Exam Re-Take

Uncertain Data Management
Université Paris-Saclay, M2 Data&Knowledge

March 13th, 2018

This is the re-take of the final exam for the Uncertain Data Management class. The grade in this exam will replace your grade of the first session of the final exam, and will become your final grade for the class.

The exam consists of 4 independent exercises: exercises 1 and 2 are about uncertain data management (Antoine's part), and exercises 3 and 4 are about social data management (Silviu's part). **You must write your answer to exercises 3 and 4 on a *separate sheet* of paper.**

No additional explanations will be given during the exam, and no questions will be answered. If you think you have found an error in the problem statement, you should report on your answer sheet what you believe to be the error, and how you chose to interpret the intent of the question to recover from the alleged error.

You are allowed up to two A4 sheets of personal notes (i.e., four page sides), printed or written by hand, with font size of 10 points at most. If you use such personal notes, you must hand them in along with your answers. You may not use any other written material.

The exam is strictly personal: any communication or influence between students, or use of outside help, is prohibited. No electronic devices such as calculators, computers, or mobile phones, are permitted. Any violation of the rules may result in a grade of 0 and/or disciplinary action.

Exercise 1: Knights of the Round Table (4 points)

Consider the following BID table representing the uncertain location of knights of the round table. The key attribute of the BID table is **knight**.

T

knight	**location**	
Lancelot	Camelot	0.4
Lancelot	Brocéliande	0.4
Galahad	Camelot	0.6

Question 1 (0.5 point). Which of the following two tables is a possible world of T? (No justification is expected.)

T_1

knight	**location**
Lancelot	Camelot
Galahad	Camelot

T_2

knight	**location**
Lancelot	Camelot
Lancelot	Brocéliande

Question 2 (0.5 point). Consider the query Q_1 $\pi_{\textbf{location}}(\sigma_{\textbf{knight}=\text{``Galahad''}}(T))$ asking for the location of Galahad. Give a TID instance representing the result of evaluating Q_1 on the table T.

Question 3 (1 point). Consider the query Q_2 asking for the knights located at Camelot. The query Q_2 should return a table with only one attribute, named **knight**. Write down the query Q_2 in the relational algebra.

Question 4 (1 point). Consider the evaluation of Q_2 on the table T. Can the result be represented as a TID instance? If yes, give a suitable TID instance and justify; if not, prove that no TID instance can represent it.

Question 5 (1 point). Is there a query Q_3 such that the result of evaluating Q_3 on the table T cannot be represented as a pc-instance? If yes, give such a query Q_3 and justify; if not, prove that no such query Q_3 exists.

Exercise 2: Numbers of Possible Worlds (6^+ points)

In this exercise, recall that a *row* of a table is simply a line of the table, e.g., the BID instance T given in Exercise 1 has 3 rows.

Question 1 (1 point). Give an example of a TID instance that has at least 42 possible worlds.

Question 2 (1 point). If a TID instance has n rows, how many possible worlds can it have at most? Give the best possible bound, and justify why it is correct.

Question 3 (1 point). If a BID instance has n rows, how many possible worlds can it have at most? Give the best possible bound, and justify why it is correct.

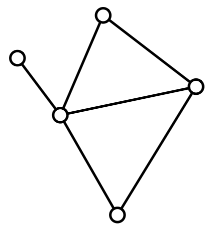
Question 4 (1 point). Is there a TID instance having *exactly* 42 possible worlds? If yes, give an example; if not, explain why.

Question 5 (2 points). Is there a BID instance having *exactly* 42 possible worlds? If yes, give an example; if not, explain why.

Question 6 (bonus). We consider a positive integer n, and search for a BID instance having exactly n possible worlds (if one exists) and whose number of rows is as small as possible. We call $r(n)$ the minimal number of rows of a BID instance having n possible worlds. Explain how to compute $r(n)$ as a function of n.

Exercise 3: Graph Measures (6 points)

In this exercise, we will work with some graph analysis measures, as defined in the course. Consider the following *undirected* graph G:



Question 1 (0.5 point). Write down the degree distribution of the graph G.

Question 2 (1 point). What is the average degree $\langle k \rangle$ in the graph G? Explain how it can be computed from the degree distribution.

Question 3 (0.5 point). How many triangles are there in the graph G? *Reminder*: a triangle is a subgraph of size 3 that is complete.

Question 4 (1 point). We work now with the Erdős-Rényi random graph model, and we wish to get random graphs having the same average degree as G. What is the resulting parameter p – the probability of an edge existing – in such a graph?

Question 5 (3 points). Compute the expected number of triangles in the random graph model, as a function of the number of nodes $|V| = n$ and of the parameter p. Now compute this value using the value of p obtained in Question 4, and using the same number of vertices as G, i.e., 5. How does it compare to the number of triangles in the example graph above?

Hint: you will need to use the formula for the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Exercise 4: Random Graphs Modeled as Uncertain Graphs (4 points)

In this exercise, we will work with the Erdős-Rényi random graph model again. A random graph with parameter p can also be thought of as an uncertain graph $\mathcal{G} = (V, E, p)$, where the underlying graph is complete, i.e., there exists an edge between every node in both directions, and each edge has the same probability of existing, equal to p.

Question 1 (2 points). Take the random graph having $|V| = 3$ and $p = 0.2$, and take two distinct nodes s and t of this graph. Compute the reachability probability between s and t. *Hint*: the probability is the same no matter our choice for s and t.

Question 2 (2 points). Now consider the general case, where we write $n := |V|$ and write p the probability parameter. We assume that $n > 1$, and choose two distinct nodes s and t. Compute the probability that s and t are at distance *exactly* 2: that is, to reach t from s, the shortest path passes through one intermediary node x (different from s and t).