SOLUTIONS for MIDTERM EXAMINATION
Networked Life (NETS 112)
October 21, 2014
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: _________ **/10**

Problem 3: _________ **/10**

Problem 4: _________ **/15**

Problem 5: _________ **/10**

Problem 6: _________ **/15**

Problem 7: _________ **/10**

Problem 8: _________ **/20**

Total: _____________ **/100**

Problem 1 (10 points: Graded by Ryan) Indicate whether the following statements are *True* or *False*.

(a) Any planar graph can be colored using at most 3 colors.
F

(b) If a social network permits efficient navigation from only local, distributed information, it necessarily has small diameter.
T

(c) PageRank is the single most important element in determining relevance in Google's search engine.
F

(d) Prisoner's Dilemma has exactly one pure-strategy Nash equilibrium.
T

(e) As long as the housing preferences of each individual are mild, there is a global solution in which everyone is happy.
F

(f) Klienberg's navigation model might appeal more to mathematicians, and the Watts, Dodds, Newman navigation model might appeal more to sociologists.
T

(g) The Backstrom et al. Facebook diameter study showed strong evidence of increasing diameter over time.
F

(h) Having a heavy-tailed degree distribution is a monotone network property.
F

(i) If the fraction of people who want to participate next time increases with the fraction participating this time, the only equilibrium is 100% participation.
F

(j) Each time we experimented with tennis ball navigation in networks, the class found a near-shortest path.
F

Problem 2 (10 points: Graded by Shahin) Compute the clustering coefficient for the network given in Figure 1. Be sure to write the clustering coefficient for each node along with your work.

Solution. We have

$$cc(A) = 0 \quad cc(D) = 2/3 \quad cc(E) = 3/10$$

$$cc(B) = cc(C) = 1/3 \qquad cc(G) = cc(F) = 1$$

So the clustering coefficient of the whole network is $\frac{0+2/3+3/10+2/3+2}{7} = 109/210$.

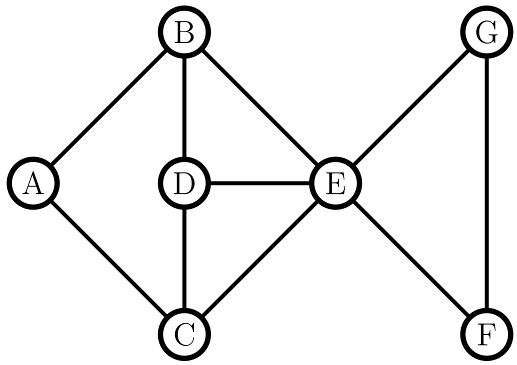

Figure 1: Compute the Clustering Coefficient for this Network

For grading: one point for the clustering coefficient of each node and 3 points for the clustering coefficient of the network. □

Problem 3 (10 points: Graded by Shahin) Consider the network in Figure 2. Determine the nodes with the smallest and largest rank after running the PageRank algorithm on the network. Briefly justify your answer.

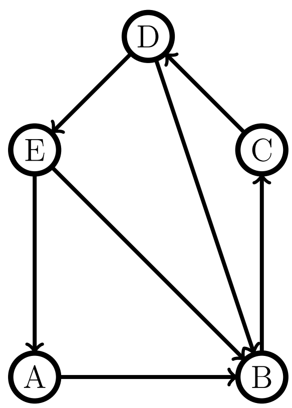

Figure 2: Page Rank

Solution. Node B has the largest page rank because it gets in-flows from nodes A, E and D. A has the smallest page rank because its in-flow comes from the stream connecting D to E and then A and half of this flow is already going to B in each step before getting delivered to A.

Common mistake 1: B, C and D all have the same page rank. (I deducted two points for this) Although their page ranks are close to B, but C and D does not have the same page rank as B. C and D do not have the same in-degree compared to B. Furthermore, although their getting a flow from B but as we are getting further away from B, the page rank decreases. So $B > C > D$ if we sort by page rank.

Common mistake 2: E and D have the smallest page rank because they have the largest out-degree. Page rank does not only depend on in and out degree. It also depends on what other nodes a node is connected to and how important those nodes are (where importance means having a high rank). I gave no partial credit for this answer. □

Problem 4 (15 points: Graded by Ryan) For this problem, consider the online graph coloring experiments you participated in.

(a) (5 points) Briefly but precisely describe the graph coloring problem.

(b) (5 points) In class there was a discussion of strategies employed by students in the class. Briefly but precisely describe at least two different strategies that were articulated.

