MIDTERM EXAMINATION
Networked Life (NETS 112)
October 1, 2013
Prof. Michael Kearns

This is a closed-book exam. You should have no material on your desk other than the exam itself and a pencil or pen. If you run out of room on a page, you may use the back, but be sure to indicate you have done so.

Name: ____

Penn ID: ____

Problem 1: ________/10

Problem 2: ________/20

Problem 3: ________/20

Problem 4: ________/10

Problem 5: ________/15

Problem 6: ________/25

TOTAL: _________/100

Problem 1 (10 points) Answer "True" or "False" for each of the following assertions.

a. The largest possible number of edges in an undirected network of N vertices grows roughly as N^2/2.
True. The largest possible number of edges is N(N-1) / 2 ~ N^2 / 2.

b. The distance between any two vertices is always less than the average-case diameter.
False. Note that the question specified "average-case" diameter.

c. Typically, in large-scale social networks, there are a reliable number of vertices with degree much higher than the average.
True. Most large-scale social networks exhibit a heavy-tail degree distribution.

d. If we remove an edge from a graph, the diameter might decrease.
False. Removing an edge can only increase the distance between two vertices.

e. In the forest fire demo, the more vertices we delete from the grid, the greater the fraction of the grid that will burn.
False. Deleting vertices turns them into parking lots, which decreases the extent of contagion.

f. In the viral spread demo, an infected vertex will always pass the infection to each of its neighbors.
False. Not always -- depends on the value of the "stickiness" parameter.

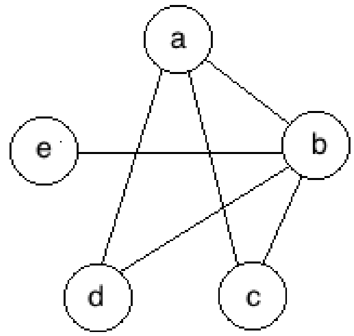
g. In the altruistic contagion model, if we add an edge between vertices A and B, the equilibrium wealths of vertices A and B will increase, and the equilibrium wealths of all other vertices will decrease.
True. The degree of A and B will increase, which increases their wealth. The sum of all degrees also increases (i.e. the denominator term), which decreases everyone else's wealth.

h. As long as short paths exist in a network, it is possible to solve the navigation or "small world" problem.
False. The existence of short paths is necessary, but not sufficient. An algorithm still has to be able to find those short paths.

i. It is impossible for a network to have 20 vertices, a maximum degree of 2, and a diameter of 3.
True. 20 > 2^3. Note that the question said "impossible."

j. In Kleinberg's model, we consider navigation to be efficient if the number of steps required is about the square root of N.
False. The number of steps needs to be about log(N) or less.

Problem 2 (20 points) Consider the network below.



Graph G

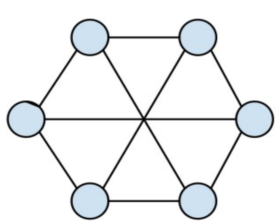
a. (5 points) Calculate the average-case diameter, which is defined as the average distance between pairs of vertices.

Sum of distances = 14
Number of pairs = 10
Diameter = 14/10 = 1.4

b. (5 points) Calculate the equilibrium wealth of each vertex in the altruistic contagion model considered in class.

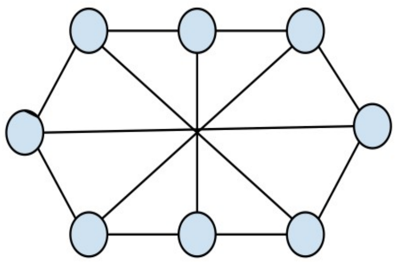
Wealth = (degree(vertex) / sum of all degrees) * total wealth
Sum of all degrees = 12
Total wealth = 5
Wealth(a) = 3*5/12
Wealth(b) = 4*5/12
Wealth(c) = Wealth(d) = 2*5/10
Wealth(e) = 1*5/12

c. (5 points) Draw a network with 6 vertices in which the maximum degree is 3 and the worst-case diameter is 2.



Partial credit was given for networks which met at least two of the requirements.

d. (5 points) Is it possible for a network to have more than 6 vertices and still have a maximum degree of 3 and worst-case diameter of 2? If so, draw such a network.



Partial credit was given for showing some work.

Many students applied the equation $N \leq \Delta^D$ to show that this should be possible ($6 \leq 3^2$), but this equation doesn't actually apply, since the question was asking about worst-case diameter. The equation tells you that an average-case diameter of 2 is possible, but this does not imply that a worst-case diameter of 2 is also possible.

