Chapter 10

COMPLEX NETWORKS

The study of complex networks represents a non-traditional application of non-equilibrium statistical physics. As we shall discuss, the tools of the field seem particularly appropriate to quantify basic properties of complex networks, such as percolation transitions and many geometrical properties. In this chapter, we will present some of the simplest complex network models and apply the master equation to quantify many of their basic features through a dynamical approach.

10.1 Erdös-Rényi Random Graph

A simple and classic starting example is the *Erdös-Rényi (ER) random graph.* This graph consists of N nodes in which each node pair may be joined by a link according to a connection probability that we define as p/N, with $0 \leq p \leq N$. Since any pair of nodes may be connected equiprobably, the ER graph has no spatial structure. A striking feature of the ER graph is the existence of a phase transition, in the limit $N \to \infty$, at $p = 1$. For $p < 1$, a finite network consists of disconnected clusters whose a maximum size is of the order of $\ln N$. At $p = 1$ the size of the largest cluster beccomes of the order of $N^{2/3}$. In the $N \to \infty$ limit, this largest cluster becomes the incipient infinite cluster; the term incipient refers to the fact that this cluster comprises a vanishing fraction of all nodes as $N \to \infty$. Finally, for $p > 1$ the largest cluster consists of a finite fraction of all nodes. For $p = N$, all $N(N-1)/2$ pairs of nodes are connected to give the *complete graph.*

While the ER graph is usually defined as a static problem — each link is independently present with probability p/N — we recast the ER graph *dynamically* by creating links between nodes at a constant rate. Within this formulation, the master equation provides in a simple way to determine the structure of the ER graph. A similar dynamical perspective was used to determine the coverage evolution in irreversible adsorption (Chapter 6), from which the final coverage — an ostensibly static quantity — emerged as a simple byproduct.

mention ER graph is like a tree with $z^g \approx N$ generations.

Degree distribution

We build the ER graph by starting with N disconnected nodes and then introducing links one by one between randomly-selected node pairs. The two nodes may be the same, and also, more than one link may be created between a pair of nodes. However, these two processes occur with a vanishingly small probability when $N \to \infty$ and may be ignored. For convenience, we define the rate at which each link is introduced as $(2N)^{-1}$. The total number of links at time t is then $Nt/2$ and the average degree equals t. Here degree is the number of links that are attached to a node. Thus the average degree evolves by a stochastic process in which $k \to k+1$ at rate 1.

As we shall see, the distribution of degrees is an important characteristic of complex networks. We define the degree distribution as n_k, the fraction of nodes of degree k. Nodes of degree k are created from nodes of degree $k-1$ at rate 1, and nodes of degree k are also lost at rate 1 due to the creation of nodes of degree

$k+1$. The degree distribution therefore satisfies the master equation of the Poisson process

$$\frac{dn_k}{dt} = n_{k-1} - n_k\,, \tag{10.1}$$

which applies for all $k \geq 0$ if we impose the additional condition $n_{-1} \equiv 0$. For a network with no links at $t = 0$, the initial condition is $n_k(0) = \delta_{k,0}$. Eqs. (10.1) may then be solved one by one starting with n_0 (see problem 10.1), and the degree distribution is

$$n_k = \frac{t^k}{k!}\, e^{-kt}\ . \tag{10.2}$$

From this expression, 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 degree distribution becomes sharp in the thermodynamic limit.

The percolation transition

We probe the percolation transition in the ER graph by studying the cluster size distribution. Here a cluster is defined as the set of all nodes that are connected by links into a single connected component. Initially the network consists of N clusters of size 1. As links are added, clusters merge and their number systematically decreases while their mean size grows. Since a link occurs equiprobably between any pair of nodes, there are $i \times j$ ways to join disconnected clusters of sizes i and j to create a cluster of size $i+j$; consequently, the overall rate for this event is $ij/(2N)$. This process is precisely product kernel aggregation discussed in chapter 4, and we can make use of the results derived therein to determine the cluster size distribution.

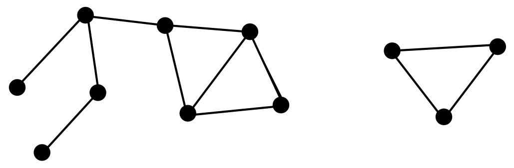


Figure 10.1: A realization of the ER graph that consists of two clusters: one of size 8 and one of size 3.

Let $c_k(t)$ be the density of clusters that contain k nodes at time t. The cluster size distribution obeys the master equation

$$\frac{dc_k}{dt} = \frac{1}{2}\sum_{i+j=k} (ic_i)(jc_j) - k\, c_k, \tag{10.3}$$

with $c_k(0) = \delta_{k,1}$ for the disconnected initial condition. The gain term accounts for the merger between two clusters whose sizes sum to k, and the loss term accounts for the loss of clusters of size k due to their linking with other clusters. As a preliminary, it is useful to study moments of the size distribution, $M_n = \sum_k k^n c_k$. The first moment, M_1, is just the fraction of nodes that belong to finite clusters; this quantity therefore equals 1 as long as there is no gelation, a condition that holds for $t < 1$ (see Sec. 4.1 for a detailed discussion of this point). As derived in Eq. (4.30), the second moment, which gives the mean cluster size, obeys the rate equation $\dot{M}_2 = M_2^2$ for $t < 1$. With the initial condition $M_2(0) = 1$, the solution is simply

$$M_2 = (1-t)^{-1}, \tag{10.4}$$

which shows that an infinite cluster forms at $t = 1$ when the number of nodes of the ER graph is infinite. As t increases beyond the percolation point, this infinite cluster contains a finite fraction of all nodes and eventually engulfs the entire system.

