5. Datalog

Instructions: Work in groups of 2 students. You can write up your answers by hand (provided your handwriting is legible) or using a word processing system.

The first two exercises are a warm-up to become familiar with writing queries in datalog. The others address questions about the expressive power of datalog.

1. Metro Reachability

Over the Metro database, consider the predicates `reachableFromOne`/3 and `reachableFromBoth`/3, with the following meaning for stations a, b, and c:

1. `reachableFromOne`(a, b, c) holds if c is reachable from one of a, b;
2. `reachableFromOne`(a, b, c) holds if c is reachable from either a or b.

Write datalog rules that define these predicates.

(5 Points)

2. Black and White Paths

Suppose you are given two directed graphs G_{black} and G_{white} over the same set of vertices, represented as binary relations. Write a datalog program that computes the set of pairs $\langle a, b\rangle$ of vertices such that there exists a path from a to b where black and white edges alternate, starting with a white edge.

(3 Points)

3. Properties Definable by datalog Programs

In this exercise we always mean "directed graphs" when we talk about graphs.

Show that graph properties definable by datalog programs are preserved under extensions and homomorphisms. That is show

Let P be a property of graphs, definable by a datalog program. Then

1. *every supergraph of G satisfies P;*
2. *if h is a graph homomorphism, then $h(G)$ satisfies P.*

(6 Points)

Hint: Suppose there is an EDB schema that has the binary predicate `edge` as its only element and that P is defined for some IDB predicates, among which there is a nullary predicate `ans`. A graph G is given by an instance $\mathbf{I}_G$ of the EDB. A graph G has property P if program P returns the answer $\texttt{ans}()$ over the instance $\mathbf{I}_G$.
For the first part of the question, you consider two graphs G, H, where H is a supergraph of G. For the instances, this means that $\mathbf{I}_G \subseteq \mathbf{I}_H$. The task is to show that P returns $\texttt{ans}()$ over $\mathbf{I}_H$ if P returns $\texttt{ans}()$ over $\mathbf{I}_G$.
For the second part we assume there is a surjective function $h\colon adom(\mathbf{I}_G) \to adom(\mathbf{I}_H)$ such that $\texttt{edge}(c', d') \in \mathbf{I}_H$ if and only if there are c, d such that $c' = h(c)$, $d' = h(d)$ and $\texttt{edge}(c, d)$ belongs to I_G. Again, the task is to show that P returns $\texttt{ans}()$ over $\mathbf{I}_H$ if P returns $\texttt{ans}()$ over $\mathbf{I}_G$. An approach to showing this could be to use the definition of datalog semantics by proof trees.

4. Which Properties are Definable by a datalog Program?

Which of the following properties of directed graphs is definable by a datalog program?

1. There is a trivial cycle (a trivial cycle consists of a single node a and an edge $\langle a, a\rangle$).
2. There is a nontrivial cycle.
3. For the two distinguished nodes a and b, there is a path between a and b.
4. For the two distinguished nodes a and b, there is no path between a and b.
5. The number of nodes is even.
6. There is a Hamiltonian path.

(12 Points)

Hint: To show that a property is definable, give a suitable datalog program. To show that a property is not definable, use the previous exercise.

Submission: 1 June 2010, 14:00 am, at the revision lecture or by email