Graphs and Networks Lecture 11

Random Graphs : Markov's Inequality

Lecturer: Daniel A. Spielman October 11, 2007

11.1 Introduction

In this lecture, we will consider the Erdos-Renyi model of random graphs. Our motivation is *not* to present them as a model of graphs that occur in real life—as it is rare to find graphs that behave like Erdos-Renyi graphs. Rather, we present them because they have many counter-intuitive properties, and they provide the most important family of counter-examples to natural conjectures about graphs. Also, their study provides great vehicle for teaching probabilistic analysis.

In this lecture, we will encounter the following quantities associated with graphs.

- Girth. We write $g(G)$ to denote the girth of a graph G. It is the length of the *shortest* cycle in G
- Clique number. We write $\omega(G)$ to denote the number of vertices in the largest clique in G. That is, the largest k for which there exists a set $S \subseteq V$ for which all edges between pairs of vertices in S are in G.
- Independence number. Written $\alpha(G)$, the independence number is the size of the largest set of vertices in G that has no edges. It is the clique number of the complement graph of G (the graph that has edges exactly where G does not).
- Chromatic number, written $\chi(G)$. A graph is said to be k-colorable if there is a mapping $f : V \to \{1, \dots, k\}$ so that for every edge (u, v), $f(u) \neq f(v)$. The chromatic number of G is the least k for which G is k-colorable. For example, a bipartite graph is 2-colorable.

Intuitively, one might think that a graph of large girth can be colored with few colors. At the end of lecture, we will see a result of Erdos which tells us this is not true. We will construct the counter-example by the probabilistic method. That is, we will describe a randomized process for constructing a graph, and prove it has the desired properties with non-zero probability. This implies that a graph with the desired properties exists.

11.2 Erdos-Renyi Model

The Erdos-Renyi model is specified by two parameters: the number of vertices in the graph n, and the probability of an edge p. Given n and p, we choose a graph on n vertices by including an edge between each pair of vertices with probability p, independently for each pair. Think of this as flipping a coin for each possible edge. I will write $\mathcal{G}(n, p)$ to denote this distribution, and

$$G \leftarrow \mathcal{G}(n, p)$$

to indicate that G is a random graph chosen from this distribution.

11.3 Markov's Inequality and Expectation

In this lecture, we will focus on using expectations of random variables. Recall that if a variable X has the distribution

$$\Pr[X = x_i] = p_i,$$

then

$$\mathbf{E}[X] = \sum_i x_i p_i.$$

The most important property of expectation is that the expectation of the sum of two variables is always the sum of their expectations:

$$\mathbf{E}[X_1 + X_2] = \mathbf{E}[X_1] + \mathbf{E}[X_2].$$

Note that this assertion requires no assumptions! In particular, X_1 and X_2 *do not* need to be independent. This is what makes it so powerful.

If X is a random variable that can never be negative, then Markov's inequality tells us that for all k

$$\Pr[X \geq k] \leq \mathbf{E}[X]/k.$$

To see why this should be true, note that if the probability that X is greater than k is p, then the expected value of X would have to be at least pk.

We will mainly use the following corollary of Markov's inequality:

$$\Pr[X \geq 1] \leq \mathbf{E}[X].$$

11.4 Other Facts from Probability

We will also use the *union bound*, which says that for events A and B,

$$\Pr[A \text{ or } B] \leq \Pr[A] + \Pr[B].$$

For A and B independent, recall that

$$\Pr[A \text{ and } B] = \Pr[A]\Pr[B].$$

11.5 Clique Number

We will now show that for $p = 1/2$, the clique number of G is at most $(2+\epsilon)\log_2 n + 1$ with high probability, for all $\epsilon > 0$. To do this, fix some $\epsilon > 0$, fix $k = \lceil (2+\epsilon)\log_2 n + 1 \rceil$, and let $S_1, \ldots, S_z$ be the subsets of vertices of size k. So,

$$z = \binom{n}{k}.$$

Let X_i be a random variable that is 1 if S_i is a clique in G. Let

$$X = \sum_i X_i.$$

If $X < 1$, then the largest clique in G has size less than k. To show that this is probably the case, we will prove that $\mathbf{E}[X]$ is very small. To do this, we will prove that $\mathbf{E}[X_i]$ is small for each i. As X_i can only take the values 0 and 1,

