3/1/10

Networked Life (CIS 112)

<u>Homework 1 Class Solutions</u>

Questions 1,5, and 6 were graded by Mickey Brautbar.

Questions 2,3, and 4 were graded by Kareem Amin.

1. For all parts of Problem 1 there are a large number of possible right answers. One possible set, based on your submitted solutions, is shown below.

    a. "epidemic"—"social"—"voting".

    "social epidemic" has 58,000 results

    "social voting" has 17,000 results

    b. "equilibrium"—"economic"—"contagion".

    " economic equilibrium" has 185,000 results

    "economic contagion" has 17,900 results

    c. "clique"—"news"—"suicide".

    "news clique" has 21,100 results

    "news suicide" has 34,000 results

    d. "obesity"—"family" —"epidemic".

    "family obesity" has 23,000 results

    "family epidemic" has 5680 results

    <u>Points were reduced for the following mistakes:</u>

    - Using a work on the path that is not a noun or an adjective (such as an article) [up to 5 points for each question part].
    - Giving a path of length 4 or more (the shortest path is 2) [up to 4 points for each question part].
    - Not using quotation marks when checking a pair of words "AB" in Google [up to 3 points for each question part].

- Using a phrase instead of a word as vertex on the path [up to -5 for each question part].

2. A common short path provided by many students was: Life->Pattern->System->Network

   Interestingly, an overwhelming majority of paths went through System on the final hop.

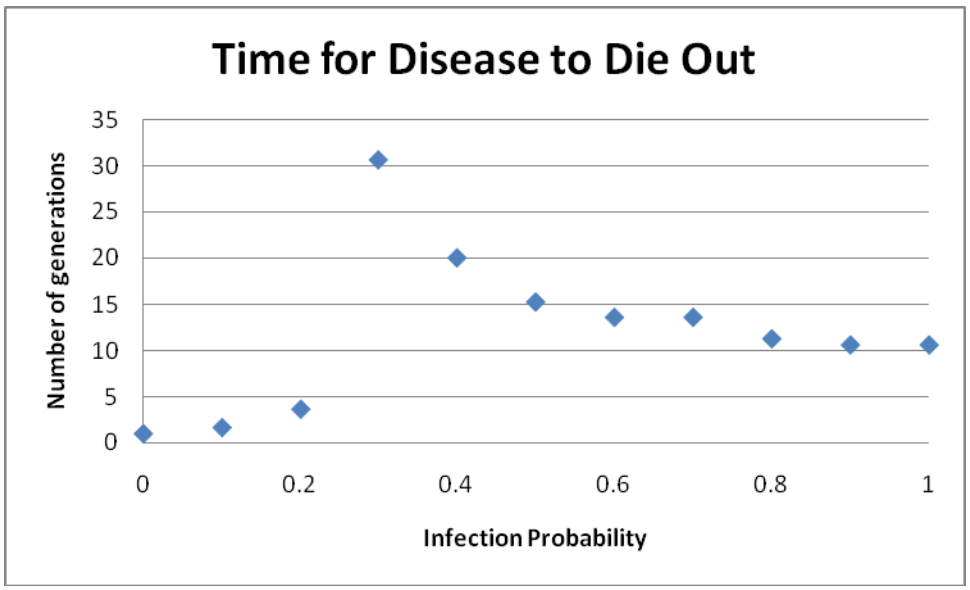
   <u>Points were reduced for the following mistakes:</u>

   - 5 points for finding a path in the wrong direction. Some people made a mistake on this problem by specifying a path from Network to Life. Since the network is directed, it is unlikely that such a path is also a valid path from Life to Network.

   - 1 point for a single invalid hop.

3. A sample set of trials follows below. Obviously randomness is involved, so you shouldn't have gotten the same set of answers, but your plot should have the same shape, with 0.2 or 0.3 (in very rare cases 0.4) as the peak. Note the exponential growth and then decay. Below 0.2, the disease dies out very quickly. Above .3, the disease spreads through the entire population, but kills everyone faster and faster, resulting in slightly fewer generations.

Infection Probability	Trial 1	Trial 2	Trial 3	Average
0	1	1	1	1
0.1	2	1	2	1.666667
0.2	1	2	8	3.666667
0.3	27	35	30	30.66667
0.4	20	20	20	20
0.5	15	14	17	15.33333
0.6	15	13	13	13.66667
0.7	15	14	12	13.66667
0.8	11	12	11	11.33333
0.9	10	11	11	10.66667
1	10	11	11	10.66667



The disease lasts longest when the infection probability is 0.3.

Points were reduced for the following mistakes:

- 2 points for an inconsistent solution to (b). Some people answered part (b) in a way that was inconsistent with their graph in part (a).

- 5 points for a graph in (a) that didn't qualitatively match the graph above (peaking at 0.2-0.4, then dipping and plateauing)

4. Being essays, there is no absolute guide for how they were graded. However, students who received high grades tended to:

a. Cite specific examples demonstrating that they actually read and understood non-trivial amounts of each book.

b. Actually address the questions asked, citing ways in which the books were similar, as well as different.