From our earlier discussion of product-kernel aggregation, the cluster size distribution is [see Eq. (4.38)]

$$c_k(t) = \frac{k^{k-2}}{k!}\, t^{k-1}\, e^{-kt}\ . \tag{10.5}$$

While this exact distribution seems to have a smooth time dependence, there is a dramatic change in behavior as t passes through 1 that can be seen by using Stirling's approximation to give the asymptotic behaviors:

$$c_k(t) \sim \begin{cases} e^{-k(t-\ln t-1)} & t<1; \\ (2\pi)^{-1/2}k^{-5/2} & t=1. \end{cases} \tag{10.6}$$

The existence of a power-law distribution at $t=1$ signals the percolation transition where the mean cluster size diverges.

What happens when the ER graph is finite? The transition is no longer sharp, and singular behavior is replaced by finite-size scaling laws. For example, for a finite network, the incipient infinite cluster now becomes just the largest cluster. Its size, M, can be estimated by the extremal criterion,

$$N\sum_{k\geq M}^{\infty} c_k = 1, \tag{10.7}$$

which states that there should be a single cluster whose size is in the range $[M,\infty]$. Using the asymptotic forms in Eqs. (10.6) for $c_k(t)$ in the extremal criterion, and approximating the sum by an integral we obtain

$$M \sim \begin{cases} \ln N & t<1; \\ N^{2/3} & t=1. \end{cases} \tag{10.8}$$

Thus clusters are at most of size $\ln N$ below percolation, while a "giant" cluster of size $N^{2/3}$ emerges as percolation is approached. When does this cluster appear? Close to the percolation time, we may estimate the typical cluster mass as

$$M_2 \approx \sum^{M} k^2 c_k \sim \int^{M} k^2\, k^{-5/2}\,dk \sim M^{1/2},$$

and equating this result to $M_2=(1-t)^{-1}$ in Eq. (10.4), we obtain $M\sim(1-t)^{-2}$. Since M also scales as $N^{2/3}$, we obtain the percolation time in a finite network:

$$t \sim 1-N^{-1/3}. \tag{10.9}$$

Paths and cycles

A deeper characterization of the ER graph may be obtained by studying paths and cycles in the network. A pair of nodes that are connected by a sequence of links forms a *path*. How do paths evolve with time? When a newly-added link connects the ends of two paths of lengths n and m, the result is a path of length $n+m+1$. For $\ell>0$, the density of *distinct* paths that contain l links at time t, $P_\ell(t)$, evolves as

$$\frac{dP_\ell}{dt} = \sum_{n+m=\ell-1} P_n P_m. \tag{10.10}$$

The initial condition is $P_\ell(0)=\delta_{l,0}$ and we define $P_0(t)=1$. The solution of (10.10) is simply

$$P_\ell = t^\ell. \tag{10.11}$$

For example, $P_1=t$ corresponds to the link density being equal to $t/2$ and that every link corresponds to two distinct paths of length 1 **ambiguous**. From this path length distribution, the total density of paths and the typical path length,

$$\begin{aligned} P_{\rm tot} &\equiv \sum_\ell P_\ell = \frac{1}{1-t}\ , \\ \langle \ell\rangle &= \frac{\sum_\ell \ell P_\ell}{\sum_\ell P_\ell} = \frac{t}{1-t}\ , \end{aligned} \tag{10.12}$$

respectively, both diverge at $t = 1$.

When a link directly joins two nodes that are already on the same path, a cycle forms. Let the average number of cycles of length ℓ at time t be $Q_\ell(t)$. This quantity is coupled to the path length density through the rate equation

$$\frac{dQ_\ell}{dt} = \frac{1}{2}\,P_{\ell-1}. \tag{10.13}$$

The right-hand side equals the link creation rate $1/(2N)$ times the total number of paths $NP_{\ell-1}$. Solving this equation, the cycle length distribution is

$$Q_\ell = \frac{t^\ell}{2\ell}. \tag{10.14}$$

Consequently, the total number of cycles in the system is $Q_{\rm tot} \equiv \sum_\ell Q_\ell = \frac{1}{2}\ln\frac{1}{1-t}$, which diverges weakly as $t \to 1$.

10.2 Sequentially Growing Networks

Sequential growth describes the evolution of many networked systems, such as the Internet and the world-wide web, where new routers or websites are added incrementally. It is natural to model such growing networks, by adding nodes one by one with each new node attaching to a "target" node, or a set of target nodes, with attachment rate A_k that depends only on the degree of the target (Fig. 10.2). The number of nodes N therefore plays the role of a time-like variable and we sometimes refer to N as the "time". The case where the attachment rate A_k increases with k defines *preferential attachment*, which encapsulates the notion of the "rich get richer". For example, in the context of scientific citations, preferential attachment means that a currently well-cited paper is more likely to be well cited in the future simply by virtue of being well cited now. We now apply the master equation to elucidate the structure of such growing networks.

Uniformly Growing Tree

As a starting example, we study the simpler case of the *uniformly growing tree* (UGT), also known as the random recursive tree. The growth rules of a UGT at any stage of its evolution are:

1. Pick one of the nodes of the UGT with uniform probability.
2. Introduce a new node that links to the target node.

Starting with the initial state of a single node, these steps repeated until the tree reaches a desired number of nodes N. Each node is distinguished by the order in which it is introduced so that there are $N!$ distinct trees of N nodes. Since each newly-introduced node has a single link, no closed loops can be generated. Thus if the graph initially is a tree, it remains a tree.

