Chapter 9

GROWING NETWORKS

Understanding the properties of growing networks with popularity-based construction rules has recently (as of this writing) become an extremely active and fruitful research area. Part of the appeal of such models is that they provide a natural way to understand a broad range of complex systems, such as the topology of the Internet and the world-wide web, the structure and function of metabolic networks, the spread of a contagion through a susceptible population, and the statistics of scientific citations. A considerable amount of data about these systems has become available in electronic form only recently, the existence of which has helped fuel much of the current interest in model building. Part of the attraction of complex networks models is their simplicity that belies the term "complex" and the sense that such models capture certain fundamental aspects of nature that heretofore were only vaguely appreciated. In this chapter, we will present some of the classic models of complex networks and apply the machinery of the master equation to solve many of the structural features of such networks.

9.1 Erdös-Rényi Random Graph

An appealingly simple starting example is the *Erdös-Rényi (ER) random graph*. This graph consists of a set of N nodes in which each node pair is joined by a link with a fixed probability. A nice feature of the ER graph is that there is absolutely no spatial structure, as a connection between any pair nodes occurs equiprobably. If all of the $N(N-1)/2$ pairs of nodes is connected, we obtain the *complete graph* in which each node is connected to all $N-1$ other nodes. This graph represents a concrete realization of mean-field theory and is particularly useful in this context. It is conventional to define the connection probability between nodes as p/N, so that $p = N$ corresponds to the complete graph.

For general values of $0 < p < N$, the ER graph has many remarkable properties. Perhaps the most striking aspect of the ER graph is the existence of a phase transition at $p = 1$. For $p < 1$, the network consists of a collection of disconnected components with a maximum cluster size that is of the order of $\ln N$. At $p = 1$, there is a percolation transition that marks the first appearance of an infinite cluster that consists of $N^{2/3}$ nodes, which is therefore a tiny fraction of the nodes that belong to clusters of any size. For $p > 1$, an infinite cluster exists that now consists of a finite fraction of all nodes in the network.

While we have introduced the ER graph as a static problem—each link between any pair of nodes is independently present with probability p/N, we can recast the ER graph as a *dynamic* problem by allowing links to be created between nodes at a constant rate. Thus as time increases, more links are created and the network becomes progressively better connected. By formulating the ER graph as a dynamic problem, we can use the master equation and related tools to determine the structure of the ER graph.

Degree distribution

We build up the ER random graph by the following dynamic model. Starting with no links and N disconnected nodes, links are sequentially added *ad infinitum* between randomly selected pairs of nodes. For convenience in this discussion, we define the rate at which a link is introduced between pairs of nodes as $(2N)^{-1}$. The two nodes selected for linking may be the same, and additionally, more than one link may be

created between a pair of nodes. However, both of these processes occur with a vanishingly small probability when $N \to \infty$. In this infinite-size limit we can also ignore statistical fluctuations in the geometrical properties of the ER graph.

Let us first determine the degree distribution, namely the probability that a node has a given degree; here degree is the number of links that are attached to the node. At time t, the total number of links is on average $Nt/2$, and therefore the average number of links per node k, the degree, equals the time t. Thus the degree undergoes the additive stochastic process $k \to k+1$ with rate 1. Thus , the degree distribution, the probability n_k that the degree of a node equals k, satisfies

$$\frac{dn_k}{dt} = n_{k-1} - n_k, \tag{9.1}$$

with the initial condition $n_k(0) = \delta_{k,0}$. Therefore the degree distribution is the Poissonian

$$n_k = \frac{t^k}{k!}\, e^{-kt}, \tag{9.2}$$

from which the mean degree equals the time, $\langle k\rangle = t$, while the standard deviation $\sqrt{\langle k^2\rangle - \langle k\rangle^2} = \sqrt{t}$. Thus the distribution of degrees becomes sharp in the thermodynamic limit.

Component size distribution

We define a component as a set of nodes that are connected by links. As the number of links in the graph increases, disconnected components will merge when a link is created that joins a site of one component to a site in the other component. Since a link can occur equiprobably between any pair of nodes, there are $i \times j$ ways to join disconnected components of size i and j. Hence components undergo the aggregation process $(i, j) \to i + j$ with aggregation rate $ij/(2N)$. This process is precisely product kernel aggregation that we treated in chapter 3, and we can make use the results derived therein to determine the component size distribution in the ER random graph.

Let $c_k(t)$ be the density of components containing k nodes at time t. The component size distribution obeys the master equation

$$\frac{dc_k}{dt} = \frac{1}{2}\sum_{i+j=k} (ic_i)(jc_j) - k\, c_k, \tag{9.3}$$

with the initial condition $c_k(0) = \delta_{k,1}$. The gain term accounts for the merger between two components whose sizes sum to k and the loss term accounts for the loss of components of size k due to their linking with other components.

