Summary of Lecture 2

In which we study Max Cut in $G_{n,1/2}$ random graphs, and begin to study Independent Set in $G_{n,p}$ random graphs.

In the previous lecture we discussed three results concerning clique in $G_{n,1/2}$ random graphs: that, up to lower order terms, the probable value of the optimum is $2\log n$; that the solution found by a simple greedy algorithm has cost $\log n$; and that the best known way to certify upper bounds for most graphs gives an upper bound of the order of $\sqrt{n}$. We reviewed one way of finding $O(\sqrt{n})$ upper bound certificate based on linear algebra considerations. Today we prove three results of the same kind for Max Cut.

The probable value of the optimum is at most $\frac{n^2}{8} + O(n^{1.5})$, which can be proved by using Chernoff bounds and a union bound.

Then we considered the following greedy algorithm: initialize two sets A, B to be empty, then process each node of the graph and put it in the set in which it has fewer neighbors. The algorithm always cuts at least half of the edges. The number of cut edges minus half the number of edges is equal to the sum for $v \in V$ of the absolute value of the difference between the number of neighbors that v has in A compared to B at the time that v is processed. For the vertex processed at step i, the expectation of the absolute value of the difference is $\Omega(\sqrt{i})$ and the variance is $O(i)$, so the advantage over half the edges has mean $\Omega(n^{1.5})$ and standard deviation $O(n)$, meaning that it is $\Omega(n^{1.5})$ with $1-O(1/n)$ probability.

Finally, we see that the max cut of G is at most

$$\frac{n^2}{8} + O(n||A - J/2||)$$

which gives a $n^2/8 + O(n^{1.5})$ upper bound certificate whenever $||A - J/2|| \leq O(\sqrt{n})$, which, as we mentioned last time, happens with high probability.

Thus, probable optimum, cost of the solution found by the greedy algorithm, and polynomial time certifiable upper bounds are all $n^2/8 + \Theta(n^{1.5})$ with high probability.

Next, we begin to study the $G_{n,p}$ model, where we allow $p = p(n)$ to depend on n. For example we may have $p = 1/\sqrt{n}$ or $p = 100/n$.

The Maximum Independent Set problem in graphs sampled from $G_{n,p}$ is the same as the Maximum Clique problem in graphs sampled from $G_{n,1-p}$, so we already understand the Independent Set problem in $G_{n,1/2}$. We will now look at the case of smaller p (including p going to zero with n), beginning with an upper bound to the probable value of the optimum.

The expected number of independent sets of size k in $G_{n,p}$ is

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

so we would like to find what are values of k (dependent on n and $p(n)$) for which the above quantity goes to zero. Some manipulation shows that we want

$$k > 2 \log_{1+\frac{1-p}{p}} \frac{n}{k}$$

The above is true for $k > \frac{2}{p} \ln n$, so $\frac{2}{p} \ln n$ is an upper bound to the probable optimum, but it is not a tight one. (Note that it can be $\Omega(n \log n)$ when $p(n) = O(1/n)$, even though the optimum can never be more than n.)

A tighter sufficient condition (up to lower order terms) is to have $k > \frac{3}{p} \ln 2np$. Replacing $d := pn$, we have that the probable value of the optimum is $O\left(\frac{n}{d} \log d\right)$.

Note that there is a simple greedy algorithm that, in every graph of average degree d, finds an independent set of size $O(n/d)$. Next time we will show how to gain an extra factor of $\log d$ in random graphs of average degree d, and we will discuss how to certify upper bounds.