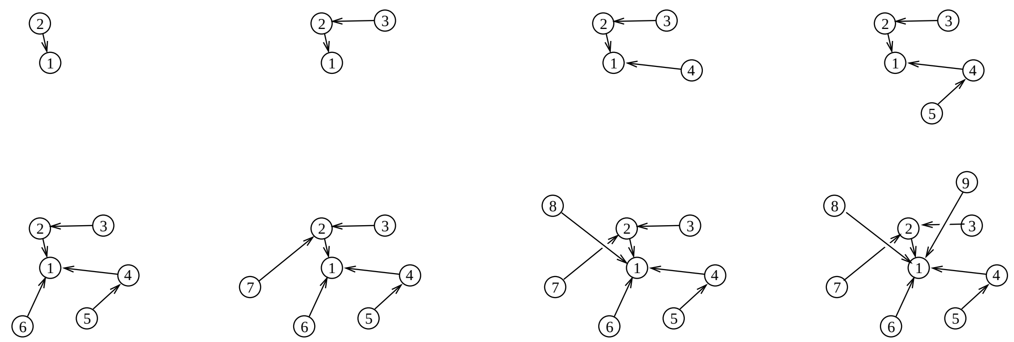


Figure 10.2: Evolution of one realization of a uniformly growing tree (upper left to lower right).

The degree distribution

Let's determine the degree distribution of the UGT, namely, the number of nodes of degree k when the network contains N nodes, $N_k(N)$. This distribution is distinct for each realization of the UGT, and, from the statistical physics perspective, the interesting quantity is degree distribution *averaged* over all realizations of trees with fixed N. Although the average distribution does not have a simple form for small N, it simplifies considerably when N is large. In this limit, the average number of nodes of degree k evolves according to

$$\frac{dN_k}{dN} = \frac{N_{k-1} - N_k}{N} + \delta_{k1}, \tag{10.15}$$

where we denote the average degree distribution by N_k **average notation**. This master equation is essentially the same as that for the ER graph, Eq. (10.1), except for the additional delta-function term that accounts for the single outgoing link of the new node.

To get a feeling for the solution, let's start by solving the master equations (10.15) one by one. With the understanding that $N_{-1} = 0$, the master equations are, explicitly:

$$\begin{aligned}
\dot{N}_0 &= -\frac{N_0}{N} \\
\dot{N}_1 &= \frac{N_0 - N_1}{N} + 1 \\
\dot{N}_2 &= \frac{N_1 - N_2}{N} \\
\dot{N}_3 &= \frac{N_2 - N_3}{N},
\end{aligned}$$

etc., where the overdot denotes differentiation with respect to N. The solution to the first equation is simply $N_0 = 1/N$. We now rewrite the equation for N_1 as $\dot{(N_1N)} = N + N_0$, with asymptotic solution $N_1 \sim N/2$. By the same method, the equation for N_2 becomes $\dot{(N_2N)} = N_1$, from which $N_2 \sim N/4$. This pattern of behavior continues for all k so that we conclude that all the N_k are proportional to N.

It therefore is convenient to work with the density of nodes of degree k, $n_k \equiv N_k/N$. In terms of this quantity, Eq. (10.15) reduces to

$$n_k = n_{k-1} - n_k + \delta_{k1}, \tag{10.16}$$

which are trivially soluble. Starting with $n_0 = 0$, we obtain $n_1 = \frac{1}{2}$, $n_2 = \frac{1}{4}$, *etc.*, and the general solution is simply $n_k = 2^{-k}$. Thus the UGT has a rapidly decaying degree distribution in which the average degree equals 2 and the largest degree is of order $\ln N$ for a network of N nodes.

Genealogical tree

It is revealing to study the genealogy underlying a UGT. We build this genealogy by taking generation $g = 0$ to be the initial node. Nodes that attach to those in generation g form generation $g + 1$. For example, in the final network of Fig. 10.2, node 1 is the "ancestor" of 2, while nodes 3 and 7 are the "descendants" of 2. There are 5 nodes in generation $g = 1$ and 3 in $g = 2$, leading to the genealogy of Fig. 10.3.

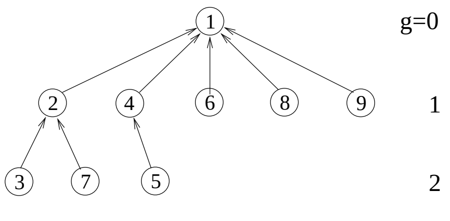


Figure 10.3: Genealogy of the network in Fig. 10.2 with nodes arranged according to generation number. The node indices indicate when each is introduced.

How many generations are there in a tree of N nodes? What is the size of the g^{th} generation, $L_g(N)$? To determine $L_g(N)$, note that $L_g(N)$ increases when a new node attaches to a node in generation $g-1$, an event that occurs with probability L_{g-1}/N. This gives the evolution equation $\dot{L}_g(N) = L_{g-1}/N$, with solution $L_g(\tau) = \tau^g/g!$, where $\tau = \ln N$. Using Stirling's approximation, we see that 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 tree of N nodes. This latter result also determines the diameter of the tree, since the diameter (also the maximum distance between any pair of nodes) is twice the distance from the root to the last generation. Therefore the diameter of the tree scales as $2e\tau \approx 2e\ln N$; this is the same dependence on N as in the Erdös-Rényi random graph.

Redirection

We now generalize the UGT to incorporate the mechanism of *redirection*. In redirection a new node **n** is introduced and an earlier node **x** is uniformly selected as a target. 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. 10.4).

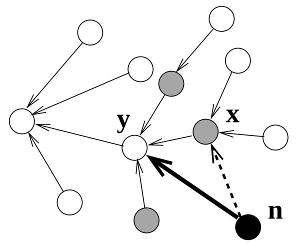


Figure 10.4: Illustration of redirection. The new node (solid) selects a target node **x** uniformly at random. With probability $1-r$ a link is established to this target (dashed arrow), while with probability r the link is established to **y**, the ancestor of **x** (thick solid arrow). The rate at which attachment to **y** occurs by redirection is proportional to the number of its upstream neighbors (shaded).

