  




1. [Graph pattern matching/subgraph isomorphism] (30) Consider the following biomedicine knowledge graph, describing proteins and the relevant diseases that may be, for example, treated, caused, or expressed by them.

A diagram of a disease

Description automatically generated

In the lecture, we introduced graph patterns and subgraph isomorphism. For the following natural language questions, (1) draw corresponding graph patterns that can express them, and (2) give a list of all the answers (use the labels on the node as their identifier, for example, “PIM1”), which are found by subgraph isomorphism mapping between your pattern and the given knowledge graph. Recall that to define the output of the queries, you can “parameterize” the pattern by adding a question mark to some pattern node. Also recall that a pattern can have multiple nodes with the same type (color coding, in the above graph).

1. “Find all the proteins that are associated with those diseases that can be treated by Arimidex.”
2. “What are the proteins that are associated with the adverse events which are caused by ‘Fulvestrant’? ”
3. “Find all the proteins that form a 3-clique”. A 3-clique in a directed graph refers to a subgraph that contains three nodes that are pairwise connected.
4. “Find the proteins that are associated with some drugs that may cause at least three different advert events. “
5. Find all the entities, regardless of their type (use a wildcard “\_” in the nodes of the pattern graph), that can form a triangle.
6. \*(bonus, +5 credits) A “knowledge hacker” may try to confuse people and inject fake facts by adding or removing an edge in a knowledge graph. If a new edge “treat” is added between Arimidex and PIM1, how will the answers to any of the above questions change? Give the updated answer, if any.
7. [RDF and RDF schema] (30) Consider the following figure, a graph representation of an RDF for a movie knowledge base.

A diagram of a movie

Description automatically generated

1. [10] This RDF graph contains the “schema” and “non-schema” parts. Circle out those nodes that belong to the non-schema part.
2. [20] Consider the subgraph of the above graph that contains the nodes { “jackson”, “fellowship”, “mckellan” and “Ian McKellan”}. Translate this graph into RDF/XML syntax.
3. [SPARQL](25) Consider the graph in Question 1, and assume it has an RDF graph representation in place (not given, as this doesn’t affect SPARQL query construction for this question). Write SPARQL queries to express the patterns you provided for Questions 1 (a) – 1 (e).
4. [Knowledge Graphs] [15] Among the advantages of representing facts with (knowledge) graphs is its capability of supporting (logic) reasoning. Consider 6 people M, N, O, P, Q and R have the following attributes.
   * M is richer than N, but shorter than R.
   * N is richer than Q but taller than R.
   * O is poorer than P and taller than M.
   * P is poorer as well as shorter than Q.
   * Q is poorer as well as shorter than M.
   * R is richer than M and taller than O.
5. Draw a directed graph that can encode the above facts. For convenience, consider “reverting” the relation ‘poorer’ to ‘richer’ by changing the edge direction.
6. An (s, r, o) triple in an RDF statement is also simply denoted as r(s,o). A relation r is transitive if r(A,B) and r(B,C) implies r(A,C). Clearly, “richer” and “taller” are transitive. How may you use the graph and transitivity to find the answer to the following questions:
   1. Who is the richest person?
   2. Who is the tallest person?