Networked Life
CSE 112
Prof. Michael Kearns
Midterm Examination
March 1, 2007

NAME: ___________________________________

PENN ID: __________________________________

Exam Score:

Problem 1:__________/10
Problem 2:__________/15
Problem 3:__________/12
Problem 4:__________/12
Problem 5:__________/15
Problem 6:__________/15
Problem 7:__________/9
Problem 8:__________/12

TOTAL:____________/100

1. (10 points) Answer “True” or “False” to each of the following assertions.

One point each, no partial credit

a) The maximum number of edges possible in an undirected network of N vertices is N^2.

False

b) In Kleinberg’s model for navigation in social networks, a value of r=2 is the only value that permits rapid navigation.

True

c) Corporate portals are an example of web pages that might be part of the component IN.

False

d) A hub is a page that points to a lot of other good hubs.

False

e) A chromatic number greater than 4 is an example of a monotone graph property.

True

f) A clustering coefficient greater than 0.4 is an example of a monotone graph property.

False

g) It is more “prestigious” to have a high Erdos number than a low one.

False

h) In the contagion or “random walk” model of economic exchange, each vertex’s wealth will depend solely on their clustering coefficient.

False

i) When using a log-log plot, if the data appears linear, it is normally distributed.

False

j) The α-model of network formation was developed to provide a better explanation of naturally observed degree distributions than the Erdos-Renyi model.

False

2. (15 points) Consider the version of the Erdos-Renyi in which at each step, we choose two vertices that are not already connected uniformly at random, and then add the edge between them. Suppose that at some point during this process, there are K distinct connected components of size C_1 > C_2 > C_3 > …> C_K.

5 points for each part; partial credit as described.

a) Explain why it is difficult for two very large components to coexist in this process.

By definition of a component, two different components have no edges between them. But if they are both very large, the number of possible (or missing) edges between them must be very large (on the order of the product of the sizes of the two components). Thus, it quite likely that a randomly chosen missing edge will connect the two components into one. This argument can be made mathematically precise, but full credit for saying something close to this line of thought.

b) What can you say about the relative likelihood of the growth of the different components? What general type of process discussed in class does this remind you of?

Similar to the reasoning in part a), the larger a component is, the more "missing edges" it has to the rest of the network. Thus larger components are more likely to grow than smaller components. This is obviously reminiscent of "rich get richer" processes generally, and of preferential attachment more specifically. Full credit for both saying that large components are advantaged in growth, and mentioning one of these two processes; 2 points for only describing the growth properties or only mentioning one of the two processes.

c) Based on your answers above and discussions in class, what do you think the distribution of component sizes looks like in this Erdos-Renyi process, and how does it contrast with the degree distribution?

The distribution of component sizes will be heavy-tailed, and specifically a power law, while the degree distribution we know from class to be sharply peaked (Poisson). Full credit for saying both of these; 1 point for only recalling the degree distribution correctly.

3. (12 points) Consider very large networks in which both the degree of any vertex and the network diameter are relatively small.

a) Discuss, drawing as much as possible on course material and readings, why we would be interested in such networks.

Relevant materials and ideas:

- Obviously we are interested in large networks because so many of the ones of interest are enormous. Examples include social networks, the Internet, the web, the human brain, and so on.
- We are interested in small-diameter networks because of the many large-scale networks that have been documented to have rather small diameter. Relevant course citations include the six degrees of Travers and Milgram, the many small-diameter networks measured by Watts (Kevin Bacon graph, C. Elegans, North America power grid), and a number of others.
- We are interested in bounded-degree networks because of the belief that many networks, especially social networks, do have some fundamental limits on the degree --- i.e. you can only have so many friends because time is limited. Relevant course concepts include Gladwell's Magic Number 150 and the cortex ratio experiments of Dunbar.

2 points each for giving reasonable justifications (not necessarily those above) for each of these three parts.

b) Are there any mathematical limitations to arbitrarily large networks of small degree and small diameter? If your answer is "no", briefly explain why not. If your answer is "yes", give specific values for the network size, diameter and degree that cannot be simultaneously achieved, and explain why.

Yes, there are clear limitations. For instance, it is impossible to have a network with 100 vertices in which every vertex has degree at most 2, yet the network has diameter 5. The most efficient arrangement would be a 100-cycle, and clearly on average a pair of vertices is much more than 5 steps apart. 2 points for answering Yes; 4 additional points for clearly describing any scenario that cannot be achieved.

4. (12 points) For each of the networks shown below, indicate which vertex (A or B) has the **higher** PageRank value.

