Lecture 3
Preferential Attachment

3.1 Barabasi-Albert Model

The ER network is completely random and lacking any internal structure. In a sense, it's the ultimate equitable network; Every node has roughly the same degree and properties. The exponential random graph is equitable in a different sense. As it is developing, each node has the same probability of attracting new edges and, as a consequence, the probability that any one node attracts more edges than any other is exponentially small.

However, we know that real networks are not so equitable. Consider for example, the World Wide Web, it contains personal blogs that no one links to, and sites such as CNN, Google, etc... that everybody links to. Such diversity can never be explained by a model as simple as the ER model.

A simple mechanism that was proposed to study real world networks is known as preferential attachment, or "the rich get richer". Under this model, new nodes are continuously added to the network. Each new node creates a link to one of the older nodes. However, the endpoint of this new link is chosen proportionally to its current degree. Before we calculate the final degree distribution of the network created by this process, let us consider how the degree of a given node evolves. In particular, what is the probability that node s has degree k at time $t+1$? Since we only add one new edge at each time step, there are only two possibilities:

- Node s has degree k at time t, and it fails to attract the new edge
- Node s has degree $k-1$ at time t, and it receives the new edge

Mathematically, this can be described as:

$$p(s,k,t+1) = \left[1 - \frac{k}{2t}\right] p(s,k,t) + \frac{k-1}{2t} p(s,k-1,t) \tag{3.1}$$

where $p(s,k,t+1)$, represents the probability that node s has degree k at time $t+1$, and the factor of $2t$ is given by the normalization condition for the probability of receiving an edge:

$$\sum_i k_i = 2t$$

since at time t we have t edges in the network. Summing both sides of Eq. 3.1 over all values of s:

$$\sum_s p(s,k,t+1) = \left[1 - \frac{k}{2t}\right] \sum_s p(s,k,t) + \frac{k-1}{2t} \sum_s p(s,k-1,t)$$

we can write:

$$(t+1)\,P(k,t+1)-\delta_{k,1}=\left[1-\frac{k}{2t}\right]tP(k,t)+\frac{k-1}{2t}tP(k-1,t)$$

Simplifying, we obtain:

$$(t+1)\,P(k,t+1)-tP(k,t)=\frac{1}{2}\left[(k-1)\,P(k-1,t)-kP(k,t)\right]+\delta_{k,1}$$

Taking the infinite time limit as above and simplifying:

$$P(k)+\frac{1}{2}\left[kP(k)-(k-1)\,P(k-1)\right]=\delta_{k,1}$$

Now we identify the term in square brackets as the discrete version of the derivative:

$$\lim_{\Delta\to 0}\frac{f(x+\Delta)-f(x)}{\Delta}=\frac{df(x)}{dx}$$

Our implicit expression for the degree distribution is then:

$$P(k)=-\frac{1}{2}\frac{d\left[kP(k)\right]}{dk} \tag{3.2}$$

This form of the expression, relating the function with it's own derivative times the variable itself indicates that the solution must be of the form:

$$P(k)=Ak^{\gamma}$$

Substituting in Eq. 3.2, we obtain:

$$Ak^{\gamma}+\frac{1}{2}A\,(\gamma+1)\,k^{\gamma}=0$$

Canceling similar terms:

$$\gamma+1=-2$$

Giving us the value of the degree exponent:

$$\gamma=-3$$

and the expression for the degree distribution becomes:

$$P(k)=Ak^{-3}$$

The value of the constant A can be found using the normalization condition. In the limit of an infinite network, we have:

$$A\int_1^{\infty}k^{-3}dk=-\frac{A}{2}\,k^{-2}\Big|_1^{\infty}\Rightarrow A=2$$

and finally:

$$P(k)=2k^{-3}$$

Taking the logarithm of both sides:

$$\log\left[P\left(k\right)\right] = \log\left(2\right) - 3\log\left(k\right)$$

and defining:

$$\begin{aligned} x = &\ \log\left(k\right) \\ y = &\ \log\left[P\left(k\right)\right] \end{aligned}$$

we find that when we plot the degree distribution on a Log-Log scale, we expect to see a straight line $y = a - \gamma x$ where the slope is the exponent.

