Homework 1
Networked Life (NETS 112)
Fall 2019
Prof Michael Kearns

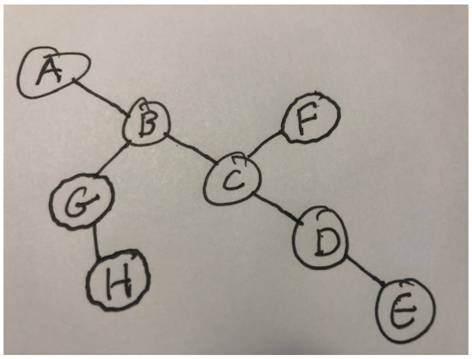
Posted September 29, 2016. Due in hard-copy format at the start of lecture on Tuesday, October 15. Please don't forget to write your name and staple the pages together.

Collaboration of any kind is NOT permitted on the homework.

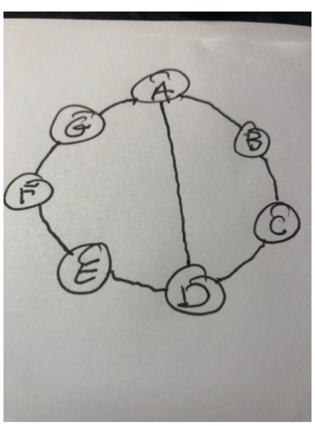
Your Name:

Problem 1. For each of the following networks, compute the worst-case diameter, the average-case diameter, the edge density (number of edges divided by the maximum possible), and plot the degree distribution. Show your work where appropriate.

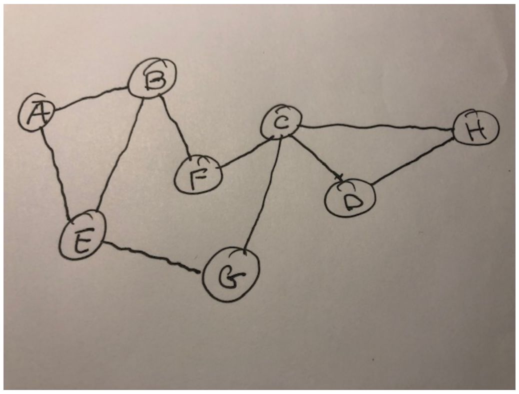
(a)



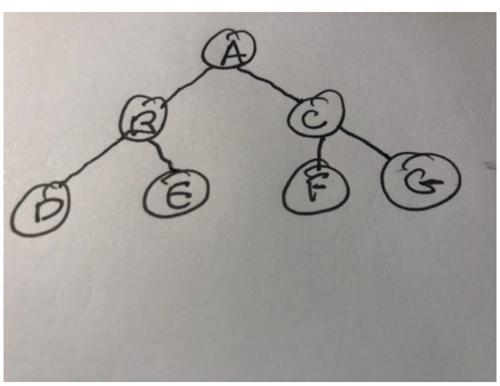
(b)



(c)



(d)



Problem 2. A *cycle* in a network is a sequence of vertices v1, v2, v3…,vn, v1 such that each vertex is distinct (other than the first and last), and each adjacent pair in the sequence are neighbors (i.e. v1 and v2 are neighbors, v2 and v3 are neighbors, etc. and vn and v1 are neighbors). Since we start at v1 and end at v1, and all vertices in between are distinct, we've taken a tour of the vertices in the sequence. Note that other edges may be present as well (e.g. v1 and v7 might be neighbors). As an example, in the network for problem 1(c), the sequence A, B, F, C, G, E, A is a cycle of length 6 edges.

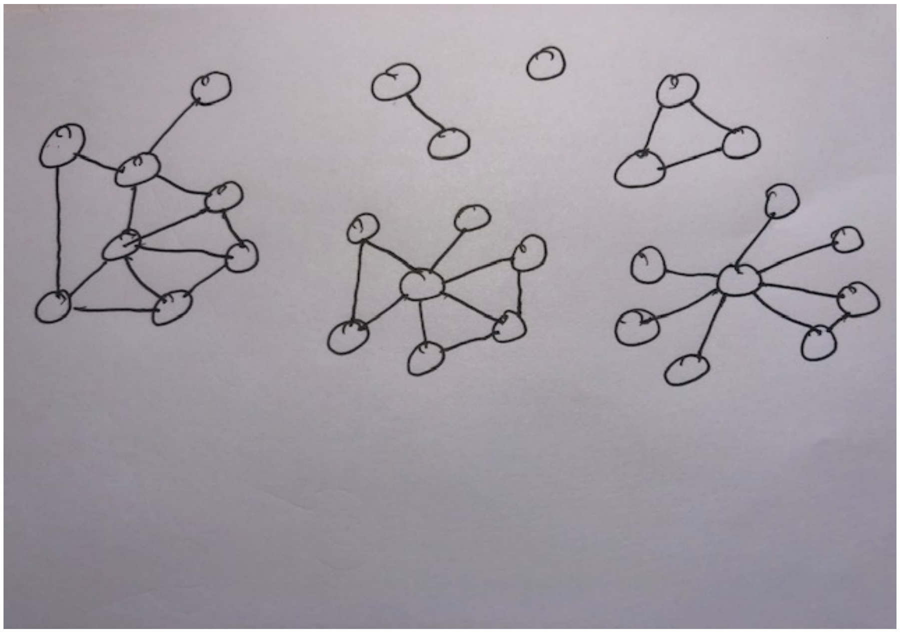
We can also consider cycles in the directed network of the Web. For instance, the following sequence of URLs:

https://www.cis.upenn.edu/~mkearns/
https://www.cis.upenn.edu/~aaroth/
https://www.cse.huji.ac.il/~katrina/
https://www.cse.huji.ac.il/~katrina/research.html
http://www.cs.cmu.edu/~avrim/
http://hunch.net/~jl/
http://hunch.net/~jl/projects/projects.html
http://hunch.net/~jl/projects/RL/RL.html
https://www.cis.upenn.edu/~mkearns/

is a cycle of length 8: each of the pages above has a link to the next one in the sequence, and we end up back where we started. Note that the hard part of finding a long cycle is not simply finding a long list of links; it's then being able to find your way "home".

