3. Evaluation and Containment of Conjunctive Queries

Instructions: Work in groups of 2 students. You can write up your answers by hand (provided your handwriting is legible) or use a word processing system like Latex or Word. However, experience shows that Word is in general difficult to use for this kind of task. If you write by hand, please, scan your solution as a PDF file. Submit by email. Please, include your name and matriculation number in your submission.

1. Evaluation of Conjunctive Queries with Unary Relation Symbols

Recall that relational conjunctive queries have only relational atoms in their body, and no equalities or inequalities. We know that the combined complexity of evaluating relational conjunctive queries is NP-complete. However, the reduction used queries with binary relation symbols.

What can you say about the difficulty of evaluating relational conjunctive queries that have only unary relations in their body (that is, relations of arity 1)? Is this an NP-hard, a coNP-hard, or a polynomial time problem?

Distinguish between the problem for

1. relational queries (that is, queries without built-in predicates),
2. general conjunctive queries, which may contain the built-in predicates "$<$", "$\leq$", and "$\neq$".

To prove NP-hardness or coNP-hardness of a problem, provide a reduction from a known NP-hard or coNP-hard problem to the new one. To prove that it is in polynomial time, give an algorithm, show that it solves the problem, and explain why it runs in polynomial time.

(5 + 6 Points)

2. Containment of Relational Conjunctive Queries without Self Joins

A conjunctive query has a *self join* if its body contains two relational atoms with the same relation symbol. Thus, in the body of a query without self join, any two relational atoms have distinct relation symbols.

We know that containment is NP-complete for arbitrary relational conjunctive queries.

Question: How difficult is it to decide containment of relational conjunctive queries that have no self join? Can this problem be solved in polynomial time? Or is it NP-complete?

(4 Points)

3. Reducing the Hamiltonian Path Problem to Containment

Let $G = (V, E)$ be an undirected graph, where V is a finite set, the elements of which are called *vertices,* and $E \subseteq \mathcal{P}_2(V)$ is a collection of two-element subsets of V, the elements of which are called the *edges* of G.
A path in G is a sequence $v_1, \ldots, v_n$ of vertices such that $\{ v_i, v_{i+1} \} \in E$ for $i = 1, \ldots, n-1$ (that is, each node is connected to the next by an edge). A path $v_1, \ldots, v_n$ is *Hamiltonian* if in addition we have

1. $v_i \neq v_j$ if $i \neq j$ (that is, all vertices on the path are distinct)
2. $\{ v_1, \ldots, v_n \} = V$ (that is, $v_1, \ldots, v_n$ enumerates all vertices of G).

The Hamiltonian Path Problem is defined as follows:

Given: An undirected graph $G = (V, E)$.

Question: Does there exist a Hamiltonian Path in G?

This problem is known to be NP-complete.

Show that containment of relational conjunctive queries is NP-hard by reducing the Hamiltonian Path Problem to Query Containment.

Proceed in the following three steps:

1. Construct, for each graph G, a pair of queries Q, Q' such that $Q \subseteq Q'$ if and only if there is a Hamiltonian path in G.

(5 Points)

2. Prove that for a graph G and the corresponding pair of queries Q, Q', the containment $Q \subseteq Q'$ implies the existence of a Hamiltonian path.

   (5 Points)

3. Prove that from the existence of a Hamiltonian path in G one can conclude the containment of $Q \subseteq Q'$.

   (5 Points)

Hints: You may want to encode edges using a binary relation `edge`. Note that you have to make sure that any two nodes of a Hamiltonian path are distinct, for which another binary relation may come in handy.

Submission: 16 May, 10:30 am, by email