According to the defining processes of redirection shown in Fig. 10.4, the degree distribution $N_k(N)$ evolves according to

$$\frac{dN_k}{dN} = \frac{1-r}{N}\left[N_{k-1} - N_k\right] + \delta_{k1} + \frac{r}{N}\left[(k-2)N_{k-1} - (k-1)N_k\right]. \tag{10.17}$$

The first three terms correspond to the growth processes of the UGT, whose master equation (10.15) is recovered for redirection probability $r = 0$. 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 of degree $k-1$ receives the newly-redirected link is proportional to the number of its incoming links, which equals $k-2$ (shaded nodes in Fig. 10.4). A similar argument applies for the redirection-driven loss term.

Combining the terms in Eq. (10.17), the master equation becomes

$$\frac{dN_k}{dN} = \frac{r}{N}\left\{\left[k-1+\left(\frac{1}{r}-2\right)\right]N_{k-1} - \left[k+\left(\frac{1}{r}-2\right)\right]N_k\right\} + \delta_{k1}. \tag{10.18}$$

Thus uniform attachment, in conjunction with redirection, generates a growing network in which the attachment rate to a node of degree k is a *linear* function of k, albeit with an additive shift. To determine the degree distribution, we now study preferential attachment networks systematically, from which the solution of the redirection model follows easily.

Preferential attachment networks

The degree distribution

Let us again consider networks in which each new node attaches to one pre-existing node of degree k with rate A_k. The master equation for the degree distribution is, in analogy with Eq. (10.15) for the uniformly growing tree,

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A(N)} + \delta_{k1}. \tag{10.19}$$

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 equals 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 $A_{k-1}N_{k-1}/A(N)$ 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, $A_k N_k/A(N)$ 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.

For attachment rates that do not grow faster than linearly with k, both the degree distribution and $A(N)$ grow linearly with time (see problem 10.x). This fact suggests making the substitutions $N_k(N) = N\, n_k$ and $A(N) = \mu N$ in Eq. (10.19). With this step, the overall dependence on N cancels out, leaving behind the recursion relations

$$n_k = \frac{A_{k-1}n_{k-1} - A_k n_k}{\mu} \qquad k > 1, \qquad \text{and} \quad n_1 = -\frac{A_1 n_1}{\mu} + 1, \tag{10.20}$$

with formal solution

$$n_k = \frac{\mu}{A_k}\prod_{j=1}^{k}\left(1 + \frac{\mu}{A_j}\right)^{-1}. \tag{10.21}$$

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. (10.21), we obtain the condition

$$\sum_{k=1}^{\infty}\prod_{j=1}^{k}\left(1 + \frac{\mu}{A_j}\right)^{-1} = 1, \tag{10.22}$$

which shows that the amplitude μ depends on the functional form of the attachment rate. When $A_k = k^\gamma$ with $0 \leq \gamma \leq 1$, a numerical solution of Eq. (10.22) shows that μ varies smoothly between 1 and 2 as γ increases from 0 to 1.

For sublinear attachment rates, $\gamma < 1$, we rewrite the product Eq. (10.21) as the exponential of a sum, convert the sum to an integral, and then expand the logarithm inside the integral in a Taylor series. These steps lead to

$$n_k \sim \begin{cases} k^{-\gamma}\exp\left[-\mu\left(\frac{k^{1-\gamma}-2^{1-\gamma}}{1-\gamma}\right)\right] & \frac{1}{2} < \gamma < 1, \\[2ex] k^{(\mu^2-1)/2}\exp\left[-2\mu\sqrt{k}\right] & \gamma = \frac{1}{2}, \\[2ex] k^{-\gamma}\exp\left[-\mu\,\frac{k^{1-\gamma}}{1-\gamma} + \frac{\mu^2}{2}\,\frac{k^{1-2\gamma}}{1-2\gamma}\right] & \frac{1}{3} < \gamma < \frac{1}{2}, \end{cases} \tag{10.23}$$

etc. 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 Eq. (10.21).

For the strictly linear attachment rate, $A_k = k$, the total event rate is $A = \sum_k A_k N_k = \sum_k kN_k = 2N$. Substituting this value for $A = \mu N$ in Eq. (10.20) and solving the resulting recursion gives the discrete power-law form

$$n_k = \frac{4}{k(k+1)(k+2)} = \frac{4\Gamma(k)}{\Gamma(k+3)} \sim \frac{4}{k^3}. \tag{10.24}$$

The main feature of this result is that there is no natural degree scale. For this reason, such networks have been dubbed *scale free*, and they stand in stark contrast to the delta-function degree distribution of regular lattices and the Poisson degree distribution of the Erdös-Rényi random graph.

A surprising feature of linear preferential attachment is that the exponent of the power-law degree distribution is *non-universal*. The asymptotic evaluation of the product in Eq. (10.21) generally leads to a degree distribution exponent 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 not depend on microscopic model details.

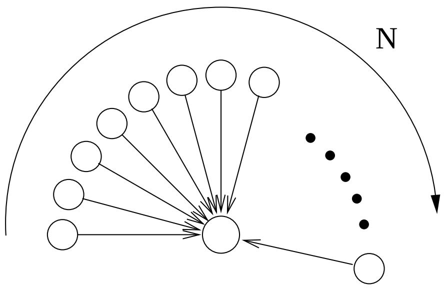


Figure 10.5: Creation of a "bible" for attachment rate $A_k \sim k^\gamma$ with $\gamma > 2$. In the configuration shown each new node attaches only to the bible.