The average degree is then:

$$\langle k\rangle = \int_1^\infty kP\left(k\right) = 2\int_1^\infty k^{-2}dk = 2$$

but now we can ask how representative this degree is. In particular, what is the standard deviation of the degree distribution? The standard deviation is given by:

$$\sigma\left(k\right) = \langle k^2\rangle - \langle k\rangle^2$$

and since we already know what the average degree is, we need only calculate the average degree squared:

$$\langle k^2\rangle = 2\int_1^\infty k^{-1}dk = 2\log\left(k\right)|_1^\infty = \infty$$

meaning the the standard deviation is infinite! What this means is that the distribution is so broad, that the average value is meaningless. But what exactly, do we mean by broad? Let us calculate the probability of finding a node with degree equal to $10\langle k\rangle$ as we did above for the ER network. In this case we have:

$$P\left(10\langle k\rangle\right) = 2\left(10\langle k\rangle\right)^{-3} = 2.5\times 10^{-4}$$

22 orders of magnitude larger than for the ER network!

3.2 Non-Linear Preferential Attachment

The Barabasi-Albert model, while a powerful first step in the right direction, is limited in one important aspect: It only allows for a single exponent of the degree distribution. However, it will well known that Networks in the real world have varying degree distributions, usually in the range between 2 and 3, but with larger exponents in several cases. How then can we develop a model that is able to account for arbitrary degree distribution exponents? Perhaps the answer lies in a non-linear form of the preference function. Let us consider the case where nodes are chosen preferentially to a power, α, of the degree of each node. In this case, we may write

$$p\left(s,k,t+1\right) = \left[1 - \frac{k^\alpha}{M_t}\right]p\left(s,k,t\right) + \frac{\left(k-1\right)^\alpha}{M_t}p\left(s,k-1,t\right)$$

where M_t is a time dependent normalization constant:

$$M_t = \sum_i k_i^\alpha \equiv \sum_k N\left(k,t\right)k^\alpha = ?$$

which depends directly on the number of nodes of degree k, $N(k,t) = N \cdot P(k,t)$. Unfortunately, there is no closed form expression for this normalization constant, but progress has been made numerically. Two regimes have to be taken into account. For $\alpha < 1$, M_k can be approximated by:

$$M_t \approx \mu t$$

resulting in a degree distribution of the form:

$$P(k) = \frac{\mu}{k^\alpha} \prod_{j=1}^{k} \left(1 + \frac{\mu}{j^\alpha}\right)^{-1}$$

that, numerically, is similar to a stretched exponential.

For $\alpha > 1$, there is no way to approximage M_t analytically. Numerical results show that this form of preferential attachment causes a small number of nodes to attract all the incoming edges. For α in the range:

$$\frac{l+1}{l} < \alpha < \frac{l}{l-1}$$

the number of nodes with degree larger than l is finite and the number of nodes of degree k is given by:

$$N(k) \sim t^{k(k-1)\alpha}$$

for $k \leq l$.

3.3 Attractiveness Model

We have just shown that the scale free nature of the resulting network is extremely sensitive on a linear form of preferential attachment, and that a simple dependence of the form suggested by Barabasi and Albert will necessarily lead us to a fixed exponent of -3.

Even though the form of preferential attachment has to be linear, there is nothing stopping us from adding a constant factor, A, to the attractiveness of each node. In this case, we have:

$$p(s,k,t+1) = \left[1 - \frac{k+A}{(1+A)t}\right] p(s,k,t) + \frac{(k-1)+A}{(1+A)t} p(s,k-1,t)$$

Following a similar approach to the one above, we find, for the time dependent degree distribution:

$$(t+1)P(k,t+1) - \delta_{k,1} = \left[t - \frac{k+A}{1+A}\right] P(k,t) + \frac{(k-1)+A}{1+A} P(k-1,t)$$

In the large time limit, when the degree distribution has reached a stationary state, we obtain:

$$P(k) = \frac{(k-1)+A}{1+A} P(k-1) - \frac{k+A}{1+A} P(k) + \delta_{k1}$$

where we identify once again the discrete derivative with respect to k:

$$P(k) + \frac{1}{1+A} \frac{d}{dk} [(k+A)P(k)] = \delta_{k,1}$$

suggesting an ansatz of the form:

$$P(k) = B(k+A)^\alpha$$

Substituting, we obtain:

$$B\,(k+A)^{\alpha} + B\,(k+A)^{\alpha}\,\frac{1+\alpha}{1+A} = \delta_{k,1}$$

and cancelling similar terms (for $k > 1$):

$$1 + \frac{1+\alpha}{1+A} = 0$$

which finally results in:

$$\alpha = -(2+A)$$

meaning that we can obtain an arbitrary degree distribution exponent by simply varying the value of the attractiveness, A.