Growth of Networks

Rishabh Singh 04CS1015, Kota Nagarjun Rao 05CS1034

Indian Institute of Technology, Kharagpur

1 Introduction

We examine a class of models which helps explain the observed properties of real-world networks, such as degree sequences or transitivity. The models help us to understand how the network has evolved to have these properties. In these models, the network grows by gradual addition of vertices and edges in some manner which tries to model the growth processes taking place in the real-world networks, which in turn lead to the characteristic structural properties of the network.

In the past, a lot of work has been done for models of network transitivity that uses *triadic closure* processes. These models consider preferential attachment of edges between the pair of vertices such that there exists a vertex which is a common neighbor to both the vertices. In other words, the edges are added so as to complete the triangles, thereby increasing the amount of transitivity in the network.

But the models which aim at explaining the evolution of the networks with highly skewed degree distributions are the best studied class of the network growth models. We first describe the archetypal model of Price [1] based on the previous work by Simon [2]. Then we consider one of the most influential growth models for complex networks, given by Barabási and Albert [3]. We also look at the generalization of Barabási and Albert model [16] where concepts of continuum theory, sublinear and superlinear attachment are considered, Krapivisky's model [6] and the Dorogovtsev and Mendes model [8,9]. Then we describe some of the vertex copying models and the Kleinberg's model [12,13].

2 Price's Model

In 1965, Derek de Solla Price described the first instance of a scale-free network: the citation network. He found out that both the in-degrees and out-degrees of the network of citations between scientific papers, for example as in Fig.1, followed a power-law distribution[7]. Some years later, Price published his explanation for the arising power-law degree distributions[1]. He built up his work on the ideas of Herbert Simon[2,5], developed in 1950s, which showed that power law arises when "the rich get richer", when the amount you get goes up with what you already have. Price coined the term *cumulative advantage* to denote this phenomenon. Today this phenomenon is known as *preferential attachment*, coined by Barabási and Albert[3].

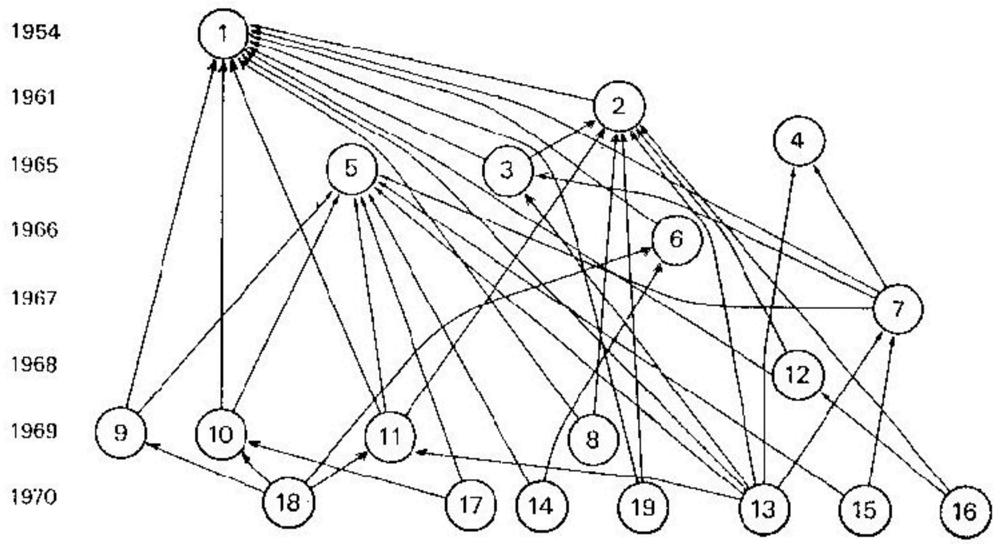

Fig. 1. Citation Network of Articles about "Peturbation of Ion Transport," 1954 to 1970

Price was the first to discuss cumulative advantage specifically in the context of networks and in particular the citation network between papers and its in-degree distribution. He reasoned that the number of citations a paper receives is proportional to the number of citations it already has. This can be understood intuitively as follows. The probability that one comes across a paper will increase with the number of citations it has received from other papers, and hence higher will be the probability that it again gets cited. The same argument follows for a lot of other networks, like the World Wide Web, airline networks, collaboration networks, etc.

Consider a directed graph G of n vertices. Let p_k be the number of vertices with in-degree k. For a citation network, this can be taken as the number of papers that cite a given paper. New vertices are added to the network not necessarily at a constant rate. The out-degree of the nodes in the network may vary but the average out-degree of all the nodes denoted by m is constant over time. Since the out-degree can vary over vertices, m can take non-integer values, including values less than 1. The value m is also the mean in-degree of the network : $\Sigma k p_k = m$.

In the cumulative advantage process, the probability that the newly appearing edge connects to an old vertex, i.e. a new paper cites a previous paper, is directly proportional to the in-degree k of the old vertex. But all the vertices initially start with zero in-degree and hence the probability of adding new edges to this node becomes zero. To overcome this problem, Price suggested that the probability of attachment is proportional to $k + k_0$, where k_0 is a constant. All his mathematical developments are for $k_0 = 1$, which he justifies in the case of citation networks by suggesting that a new paper can be considered as a citation to itself. So the probability of getting a new citation for a paper with k earlier citations become $k + 1$.

The probability that a new edge gets connected with any vertex with degree k is given by

$$\frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1} \tag{1}$$

since $\Sigma k p_k = m$ and $\Sigma p_k = 1$.

