M375T/M396C: Topics in Complex Networks	**Spring 2013**
Lecture 04 — 01/24/13	
Lecturer: Ravi Srinivasan	*Scribe: George Wang*

4.1 Erdős-Rényi random graphs

In the previous lectures, we considered branching processes as an elementary model of a random graph. However, this only produces trees, which lack sufficient structure to describe most networks. To allow for non-tree structures we will consider another simple mechanism for producing undirected random graphs: the Erdős-Rényi model.

To begin, define $G(n,p)$ be the random graph with n nodes such that each of the $\binom{n}{k}$ possible edges exists with fixed probability $0 \leq p \leq 1$, independently of all other edges. That is, if

$$I_{(i,j)} = \begin{cases} 1 & \text{if } (i,j) \in E_n \\ 0 & \text{else} \end{cases}$$

is the indicator function that edge (i,j) exists, we have that $\left\{I_{(i,j)}\right\}_{1\leq i<j\leq n}$ are iid Bernoulli(p) random variables. The average number of edges in $G(n,p)$ is therefore

$$\mathbb{E}\left[\sum_{1\leq i<j\leq n} I_{(i,j)}\right] = \sum_{1\leq i<j\leq n} \mathbb{E}\left[I_{(i,j)}\right] = \binom{n}{k}p.$$

We can think of this as a global property of the random graph.

For a more local version, let us first introduce the following definition:

Definition. For a given node $u \in V_n$, let

$$N(u) = \{v \in V_n \backslash \{u\} : (u,v) \in E_n\}$$

be the set of *neighbors* of u. The *degree* of u is defined as $D_u = |N(u)| = \sum_{v\in V_n\backslash\{u\}} I_{(u,v)}$.

In the Erdős-Rényi model, we see that for any node i in $G(n,p)$,

$$D_i = \sum_{j\neq i} I_{(i,j)} \sim \text{Binomial}(n-1,p).$$

That is,

$$\mathbb{P}\left(D_i = k\right) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

and the average degree of a node is

$$\mathbb{E}[D_i] = (n-1)p.$$

Now note that if p is fixed and nonzero, the average degree goes to infinity as $n \to \infty$. This is obviously incompatible with the local structure of any realistic network. Since we will be interested in the limit of large graphs, to obtain a model in which nodes have finite average degree we must allow p to depend on n in such a manner that $p = p(n) \to 0$ as $n \to \infty$. In particular, if $p(n) = \lambda/n$ for some constant $\lambda > 0$ then $\mathbb{E}[D_i] \to \lambda$ and

$$\mathbb{P}(D_i = k) \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{as} \quad n \to \infty.$$

Therefore, the degree of a node has Poisson(λ) distribution in the limit. For this reason, Erdős-Rényi random graphs are also known as *Poisson random graphs*.

4.1.1 Threshold functions and phase transitions

As we have seen, random graphs with nontrivial local structure requires that we take the edge probabilities to zero at a rate that depends on n. In this framework, how can we rigorously establish the existence of phase transitions?

Definition. $t(n)$ is a *threshold function* for a given property A of $G(n,p)$ if

1. $\mathbb{P}(A) \longrightarrow 0$ if $p(n) = o(t(n))$ as $n \to \infty$
2. $\mathbb{P}(A) \longrightarrow 1$ if $p(n) = \omega(t(n))$ as $n \to \infty$.

Recall that $p(n) = o(t(n))$ if and only if $p(n)/t(n) \to 0$, and that $p(n) = \omega(t(n))$ if and only if $p(n)/t(n) \to +\infty$. Therefore, $t(n)$ is a threshold if A never occurs when $p(n)$ goes to zero faster than $t(n)$, and A always occurs when $p(n)$ goes to zero slower than $t(n)$. This allows us to say that a phase transition of A occurs at $t(n)$ as $n \to \infty$. For example, one can show the following hierarchy of thresholds occur:

A	$t(n)$
an edge exists	n^{-2}
a tree of with 3 nodes exists	$n^{-3/2}$
a tree with r nodes exists	$n^{-r/(r-1)}$
a cycle of length r exists	n^{-1}

More generally, threshold functions exist for any *monotone properties* of $G(n,p)$—property A is said to be monotone if when G' is a subgraph of G that satisfies A, we have that G itself satisfies A. It is easy to see that all of the properties mentioned above are monotone.

