Discrete Structures	Problem Set 11
CS 280 Spring 2005	Due Wednesday, May 4, 2005

Reading: Rosen edition 5: 8.1-5,8.5.7-8. and Section 9.4 Or from edition 4: Sections 7.1-4, 7.7-8, and Section 8.5.

(1) To understand big graphs like the Web graph, people study random processes to create graphs that share many of the properties of the graph that we want to understand. The simplest such random graph definition is the following. For a probability $0 < p < 1$ we obtain a random graph G on n nodes by adding each of the possible $\binom{n}{2}$ edges independently with probability p.

(a) What is the probability that we obtain a graph with m edges via this process, for a given integer m?

(b) Argue that the expected degree of a node in this graph is $\mu = p(n-1)$. Let $0 < k < 1/p$, and use the Chebyshev bound to bound the probability that a node has degree $(k+1)\mu$, that is, $(k+1)$ times the expectation.

(c) Unfortunately, this simple graph creation process is very unlikely to generate graphs that are similar to the Web or to the down-load graph we considered on the previous problem set. The Web and also our down load graph has a lot of high degree nodes (popular Web sites like, CNN, Google, etc are pointed to by lots and lots of other pages, and as a result there are a good number of nodes whose degree far exceeds the expectation.)

Show that the the expected number of nodes whose degree is at least $(k+1)\mu$ in the random graph we consider here is at most $\frac{1}{k^2 p}$.

(d) Let v and w be two nodes ($v \neq w$), and let X_v and X_w be the degree of v and w respectively. Note that X_v and X_w are random variables. Are these two random variables independent? Explain why your answer is correct.

(2) We say that a graph is a simple path, if the graph has k nodes for some k, and $k-1$ edges that connect the graph as a simple path, as shown in the figure on the next page.

The following proof claims that a graph that has exactly 2 nodes of degree 1 and all other nodes have degree 2 is a simple path. The proof is by induction on the number of nodes n. Since the graph must have 2 nodes of degree 1 by assumption, we must have $n \geq 2$.

Base case: $n = 2$. The only simple graph on 2 nodes that has two degree 1 nodes has a single edge, and is indeed a simple path.

Induction step (by weak induction). Consider an n node graph G that satisfy our conditions. We know by the induction hypothesis that this graph G is a simple path. Assume the nodes occur on the path in order $v_1, \ldots, v_n$. Now we need to add one extra node v_{n+1} to

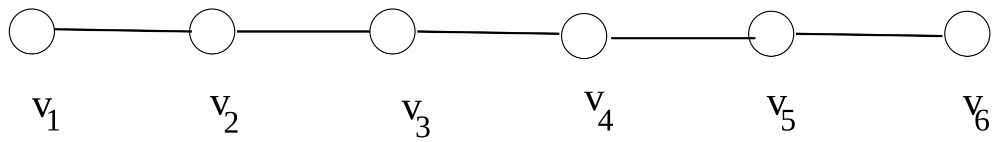

Figure 1: A simple path graph.

make an $n+1$ node graph. Now node v_{n+1} cannot have degree 0, so we need to add some edges to G. If we add an edge connecting node v_{n+1} to an internal node of the path, then the internal node would have degree 3, hence violating our assumption. The only way to connect v_{n+1} is to connect it to one of the ends v_1 or v_n, and in either case, the resulting graph is a path.

(a) Show that this claim is wrong by showing an example of a graph that satisfies the condition, and yet is not a path.

(b) Explain what is wrong in the proof above.

(3) A *binary tree* is a rooted tree in which each node has at most two children. Show by induction that in any binary tree the number of nodes with two children is exactly one less than the number of leaves. (Leaves of a tree are all nodes of degree 1, expect the root.)

(4) [From Kleinberg-Tardos] A number of stories in the press about the structure of the Internet and the Web have focused on some version of the following question: How far apart are typical nodes in these networks? If you read these stories carefully, you find that many of them are confused about the difference between the *diameter* of a network and the *average distance* in a network — they often jump back and forth between these concepts as though they're the same thing.

As in the text, we say that the *distance* between two nodes u and v in a graph $G=(V,E)$ is the minimum number of edges in a path joining them; we'll denote this by $dist(u,v)$. We say that the *diameter* of G is the maximum distance between any pair of nodes; and we'll denote this quantity by $diam(G)$.

Let's define a related quantity, which we'll call the *average pairwise distance* in G (denoted $apd(G)$). We define $apd(G)$ to be the average, over all $\binom{n}{2}$ sets of two distinct nodes u and v, of the distance between u and v. That is,

$$apd(G) = \left[\sum_{\{u,v\}\subseteq V} dist(u,v)\right] / \binom{n}{2}.$$

Here's a simple example to convince yourself that there are graphs G for which $diam(G) \neq apd(G)$. Let G be a graph with three nodes u,v,w; and with the two edges $\{u,v\}$ and $\{v,w\}$.

Then

$$diam(G) = dist(u, w) = 2,$$

while

$$apd(G) = [dist(u, v) + dist(u, w) + dist(v, w)]/3 = 4/3.$$

Of course, these two numbers aren't all *that* far apart in the case of this 3-node graph, and so it's natural to ask whether there's always a close relation between them. Here's a claim that tries to make this precise.

> *Claim: There exists a positive natural number c so that for all connected graphs G, it is the case that*
> $$\frac{diam(G)}{apd(G)} \leq c.$$

Decide whether you think the claim is true or false, and give a proof of either the claim or its negation.

(5) [From Kleinberg-Tardos] *Fault tolerance* is an important property of networks, and understanding what makes networks fault tolerant is subject of a lot of research. For example, it is important to know if a construction crew accidentally cutting thought a phone cable will disconnect the phone network, cause major outages in electricity, or bring down the Internet.

To study this issue, we can study this issue in terms of graph properties. Consider a connected graph $G = (V, E)$. We say that an edge $e \in E$ is a *bridge* if deleting e results in a graphs that is no longer connected, that is $(V, E - \{e\})$ is not a connected graph. We say that a graph G is *bridge-free* if G is connected and has no bridges. An alternate notion of network robustness is defined using spanning tree. We say that G is *robust* if there are two spanning trees T_1 and T_2 of G that have no edges in common.

Both of these notions somehow formulate the resilience of the network in terms of a graph property. Here we try to understand the relation between these two properties. For both of the questions below decide if the implication is true, and prove that your answer is correct.

(a) Are all bridge free graphs robust?

(b) Are all robust graphs bridge free?

(6) [From Kleinberg-Tardos] There's a natural intuition that two nodes that are far apart in a communication network — separated by many hops — have a more "tenuous" connection than two nodes that are close together. There are a number of algorithmic results that are based to some extent on different ways of making this notion precise. Here's one that involves the susceptibility of paths to the deletion of nodes.

Suppose that an n-node undirected graph $G = (V, E)$ contains two nodes s and t such that the distance between s and t is strictly greater than $n/2$. Show that there must exist some node v, not equal to either s or t, such that deleting v from G destroys all s-t paths. (In other words, the graph obtained from G by deleting v contains no path from s to t.)

(7 optional) Suppose a simple undirected connected graph has the property that from some node s the DFS tree with root s is the same as the BFS tree with root s. What can you say about the graph? Prove your claim.