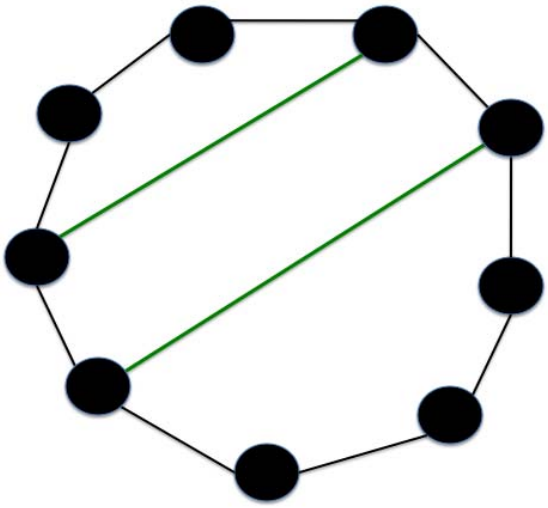
c. Have coherent theses (to the extent that a single-page essay allows).

d. Adhere to the rules of standard written English.

e. Write essays that were fun to read.

An strong essay in no way required all these properties. Property (a) was perhaps the most important.

5 a) We can place the 9 vertices on a cycle network. The clustering coefficient of each of the vertices would then be zero. We can then add two cross edges (colored green in the figure below) to get a network with 11 edges. Note that the network is connected since it is placed on a cycle. Moreover, we never create a triangle in the network so the clustering coefficient is zero for all vertices. Therefore, the average clustering coefficient is also zero.



b) By creating three isolated triangle networks we can make the total clustering coefficient 1. However the question asked us to give a connected network, which also has 11 edges. We can then add two more edges to connect the triangles together to get the final network as in the figure below.

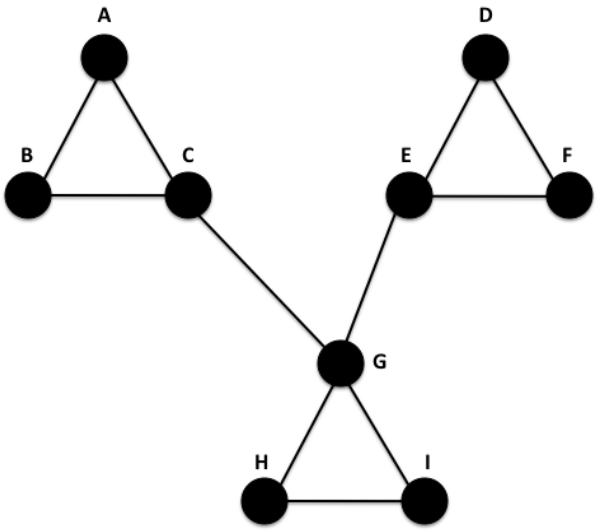
The clustering coefficient values of the vertices are then:

CC(A)=CC(B)=CC(D)=CC(F)=CC(H)=CC(I)= 1/[(2*1)/2=1

CC(C)=CC(E)=1[(3*2)/2]=1/3

CC(G)=1/[(4*3)/2]=1/6

So the average clustering coefficient is (6*1+2*1/3+1*1/6)/9 = 0.759



Points were reduced for the following mistakes:

- Drawing a network with less (or more) than 11 edges [up to -7 points for each question part].
- Not showing/explaining your work [up to -5 points for each question part].
- Wrong calculations [up to -5 points for each question part].
- Providing a network for part (a) that has an average clustering coefficient more than zero [-2 points].
- Providing a network for part (b) that has an average clustering coefficient less than half [-2 points].

6 a) The size of the largest clique I could find was 6. I found it using the Friend wheel Facebook application. I started with a small clique of three friends and kept pulling in Facebook friends who were friends with **all** individuals in the clique created up to that point (if there was even one person in the clique so far that wasn't connected to the pulled individual then I would break the clique structure).

All of the individuals in the final clique are graduate students at Penn studying theoretical computer science. This fact may explain why they are all connected to each other.

An alternative answer: guess and check. Use your knowledge about your friends to come up with a specific set of individuals that should form a clique. Then verify and check, over all pairs of individuals in this group (no exceptions), that they are indeed connected.

b) The largest independent set I could find had 20 people. I found it using the Friend wheel. I started with two friends that do not know each other and each time kept pulling in a new Facebook friend who wasn't friend with any of the individuals in the independent set created up to that point.

[Note: if there was even one person in the independent set so far that was connected to the pulled individual then I would break the independent set structure].

Each of the individuals in this independent set belongs to a different category (family, work, high school, college, etc.) which may explain why this group is so unconnected to each other.

c) The longest proper cycle I could find was of length 5. I again used the Friend wheel app and leveraged the fact that an elementary school friend knows people I went to college and high school with. The proper cycle then looks like:

college friend A— college friend B— college friend C—high school friend D— elementary school friend E— college friend A

Dragging all these 5 vertices aside in the Friend Wheel app allowed me to verify that the cycle is indeed proper and doesn't contain a shortcut inside it.

[Note: if I had included myself as one of the links of the cycle, then it would be impossible for me to form a proper cycle of length greater than 3. This is because I am friends with everyone in my immediate Facebook neighborhood and thus always create a "shortcut."]

d) The individual who shares the most Facebook friends with me is another computer science graduate student. We share 16 Facebook friends. I used the Friend wheel app to find the friend that shares the biggest number of friends with me. This person has 410 friends on Facebook while I have 82.

Points were reduced for the following mistakes:

- For part (a) – Not explaining / showing the group of individuals you find is indeed a clique, that is, **all** the connections between any two group members exist [up to 3 points].
- For part (b) – Not explaining / showing the group you find is indeed an independent set, that is, **no** connection between any two group members exist [up to 3 points].
- For part (c) - Providing an illegal proper cycle [up to 2 points].
- For part (c) - showing/explaining why the cycle you found is indeed proper [up to 2 points].
- For part (d) - giving only partial statistics (for example not including the number of friends you have in Facebook) [up to 3 points].
- Not explaining how you came up with the requested group (for example a clique) [up to -4 points].