MIDTERM EXAMINATION
Networked Life (NETS 112)
October 20, 2016
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. You may also make annotations directly on any diagrams given.

Name:

Problem 1: ________/10

Problem 2: ________/20

Problem 3: ________/20

Problem 4: ________/20

Problem 5: ________/15

Problem 6: ________/15

TOTAL: _________/100

Problem 1 (10 points). Answer “true” or “false” to each of the following assertions.
[Graded by Adel]

(a) The web-based version of the Travers and Milgram experiment demonstrated that people forwarded messages based on professional ties early in the chain, and geographic proximity later in the chain.

False

(b) An Erdos Number of 5 or less is rare among published mathematicians.

False

(c) The squash network exhibited homophily of ratings.

True

(d) The value of the exponent r in Kleinberg’s model that permits efficient navigation is 3.

False

(e) Being shown real-world friends who are poorly connected made people more likely to join Facebook.

True

(f) The main research are of Paul Erdos was statistical physics.

False

(g) Structural diversity refers to the many ways content can spread through social media.

True

(h) Cascades started by celebrity tweets tend to be broad but shallow.

True

(i) Gradually increasing the fraction of forest gradually increased the amount burned.

False

(j) A photo of Daniel Ellsberg appears on the course home page.

True

Problem 2 (20 points). This problem refers to the assigned reading "Can Cascades Be Predicted?" [Graded by Adel]

(a) What exactly were the authors trying to predict? Be as precise as you can.

Predicting ultimate size of viral cascade based on a number of variables that capture cascade properties; percentage of cascades that exceed f(k) if is k is median cascade size

(b) What were the five categories of features from which the authors tried to make their predictions? Give an example of a feature in each category.

Content: Type of picture
Root/author: who originally posted it
"Resharer": type of originator (page vs person)
Structural: network structure (depth vs breadth)
Temporal: how fast the shares spread

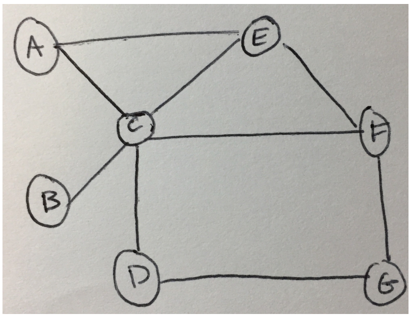
(c) What category of features was most predictive? What category was least predictive?

Most = temporal, least = content

(d) Can you draw any inferences from the answer to part (c) regarding how effective it might be to "engineer" viral content? Why or why not?

The paper doesn't provide negative evidence for engineering viral content because it was not a controlled experiment; in particular, there's no reason to think that most of the photos in the dataset were engineered for virality, so the fact that content features were not predictive doesn't mean they couldn't be.

Problem 3 (20 points). For the network shown below, numerically calculate each of the following quantities.
[Graded by MP]



(a) Diameter (average-case)

Sum of each possible path between all pairs of vertices divided by the total number of paths = 35/21 = ***5/3***

(b) Edge density

Number of edges present in network / total possible number of edges = 9/21 = ***3/7***
[Total possible number of edges = N(N-1)/2 = 7(6)/2 = 21]

(c) Clustering coefficient (assume degree 1 vertices have clustering coefficient 1)

cc(A) = 1
cc(B) = 1
cc(C) = 1/5
cc(D) = 0
cc(E) = 2/3
cc(F) = 1/3
cc(G) = 0
→ Global clustering = sum of individual clustering coefficients / number of nodes = ***16/35***

(d) Equilibrium wealth in the model discussed in class

Assume every vertex starts with $x. So total wealth = $7x
*For every vertex I, equilibrium wealth, W(i), equals Total Wealth * (degree(i) / total degree)*
W(A) = 7x(2/18) = 7x*(1/9)*
W(B) = 7x(1/18)*
W(C) = 7x(5/18)*
W(D) = 7x(2/18) = 7x*(1/9)*
W(E) = 7x(3/18) = 7x*(1/6)*
W(F) = 7x(3/18) = 7x*(1/6)*
W(G) = 7x(2/18) = 7x*(1/9)*

[We awarded to any answer that gave these ratios and a logical values for initial and total

wealth states.]

Problem 4 (20 points). In recent lectures, we have articulated five "universal" empirical properties of large-scale social networks.
[Graded by Brad]

(a) Briefly but clearly name and describe/define each of these properties. Be as precise as you can, using expressions involving the population size N where appropriate.

<u>*Small diameter*</u>*: diameter << N (# of vertices), log(N), or some constant*
<u>*Heavy-tailed degree distribution*</u>*: largest degree >> average degree; polynomial decay of network degrees; modeled by power law*
<u>*Giant component*</u>*: existence of a large connected component that is much larger than the second largest component*
<u>*Sparsity*</u>*: average or typical # edges << possible # edges; or, average # of edges grows linearly in N edges added*
<u>*Clustering*</u>*: density of connectivity w/i a community >> connectivity b/w communities, local edge density much larger than global edge density*

(b) One of the five properties is apparently "challenging", in the sense that it's not obvious how to satisfy it and all or some of the other properties simultaneously. State what the challenging property is, and justify your answer.

Sparsity or small diameter was accepted. Credit assigned depended on argument given.

Problem 5 (15 points). Consider the network formation model in which start with N vertices and no edges, and at each step, we pick a pair of vertices not already connected by an edge randomly among all such pairs, and add the edge between them. Suppose we run this process for E steps, at which point we will have added exactly E edges to the network.
[Graded by MP]

(a) For approximately what value of E would you expect the network to first have a giant component? Explain your reasoning by arguing why it is difficult for two large components to coexist at your chosen value for E.

Any answer on order N; the probability of adding an edge between two vertices and having one of them not in the connected component is very small

(b) Do you think there is a value of E at which the clustering coefficient of the network will be much higher than the overall edge density? Why or why not?

No, as discussed in lecture, there is simply no force in this model that would cause clustering — the expected clustering coefficient of a vertex is exactly the background edge density.

(c) Do you think there will be a value of E at which the degree distribution is heavy-tailed? Why or why not?

No, edges are just added at random so there is no "richer get richer process" in terms of individual degree.

Problem 6 (15 points). Consider the following model of network formation. We start with two vertices connected by an edge. At each step, we add a new vertex u with a single edge to the current network. The vertex v that we connect to u is determined as follows: with probability ½, we choose v to be the vertex that currently has the highest degree (breaking ties randomly); with probability ½, we choose v randomly among all vertices in the network so far.
[Graded by Brad]

(a) Do you think this model generates networks with small diameter? Why or why not? What do you think the diameter will be for large N?

Yes, there is a high probability that most vertices will be within 1 or 2 steps of the hub or highest degree vertex.

(b) Do you think this model generates networks with high clustering coefficient? Why or why not?

No, there will be one hub with none of its neighbors not connected to each other, implying low clustering.

(c) As precisely as you can, describe the degree distribution for large N.

Hub in middle will have ~ E/2 (or ~N/2) as degree and all other vertices will have degree values significantly less than ~E/2 or N/2 and their degrees will approximately have a Poisson distribution and not have a power-law distribution