(c) (5 points) Recall that for each network, there were additional points awarded for the three fastest completion times. Discuss the distribution of these additional points across the class, including a comparison to the distribution expected if all students were equally skilled at graph coloring. Based on this distribution, do you think graph-coloring ability would be better modeled as a normal or heavy-tailed distribution?

Solution.

(a) We are given a graph G with nodes V and edges E. We want to color each node of the graph such that no two nodes of the same color share an edge. We want to do this in as few colors as possible, or color each node from a predefined set of colors as the web app had.

(b) Color a high degree node one color, say Red, and try to color as many of its neighbors the same color that is different from Red. Find triangles and color each node of the triangle differently. Try to partition the nodes into different sets so that each set has no nodes that share an edge, then color each set differently.

(c) We consider the Ranked Points Total plot given in lecture. In this plot, the blue graph represents what would happen if everyone had the same coloring ability and completion times for everyone was normally distributed around the same mean for everyone. The red represents the real data. Note that the top ranked students were several times higher than what we would have expected from the blue plot. Recall that the top ranked student got 26 points and the points dropped pretty quick after that.

This problem involved looking at the points distribution and the ability distribution. You were asked to conclude how the ability distribution looked given how the points distribution looked. From the lecture slides, you are only given the ranked points distribution plot. From this, you can tell that the points distribution has a heavy tail because few people got way above zero medals.

In order to see a points distribution with a heavy tail, we may expect there to be just a few people that are slightly above average at graph coloring ability. Note that we do not need someone many many times above average in coloring ability for her to accrue a majority of the points, but merely slightly above average. We are not asking about the finish times distribution but the points distribution. This would be an arguement for why a *normal* ability distribution would make sense, because maybe a student is slightly above average at skill and so would gain a majority of the points every time.

You could also argue that the ability distribution could be heavy tailed. This would mean that there are very few people that are way above average in coloring ability. Hence, of course those people would get all the medals, so everytime there would be always the same awesome colorer getting gold, then the second getting silver and third getting bronze for every graph. In our data we did see people get a majority of the points, but more than 3 people got additional points. This may be a reason against ability being heavy tailed.

This problem was graded *very* leniently. If you were clear what distribution you were talking about (points or ability) then you typically got full credit.

□

Problem 5 (10 points: Graded by Ryan) Consider the curve in Figure 3 where the horizontal axis shows the percentage of the students who attend today's class and the vertical axis shows the expected percentage of student who will attend the next class.

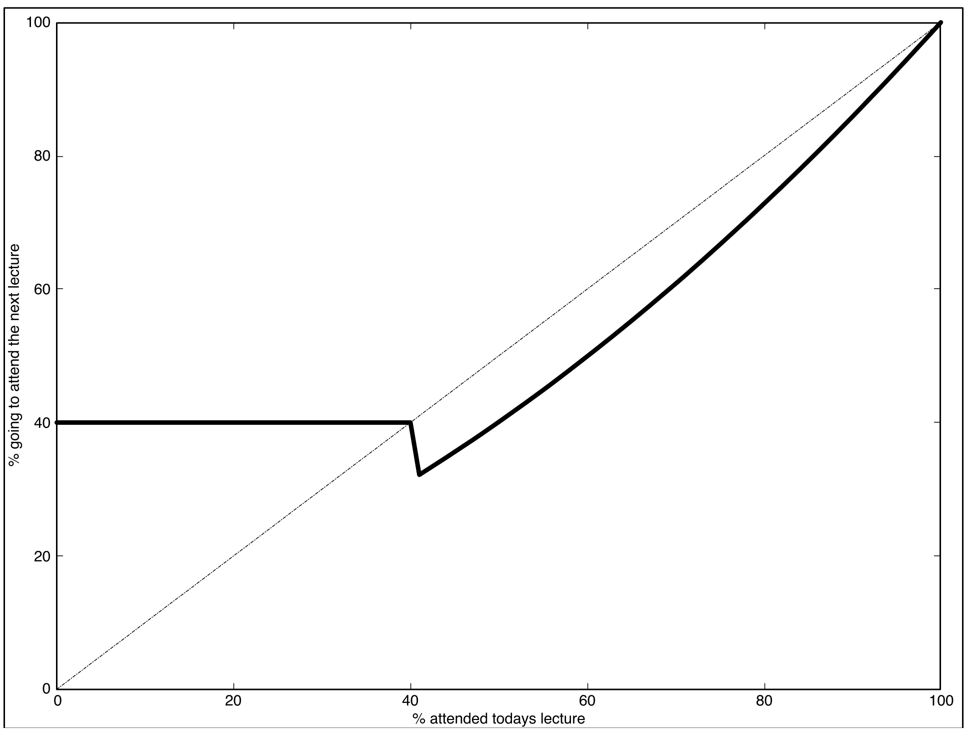

Figure 3: Equilibrium Analysis

(a) (5 points) Are there any equilibria in the system? If so, mark them on the diagram and mention whether the equilibria is stable or not?