The mean number of new citations adder per vertex is m (as m is the mean in-degree and out-degree), therefore the number of new citations to vertices with in-degree k is $\frac{(k+1)p_k m}{m+1}$. We assume that the addition of a vertex to the network with mean out-degree m increases the out-degree of existing vertices by at most one. Thus, the number of vertices with in-degree k reduces by an amount equal to $\frac{(k+1)p_k m}{m+1}$ as the vertices which get new citations become vertices of degree $k + 1$. But the number of vertices with in-degree k increases as well due to the vertices previously of degree $k - 1$ and receives new citations. If we denote by $p_{k,n}$, the value of p_k when the graph G has n vertices, then the net change in np_k, i.e the number of nodes with in-degree k per vertex added is

$$(n+1)p_{k,n+1} - np_{k,n} = \begin{cases} [kp_{k-1,n} - (k+1)p_{k,n}]\dfrac{m}{m+1} & \text{for } k >= 1 \\ 1 - p_{0,n}\dfrac{m}{m+1} & \text{for } k = 0 \end{cases} \tag{2}$$

This equation is called as the *Master Equation*. Looking for stationary solutions $p_{k,n+1} = p_{k,n} = p_k$, we find

$$p_k = \begin{cases} [kp_{k-1} - (k+1)p_k]m/(m+1) & \text{for } k \geq 1 \\ 1 - p_0 m/(m+1) & \text{for } k = 0 \end{cases} \tag{3}$$

After rearranging the terms, we find $p_0 = (m+1)/(2m+1)$ and $p_k = p_{k-1}k/(k+2+1/m)$ or

$$p_k = \frac{k(k-1)\ldots 1}{(k+2+1/m)\ldots(3+1/m)} p_0 = (1+1/m)B(k+1, 2+1/m), \tag{4}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is Legendre's beta function, which goes asymptotically as a^{-b} for large a and fixed b, and hence

$$p_k \sim k^{-(2+1/m)} \tag{5}$$

For large values of n, the degree distribution has a power-law distribution with exponent $\alpha = 2 + 1/m$. This typically gives exponents in the range between 2 and 3 which is in agreement with the values seen in real-world networks.

Price's assumption that $k_0 = 1$ can be justified because the value of the exponent does not depend on k_0. The argument about can be easily generalized to the case when $k_0 \neq 1$, we get

$$p_k = \frac{m+1}{m(k_0+1)+1}\frac{B(k+k_0, 2+1/m)}{B(k_0, 2+1/m)}, \tag{6}$$

and hence $\alpha = 2 + 1/m$ again for large k and fixed k_0.

3 Barabási and Albert Model

Barabási and Albert[3] gave cumulative advantage a new name of preferential attachment. They proposed a network growth model of the Web which to very similar to that of Price's but with one important difference.

In the Barabási and Albert model, the vertices are added to the network with degree m, the other end of each edge being attached to another vertex with probability proportional to the degree of the vertex. But in this model the edges are undirected, so there is no distinction between the in-degree and out-degree of the vertex. On one hand, both Web and citation networks are essentially a directed network, so any undirected graph modelling them is missing a crucial feature of these networks. But on the other hand, by ignoring the direction of the edges the model overcomes the problem of Price's problem of how a paper gets its first citation or a Web site gets its first link. Every new vertex appears with initial degree m which is also the mean degree of the network, and hence automatically has a non-zero probability of receiving new links. Here m must be an integer and must always have a value $m \geq 1$. Overall, the best way to look at the model is that it sacrifices some of the realism of Price's model in favor of simplicity.

The Barabási and Albert model can be solved using the Master Equation method in the limit of a large graph size. The probability that a new edge attaches to a vertex of degree k is

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}. \tag{7}$$

The sum in the denominator is equal to the mean degree of the network, which is $2m$, since for every vertex added there are m edges, and each edge being undirected contributes two ends to the degree of the network vertices. Now the mean number of vertices of degree k that gain an edge when a new vertex with m edges is added is $m \times kp_k/2m = kp_k/2$, independent of m. The number np_k of vertices with degree k thus decreases by this same amount, since they now become vertices of degree $k+1$. The number of vertices of degree k also increases due to attachment of edges to the vertices of degree $k-1$. 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

for $k > m$,

$$(n+1)p_{k,n+1} - np_{k,n} = \frac{1}{2}(k-1)p_{k-1,n} - \frac{1}{2}kp_{k,n}, \tag{8}$$

for $k = m$,

$$(n+1)p_{m,n+1} - np_{m,n} = 1 - \frac{1}{2}mp_{m,n} \tag{9}$$

Looking for stationary solutions $p_{k,n+1} = p_{k,n} = p_k$, we find

$$p_k = \begin{cases} \frac{1}{2}(k-1)p_{k-1} - \frac{1}{2}kp_k & \text{for } k > m \\ 1 - \frac{1}{2}mp_m & \text{for } k = m \end{cases} \tag{10}$$

Rearranging for p_k once again, we find $p_m = 2/(m+2)$ and $p_k = p_{k-1}(k-1)/(k+2)$, or

$$p_k = \frac{(k-1)(k-2)\ldots m}{(k+2)(k+1)\ldots(m+3)}p_m = \frac{2m(m+1)}{(k+2)(k+1)k}. \tag{11}$$

In the limit of large k this gives a power law degree distribution $p_k \sim k^{-3}$, with only the single fixed exponent $\alpha = 3$.

Krapivsky and Redner [6] have conducted a thorough analytic study of the model, showing two important types of correlations among other things. First, they showed that there is a correlation between the age and degree of the vertices, with older vertices having higher mean degree. For the case $m = 1$, the distribution of the degree of a vertex i with age a, is

$$p_k(a) = \sqrt{1 - \frac{a}{n}}(1 - \sqrt{1 - \frac{a}{n}})^k. \tag{12}$$

Thus for specified age a the distribution is exponential, with a characteristic degree scale that diverges as $\sqrt{1 - \frac{a}{n}}$ as $a \to n$. The older vertices have substantially higher expected degree than the vertices added later, and the overall power-law degree distribution of the whole graph is a result primarily of the influence of these earliest vertices.

This correlation between the degree and age of the vertices has been used by Adamic and Huberman [7] to show that in actual World Wide Web network data, there is no such correlation present in the real Web. It seems that the dynamics of the Web must be more complicated than this simple model to account for the observed age distribution.

Second, Krapivsky and Redner [6] show that there are correlations between the degrees of adjacent vertices in the model. Looking at the special case of $m = 1$, they show that the quantity e_{jk}, the number of edges that connect vertex pairs with total degrees j and k, is

$$e_{jk} = \frac{4j}{(k+1)(k+2)(j+k+2)(j+k+3)(j+k+4)} \tag{13}$$

$$+\frac{12j}{(k+1)(j+k+1)(j+k+2)(j+k+3)(j+k+4)} \tag{14}$$

They regard the network as being directed, with edges leading from the vertex just added to the pre-existing vertex to which they attach.

