# A general model of undirected web graphs

Colin Cooper[1] and Alan M. Frieze[2]

[1] School of Mathematical and Computing Sciences, Goldsmiths College, University of London, London SE14 6NW, UK.
[2] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, U.S.A.

**Abstract.** We describe a general model of a random graph whose degree sequence obeys a power law. Such laws have recently been observed in graphs associated with the world wide web.

## 1 Introduction

We describe the evolution of an undirected random (multi-)graph $G(t)$ which is an example of the type of model referred to as a web graph. Such graphs (or digraphs) evolve by the addition of new vertices and/or edges at each step $t$. See eg: [1], [2], [3], [4], [5], [7], [10], [11] for further details.

Initially, at step $t = 0$, there is a single vertex $v_0$. At step $t = 1, 2, ..., T, ...$ either a procedure NEW is followed with probability $1 - \alpha$, or a procedure OLD is followed with probability $\alpha$. In procedure NEW, a new vertex $v$ is added to $G(t-1)$ with one or more edges added between $v$ and $G(t-1)$. In procedure OLD, an existing vertex $v$ is selected and extra edges are added at $v$.

The recipe for adding edges typically permits the choice of initial vertex $v$ (in the case of OLD) and the terminal vertices (in both cases) to be made either u.a.r or according to vertex degree, or a mixture of these two based on further sampling. The number of edges added to vertex $v$ at step $t$ by the procedures (NEW, OLD) is given by distributions specific to the procedure. The details of these choices are given below.

A question arises in our model: Should we regard the edges as directed or undirected in relation to the sampling procedures NEW, OLD? We note that the edges have an intrinsic direction arising from the way they are inserted, which we can ignore or not as we please. In the undirected model, the variable we estimate is $\widehat{d}_k$, the expected proportion of vertices of degree $k$ at step $t$, and we use sampling procedures based on vertex degree. Due to limitations of space we will only consider the case of the undirected model here. For equivalent directed versions of this model see [6].

For the undirected model, we prove that $\widehat{d}_k$ is asymptotically independent of $t$ (provided $t \to \infty$), and provided some degree based sampling occurs, $\widehat{d}_k$ is asymptotic (in $k$) to $Ck^{-x}$ where $x > 1$ is a function of the parameters of the model. The explicit functional form of $x$ is given by (1). If all the vertex sampling is u.a.r then $\widehat{d}_k$ is a geometric series in powers of $k$. This model is well understood and we do not consider it further here.

Our undirected model $G(t)$ has sampling parameters $\alpha, \beta, \gamma, \delta, \boldsymbol{p}, \boldsymbol{q}$ whose meaning is given below:

Choice of procedure at step $t$.

$\alpha$: Probability that an OLD node generates edges.

$1-\alpha$: Probability that a NEW node is created.

Procedure NEW

$\boldsymbol{p} = (p_i : i \geq 1)$: Probability that new node generates $i$ new edges.

$\beta$: Probability that choices of terminal vertices are made uniformly.

$1-\beta$: Probability for choice of terminal vertices according to degree.

Procedure OLD

$\boldsymbol{q} = (q_i : i \geq 1)$: Probability that old node generates $i$ new edges.

$\delta$: Probability that the initial node is selected uniformly.

$1-\delta$: Probability that the initial node is selected according to degree.

$\gamma$: Probability that choices of terminal vertices are made uniformly.

$1-\gamma$: Probability for choice of terminal vertices according to degree.

The models we study here require $\alpha < 1$, always. We also assume a *finiteness* condition for the distributions $\{p_j\}, \ \{q_j\}$. This means that there exists $j_0, j_1$ such that $p_j = 0, \ j \geq j_0$ and $q_j = 0, \ j > j_1$.

The model creates edges in the following way: An initial vertex $v$ is selected. If the terminal vertex $w$ is chosen u.a.r, we regard the newly inserted edge $vw$ as directed $(v, w)$ and say $v$ is *assigned* to $w$. If the terminal vertex $w$ is chosen according to its vertex degree, we say $v$ is *copied* to $w$.

