
010000	50	0	R_4
010001	—	—	—
010010	—	—	—
010011	—	—	—
010100	—	—	—
010101	—	—	—
010110	—	—	—
010111	—	—	—
011000	600	600	$R_3 \bowtie R_4$
011001	—	—	—
011010	—	—	—
011011	—	—	—
011100	6000	6400	$(R_2 \bowtie R_3) \bowtie R_4$
011101	24000	24800	$(R_0 \bowtie R_2) \bowtie (R_3 \bowtie R_4)$
011110	—	—	—
011111	24000	24820	$((R_0 \bowtie R_1) \bowtie R_2) \bowtie (R_3 \bowtie R_4)$
100000	55	0	R_5
100001	—	—	—
100010	—	—	—
100011	—	—	—
100100	550	550	$R_2 \bowtie R_5$
100101	2200	2400	$(R_0 \bowtie R_2) \bowtie R_5$
100110	—	—	—
100111	2200	2420	$((R_0 \bowtie R_1) \bowtie R_2) \bowtie R_5$
101000	—	—	—
101001	—	—	—
101010	—	—	—
101011	—	—	—
101100	4400	4800	$(R_2 \bowtie R_3) \bowtie R_5$
101101	17600	19400	$(R_0 \bowtie R_2) \bowtie R_3) \bowtie R_5$
101110	—	—	—
101111	17600	19420	$(((R_0 \bowtie R_1) \bowtie R_2) \bowtie R_3) \bowtie R_5$
110000	—	—	—
111011	—	—	—
111100	66000	67100	$(R_3 \bowtie R_4) \bowtie (R_2 \bowtie R_5)$
111101	264000	267000	$(R_3 \bowtie R_4) \bowtie ((R_0 \bowtie R2) \bowtie R_5)$
111110	—	—	—
111111	264000	267020	$(R_3 \bowtie R_4) \bowtie (((R_0 \bowtie R1) \bowtie R_2) \bowtie R_5)$

FullDP

$\bowtie_{264000}^{267020}$

$\bowtie_{2200}^{2420}$ $\bowtie_{600}^{600}$

$\bowtie_{200}^{220}$ R_5 R_3 R_4

$\bowtie_{20}^{20}$ R_2

R_0 R_1

In this case linearizedDP find the optimal solution

Lin DP

$\bowtie_{264000}^{267020}$

$\bowtie_{2200}^{2420}$ $\bowtie_{600}^{600}$

$\bowtie_{200}^{220}$ R_5 R_3 R_4

$\bowtie_{20}^{20}$ R_2

R_0 R_1