An insightful and simple way to understand the dynamics is through moments of the component size distribution, $M_n = \sum_k k^n c_k$. For example, the mean component size, or equivalently, the second moment obeys

$$\begin{aligned}
\dot{M}_2 = \sum_k k^2\, \dot{c}_k &= \frac{1}{2}\sum_i \sum_j \left[(i+j)^2 (ic_i)(jc_j) - k^3 c_k\right] \\
&= \frac{1}{2}\sum_i \sum_j \left[(i^3 c_i)(jc_j) + (ic_i)(j^3 c_j) + (i^2 c_i)(j^2 c_j) - k^2 c_k\right] \\
&= M_2^2,
\end{aligned}$$

as long as there is no gelation so that $M_1 = 1$, a condition that holds for $t < 1$. With the initial condition $M_2(0) = 1$, the solution for the second moment is simply

$$M_2 = (1-t)^{-1}, \tag{9.4}$$

for $t < 1$. The divergence shows that the system undergoes a percolation transition at a finite time $t_g = 1$. At t_g, the giant component forms. Beyond the percolation point, the giant component contains a finite fraction of the nodes, and eventually it engulfs the entire system.

From our discussion of aggregation with the product kernel in chapter 3, the component size distribution is

$$c_k(t) = \frac{k^{k-2}}{k!}\, t^{k-1}\, e^{-kt}. \tag{9.5}$$

The asymptotic behaviors of this size distribution for $t < 1$ and $t = 1$ are given by

$$c_k(t) \sim \begin{cases} e^{-k(t-\ln t-1)} & t < 1; \\ (2\pi)^{-1/2} k^{-5/2} & t = 1. \end{cases}$$

Thus below from the percolation point, this distribution decays exponentially with size, while at the percolation point the distribution has a power law decay. These asymptotic distribution can be used to infer the size of the largest cluster. We use the extremal criterion

$$N \sum_{k_{\max}}^{\infty} c_k = 1$$

that mandates that there is one cluster whose size is in the range $[k_{\max}, \infty]$, as the condition to estimate $k_{\max}$. Using the above asymptotic forms for the component size distributions and approximating the sum by an integral in the above extremal criterion, we obtain

$$k_{\max} \propto \begin{cases} \ln N & t < 1; \\ \\ N^{2/3} & t = 1. \end{cases}$$

9.2 Preferential Attachment Networks

In preferential attachment networks, nodes are added one by one and each attaches to pre-existing nodes of the network according to an attachment rate A_k that depends only on the degree of the "target" node (Fig. 9.1). The example that has received much attention is *preferential attachment* in which A_k is an increasing function of k. This rule encapsulates the intuitive notion that the "rich get richer". As a practical example in the context of scientific citations, preferential attachment means that if a paper is currently well cited, it is likely to be cited at a higher rate in the future merely by the virtue of being well cited. Similarly, in creating hyperlinks between websites, it is more likely that these links will go to popular websites. The special case of *linear* preferential attachment has generated the most interest because this rule leads to a *scale free* network in which the distribution of node degrees has a power-law tail. A power-law degree distribution stands in stark contrast to regular lattices, where this distribution is simply a delta function, and to the Poisson degree distribution of the Erdös-Rényi random graph. In addition to providing a natural mechanism for a power law degree distribution, linear preferential attachment seems to capture some of the truly compelling aspects of complex networks.

In this section, we apply the master equation to elucidate the structure of growing networks. We will attempt to convince the reader, that the master equations are a simple, incisive, and powerful theoretical tool to analyze growing network systems. In addition to providing comprehensive information about the node degree distribution, the master equation approach can be easily adapted to elucidate many other important structural features of complex networks.

The degree distribution

The degree distribution is perhaps the most prominent geometrical feature that distinguishes complex networks from simple networks. We write the degree distribution as $N_k(N)$, the number of nodes of degree k when network contains N total nodes. To determine the evolution of $N_k(N)$, we write the master equation that accounts for the change in the degree distribution after each node addition event. When nodes are added one by one, with each new node attaching to only one pre-existing node, the master equation for the degree distribution is

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A} + \delta_{k1}. \tag{9.6}$$

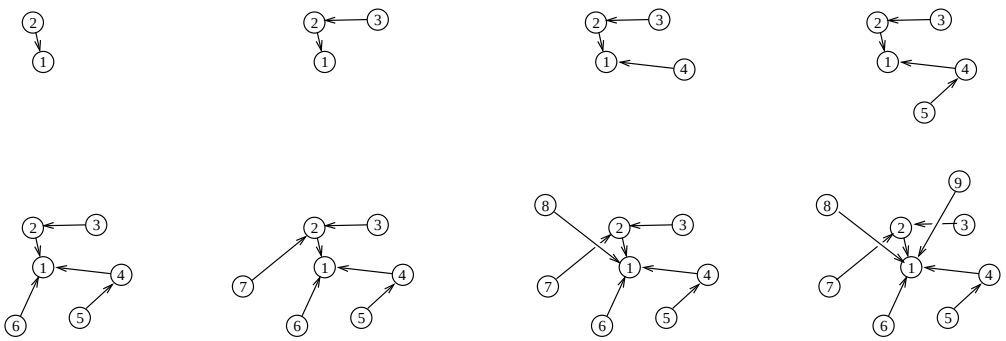

Figure 9.1: Evolution of a preferential attachment network for the case where each new node links to m other nodes with $m = 1$ (upper left to lower right).