(b) (5 points) Suppose 80% of the students attended the first lecture. Briefly describe the dynamic of participation for the future classes.

Solution.

(a) There are two equilibria, one at 40% which is stable and one at 100% which is not stable.

(b) The participation decreases until it reaches somewhere below 40% (corresponding to the bottom of the v-shape in the diagram), then it increases to 40% and stabilize at 40%.

□

Problem 6 (15 points: Graded by Ryan) Consider the following 2-player game matrix:

Row Player/Column Player	A	B
A	-1,-1	-1,+1
B	+1,-1	-10,-10

(a) (5 points) Does this game have any pure-strategy equilibria? If so, what are they?

(b) (5 points) Do you think there is a mixed-strategy equilibrium that is different from any pure-strategy equilibria? If so, broadly describe what the mixed-strategy equilibrium is. If not, why not?

(c) (5 points) Imagine that the game describes the contest known as Chicken: two players drive towards each other at high speed, and must choose either to keep going straight or swerve to their right at the last second. Indicate which of actions A and B correspond to straight and swerve, and explain why the payoffs in the table model this contest.

Solution.

(a) It has 2: Row Player plays A, Column Player plays B; Row Player plays B, Column Player plays A.

(b) There is a mixed strategy Nash equilibrium. Note that there are two pure strategy Nash equilibria, one where the Row Player is happy and the Column Player is unhappy, and one where the Row Player is unhappy and the Column Player is happy. However, we could imagine that the Column Player randomizes between his strategies to try to sometimes get the payoff where she is happy and in response to this the Row Player will also randomize between his strategies so that he will sometimes get the better payoff. In the mixed strategy Nash equilibrium, each player would need to ensure that the chance that they both end up playing B is small, because this is where both players are *very* unhappy.

(c) B would be going straight and A would be swerving. It makes sense because if both go straight, this corresponds to entry (B,B) and both end up with lots of damage to each other's cars (not to mention the medical bills). However, if one swerves and the other goes straight, then there is no collision and the one that goes straight gets "steet cred" worth utility 1, while the other person that swerves will always be referred to as "chicken." In the last scenario when both swerve, there is no collision, but also both are called "chicken."

□

Problem 7 (10 points: Graded by Shahin) Describe, as precisely as possible, a network in which there is a special vertex v that you would argue is important, despite having very low degree. Give a schematic diagram of this network with v clearly identified. Describe, as precisely as possible, a general measure of vertex importance that would identify v as important in your network.

Solution. The network given in Figure 4 has a node v that has a low degree compared to all the other nodes but may be referred to as important because it connects two large connected components (in this case two cliques) that otherwise would not be connected. (5 points)

