Foundations of Databases

Exercises – Complexity and Expressiveness of Query Languages

April 5, 2005

1. Determine which of the following queries on a directed graph G, stored by its edges e, are generic:
   - There is a non-trivial path in G from node a to itself.
   - G has not a unique largest clique (a clique C is a subgraph of G which has an edge between any pair of nodes).
   - Assuming that *edb* contains a linear ordering $\leq$ of the nodes, return the first node in G which has no outgoing edge.
   - G is a tree.
   - G is 2-colorable.
2. Discuss the time complexity of the queries in Ex. 1.
3. Which of the queries in Ex. 1 can be expressed in (a) $\text{datalog}^{\neg}_{strat}$, (b) While^{+}, (c) While? If not, argue why. Would ordering help in case ?
4. Give a detailed calculation of an upper bound for the number of steps which is needed to evaluate a given datalog program P on a given input database $\mathbf{I} \in inst(edb(P))$ (i.e., deciding QOT).

   What changes if $\mathbf{I}$ is fixed?
5. Verify that for generic queries, the choice of the enumeration α of **dom** or the encoding $enc_{\alpha}(\mathbf{I})$ of an input database $\mathbf{I}$ to a query does not matter.
6. Suppose that $\mathbf{dom} = \{0, 1, 2, \ldots\}$ is the set of nonnegative numbers and has the usual ordering $\leq$. Develop a natural definition of "generic" for this setting, and describe a non-generic query if there is one.

7. Consider the problem of recognizing the whole output of datalog query, i.e., whether for given databases $\mathbf{I}$ and $\mathbf{J}$ it holds that $P(\mathbf{I}) = \mathbf{J}$. What is the complexity of this problem (a) if P is fixed, and (b) if P is part of the input?

8. Determine whether the following properties of graphs are almost surely true or almost surely false:

   - Existence of a cycle of length three
   - Disconnectivity
   - Being a tree

9. Consider the class of datalog programs in which (a) the arity of each relation (b) the number of different variables is bounded by a constant number r. Analyze the data and expression complexity of such programs.

10. Adapt the Turing-Machine encoding from the lecture for computations without output, such that the output string computed by T is provided (in similar encoding as the input) in some designated output relation(s) of P.

11. Spell out the details in the modification of the propositional Turing-Machine encoding for showing the expression complexity hardness.

12. Consider the simulation of a Turing-Machine which evaluates a Boolean query $f : \mathbf{EDB} \rightarrow \{p\}$ on all input instances $\mathbf{I}$, encoded as strings $I = enc(\mathbf{I})$ from the lecture. Assuming that $|adom(\mathbf{I})| > 1$, that all relations $R \in \mathbf{EDB}$ have the same arity, and that that $enc(\mathbf{I})$ is a bitmap, describe rules to compute the facts $input_\sigma(\pi)$, representing $\mathbf{I}$ padded with blanks "␣".

   .