Graphs and Networks	Lecture 5
The Giant Component and Diameter	
Daniel A. Spielman	September 16, 2010

5.1 Overview

I have three goals for this lecture:

1. Sketch the proof of the existance of the Giant Component in Erdös-Rényi-random graphs.
2. Introduce random graphs with given degree distributions.
3. Prove that such graphs can have logarithmic diameter.

5.2 Review of Galton-Watson

Let me recall what we proved about the Galton-Watson process, and one thing that I forgot to mention. In particular, I will be concerned with the case in which a cell splits into k children, each of which survives to split themselves with probability $p = c/k$.

Theorem 5.2.1. *Let Y be the number of descendants of the first organism that survive to reproduce (not counting the first organism). If $c < 1$ then for every $u > 0$*

$$Pr\left[Y \geq u\right] \leq \exp\left(\frac{(1-c)^2}{3c}\right)^{-u}.$$

Compared to the variable Z I defined last lecture, we have $Y = Z - 1$.

Here is a theorem that I forgot to state. It's proof is analogous to the proof in the case $k = 2$, and will appear in the first problem set.

Theorem 5.2.2. *Let Y be the number of descendants of the first organism that survive to reproduce (not counting the first organism). If $c > 1$ then there exists a constant β_c, which does not depend upon k, such that*

$$Pr\left[Y = \infty\right] \geq \beta_c.$$

We also proved that in this case Y is unlikely to be moderately large.

Theorem 5.2.3. *Let Y be the number of descendants of the first organism that survive to reproduce (not counting the first organism). If $c > 1$ then for every $u > 0$*

$$Pr\left[u \leq Y < \infty\right] \leq \exp\left(\frac{(c-1)^2}{2c}\right)^{-u}.$$

5.3 $c < 1$: all small components

We finally turn to the examination of the sizes of components in Erdös-Rényi-random graphs $\mathcal{G}(n, p)$. We will begin by considering the case $p = c/(n-1)$ for $c < 1$. In this case, we will see that every component is probably small. We do this by exploring the comonent of some vertex in a breadth-first fashion. To start, pick any vertex. Say v. There are $n-1$ vertices that could be its neighbors, and each is a neighbor with probabilty p.

We identify vertex v with the original organism in a Galton-Watson branching process with $k = n-1$. We then view the neighbors of v as the children of this original organism that survive. It is clear that the number of these is identically distributed. Now, let w be a neighbor of v that we have just identified, and let j be the number of neighbors of v. The number of vertices other than v and its neighbors is $n-j-1$. Each of these has a probability p of being a neighbor of w. We view these as being the children of organism w that survive. Here we see a difference with the Galton-Watson process: cell w splits into $n-j-1$ parts rather than $n-1$. This will only help us as we want to show that the component of v is small. The fact that w has fewer than $n-1$ children only makes the component smaller. After we do this for w, we do it for another neighbor of v. But, we only consider the nodes that are not yet known to be in the component of v as potential children.

If we proceed in this way, we see that the size of the component of vertex v is distributed in the same way as the number of descendants of an organism in a modified Galton-Watson process in which we decrease the number of children that organisms can have. But, we keep p fixed as we decrease this number. Let $C(v)$ be the random variable giving the number of vertices in the component of v, other than v itself. For every $u > 0$ we have

$$\Pr\left[C(v) \geq u\right] \leq \Pr\left[Y \geq u\right] \leq \exp\left(\frac{(c-1)^2}{2c}\right)^{-u}.$$

For some constant α we can set $u = \alpha \ln n$ and make this probability $1/n^2$. We could then say

$$\Pr\left[G \text{ has a component with more than } \alpha \ln n \text{ vertices}\right] \leq \sum_{v \in V} \Pr\left[C(v) \geq \alpha \ln n\right] \leq 1/n.$$

5.4 $c > 1$: the giant component

I will now sketch the argument that when $p = c/n$ the graph has a giant component. For simplicity, we will just consider the case $c \leq 2$. We begin in the previous section by examining the component of some vertex, say v. We again identify our exploration of the component of v with a modified Galton-Watson process. The difficult here is that the modified process decreases the size of components, while we now want to show that there is a big component. To handle this, set $d = \sqrt{c}$. As $c > 1$ we also know that $d > 1$. We will compare the exploration from v with a modified Galton-Watson process with $k = (2-d)n$ and probability $p' = d/(n-1)$ of cell survival. So long as our process has explored fewer than $(d-1)n$ vertices, there will be at least $(2-d)n$ vertices available to serve as neighbors of each vertex we explore. Thus, the exploration of the neighborhood of v resembles modified Galton-Watson process in which some cells split into more than k children and have a larger chance of survival. This means that they are only more likely to have a large component. This analysis breaks down after we have explored $(d-1)n$ vertices. But we will probably have identified a giant component by this point.

Order the vertices and explore the component of the first vertex. There is a constant chance that its Galton-Watson process suggests an infinite component, in which case we identify a component of size at least $(d-1)n$. Otherwise, the chance that its component has size more than $\alpha \ln n$ is at most $1/n^2$. So, we can safely assume that this is the case. In this case, we proceed with the next vertex that has not yet been placed in a component. After doing this some logarithmic number of times, it is highly likely that we will have found a large component. Moreover, a logarathmic number of small component will only remove $O(\log^2 n)$ vertices from the process, which is negligible for large n. So, they just won't matter.

Finally, the reason that it is unlikely that there are two large component is that they would probably have an edge between them.

5.5 Graphs with given degree distribution

Recall that Erdös-Rényi-random graphs have binomial degree distributions, whereas most graphs we observe have degree distributions with heavy tails. That is, they probably have some vertices of very high degrees. In this lecture and in one lecture next week we will explore models that produce such distributions.

In this lecture, we will do it by just fixing the degree distribution. That is, we will set number k_i and choose a random graph that has k_i vertices of degree i, for each i. Of course, we require that

$$\sum_i k_i = n.$$

You may view this as cheating. But, graphs chosen in this way have been very helpful in a number of applications.

Here's an idea of how you might choose such a graph. Think of each edge in a graph as a wire with two ends, each of which is plugged into a vertex. So, a vertex of degree d should have d sockets. To choose a random graph having a specified degree distribution, we create k_i vertices with i sockets, for each i. We then start adding in edges. We pick two available sockets uniformly at random, and connect them by an edge. This almost works. The only things that could go wrong is that we could create loops or multiple edges. There are two ways of avoiding this problem. The simplest is to hope that this is unlikely and just re-sample if it happens. Alternatively, one could pick such a problematic edge, pick another edge at random, and then swap a pair of their connections. Unless there are vertices of very high degree one shouldn't need to do this too many times to fix the problem (reference Kim and Wormald).

5.6 Diameter

It has been observed that many graphs encountered "in practice" have small diameters. We will now observe that random graph chosen with fixed degree distributions probably have logarithmic diameter, as long as the minimum degree is not too small. Minimum degree at least 3 is sufficient. We will prove it for minimum degree at least ?.

Our proof will work by establishing that these graphs have a much stronger property: they are probably expanders. To define this, consider any subset of vertices S. We will prove that if $|S| \leq n/4$ then the size of the union of S with its neighbors probably has size at least $2\,|S|$. That is, each vertex probably has at least 1 neighbor, these together probably have at least 2 neighbors, and so on. So, every vertex probably has at least 2^k other nodes within distance k of itself, provided that $2^k < n/2$.