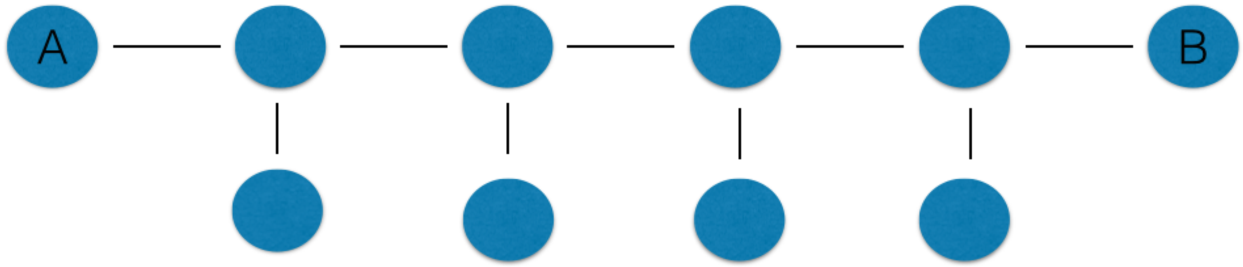
1. (Rohan) 10 points: (2 points awarded for meeting each of the 5 requirements)



2. (Chris) 15 points: The goal was to compare the (a) motivations, (b) main results, and © limitations of the three articles in the “Contagion in Networks” section of the course page. The problem asks that you not simply summarize the content. If an effort was made to find meaningful connections between the three articles in each of the three areas above, then full credit was given.

3. (Chris) 15 points: (2 points awarded for correct answers to parts a-c)

d. 3 points: Points awarded for reasonable methodology that comes up with a very low estimate. Example method: estimate the average number of edges per node and multiply by estimate of number of nodes/2 to get estimate for number of edges. Then divide by the estimated number of nodes choose 2.

e. 3 points: Answer should be about 2 times the largest Erdos number. Since any node can reach any other node by going through Erdos.

f. 3 points: Erdos’ Graham number is 1. Thus, any mathematician’s Erdos number and Graham number differ by at most 1.

4. (Rohan) 15 points: Answers should compare/contrast to the models studied in class (e.g. the forest fire contagion model assumes every neighbor is infected, which isn’t true of viral marketing campaigns), describe concrete examples of viral marketing, explore the algorithmic/computational challenges involved (e.g. finding seed nodes, quantifying ‘stickiness’ of an advertisement) and cite sources. More generally, responses should demonstrate understanding of viral marketing techniques as they relate to the study of networks.

5. (Chris) 15 points: Except for a few exceptions, there were three correct answers:

a. It was mentioned in class that a tree is the densest packing of nodes of a given degree. In this case, we would like a binary tree with a largest shortest distance of 33. This requires a depth of about 16. Thus, the estimate for the largest possible number of nodes is on the order of 2^16.
b. In the notes you can find the formula N < \delta^D, which gives an upper bound of 3^33.
c. By doing some independent research you can find Moore’s formula which gives an upper bound for the number of nodes in a network.

6. 15 points:

a. (Rohan) In general, random routing would be the least effective for two or more densely connected parts of the network that are linked by sparse bridges (think of the social network that includes only people in Philly, Chicago and LA - how are they within cities vs. between cities?). Messages would have a hard time finding the bridge vertex and then traversing the bridge into another area of the network, and would tend to get 'stuck' near the source node. A common answer was that a network with one densely connected component would make random forwarding difficult, but this is actually the best scenario for which we could randomly route a message through a network with small diameter. In the extreme case, consider a complete graph - at each step, the message will have a 1/(n-1) chance of finding its target, so we expect that it will take roughly ~n steps for a message to travel from source to target (see expectation of a geometric random variable for more insight).

b. (Chris) The random forwarding will work best if there is one giant component without any "bridges" separating any group of nodes from another group. For example, the network of math coauthorships has this geometry.

7. (Rohan) 15 points (7 for histogram, 8 for connection to Where's George study): Histograms should demonstrate a reasonably random sampling methodology and clearly labeled axes. Ideally these histograms should have the distance from Philly on the x axis and the # of friends at each distance on the y axis to model the distribution of friends as a function of relative location. You should then compare and contrast your findings with the exponential decay observed in the Where's George study. If your data matches the distribution, explain why that is (e.g. the movement of dollar bills is a good measure of human travel which in turn might be related to the distribution of a person's social network). If it doesn't, hypothesize as to why that might be (e.g. you are from the west coast and setting LA as the origin might have yielded a more obvious heavy tailed distribution).