$$\mathbf{E}[X_i] = \Pr[X_i = 1].$$

We have $X_i = 1$ only if S_i is a clique, which happens exactly when all of the $\binom{k}{2}$ edges between vertices in S_i appear in the graph. This happens with probability

$$(1/2)^{\binom{k}{2}} = \left((1/2)^{(k-1)/2}\right)^k = \left((1/2)^{(2+\epsilon)\log_2 n/2}\right)^k = \left((1/2)^{(1+\epsilon/2)\log_2 n}\right)^k = \left(n^{-(1+\epsilon/2)}\right)^k.$$

So,

$$\mathbf{E}[X] = \sum_i \mathbf{E}[X_i] = \binom{n}{k}\left(n^{-(1+\epsilon/2)}\right)^k \leq n^k \left(n^{-(1+\epsilon/2)}\right)^k = \left(n^{-(\epsilon/2)}\right)^k = n^{-\epsilon k/2} \to 0.$$

as n goes to infinity. So, in summary

$$\Pr_{G \leftarrow \mathcal{G}(n,1/2)}[\omega(G) \geq (2+\epsilon)\log_2 n + 1] \leq n^{-\epsilon k/2} \to 0.$$

We could of course carry this argument out for general p. This would give

$$\Pr_{G \leftarrow \mathcal{G}(n,p)}[\omega(G) \geq k] \leq \left(np^{(k-1)/2}\right)^k.$$

11.6 Girth

In $\mathcal{G}(n,p)$, the expected number of edges attached to each vertex is $p(n-1)$. Next lecture, we will see that it must be close to this for all vertices. For now, let's set $d = p(n-1)$, and ask ourselves how large the girth can be of a graph in which every vertex has degree d. We will then show that this bound is almost achieved.

For a vertex v, let $N(v)$ be the set of neighbors of v, and let $N^{(k)}(v)$ be the set of vertices that can be reached from v by a path of length at most k. We have $|N(v)| = d + 1$. If $g(G) > 4$, then

$$\left|N^{(2)}\right|(v) = d(d-1) + |N(v)| = d^2 + 1.$$

Similarly, if $g(G) \geq 2k+1$, then

$$\left|N^{(k)}(v)\right| = d(d-1)^k + \left|N^{(k-1)}(v)\right| = d^k + 1.$$

But, there are n vertices, so

$$n \geq \left|N^{(k)}(v)\right| = d^k + 1,$$

which implies

$$k < \log_d n = \frac{\log n}{\log d}.$$

If $d = n^{1/j}$, this gives the bound $g \leq 2j+1$. We will now show that this is approximately tight.

Set $p = n^{(1-\epsilon)/g}/n$, for any $\epsilon > 0$, and choose $G \leftarrow \mathcal{G}(n,p)$. We will prove that few of the vertices of G will be in cycles of length g. It would be unreasonable to hope that there are no short cycles.

(In fact, the analysis from the previous section tells us that the expected number of triangles is $\binom{n}{3}p^3 = n^{3(1-\epsilon)/g}/6 > 1$.)

A g-cycle is described by a sequence of g vertices, giving the first vertex in the cycle, the second, and so on. Actually, each g-cycle has $2g$ descriptions of this form: there are g choices for the first vertex, and two directions in which the cycle can be traversed. Either way, we know that there are most

$$n(n-1)\cdots(n-g+1) \leq n^g$$

possible g-cycles. The probability that any given possible g-cycle appears in G is p^g. So, the expected number of g-cycles is at most

$$n^g p^g = (np)^g = \left(n^{(1-\epsilon)/g}\right)^g = n^{1-\epsilon}.$$

One can show that the expected number of j cycles for $j < g$ is lower. So, the expected number of cycles of length at most g is at most

$$gn^{1-\epsilon}.$$

By Markov's inequality, this means that the probability that G has more than $2gn^{1-\epsilon}$ cycles of length at most g is at most $1/2$, and that the probability that G has more than $n/2g$ cycles of length up to g is at most

$$gn^{1-\epsilon}/(n/2g) = 2g^2/n^{\epsilon}.$$

So, we may conclude that the probability that more than $n/2$ of the vertices are involved in cycles of length up to g is at most

$$2g^2/n^{\epsilon}.$$

Note that all of these bounds go to zero as n grows and g stays fixed.