The first term on the right accounts for processes in which the new node connects to a node that already has $k-1$ links, thereby increasing N_k by one. Since there are N_{k-1} nodes of degree $k-1$, the total rate at which such processes occur is proportional to $A_{k-1}N_{k-1}$. The factor $A(N) \equiv \sum_{j\geq 1} A_j N_j(N)$ is the total rate for any event to occur, so that $A_{k-1}N_{k-1}/A$ is the probability to attach to a node of degree $k-1$. A corresponding role is played by the second (loss) term on the right-hand side; namely, $A_k N_k/A$ is the probability that the new node connects to a node with k links, thus leading to a loss in N_k by one. The last term accounts for the new node itself that has one outgoing link and no incoming links. While we are keeping track of the total degree in the above master equation, it is straightforward to generalize and keep track both the in-degree (the number arrowheads pointing to a node) and the out-degree. For the model defined by Fig. 9.1, the out-degree of each node equals one, but networks with more than one outgoing link at each node can easily be studied.

To solve the degree distribution for a network with one outgoing link per node, we now need to specify the attachment rate A_k. We consider the generic preferential attachment form $A_k = k^\gamma$, with γ arbitrary. Note that the amplitude in A_k is irrelevant as in cancels out in the numerator and denominator of the master equation; consequently, we take this amplitude to be one. The case $\gamma = 1$ corresponds to linear preferential attachment, but there is much insight to be gained by studying arbitrary γ. Let's start by solving for the time dependence of the moments of the degree distribution $M_n(N) \equiv \sum_{j\geq 1} j^n N_j(N)$. When nodes are added one by one, N plays the role of a time-like variable and we will sometimes refer to N as the "time". By summing Eqs. (9.6) over all k, we immediately obtain $\dot{M}_0(N) = 1$, where the overdot denotes differentiation with respect to N, so that $M_0(N) = N$. We could also use the fact that $M_0 = \sum_k N_k$ is simply the total number of network nodes and thus equals N by construction. Similarly, the first moment obeys $\dot{M}_1(N) = 2$, with solution $M_1(N) = 2N$. Alternatively, the first moment equals the total number of link ends. Clearly, this number is $2N$, since introducing a new node adds two link ends to the network. Therefore the first two moments grow linearly with time, independent of the attachment rate A_k. On the other hand, higher moments and the degree distribution itself depend on the form of the attachment rate.

For $A_k = k^\gamma$ with $0 \leq \gamma \leq 1$, let us assume that both the degree distribution and $A(N)$ grow linearly with time. This hypothesis can be easily verified by solving for the first few N_k explicitly for attachment rates that do not grow faster than linearly with k. Thus writing $N_k(N) = N\, n_k$ and $A(N) = \mu N$ in Eq. (9.6), the overall dependence on N cancels out, leaving behind the recursion relations

$$n_k = \frac{A_{k-1}n_{k-1} - A_k n_k}{\mu} \qquad k > 1, \qquad \text{and} \quad n_1 = -\frac{A_1 n_1}{\mu} + 1. \tag{9.7}$$

The formal solution to this recursion is

$$n_k = \frac{\mu}{A_k} \prod_{j=1}^{k} \left(1 + \frac{\mu}{A_j}\right)^{-1}. \tag{9.8}$$

To make this solution explicit, we need the amplitude μ in $A(N) = \mu N$. Using the definition $\mu = \sum_{j\geq 1} A_j n_j$

in Eq. (9.8), we obtain the implicit relation

$$\sum_{k=1}^{\infty}\prod_{j=1}^{k}\left(1+\frac{\mu}{A_j}\right)^{-1}=1, \tag{9.9}$$

which shows that the amplitude μ depends on the functional form of the attachment rate. When $A_k = k^\gamma$ with $0 \le gamma \le 1$, a numerical solution of Eq. (9.9) shows that μ varies smoothly between 1 and 2 as γ increases from 0 to 1. However, for $\gamma > 1$ the left-hand side of (9.9) diverges, which implies that μ must also diverge.

To obtain an explicit solution for the degree distribution, we now substitute $A_k = k^\gamma$ into Eq. (9.8) and then use standard asymptotic methods to find the main result of this section:

$$n_k \sim \begin{cases} k^{-\gamma}\exp\left[-\mu\left(\frac{k^{1-\gamma}-2^{1-\gamma}}{1-\gamma}\right)\right], & 0\le\gamma<1;\\ k^{-\nu}, \quad \nu>2, & \gamma=1;\\ \text{“best seller”} & 1<\gamma<2;\\ \text{“bible”} & 2<\gamma. \end{cases} \tag{9.10}$$

Only the first two lines are actually obtained by the asymptotic analysis outlined above; however, we also write the results for $\gamma > 1$ for completeness.