The papers [10], [11] introduce an alternative copying model in which a new vertex $v$ chooses an old vertex $w$ and selects (copies) a randomly chosen set of out-neighbours of $w$ to be its own out-neighbours. This construction leads to a larger number of small complete bipartite subgraphs than would be obtained by purely random selection of endpoints. This suggests an explanation the occurrence of the number of small complete bipartite subgraphs found in trawls of the web.

Since our focus is on the degree sequence and since the construction above does not lead to a fundamentally different recurrence (see [6] for details) we will not explicitly use this method of generating edges. We will continue to use the term copy to mean that vertices are chosen with probability proportional to degree.

Let $\mu_p = \sum_{j=0}^{j_0} j p_j$, $\mu_q = \sum_{j=0}^{j_1} i q_j$ and let $\theta = 2((1-\alpha)\mu_p + \alpha\mu_q)$. To simplify subsequent notation, we transform the parameters as follows:

$$
\begin{aligned}
a &= 1 + \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha} + \frac{\alpha\delta}{1-\alpha}, \\
b &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta} + \frac{\alpha(1-\delta)}{\theta}, \\
c &= \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha}, \\
d &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta}, \\
e &= \frac{\alpha\delta}{1-\alpha}, \qquad\qquad f = \frac{\alpha(1-\delta)}{\theta}.
\end{aligned}
$$

Note that $c+e=a-1$ and $b=d+f$.
The main quantity we study is the random variable $D_k(t)$, the number of vertices of degree $k$ at step $t$. We let $\overline{D}_k(t)=\mathbf{E}(D_k(t))$ and then $\widehat{d}_k=t^{-1}\mathbf{E}(D_k(t))$.

**Theorem 1.** *Assume that $t\to\infty$. Then provided $k=o(\log t/\log\log t)$, we have that with probability 1-o(1),*

**(i)** $Ak^{-\gamma}\leq\widehat{d}_k\leq Bk^{-\gamma/j_1}$ *where* $\gamma=(1+d+f\mu_q)/(d+f)$.
**(ii)** *If* $j_1=1$ *then* $\widehat{d}_k\sim Ck^{-(1+1/(d+f))}$.
**(iii)** *If* $f=0$ *then* $\widehat{d}_k\sim Ck^{-(1+1/d)}$.
**(iv)** *If the* SOLUTION CONDITIONS *hold then*

$$\widehat{d}_k=C\left(1+O\left(\frac{1}{k}\right)\right)k^{-x},$$

*where $C$ is constant and*

$$x=1+\frac{1}{d+f\mu_q}. \tag{1}$$

We say that $\{q_j:j=1,...,j_1\}$ is *periodic* if there exists $m>1$ such that $q_j=0$ unless $j\in\{m,2m,3m,\dots\}$.
Let

$$\phi_1(y)=y^{j_1}-\left(\frac{d+q_1f}{b}y^{j_1-1}+\frac{q_2f}{b}y^{j_1-2}+\cdots+\frac{q_{j_1}f}{b}\right).$$

Our SOLUTION CONDITIONS are:

**S(i)** $f>0$ and either (a) $d+q_1f>0$ or (b) $\{q_j\}$ is not periodic.
**S(ii)** The polynomial $\phi_1(y)$ has no repeated roots.

We can also prove the following concentration result (see [6] for details).

**Theorem 2.** *For any $u>0$,*

$$\mathbf{Pr}(|D_k(t)-\overline{D}_k(t)|\geq u)\leq\exp\left\{-\frac{2u^2}{tj_{\max}^2}\right\}$$

*where* $j_{\max}=\max\{j_0,j_1\}$.

## 2 Evolution of the degree sequence the graph $G(t)$

Let $\nu(t)=|V(t)|$ be the number of vertices and let $\eta(t)=|2E(t)|$ be the total degree of the graph at the end of step $t$. $\mathbf{E}\nu(t)=(1-\alpha)t$ and $\mathbf{E}\eta(t)=\theta t$. The random variables $\nu(t)$, $\eta(t)$ are sharply concentrated provided $t\to\infty$. Indeed $\nu(t)$ has binomial distribution $B(t,1-\alpha)$ and so by the Chernoff bounds,

$$\mathbf{Pr}(|\nu(t)-(1-\alpha)t|\geq t^{1/2}\log t)=O(t^{-K}) \tag{2}$$

for any constant $K > 0$.
Similarly, $\eta(t)$ has expectation $\theta t$ and is the sum of $t$ independent random variables, each bounded by $\max\{j_0, j_1\}$. Hence, by Hoeffding's theorem [8],