Problem 3 (20 points)

a. (5 points) In the forest fire demo, what is the parameter that we varied between the different trials? What is the effect of this parameter on the extent of contagion?

We varied a *connectivity* parameter that determines the probability that a vertex will be forest. The higher this parameter, the greater the contagion. When we increase the probability that a vertex will be forest, we increase the expected size of a vertex's connected component, and therefore increase the expected amount of the grid that will burn.

b. (5 points) In the viral epidemic demo, what are the two parameters that we varied? What is the effect of each parameter on the extent of contagion?

We varied a *connectivity* parameter that determines the probability of rewiring local connections to random long-distance connections, and we varied a *stickiness* parameter that determines the probability that an infected vertex will pass on the infection to its neighbors. The higher the stickiness parameter, the greater the contagion – this relationship is fairly straightforward. Increasing the connectivity parameter also increases the contagion, to an extent – ideally you want some, but not all, connections to be rewired. This creates the necessary mix of local and long-distance edges.

c. (5 points) What do we mean by the "tipping" or "threshold" phenomenon that was observed in these demos?

We observe a tipping phenomenon when there is some value q of a parameter such that, below this value, contagion is very limited / contained, and above this value, contagion is nearly complete.

Some students stated that the tipping point is the point at which contagion is nearly complete. This is almost, but not quite, correct. The key point is that there is an *exponential increase* at the tipping point, rather than a *linear* increase. We would *not* observe a tipping phenomenon if contagion increased steadily and eventually affected the majority of the network.

d. (5 points) Compare and contrast the viral epidemic demo with Kleinberg's model. Consider both the network formation processes, and the network dynamics (i.e. the way in which information spreads through the network). Comment on the extent to which there is a "purpose" for the participants each dynamic.

Network formation: In the viral demo, we *rewire* edges from one vertex to a *random* destination vertex. In Kleinberg's model, we *add* edges from one vertex to a destination vertex based on the *distance* between the two vertices. Both models involve creating long-distance edges, but the main differences are the 1) rewiring vs. adding, and 2) distance is a factor in Kleinberg, but not in the viral epidemic.

Network dynamics: In the viral demo, contagion spreads *probabilistically* through the network. Vertices have no "purpose" or "choice" – they *passively* spread the contagion to their neighbors. In Kleinberg's model, information is spread with the purpose of reaching a target. Each vertex *actively* chooses a neighbor (the one geographically closest to the target) to pass the message to. Here, dynamics are controlled by an algorithm, whereas in the viral demo, dynamics are purely probabilistic.

Problem 4 (10 points)

In The Tipping Point, Gladwell gives some cases where people applied the ideas of the book to real situations. 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?

Most answers that demonstrated an understanding of the reading were given full credit. Answers that did not use ideas from the book, but rather from class, were given partial credit, based on how “reasonable” they were.

A surprising number of students chose the payments app Venmo (created by Penn grads) as the product they would like to see become more widely used. I've never tried it before, but these students were quite effective salesmen, and convinced me to download it. I suggest that the entire class do so, and we can leverage the law of the few to turn Venmo into the next big thing!

Problem 5 (15 points)

Consider the in-class biased voting experiments.

High level note – some of you were not present in class for these experiments, and had trouble with this question. Note that summary slides were posted on the website, and you should have been able to answer (at least parts b and c) based on those slides. Class attendance is mandatory, and you are responsible for anything that happens in class, even if you are absent.

a) (5 points) Summarize the rules and incentives or payoffs for these experiments.

The goal of the class is to converge on one of two colors. Each student is given a preference color. If a significant majority of the class chooses one color, then anyone who chose that color gets points, as follows: if the color is your preference, you receive 2 points, otherwise 1. Anyone who chooses the minority color gets 0 points. If the class does not converge, no one receives points.

b) (5 points) In one of the experiments, the network was defined by connections between students of the same gender, and then there were 5 students who had connections to everyone. How did the existence of these connectors affect the dynamics of the experiment? What was the outcome of this experiment?

