CS5984: Urban Computing
Assignment 2
Date Assigned: Oct 28, 2015
Date Due: Nov 9, 2015

1) **(35 points) Road Networks, Pagerank, and HITS:** An application of the Pagerank and HITS algorithms in the context of urban computing is in modeling traffic flow. The discovery of important intersections can be used for transportation planning.
Download the three directed, weighted, road networks from:
https://www.dropbox.com/s/7a564p60zu1s5q8/weighted_road_networks.zip?dl=0

    a. Compute the Pagerank score and rank the nodes according to the score.
    b. Compute the HITS scores of the graph.
    c. Is there a difference in the ranking? If yes, can you quantify/justify what and why?

2) **(30 points) Computational Epidemiology:** Syndromic surveillance using the web in both social media and search activities online have become popular. For instance, Google Flu Trends uses a list of keywords to track the flu season.
    a. If you were to develop a keyword list that one could use to forecast the flu season using such social media and search data, how would you go about it?
    b. Geography and language: Can you modify the list based on geography and language? For instance, even within the United States – some refer to soft drinks like Coke and Pepsi as 'soda' whereas some refer to it as 'pop'. Moreover, how will such a list be modified for language (e.g., English vs Spanish)?

3) **(35 points) Computational Epidemiology:** We have studied several epidemiological models and understanding the spread process via simulations is critical to policy and decision making. Towards that end, in this question, consider the Portland contact graph from the NDSSL dataset that was downloaded for the previous assignment.
    a. Implement the SIS, SIR, and SIRS models by setting infectivity parameter to be 0.3 and recovery parameter to be 0.5.
        i. Let the number of *seed* nodes (*patient zeroes*) be (i) 5 (ii) 50 (iii) 100 and plot the *epi-curve* (# of individuals infected as a function of time) for each of the spread models. Show how the epi-curve is different for each of the different seed set size.
    b. Consider a scenario where there are three infections I_1, I_2, and I_3 and an extension of the SIS model. Now this becomes a $SI_1I_2I_3S$ model. Assume that each of the infections have mutual immunity i.e. if a node is infected with the I_1 strain, it cannot contract the other two. However, once the node has recovered (gone back to being susceptible) from I1 and one of it's neighbors is attempting to infect it with I2 or I3 it can contract those strains. Consider the following infectivity and recovery parameters:
        i. I1: ß = 0.3 and $\partial = 0.5$
        ii. I2: ß = 0.4 and $\partial = 0.6$

iii. I3: ß = 0.3 and $\partial = 0.7$

Plot the epicurve for each of the infections and observe what happens. Is there any property that you can observe as a function of time? For instance, do all three viruses exist in the system as time goes by. Explain your findings.