There are many intriguing aspects of these fundamental results. First, we emphasize that a stretched exponential degree distribution arises for all $0 < \gamma < 1$ (with pure exponential decay for $\gamma = 0$). This range of γ should be viewed as the universal regime. On the other hand, ultra-singular behavior occurs for $\gamma > 2$ in which one node has a non-zero probability to be linked to *every* node in an infinite network, while only a finite number of links exist between all other nodes—the “bible” phase. It is easy to show that such a highly-connected node exists when $\gamma > 2$. Consider the specific network in which each new node links only to the initial node. After there are $N+1$ nodes in the network, the probability that the next new node links to the initial node is $N^\gamma/(N+N^\gamma)$. The probability that this connection pattern continues indefinitely is then

$$\mathcal{P}=\prod_{N=1}^{\infty}\frac{1}{1+N^{1-\gamma}},$$

so that $\mathcal{P} = 0$ for $\gamma \le 2$ and $\mathcal{P} > 0$ for $\gamma > 2$. Thus for $\gamma > 2$, there is the possibility of a bible—a node that is linked to every other node of the network, while all other nodes have a finite number of links, even in an infinite network.

When $1 < \gamma < 2$, singular behavior still arises in which one node is linked to all but a small number of other nodes. There is a also an infinite sequence of subtle connectivity transitions in the behavior of the number of low-degree nodes. For $3/2 < \gamma < 2$, the number of nodes of degree 2 grows as $N^{2-\gamma}$, while the number of nodes with degree > 2 remains finite. For $4/3 < \gamma < 3/2$, the number of nodes of degree 3 grows as $N^{3-2\gamma}$ and the number with degree > 3 is finite. Generally for $\frac{m+1}{m} < \gamma < \frac{m}{m-1}$, $N_k \sim N^{k-(k-1)\gamma}$ for $k \le m$, while the number of nodes with degree greater than m links is finite.

Finally, let us study the linear attachment rate. Consider first *strictly linear* attachment, $A_k = k$. In this case, the total event rate is $A = \sum_k A_k N_k = \sum_k kN_k = 2N \equiv \mu N$. Substituting this value $\mu = 2$ into Eq. (9.7) and solving the resulting recursion gives the pretty result

$$n_k=\frac{4}{k(k+1)(k+2)}=\frac{4\Gamma(k)}{\Gamma(k+3)}\sim\frac{4}{k^3}. \tag{9.11}$$

This function is a discrete power-law—the appropriate definition of a power-law when the function that is defined only on the positive integers. More germane to our discussion is that for the strictly linear attachment rate the degree distribution exponent $\nu = 3$. We can easily generalize this result to the case where each new node links to m pre-existing nodes, with every target node chosen by linear preferential attachment. Now the master equation is

$$\frac{dN_k}{dt}=\frac{m}{M_1}\left[(k-1)N_{k-1}-kN_k\right]+\delta_{km}\,, \tag{9.12}$$

and applying the same method of solution as that used for Eq. (9.6), we obtain the degree distribution

$$n_k = \frac{2m(m+1)}{k(k+1)(k+2)} \qquad k \geq m. \tag{9.13}$$

Thus the exponent of the degree distribution is again 3. However, the global nature of the network does depend on m. With $m > 1$ outgoing links, the ensuing network has closed loops while for $m = 1$ the networks as a tree topology.

Perhaps the most surprising feature of linear preferential attachment at first sight is that the exponent of the power-law degree distribution (second line of Eq. (9.10)) is actually *non-universal*. The asymptotic evaluation of the product in Eq. (9.8) generally leads to a degree distribution exponent $\nu = 1 + \mu$ that can assume any value greater than 2. All that is required is that the attachment rate is *asymptotically linear*, $A_k \sim k$, rather than strictly linear, $A_k = k$. This non-universal behavior is counter to the conventional wisdom of critical phenomena in which power laws, by their very nature, should be universal. Nearly everything about linear preferential attachment is counter to this dogma.

Redirection

To illustrate the vagaries of a network with an asymptotically linear attachment rate, consider the shifted linear rate $A_k = k + \lambda$. A nice way to realize this model is by random network growth that is augmented by *redirection*. In redirection a new node **n** is added and an earlier node **x** is selected *uniformly* from the set of all nodes as a possible target for attachment. With probability $1 - r$, the link from **n** to **x** is created. However, with probability r, the link is redirected to the ancestor **y** of node **x** (Fig. 9.2).

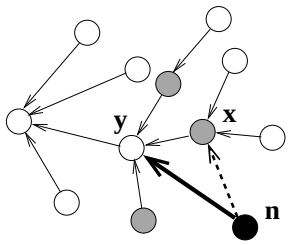

Figure 9.2: Illustration of redirection. The new node (solid) selects a target node **x** at random. With probability $1 - r$ a link is established to this target (dashed arrow), while with probability r the link is established to **y**, the ancestor of **x** (thick solid arrow). The rate at which attachment to **y** occurs by redirection is proportional to the number of its upstream neighbors (shaded).