What happens for superlinear attachments rates? Now an analog of gelation occurs in that nearly all the links in the network condense onto a single node, while all other nodes are attached to a small number of links. Especially singular behavior occurs for $\gamma > 2$ because one node there is a non-zero probability that a single node links to *every* node in an infinite network, while only a finite number of links exist between all other nodes. We call such a highly-linked node a "bible". It is easy to see that the probability for a bible to exist is non zero when $\gamma > 2$. Suppose that there is a bible after then network contains $N+1$ nodes (1 bible and N citing nodes). The probability that the next node links to the initial node is then $N^\gamma/(N+N^\gamma)$, and the probability that this connection pattern to continue indefinitely is

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

Evaluating this product by the standard steps of writing the product as the exponential of a sum, approximating the sum as an integral, and expanding the logarithm in the integrand to first order, we find that $\mathcal{P} = 0$ for $\gamma \le 2$ and $\mathcal{P} > 0$ for $\gamma > 2$. Thus for $\gamma > 2$, there a non-zero probability for a bible to exist 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.

Node attractiveness

In many real networked systems, such as the world-wide web, book sales by individuals, scientific publications, *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. The master equation approach easily gives the degree distribution for this natural generalization of preferential attachment.

Each node is assigned an initial “attractiveness” $\eta > 0$ that is chosen from a specified distribution $p_0(\eta)$. The attachment rate for a node with degree k and attractiveness η is defined as $A_k(\eta)$. To characterize how nodes evolve, we need to monitor their degree and keep track of their attractiveness. Thus let $N_k(\eta)$ be the number of nodes of degree k and attractiveness η. The evolution of this joint degree-attractiveness distribution is governed by the master equation

$$\frac{dN_k(\eta)}{dN} = \frac{A_{k-1}(\eta)N_{k-1}(\eta) - A_k(\eta)N_k(\eta)}{A} + p_0(\eta)\delta_{k1}, \tag{10.25}$$

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. (10.19), we substitute $A = \mu N$ and $N_k(\eta) = Nn_k(\eta)$ into Eq. (10.25). The solution of the resulting recursion relation is

$$n_k(\eta) = p_0(\eta)\frac{\mu}{A_k(\eta)} \prod_{j=1}^{k} \left(1 + \frac{\mu}{A_j(\eta)}\right)^{-1}. \tag{10.26}$$

As a simple and generic example, consider the case where $A_k(\eta) = \eta k$, that is, the attachment rate is linear in degree *and* in attractiveness. Applying the same analysis as in the homogeneous network, we obtain the degree distribution

$$n_k(\eta) = \frac{\mu\, p_0(\eta)}{\eta}\, \frac{\Gamma(k)\,\Gamma\left(1+\frac{\mu}{\eta}\right)}{\Gamma\left(k+1+\frac{\mu}{\eta}\right)}\,. \tag{10.27}$$

Thus for nodes with a fixed attractiveness η, the asymptotic degree distribution is the power law $n_k(\eta) \sim k^{-1-\mu/\eta}$. What is perhaps more relevant, however, is the degree distribution averaged over the attractiveness distribution. For this purpose, we need the amplitude μ. We therefore substitute (10.27) 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 that determines μ:

$$1 = \int d\eta\, p_0(\eta)\, \left(\frac{\mu}{\eta} - 1\right)^{-1}. \tag{10.28}$$

The above condition leads to two alternatives: in the pathological case where the support of η is unbounded so that arbitrarily attractive nodes can exist, the integral diverges and there is no solution for μ. In this case, 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 exponent $\nu(\eta) = 1 + \mu/\eta$. However the degree distribution averaged over all attractiveness, $\langle n_k\rangle = \int d\eta\, n_k(\eta)$, 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

$$n_k \sim k^{-(1+\mu/\eta_{\max})}\,(\ln k)^{-\omega}. \tag{10.29}$$

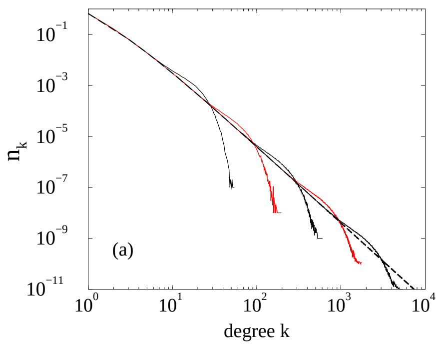
10.3 Finite Networks

Since real networks are necessarily finite, it is wothwhile to ask: what is the role of finiteness on the properties of growing networks? Clearly, finiteness imposes a cutoff on the power-law tail degree distribution (Fig. 10.6), and we wish to quantify this cutoff and related manifestations of finitness. For finite N, the state of a network can be more generally characterized by the set $\mathbf{N} = \{N_1, N_2, \ldots\}$ that specifies the number of nodes N_k of degree k. Each time a new node is introduced into the network, its state $\mathbf{N}$ evolves by:

$$\begin{aligned}(N_1, N_2) &\to (N_1, N_2 + 1),\\ (N_1, N_k, N_{k+1}) &\to (N_1 + 1, N_k - 1, N_{k+1} + 1).\end{aligned}$$

The first process arises when the new node attaches to an existing node of degree 1; in this case, the number of nodes of degree 1 does not change while the number of nodes of degree 2 increases by 1. The second line accounts for the case where the new node attaches to a node of degree $k > 1$. From these processes, it is straightforward, in principle, to write the master equation for the joint probability distribution $P(\mathbf{N})$. However, such an equation would provide much more information than is of practical interest. Here we focus on the degree distribution and fluctuations in the degree distribution. As we shall show, for a finite network the number of nodes of fixed degree, $N_k(N)$, are random variables that become sharply peaked about their average values in the $N \to \infty$ limit.

