

Due Wednesday, January 27 by the end of the day

Questions may continue on the back. Please create a PDF file and submit via Blackboard.

*I have adhered to the Duke Community Standard in completing this assignment.*

Signature: \_\_\_\_\_

**1. (+1 pts.) Basics**

Complete survey labeled *Beginning of Semester Survey* on Blackboard.

**2. (15 pts.) Book problems**

- (a) §2.5, Exercise 1
- (b) §2.5, Exercise 2
- (c) §2.5, Exercise 3

**3. (10 pts.) Network Construction**

*Taken from Prof. Michael Kearns Networked Life course.*

You should identify a specific source of real-world data, the precise definition of the network (vertices and edges) you plan to extract from this data, and the methodology by which you will extract it.

We will be generous with the term "real-world", which could include data from the domains of biology, sociology, economics and finance, technology, etc. However, the data must come from a well-defined, objective data source gathered by a third party. An example of an entirely acceptable data source is the corpus of emails exchanged by Enron executives, where it would be natural to examine the network of whom exchanged email with whom. An example of an unacceptable data source and network would be "I wrote down a list of all my friends and then connected any pair of them that I thought shared a lot of common interests". This example is too subjective and the data is not gathered by a third party.

To be sure there is some minimal level of complexity to your network, we require that the number of vertices in the network be at least 25, and the total number of edges in the network to be at least 25. However, considerably more ambitious networks are encouraged.

By the "methodology" by which you will extract your network, we mean how you plan to go from the raw data source and your defined network to an actual representation of your network in our simple format (see below, but essentially nothing more than a list of all the vertices in your network, followed by a list of all those pairs of vertices that are connected by an edge).

For this part, you should submit a brief write-up detailing the information described above for your network. If your data source is online, please provide the URLs for the source; feel free to include a small portion of the raw data in your write-up if it would be helpful to do so. Be sure to be as precise as possible in all aspects of your write-up, from network definition to methodology. As an informal test, your write-up should be sufficiently precise that a third party could independently create the same network you will from your description.

As a rough guideline, I would expect most write-ups to be at least a page long.