There are many ways to simulate these types of network models. A naive implementation of preferential attachment algorithm is very inefficient. We need to look for degrees for each of the vertex when a vertex is added. The algorithm takes $O(n)$ time for each step and $O(n^2)$ time in total. An efficient procedure that works in $O(1)$ time per step and $O(n)$ time overall is as follows. We maintain a list that includes k_i entries for value i for each vertex i. For example, in a network we have 4 vertices labelled $1, 2, 3$ and 4 with degrees $1, 2, 1$ and 4 respectively could be represented by the array $(1, 2, 2, 3, 4, 4, 4, 4)$. To select the target node for a new edge for preferential attachment, we need to choose a random number from the list. Models such as Price's in which there is an offset k_0 in the probability of selecting a vertex can be treated with the same method, we just need to choose a vertex with preferential attachment and otherwise we choose uniformly from the set of all vertices.

An alternative method has been suggested by Krapivsky and Redner [6]. The model is regarded as a directed network with each node having m edges pointing towards other nodes. We first select a vertex at random and with some probability decide to either keep the vertex or move to one of its neighbor. Since every vertex has exactly m outgoing edges, the latter operation is equivalent to selecting an edge at random from the graph and following it. So the total probability of selecting any vertex is proportional to $j + c$, where c is some constant and j is the in-degree of the target vertex. Since the out-degree of all the vertices is m, the total degree is $k = j + m$, and the selection probability is proportional to $k - m + c$. By suitably choosing the redirection probability, we can adjust the constant c to be equal to m and hence the probability of selecting a vertex becomes proportional to k. Since this doesn't require an array for simulation, this is more memory efficient but it is slightly more complicated to implement.

4 Generalizations of Barabási and Albert Model

Barabási and Albert Model has been studied and worked upon extensively in the literature. Many authors have, in addition to performing the model's analytical and numerical studies, have proposed extensions and modifications of the model that alter its behavior or make it a more realistic representation of process taking place in real-world networks.

Dorogovtsev *et al.* [8] and Krapivsky and Redner [6] have examined the model in which the attachment probability of an edge with a vertex of degree k is proportional to $k + k_0$, where the offset k_0 is a constant $(-m < k_0 < \infty)$. In other words, the attachment trend is considered to be a superposition of preferential attachment and random attachment. Thus, the probability that a new edge attaches to a vertex of degree k is

$$\frac{(k + k_0)p_k}{\sum_k (k + k_0)p_k} = \frac{(k + k_0)p_k}{2m + k_0}. \tag{15}$$

where $\sum_k kp_k = 2m$ as in the naive Barabási-Albert model. Now, the mean number of vertices of degree k that gain an edge when a new vertex with degree m is added is given by $m \times (k+k_0)p_k/(2m+k_0)$, which is no longer independent of m. Accordingly, the Master equation is re-written as follows:
for $k > m$,

$$(n+1)p_{k,n+1} - np_{k,n} = \frac{m}{2m+k_0}[(k-1+k_0)p_{k-1,n} - (k+k_0)p_{k,n}], \tag{16}$$

Assuming that k is large, the terms proportional to k_0 can be ignored. Hence, the above equation can be approximated to the following:
for $k > m$,

$$(n+1)p_{k,n+1} - np_{k,n} = \frac{m}{2m+k_0}[(k-1)p_{k-1,n} - kp_{k,n}], \tag{17}$$

for $k = m$,

$$(n+1)p_{m,n+1} - np_{m,n} = 1 - \frac{m^2}{2m+k_0}p_{m,n} \tag{18}$$

Solving these equations for the steady-state, i.e. $p_{k,n+1} = p_{k,n} = p_k$, we get

$$p_k = \begin{cases} [(k-1)p_{k-1} - (k)p_k]m/(2m+k_0) & \text{for } k > m, \\ 1 - p_m m^2/(2m+k_0) & \text{for } k = m, \end{cases} \tag{19}$$

After rearranging the terms, we find that $p_m = \dfrac{2m+k_0}{m^2+2m+k_0} = \dfrac{2+k_0/m}{m+2+k_0/m}$ and $p_k = \dfrac{(k-1)p_{k-1}}{k+2+k_0/m}$. Thus, on expansion,

$$p_k = \frac{(k-1)\dots m}{(k+2+k_0/m)\dots(m+3+k_0/m)}p_m = \frac{B(k, 3+k_0/m)}{B(m, 2+k_0/m)}, \tag{20}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is again the Legendre's beta function. This gives a power law for large k with exponent $\alpha = 3 + k_0/m$. The negative values of k_0 could be the explanation for the values $\alpha < 3$ seen in real-world networks.

Krapivsky et al. [6] also consider the generalization of the model where the attachment probability of an edge to a vertex is not linearly proportional to its degree k, instead as some general power k^γ. Three general classes of behavior are found out. For $\gamma = 1$, we get the normal linear preferential attachment and power-law degree sequences. For $\gamma < 1$, the degree distribution is product of power law and a stretched exponential, whose exponent is a complicated function of γ. For $\gamma > 1$, there is a *condensation* phenomenon, where a particular vertex gets a finite fraction of all the network connections, and for $\gamma > 2$, there is a non-zero probability that this *gel node* is connected to all the vertices on the graph. The other vertices have exponentially decaying degree distribution. The behavior of the network is shown analytically in Section 4.4.

Another variation over this network growth model is to vary the mean degree over time. There is evidence available to suggest that in the World Wide Web, the average degree of a vertex is increasing with time. Dorogovtsev and Mendes [8,9] have studied a variation of the model that has a varying m parameter. They assume m increases with network size n as $m = n^a$ where a is a constant, and that the attachment probability is proportional to $k + Bn^a$ for constant B. They have shown the resulting degree distribution follows a power law with exponent $\alpha = 2 + B(1 + a)/(1 - Ba)$.

4.1 Dorogovtsev and Mendes Model

A wide class of developing and decaying networks have scaling properties similar to what Barabási and Albert proposed. The networks evolve in Dorgovtsev and Mendes model[9] according to two rules: i) The probability of connection of a new site to an old site is proportional to its connectivity. ii) New links appear between old sites with probability proportional to the product of their respective connectivities and some links between the old sites are removed with the same probability.

The BA model describes only a particular type of evolving networks but in reality, there are a lot more events taking place. In real networks like Internet links are not only added, but they may break from time to time. Some new links might also appear between the old sites. They proposed a model of developing and decaying networks with undirected links showing scaling behavior. They considered the following network evolution reasons. First they grow like in the BA model, at every instant a new site comes up and is connected with an old site with a probability proportional to its connectivity k. In addition they introduced a new parallel component of the evolution - the permanent evolution of new undirected links between the old sites, and the permanent removal of some old links.

