Pan Wu CS 784 Homework 1

A: Query Containment

1. (a). $q \subseteq q'$. Because there is a homomorphism h from q' to q,
   h from q' to q: $(y \to x, z \to z, t \to z, w \to y)$
   But there is not homomorphism from q to q'.

   (b) $q \subseteq q'$. Because there is a homomorphism h from q' to q,
   h from q' to q: $(x \to x, y \to y, z \to z, u \to x, v \to y)$
   But there is no homomorphism from q to q'.

   (c) None. Because there is no homomorphism from q' to q,
   And there is no homomorphism from q to q'.

According to Homomorphism Theorem.

2. $q(x) :- R(x, z_1)$

Firstly: $q \subseteq S_k$, because there is a homomorphism h from S_k to q,
h from S_k to q: $(x \to x, z_1, z_2, \cdots, z_k \to z_1)$
Thus $q \subseteq S_k$ according to Homomorphism Theorem.

Secondly: $S_k \subseteq q$, because there is a homomorphism h from q to S_k,
h from q to S_k: $(x \to x, z_1 \to z_1)$
Thus $S_k \subseteq q$ according to Homomorphism Theorem.

Thus $S_k \equiv q$, and obviously, q is the minimal CQ which is equivalent to S_k, since q only has one atom.

3. There is no such q.

Because if $q_1 \subsetneq q \subsetneq q_2$, then there is a homomorphism h from q_2 to q_1, ~~But q(x) :- R(x,y) R(z,z)~~ and q is not equivalent to q_2, so q must be like $q(x) :- R(x,y)\, R(\ \ )\cdots$, there are 3 cases :

1° If $q(x) :- R(x,y)\, R(z)\cdots$, then there is no homomorphism from q to q_1,

2° If $q(x) :- R(x,y)\, R(z,z)\cdots$, then there is no homomorphism from q to q_1,

3° If $q(x) :- R(x,y)\, R(z,t)\cdots$, then there is a homomorphism from q_1 to q which means $q \subseteq q_1$, which is not allowed.

Thus there is no such q, that will lead to $q_1 \subsetneq q \subsetneq q_2$.

4. Let $q_1 = q = R(x,10),\ x<10$ Then $\underline{q = q_1 \cup q_2}$

$q_2 = q = R(x,10),\ x<10$

Let $q_1' = R(y,10), R(9,y), y<10$

$q_2' = R(y,10), R(9,y), y=10$ Then $\underline{q' = q_1' \cup q_2'}$

$q_1' \subseteq q_1$, since there is a homomorphism from q_1 to q_1'

h from q_1 to q_1' : $(x \to y)$

$q_2' \subseteq q_2$, since there is a homomorphism from q_2 to q_2'

h from q_2 to q_2' : $(x \to 9)$

Because $q_1' \subseteq q_1$, $q_2' \subseteq q_2$, $q = q_1 \cup q_2$, $q' = q_1' \cup q_2'$

So $q' \subseteq q$.

5. To determine whether q' contain q, ~~we just~~ if and only if there exises a homomorphism from q' to q. To find the homomorphism, we can check the variables in q' one by one and define a h to map the corresponding variable in q' to the one in q. If q ~~has~~ has no self join, the time is linear to the number of variables in q and q'. After this process, if we end with a homomorphism, then $q \subseteq q'$, otherwise $q \subsetneq q'$. Thus the algorithm is in polynomial time.

B. Query Complexity

1. (a). $O(|q| \cdot |I|)$ $|q|$ is the size of query, $|I|$ is size of dataset.
Because in this case, we can join the variables one by one, and for each variable, we join all the corresponding atoms one by one.

(b) NP-complete . Because there is a reduction to 3-coloring problem.
For an undirected graph $G = (V, E)$, for each vertex $v_i \in V$, has an atom $R_i(v_i)$, and the table of R_i contains 3 possible values 1, 2, 3. For each edge (v_i, v_j), $v_i \neq v_j$. Then this problem become solving the 3-coloring problem. Thus the combined complexity is NP-complete.

2. (a) According to GYO algorithm, Yes, q_1 is acyclic.

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

Since $R(x, t_1, y)$ is ear, remove t_1 and remove R,
Then $S(y, t_2, z)$ is ear, remove t_2 and remove S,
Then $T(z, t_3, x)$ is ear, remove t_3 and remove T,
Lastly, $U(x, y, z)$ is ear, remove x, y, z and remove U,
H is the empty hypergraph. Thus q_1 is acyclic and the join tree of q_1 is

(b). No. Since R, S, T, U are all not ear, we can not remove any one of them. Thus q_2 is not acyclic.

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

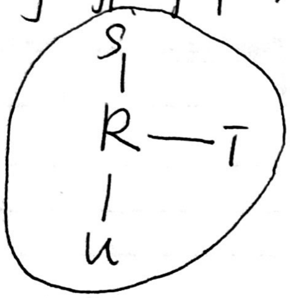
Since $S(x,y)$ is ear, remove S

Then $T(y,z)$ is ear, remove T

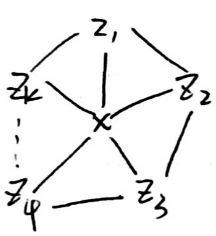
Then $U(z,x)$ is ear, remove U

At last, $R(x,y,z)$ is ear, remove x, y, z and R.

H is the empty hypergraph, Thus q_3 is acyclic and the join tree of q_3 is



3. (a)



GHD with the smallest possible ghw:

$\chi(v_i) = \{x, z_1, z_i, z_{i+1}\}$ $(i = 1, 2, \cdots, k)$

ghw = 3

(b). GHD with smallest possible ghw:

$\chi(v_i) = \{x_1, x_i, x_{i+1}\}$ or $\{x_1, z_i, z_{i+1}\}$ $(i = 1, 2, \cdots, k)$.

ghw = 2

4. the maximum possible ghw of CQs with n atoms in the body is

$$\frac{1+\sqrt{1+8n}}{2} - 1$$

The class of queries achieves this ghw is a complete connected graph, for that graph, it has $\frac{1+\sqrt{1+8n}}{2}$ vertices, and each couple of vertices is connected with each other.