Graphs and Networks **out:** October 18, 2007

Problem Set 3

Lecturer: Daniel A. Spielman **due:** November 1, 2007

1 Introduction

There are 3 sections to this problem set. The first contains the 3 problems for the undergraduates. The second contains an additional problem for the graduate students. The third contains the optional experimental portion. Each experimental problem may be substituted for one ordinary problem.

2 Homework Policy

You may discuss the problems with other students. But, you must write your solutions independently, drawing on your own understanding. You should cite any sources that you use on the problem sets other than the textbook, TA and instructor. This means that you should list your collaborators.

You **may not** search the web for solutions to similar problems given out in other classes. If you think this policy needs any clarification, please let me know.

3 Useful inequalities

Here are some inequalities you may find useful when doing this problem set.

$$(1-p) \leq e^{-p} \tag{1}$$

$$(1-p) \geq (1-p)^p e^{-p} \quad \text{for } p < 1 \tag{2}$$

$$(1-p)^k \geq (1-pk) \quad \text{for } k \geq 1 \tag{3}$$

$$\binom{n}{k} \leq \frac{n^k}{k!} \tag{4}$$

$$\binom{n}{k} \geq \left(\frac{n}{k}\right)^k \tag{5}$$

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k \tag{6}$$

$$n! \geq \left(\frac{n}{e}\right)^n \tag{7}$$

4 Undergraduate Problems

1. **Threshold for connectivity**

   Consider the random graph model $G(n, p)$, where $p = 2 \ln n/n$. We will prove that a graph chosen from this distribution is almost certainly connected. (Hint: this problem does not require any technique more sophisticated than the union bound or Markov's inequality.)

   (a) Prove that the probability that there is a vertex with no neighbors is at most $2/n$, for n sufficiently large (this is just warm-up).

   (b) The graph is disconnected if and only if there exists a subset of the vertices $\emptyset \subset S \subset V$ such that G contains no edges between S and $V - S$. Prove for each $0 < k \leq n/2$ that it is unlikely there is any set S of size k such that G contains no edges between S and $V - S$. (that is, just prove it for each particular k)

   Hint: you may want to use a different approach for small k, say $k \leq n/4$ than for large k, say $k \geq n/2$.

   (c) Prove that it is unlikely that a graph chosen from the distribution $G(n, p)$ is disconnected. (that is, sum the bound from part (b) over k)

   (d) **[Extra Credit]** Prove this for $p = (1 + \epsilon) \ln n/n$.

2. **Degree-3 vertices**

   Consider a random graph from the distribution $G(n, p)$, with $p = 1/n$. We will show that it is very likely that such a graph contains a vertex of degree at least 3. (Actually, we could do this for much lower values of p)

   (a) Prove that there is a constant c such that the expected number of vertices of degree at least 3 is at least cn. (it suffices to do this for n large)

   Hint: Compute a lower bound on the probability that a vertex has degree at least 3. You could do this by computing the probability that a vertex has degree exactly 3, or by computing the probability that a vertex has degree 0, 1, or 2.

   (b) Compute an upper bound on the variance of the number of vertices of degree at least 3. (it suffices to do this for n large)

   (c) Use the variance bound from part (b) to prove that the probability there is vertex of degree at least 3 goes to 1 as n goes to infinity.

   We will now see an alternate way to prove that the probability of a vertex of degree at least 3 goes to 1.

   (d) Assume n is even and arbitrarily divide the vertices in half, into sets A and B where $|A| = n/2$. Find a constant c such that the probability that a vertex in A has at least 3 neighbors in B is at least c. Note that for distinct vertices in A, the events that they have 3 neighbors in B are independent. Now, prove that the probability that there is no vertex in A with at least 3 neighbors in B goes to zero as n goes to infinity.

3. **You can't cluster a random graph.** Consider a random graph distributed according to $G(n, p)$, with $p = 100 \ln n/(n-1)$ (I've chosen 100 just for overkill). Prove that there exists an absolute constants $\alpha > 0$ and $\beta > 0$ such that

$$\Pr\left[\min_S \mathrm{sp}(S) < \alpha\right] < \beta,$$

where

$$\mathrm{sp}(S) = \frac{|\partial(S)|}{\min(|S|, |V - S|)}.$$

Hint: this is like Problem 1, but you will have to use Chernoff bounds instead of Markov's inequality to prove the analog of part (*b*).

5 Graduate Problem

You can't cluster a cycle plus a matching.

Let G be a cycle plus a random matching. Prove that there exists an absolute constant $\alpha > 0$ such that

$$\Pr\left[\min_S \mathrm{sp}(S) < \alpha\right] \to 0,$$

as n goes to infinity, where the minimum is taken over all sets of at most half the vertices.

a. Let S be a set k vertices. Let $j \leq k$. Figure out how many matchings there are on V that have exactly j edges leaving S. Note that the number of matchings on a graph with $2n$ vertices is $(2n-1)!! \stackrel{\text{def}}{=} (2n-1)(2n-3)(2n-5)\cdots(1)$.

b. Prove that it is unlikely that any particular set of $k \leq n/2$ vertices has fewer than αk matching edges leaving it.

c. Now, if you had to sum over all sets of size k, there is no way you could prove the claimed result. But, you don't have to! It is only necessary to consider sets S with fewer than αk cycle edges leaving. For each k, prove an upper bound on the number of sets of vertices with fewer than αk cycle edges on their boundary.

d. Finally, prove the claimed result.

6 Experimental Problems

There are two experimental problems. Each may be substituted for one of the other problems.

1. Compute the assorativity by degree and clustering coefficient of your graph. (You may use any reasonable definition of clustering coefficient from Lecture 2).

2. Implement one of the following two models for random graphs: Erdos-Renyi graphs with $p = 2 \ln n / n$, or a cycle plus a random matching. Now, try to partition these graphs by either Grady's method or the spectral method. If you use Grady's method, try using many pairs s and t.

   a. Try this with at least 10 graphs from many sizes, including $n = 100, 1000, 10000$.

   b. For each run, report the number of edges in the sparsest cut you found.