They consider two cases : a) A developing network: At every instant, new $c(c \geq 0)$ links are added between the old sites i and j with probability that is proportional to the product of their connectivities $k_i k_j$. b) A decaying structure: Each instant, some links between the old sites also gets removed with equal probability $(c \leq 0)$.

The following one-site quantities of the structures was studied: the total distribution of connectivities at long times, $P(k)$, and the average connectivity of a site of an age s at long time t, $\bar{k}(s,t)$ and their scaling exponents $P(k) \propto k^{-\gamma}$ and $\bar{k}(s,t) \propto (s/t)^{-\beta}$.

They found out that both $P(k)$ and $\bar{k}(s,t)$ follow power law distributions for the developing networks for all $c \geq 0$. On the other hand for decaying networks, only $\bar{k}(s,t)$ demonstrates the power law behavior in the whole range $-1 < c < 0$. The power law dependence for $P(k)$ is observed only close to $c = 0$.

For developing networks, they found

$$\beta = \frac{1 + 2c}{2(1 + c)} \tag{21}$$

$$\gamma = 2 + \frac{1}{1 + 2c} \tag{22}$$

4.2 Adamic and Huberman Model

According to Barabási and Albert a vertex that acquires more connections than another will increase its connectivity at a higher rate; thus an initial difference in the connectivity between the two vertices will increase further as the network grows. BA model predicts that older vertices increase their connectivity at the expense of younger ones, a 'rich-get-richer' phenomenon.

Adamic and Huberman[7] studied a crawl of $260,000$ sites, each representing a separate domain name. They counted the number of links the sites received from the other sites and found the distribution of links to follow a power law. They then queried the InterNIC database for the date on which the sites were originally registered. Whereas the BA model suggests that since older vertices have a lot more time to acquire and gather links at a much faster rate than the newer sites, the results of their search showed that there is no correlation between the age of a site and its number of links.

The absence of correlation between age and a node degree is not that surprising. Every site is not created equal. They suggest that this is because vertex degree is in addition also a function of their intrinsic worth, e.g. some web sites caters to need of more people than others and so they gain links at a much higher rate.

They proposed a model in which the number of new links a site receives at each time step is a random fraction of the number of links the site already has. New sites, each with a different growth rate, appear at an exponential rate. This model can produce any power-law exponent $\gamma > 1$.

4.3 Continuum Theory and the Extended Barabási Model

The rate equations constructed so far are discrete, i.e. they assume that in a single time unit, only one vertex is added to the entire network. However, this assumption might not hold true for large networks such as the World Wide Web, where several nodes may be added to the network within the same time frame. Hence, an alternative question that can be asked is : within a time-step Δt, by how much does the degree of each node increase ?

Let $N(t)$ be the number of nodes in the network at time t. Then, the master equation can be expressed as the rate of change of degree of a node i.as follows: $\frac{\Delta k_i}{\Delta t} = m \times \frac{k_i}{\Sigma k_i}$, where a new node arrives in time Δt and initial degree m. Since, m is also the average degree of the network, we can write Σk_i as $2e(t) = 2N(t)m$, where $e(t)$ is the total number of edges at time t. The master equation becomes continuous as $\Delta t \to 0$:

$$\frac{dk_i}{dt} = m \times \frac{k_i}{2N(t)m} = \frac{k_i}{2N(t)} \tag{23}$$

We take $N(t) = t + n_0$, where n_0 is the initial number of nodes. Inserting this relation in the above differential equation and solving, we get $k_i(t) = C\sqrt{t + m_0}$.

If node i was introduced in the network at time t_i, then $k_i(t_i) = m$. Thus, $k_i(t) = m\sqrt{\frac{t+n_0}{t_i+n_0}}$, which can be approximated for large values of t as $k_i(t) = m\sqrt{\frac{t}{t_i}}$.

In order to obtain the degree distribution of the network, we first evaluate the expression $P(k_i < k)$. From the degree distribution plot, we can see that the value of this expression is equal to the area under the degree distribution curve between $k_i = 0$ and $k_i = k$. Thus,

$$P(k_i < k) = \int_0^k p_{k_i} dk_i \Rightarrow p_{k_i} = \frac{d(P(k_i < k))}{dk_i} \tag{24}$$

Now, $P(k_i < k) = 1 - P(k_i > k) = 1 - P(t > m^2 t/k^2) = 1 - m^2/k^2$. Differentiating this expression with respecting to k, we get $p_k = 2m^2k^{-3}$, which is structurally similar to the result obtained by using a discrete master equation for the naive Barabási model. In order to generalize this result, we can proceed by using the equation $\frac{dk_i}{dt} = m \times \frac{k_i+\delta}{\Sigma(k_i+\delta)}$, where δ is an offset.

The above analysis is based on the assumption that in a large-scale complex network, the only possible way of addition of new edges is *network growth*, i.e. addition of new nodes with initial degree m. However, this does not hold true for real-life networks. The extended Barabási model [16] suggests that other events can also occur in a network, such as creation of new links between existing nodes (*edge addition*) and replacement of ties among exisiting nodes (*rewiring*). Each of these events can occur continuously and in parallel, throughout the lifetime of the network.

Suppose m new edges (links) are added to the network by a combination of the above events. The master equation can again be expressed at the rate of change of degree of a node i as follows:

1. Addition of m edges with probability p:

$$(\frac{dk_i}{dt})_1 = pA\frac{1}{N(t)} + pA\frac{k_i+1}{\Sigma_i k_i + 1} \tag{25}$$

   where $N(t)$ is the number of nodes in the network at time t. The first term on the RHS denotes the addition of A edge-heads randomly and the second term denotes the selection of the corresponding tails preferentially. Since the net change in connectivity of the network is $\Delta k = 2m$, we have $A = m$.

2. Rewiring of m edges with probability q:

$$(\frac{dk_i}{dt})_2 = -qB\frac{1}{N(t)} + qB\frac{k_i+1}{\Sigma_i k_i + 1} \tag{26}$$

   where $N(t)$ is the number of nodes in the network at time t. The first term on the RHS denotes the deletion of B edge-tails randomly and the second term denotes the reconnection of the corresponding tails preferentially. Although the net connectivity of the network does not change, we have

$B = m$ obtained by splitting the rewiring event into random deletion and preferential attachment.

