CS5984: Urban Computing

Assignment 1

Posted: Sep 18, 2015

Due Date: Oct 2, 2015

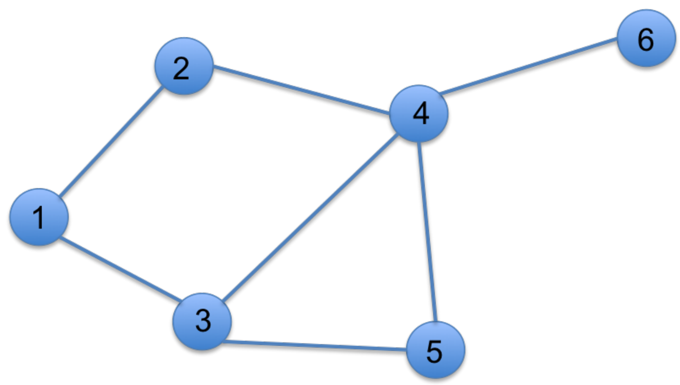
Submission guidelines will be posted on piazza

1) **(15 points)** The D.C. metro map is essentially a network with nodes denoting various stations; an edge between two nodes indicates that those nodes are successive stations on some line. Moreover, there are 6 different lines on this map that connect the tri-state area (D.C., Virginia, and Maryland) viz. silver, orange, blue, yellow, green, and red. In order to maintain connectedness, some nodes serve as transfer points. For instance, to travel from Wiehle-Reston to Bethesda, one would take the silver line to Metrocenter and then take the red line to Bethesda.
    a. **Recolor the lines:** While retaining the exact same stations and the exact same network structure (so that new tracks need not be laid), how can the different lines be reorganized to optimize traffic flow? For instance, perhaps a new line can be created that goes from Wiehle-Reston East to Glenmont (instead of Largo). What information would you ask WMATA (the government authority in charge of the DC metro) to answer this question? What would be your objective? Use graph-theoretic concepts covered in class to describe your solution. Your proposed solution must also have six lines as in the original map. For a picture of an uncolored map, please see: http://www.wmata.com/rail/docs/bw_map_silverline.pdf
    b. **Express Trains:** All lines currently run local (i.e., they stop at all stations on the path). Suppose WMATA decides to add *three express trains* to account for volume. For instance, if there is a huge inflow in East Falls Church and outflow in Rosslyn, connecting them could optimize flow. How will you identify the start/stop destinations for these three new express trains? What information would you ask WMATA to answer this question? Again, use graph-theoretic concepts covered in class.

2) **(15 points) Immunization:** One of the important problems in epidemiology is to select nodes to immunize in a network such that the spread of disease is limited. Given a network of people, provide two strategies (and explain them) to immunize the minimum number of nodes such that the least number of nodes in the overall network are infected. Define what you mean by "limiting the spread of disease". Use graph-theoretic concepts covered in class.

3) **(15 points) Sensor Placement:** Consider a power network where each node is a substation and the edges between them are high voltage transmission lines. Using a specific sensor, a node can monitor its own state (whether it is working or has failed) and the state of its neighbors. Propose a strategy to place the least number of sensors such that the the state of the entire power network can be monitored effectively.

4) **(25 points)** In the three road networks given at:
i) http://snap.stanford.edu/data/roadNet-CA.html
ii) http://snap.stanford.edu/data/roadNet-PA.html
iii) http://snap.stanford.edu/data/roadNet-TX.html
   a. Calculate and plot the degree distributions of each of these networks. Do they follow a power law distribution?
   b. Are these networks connected? Is there a rigorous mathematical way to test connectedness given any network? if yes, present an algorithm for it.
   c. If the networks are connected, what is the least number of nodes that need to be removed to make them disconnected? How will you identify such nodes?
   d. A *wedge* in a graph is a two-hop path. More colloquially, it is an *open triangle*. (A triangle is also referred to as a closed wedge.) In the figure (attached here), 5-4-6 is a wedge and 3-4-5 is a closed wedge. Compute the total number of *closed and open wedges* in the above three networks.



   e. The wedge center is the node that forms the middle vertex of the wedge. For example, 5-4-6 is centered at 4. The wedge center ratio is defined as the total number of wedge center nodes divided by the total number of nodes in the graph. Compute the *wedge center ratio* of the three networks above. What use is this information in the context of the underlying road networks?

5) **(30 points)** Download the Portland activities dataset, which is a bipartite graph, from http://ndssl.vbi.vt.edu/synthetic-data/download.html (documentation describing the data is available in the link)
   a. Induce a 1-mode network such that two individuals are connected if they spend at least two hours in the same location.
   b. Plot the degree distribution of the induced network.
   c. For the induced network, compute the different centrality measures learnt in class – degree, betweenness, and closeness.
   d. Is the resulting network a connected network? If yes, give the least number of nodes that need to be removed to make this into four disconnected components. If the resulting network is not connected, how many connected components exist?

6) **(20 points) Bonus Question:** In a Cayley tree, also known as a symmetric k-regular tree, each node at distance d from the central node has degree k, until we reach the leaves. (A Cayley tree for k=3 given below)

   a. How many nodes are reachable from the central node in *t* steps?
   b. Calculate the degree distribution of the network (in terms of k and n, where k is the degree and n is the number of nodes).
   c. What is the diameter in terms of *k* and *n*?
   d. What does this notion have to do with urban computing? Where would it be useful?