Networked Life (CIS 112), Spring 2010

Prof Michael Kearns

Homework 1

Issued Feb 9, 2010. Due as ***hardcopy*** in class Tue Feb 23. Don't forget to ***staple and write your name.***

1. (20 points) Consider an undirected network where each vertex corresponds to an English word. There is an edge between two words **A** and **B** if either the exact phrase "**A B**" or "**B A**" has more than 1,000 results from Google. *Be sure to put quotation marks around all phrases.* This ensures that Google only returns pages in which the words appear immediately adjacent to each other. You can see the count of how many pages were found in the upper-right corner of the Google search results page.

   For example, if you search for "devastating news" (including the quotation marks), there are about 159,000 results returned, so there's an edge between "devastating" and "news"; "dimly brilliant" has only 210 results and "brilliant dimly" has only 2 results, so there is *not* an edge between "brilliant" and "dimly".

   a. Find the shortest path you can in this network between "epidemic" and "voting". Do not go through or use anything but nouns and adjectives (e.g. no articles such as "an", "the", etc.) at every step.

   b. Find the shortest path you can between "equilibrium" and "contagion". Do not go through or use anything but nouns and adjectives (e.g. no articles such as "an", "the", etc.) at every step.

   c. Find the shortest path you can between "clique" and "suicide". Do not go through or use anything but nouns and adjectives (e.g. no articles such as "an", "the", etc.) at every step.

   d. Find the shortest path you can between "obesity" and "topology". Do not go through or use anything but nouns and adjectives (e.g. no articles such as "an", "the", etc.) at every step.

2. (15 points) Consider the ***directed*** network of words in which there is an edge from word **A** to word **B** if word **B** occurs in the dictionary definition of word **A.** Find a path starting from the word "life" to the word "network" using the online dictionary:

   http://www.merriam-webster.com/

3. (15 points) Experiment with the Networked Life "Disease Demo" available at: http://www.cis.upenn.edu/~mkearns/teaching/NetworkedLife/demos/Epidemic.html .

   Leave the rewiring rate at the default of 0.1 and the random seed at the default of 73.

For each infection probability in {0.0, 0.1, 0.2, 0.3,…,1.0}, run three trials of the simulation. For each of those infection probabilities, record the average number of generations your simulations lasted. The number of generations corresponds to how long it takes the disease to die out, and is displayed in the upper right corner.

a. Using Excel, your other favorite plotting software, or careful handwriting, create a plot of your data. The infection probability should be on the x-axis and the average number of generations should be on the y-axis.

b. At what infection probability does the disease last the greatest average number of generations?

4. (15 points) The books "The Tipping Point" and "Connected" cover much of the same thematic material, but from different perspectives and with different emphases. Write a brief (about 1 page) essay in which you compare and contrast the approaches in the two books. On what common topics are the emphases taken most different? Most similar? Be as specific as you can, citing passages by page numbers to support your comparisons. (However, don't quote lengthy passages within your essay itself.)

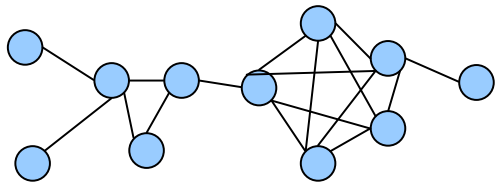
5. (15 points) Recall that the ***clustering coefficient*** of a vertex v in an undirected network is the fraction of pairs of v's neighbors that are connected. If a vertex has 0 or 1 neighbors, we define its clustering coefficient to be 0. The clustering coefficient of the entire network is the average of all the vertex clustering coefficients.

a. Draw an undirected network that is connected (that is, has only one component), has exactly 9 vertices, exactly 11 edges, and has the smallest clustering coefficient you can arrange. Explicitly compute the clustering coefficient of your network, showing your work.

b. Draw an undirected network that is connected (that is, has only one component), has exactly 9 vertices, exactly 11 edges, and has the largest clustering coefficient you can arrange. Explicitly compute the clustering coefficient of your network, showing your work.

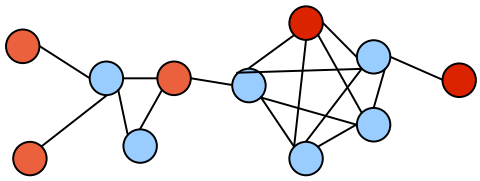
6. (20 points) For this question you will need a Facebook account. Define your immediate neighborhood on Facebook to be yourself plus all the people you are linked to as friends.

a. A ***clique*** in a network is a subset of vertices, every pair of which are directly connected to each other by edges (that is, all pairs of a clique are neighbors). The size of the clique is just the number of vertices in the clique. For example, the network below has cliques of size 5 and size 3, as well as many cliques of size 2 (which are just the edges):



What is the size of the largest clique you can find in your immediate Facebook neighborhood? How did you find it? What reasons can you think of which make this group of individuals so highly interconnected?

b. An ***independent set*** in a network is just the opposite of a clique. An independent set is a subset of vertices ***none of which*** are connected by an edge. In the network below, the red vertices form an independent set since of size 5 since none of them are connected by an edge:



What is the size of the largest independent set you can find in your immediate Facebook neighborhood? How did you find it? What reasons can you think of which would make this group of individuals so unconnected to each other?

c. A ***cycle*** in a network is a set of vertices that are connected in a ring; for instance:

A connected to B
B connected to C
C connected to D
D connected to E
E connected to A

where A, B, C, D and E are all different individuals, is a cycle of length 5. Call the cycle a ***proper*** cycle of length 5 if there are no short-cuts "through the middle." In other words:

A does not link to C or D
B does not link to D or E
C does not link to E or A
D does not link to A or B
E does not link to B or C.

What is the longest proper cycle you can find in your immediate Facebook neighborhood? How did you find it?

d. Lastly, try to find the individual in your immediate Facebook neighborhood who has the most Facebook friends in common with you. Who is this individual and how did you find him/her? How many Facebook friends do each of you have, and how many friends do you have in common?