2 points each, no partial credit

a) Circle A or B

B

b) Circle A or B

B

c) Circle A or B

A

d) Circle A or B

A

e) Circle A or B

A

f) Circle A or B

B

5. (15 points) This question refers to the following class readings:

"*An Experimental Study of the Small-World Problem*".Travers, Milgram.
"*Navigation in a Small World*". Kleinberg.
"*Identity and Search in Social Networks*". Watts, Dodds, Newman.

Write a brief essay in which you describe what phenomenon Kleinberg and Watts, Dodds and Newman are attempting to explain *above and beyond* the findings of Travers and Milgram. Be sure to not only describe this phenomenon, but to indicate why its empirical presence requires explanation (that is, why might it be difficult to achieve). Finally, briefly contrast the different models and answers proposed by Kleinberg and Watts, Dodds and Newman.

- Whereas Travers and Milgram emphasized that large social networks have small diameter, the later works focused on the fact that people could actually *find* the short paths in a distributed fashion, using only very local information about the overall network topology.
- The reason this requires explanation is that sometimes simple "local" algorithms can fail to find the short paths even though they exist. An example are networks in which one must travel "away" from the target destination in order to travel the shortest path. For instance, we all know that the fastest way to travel to a geographic location is not to always try to move in its direction at all times --- we might need to go south to the airport (a "long distance link") in order to reach a northern destination most quickly.
- Kleinberg's model permits navigation solely through grid coordinates --- his local algorithms always forward messages to the neighbor whose grid address is closest to that of the target. Watts et al. have a richer model of social navigation that allows navigation via multiple "dimensions" --- geography, profession, religion, etc. Mathematically, Kleinberg's result shows a "knife's edge" requiring very particular conditions for efficient navigation, while Watts et al. have a more "robust" model that permits fast navigation under a wider range of circumstances.

5 points each for showing a clear grasp of each of the above 3 issues.

6. (15 points) This problem refers to the definition of the clustering coefficient discussed in class.

5 points each.

a) Briefly give the definition of the clustering coefficient of an individual vertex, and of an entire network.

The clustering coefficient of a vertex v is computed by first taking the neighbors of v. Letting the number of neighbors be k, we then count how many edges there are between the neighbors of v (ignoring v itself, which is used only to find its neighbors). If r is the number of such edges, the clustering coefficient of v is then $r/[k(k-1)/2]$, which is simply the fraction of possible edges appearing among v's neighbors.

The clustering coefficient of the entire network is simply the average of the clustering coefficients of all the vertices.

b) Briefly describe the criterion used to determine whether a network is highly clustered or not.

Let c be the clustering coefficient of the network. We then compute the fraction of all possible edges present in the entire network --- i.e. if the entire network has E edges and N vertices, we compute $p = E/[N(N-1)/2]$. We then compare c and p. If c is much larger than p we consider the network to be highly clustered; if, on the other hand, c is close to p then we do not consider it so.

c) For each of the following network formation models, briefly say whether they are highly clustered or not: Erdos-Renyi, the α-model, preferential attachment.

Only the α-model shows high clustering (at only at certain values of α, though it's not required to say this). Neither Erdos-Renyi nor preferential attachment show high clustering.

7. (9 points) For each of the network formation models below, write all of the following properties that the model exhibits: small diameter, giant component, high clustering, heavy-tailed degree distribution. If you think a model has none of these properties, write “none”.

2 points off for each missing/incorrect answer (min 0, max 9)

a) Preferential Attachment

small diameter
giant component (by definition PA always generates only a single component)
heavy-tailed degree distribution

b) Erdos-Renyi with p = 1/(2N)

none, p is too small

c) Erdos-Renyi with p = 150/N

giant component

small diameter is ambiguous, so will accept either answer

8. (12 points) The image on the last page of this exam was discussed in class and was generated from one of the February 16 human-subject consensus experiments. (Notice that it is rotated; be sure to read it with the numbers 5, 10,…, 35 going down the vertical axis.)

a) Briefly but precisely describe what this figure is showing, describing what the vertical and horizontal coordinates are illustrating.

The figure is showing the progression of a consensus experiment. For each of the 36 players, there is a row of colored bars; these are arranged on the vertical axis. The horizontal axis represents the elapsed time in the experiment, up to the maximum of 180 seconds. The color at row i at time t is the color currently chosen by player i at that moment of the consensus experiment.

4 points for showing a clear understanding of the diagram.

b) Write a brief analysis of the image, pointing out interesting examples of both individual and collective behavior. What was the final outcome of this experiment? Feel free to make annotations on the image and refer to them in your analysis.

2 points for indicating this is a failed consensus experiment
3 points for at least one clear example/discussion of individual behavior
3 points for at least one clear example/discussion of collective behavior