Let's solve this model within the master equation framework. According to the defining processes of the model (Fig. 9.2), the degree distribution $N_k(N)$ evolves by the master equation

$$\frac{dN_k}{dN} = \frac{1-r}{M_0}\left[N_{k-1} - N_k\right] + \delta_{k1} + \frac{r}{M_0}\left[(k-2)N_{k-1} - (k-1)N_k\right]. \tag{9.14}$$

For redirection probability $r = 0$, the first three terms on the right-hand side of Eqs. (9.14) are the same as in the preferential attachment network with $A_k = 1$ (equiprobable attachment to any node). The last two terms account for the change in N_k due to redirection. To understand their origin, consider the gain term. Since the initial node is chosen uniformly, if redirection does occur, the probability that a node with $k-1$ pre-existing links receives the newly-redirected link is proportional to the number of pre-existing incoming links $k-2$ (shaded nodes in Fig. 9.2). A similar argument applies for the redirection-driven loss term. Finally, by combining the terms in Eq. (9.14), the master equation reduces to Eq. (9.6) with $A_k = r[k - 1 + (1 - r)/r]$, except for a difference of an overall factor of 2 due to slightly different denominators in the two master equations. Scaling out an irrelevant overall factor r, A_k becomes the shifted linear attachment rate $k + \lambda$, with $\lambda = \frac{1}{r} - 2$.

To determine the degree distribution for the shifted linear attachment rate, we use the fact that $A(N) = \sum_j A_j N_j(N)$ now equals $A(N) = M_1(N) + \lambda M_0(N)$. Then making use of the relations $A = \mu N$, $M_0 = N$

and $M_1 = 2N$, we obtain $\mu = 2 + \lambda$. Finally, we substitute the result $\mu = 2 + \lambda$ in Eq. (9.8) to obtain the degree distribution

$$n_k = (2+\lambda)\,\frac{\Gamma(3+2\lambda)}{\Gamma(1+\lambda)}\,\frac{\Gamma(k+\lambda)}{\Gamma(k+3+2\lambda)}. \tag{9.15}$$

Asymptotically, this distribution decays as $k^{-\nu}$, with $\nu = 3 + \lambda = 1 + \frac{1}{r}$, a value that can be tuned to any value larger than 2. Thus a simple additive shift in the attachment rate profoundly affects the asymptotic degree distribution. Amusingly, for $r = 1/2$, the redirection process is equivalent to strictly linear preferential attachment. While it might seem surprising that a small additive shift in the attachment rate affects the asymptotic degree distribution, this result has a natural explanation. The actual degree of a node is the result of the product of attachment rates as the network is built. The fact that the degree of a node arises from a product of random variables is the mechanism that allows an additive shift to the rate to play such a large role.

Node attractiveness

A natural generalization of preferential attachment from a practical perspective is to endow each node with an intrinsic "attractiveness". This attribute accounts for the obvious fact that in many real settings (scientific publications, websites, *etc.*) not all nodes are equivalent, but rather, some are more attractive than others at their inception. Thus it is natural that the subsequent attachment rate to a node should be a function of both its degree and its attractiveness. For this generalization of preferential attachment, the master equation approach easily gives complete results for the degree distribution.

We assign each node an initial attractiveness $\eta > 0$ that is chosen from a specified distribution $p_0(\eta)$. We define attractiveness to modify the node attachment rate as follows: for a node with degree k and attractiveness η, the attachment rate is simply $A_k(\eta)$. To characterize how nodes evolve, we now need to characterize nodes both by their degree and their attractiveness. Thus let $N_k(\eta)$ be the number of nodes with degree k and attractiveness η. The evolution of this joint degree-attractiveness distribution is governed by the master equation

$$\frac{dN_k(\eta)}{dN} = \frac{A_{k-1}(\eta)N_{k-1}(\eta) - A_k(\eta)N_k(\eta)}{A} + p_0(\eta)\delta_{k1}, \tag{9.16}$$

where $A = \int d\eta \sum_k A_k(\eta)N_k(\eta)$ is the total rate. Following the same approach as that used to analyze Eq. (9.6), we substitute $A = \mu N$ and $N_k(\eta) = N n_k(\eta)$ into Eq. (9.16), and the solve the resulting recursion relation to give

$$n_k(\eta) = p_0(\eta)\frac{\mu}{A_k(\eta)}\prod_{j=1}^{k}\left(1 + \frac{\mu}{A_j(\eta)}\right)^{-1}. \tag{9.17}$$

For concreteness, let's study the generalized linear rate $A_k(\eta) = \eta k$—linear in the degree *and* in the attractiveness. Applying the same analysis as in the homogeneous network, we obtain the degree distribution

$$n_k(\eta) = \frac{\mu\, p_0(\eta)}{\eta}\,\frac{\Gamma(k)\,\Gamma\left(1+\frac{\mu}{\eta}\right)}{\Gamma\left(k+1+\frac{\mu}{\eta}\right)}. \tag{9.18}$$