3. Addition of a new node with initial degree m:

$$(\frac{dk_i}{dt})_3 = (1-p-q)C\frac{k_i+1}{\Sigma_i k_i + 1} \tag{27}$$

where $(1-p-q)$ is the probability of a new node being added to the network. The RHS term is similar to the one obtained using the naive assumption. Since, the number of new edges introduced is equal to m, $C = m$.

Thus, the overall master equation is obtained by summing the above three equations.

$$\frac{dk_i}{dt} = (p-q)\frac{m}{N(t)} + m\frac{k_i+1}{\Sigma_i k_i + 1} \tag{28}$$

where $N(t) = (1-p-q)t + n_0$ and $\Sigma_i k_i = 2e(t) = (1-q)2mt - m$. The constant terms can be ignored for large values of t and hence, the above equation becomes

$$\frac{dk_i}{dt} + k_i\frac{-1}{(1-q)2t} = \frac{m(p-q)}{(1-p-q)t} + \frac{1}{(1-q)2t} \tag{29}$$

which is a first-order linear ordinary differential equation solvable by standard methods. The solution to this equation is given by

$$k_i(t) = (A(p,q,m) + m + 1)(\frac{t}{t_i})^{\frac{1}{B(p,q,m)}} - A(p,q,m) - 1 \tag{30}$$

where

$$A(p,q,m) = (p-q)(\frac{2m(1-q)}{1-p-q} + 1),$$
$$B(p,q,m) = \frac{2m(1-q) + 1 - p - q}{m}$$

Further, the probability that a node has a connectivity $k_i(t)$ smaller than k, i.e. $P(k_i(t) < k)$ can be written as $P(k_i(t) < k) = P(t_i > C(p,q,m)t)$, where

$$C(p,q,m) = (\frac{m + A(p,q,m) + 1}{k + A(p,q,m) + 1})^{B(p,q,m)} \tag{31}$$

. For $P(k_i(t) < k)$ to be calculated, the following conditions must be satisfied:

1. Since t_i/t must lie between 0 and 1, $C(p,q,m)$ must also lie between 0 and 1. Thus, the condition $m < k$ must be satisfied.

2. Also, $C(p,q,m)$ must be real which implies that the condition $A(p,q,m)+m+1>0$ must also be satisfied.

Thus, we have

$$P(k_i(t)<k) = 1 - C(p,q,m)\frac{t}{n_0+t} \tag{32}$$

from which, using $p_k = \dfrac{dP(k_i(t)<k)}{dk}$, we obtain

$$p_k = \frac{t}{t+n_0}D(p,q,m)(k+A(p,q,m)+1)^{-1-B(p,q,m)} \tag{33}$$

where $D(p,q,m) = (m+A(p,q,m)+1)^{B(p,q,m)}B(p,q,m)$.

Thus, the degree distribution of a scale-free network, as predicted by continuum theory takes the general form

$$p_k \propto (k+\kappa(p,q,m))^{-\gamma(p,q,m)} \tag{34}$$

where $\kappa(p,q,m) = A(p,q,m)+1$ and $\gamma(p,q,m) = B(p,q,m)+1$.

4.4 Sublinear and Superlinear attachment

Krapivsky et al.[6] suggested that the generalised rate equation of a complex network be written as follows:

$$\frac{dN_k}{dt} = A^{-1}[A_{k-1}N_{k-1} - A_kN_k] + \delta_{kl} \tag{35}$$

where A is attachment normalization constant given by $\Sigma_j(A_jN_j)$,
A_k is the probability that a new node attaches to an existing node of degree k,
N_k is the total number of nodes in the network with degree k,

$$\delta_{kl} = \begin{cases} 1 \text{ if k = 1,} \\ 0 \text{ otherwise} \end{cases} \tag{36}$$

Here, we assume that $m=1$. Also, $N_0=0$ is set so that the generalised rate equation applies to all $k \geq 1$.

Given this node distribution, we proceed to compute the low-order moments of the distribution, which are defined as $M_n(t) = \Sigma_{j\geq1}j^nN_j(t)$. Correspondingly, $\dot{M}_n(t) = \Sigma_{j\geq1}j^n\dfrac{dN_j}{dt}$. For $n=0$, the expression reduces to $\Sigma_{j\geq1}\dfrac{dN_j}{dt} = 1 - \dfrac{A_1N_1}{A} + \Sigma_{j>1}[\dfrac{A_{j-1}N_{j-1}}{A} - \dfrac{A_jN_j}{A}] = 1$, as all the terms in the expanded summation cancel out. Thus, $\dot{M}_0(t)=1$, implying that $M_0(t) = M_0(0)+t$. For the first moment, which is equal to the total number of link endpoints in the network, the expression for $\dot{M}_1(t)$ becomes

$\Sigma_{j\geq 1} j \dfrac{dN_j}{dt} = 1 - \dfrac{A_1 N_1}{A} + \Sigma_{j>1} j [\dfrac{A_{j-1}N_{j-1}}{A} - \dfrac{A_j N_j}{A}] = 1 + 1 = 2$. Correspondingly, $M_1(t) = M_1(0) + 2t$ and thus, only the first two moments are independent of the attachment probabilities A_k.

For a typical linear attachment model, we have $A_k = k$, thus $M_1(t) = 2t$ coincides with $A(t)$. Solving the rate equation in this case for an arbitrary initial condition gives $N_1(t) = 2t/3$. Also, the frequency of all other degree nodes varies linearly with t, so that $N_k(t) = tn_k$. Solving the rate equation therefore yields $n_k = n_{k-1}\dfrac{k-1}{k+2} = \dfrac{2}{3}B(k,3)$ which for large k becomes $\propto Ck^{-3}$. This result is the same as the degree distribution obtained by a naive Barabási model.

