



Database Theory

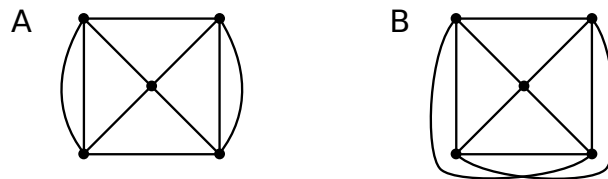
Summer Semester 2016

Exercise Sheet 8 – FO Expressivity and Introduction to Datalog

13th June 2016

Dr. rer. pol. Markus Krötzsch & Dipl.-Math. Francesco Kriegel

Exercise 8.1 A graph is *planar* if it can be drawn on the plane without intersections of edges. For example, the following graph A is planar, while graph B is not:



Can the graphs A and B be distinguished by a first-order query? Show that planarity is not FO-definable by using locality.

Exercise 8.2 Consider the example Datalog program from the lecture:

- father(alice, bob) (1)
- mother(alice, carla) (2)
- mother(ewan, carla) (3)
- father(carla, david) (4)
- Parent(x, y) \leftarrow father(x, y) (5)
- Parent(x, y) \leftarrow mother(x, y) (6)
- Ancestor(x, y) \leftarrow Parent(x, y) (7)
- Ancestor(x, z) \leftarrow Parent(x, y) \wedge Ancestor(y, z) (8)
- SameGeneration(x, x) (9)
- SameGeneration(x, y) \leftarrow Parent(x, v) \wedge Parent(y, w) \wedge SameGeneration(v, w) (10)

- (a) Give a poof tree for SameGeneration(ewan, alice).
- (b) Compute the sets $T_P^0, T_P^1, T_P^2, \dots$. When is the fixed point reached?

Exercise 8.3 Consider databases that encodes a labelled, directed graph by means of a ternary EDB predicate e ("edge"). The two parameters are the source and target nodes of the edge, while the third parameter is its label. For example, the edge $n_1 \xrightarrow{a} n_2$ would be represented by the fact $e(n_1, n_2, a)$. Moreover, assume that only constants a and b are used as labels.

Can you express the following queries using Datalog?

- (a) "Which nodes in the graph are reachable from the node n ?"
- (b) "Are all nodes of the graph reachable from the node n ?"
- (c) "Does the graph have a directed cycle?"

- (d) "Does the graph have a path that is labelled by a palindrome?"
(a palindrome is a word that reads the same forwards and backwards)
- (e) "Is the connected component that contains the node n 2-colourable?"
- (f) "Is the graph 2-colourable?"
- (g) "Which pairs of nodes are connected by a path with an even number of a labels?"
- (h) "Which pairs of nodes are connected by a path with the same number of a and b labels?"
- (i) "Is there a pair of nodes that is connected by two distinct paths?"

Exercise 8.4 Consider a UCQ of the following form

$$(r_{11}(x) \wedge r_{12}(x)) \vee \dots \vee (r_{\ell 1}(x) \wedge r_{\ell 2}(x))$$

Find a Datalog query that expresses this UCQ. How many rules and how many additional IDB predicates does your solution use (depending on ℓ)?

Exercise 8.5 Consider a Datalog query of the following form:

$$\begin{array}{lll} A_1(x) \leftarrow r_{11}(x) & \dots & A_\ell(x) \leftarrow r_{\ell 1}(x) \\ A_1(x) \leftarrow r_{12}(x) & \dots & A_\ell(x) \leftarrow r_{\ell 2}(x) \end{array}$$

$$\text{Ans}(x) \leftarrow A_1(x) \wedge \dots \wedge A_\ell(x)$$

Find a UCQ that expresses this Datalog query. How many CQs does your solution contain (depending on ℓ)?

Exercise 8.6 Show that T_P^∞ is the least fixed point of the T_P operator.

- (a) Show that it is a fixed point, i.e., that $T_P(T_P^\infty) = T_P^\infty$.
- (b) Show that every fixed point of T_P must contain every fact in T_P^∞ .