MIDTERM EXAMINATION
Networked Life (CIS 112)
March 4, 2010
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.

Name: _____

Penn ID: _____

Problem 1: _____/10

Problem 2: _____/20

Problem 3: _____/10

Problem 4: _____/10

Problem 5: _____/10

Problem 6: _____/10

Problem 7: _____/10

Problem 8: _____/20

TOTAL: _____/100

Problem 1 (10 points) For each of the following statements, simply write "TRUE" or "FALSE"

a. The Preferential Attachment network formation model explains all of the "universal" structural properties we examined in class.

b. The paper "Graph Structure in the Web" divides the pages on the Web into 7 distinct categories.

c. There is both mathematical and neuroscience evidence for the notion that there are limits to how many friendships we can maintain.

d. In controlled experiments in routing or navigation in social networks, it appears that people use geographic information mainly towards the very end of a chain.

e. If you have K friends or neighbors in a social network, the number of possible friendships among your friends grows roughly like the square root of K.

f. The PageRank algorithm can be viewed as spreading influence to the pages a particular web page points (hyperlinks) to.

g. In Kleinberg's "Hubs and Authorities" algorithm, a web page consisting of only hyperlinks to informative pages on mountain biking might obtain high authority weight for that topic.

h. "Connected" authors Christakis and Fowler are computer scientists.

i. The clustering coefficient measures how close "similar" vertices are in a network.

j. In Gladwell's terminology, a "maven" in a social network has high degree.

Problem 2 (20 points) In class we discussed the paper "The Scaling Laws of Human Travel", which makes use of data from the web site www.wheresgeorge.com.

a. Briefly describe the service provided by this web site.

b. Briefly describe the reasons for the paper's authors' interest in the data from the site.

c. Briefly describe the main finding of the paper that was discussed in class.

d. Briefly describe the connection drawn in lecture between this finding and the theoretical results of Kleinberg on navigation in social networks.

Problem 3 (10 points)

a. Let P be a monotone property of networks as defined in class. As precisely and succinctly as possible, give the definition of what it means for P to have a tipping point in the Erdos-Renyi model of network formation.

b. Name three specific monotone properties that have tipping points in the Erdos-Renyi model, and name them in the order (first to last) in which they would first appear in a network generated by Erdos-Renyi.

Problem 4 (10 points) Name two structural properties of social networks that frequently occur simultaneously, yet appear to be "in tension" with each other, in the sense that it is not obvious there should be simple mathematical models for network formation that can produce these two properties together. Then briefly describe a model we have studied or read about that indeed can do so.

Problem 5 (10 points) The assigned recent *Wired* magazine article “How Google’s Algorithm Rules the Web” discusses at length the many “signals” that inform Google’s algorithm --- perhaps individually tiny, but collectively important, contextual cues that Google uses to determine page relevancy for a given query. Briefly describe three such signals mentioned in the article, and suggest why they might be helpful in web search.

Problem 6 (10 points) In class we discussed “Rich Get Richer” processes, and the fact that they often lead to heavy-tailed distributions of whatever quantity is being allocated. The Preferential Attachment network formation model is one example of such a process, where connectivity (degree distribution) is being allocated. Give two other examples of quantities (not necessarily having to do with networks) approximately obeying a heavy-tailed distribution, and for each one briefly describe a natural “Rich Get Richer” process that might explain it.

Problem 7 (10 points)

a. Draw a connected network with exactly 10 vertices, exactly 9 edges, and the smallest diameter possible. Compute the clustering coefficient of this network.

b. Draw a connected network with exactly 10 vertices, exactly 9 edges, and the largest diameter possible. Compute the clustering coefficient of this network.

c. Draw a network (which may have multiple connected components) with exactly 10 vertices, exactly 9 edges, and the largest clustering coefficient possible.

Problem 8 (20 points) The book “Connected” and the associated articles discussed in class describe the research methodology and findings of authors Christakis and Fowler.

a) Briefly discuss the primary data source that is the basis for much of the authors’ research, and describe some of the properties of it that are different from what one might get from online social networks such as Facebook.

b) Much of the authors’ work meticulously establishes that certain behaviors, mental and physical states exhibit contagion in social networks. Name three things that are the subject of such contagion studies by Christakis and Fowler.

c) Briefly but carefully describe what the authors mean by “Three Degrees of Influence”.