CSE 112: Networked Life
Problem Set 1

Due Tuesday, February 7, Start of Class

1. For each of the following general categories of networks discussed in class, give a real-world example of such a network other than those given in class. You should be as clear as possible about what the vertices (nodes) are and about what relationship between two vertices determines whether or not there is an edge (link) between them. For each of your networks, comment in a sentence or two on its static structure, its dynamics, and its formation process:

   (a) Social Networks

   (b) Content Networks

   (c) Business or Economic Networks

   (d) Physical Networks

   (e) Biological Networks

2. In class we mentioned that random connectivity (e.g. in which each possible edge appears randomly and independently with some probability) is not a good fit for the social networks that arise in many natural social networks, such as those of the Facebook, mathematical collaboration, and real-world acquaintanceship. In what ways do such random connectivity models fail to explain real social networks? (You may cite both properties mentioned in class, as well as any others you think apply.) What model of network generation might be a better fit? You can explain your model mathematically or in English, but be precise either way.

3. Think about a fad or trend you have observed. This could be a large scale nation wide trend or just something that was popular in your high school or dorm, but should not be a trend that was mentioned specifically in The Tipping Point or in class.

   (a) Briefly describe this trend.

   (b) Who were the connectors, mavens, and salesmen who caused this trend to spread?

   (c) What made this trend stick?

   (d) Was the spread of this trend gradual and steady or did it exhibit tipping phenomena. Why do you think this is the case?

4. Complete the following tasks in the order in which they are listed.

(a) List five entities (e.g. people, departments, universities, programs, companies, locations, etc.) that you believe are related to Penn's School of Engineering and Applied Science (SEAS) and call this the "My SEAS" list.

(b) Now run a Google search for "related:www.seas.upenn.edu" – This should give a ranked list of entities that Google thinks are "related" to SEAS. List the top five results. We'll call this "Google SEAS" list.

(c) Are there any entities in the "My SEAS" list that do not appear in "Google SEAS" list? If so, briefly state why you included these entities in your list.

(d) Now, list five entities that you think are related to Penn's Wharton School and call this "My Wharton" list.

(e) Run a Google search for "related:www.wharton.upenn.edu" – This should give a ranked list of entities that Google thinks are "related" to Wharton. List the top five results. We'll call this the "Google Wharton" list.

(f) Are there any entities in the "My Wharton" list that do not appear in the "Google Wharton" list? If so, briefly state why you included these in your list.

(g) Compare the "Google SEAS" and "Google Wharton" lists. Do you see any difference in Google's concept of "related" in the two cases? Very briefly give an explanation for this. What is/are the concept(s) of relatedness that Google is using?

(h) Visit the Google TouchGraph tool at `http://www.touchgraph.com/TGGoogleBrowser.html`. This tool provides a graphical method of viewing websites that are related according to Google. Enter the SEAS website as a start URL. Can you find a site within three or four steps of SEAS that does not seem related to SEAS at all? What is the path from SEAS to this site?

5. For this problem you will need to have an account on the Facebook. For those of you who do not already have one, you can obtain one by going to `www.thefacebook.com` and following the sign-up instructions. Note that in order to complete this problem, there is no need for you to create a publicly viewable profile, nor to have any connections (friends).

   (a) Find the individual in the Penn community Facebook with the largest number of friends (highest degree) that you can. You can use any method that you like, including those that use information external to the Facebook. Write the name of the individual and the value of their degree, and briefly describe your method of finding them.

(b) Repeat part a, but instead finding the individual with the lowest degree (other than yourself) . Again write the name of the individual and their degree, and describe your method of finding them.

(c) By using the Facebook's search facility, and using the "Search All Fields" option, you can find all users who have a chosen word, name, phrase, etc. in their profiles. Pick any phrase such that the resulting search produces at least 30 and at most 100 individuals. Then generate a histogram of the degrees of these individuals; you may use Excel or any other plotting software that you like. Compute and provide the minimum, maximum, and average degrees of this population.

(d) Discuss the extent to which your degree histogram from part c does or does not seem to exhibit the properties discussed in class for such distributions in natural social networks.

(e) (Extra Credit) 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 A

where A, B, C and D are all different individuals, is a cycle of length 4. Note that there may or may not be "chords" that cut across this cycle (e.g. B and D might be directly connected). Find the longest cycle you can in the Facebook of length at least 3, and write down the names of the individuals in the order they appear in this cycle. Describe your method for discovering this cycle.

6. (a) Draw an example of an undirected network of 10 vertices in which every vertex's degree is exactly 3, and the diameter (maximum shortest-path distance between any pair of vertices) is as small as you can make it. Indicate the diameter, and mark the two vertices farthest apart in your graph.

   (b) Draw an example of an undirected network of 10 vertices in which every vertex's degree is exactly 3, and is colorable with only 2 colors. Annotate your graph with the proper coloring.

7. This question is about the graph coloring experiments you participated in on January 24 and/or 25. In each experiment in those sessions, you collectively attempted to color a graph whose underlying structure was either fully or only partially revealed to you. While there were many factors that may have contributed to which experiments were relatively "easy" or "hard" for the group (including the information condition and incentive scheme), one of these factors was certainly the actual structure of

the graph. For each of the potential structural properties of graphs given below, indicate whether you think the property would generally make such graph-coloring problems for human subjects harder or easier, and why.

(a) a very dense graph (many edges)

(b) a very sparse graph (few edges)

(c) presence of “connectors” (high-degree vertices)

(d) highly regular “local” structure (e.g. grid, cycle)

(e) random connectivity pattern

(f) small diameter