Consider a generalised attachment model where both the degree distribution and the attachment normalization factor $A(t)$ grow linearly in time. We also assume that this hypothesis holds true for sublinear attachment, i.e. the attachment probability A_k does not grow faster than linearly with k. Then the following relations hold:

$$\begin{aligned} A(t) &= \mu t; \\ N_k(t) &= tn_k; \\ A_k &\propto k^{\gamma}, 0 < \gamma < 1 \end{aligned}$$

Using the above mentioned results into the generalised rate equation, we get

$$\begin{aligned} n_k = \frac{1}{\mu}(A_{k-1}n_{k-1} - A_k n_k), \\ \Rightarrow n_k = n_{k-1}\frac{A_{k-1}}{\mu + A_k} \text{ where } k > 1, \\ n_1 = \frac{\mu}{\mu + A_k} \end{aligned}$$

Expanding the above recursive relation we get

$$n_k = \frac{\mu}{A_k}\prod_{j=1}^{k}(1 + \frac{\mu}{A_j})^{-1} \tag{37}$$

where the amplitude μ is still unknown. However, we know that $\dfrac{A(t)}{t} = \mu = \Sigma_k (A_k n_k)$. Therefore,

$$\Sigma_{k=1}^{\infty}\prod_{j=1}^{k}(1 + \frac{\mu}{A_j})^{-1} = 1 \tag{38}$$

Thus, the amplitude μ always depends on the entire set of attachment probabilities. However, the behaviour of the degree distribution varies largely, depending on whether the attachment is asymptotically linear, sublinear or superlinear.

For sublinear attachment, the degree distribution is robust and depends only gross features of the network. Putting $A_k \propto k^{\gamma}$ into equation (37), writing the extended product as the exponential of a sum, converting the sum into an integral and performing the integral, we obtain

$$n_k \propto \begin{cases} k^{-\gamma} exp[-\mu(\frac{k^{1-\gamma}-2^{1-\gamma}}{1-\gamma})] & 1 > \gamma > \frac{1}{2}, \\ k^{\mu^2-1/2} exp[-2\mu\sqrt{k}] & \gamma = \frac{1}{2}, \\ k^{-\gamma} exp[-\mu \frac{k^{1-\gamma}}{1-\gamma} + \frac{\mu^2}{2}\frac{k^{1-2\gamma}}{1-2\gamma}] & \frac{1}{2} > \gamma > \frac{1}{3}, \end{cases} \tag{39}$$

etc. The pattern observed in the above equation continues *ad infinitum*: Whenever γ decreases below $1/m$ with m a positive integer, an additional term in the exponential arises from the now relevant contribution of the next-higher order term in the expansion of the product in equation (37) [6]. Also, it has been found that in this case, the amplitude μ does not have a closed-form expression, but varies smoothly between 1 and 2 as γ varies between 0 and 1. Also, for the limiting cases, the degree distributions are observed to be $n_k = 2^{-k}$ and $n_k = \frac{2}{3}B(k,3)$ for γ equal to 0 and 1, respectively.

For asymptotically linear attachment, $A_k \sim k$ as $k \to \infty$. Inserting this expression in equation () and proceeding as in the case of sublinear attachment, we get an asymptotic power law

$$n_k \sim k^{-\nu}, \nu = 1 + \mu \tag{40}$$

It must be noted that the exponent ν can be tuned to any value greater than 2. Also, the amplitude μ can be computed as a function of $A_1 = \alpha$. (Detailed proof of this is provided by Krapivsky et al in [6]) Inserting this value of ν in terms of α, we get the degree distribution as follows:

$$n_1 = \frac{\mu}{\mu + \alpha}, n_k = C.B(k, 2 + \mu) \tag{41}$$

where $k \geq 2$ and $C = \frac{(\mu\alpha)(2+\mu+k)}{\mu+\alpha}$.

For superlinear attachment, $A_k = k^{\gamma}$, with $\gamma > 1$. For such a network, a "winner takes all" phenomenon arises, i.e. there emerges a "gel" node which is linked to almost every other node. A particularly singular behaviour occurs for $\gamma > 2$, where there is a non-zero probability that this gel node is connected to every other node in the network.

Let us proceed by determining the probability that this gel node connects to all other nodes. We consider a discrete growth model where one node is introduced at each time step and always links to the initial node. If the initial degree of every node in the network is equal to 1, then the attachment probability is given by

$$A_{gel} = \frac{k_{gel}^{\gamma}}{\Sigma_{j=1}^{N} k_j^{\gamma}} = \frac{N^{\gamma}}{N + N^{\gamma}} = \frac{1}{1 + N^{1-\gamma}} \tag{42}$$

The probability that this connectivity pattern continues indefinitely is equal to

$$P = \prod_{N=1}^{\infty} \frac{1}{1 + N^{1-\gamma}} \tag{43}$$

which is clearly equal to 0 for $\gamma \leq 2$ and non-zero otherwise.

In order to determine the degree distribution for $\gamma > 1$, we first need to find out the asymptotic time dependence of M_γ. Using the rate equation in (35), we obtain

$$n_k(k) = (k-1)^\gamma \frac{n_{k-1}(k-1)}{M_\gamma k - 1} = n_2(2) \prod_{j=2}^{k-1} \frac{j^\gamma}{M_\gamma(j)} \tag{44}$$

Further, Krapivsky et al. [6] rigorously prove that $M_\gamma(t) \sim t^\gamma$ and use this result to recursively obtain all N_k as

$$N_k(t) = J_k t^{k-(k-1)\gamma}, k \geq 1 \tag{45}$$

with $J_k = \prod_{j=1}^{k-1} \dfrac{j^\gamma}{1 + j(1-\gamma)}$ and $k < \dfrac{\gamma}{\gamma - 1}$. Thus, superlinear complex networks undergo an infinite sequence of connectivity transitions as a function of γ. For $\gamma > 2$, all but a finite number of nodes are linked to the "gel" node, which has the rest of the links in the network. This is the "winner takes all" situation. For $\dfrac{3}{2} < gamma < 2$, the number of nodes with two links grows as $t^{2-\gamma}$, while the number of nodes with more than two links is again finite. In general, for $(m+1)/m < \gamma < m/(m-1)$, the number of links with more than m links is finite, while $N_k \sim t^{k-(k-1)\gamma}$ for $k \leq m$.

4.5 Bianconi and Barabási Model

The BA model neglects an important aspect of the competitive systems that not all the nodes are equally successful in acquiring the links. The model predicts that all nodes increase their connectivity in time as $k_i(t) = (t/t_i)^\beta$, where $\beta = 1/2$ and t_i is the time of addition of node i into the system. Therefore, the oldest node is expected to have the highest number of links pertaining to the longest timeframe they have to acquire them.