When the number of nodes N is finite, there will also necessarily be a maximal degree $k_{\rm max}$. Thus predictions for the degree distribution that are implicitly based on an infinite network must eventually break down as k approaches $k_{\rm max}$. The maximal degree may be determined by the extremal criterion $\sum_{k\geq k_{\rm max}} \langle N_k(N)\rangle \approx 1$ which states that there should be one node whose degree is in the range $(k_{\rm max}, \infty)$. This criterion yields $k_{\rm max} \sim N^{1/(\nu-1)}$ when the degree distribution of an infinite network asymptotically decays as $k^{-\nu}$.



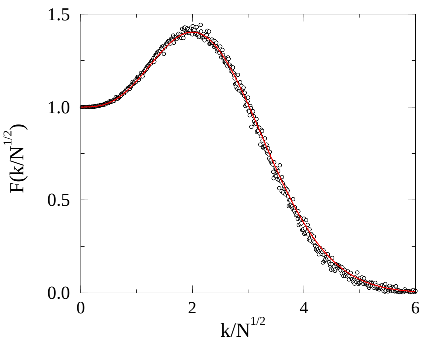


Figure 10.6: (left) Normalized degree distributions for strictly linear preferential attachment networks ($A_k = k$) with $10^2, 10^3, \ldots, 10^6$ links (upper left to lower right). The dashed line is the asymptotic result $n_k = 4/[k(k+1)(k+2)]$. (right) The scaling function $F(\xi)$, with $\xi = k/N^{1/2}$ from Eq. (10.39). The circles give the simulation data of 10^6 realizations of a network with $N = 10^4$ links for the dimer initial condition.

Nodes of Fixed Degree

Degree 1

To appreciate the role of finiteness in the simplest possible setting, consider the number of nodes of degree 1, $N_1(N)$. For the case of strictly linear preferential attachment, $A_k = k$, the average number of such nodes evolves according to

$$\langle N_1(N+1)\rangle = \left\langle N_1(N) \times \frac{N_1(N)}{2N}\right\rangle + \left\langle (N_1(N)+1) \times \left(1 - \frac{N_1}{2N}\right)\right\rangle .$$

The first term on the right accounts for the new node attaching to a node of degree one, an event that occurs with probability $N_1/2N$. In this case, the number of nodes of degree one does not change. The second term accounts for the new node attaching to a node of degree greater than one with probability $(1 - N_1/2N)$. For this event N_1 increases by one. Simplifying, we obtain the recursion

$$\langle N_1(N+1)\rangle = 1 + \left(1 - \frac{1}{2N}\right)\langle N_1(N)\rangle. \tag{10.30}$$

We take the initial condition $\langle N_1(1)\rangle = N_1(1) = 2$. **why?**

To solve this recursion, we multiply (10.30) by Nw^{N-1} and sum over $N \geq 1$ to convert it into the differential equation for the generating function $\mathcal{X}_1(w) = \sum_{N\geq 1}\langle N_1(N)\rangle\, w^{N-1}$,

$$\frac{d\mathcal{X}_1}{dw} = \frac{1}{(1-w)^2} + \frac{1}{2}\,\mathcal{X}_1 + w\,\frac{d\mathcal{X}_1}{dw}\,. \tag{10.31}$$

Solving Eq. (10.31) subject to the initial condition $\mathcal{X}_1(0) = 2$ gives $\mathcal{X}_1(w) = \frac{2}{3}(1-w)^{-2} + \frac{4}{3}(1-w)^{-1/2}$, and expanding this solution in a Taylor series in w leads to

$$\langle N_1(N)\rangle = \frac{2}{3}\,N + \frac{4}{3\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)}\,. \tag{10.32}$$

Thus the fraction of nodes of degree 1 approaches the expected value of 2/3 [see Eq. (10.24)], but with corrections that vanish as $N^{-1/2}$.

Degree Greater Than One

Following the same reasoning as in the case of nodes of degree 1, the number of nodes of degree $k \geq 2$, $N_k(N)$ evolves as

$$\langle N_k(N+1)\rangle = \left\langle N_1(N) \times \frac{N_1(N)}{2N}\right\rangle + \left\langle (N_1(N)+1) \times \left(1 - \frac{N_1}{2N}\right)\right\rangle\,.$$

The first term on the right accounts for the new node attaching to a node of degree k with probability $N_k/2N$, after which the N_k decreases by 1. The second term accounts for the new node attaching to a node of degree $k-1$ with probability $N_{k-1}/2N$), after which N_k increases by 1. The last term accounts for attachment to all other nodes, which leads to no change in N_k. This then gives the recursion

$$\langle N_k(N+1)\rangle = \langle N_k(N)\rangle + \left\langle \frac{(k-1)N_{k-1}(N) - kN_k(N)}{2N}\right\rangle. \tag{10.33}$$

The first term on the right accounts for the new node attaching to a node of degree $k-1$ with probability $N_{k-1}/2N$. The second term accounts for the new node attaching to a node of degree k with probability $N_k/2N$. For this event N_1 increases by one.

