Advanced Algorithms, Fall 2012

Prof. Bernard Moret

“Homework” Assignment #4

NOT DUE (just for reference for the test)

Question 2.
Recall that K_n denotes the complete (undirected) graph on n nodes, i.e., the graph with n nodes and all possible $\binom{n}{2}$ edges. Consider coloring the edges of K_n from a set of two colors, say red and blue. Thus each edge is assigned a single color; note that this is not what is termed a *coloring* of the graph (where coincident edges must have distinct colors). We are interested in the largest complete subgraph of K_n that is entirely red or entirely blue.

- Prove that there exists a red-blue coloring of K_n that does not induce a complete subgraph K_k that is entirely red nor one that is entirely blue whenever we have

  $$\binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

  Prove it by showing that the event [no blue K_k and no red K_k] has non-zero probability.

- Next (as an exercise in asymptotics), find the largest n (as a function of k) for which the inequality above holds.

- Let us examine a random coloring of K_n, where each edge is independently colored red with probability $\frac{1}{2}$ and colored blue with probability $\frac{1}{2}$. Then the probability that a randomly chosen subset of k vertices, call it S_k, induces a monochromatic subgraph on these k vertices and their $\binom{k}{2}$ associated edges is

  $$Pr[S_k \text{ is monochromatic}] = Pr[M(S_k)] = 2 \cdot 2^{-\binom{k}{2}}$$

  (the factor of 2 is due to the presence of two possible colors). We now want to consider the probability of the union of all events $M(S_k)$, $Pr[\bigcup_{S(k)} M(S_k)]$. These are certainly not independent events, but we can still bound the probability of the union by the sum of the probabilities:

  $$Pr[\bigcup_{S(k)} M(S_k)] \leq \binom{n}{k} 2^{1-\binom{k}{2}}$$

  We have assumed that this quantity was strictly less than 1; hence the probability of the complemented event, $Pr[\bigcap_{S(k)} \overline{M(S_k)}]$, is strictly larger than zero. But this is the probability that our randomly colored K_n has no monochromatic subgraph of size k. Thus there must exist at least one coloring of K_n that possesses this property, proving our assertion.

- Now we turn our attention to the equality

  $$\binom{n}{k} 2^{1-\binom{k}{2}} = 1$$

If k is the variable, how can we solve for n, at least so as to bound it? Using our standard approximation for $\binom{n}{k}$, we want

$$\left(\frac{ne}{k}\right)^k \approx 2^{\binom{k}{2}}$$

or, after taking the kth root, $ne/k \approx 2^{(k-1)/2}$, yielding finally $n \approx ke^{-1}2^{(k-1)/2}$, or

$$n \approx \frac{k}{e}\sqrt{2^{k-1}}$$

Because we took a kth root, the approximations introduced only negligible error, on the order of $k^{1/k}$.

Question 2.
Let G be a random bipartite graph on two independent sets of vertices U and V, each with n vertices. For each pair of vertices $u \in U$ and $v \in V$, the probability that the edge $\{u, v\}$ exists is $p(n)$, a value that is independent of all other edges. Let $p(n) = (\ln n + c)/n$ for some positive real value c. Show that the probabiility that G contains an isolated vertex is asymptotically equal to $1 - e^{-2e^{-c}}$.

If p is the probability that an edge exists in the bipartite graph, then the probability that some given vertex is isolated is just $(1-p)^n$, since there are n potential neighbors for i. We have

$$(1-p)^n = \left(1 - \frac{\ln n + c}{n}\right)^p \approx \frac{1}{n}e^{-c}$$

where the approximation is asymptotically exact. If we let X_i be an indicator variable set to 1 is vertex i is isolated and to 0 otherwise, then $\sum X_i$ is the number of isolated vertices and