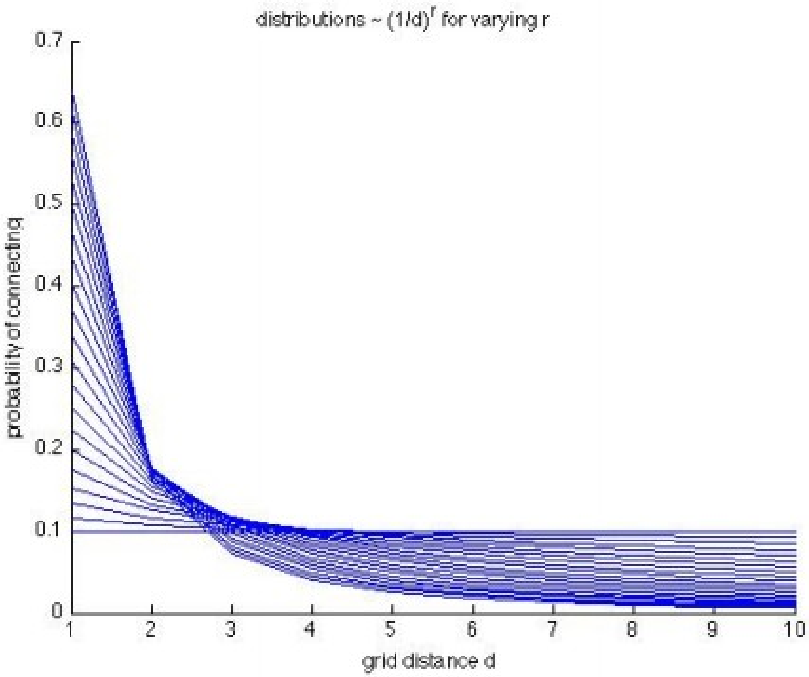
The connectors had a position of "power" or "influence" because they could communicate with everyone, and therefore had access to the most information. They facilitated the transfer of information between the two gender components, and therefore helped the class reach a consensus. Four of the five connectors had the same preference color, and managed to convince the class to converge to that color.

c) (5 points) In another one of the experiments, the network was defined by connections between students who were born in adjacent months. Furthermore, color preferences were chosen so that a student's preference was different from that of his neighbors. How did this network structure and preference rule affect the dynamics of the experiment? What was the outcome of the experiment?

The set-up of this experiment – namely, the lack of connectors / long-distance edges – made it very difficult for different parts of the network to communicate with each other. Information remained local, contained in "month" clusters, rather than spreading. In the end, most students simply chose their preference color, and consensus was not reached.

Problem 6 (25 points)

The image below is taken from the slides and illustrates the effect of *r* on the networks generated by the Kleinberg model.



a. (4 points) What is the main point that this graph is making about the effect of *r*?

Each line represent a different value of *r*. The flat line at 0.1 represents $r = 0$. As *r* increases, the lines get successively steeper. This illustrates the fact that, for small values of *r*, distance does not make a large difference in the probability of connection. We are about equally likely to make a long-distance connection as a local connection. But at large values of *r*, distance makes a big difference, and we are much less likely to make long-distance connection as compared to a local connection.

Many students thought that the graph was illustrating the fact that $r = 2$ is the only value that permits efficient navigation. Although this is a true fact, it is *not* illustrated by the graph. The graph is not plotting anything related to the efficiency of navigation.

b. (6 points) If *r* is too large, why might efficient navigation be difficult?

At large values of r, we are more likely to add local edges, rather than long-distance edges, to the network. The lack of long-distance edges makes it difficult to make “large hops” across the network. Indeed, short paths between vertices may not even exist.

c. (6 points) What about if *r* is too small?

At small values of r, we are equally likely to add long-distance edges as local edges. So there are a fair number of long-distance connections, and short paths do exist in the network. However, it may be hard to find these short paths using only local information. For instance, maybe we take a big hop towards the target, and end up in its general vicinity. But the short path requires us now to take a big hop *away* from the target, in order to get back to it. Since we only have local information, we don't know this. So instead, we take lots of small steps to get the rest of the way there, and this is inefficient.

Suppose we are generating a network in Kleinberg's model, and we are about to add a long distance edge to vertex u. The distance between vertex u and vertex v is 5, and the distance between vertex u and vertex w is 10. Let p denote the probability that we add the u-v edge, and let q denote the probability that we add the u-w edge.

d. (3 points) If $r = 0$, what is numerical value of the ratio p/q?

p ~ (1/5)^0
q ~ (1/10)^0
p/q = 1

e. (3 points) What is the ratio if $r = 1$?

p ~ (1/5)^1 = 1/5
q ~ (1/10)^1 = 1/10
p/q = 2

e. (3 points) What is the ratio if $r = 2$?

p ~ (1/5)^1 = 1/25
q ~ (1/10)^1 = 1/100
p/q = 4