CMPSCI 645: Homework 2

Due: Monday, April 6 2015, 11:59pm

Name: Patrick Pegus

Question	Points	Score
1	30	
2	35	
3	35	
Total:	100	

Please turn in this homework electronically, as a PDF, through Moodle. You may handwrite your solutions, and then scan the document, or type directly into the PDF form.

1. **Normalization** [30 points]

Consider the following relational schema and set of functional dependencies.

$$R(A,B,C,D,E) \qquad \begin{array}{l} CD \rightarrow E \\ A \rightarrow B \end{array}$$

(a) [10 points] List **all** superkey(s) for this relation. Justify your answer in terms of functional dependencies and closures.

(A, B, C, D, E)

(A, B, C, D) since $(A, B, C, D)^+ = (A, B, C, D, E)$ given $CD \rightarrow E$

(A, C, D, E) since $(A, C, D, E)^+ = (A, B, C, D, E)$ given $A \rightarrow B$

(A, C, D) since $[(A, C, D)^+ = (A, C, D, E)$ given $CD \rightarrow E]$ and (A, C, D, E) is a superkey

(b) [5 points] Which of these superkeys form a key (i.e., a minimal superkey) for this relation?

(A, C, D)

(c) [15 points] Decompose R into BCNF. Show your work for partial credit. Your answer should consist of a list of table names and attributes and an indication of the keys in each table (either underline the corresponding attributes, or explicitly state the keys).

$R(\underline{A}, B, \underline{C}, \underline{D}, E)$

$A \neq A^+ = (A, B) \neq (A, B, C, D, E)$

$\rightarrow R_1(\underline{A}, B)$

$\rightarrow R_2(\underline{A}, \underline{C}, \underline{D}, E)$

$(C,D) \neq (C, D)^+ = (C, D, E) \neq (A, C, D, E)$

$\rightarrow R_{21}(\underline{C}, \underline{D}, E)$

$\rightarrow R_{22}(\underline{A}, \underline{C}, \underline{D})$

R decomposed into BCNF is R_1, R_{21}, R_{22}

2. **Datalog** [35 points]

   (a) [15 points] Consider the following two datalog programs computing the transitive closure:

P1:

```
T(x,y) :- R(x,y)
T(x,y) :- T(x,z),R(z,y)
```

P2:

```
T(x,y) :- R(x,y)
T(x,y) :- T(x,z),T(z,y)
```

Suppose R is a graph that consists of a single path: $R(a_0, a_1), R(a_1, a_2), \ldots, R(a_{n-1}, a_n)$. Thus, the transitive closure T computed by both programs consists of all $\binom{n}{2}$ ground facts of the form $T(a_i, a_j)$, for $1 \le i < j \le n$. Assume that we evaluate both programs using the semi-naive evaluation algorithm.

i. For a fixed $m = 1, \ldots, n-1$, how many times will the fact $T(a_1, a_{m+1})$ be discovered by **P1**?

i. 1

Explain your answer:

To discover path $T(a_1, a_{m+1})$, P1 must begin at edge $R(a_1, a_2)$ and, in general, can only expand the path $T(a_1, a_i)$ computed in the previous round by adding edge $R(a_i, a_{i+1})$.

ii. How many times will the fact $T(a_1, a_{m+1})$ be discovered by **P2**?

ii. 1 if $m=1$

$m-1$ otherwise

Explain your answer:

P2 will discover path $T(a_1, a_{m+1})$ in the following $m-1$ ways:

$T(a_1, a_2)\ T(a_2, a_{m+1})$

$T(a_1, a_3)\ T(a_3, a_{m+1})$

$\vdots$

$T(a_1, a_{m-1})\ T(a_{m-1}, a_{m+1})$

$T(a_1, a_m)\ T(a_m, a_{m+1})$

(b) [20 points] Consider a graph where each node x is either a leaf, or has two outgoing edges (x•y)•(x•z). In the former case, we store x in a relation L(x); in the latter case we store the triple (x•y•z) in a relation T. (Thus, x is a key in T(x•y•z).) Consider the following game with two players. Players take turns in moving a pebble on the graph. If the pebble is on a node x, then the player whose turn it is may move it to one of the two children, y or z. A player wins if it is her turn to move and the pebble is on a leaf. Write a datalog program that computes the set of starting nodes from which player 1 has a winning strategy. That is, your program should compute a relation P1(x) that returns all nodes x such that, if player 1 starts the game on x (and plays smartly!) then she is guaranteed to win the game.