There are many examples that indicate that in real networks, a node's connectivity and growth rate does not depend on its age alone. For example, some individuals acquire more social links than others. On the www, some webpages through a combination of good content and marketing acquire a large number of links in a very short span of time. Some ground-breaking research papers in a short timeframe acquire a very large number of citations. Bianconi and Barabási[14,15] attributed these differences to some intrinsic quality of the nodes, such as the social skills of an individual, content of a web page, or the content of a scientific article. They gave this the term *node's fitness*, describing its ability to compete for links at the expense of other nodes.

The Fitness Model : They assign a fitness parameter η_i which is time invariant. At each timestamp, a new node i is added with fitness η_i, where η is chosen from the distribution $\rho(n)$. m new links are added to the system with the addition of each node. The probability that the newly added node gets connected to an old node i, Π_i depends on the connectivity k_i and its fitness η_i.

$$\Pi_i = \frac{\eta_i k_i}{\Sigma_j \eta_j k_j} \tag{46}$$

A node i increases its connectivity k_i at a rate that is proportional to the probability that a new node will attach to it, giving

$$\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{\Sigma_j k_j \eta_j} \tag{47}$$

The factor m accounts for the fact that we add m links to the system with every addition of a new node. Due to multiscaling in the system, the dynamic exponent depends on the fitness η_i,

$$k_{\eta_i}(t, t_0) = m \left(\frac{t}{t_0} \right)^{\beta(\eta_i)} \tag{48}$$

where t_0 is the time at which the node i was born. The dynamic exponent $\beta(\eta)$ is bounded, i.e. $0 < \beta(\eta) < 1$ because a node always increases the number of links in time and $k_i(t)$ cannot increase faster than t. They found

$$\beta(\eta) = \frac{\eta}{C} \tag{49}$$

where $\eta_m ax < C < 2\eta_m ax$.

The degree distribution $P(k)$ can be calculated which gives the probability that a node has k links. If there is a single dynamic exponent β, the distribution follows the power law $P(k) \propto k^\gamma$, where the connectivity exponent is given by $\gamma = 1/\beta + 1$.

In this model, we have a spectrum of dynamic exponents $\beta(n)$, thus $P(k)$ is given by a weighted sum over different power laws. The connectivity distribution is given by

$$P(k) \propto \int d\eta \rho(\eta) \frac{C}{\eta} \left(\frac{m}{k} \right)^{\frac{C}{\eta}+1} \tag{50}$$

Scale-free model : We have $\rho(\eta) = \delta(\eta - 1)$, which gives $C = 2$ and $\beta = 1/2$. We get $P(k) \propto k^{-3}$, the known scaling of the scale-free model.

Uniform fitness distribution: The fitness distribution $\rho(\eta)$ is chosen uniformly from the interval $[0, 1]$. The constant C is calculated from

$$e^{(-2/C)} = 1 - 1/C \tag{51}$$

whose solution is $C^* = 1.255$. Accordingly each node will have a different dynamic exponent, given by $\beta(\eta) \propto \frac{\eta}{C^*}$. We obtain

$$P(k) \propto \frac{k^{-(1+C^*)}}{log(k)} \tag{52}$$

i.e. the connectivity distribution follows a generalized power law, with an inverse logarithmic correction.

Exponential fitness distribution : If the $\rho(\eta)$ distribution has an infinite support, the integral has a singularity at $\eta = C$. For such systems, they studied numerically the case $\rho(\eta) = e^{-\eta}$. The numerical simulations demonstrated that $P(k)$ follows a stretched exponential.

5 Criticism of Barabási-Albert Model and Price's Model

The Barabási-Albert model lacks a number of features present in the real World Wide Web network :

- It assumes an undirected network whereas the real Web is a directed network.
- One can regard the model as a directed network with k as the sum of in-degree and out-degree of a node. But in this case the attachment probability should be in proportion to only the in-degree value and not k.
- If we regard the model as producing a directed network, then it generates acyclic graphs which are a poor representation of web.
- All vertices belong to a single connected component, whereas in the real Web there are many separate components.
- The out-degree distribution of the Web follows a power law, whereas it is a constant in the model.

Most of these are true for Price's model as well. But Price's model is intended to model citation networks and they are in practice directed, acyclic and bears a good approximation to a network where all vertices belong to a single component, unless a paper cites and does not get cited by any one else. For the World Wide Web authors have suggested new growth models to address the above issues.

Let us consider the issue of the single component structure of the network. In both Price and Barabási-Albert model each appearing vertex joins with at least one other vertex. So if no edges are removed in the future, we get a single (weakly-connected) component. However this is not true for real Web network. To address this, Callaway *et al.* [10] proposed that the vertices are added to the network one by one as before, and a mean number m of undirected edges are added with each vertex. The value of m is only an average, the actual number of edges added per step can vary and So m is not restricted to integer values.

One important difference between this model and previous model is that edges may not necessarily attach with the newly added vertex. The edges are attached with two randomly chosen vertices without any preferential attachment. The model does not show power law degree distribution and in fact the degree distribution can be shown to follow exponential. The older vertices in the network tend to be connected to one another, so we have a network with a cliquish core of old nodes surrounded by a sea of younger nodes. The network has a finite value of m at which a giant component appears that occupies a fixed fraction of the volume of network as $n \to \infty$. To demonstrate this, Callaway used a similar master-equation approach. p_s is defined as the probability that a randomly chosen vertex lies in a component of size s. We can write the equations which denote the change in p_s when a single vertex and m edges are added to the network. Looking for stationary solutions, in the limit of large graph size

$$p_s = \begin{cases} ms\sum_{j=1}^{s-1} p_j p_{s-j} - 2msp_s & \text{for } s > 1 \\ 1 - 2mp_1 & \text{for } s = 1. \end{cases} \tag{53}$$

The equations being nonlinear in p_s are harder to solve and indeed no exact solution has been found. We can see that a giant component must form by defining a generating function for the component size distribution $H(x) = \sum_{s=0}^{\infty} p_s x^s$. We get

$$\frac{dH}{dx} = \frac{1}{2m}[\frac{1 - H(x)/x}{1 - H(x)}]. \tag{54}$$

If there is no giant component, then $H(1) = 1$ and the average component size is $\langle s \rangle = H'(1)$. Taking limit $x \to 1$, we find that $\langle s \rangle$ is a solution of the quadratic equation $2m\langle s \rangle^2 - \langle s \rangle + 1 = 0$, or

$$\langle s \rangle = \frac{1 - \sqrt{1 - 8m}}{4m}. \tag{55}$$

From the equation, we can see that the solution exists only up to $m = \frac{1}{8}$ and hence above this point there must be a giant component. A proof that the transition in fact falls precisely at $m = \frac{1}{8}$ was later given by Durrett [11].

