

Networked Life (CIS 112), Spring 2009

Prof Michael Kearns

Homework 1

Issued Feb 4, 2009; Due as *hardcopy* in class Tue Feb 17; don't forget to staple and write your name.

1. Consider a 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 page in which the words appear immediately next 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 "own website" (including the quotation marks), there are about 8 million results returned, so there's an edge between "own" and "website"; "science melodious" has only 2 results and "melodious science" has only 235 results, so there is *not* an edge between "science" and "melodious".

- a. Find the shortest path you can between "epidemic" and "network". Do not go through "a", "an", or "the".
- b. Find the shortest path you can between "connector" and "contagion". Do not go through "a", "an", or "the".
- c. What parts of speech are the intermediate word(s) you used? What kind of words tend to act as Connectors in this network of words?

Now assume that the network is directed, i.e. there is an edge from **A** to **B** if the search for "**A B**" has more than 1,000 results, but there is not necessarily an edge from **B** to **A**. Do not go through "a", "an", or "the". In this directed network, find the shortest paths for:

- d. "network" to "epidemic"
- e. "epidemic" to "network"
- f. "connector" to "contagion"
- g. "contagion" to "connector"

Extra credit: 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/>

2. Experiment with the Networked Life Disease Demo

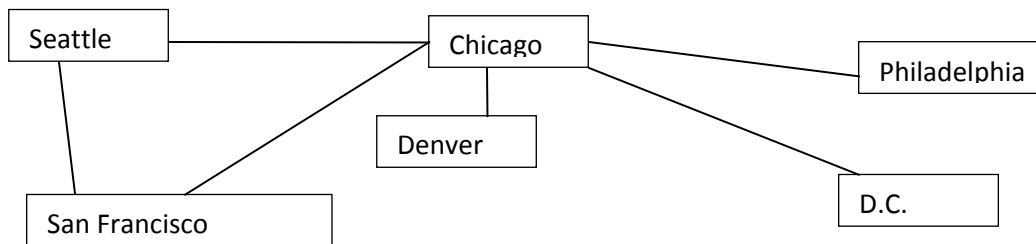
<http://www.cis.upenn.edu/~mkearns/teaching/NetworkedLife/demos/Epidemic.html> .

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

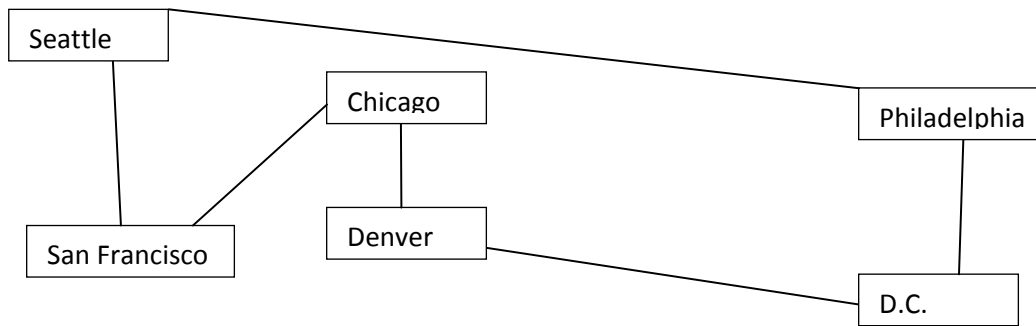
For each infection probability in  $\{0.0, 0.1, 0.2, 0.3, \dots, 1.0\}$ , run three trials of the simulation. For each of those infection probabilities, record the average number of generations your simulations lasted. The number of generations corresponds to how long it takes the disease to die out, 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 number of average generations?
3. Let's say you're creating a new airline company that will serve six cities: Philadelphia, D.C., Chicago, Denver, San Francisco, and Seattle. You can only afford six routes. Your consultants have two competing network designs.

