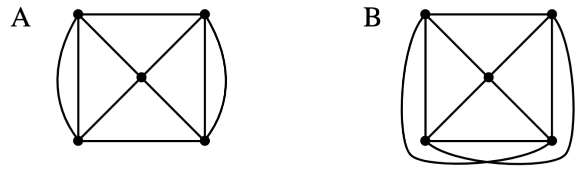
Exercise Sheet 8: Datalog

Maximilian Marx, David Carral

Database Theory, 2020-06-10, Summer Term 2020

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:

$$\mathsf{father}(\mathsf{alice}, \mathsf{bob}) \tag{0.1}$$
$$\mathsf{mother}(\mathsf{alice}, \mathsf{carla}) \tag{0.2}$$
$$\mathsf{mother}(\mathsf{evan}, \mathsf{carla}) \tag{0.3}$$
$$\mathsf{father}(\mathsf{carla}, \mathsf{david}) \tag{0.4}$$
$$\mathsf{Parent}(x, y) \leftarrow \mathsf{father}(x, y) \tag{0.5}$$
$$\mathsf{Parent}(x, y) \leftarrow \mathsf{mother}(x, y) \tag{0.6}$$
$$\mathsf{Ancestor}(x, y) \leftarrow \mathsf{Parent}(x, y) \tag{0.7}$$
$$\mathsf{Ancestor}(x, z) \leftarrow \mathsf{Parent}(x, y) \wedge \mathsf{Ancestor}(y, z) \tag{0.8}$$
$$\mathsf{SameGeneration}(x, x) \tag{0.9}$$
$$\mathsf{SameGeneration}(x, y) \leftarrow \mathsf{Parent}(x, v) \wedge \mathsf{Parent}(y, w) \wedge \mathsf{SameGeneration}(v, w) \tag{0.10}$$

1. Give a poof tree for $\mathsf{SameGeneration}(\mathsf{evan}, \mathsf{alice})$.

2. Compute the sets T_P^0, T_P^1, T_P^2, … 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?

1. "Which nodes in the graph are reachable from the node n?"

2. "Are all nodes of the graph reachable from the node n?"

3. "Does the graph have a directed cycle?"

4. "Does the graph have a path that is labelled by a palindrome?"
(a palindrome is a word that reads the same forwards and backwards)

5. "Is the connected component that contains the node n 2-colourable?"

6. "Is the graph 2-colourable?"

7. "Which pairs of nodes are connected by a path with an even number of a labels?"

8. "Which pairs of nodes are connected by a path with the same number of a and b labels?"
9. "Is there a pair of nodes that is connected by two distinct paths?"

Exercise 8.4. Consider a UCQ of the following form

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

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:

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

$$\mathsf{Ans}(x) \leftarrow A_1(x) \wedge \ldots \wedge A_\ell(x)$$

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.

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