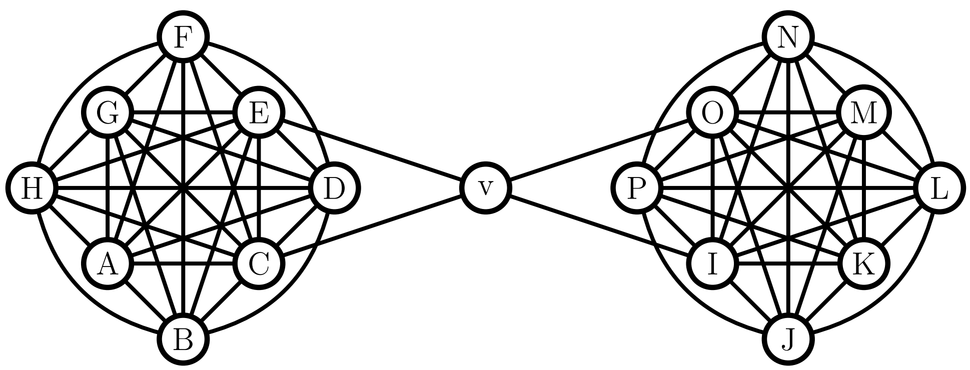

Figure 4: Solution for Question 7

Perhaps one definition for importance that would make v important compared to the other nodes would be to compute the size of the largest connected component in a graph G with and without the node v and compare these two numbers together. Hence, in Figure 4 v is important because the size of the largest connected component will drop from 17 to 8 after v being removed. Note that the deletion of no other node will decrease the size of the largest connected component. Another definition of importance would be as follows. Consider any two pairs of nodes in the graph and compute the shortest path between them. For each node v, computer the total number of shortest path among all the pairs that involves node v and refer to this number as $\mathrm{Im}(v)$. A node is important if it has a low degree and the $\mathrm{Im}(v)$ is high *e.g.* in Figure 4, $\mathrm{Im}(v) = 48$ which is much higher than other nodes. Note that E, C, O and I also have high Im but they are not of low degree. (5 points) **Common mistake:** define a node important if it has a low degree and removing the node makes the diameter graph disconnected (or increase the worst case diameter to infinity). Although v in Figure 4 indeed has such property, it is not hard to see why this is not a good notion of importance *e.g.*, node Q in Figure 5 will be defined as important with the latter definition but it is clearly not an important node in the network.

□

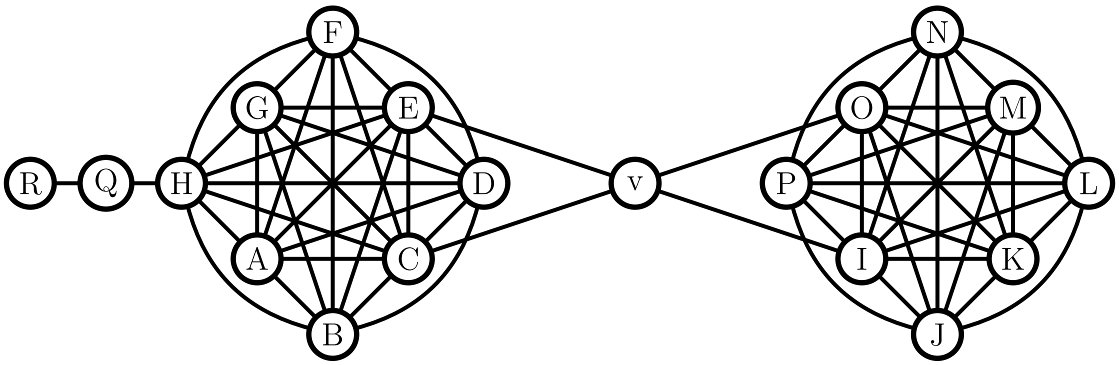

Figure 5: Is the node Q important?

Problem 8 (20 points: Graded by Shahin) In a recent lecture, we played a participatory game based on Schelling's segregation model.

(a) (5 points) Briefly but precisely describe the rules of the game.

(b) (5 points) Which of the following best describes the resulting outcome we observed: full segregation; partial segregation; full integration.

(c) (5 points) Briefly but precisely describe what was shown in the corresponding computer simulation in the same lecture.

(d) (5 points) Briefly but precisely describe the main points that this simulation is meant to illustrate.

Solution.

(a) We considered the network where students were nodes and a student's neighbors were all those that sat in the 8 adjacent seats around the student (1 point). We considered a student A to be *happy* if A had at least three neighboring seats that were occupied (loneliness) (1 point) and at least two of A's neighbors were of the same gender as A (homophily) (1 point). We then had the students move until all the students were *happy* (2 points).

(b) We saw partial segregation. (No partial credit)

(c) We again have a grid network where every node has 8 neighbors which are adjacent to it, and each node may be empty or be one of two colors. Of the nonempty nodes, 50% were one color, Red, and the other 50% Green. Each node desire a certain percentage p of its 8 neighbors to be of the same color and we can adjust this percentage. The nodes then can move to empty cells until all the nodes have at least p% of the same color.

The outcome of the simulation was an equilibrium: given what everyone else is doing, some nodes may want to move or not. We saw a tipping point when we set the p to be near 51%, just over a majority. This caused nearly everyone (95%) to be segregated from nodes with different colors. Where as when p was smaller, we observed partial segregation. Also as we increased the p it took more time for nodes to reach an equilibrium.

For grading, I deducted points for not mentioning the details of the experiment as in part (a). Also, I deducted points if you have not mentioned the tipping point and the threshold between partial and full segregation. I did not deduct any point if you mentioned the tipping point in the answer to part (d). Please refer to my comments in your exams.

(d) The main point of the simulation was that we could not infer the individual preferences from the outcome of the simulation *i.e.,* we saw almost full segregation even when the preferences of the individuals were mild (p was close to 50%). Also, there might be other equilibria for small values of p with less segregation, but since the nodes were not cooperating with all the other nodes in the network, we did not achieve those equilibria. Again for grading, please refer to my comments in your exams.

$\square$