Consider removing all the vertices in G that are involved in cycles of length up to g. With probability $1 - 2g^2/n^{\epsilon}$, at least $n/2$ vertices remain, and the remaining graph has girth at least g. You might be wondering how many edges are left in the graph. We will later learn techniques that show that with high probability at least $1/4$ of the edges remain.

11.7 High Girth and Chromatic Number

Theorem 11.7.1 (Erdos). *For every g and x, there exists a graph G of girth at least g and chromatic number at least x.*

Proof. As in the previous section, set $p = n^{(1-\epsilon)/g}/n$, and choose G from $\mathcal{G}(n, p)$. Use $\epsilon = 1/2$, so $p = n^{1/2g}/n$.

Then, remove all vertices from G in cycles of length up to g, and call the resulting graph G'. We will show that G' probably has high chromatic number.

We would like to say something like "$\chi(G') \geq \chi(G)$", but I see no reason it should be true. Instead, we use the inequalities

$$\alpha(G) \geq \frac{n}{\chi(G)}, \tag{11.1}$$

and

$$\alpha(G') \leq \alpha(G). \tag{11.2}$$

The first follows from the fact that each color class in a coloring is an independent set, and the largest must have size at least $n/\chi(G)$. The second is because every independent set in G' is also an independent set in G. If we let n' be the number of vertices in G', we may combine these inequalities to find

$$\chi(G') \geq \frac{n'}{\alpha(G')} \geq \frac{n'}{\alpha(G)}.$$

Let's see what we can say about $\alpha(G)$. As $\alpha(G)$ is the clique number of the complement graph of G, we may apply the results from the first section to count the expected number of independent set in G of size a. Let X be the number of independent set in G of size a. We get

$$\mathbf{E}[X] \leq \left(n(1-p)^{(a-1)/2}\right)^a.$$

To estimate this, we will use one of the most important inequalities in probability:

$$1 - p \leq e^{-p}.$$

I suggest you memorize it.

We then compute

$$\begin{aligned}\mathbf{E}[X] &\leq \left(n(1-p)^{(a-1)/2}\right)^a \\ &\leq \left(n(e^{-p})^{(a-1)/2}\right)^a \\ &= \left(n(e^{-p(a-1)/2})\right)^a.\end{aligned}$$

If we set

$$a = \frac{4n \ln n}{n^{1/2g}} + 1,$$

then we get

$$\left(n(e^{-p(a-1)/2}\right)^a = \left(n(e^{-2\ln n}\right)^a = \left(nn^{-2}\right)^a = n^{-a}.$$

So, the probability that $\alpha(G)$ exceeds a is at most n^{-a}. Now, if $\alpha(G) \leq a$ and $n' \geq n/2$, then

$$\chi(G') \geq \frac{n'}{\alpha(G)} \geq \frac{n/2}{\frac{4n\ln n}{n^{1/2g}} + 1} = \frac{n^{1/2g}}{8\ln n + n^{1/2g-1}} \geq \frac{n^{1/2g}}{9\ln n}.$$

If we fix g and let n grow, then this quantity grows as well, and so eventually becomes bigger than x. So, to establish the existence of the desired graph G', we just need to show that with some reasonable probability, $n' \geq n/2$ and $\alpha(G) \leq a$. To do this, we examine the probability of failing. We have

$$\Pr\left[n' \leq n/2 \text{ or } \alpha(G) > a\right] \leq \Pr\left[n' \leq n/2\right] + \Pr\left[\alpha(G) > a\right] \leq 2g^2/n^{\epsilon} + n^{-a} \to 0,$$

as n grows. So, the probability of G' having the desired properties tends to 1, and so the desired graph exists.

In fact, we only needed to show that the probability of G' having the desired properties is greater than 0 □