Thus for nodes with a given attractiveness η, the asymptotic degree distribution is the power law $n_k(\eta) \sim k^{-1-\mu/\eta}$. To determine the full distribution, we need the amplitude μ. We therefore substitute (9.18) into the definition $\mu = \int d\eta \sum_{k\geq 1} A_k(\eta)\, n_k(\eta)$ and use the identity

$$\sum_{k=1}^{\infty}\frac{\Gamma(k+u)}{\Gamma(k+v)} = \frac{\Gamma(u+1)}{(v-u-1)\,\Gamma(v)}$$

to yield the implicit relation

$$1 = \int d\eta\, p_0(\eta)\,\left(\frac{\mu}{\eta} - 1\right)^{-1}. \tag{9.19}$$

This condition on μ leads to two alternatives: in the pathological case where the support of η is unbounded and arbitrarily attractive nodes can exist, then the integral diverges and there is no solution for μ. Here the most attractive node is connected to a finite fraction of all links. Conversely, if the support of η is bounded, then the degree distribution for fixed η is simply the power law $n_k(\eta) \sim k^{-\nu(\eta)}$, with an attractiveness-dependent decay exponent $\nu(\eta) = 1 + \mu/\eta$. However the degree distribution averaged over all attractiveness values, $\langle n_k \rangle = \int d\eta\, n_k(\eta)$, is no longer a power law, but rather $\langle n_k \rangle$ is governed by properties of the initial attractiveness distribution near the upper cutoff. For example, if $p_0(\eta) \sim (\eta_{\max} - \eta)^{\omega - 1}$ (with $\omega > 0$ to ensure normalization), the total degree distribution is

$$n_k \sim k^{-(1+\mu/\eta_{\max})}\, (\ln k)^{-\omega}. \tag{9.20}$$

Correlations

The master equation can be easily adapted to provide more detailed network properties. One such example is the correlation between the degrees of connected nodes. Correlations naturally develop because a node with large degree is likely to be old. Thus its ancestor is also old and hence also has a large degree. To quantify these correlation, define $C_{kl}(N)$ as the number of nodes of degree k that attach to an ancestor node of degree l (Fig. 9.3). For example, in the final network of Fig. 9.1, there are $N_1 = 6$ nodes of degree 1, with $C_{12} = 1$, $C_{13} = 2$, and $C_{15} = 3$. There is also $N_2 = 1$ node of degree 2, with $C_{25} = 1$, and $N_3 = 1$ nodes of degree 3, with $C_{35} = 1$.

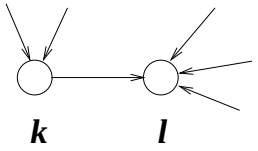

Figure 9.3: Definition of the node degree correlation C_{kl} for the case $k = 3$ and $l = 4$.

For simplicity, we specialize to the strictly linear attachment rate. The degree correlation function $C_{kl}(N)$ evolves according to the master equation

$$M_1 \frac{dC_{kl}}{dt} = (k-1)C_{k-1,l} - kC_{kl} + (l-1)C_{k,l-1} - lC_{kl} + (l-1)C_{l-1}\,\delta_{k1}. \tag{9.21}$$

The processes that correspond to each of the terms in this equation are illustrated in Fig. 9.4. The first two terms on the right account for the change in C_{kl} due to the addition of a link onto a node of degree $k-1$ (gain) or k (loss) respectively, while the second set of terms gives the change in C_{kl} due to the addition of a link onto the ancestor node. Finally, the last term accounts for the gain in C_{1l} due to the addition of a new node. A crucial point is that these equations are closed, as they do not involve any higher-order correlations.

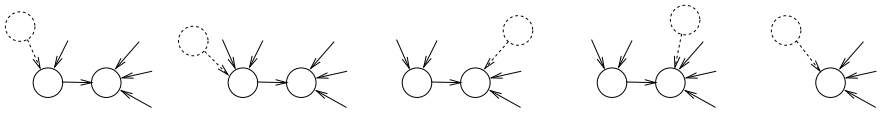

Figure 9.4: The processes that contribute ((i)–(v) in order) to the various terms in the master equation (9.21). The newly-added node and link are shown dashed.

