BASIC ASSIGNMENT 5
CORE METHODS IN EDUCATIONAL DATA MINING
PROFESSOR BAKER
SOCIAL NETWORK ANALYSIS
DUE NOON, WEDNESDAY NOVEMBER 5

In this assignment, I've taken the data from the discussion forums during week 1 of "Big Data and Education", taught on Coursera in Fall 2013. This data was pulled from the discussion forum, which was publicly accessible to any student taking the course (and to anyone who enrolls in the course today, even though the course ended almost a year ago).

The data sheet has four columns: the post ID, the poster's (deidentified) ID, the thread the post was in, and which post this post was a response to.

You can complete this assignment using any social network analysis tool (people tend to like Gephi), or in Microsoft Excel, or any other tool you like.

1) How many possible connections are there between individuals in this graph, not counting self-connections, and assuming that link direction does not matter?

2) Only counting connections where poster A actually responds directly to poster B (and not connections where they both posted in the same thread), how many connections are there between individuals in this graph? (don't count self-connections and assume that link direction does not matter)

3) What is the density of the social network graph? (don't count self-connections and assume that link direction does not matter)

4) What is the geodesic distance between 12345 and 24601?

5) What is the geodesic distance between 3903 and 1588?

6) Which of these nodes is not reachable?

A) 87908
B) 22986
C) 74577
D) 81734

7) Which individual has the most posts?

8) Which individual is most often responded to by other posters? (the in-degree)

9) Who do you think the individuals from #7 and #8 are?