Thus recursion can be again solved by the generating function method (problem 10.x). Expanding this generating function in a Taylor series we then obtain $\langle N_k(N)\rangle$, and the explicit results for the first few k are:

$$\begin{aligned}
\langle N_1(N)\rangle &= \frac{2}{3}\,N + \frac{4}{3\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)},\\
\langle N_2(N)\rangle &= \frac{1}{6}\,N + \frac{4}{3\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} - \frac{3}{2}\,\delta_{N,1},\\
\langle N_3(N)\rangle &= \frac{1}{15}\,N + \frac{4}{3\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} - \frac{4}{5\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{3}{2}\right)}{\Gamma(N)} - 3\,\delta_{N,1}
\end{aligned}$$

Thus as N increases the number of nodes N_k with fixed degree k approaches the value obtained for the infinite system, $\frac{4}{k(k+1)(k+2)}$, but with corrections whose leading behavior is proportional to $N^{-1/2}$.

Nodes of Arbitrary Degree

More generally, what is the dependence of N_k on *both* k and on N? The existence of this maximal degree suggests that the degree distribution should be described by the finite-size scaling form

$$\langle N_k(N)\rangle \simeq N n_k F(\xi), \qquad \xi = k/k_{\max}. \tag{10.34}$$

This scaling function has a well-defined peak for $\xi \approx 2$ (Fig. 10.6), and through the use of the generating function method, it is possible to determine the full behavior of the scaling function. To start, we need the dependence of $N_k(N)$ on both N nd k. It is therefore useful to introduce the two-variable generating function

$$\mathcal{N}(w,z) = \sum_{N=1}^{\infty}\sum_{k=1}^{\infty}\langle N_k(N)\rangle\, w^{N-1}\, z^k\,. \tag{10.35}$$

Taking Eq. (10.33), multiplying by $w^{N-1}z^k$ and summing over all N and k, the generating function $\mathcal{N}(w,z)$ satisfies

$$\left(2(1-w)\,\frac{\partial}{\partial w} + z(1-z)\,\frac{\partial}{\partial z} - 2\right)\mathcal{N} = \frac{2z}{(1-w)^2}\,. \tag{10.36}$$

To simplify this equation, we introduce the rotated variables x, y defined by $x+y = -\frac{1}{2}\ln(1-w)$ and $x-y = \ln\frac{z}{1-z}$, to recast Eq. (10.36) into

$$\left(\frac{\partial}{\partial x} - 2\right)\mathcal{N}(x,y) = \frac{2\,e^{5x+4y}}{e^x+e^y}\,, \tag{10.37}$$

whose general solution is

$$\mathcal{N}(x,y) = e^{4x+4y} - 2e^{3x+5y} - e^{2x+2y} + 2e^{2x+4y} + 2\,\frac{e^{2x+2y}}{1+e^{2y}} + 2e^{2x+6y}\ln\left(\frac{e^{x+y}+e^{2y}}{1+e^{2y}}\right). \tag{10.38}$$

We may then extract the scaling function $F(\xi)$ from this generating function, and the final result is

$$F(\xi) = \mathrm{erfc}\left(\frac{\xi}{2}\right) + \frac{2\xi+\xi^3}{\sqrt{4\pi}}\; e^{-\xi^2/4}\,, \tag{10.39}$$

where $\mathrm{erfc}(x)$ is the complementary error function. The most important feature of this result is that the exact average degree distribution has a Gaussian large-degree tail

$$\langle N_k(N)\rangle \to \frac{2}{\sqrt{\pi N}}\; e^{-k^2/4N}\,, \tag{10.40}$$

and moreover, the scaling function in Eq. (10.39) completely describes the finite-size correction to the degree distribution (Fig. 10.6).

Higher Moments and Their Fluctuation

We now turn to higher moments of the degree distribution. While the zeroth and first moments of the degree distribution are simply related to the total number of links for *any* network topology, the higher moments are not so simply characterized, but instead reflect the power-law tail of the degree distribution.

Using the exact expression (10.38) for the generating function, the second moment of the degree distribution is obtained from

$$\left(z^2\,\frac{\partial}{\partial z}\right)^2 \mathcal{N}(w,z)\Big|_{z=1} = \frac{4-2\ln(1-w)}{(1-w)^2}\,.$$

We now expand the right-hand side in a series in w to yield

$$\langle k^2\rangle \equiv \sum_{k=1}^{\infty} k^2\langle N_k\rangle = 2NH_N, = 2N\ln N + 2\gamma N + 1 - \frac{1}{6N} + \dots\,, \tag{10.41}$$

where $H_N = \sum_{1\le j\le N} j^{-1}$ is the harmonic number and $\gamma \approx 0.5772166$ is Euler's constant.

For the third moment we find

$$\langle k^3\rangle = \frac{32}{\sqrt{\pi}}\,\frac{\Gamma\left(N+\frac{3}{2}\right)}{\Gamma(N)} - 6NH_N - 16N\,. \tag{10.42}$$

More generally, the dependence of the moments on N stems from the power-law tail of the degree distribution $\langle N_k\rangle \propto N/k^3$. From this asymptotic distribution, a suitably normalized set of measures for the mean degree

$$\mathcal{M}_n = \left(\frac{\langle k^n\rangle}{\langle k^0\rangle}\right)^{1/n}, \tag{10.43}$$

has the following N dependence:

$$\mathcal{M}_n \propto \begin{cases} \text{const.} & n<2 \\ \ln N & n=2 \\ N^{(n-2)/2} & n>2 \end{cases} \tag{10.44}$$

Sensitivity

To illustrate the crucial role of the initial condition, let's 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 nodes. For linear preferential attachment, $A_k = k$, this probability obeys the master equation

$$P(k,N+1) = \frac{k-1}{2N}\,P(k-1,N) + \frac{2N-k}{2N}\,P(k,N). \tag{10.45}$$

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 (10.45) for the "dimer" initial condition $\bullet\!\!-\!\!\bullet$ is

$$P(k,N) = \frac{1}{2^{2N-k-1}}\,\frac{(2N-k-1)!}{(N-k)!\,(N-1)!} \longrightarrow \frac{1}{\sqrt{\pi N}}\,e^{-k^2/4N}, \tag{10.46}$$

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 largest degree; this fact 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. (10.45), the average degree of the initial node satisfies the recursion relation

$$\langle k\rangle_{N+1} = \langle k\rangle_N \left(1+\frac{1}{2N}\right), \tag{10.47}$$

whose solution is

$$\langle k\rangle_N = \Lambda\,\frac{\Gamma\left(N+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma(N)} \sim \frac{\Lambda}{\sqrt{\pi}}\,N^{1/2}. \tag{10.48}$$

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.

