Part 1: Data Exchange

Question 1 (Core Computation)

A coloring of a graph is an assignment of colors to the vertices, such that no adjacent nodes are assigned the same color. A graph is $k$-colorable if we can color it using a collection of (at most) $k$ colors. It is known that determining whether a given graph is $k$-colorable is NP-complete for every $k > 2$. Consider the schema $S$ that has a single relation symbol $R$ and no constraints. Using a reduction from 3-colorability, show that one cannot find the core of a $v$-instance over $S$, unless P = NP.

Question 2 (Core of Solutions)

As in the lecture, we restrict our attention to data exchange settings where the constraints belong to the classes of st-TGDs, t-TGDs, and t-EDGs.

Definitions. Let $J$ be a $v$-instance, and let $\mu$ be an endomorphism over $J$. Denote by $\mu(J)$ the image of $J$ under $\mu$, that is, the sub-instance of $J$ that is obtained from $J$ by replacing every variable $x$ with $\mu(x)$. Let us say that $\mu$ is a core endomorphism if $\mu(J)$ is a core of $J$. For a natural number $m$, we denote by $\mu^m$ the composition of $\mu$ with itself $m$ times; that is:

$$
\mu^m(x) \overset{\text{def}}{=} \begin{cases} 
\mu(x) & \text{if } m = 1; \\
\mu(\mu^{m-1}(x)) & \text{if } m > 1.
\end{cases}
$$

A fixed point of $\mu$ is a value $x$ such that $\mu(x) = x$. Finally, we will say that $\mu$ is image preserving if every value in the image of $\mu$ is a fixed point (that is, $\mu$ is the identity function over $\mu(J)$).

Part A. In the lecture, we have seen an example of a solution $J$ and an endomorphism $\mu$ over $J$, such that $\mu(J)$ is not a solution (Lecture 4, page 35). Explain why $\mu(J)$ is indeed not a solution.

In Part B, you will prove the following theorem, originally established by Fagin et al. [1].

Theorem 1 Let $(S, T, \Sigma)$ be a schema mapping, let $I$ be a source instance, and let $J$ be target instance. If $J$ is a solution, then every core of $J$ is a solution.

Part B. Prove the following.

1. If $\mu$ is a core endomorphism, then $\mu^m$ is a core endomorphism for all $m > 1$.

2. If $\mu$ is a core endomorphism, then $\mu$ is an isomorphism over $\mu(J)$.

3. Suppose that (1) $\mu$ is a core endomorphism, and (2) $\mu$ is not image preserving. There exists some $m > 0$ such that $\mu^m$ has a fixed point that $\mu$ does not have.
4. There exists a core endomorphism over J that is image preserving.

5. If J is a solution and µ is a core endomorphism that is image preserving, then µ(J) is a solution.

6. If K is a core of a solution, then K is a solution.

**Part 2: Inconsistent Databases**

**Question 1 (Trichotomy)**

Consider the following CQ:

\[ R(x,y), S(y,z), T(z,x), U(x,u), V(x,u,v) \]

1. In the attack graph, is there an edge from R to T? Please explain your answer.

2. Which of the three categories of the trichotomy contains this CQ?

(Hint: you can find the answer without instantiating the entire attack graph.)

**Question 2 (Aggregates on Repairs)**

Consider a schema \((S, \Sigma)\) where \(S\) consists of a single relation schema \(R(A,B)\). The following are aggregate queries in SQL that produce a number as a result.

- \(Q_1: \text{SELECT COUNT(\ast) FROM R}\)
- \(Q_2: \text{SELECT COUNT(DISTINCT A) FROM R}\)
- \(Q_3: \text{SELECT COUNT(DISTINCT B) FROM R}\)
- \(Q_4: \text{SELECT MIN(B) FROM R}\)

Given an inconsistent instance \(I\) over \((S, \Sigma)\), we are interested in the extremal (minimal and maximal) values for each query \(Q\):

\[
Q_{\text{min}}(I) \overset{\text{def}}{=} \min \{Q(J) \mid J \in \text{Repairs}_\Sigma(I)\}
\]

\[
Q_{\text{max}}(I) \overset{\text{def}}{=} \max \{Q(J) \mid J \in \text{Repairs}_\Sigma(I)\}
\]

1. Assume that \(\Sigma\) consists of a single FD: \(R: A \rightarrow B\). For each of the four queries \(Q_i\), give an efficient algorithm for, or prove hardness of, computing \(Q_{\text{min}}^i(I)\) and \(Q_{\text{max}}^i(I)\).

2. Do the same with \(\Sigma = \{R_1: A \rightarrow B, R_2: B \rightarrow A\}\).

Hint: find out about the following concepts:

- Maximum network flow
- Maximal matching
- Minimum maximal matching
• Hitting set

Good luck!

References