CS 784 Homework 2 Tien-Lung Fu

A.

1.

(a) $q(x,y,z) :- R(x), T(y), U(z), S(x,y,z)$

the Fractional edge covers: $(u_R, u_T, u_U, u_S) = (0,0,0,1)$

the Output size upper bound is $N^0 \cdot N^0 \cdot N^0 \cdot N^1 = N$

(b) $q(x,y,z,w,t) :- R(x,y), S(y,z), T(z,w), U(w,x), V(x,t)$

Fractional edge covers: $(u_R, u_S, u_T, u_U, u_V) = (0,1,0,1,1)$

Output size upper bound is $N^0 \cdot N^1 \cdot N^0 \cdot N^1 \cdot N^1 = N^3$

(c) $q(x,y,z,w) :- R(x,y,z), S(x,z,w), T(x,y,w), U(y,z,w)$

Fractional edge covers: $(u_R, u_S, u_T, u_U) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

Output size upper bound is $(N^{\frac{1}{3}})^4 = N^{\frac{4}{3}}$

2.

Given $k \in \mathbb{R}$ and $0 \le k \le 1$, the fractional edge covers could be

$(u_{R_1}, u_{R_2}, u_{R_3}, u_{R_4}) = (1, k, 1-k, 1)$

Then the maximum output size is $N_1 \cdot N_4 \cdot N_2^{k} \cdot N_3^{1-k}$

1°. if $N_2 \ge N_3$, choose $k=1$, the answer is $\underline{N_1 \cdot N_4 \cdot N_2}$

2°. if $N_3 > N_2$, choose $k=0$, the answer is $\underline{N_1 \cdot N_4 \cdot N_3}$

B.

1. Yes, it is equivalent to a UCQ query

$q_B = L(x,Y) \cup \pi_{x,Y}(T(x) \bowtie B(z,Y))$

2.

(a) The datalog program P that performs transitive closure could be as:

$T(x,y) :- R(x,y)$ —— (1)

$T(x,y) :- T(x,w), R(w,y)$ —— (2)

In (2), each iteration is to find reachable vertices of next level.
Given n the number of vertices, the longest path to reach from an initial node should be n, and the step number of P is also proportional to n. Thus, P is not bounded.

(b) Below Datalog program ends in O(1) step.

$T(y,z) :- R(y,z)$

$T(y,z) :- T(y,z)$

4. $T^{bf}(x,y) :- [sup_0^1(x)]\ F(x,y)\ [sup_1^1(x,y)]$

$T^{bf}(x,y) :- [sup_0^2(x)]\ up(x,z_1), [sup_1^2(x,z_1)]\ T^{bf}(z_1,z_2), [sup_2^2(x,z_2)]\ F(z_2,z_3), [sup_3^2(x,z_3)]\ T^{bf}(z_3,z_4), [sup_4^2(x,z_4)]\ down(z_4,y)\ [sup_5^2(x,y)]$

$g(y) :- T^{bf}(a,y)$

⇓

$sup_0^1(x) :- in\_T^{bf}(x)$

$sup_1^1(x,y) :- sup_0^1(x), F(x,y)$

$T^{bf}(x,y) :- sup_1^1(x,y)$

$sup_0^2(x) :- in\_T^{bf}(x)$

$sup_1^2(x,z_1) :- sup_0^2(x), up(x,z_1)$

$sup_2^2(x,z_2) :- sup_1^2(x,z_1), T^{bf}(z_1,z_2)$

$sup_3^2(x,z_3) :- sup_2^2(x,z_2), F(z_2,z_3)$

$sup_4^2(x,z_4) :- sup_3^2(x,z_3), T^{bf}(z_3,z_4)$

$sup_5^2(x,y) :- sup_4^2(x,z_4), down(z_4,y)$

$T^{bf}(x,y) :- sup_2^2(x,y)$

initialize the input relations:

$in\_T^{bf}(x) :- T^{bf}(x,y)$

$in\_T^{bf}(a) :-$

$g(y) :- T^{bf}(a,y)$

5. There is only one stratification which partitions the Datalog program into P1, P2, P3 as:

P1: $\begin{cases} T(x) :- S(x), \neg R(x) \\ S(x) :- T(x), \neg R(x) \end{cases}$

P2: $U(x) :- R(x), \neg T(x), \neg S(x)$

P3: $V(x,y) :- V(x,y), \neg U(x), \neg U(y)$