

## Project Statement for Milestone 2

(Optional) You (team) name

Team member names

In this report you should focus on the datasets description, dataset preparation and formatting, Description of data collection and the tools you use. Usually you will write a parser to extract the information you need to the data structure/platform you will be using.

### 1. Datasets and tools

- a. Data model: what's the data model you will be use to represent the dataset? Why?
- b. Datasets statistics: how large is the dataset you will be dealing with? Report the following statistics for your datasets:
  - i. If you are using relational data model: how many tuples? How many attributes? Any data constraints (FDs?) What's the physical storage size (in KB/MB/GB).
  - ii. If you are using a graph data set: how many nodes and edges? How many attributes are there for the nodes/edges? Is it labelled? Directed? What's the average degree of the nodes? What's the density of the graph/network data? Understand these statistics help you design the algorithm in M3.
- c. In data processing, you may need to develop a parser to transform the raw data into the format/tools you are using. Briefly describe the functions of the parser you implemented.
- d. You should have successfully loaded the data into the tools you use. What's the estimated loading time, if you have the result?

### 2. Data structure and auxiliary structure. What data structure you developed of your own to represent the data, if any? For example, if you are using graphs, put up pseudo code that represent the data. Do you use any types of indexes? If so, briefly describe the indexes you developed for fast access the data.

### 3. What's your plan for Milestone 3? In Milestone 3, you will be describing the detailed algorithm for your problem. Give a brief description about how you plan to design the algorithm. Provide any related work you plan to read if any.