Graphs and Networks Lecture 3

Erdös-Rényi Random Graphs: Warm Up

Daniel A. Spielman September 9, 2010

3.1 Introduction

In this lecture we will introduce the Erdös-Rényi model of random graphs. Erdös and Rényi did not introduce them in an attempt to model any graphs found in the real world. Rather, they introduced them because they are the source of a lot of interesting mathematics. In fact, these random graphs have many properties that we do not know how to obtain through any efficient explicit construction of graphs. They are also the source of many graphs having counter-intuitive properties.

Over the next two lectures, will see that Erdös-Rényi random graphs have many properties in common with graphs encountered in the real world, and many properties that are very different. We study these graphs for three reasons:

1. we need warm up our probabilistic analysis skills,
2. these provide an elementary model upon which we can build, and
3. they will improve our understanding of what graphs can look like.

Our motivation is *not* to present them as a model of graphs that occur in real life.

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

- Girth. We denote the girth of a graph G by $\gamma(G)$. It is the length of the *shortest* cycle 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. That is, a set of vertices S is *independent* if for all $\{u, v\} \subseteq V$, $(u, v) \notin E$.

$$\alpha(G) = \max \{|S| : S \text{ is an independent set}\}.$$

- Chromatic number, written $\chi(G)$. A coloring of a graph G is a mapping $f : V \to \{1, \ldots, k\}$ so that for every edge (u, v), $f(u) \neq f(v)$. A graph is said to be k-colorable if it has a coloring using only k colors. The chromatic number of G is the least k for which G is k-colorable. For example, a bipartite graph is 2-colorable. Planar graphs are 4-colorable.

These quantities are related. If f is a k-coloring of G, then the set of vertices of a given color are independent. As the largest class must have at least n/k vertices, we know that

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

There are obvious obstacles to coloring a graph with few colors. For example, if there is a set of k vertices that are all attached to each other then at least k colors will be required to color the graph. Such a set of vertices is called a k-clique, and it is an independent set of size k in the complement of the graph G (that is the graph that has edges precisely where G does not). But, there are obstacles that are not as easy to spot. For example, if a graph has an odd cycle then it is not two-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 Erdös which tells us this is not true. We will construct the 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.

This lecture is based on the section entitled "The Probabilistic Lens: High Girth and High Chromatic Number" from the book "The Probabilistic Method" by Noga Alon and Joel Spencer.

3.2 Erdös-Rényi Model

The Erdös-Rényi 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 with bias p 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.

3.3 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

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

For example, the expected degree of a vertex in a graph drawn from the distribution $\mathcal{G}(n, p)$ is $p(n-1)$. Next lecture we will see that the degrees of vertices are tightly concentrated around their expectations. This is one thing that differentiates Erdös-Rényi random graphs from most graphs encoutered in the real world.

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

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

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

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

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:

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

3.4 The union bound

Recall that

$$\Pr\left[A \text{ or } B\right] = \Pr\left[A\right] + \Pr\left[B\right] - \Pr\left[A \text{ and } B\right].$$

We will often use the following elementary consequence of this fact, called the *union bound*:

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

3.5 Independence Number

If every vertex in a graph has degree at most d, then it is easy to find an independent set in the graph of size at least $n/(d+1)$. Choose an arbitrary vertex and add it to S. Now, remove all of its at most d neighbors and their attached edges. At least $n-(d+1)$ vertices remain. Repeat this process until no more vertices remain. The resulting set S will be independent and will have at least $n/(d+1)$ vertices. We will now see that random graphs drawn from the distribution $\mathcal{G}(n,p)$ do not have independent sets that are significantly larger than this bound, where we take $d = p(n-1)$.

We will now show that for $p = 1/2$, the independence number of G is at most $3\log_2 n + 1$ with high probability. If this were a social network, it would mean that it is impossible to find more than $3\log_2 n + 1$ people none of whom know each other. Of course, each node in this graph is connected to approximately half of the others, so it does not resemble a large social network.

To be concrete,fix some $\epsilon > 0$, fix $k = \lceil 3\log_2 n + 1\rceil$, and let $S_1, \ldots, S_z$ be all 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 an independent set in G. Let

$$X = \sum_i X_i.$$

If $X < 1$, then the largest independent set 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 an independent set, which happens exactly when all of the $\binom{k}{2}$ edges between vertices in S_i appear in the graph. This happens with probability

$$(1/2)^{\binom{k}{2}} = \left((1/2)^{(k-1)/2}\right)^k = \left((1/2)^{3\log_2 n/2}\right)^k = \left((1/2)^{(3/2)\log_2 n}\right)^k = \left(n^{-(3/2)}\right)^k.$$

So,

$$\mathbf{E}[X] = \sum_i \mathbf{E}[X_i] = \binom{n}{k}\left(n^{-(3/2)}\right)^k \leq n^k \left(n^{-(3/2)}\right)^k = \left(n^{-(1/2)}\right)^k = n^{-k/2} \to 0.$$

as n goes to infinity. So, in summary

$$\Pr_{G \leftarrow \mathcal{G}(n,1/2)}[\alpha(G) \geq (3\log_2 n + 1] \leq n^{-k/2} \to 0.$$

In fact, one can show that $\alpha(G)$ is tightly concentrated around $2\log_2 n$.