6 Vertex Copying Models

There are a number of networks which seem to have a power-law degree distributions, but preferential attachment is clearly not a good explanation for them.

Good examples are biochemical interaction networks of various kinds. The interaction networks of proteins, where the vertices are proteins and the edges represent reactions. They change on very long time-scales because of biological evolution, but we can't assume that they grow obeying just preferential attachment model. But it appears that degree distribution of these networks too obey roughly a power law.

Kleingberg et al. [12,13] proposed a possible explanation for this observation that these type of networks grow, at least in part, by the copying of vertices. Their model is characterized by four stochastic processes- *creation* processes C_v and C_e for node and edge creation, and *deletion* processes D_v and D_e for node and edge deletion. Each process is discrete-time process and is a function of the time step and of the current graph.

A simple node-creation process is as follows. At each time step, create a node with probability $\alpha_c(t)$. We have a similar Bernoulli model with probability $\alpha_d(t)$ for node deletion; upon deleting a node, we also delete all its incident edges.

At each step, we sample a probability distribution to determine a node v to add edges out of, and k the number of edges to be added. With probability β, we add k edges from v to nodes which are chosen uniformly and independently at random. With probability $1 - \beta$ we copy k edges from a random vertex u, i.e. for every edge (u, w) in the network, we add the edge (v, w). If u has exactly k number of edges, we copy all of them. If u has greater than k edges, we take a subset of size k of them at random and copy. If u has less than k number of edges, we copy all of its edges and repeat the same process by selecting another vertex u' at random until we get k number of edges to copy.

The edge-deletion process D_e is Bernoulli process in each time step, we delete a randomly chosen node with probability δ. The probability that a particular node v gets deleted is non-increasing in its in-degree.

This copying mechanism gives rise to power-law degree distributions. The mean probability that an edge from a randomly chosen vertex will lead to a particular other vertex with in-degree k is proportional to k and hence the rate of increase of a vertexs degree is proportional to its current degree. This mechanism does not add new edges to vertices that currently have degree zero, so Kleinberg also include a finite probability that the target of a newly added edge is chosen at random, so that the zero-degree vertices also get a chance to gain edges. The model results in a degree distribution which obeys a power law with exponent $\alpha = (2 - a)/(1 - a)$, where a is the ratio of the number of edges added whose targets are chosen at random to the number of edges whose targets are copied from other vertices. For small values of a, between 0 and $\frac{1}{2}$, for models in which most target vertices are selected by vertex copying, this produces exponents $2 \leq \alpha \leq 3$ which is the range observed in most real world networks.

7 Conclusion

In this survey we have looked at various network models which try to model the growth processes taking place in real-world networks. We started off with the

Price Model and the Barabási Albert model which considers that the new nodes get attach to the network preferentially, i.e. nodes having a higher degree have a higher chance of getting new edges. Many real-world networks are found which do not directly follow this simple model. Then we looked at some generalizations of the BA model which try to model the anomalies found from the simple BA model. Dorgovtsev and Mendes also takes into consideration that edges appear and disappear between the old nodes while the network is growing. Adamic and Huberman's model gives an intrinsic worth to each node for attracting newer edges to get attached to it, instead of just depending on its degree. Bianconi and Barabási attach to each vertex a fitness value which represents its attractiveness to accrue more links in the future. Then we looked at the various criticisms that the BA model and Price model have had to face. In the end, we looked at the vertex copying models which try to model networks following power-law degree distributions but for which preferential attachment does not seem to be a good explanation.

8 References

[1] Price, D. J. de S., A general theory of bibliometric and other cumulative advantage processes, *J. Amer. Soc. In- form. Sci.* **27**, 292306 (1976).

[2] Simon, H. A., On a class of skew distribution functions, *Biometrika* **42**, 425440 (1955).

[3] Barabasi, A.-L. and Albert, R., Emergence of scaling in random networks, *Science* **286**, 509512 (1999).

[4] Ball, F., Mollison, D., and Scalia-Tomba, G., Epidemics with two levels of mixing, *Annals of Applied Probability* **7**, 4689 (1997).

[5] Bornholdt, S. and Ebel, H., World Wide Web scaling exponent from Simons 1955 model, *Phys. Rev. E* **64**, 035104 (2001).

[6] Krapivsky, P. L. and Redner, S., Organization of growing random networks, *Phys. Rev. E* **63**, 066123 (2001).

[7] Adamic, L. A. and Huberman, B. A., Power-law distribution of the world wide web, *Science* **287**, 2115 (2000).

[8] Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N., Structure of growing networks with preferential linking, *Phys. Rev. Lett.* **85**, 46334636 (2000).

[9] Dorogovtsev, S. N. and Mendes, J. F. F., Scaling behaviour of developing and decaying networks, *Europhys. Lett.* **52**, 3339 (2000).

[10] Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E. J., and Strogatz, S. H., Are randomly grown graphs really random?, *Phys. Rev. E* **64**, 041902 (2001).

[11] Durrett, R. T., Rigorous results for the Callaway HopcroftKleinbergNewmanStrogatz model, Preprint, Department of Mathematics, Cornell University (2003).

[12] Kleinberg, J. M., Kumar, S. R., Raghavan, P., Rajagopalan, S., and Tomkins, A., The Web as a graph: Measurements, models and methods, in *Proceedings of the International Conference on Combinatorics and Computing*, no. **1627** in Lecture Notes in Computer Science, pp. 118, Springer, Berlin (1999).

[13] Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A. S., and Upfal, E., Stochastic models for the Web graph, in *Proceedings of the 42st Annual IEEE Symposium on the Foundations of Computer Science*, pp. **5765**, Institute of Electrical and Electronics Engineers, New York (2000).

[14] Bianconi, G. and Barabasi, A.-L., BoseEinstein condensation in complex networks, *Phys. Rev. Lett.* **86**, 5632-5635 (2001).

[15] Bianconi, G. and Barabasi, A.-L., Competition and multiscaling in evolving networks, *Europhys. Lett.* **54**, 436-442 (2001).

[16] Albert, R. and Barabasi, A. -L., Topology of evolving networks : local events and universality, *Phys. Rev. Lett.* **85**, 5234-5237 (11th December 2000).