Problems

Section 10.1

1. Solve the master equations for the degree distribution of the ER graph

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

one by one for the initial condition $n_0(t=0) = 1$ and show that the solution is

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

Section 10.2

1. Compute the degree distribution for the uniformly growing tree for a few values of $N = 1, 2, 3, \ldots$ within the (i) continuous and (ii) discrete approach. For the continous formulation, solve the system of differential equations (10.15). For the discrete approach, solve the exact recursion formulae

$$N_k(N+1) - N_k(N) = \frac{N_{k-1}(N) - N_k(N)}{N} + \delta_{k,1}.$$

Compare your results with the asymptotic average degree distribution $N_k(N) \sim N2^{-k}$.

2. Show that the zeroth and first moments of the degree distribution $M_n(N) \equiv \sum_{j\geq 1} j^n N_j(N)$ have the time dependence $M_0(N) = N$ and $M_1(N) = 2N$, independent of the attachment rate A_k.

3. Verify, for attachment rates that do not grow faster than linearly with k, that both the degree distribution $N_k(N)$ and $A(N)$ both grow linearly with time.

4. Show that the limiting behavior of μ in $A(N) = \mu N$ is give by:

$$\begin{aligned} \mu &= 1 + B_0\gamma + \mathcal{O}\left(\gamma^2\right), && \gamma \downarrow 0 \\ \mu &= 2 - B_1(1-\gamma) + \mathcal{O}\left((1-\gamma)^2\right), && \gamma \uparrow 1 \end{aligned}$$

with

$$\begin{aligned} B_0 &= \sum_{j=1}^{\infty} \frac{\ln j}{2^j} = 0.5078\ldots, \\ B_1 &= 4\sum_{j=1}^{\infty} \frac{\ln j}{(j+1)(j+2)} = 2.407\ldots. \end{aligned}$$

Here γ is the exponent in the attachment rate A_k defined by $A_k = k^\gamma$.

5. Determine the degree distribution for the shifted linear attachment rate $A_k = k + \lambda$. First show that $A(N) = \sum_j A_j N_j(N)$ now equals $A(N) = M_1(N) + \lambda M_0(N)$. Using these results in the master equation show that the degree distribution is

$$n_k = (2+\lambda)\, \frac{\Gamma(3+2\lambda)}{\Gamma(1+\lambda)}\, \frac{\Gamma(k+\lambda)}{\Gamma(k+3+2\lambda)}. \tag{10.49}$$

Show that asymptotically, this distribution decays as $k^{-\nu}$, with $\nu = 3 + \lambda = 1 + \frac{1}{r}$.

6. Consider the connection kernel $A_1 = 1$ and $A_k = ak$ for $k \geq 2$. Show that the resulting degree distribution is asymptotically a power law, $n_k \sim k^{-\nu}$, with $\nu = (3 + \sqrt{1 + 8/a_\infty})/2$, which can indeed be tuned to any value larger than 2.

7. Generalize linear preferential attachment networks to the case where each new node links to m pre-existing nodes. Write the master equation for this process, and by applying the same approach as that used for Eq. (10.19), find the the degree distribution

8. Solve the recursion Eq. (10.33) for the number of nodes of degree k, $N_k(N)$ by the generating function method.

   Solution: Define the generating function as $\mathcal{X}_k(w) = \sum_{N=1}^{\infty} \langle N_k(N) \rangle \, w^{N-1}$. We multiple Eq. (10.33) by w^N to convert it to a differential equation that relates $\mathcal{X}_k$ and $\mathcal{X}_{k-1}$. This equation is further simplified by making the transformation

$$\mathcal{X}_k(w) = (1-w)^{\frac{k}{2}-1}\,\mathcal{U}_k(u), \quad u = \frac{1}{\sqrt{1-w}} - 1. \tag{10.50}$$

   The resulting equation is

$$\frac{d\mathcal{U}_k}{du} = (k-1)\,\mathcal{U}_{k-1}, \qquad k \geq 2. \tag{10.51}$$

   Rewriting our previous solution for $\mathcal{X}_1$ as

$$\mathcal{U}_1(u) = \frac{2}{3}\,u^3 + 2u^2 + 2u + 2, \tag{10.52}$$

   one can solve Eqs. (10.51) subject to the initial condition $\mathcal{U}_k(u=0) = 0$ for $k \geq 2$. The final result is

$$\mathcal{U}_k(u) = \frac{4u^{k+2}}{k(k+1)(k+2)} + \frac{4u^{k+1}}{k(k+1)} + \frac{2u^k}{k} + 2u^{k-1}\,.$$

   Using the binomial formula, we transform $\mathcal{X}_k(z)$ into the series

$$\mathcal{X}_k(w) = \frac{4}{k(k+1)(k+2)}\,\frac{1}{(1-w)^2} + \frac{4}{3}\,\frac{1}{(1-w)^{1/2}} + 2\sum_{a=1}^{k-1}(-1)^a\,\frac{a+2}{a+3}\binom{k-1}{a}(1-w)^{(a-1)/2}\,.$$