Homework 1

Due on October 6

A: QUERY CONTAINMENT **[50%]**

1. **[9%]** For each of the following pair of queries q, q' decide whether $q \subseteq q'$, $q' \subseteq q$, both or none. Explain your answer.

   (a) $q(x,y) : -R(x,z), S(z,z), R(z,y)$
   $q'(y,w) : -R(y,z), S(z,t), R(t,w)$

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

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

2. **[6%]** Find the minimal equivalent CQ to $S_k(x) : -R(x,z_1), R(x,z_2), \ldots, R(x,z_k)$ for any $k \geq 1$. Explain your answer.

3. **[10%]** Consider the following two CQs: $q_1(x) : -R(x,y), R(y,z)$ and $q_2(x) : -R(x,y)$. The notation $q \subsetneq q'$ means that $q \subseteq q'$, but that q is not equivalent to q'. Does there exist a conjunctive query q such that $q_1 \subsetneq q \subsetneq q_2$? Explain your answer.

4. **[10%]** Consider the following two queries in $\text{CQ}^<$:

$$q() : -R(x,10), x < 10$$
$$q'() : -R(y,10), R(9,y), y \leq 10$$

   Decide whether $q \subseteq q'$, $q' \subseteq q$, both or none. Explain your answer.

5. **[15%]** We say that a conjunctive query q has a *self join* if its body contains two atoms of the same relation. For example, the query $q() : -R(x,y), R(y,z)$ has a self-join, but the query $q'(x) : -R(x,y), S(y,z,w)$ does not. What is the complexity for query containment of CQs that have no self join? You should either provide a polynomial time algorithm or show that the problem is NP-complete.

B: QUERY COMPLEXITY **[50%]**

1. **[15%]** We have shown that the combined complexity of query evaluation for CQs is NP-complete. However, the reduction uses queries where the relations are binary (have arity 2).

   (a) What is the combined complexity of CQ evaluation if we restrict the relations to be *unary* (i.e. have arity exactly 1)? For example, such a query would be $q(x) = R(x), S(x), T(y)$.

   (b) What is the combined complexity of $\text{CQ}^{\neq}$ evaluation (we can add disequalities) with relations that are unary? For example, such a query would be $q() = R(x), S(y), T(z), x \neq y, y \neq z$.

2. **[15%]** For the following conjunctive queries, determine if they are acyclic or not. If the query is acyclic, construct the join tree of the query as well.

   (a) $q_1() : -R(x, t_1, y), S(y, t_2, z), T(z, t_3, x), U(x, y, z)$

   (b) $q_2() : -R(x, y, z), S(y, z, w), T(x, y, w), U(x, z, w)$

   (c) $q_3() : -R(x, y, z), S(x, y), T(y, z), U(z, x)$

3. **[10%]** For the following CQs, construct a GHD with the smallest possible ghw.

   (a) The k-th spider-web query W_k:

   $$W_k(x) : -R(x, z_1), R(x, z_2), \ldots R(x, z_k), R(z_1, z_2), R(z_2, z_3), \ldots, R(z_k, z_1)$$

   (b) $V_k() : -R(x_1, x_2), R(x_2, x_3), \ldots, R(x_k, x_1), R(x_1, z_2), R(z_2, z_3), \ldots, R(z_k, x_1)$, for any $k \geq 1$. This query can be drawn as two cycles of length k connected at variable x_1.

4. **[10%]** Among all CQs with n atoms in the body, what is the maximum possible ghw as a function of n? Describe a class of queries that achieves this ghw.

DELIVERABLES

Submit a single PDF document using `Canvas` (Homework 1). It is strongly suggested to use LaTeX to write your solution.