$$E[\sum_{i=1}^{2n} X_i] = \sum_{i=1}^{2n} E[X_i] = 2nE[X_i] \approx 2e^{-c}$$

where, again, the approximation is exact in the limit. In order to have at least one isolated vertex, we must deviate from this expectation by a factor of at least $e^c/2$, which, by Markov's inequality, occurs with probability at most $2e^{-c}$. If we fear that the bound is too sloppy, we can use the Boole-Bonferroni inequalities to get a *lower* bound on the probability. We know that the probability that one specific vertex i is isolated is $\frac{1}{n}e^{-c}$; call that event E_i. What we seek, then, is the probability of $\cup_i E_i$. By the Boole-Bonferroni inequalities, that probability is at least

$$\sum_i Pr[E_i] - \sum_{i>j} Pr[E_i \cap E_j]$$

We know the first term, but need to compute the second. The probability that both vertices i and j are isolated depends on whether these two vertices are on the same side of the bipartite graph or not; thus we write

$$\sum_{i>j} Pr[E_i \cap E_j] = 2\binom{n}{2}(1-p)^n(1-p)^n + n^2(1-p)^n(1-p)^{n-1}$$

where the first term exists twice (once for each side of the graph) and accounts for all choices of 2 vertices from the n vertices on a given side, the two being then isolated with independent probabilities (we need $2n$ edges absent), whereas the second term accounts for all $n \cdot n$ choices of one vertex from each side, with isolation of both vertices achieved with only $2n-1$ edges absent, since one is shared. So now we can state that the probability of having at least one isolated vertex is at least as large as

$$2n(1-p)^n - 2\binom{n}{2}(1-p)^n(1-p)^n - n^2(1-p)^n(1-p)^{n-1}$$

which is asymptotically equal to

$$2e^{-c} - 2e^{-2c}$$

Thus the probability of having at least one isolated vertex is asymptotically larger than $2e^{-c} - 2e^{-2c}$, but asymptotically smaller than $2e^{-c}$, confirming that our first bound was tight. Note that the deviation from independence is tiny at first and asymptotically negligible, so that we can in fact just write our probability as

$$\binom{2n}{1}e^{-c}/n - \binom{2n}{2}e^{-2c}/n^2 + \binom{2n}{3}e^{-3c}/n^3 - \cdots$$

which becomes asymptotically simply

$$2e^{-c} - 4e^{-2c}/2 + 8e^{-3c}/6 - \cdots = \sum_i (-1)^{i+1}(2e^{-c})^i/i! \approx 1 - e^{-2e^{-c}}$$

(again asymptotically exact).

Question 3.
You are given a graph G of n vertices and $nd/2$ edges. An independent set in a graph is a subset of vertices, no two of which are connected by an edge. Consider the following randomized procedure to produce an independent set: delete each vertex of G (along with its associated edges) independently with probability $1 - \frac{1}{d}$.

- Compute the expected number of vertices and edges remaining after the deletion pass.
- Infer that there exists an independent set with at least $n/2d$ vertices in G.

- Let X_i be an indicator variable set to 1 if vertex i survives the elimination, to 0 otherwise. Then we have $E[X_i] = \frac{1}{d}$; since the expected number of remaining vertices is $E[\sum_i X_i]$, by linearity of expectations, we conclude that this number is just $\frac{n}{d}$. In particular, note that a vertex of degree k, if it survives, survives with an expected degree of k/d and thus contributes k/d^2 to the expected total degree of the graph; since the sum of the degrees of all vertices in the graph is nd, the expected sum of the degrees of the surviving vertices is n/d. Now consider some arbitrary edge of the graph; it survives if and only if both of its endpoints survive, that is, it survives with probability $\frac{1}{d^2}$ (since the survival of one endpoint is independent of the survival of the other). While the survival of one edge is *not* independent of the survival of another, linearity of expectations nevertheless allows us to conclude that the expected number of surviving edges is $\frac{nd}{2} \cdot \frac{1}{d^2} = \frac{n}{2d}$.
- We claim that any graph $G = (V,E)$ with $|E| \leq \alpha|V|$, for $\alpha < 1$, has an independent set of size at least $|V|/2$; we prove it below for $\alpha = \frac{3}{4}$, but note that the proof extends easily to the general case.