As in the case of the node degree, the time dependence can be separated as $C_{kl} = Nc_{kl}$ to reduce Eq. (9.21) to the time-independent recursion

$$(k+l+2)c_{kl} = (k-1)c_{k-1,l} + (l-1)c_{k,l-1} + (l-1)c_{l-1}\,\delta_{k1}. \tag{9.22}$$

This can be further reduced to a constant-coefficient inhomogeneous recursion relation by the substitution

$$c_{kl} = \frac{\Gamma(k)\,\Gamma(l)}{\Gamma(k+l+3)}\; d_{kl}$$

to yield

$$d_{kl} = d_{k-1,l} + d_{k,l-1} + 4(l+2)\delta_{k1}. \tag{9.23}$$

Solving Eqs. (9.23) for the first few k yields the pattern of dependence on k and l from which one can then infer the solution

$$d_{kl} = 4\,\frac{\Gamma(k+l)}{\Gamma(k+2)\,\Gamma(l-1)} + 12\,\frac{\Gamma(k+l-1)}{\Gamma(k+1)\,\Gamma(l-1)}, \tag{9.24}$$

from which we ultimately obtain

$$c_{kl} = \frac{4(l-1)}{k(k+l)(k+l+1)(k+l+2)}\left[\frac{1}{k+1} + \frac{3}{k+l-1}\right]. \tag{9.25}$$

An important feature of this result is that the correlation function does not factorize, that is, $c_{kl} \neq n_k n_l$. This result is is a basic distinction between preferential attachment networks and classical random graphs.

While the solution of Eq. (9.25) is unwieldy, it greatly simplifies in the scaling regime, $k \to \infty$ and $l \to \infty$ with l/k finite, where the scaled form of the solution is

$$c_{kl} \sim k^{-4}\,\frac{4\frac{l}{k}\left(\frac{l}{k}+4\right)}{\left(1+\frac{l}{k}\right)^4} \to \begin{cases} 16\,(l/k^5) & l \ll k, \\ 4/(k^2\,l^2) & l \gg k. \end{cases} \tag{9.26}$$

For fixed large k, c_{kl} has a single maximum at $(l/k)^* = (\sqrt{33}-5)/2 \approx 0.372$. Thus a node whose degree k is large is typically linked to another node whose degree is also large; the typical degree of the ancestor is 37% that of the daughter node.

Global properties

In addition to determining the degree distribution and degree correlations, the master equation can be applied to determine global properties. One example is the *out-component* with respect to a given node **x**—the set of nodes that can be reached by iteratively following directed links that emanate from **x** (Fig. 9.5). Conversely, the *in-component* to node **x** is the set of nodes from which **x** can be reached by following a path of directed links—the progeny of **x**.

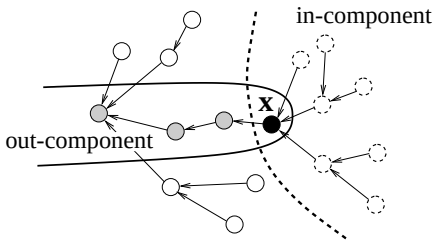

Figure 9.5: In-component and out-components of node **x**.

In-component

Let's study the in-component size distribution for a growing network with random attachment, in which a node attaches to any other node equiprobably; that is, $A_k = 1$. We consider this example because many results about network components are independent of the form of the attachment rate and it suffices to

consider the simplest situation. For a constant attachment rate, the number $I_s(N)$ of in-components with s nodes satisfies the master equation

$$\frac{dI_s}{dN} = \frac{(s-1)I_{s-1} - sI_s}{A} + \delta_{s1}. \tag{9.27}$$

The loss term accounts for processes in which the attachment of a new node to an in-component of size s increases its size by one. The rate of this attachment is simply proportional to s, even if there is more than one (disjoint) in-component. Thus the total loss rate for $I_s(N)$ is simply $sI_s(N)$. A similar argument applies for the gain term. Finally, dividing by $A(N) = \sum_j A_j N_j(N)$ converts these rates to normalized probabilities. For the constant attachment rate, $A(N) = N$. Interestingly, Eq. (9.27) is almost identical to the master equation (9.6) for the degree distribution for the linear preferential attachment network, except that the denominator equals N rather than $2N$. This change in the normalization factor is responsible for shifting the exponent of the resulting distribution from -3 to -2.

To determine $I_s(N)$, we again note, by explicitly solving the first few of the master equations, that each I_s grows linearly in time. Thus we substitute $I_s(N) = Ni_s$ into Eqs. (9.27) to obtain $i_1 = 1/2$ and $i_s = i_{s-1}(s-1)/(s+1)$. This immediately gives

$$i_s = \frac{1}{s(s+1)}, \tag{9.28}$$

and the s^{-2} tail for the in-component distribution is a robust feature that is independent of the form of the attachment rate.

Out-component

The complementary out-component from each node can be determined by constructing a mapping between the out-component and an underlying network genealogy. We build a genealogical tree for a network by taking generation $g = 0$ to be the initial node. Nodes that attach to those in generation g form generation $g+1$; the node index does not matter in this characterization. For example, in the final network of Fig. 9.1, node 1 is the "ancestor" of 2, while nodes 3 and 10 are the "descendants" of 2, and there are 5 nodes in generation $g = 1$ and 3 in $g = 2$. This leads to the genealogical tree of Fig. 9.6.

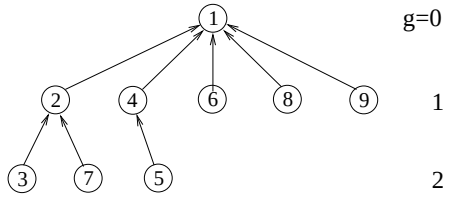

Figure 9.6: Genealogy of the network in Fig. 9.1 with nodes arranged according to generation number. The node indices indicate when each is introduced.

