Lecture 2 (October 12, 2005)

prepared by Paul Constantine

1. Scale-free Graphs

We motivate this lecture with the following question: Why does the distribution of the degrees of nodes in the graph of the Internet or world wide web follow a power law distribution? Some claim to have observed this result, while others attempt to reason to this conclusion. From observation, if the distribution of degrees does not follow a power law it is heavy-tailed, i.e. there are many vertices with small valency and few vertices with high valency. Where does this structure come from? More generally, why do we find power laws everywhere? The following process attempts to give a simple mathematical model for explaining this behavior:

Start at time 0 with a vertex connected to itself with a self-loop. Repeat the following procedure until time n: At time t, a new node arrives and attaches to m of the existing vertices. The probability of attaching to vertex i is proportional to the degree of i at time $t-1$. By simulating the above process, Barabasi and Albert [1] found out that with high probability the resulting graph will have a power law degree distribution. In their results, they also found that $\Pr(\text{degree}(i) = k)$ is proportional to $k^{-\alpha}$, where α was around 2.9. They also gave the following heuristic argument showing $\alpha = 3$.

Let us observe the evolution of the expected degree of node i. Let $K_i(t)$ equal degree of node i at time t. If we assume that we can take the derivative of the $K_i(t)$.

$$\frac{\partial K_i(t)}{\partial t} = \frac{mK_i(t)}{\sum_{i=1}^{t} K_i(t)} = \frac{mK_i(t)}{2mt}$$

Therefore

$$\frac{\partial K_i}{\partial t} = \frac{K_i}{2t} \quad \text{where} \quad K_i(t_i) = m$$

We can solve this differential equation:

$$K_i(t) = \left(\frac{t}{t_i}\right)^{1/2} m$$

[1] A. L. Barabasi and R. Albert, Emergence of scaling in random networks, **Science** 286 (1999), 509-512

Also

$$\begin{aligned}\Pr(K_i(t) > k) &= \Pr\left(\left(\frac{t}{t_i}\right)^{1/2} m > k\right)\\ &= \Pr(t_i < \frac{m^2}{k^2} t)\\ &= \frac{m^2}{k^2}\end{aligned}$$

Therefore

$$\Pr(K_i(t) = k) = \frac{2m^2}{k^3}$$

which gives α. A more rigorous analysis comes from Bollobas, Riordan, and Spencer, which is outlined here.

Define $N(k,t)$ equal to the expected number of vertices of degree k at time t. Then compute

$$N(k,t) - N(k,t-1) = \frac{N(k-1,t-1)m(k-1)}{2mt} - \frac{N(k,t-1)mk}{2mt} + \delta_{k,m}$$

where $\delta = 1$ if $k = m$, and 0 otherwise. This problem is solvable with a lot of calculations. Skipping the boring details, we find that as $t \to \infty$,

$$\frac{N(k,t)}{t} = \frac{2m(m+1)}{k(k+1)(k+2)}$$

This argument is only for expected values. Using Azuma-Hoeffding inequality, one can show that the degree sequence of the graph converges to the above limit with high probability. The main idea of the above process, also known as the *preferential attachment* is "rich gets richer". You can also see results with a similar spirit (such as Yule process and Simon's work) in Durrett's book (Chapter 4, Section 2). But the most interesting example is Polya's urn scheme.

2. Polya Urns Problem

Suppose there is an urn with r red balls and b blue balls. At each step, take a ball from the urn, and note its color. Return it to the urn with k copies of the same color. The growth of the number of red/blue balls is proportional to the number of balls in the urn.

Let's study the process for $k = 1$. We want to compute the probability of observing i red balls in n trials. Try a simple computation;

$$\begin{aligned}\Pr(\text{observing 2 red balls in 3 trials}) &= Pr(RBR) + Pr(RRB) + Pr(BRR)\\ &= \frac{r}{r+b} \times \frac{b}{r+b+1} \times \frac{r+1}{r+b+2}\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{r}{r+b} \times \frac{r+1}{r+b+1} \times \frac{b}{r+b+2} \\
&+ \quad \frac{b}{r+b} \times \frac{r+1}{r+b+1} \times \frac{r}{r+b+2} \\
&= \quad \binom{3}{2} \frac{r(r+1)b}{(r+b)(r+b+1)(r+b+2)}
\end{aligned}
$$

And in general

$$
\begin{aligned}
\Pr(\text{observing } i \text{ reds in } n \text{ trials}) &= \binom{n}{i} \frac{r(r+1)\ldots(r+i-1)b(b+1)\ldots(b+n-i-1)}{(r+b)(r+b+1)\ldots(r+b+n)} \\
&= \binom{n}{i} \frac{(r+b-1)!}{(r-1)!(b-1)!} \frac{(r+i-1)!(b+n-i-1)!}{(r+b-n-1)!}
\end{aligned}
$$

Take the limit as $n \to \infty$, and say $\frac{i}{n} \to x$ to get:

$$
\frac{(r+b-1)!}{(r-1)!(b-1)!} x^{r-1} (1-x)^{r-1} \frac{1}{n}
$$

This is a beta distribution $\beta(r, b)$. So the limit of the process is a random variable. But what does this have to do with power law graphs?

Think of the Preferential attachment process as the following. Suppose you have an urn with balls of type k, for $k = 1, ..., n$. The number of balls of type k in the urn corresponds to the degree of the k'th node. The stochastic process corresponding to the preferential attachment with $m = 1$ will be as follows: At time t, take a ball from the urn. Return it with an extra copy of the same type plus a ball of type t.

Let's study the ratio of balls of type k over the total number of balls of type $1, \ldots, k$. Say blue balls are of type $1, \ldots, k$ and red balls are of type k. This is a simple Polya Urn scheme. The rest is easy. The number of balls of type k over the total number of balls of type $1, ..., k-1$ is

$$
\psi(k) = \beta(1, 2k-1)
$$

The expected degree of vertex k is given by

$$
\phi_k = \psi_k \prod_{j=k}^{n} (1 - \psi_j)
$$

This gives a much simpler procedure for generating a graph resulting from a preferential attachment.