4.1.2 Average analysis

Although we will not actually prove the existence of some of these thresholds until the next lecture, let us at least try to justify them using what is known as an *average analysis*. Here, we try to determine the value of the threshold by studying the expected value of a quantity of interest.

For example, consider the property that an edge exists. Let $N^{(1)}$ be the number of edges (i.e., trees with 2 nodes) in $G(n, p(n))$. Then $N^{(1)} = \sum_{i \neq j} I_{(i,j)}$ and

$$E\left[N^{(1)}\right] = \binom{n}{2} p(n) = \frac{n(n-1)}{2} p(n) = \Theta\left(n^2 p(n)\right).$$

Therefore, the expected number of edges goes to zero if $p(n) = o\left(n^{-2}\right)$, while it goes to infinity if $p(n) = \omega\left(n^{-2}\right)$. This hints that the appropriate threshold for the existence of an edge is $t(n) = n^{-2}$.

Similarly, to find a threshold for the existence of a tree with 3 nodes we proceed as follows. Let $N^{(2)} = \sum_{i,j,k \text{ distinct}} I_{(i,j)} I_{(j,k)}$ be the number of trees with 3 nodes. Then

$$\begin{aligned}
\mathbb{E}\left[N^{(2)}\right] &= \sum_{i,j,k \text{ distinct}} \mathbb{E}\left[I_{(i,j)} I_{(j,k)}\right] \\
&= \sum_{i,j,k \text{ distinct}} \mathbb{E}\left[I_{(i,j)}\right] \mathbb{E}\left[I_{(j,k)}\right] \\
&= \binom{n}{3} p(n)^2 \\
&= \Theta\left(n^3 p(n)^2\right).
\end{aligned}$$

The appropriate threshold should therefore be $t(n) = n^{-3/2}$. We will leave it as an exercise for the reader to show for a tree with r nodes, an analogous calculation yields the threshold $t(n) = n^{-r/(r-1)}$.

Finally, consider the existence of a cycle of length r. With

$$N_{\text{cycles}} = \sum_{i_1, \ldots, i_r \text{ distinct}} I_{(i_1,i_2)} \cdots I_{(i_{r-1},i_r)} I_{(i_r,i_1)}$$

the number of such cycles, it is straightforward to see that

$$\begin{aligned}
\mathbb{E}\left[N_{\text{cycles}}\right] &= \sum_{i_1, \ldots, i_r \text{ distinct}} \mathbb{E}\left[I_{(i_1,i_2)}\right] \cdots \mathbb{E}\left[I_{(i_{r-1},i_r)}\right] \mathbb{E}\left[I_{(i_r,i_1)}\right] \\
&= \binom{n}{r} p(n)^r \\
&= \Theta\left(n^r p(n)^r\right).
\end{aligned}$$

Note that we now get a threshold $t(n) = n^{-1}$, independent of r! This implies that as we increase the edge probabilities p, we should expect to see arbitrarily large trees in $G(n, p)$ before even one small cycle (e.g., a triangle) appears. In particular, our argument hints that Erdős-Rényi random graphs are typically tree-like, and are thus insufficient to describe structures with clusters of mutually connected nodes. We will discuss this at length in subsequent lectures.

References

1. Acemoglu, D., & Ozdaglar A. (2009). 6.207/14.15, Networks, Fall 2009 [Lecture notes]. Retrieved from:
   `http://economics.mit.edu/faculty/acemoglu/courses/`

2. Chaintreau, A. (2012). COMS 4995-1, Introduction to Social Networks, Fall 2012 [Lecture notes]. Retrieved from:
   `http://www.cs.columbia.edu/~augustin/`

3. Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutat Int. Kzl, 5, 17-61.

4. Jackson, M. O. (2010). Social and economic networks. Princeton University Press.

Last edited: March 21, 2013