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

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

This probability will be small as long as the term inside the parenthesis is small. We will now show that it is small for

$$k = \frac{3\ln n}{p} + 1.$$

Our proof will exploit the fundamental inequality

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

I suggest you memorize it.

We have

$$n(1-p)^{(k-1)/2} \leq ne^{-p(k-1)/2} = e^{\ln n - p(3\ln n)/2p} = e^{\ln n - (3/2)\ln n} = n^{-1/2}.$$

So,

$$\Pr_{G \leftarrow \mathcal{G}(n,p)}\left[\alpha(G) \geq \frac{3\ln n}{p} + 1\right] \leq n^{-3\ln n/2p}.$$

This certainly goes to zero as n grows large.

3.6 High Girth

We will now prove that there are graphs with both high girth and high chromatic number. In particular, we will prove that for every g and x there exists a graph with chromatic number at least x and girth at least g. We will not give an explicit construction. Instead, we will describe a process that produces such a graph with non-zero probability. That is enough to show that one exists! This sort of argument motivated most of the research in probabilistic combinatorics for decades.

Our approach will be simple. We will first choose a random graph from the distribution $\mathcal{G}(n,p)$ with the carefully chosen

$$p = n^{1/2g-1}.$$

This graph may have small cycles. However, we will show that it does not have too many of them. So, we will remove one vertex from every cycle of length up to g. The remaining graph will not have any small cycles, and it will probably still have at least $n/2$ vertices. We will show that it also has large chromatic number.

The bound on the chromatic number is the easy part, so let's do that part first. Our reasoning will exploit the relation between the independence and chromatic numbers $n/\chi \geq \alpha$. Let G have n vertices and let $G' = (V', E')$ be a graph obtained by removing at most $n/2$ of the vertices of G and all of their attached edges. A set of vertices $S \subseteq V'$ is an independent set in G' if and only if it is an independent set in G. So,

$$\alpha(G') \leq \alpha(G)$$

and

$$\chi(G') \geq \frac{|V'|}{\alpha(G')} \geq \frac{(n/2)}{\alpha(G')} \geq \frac{n}{2\alpha(G)}.$$

Using the argument of the previous section we can show that it is unlikely that $\alpha(G)$ is small. By the argument from the previous section, we know that with probably approaching 1, and certainly at least 3/4,

$$\alpha(G) \leq 3n^{1-1/2g} \ln n.$$

If this happens and $|V'| \geq n/2$ then

$$\chi(G') \geq \frac{n^{1/2g}}{6 \ln n}.$$

For any fixed g $n^{1/2g}$ grows faster than $6 \ln n$. So for sufficiently large n this would give $\chi(G') \geq x$.

We will now 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 $\binom{n}{3}p^3 > 1$.)

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

$$n^g p^g = (np)^g = \left(n^{1/2g}\right)^g = n^{1/2}.$$

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

$$gn^{1/2}.$$

By Markov's inequality, this means that the probability that G has more than $4gn^{1/2}$ cycles of length at most g is at most $1/4$. So, with probability at least $3/4$ we can remove all cycles from G of length up to G by removing at most $4gn^{1/2}$ vertices from G. Call the resulting graph G'. As long as n is large enough we will have

$$4gn^{1/2} \leq n/2,$$

and so with probability at least $3/4$ G' will have at least $n/2$ vertices.

By construction, G' has girth at least g. With probability at least $3/4$, G' has at least $n/2$ vertices and with probability at least $3/4$ the independence number of G is at least $3n^{1-1/2g}\ln n$. As the probability that each of these events fails to happen is at most $1/4$, the union bounds tells us that the probability that either of them fails is at most $1/2$. So, with probability at least $1/2$ G' has at least $n/2$ vertices and $\alpha(G)$ is at least $3n^{1-1/2g}\ln n$, which we saw implies that $\chi(G) \geq x$ if n is sufficiently large.

3.7 Remark

All of these estimates can be tightened considerably. Some by means that are obvious and some by the use of fancy techniques.