Using any methods you like, find the longest cycle you can in the Web, and list the corresponding sequence of URLs. Carefully describe your methodology. We will spot-check for correctness of your cycles (i.e. that each page links to the next). Rewards of some kind will be given for the longest cycle found, and for the most creative/interesting. Try to avoid making the problem easy or trivial by having repeated links from the same top-level domain (e.g. by having all of your links begin with www.cnn.com).

Problem 3. Consider the following network, which consists of multiple connected components:



(a) Consider the process in which a vertex v is chosen at random to be infected, and the infection then spreads deterministically to kill all vertices in the connected component of v. What is the expected or average number of vertices killed for this network? Show your work.

(b) Suppose you are allowed to "immunize" exactly one vertex that can no longer be infected, and cannot infect others. Which vertex would you choose to make the average number of vertices killed as small as possible? What would the new value for this average be? Show your work.

(c) Suppose you are forced to add an edge between vertices in different connected components. Which edge would you choose to make the average number of vertices killed as small as possible? What would the new value for this average be? Show your work.

Problem 4. For each of the following parts, carefully draw an undirected network with the specified properties.

(a) Draw a network with three connected components of sizes 7, 4 and 3 vertices, in which the total number of edges is 12 and the worst-case diameter (longest shortest-path distance) of the largest component is 5.

(b) Draw a network with three connected components of sizes 5, 4 and 4 vertices, in which the total number of edges is 19 and the worst-case diameter of the largest component is 1.

(c) Draw a network with three connected components of sizes 12, 5 and 3 vertices, in which the total number of edges is 17 and the worst-case diameter of the largest component is 7.

Problem 5. In this problem you are asked to carefully define, create and visualize a network of your own choosing, ideally reflecting a personal interest. You should begin by carefully specifying what the vertices are in your network, and stating exactly what constitutes an edge between two vertices. As long as you're precise, your network can be about any topic you like; the vertices don't have to correspond to people at all. As one example, the vertices in your network could represent your favorite musical artists (people or bands), and the edges could represent pairs of artists who have collaborated on a song, for some precise definition of collaboration.

Carefully define your vertices and edge definition, making sure your network has at least 15 vertices and 20 edges. Specify any research or data sources you use to identify the edges in your network. Draw a clear visualization of your network, showing all vertices and edges. Compute the worst-case diameter, edge density and degree distribution for your network.

Problem 6. For this problem, you will need a standard 6-sided die in order to make random choices. For each of the following networks, you are to repeat the following experiment **10 times**:

- Use your die to choose randomly choose a "start" vertex in the network.
- From that start vertex, roll your die to pick a random neighbor of your start vertex, and "move" to the chosen vertex.
- Repeat this process for a total of **15 steps**, always choosing a random neighbor of your current vertex and moving to that neighbor.

Note that if your current vertex has, say, 3 neighbors, you can just use the 1, 2 and 3 faces of your die to pick your next move. Also, if your current vertex has only one neighbor, you don't need to roll your die, your next move will simply be to that neighbor.

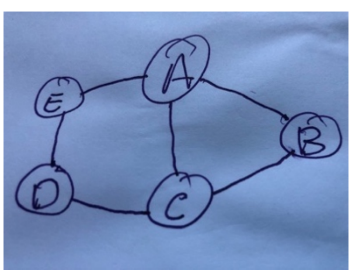
As an example, a random walk in the network for part (a) below might look like the following sequence of 15 steps:

D, E, D, C, A, B, C, B, C, D, E, A, C, D, E

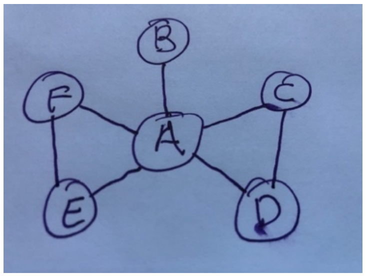
We will say that this sequence **terminates** at vertex E.

For each of the networks below, perform 10 separate 15-step random walks, and explicitly write out the resulting sequence of vertices visited. Then annotate each vertex in the network diagram with the total number of times that a sequence terminated on that vertex.

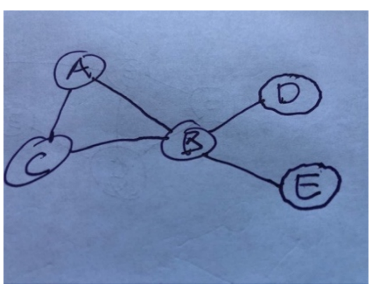
(a)



(b)



(c)



(d) Based on the counts of terminal vertices, do you see any pattern or relationship between the terminal counts and the network structure? Can you guess how the counts would compare to each other across vertices if you were to perform an infinite number of random walks?

Problem 7. Write a brief (say about a page) essay summarizing and synthesizing the readings and lectures on any one of the four course “modules” so far other than the introductory material (i.e. one of “Structural Properties of Networks”, “Contagion in Networks”, “Machine Learning and Social Networks”, and “Navigation in Networks” on the course schedule).