Assignment 5
Due July 1st, 2018

Part 1: Probabilistic Databases

**Question 1.1.** Let $S$ be a schema, and let $D$ be a finite probabilistic database over $S$. Prove that there exists a pc-instance $P$ over $S$ such that $[P] = D$.

**Question 1.2.** For each of the following Boolean CQs, determine whether the evaluation can be done in polynomial time or is #P-hard. When applicable, you can use the Dalvi-Suciu dichotomy theorem studied in class.

1. $Q_1() ::= R(x, y), S(x, z, w), T(x, z, u)$.
2. $Q_2() ::= R(x, y), S(x, z, y), T(x, z, u)$.
3. $Q_3() ::= R(x, y), R(y, x), S(x, y, z)$.

Part 2: Bounded Network Reliability

By a “network” we refer to a directed graph $G$ with two distinguished nodes: a source $s$ and a target $t$. In addition, every edge $(u, v)$ has an associated probability $p(u, v)$ of disconnecting (disappearing). We assume that edges disconnect independently. In the following assignments, we are given a network as input.

**Question 2.1.** Devise a polynomial-time algorithm computing the exact probability that there is a path of length at most 2 from $s$ to $t$.

**Question 2.2.** We would like to compute the probability that there is a path of length at most 3 from $s$ to $t$.

1. Using the Dalvi-Suciu dichotomy, prove that this problem is #P-hard.

2. Using what we have learned about approximating UCQs using the Karp-Luby estimator, prove that this problem has an FPRAS.

**Question 2.3.** Show that there is an FPRAS for computing the probability that there are at least two edge-disjoint paths of length at most 3 from $s$ to $t$.

Good luck!