Assignment 3: Algorithms and Complexity

Due June 9th, 2015

Part 1: Incomplete Databases

Question 1

Background. 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.

Assignment. Consider the schema S that has a single relation schema Follows(person1, person2) and no constraints. Using a reduction from 3-colorability, show that the following tasks cannot be performed in polynomial time, unless P = NP.

1. Given a v-instance I and an ordinary instance J, both over S, determine whether J \in [I].
2. Compute the core of I.

Part 2: Consistent Query Answering

Question 2 (Cardinality Repairs)

Background. An independent set of an (undirected) graph is a set of nodes that does not contain any edge. The following is a well known NP-complete problem, called the independent-set problem: given a graph G and a number m, does G have an independent set of size m (or larger)?

Assignment. Build a schema S = (R, \Sigma) with all of the following properties:

- R contains two relations.
- \Sigma consists of a single denial constraint.
- Repair checking is coNP-complete over S.

Question 3 (p-Repairs)

Background. Recall that x-repair checking (where x \in \{p, g\}) over a schema S is the following problem: Given an inconsistent prioritizing instance (I, >) over S and an instance J over S, determine whether J is an x-repair of (I, >).

Assignment. Let S be the schema that consists of the single relation schema R(A, B, C) and the following FDs:
• $A \rightarrow B$
• $\emptyset \rightarrow C$ (i.e., $C$ should have the same value for all tuples)

In this question the task it to write a SQL query (with [NOT] EXISTS) to solve p-repair checking over $S$. Specifically, assume the following extended schema:

• The above $R(A, B, C)$.
• $P(A, B, C, A', B', C')$ for storing a priority relation over $R(A, B, C)$. For example, the relationship $R(0,0,0) \succ R(1,1,1)$ is represented by $P(0,0,0,1,1,1)$.
• $R_J(A, B, C)$ for storing a candidate p-repair $J$.

Write a SQL query $Q$ over $\{R, P, R_J\}$ that returns a nonempty result (e.g., a single line 'true') if and only if $J$ is a p-repair of $(I, \succ)$.

Question 4 (g-Repairs)

Prove that g-repair checking is coNP-complete over the schema $S$ of the previous question. (Hint: there is a simple reduction from CNF satisfiability.)

Part 3: Probabilistic Databases

Question 5 (CQ Inference)

Background. Let $D$ be a probabilistic database, and let $Q$ be a Boolean query over the signature of $D$. The probability of $Q$ in $D$ is the probability that a random possible world $I$ in $D$ satisfies $Q(I) = \text{true}$.

Assignment. Show a polynomial-time algorithm to compute the probability of the following Boolean CQ in a tuple-independent database.

$$Q() :\neg R(x), S(x, y)$$

Question 6 (Hardness of CQ Inference)

Background. The following counting problem is known to be #P-complete [PB83].

Given a bipartite graph $(U \cup V, E)$, compute the number of subsets $W$ of $U \cup V$ such that $W$ contains at least one edge (i.e., $E$ contains a pair $\{u, v\}$ such that both $u$ and $v$ are in $W$).

Assignment. Prove that we could solve this problem in polynomial time if we could compute, in polynomial time, the probability of the following Boolean CQ is true in a tuple-independent database.

$$Q() :\neg R(x), S(x, y), T(y)$$

---

1 Recall that $\cup$ represents a disjoint union.
References


Good luck!