The Structure of Information Networks	Problem Set 2
CS 6850 Fall 2008	Due October 3, 2008

(1) Consider a random graph created as follows. We have n nodes labeled $v_1, v_2, \ldots, v_n$ and n other nodes labeled $w_1, w_2, \ldots, w_n$. For a parameter α, we include each edge of the form (v_a, w_b) independently with probability $n^{-\alpha}$. This gives us a random bipartite graph on $2n$ nodes (since edges only go between nodes of the form v_a and nodes of the form w_b).

For a fixed choice of i and j, let E_{ij} be the event that in the resulting random graph, there is a path of length 2 connecting nodes v_i and v_j.

(a) Give a formula for $\Pr[E_{ij}]$ in terms of n and α.

(b) We say that a value of α is *critical* if, by choosing the probability of each edge to be $n^{-\alpha}$, the probability of E_{ij} (for a fixed i and j) converges to a number strictly between 0 and 1 as n goes to infinity. More succinctly, α is critical if

$$\lim_{n\to\infty} \Pr[E_{ij}] = c$$

for some $0 < c < 1$. Such a choice of α is interesting because the probability of a length-2 path doesn't converge to either 0 or 1, but remains somewhere in between even as n grows arbitrarily large.

Give a value of α that is critical, and provide an explanation for your answer.

(c) Let α^* be a critical value of α, and let $\beta < \alpha^*$. Show that if we generate edges with probability $n^{-\beta}$ according to the model above, the probability that *all* pairs of nodes v_i, v_j are connected by length-2 paths converges to 1 as $n \to \infty$. *(Hint: Use the Union Bound. You can also use the fact that for any constants $c > 0$ and $\varepsilon > 0$, it is the case that $\lim_{n\to\infty} n^c e^{-n^\varepsilon} = 0$.)*

(2) Consider a long, straight road, which we model as a line segment of length n. We drop a set of k sensors randomly on this road — so each lands in a location selected uniformly and independently from the interval $[0, n]$.

Now, each sensor has a transmitting range of 2, so it can communicate with any other sensor within a distance 2 of it. This mean that the random placement of the sensors defines a random k-node graph G, in which the nodes are the sensors, and we connect two by an edge if they can communicate with each other. We'd like to choose k large enough so that G is connected with high probability, and we can do this by reasoning as follows.

(a) For an integer j from $1, 2, \ldots, n$, let E_j denote the event that no sensor lands in the interval $[j-1, j]$. Give a formula for $\Pr[E_j]$ in terms of n and k.

(b) Argue briefly that if none of the events E_j occurs, then the random graph G defined above is connected.

(c) Show that if we drop $k = 2n \ln n$ sensors at random, then with high probability the graph G will be connected. (In particular, with probability converging to 1 as $n \to \infty$.)

(3) A common goal, when analyzing a large graph G, is to try identifying a dense "core" that is internally well-connected. There are several definitions for this in practice, but one simple way to evaluate whether a subgraph H of G constitutes a good core is to look at its minimum node degree, considering H as a graph in isolation. That is, we try to find a subgraph H in which no degree is small.

Given an input graph G, with a subgraph H, let $\delta(H)$ denote the minimum degree of a node in H (when we view H as a graph in isolation). Now, suppose we are given a parameter d. We would like to decide whether G contains a subgraph H for which $\delta(H) \geq d$. Such an H would be a "core" of the type of described above.

(a) For any d and any r, describe an example of a graph G such that G contains at least r nodes each of degree at least d, but it contains no subgraph H for which $\delta(H) \geq d$. Give a brief explanation for why your example has the desired property.

(b) Give an efficient algorithm that takes an input graph G and a parameter d, and decides whether G contains a subgraph H with $\delta(H) \geq d$. Give a brief proof that your algorithm is correct.