Growth Model

Lecture delivered by: Prof. Niloy Ganguly
Scribed by: Himika Biswas (05CS6027)

19-04-2006

1 Price's Model

In this section we shall discuss about another model whose primary goal is also to explain network properties. In this model also, the networks typically grows by the gradual addition of vertices and edges in some manner intended to reflect growth processes that might be taking place on the real networks. In this model also, edges are added to the network preferentially between pairs of vertices that have another third vertex as a common neighbor.

Derek de Solla Price (1965) studied the network of citations between scientific papers and found that both in and out-degrees follow *Power-law* distributions. Based on this observation, Price explained *Power-law* degree distributions as "the rich get richer" when the amount you get goes up with the amount you already have. The important contribution of Price's work was to take the ideas of Simon and apply them to the growth of a network. His idea was that the rate at which a paper gets new citations should be proportional to the number that it already has. This can be explained as the probability that one comes across a particular paper whilst reading the literature will presumably increase with the number of other papers that cite it, and hence the probability that you cite it yourself in a paper that you write will increase similarly. The same argument can be applied to other networks also.

Consider a directed graph of n vertices, such as a citation network. Let p_k be the fraction of vertices in the network with in-degree k, so that $\sum_k p_k = 1$. New vertices are continually added to the network, though not necessarily at a constant rate. Each added vertex has a certain out-degree and this out-degree is fixed permanently at the creation of the vertex. The out-degree may vary from one vertex to another, but the mean out-degree, which is denoted m, is a constant over time.The value m is also the mean in-degree of the network: $\sum_k kp_k = m$. In the simplest form of cumulative advantage process the probability of attachment of one of our new edges to an old vertex-i.e., the probability that a newly appearing paper cites a previous paper-is simply proportional to the in-degree k of the old vertex. But since each vertex starts with in-degree zero, and hence would forever have zero probability of gaining new edges. To overcome this problem, Price suggests that the probability of attachment to a vertex should be proportional to $k + 1$, which he justifies for the citation network by saying that one can consider the initial publication of a paper to be its first citation (of itself by itself). Thus the probability of a new citation is proportional to $k+1$. The probability that a new edge attaches to any of the vertices with degree k is thus $\frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1}$. The mean number of new citations per vertex added is simply m, and hence the mean number of new citations to vertices with current in-degree k is $(k+1)p_k \frac{m}{m+1}$. The number np_k of vertices with in-degree k decreases by this amount, since the vertices that get new citations become vertices of degree $k+1$. However, the number of vertices of in- degree k increases because of the transformation of the vertices previously of degree $k-1$ to k that have also just acquired a new citation. If we denote by $p_{k,n}$ the value of p_k when the graph has n vertices, then the net change in np_k per vertex added is $(n+1)p_{k,n+1} - np_{k,n} = [kp_{k-1,n} - (k+1)p_{k,n}]\frac{m}{m+1}$, for $k \geq 1$ This equation is called **Master equation**. Solving this equation, we get

$p_k = [kp_{k-1} - (k+1)p_k]\frac{m}{m+1}$ for $k \geq 1$,
and $p_k = 1 - p_0\frac{m}{m+1}$ for $k = 0$.

Rearranging, we find $p_0 = \frac{m+1}{2m+1}$ and $p_k = \frac{p_{k-1}k}{k+2+\frac{1}{m}}$ or $p_k = \frac{k(k-1)...1}{(k+2+\frac{1}{m})...(3+\frac{1}{m})}p_0 = (1+\frac{1}{m})\beta(k+1, 2+\frac{1}{m})$, where $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is Legendre's beta-function, which goes asymptotically as a^{-b} for large a and constant b, and hence $p_k \sim k^{(2+\frac{1}{m})}$. So it can be concluded that in the limit of large n, the degree distribution has a *Power-law* tail with exponent $\gamma = 2 + \frac{1}{m}$.