4. Evaluation of Conjunctive Queries

1. Reducing the Hamiltonian Path Problem to Conjunctive Query Evaluation

Let $G = (V, E)$ be a directed graph, where V is a finite set, the elements of which are called *vertices,* and $E \subseteq V \times V$ is a collection of pairs of vertices, the elements of which called the *edges* of G.
A path in G is a sequence $v_1, \dots, v_n$ of vertices such that $(v_i, v_{i+1}) \in E$ for $i = 1, \dots, n-1$ (that is, each node is connected to the next by an edge). A path $v_1, \dots, v_n$ 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. $\{ v_1, \dots, v_n \} = V$ (that is, $v_1, \dots, v_n$ enumerates all vertices of G).

The Hamiltonian Path Problem is defined as follows:

Given: A directed 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 conjunctive queries is NP-hard by reducing the Hamiltonian Path Problem to Query Evaluation.

Proceed in the following three steps:

1. Construct, for each graph G, an instance $\mathbf{I}_G$ and a query Q_G such that $Q_G(\mathbf{I}_G) \neq \emptyset$ if and only if there is a Hamiltonian path in G.
2. Prove that for a graph G, the corresponding instance $\mathbf{I}_G$ and the query Q_G, the nonemptyness of the set of answers $Q_G(\mathbf{I}_G)$ implies the existence of a Hamiltonian path.
3. Prove that from the existence of a Hamiltonian path in G one can conclude that $Q_G(\mathbf{I}_G) \neq \emptyset$.

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.

2. 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 or a polynomial time problem?

To prove NP-hardness, provide a reduction from a known NP-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.

Hint: Distinguish between relational queries (that is, queries without built-in predicates) and general conjunctive queries, which may contain built-in predicates.