  In such a graph, the average degree of a vertex is at most 1.5. If every vertex has degree 1 or 0, then each edge has endpoints distinct from all other edges, so that selecting one endpoint from each edge plus all isolated vertices yields an independent set of size at least $|V|/2$. If vertices of degree higher than 1 exist, say a vertex of degree k, then, for each such vertex, there must be $\frac{2}{3}k-1$ vertices of degree 0, or $2k-3$ vertices of degree 1, or some combination of vertices of degree 0 and 1 to lower the average back down to 1.5 (other degrees do not help). Then place the vertices of degree 0 and as many vertices of degree 1 as possible (if two of the vertices of degree 1 are connected by an edge, we can only place one of them) in the independent set, discard the vertex of degree k and any vertices connected to the vertices of degree 1, adjust the degrees of the remaining vertices, and proceed. In the worst case (all vertices removed are of degree 1 and they come in pairs, except for one of them), this removed a total of $2k-2$ vertices and at least $2k$ edges from the graph, placing at least k vertices in the independent set. Thus the resulting reduced graph still has the desired ratio of vertices to edges and we shall obtain enough vertices from it to complete our independent set. Hence our claim is proved. Note in passing that our claim is false for $|V| = |E|$.

It remains to show that any graph with n vertices and $nd/2$ edges has a subgraph of at least $\frac{n}{d}$ vertices with at most $\frac{\alpha n}{d}$ edges for some $\alpha < 1$. The probabilistic method tells us that, since the expected number of surviving vertices is n/d, there exists at least one sequence of eliminations that yields a graph with at least $\frac{n}{d}$ surviving vertices; similarly, we can conclude that there exists at least one sequence of eliminations that yields a graph with at most $\frac{\alpha n}{2d}$ edges. The problem is that we do not know whether the same sequence of eliminations could work in both cases. So we can try to work on both processes together and establish a non-zero probability that such a graph is produced. The survival of vertices is like a series of coin tosses and is described by the binomial distribution

$$Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

where p is the success probability, in our case the probability of survival $p = d^{-1}$. We can thus conclude that the probability of observing fewer than the expected number of vertices is

$$\sum_{i=0}^{\lceil \frac{n}{d} \rceil - 1} \binom{n}{i} d^{-i} (1-d^{-1})^{n-i}$$

which is always less than $\frac{1}{2}$ and, in fact, is at most $\frac{1}{2} - \frac{1}{n}$.

When it comes to edges, however, we no longer have a Poisson process, since edges with common endpoints are interdependent; however, we can still use the Markov bound, which tells us that the probability of having more than $\alpha n/d$ edges, for $\alpha < 1$, surviving the process is at most $1/2\alpha$, which is slightly larger than $\frac{1}{2}$. We can choose $1 - \frac{2}{n+2} < \alpha < 1$, so that we have $\frac{1}{2} - \frac{1}{n} + \frac{1}{2\alpha} < 1$; hence failing both bounds occurs with probability strictly less than 1, thereby establishing that at least one subgraph exists that combines both vertex and edge bounds.

A much simpler way to show the same is to combine the two expected values into a single useful quantity. Consider the difference between the number of surviving vertices and the number of surviving edges; by linearity of expectations, the expected value of this difference is $\frac{n}{d} - \frac{n}{2d} = \frac{n}{2d}$. The probabilistic method thus tells us that any graph of n vertices and $nd/2$ edges has a subgraph with $\frac{n}{2d}$ more vertices than edges. Now consider removing all of the edges of this subgraph by removing one endpoint for each edge; at most we remove as many vertices as edges, leaving us with at least $\frac{n}{2d}$ vertices that must then form an independent set. Et voilà!