**CS 63018 & CS 73018 Probabilistic Data Management**

**Homework 4**

**Instructor:** Xiang Lian

**Due Date:** Please refer to the course website

1. Please discuss one semantic of probabilistic top-*k* query over uncertain data, including the problem definition and its real applications. **[20 points]**

2. Please discuss probabilistic similarity join operator over uncertain data, including the problem definition and its real applications. **[20 points]**

3. Please refer to the paper below and answer the following 3 questions. **[60 points]**

*J. Pei, B. Jiang, X. Lin, and Y. Yuan. Probabilistic skylines on uncertain data. In VLDB, 2007.* [*https://dl.acm.org/citation.cfm?id=1325858*](https://dl.acm.org/citation.cfm?id=1325858)

3(a). Please describe the definition of dominance between two data points (static attributes). [20 points]

3(a). Please give the definition of dominance probability between two uncertain objects. [20 points]

3(c). Please give the definitions of skyline probability and probabilistic skyline query over uncertain data. [20 points]

**Bonus Question** [extra 20 points] Describe the problem of your project (see the project template), including:

* Introduction,
* Project description,
* Background, and
* Problem definition.

**Submission**

Submit an electronic copy of your homework solution to the [Blackboard](https://learn.kent.edu/).

18