October 10, 2007
Prepared by Adam Guetz

Preferential Attachment (or Proportional Attachment)

The preferential attachment model was introduced by Barabási and Albert as a mechanism to explain the power-law degree distributions in complex real-world networks. The idea is that as new nodes arrive in the network, they attach to existing nodes with probability proportional to their degree. Intuitively, this "rich-get-richer" scheme should lead to a power-law, since most nodes will have few neighbors but a few nodes will have many neighbors.

Formally, let $G_1, G_2, \ldots, G_t, \ldots$ be a sequence of graphs. Fix $m > 0$ constant. We define intermediate graphs $G_t = \Gamma_{t,0}, \Gamma_{t,1}, \Gamma_{t,2}, \ldots, \Gamma_{t,m} = G_{t+1}$ such that $\Gamma_{t,j}$ is obtained from $\Gamma_{t,j-1}$ by adding a link between v_t and v_l where v_l is chosen randomly from $v_1, v_2, \cdots v_t$ with probability $\frac{deg_(v_l)}{2(m(t-1)+j-1)}$. We initialize the process by setting G_1 to be a single vertex with m loops.

Barabási and Albert ran Monte Carlo simulations of the preferential attachment process and estimated a power-law exponent of $\alpha = 2.9 \pm .1$. This led to the conjecture that the true exponent is $\alpha = 3$, for which they gave a clever heuristic "mean-field" argument. Bollobás, Riordan, Spencer and Tusnády were later able to give a more rigorous proof.

Theorem 3.1 *As $t \to \infty$, the probability of node $u \in G_t$ having degree k is proportional to k^{-3}.*

We begin by giving the heuristic argument of Barabási and Albert. Let $d_i(t)$ be the degree of node i at time t:

$$d_i(t) = deg(v_i), v_i \in G_t.$$

If we think of this as a continuous process and take the derivative, we get:

$$\frac{\delta d_i(t)}{\delta t} = \frac{m d_i(t)}{\sum_{j=1}^{n} d_j(t)} = \frac{d_i(t)}{2t} \tag{1}$$

Let t_i be the time that node i arrives. We can solve (1) to get:

$$d_i(t) = m\sqrt{\frac{t}{t_i}} \tag{2}$$

Assuming that t_i is distributed uniformly on $[0, t]$, then

$$\Pr\left(d_i(t) > k\right) = \Pr\left(m\sqrt{t} > k\sqrt{t_i}\right) = \Pr\left(t_i < \frac{m^2 t}{k^2} = \frac{m^2}{k^2}\right) \tag{3}$$

Differentiating (3) gives

$$\Pr(d_i(t) = k) = \frac{2m^2}{k^3}, \tag{4}$$

which is the desired exponent.

Proof: Let $Z(k,t)$ be the random variable indicating the number of vertices of degree k in G_t. Let $N(k,t) = \mathrm{E}(Z(k,t))$. Then

$$N(k,t) - N(k,t-1) = \frac{m(k-1)N(k-1,t-1)}{2m(t-1)} - \frac{mkN(k,t-1)}{2m(t-1)} + 1_{m=k} + \epsilon(k,t). \tag{5}$$

The first term corresponds to expected number of nodes of degree $(k-1)$ that will aquire an edge at time t, while the second term corresponds to nodes of degree k that will gain an edge at time t. The third term corresponds to the new node entering the network. $\epsilon(k,t)$ accounts for the possibility of multple edges and self loops, and can be bounded by

$$|\epsilon(k,t)| = O\left(\sum_{i=2}^{m} \frac{(k-i)^i N_{k-i}(t)}{(mt)^i}\right) = O\left(\frac{k}{t}\right) = O\left(t^{-1/2}\right),$$

since $k = O(t^{1/2})$.

Solving the recurrance relationship (5) gives

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

which is asymptotically power-law with exponent $\alpha = 3$.

All that is left is to show that $Z(k,t)$ is concentrated around its expectation w.h.p. To prove this, we define a martingale

$$Z_i(k,t) = \mathrm{E}(Z(k,t)|Y_1, Y_2, \ldots, Y_{i-1}),$$

where the Y_i's are the random choices made at time i. This type of martingale is known as a "Doob martingale", and is a standard tool for showing concentration results. To see that it's a martingale, note that

$$\mathrm{E}[Z_i(k,t)|Z_1,\ldots,Z_{i-1}] = \mathrm{E}[\mathrm{E}(Z(k,t)|Y_1,\ldots,Y_{i-1})|Z_1,\ldots,Z_{i-1}] = Z_{i-1}.$$

We can note that the martingale differences are bounded

$$|Z_i(k,t) - Z_{i-1}(k,t)| \leq 2m,$$

then evoke the Azuma-Hoeffding inequality to show concentration and prove the theorem:

Lemma 3.1 *(Azuma-Hoeffding inequality). Let* $(X_t)_{t=0}^n$ *be a martingale with* $|X_{t+1} - X_t| \leq c$ *for* $t = 0, \ldots, n-1$. *Then*

$$\Pr(|X_n - X_0| \geq x) \leq \exp\left(-\frac{x^2}{2c^2 n}\right).$$

■