P1(x) :- L(x)

P1(x) :- T(x,i,j), T(i,k,l), T(j,m,n), W(k), W(m)

3. **Conjunctive Queries** [35 points]

(a) [5 points] Find a full semi-join reduction for the query below.

$$q(x) : -R(x \cdot y) \cdot S(y \cdot z) \cdot T(y \cdot u)$$

$R'(x,y) = R(x,y) \ltimes S(y,z)$

$S'(y,z) = S(y,z) \ltimes R'(x,y)$

$T'(y,u) = T(y,u) \ltimes R'(x,y)$

$T^2(y,u) = T'(y,u) \ltimes S'(y,z)$

$R^2(x,y) = R'(x,y) \ltimes T^2(y,u)$

$S^2(y,z) = S'(y,z) \ltimes T^2(y,u)$

$q(x) :- R^2(x,y), S^2(y,z), T^2(y,u)$

(b) i. [20 points] Indicate for each pair of queries q, q' below, whether $q \subseteq q'$. If the answer is yes, provide a proof; if the answer is no, give a database instance I on which $q(I) \not\subseteq q'(I)$.

α)

$$q(x) \ :- \ R(x,y),R(y,z),R(z,x)$$
$$q'(x) \ :- \ R(x,y),R(y,z),R(z,u),R(u,v),R(v,z)$$

α) $q \subseteq q'$

Explain your answer:

Let h be a function such that $h(x_{q'}) = x_q$, $h(y_{q'}) = y_q$, $h(z_{q'}) = z_q$, $h(u_{q'}) = x_q$, and $h(v_{q'}) = y_q$. Then

$h(body(q')) = R(x,y), R(y,z), R(z,x), R(x,y), R(y,z) \subseteq body(q)$ and

$h(tuple(q')) = x = tuple(q)$. Therefore, $h: q' \to q$ is a homomorphism and its existence proves $q \subseteq q'$.

β)

$$q(x,y) \ :- \ R(x,u,u),R(u,v,w),R(w,w,y)$$
$$q'(x,y) \ :- \ R(x,u,v),R(v,v,v),R(v,w,y)$$

β) $q \not\subseteq q'$

Explain your answer:

Let I =

1	2	2
2	3	4
4	4	5

. Then $(1,5) \subseteq q(I)$, but since no tuple in I has all the same values, $R(v,v,v)$ cannot be satisfied. Therefore, $q'(I) = \emptyset$ and $q(I) \not\subseteq q'(I)$.

ii. [10 points] Consider the two conjunctive queries below, and notice that $q_1 \subset q_2$.

$$q_1(x) = R(x,y), R(y,z)$$
$$q_2(x) = R(x,y)$$

1. Find a conjunctive query $r(x)$ s.t. $q_1 \subset r \subset q_2$

$r(x) = R(x,y), R(b,z), R(a,b)$

- q_1 is a subset of r:
  Let $h(x_r) = x_{q_1}$, $h(y_r) = y_{q_1}$, $h(b_r) = y_{q_1}$, $h(z_r) = z_{q_1}$, and $h(a_r) = x_{q_1}$. Then $h(body(r) = R(x,y), R(y,z), R(x,y) \subseteq body(q_1)$ and $h(tuple(r)) = (x) = tuple(q_1)$

- r is not a subset of q_1:
  Let I =

1	2
3	4
5	3

. Then $q_1(I) = \{(3)\}$, but $r(I) = \{(1),(3),(5)\}$.

- r is a subset of q_2:
  Let $h(x_{q_2}) = x_r$ and $h(y_{q_2}) = y_r$. Then $h(body(q_2)) = R(x,y) \subseteq body(r)$ and $h(tuple(q_2)) = (x) = tuple(r)$

- q_2 is not a subset of r:
  Let I =

1	2

. Then $r(I) = \emptyset$, but $q_2(I) = \{(1)\}$