The genealogical tree provides a convenient way to characterize the out-component distribution. As can be easily verified from Fig. 9.6, the number O_s of out-components with s nodes equals L_{s-1}, the number of nodes in generation $s-1$ in the genealogical tree. We therefore compute $L_g(N)$, the size of generation g when the network contains N nodes, again for the constant attachment rate. We determine $L_g(N)$ by noting that $L_g(N)$ increases when a new node attaches to a node in generation $g-1$, an event that occurs with rate L_{g-1}/N. This gives the differential equation $\dot{L}_g(N) = L_{g-1}/N$ with solution $L_g(\tau) = \tau^g/g!$, where $\tau = \ln N$. The generation size $L_g(N)$ therefore grows with g, when $g < \tau$, and then decreases and becomes of order 1 when $g = e\tau$. The genealogical tree therefore contains approximately $e\tau$ generations for a network of N nodes. Finally, the number O_s of out-components with s nodes equals

$$O_s(\tau) = \tau^{s-1}/(s-1)!. \tag{9.29}$$

This result allows us to determine the diameter of the network, since the maximum distance between any pair of nodes is twice the distance from the root to the last generation. Therefore the diameter of the network scales as $2e\tau \approx 2e \ln N$; this is the same dependence on N as in the Erdös-Rényi random graph. More importantly, this result shows that the diameter of a network that grows by preferential attachment is always small–ranging from the order of $\ln N$ for a constant attachment rate, to the order of one for a superlinear attachment rate.

Fluctuations

One of the surprising aspects of linear preferential attachment networks is the sensitivity of the degree distribution exponent to fine details of the growth process itself. We can provide an understanding of this unexpected behavior by examining the essential role of fluctuations between different realizations of the network.

As an instructive illustration, we study the degree of the first node in the network. Let $P(k,N)$ be the probability that the first node has degree k in a network of N links. For strictly linear preferential attachment, $A_k = k$, this probability obeys the master equation

$$P(k,N+1) = \frac{k-1}{2N}\,P(k-1,N) + \frac{2N-k}{2N}\;P(k,N). \tag{9.30}$$

The first term on the right accounts for the situation when the first node has degree $k-1$. A new node can attach to it with probability $(k-1)/2N$, thereby increasing the probability for the first node to have degree k. Conversely, with probability $(2N-k)/2N$ a new node does not attach to the earliest node, thereby giving the second term on the right.

The solution to the master equation (9.30) for the "dimer" initial condition •—• is

$$P(k,N) = \frac{1}{2^{2N-k-1}}\,\frac{(2N-k-1)!}{(N-k)!\,(N-1)!} \longrightarrow \frac{1}{\sqrt{\pi N}}\;e^{-k^2/4N}\,, \tag{9.31}$$

where the asymptotic behavior applies in the limit $N\to\infty$, with the scaling variable $k/N^{1/2}$ being finite. Thus the average degree of the first node is $\langle k\rangle_1 = \sqrt{4N/\pi} \approx 1.228\sqrt{N}$. On the other hand, from the extremal criterion for the largest degree in the network

$$\sum_{k_{\max}}^{\infty} N n_k = 1,$$

and using asymptotic degree distribution $n_k \sim 4/k^3$, we obtain the largest degree $k_{\max} \sim \sqrt{2N} \approx 1.4142\sqrt{N}$. Thus the degree of the first node of the network is close the the largest degree; this implies that there is a substantial probability that the first node in the network is the one with the largest degree.

Although $P(k,N)$ contains all information about the degree of the first node, its moments $\langle k^a\rangle_N = \sum k^a P(k,N)$ are simpler to appreciate. Using Eq. (9.30), the average degree of the initial node satisfies the recursion relation

$$\langle k\rangle_{N+1} = \langle k\rangle_N \left(1+\frac{1}{2N}\right)\,, \tag{9.32}$$

whose solution is

$$\langle k\rangle_N = \Lambda\,\frac{\Gamma\left(N+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma(N)} \sim \frac{\Lambda}{\sqrt{\pi}}\,N^{1/2}\,. \tag{9.33}$$

The prefactor Λ depends on the initial condition, with $\Lambda = 2, 8/3, 16/5, \ldots$ for the dimer, trimer, tetramer, *etc.*, initial conditions.

This multiplicative dependence on the initial condition means that the first few growth steps substantially affect the average degree of the first node. For example, for the dimer initial condition, the average degree of the first node is, asymptotically, $\langle k\rangle_N \sim 2\sqrt{N/\pi}$. However, if the second link attaches to the first node, an effective trimer initial condition arises and $\langle k\rangle_N \sim (8/3)\sqrt{N/\pi}$. Thus small initial perturbations at the beginning of the network growth lead to huge differences in the degree of the first node.