3. Datalog

1. Reachability in Graphs

We are given two directed graphs G_{black} and G_{white} over the same set V of vertices, represented as binary relations. Write a datalog program P 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.

2. Datalog Queries

Suppose in a travel agency database there is a table with the schema

$$\texttt{flight}(\texttt{from}, \texttt{to}, \texttt{airline}),$$

where an entry $(c1,\ c2,\ l)$ means that it is possible to fly from city *c1* to city *c2* with the airline l.
Consider the following two queries:

1. Return all pairs of cities (x, y) such that is possible to travel from x to y using flights of *no more than one* airline.

2. Return all pairs of cities (x, y) such that is possible to travel from x to y using flights of *no more than two* airlines.

Which of these queries is expressible in datalog and which not? To show that a query is expressible, write down a datalog program that computes the query. To show that a query is not expressible, sketch a proof. In that proof, you may use any of the results on the expresivity of query languages that were shown in the course.

3. Transitive Closure and First-order Logic

Is it possible to express transitive closure in first-order logic? In other words, given a binary relation $R(\cdot, \cdot)$, is it possible to write a formula $\varphi(x, y)$ such that in every interpretation of R, the formula φ is satisfied exactly by those pairs of domain elements that are in the transitive closure R^+ of R?
Hint: Consider the following statements about a point a and their formulation as logical formulas:

- There exists a point reachable from a that is an R-sink, that is, no R-edge is emanating from that point.
- For each natural number k: All points reachable from a via a path of length k have an emanating R-edge.

4. Properties of Datalog Programs

Suppose P is a property of graphs, definable by a datalog program. Show that P is preserved under extensions and homomorphisms. That is, if G is a graph satisfying P, then

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

5. Definability in Datalog

Which of the following graph properties 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 nodes a and b, there is a path between a and b.
4. For the two 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.