Network A:

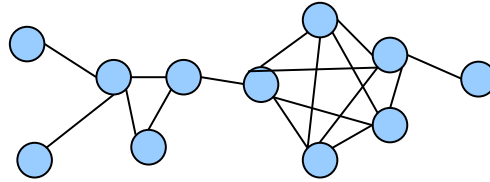


Network B:



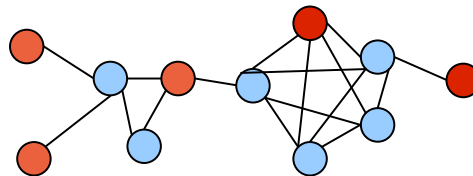
- a. What is the worst-case diameter (# of hops) in each network?
  - b. What is the average-case diameter in each network?
  - c. Both of these networks are connected. There are 15 pairs of cities, and a traveler in each city can fly to each other city. If Chicago's airport is closed, how many of the pairs of cities are still connected in network A? Network B?
4. (2 paragraphs or less) In *The Tipping point*, Gladwell gives some cases where people applied the ideas of the book to real situations (pages 262-264). Choose something (a product, a trend, an idea, a practice, etc.) that you would like to see become more widely used or adopted, and describe how the ideas in the book could be applied to make this happen. Which of the "three rules of epidemics" according to Gladwell (the law of the few, the stickiness factor, the power of context) did you use?
5. Remember that the worst-case diameter of an undirected network is the longest shortest-path distance between any pair of vertices.
- a. Draw a connected undirected network of 10 vertices with the largest possible worst-case diameter. Indicate two vertices that have the maximum shortest path.
  - b. Draw an undirected network consisting of 10 vertices and 9 edges with the smallest possible worst-case diameter. What is the diameter?
  - c. Draw an undirected network consisting of 10 vertices, each of degree 2, with the smallest possible worst-case diameter. What is the diameter?
6. In a computer network, every vertex is a computer and every edge is a network connection. The network map or topology (the vertices and the edges connecting them) is stored in every computer. Messages sent on this network contain the name of the destination computer and are passed from vertex to neighboring vertex until they reach their destination. When a computer receives a message destined for another computer, it looks up the location of the destination in the network, computes the shortest path to the destination using the map of the network stored in its memory, and forwards the message to the next vertex on the path.
- Consider two designs for such a network. The first design uses a small number of hubs connected to each other and most other vertices are connected to one of the hubs. The second design is more distributed with each vertex connected to several other vertices, but there are no vertices of extremely high degree (i.e. there are no hubs). What are the advantages and disadvantages of each design? Consider things such as message routing, cost, susceptibility to failures, congestion, speed of message delivery and anything else you can think of.
7. 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:

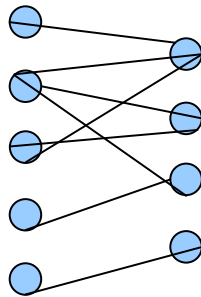
A does not link to C or D  
 B does not link to D or E  
 C does not link to E or A  
 D does not link to A or B  
 E does not link to B or C.

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?

8. Recall that networks can be either directed or undirected. An undirected network is one in which an edge connecting A to B implies the existence of an edge connecting B to A. So the system of intersections (as the vertices) and roads (as the edges) is **not** undirected since there exist one-way streets which are not accompanied by one-way streets going in the opposite direction (a two-way street in a directed network can be thought of as two one-way streets going in opposite directions).

- a. Think of a real life network not discussed in class, the readings, or this homework which is directed. Describe the network precisely. That is, define what its vertices and edges are, and describe why the network is directed rather than undirected.
- b. Do the same for a real life undirected network and describe why the network is undirected. That is, why does the existence of an edge connecting A to B also imply the existence of an edge from B to A?
- c. A “bipartite” network is one in which the vertices can be separated into two groups such that there are no edges linking vertices within the same group. For example, the following network is bipartite:



Describe a real life network which is bipartite. Again be precise in defining what its vertices and edges are. Is your network directed or undirected, and how so?