$$\mathbf{Pr}(|\eta(t) - \theta t| \geq t^{1/2}\log t) = O(t^{-K}) \tag{3}$$

for any constant $K > 0$.
$D_k(t)$ is produced by $t$ random choices and changing any one of them will change $D_k(t)$ by at most $\max\{j_0, j_1\} + 1$ and so applying the Azuma-Hoeffding martingale tail inequality, see for example Janson, Łuczak and Ruciński [9] we see that for $k \geq 0$,

$$\mathbf{Pr}(|D_k(t) - \overline{D}_k(t)| \geq t^{1/2}\log t) = O(t^{-K}) \tag{4}$$

for any constant $K > 0$.
Our main focus here is on giving asymptotic expressions for $\overline{D}_k(t)$, where $k \leq \epsilon \log t / \log\log t$ for $\epsilon$ sufficiently small.
Here $\overline{D}_0(t) = 0$ for all $t$, $\overline{D}_1(0) = 1,\ \overline{D}_k(0) = 0,\ k \geq 2$. Write $\widehat{D}_l = \overline{D}_l(t-1)$. Then, after using (2), (3) and (4) we see that

$$\overline{D}_k(t) = \overline{D}_k(t-1) + (1-\alpha)p_k + O(t^{-1/2}\log t) \tag{5}$$

$$+ (1-\alpha)\sum_{j=1}^{j_0} p_j \left(\frac{\beta j \widehat{D}_{k-1}}{(1-\alpha)t} - \frac{\beta j \widehat{D}_k}{(1-\alpha)t} + (1-\beta)\left(\frac{j(k-1)\widehat{D}_{k-1}}{\theta t} - \frac{jk\widehat{D}_k}{\theta t}\right)\right) \tag{6}$$

$$- \alpha\left(\frac{\delta \widehat{D}_k}{(1-\alpha)t} + \frac{(1-\delta)k\widehat{D}_k}{\theta t}\right) + \alpha\sum_{j=1}^{k-1} q_j \left(\frac{\delta \widehat{D}_{k-j}}{(1-\alpha)t} + \frac{(1-\delta)(k-j)\widehat{D}_{k-j}}{\theta t}\right) \tag{7}$$

$$+ \alpha\sum_{j=1}^{j_1} q_j \left(\frac{\gamma j \widehat{D}_{k-1}}{(1-\alpha)t} - \frac{\gamma j \widehat{D}_k}{(1-\alpha)t} + (1-\gamma)\left(\frac{j(k-1)\widehat{D}_{k-1}}{\theta t} - \frac{jk\widehat{D}_k}{\theta t}\right)\right). \tag{8}$$

Here (6), (7), (8) are (respectively) the main terms of the change in the expected number of vertices of degree $k$ due to the effect on: terminal vertices in NEW, the initial vertex in OLD and the terminal vertices in OLD. Rearranging the right hand side, we find:

$$\begin{aligned}
\overline{D}_k(t) &= \widehat{D}_k + (1-\alpha)p_k + O(t^{-1/2}\log t) \\
&- \frac{\widehat{D}_k}{t}\left(\beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha} + \frac{\alpha\delta}{1-\alpha} + \frac{(1-\alpha)(1-\beta)\mu_p k}{\theta} + \frac{\alpha(1-\gamma)\mu_q k}{\theta} + \frac{\alpha(1-\delta)k}{\theta}\right) \\
&+ \frac{\widehat{D}_{k-1}}{t}\left(\beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha} + \frac{(1-\alpha)(1-\beta)\mu_p(k-1)}{\theta} + \frac{\alpha(1-\gamma)\mu_q(k-1)}{\theta}\right) \\
&+ \sum_{j=1}^{j_1-1} q_j \frac{\widehat{D}_{k-j}}{t}\left(\frac{\alpha\delta}{1-\alpha} + \frac{\alpha(1-\delta)(k-j)}{\theta}\right).
\end{aligned}$$

Thus

$$\overline{D}_k(t) = \widehat{D}_k + (1-\alpha)p_k + O(t^{-1/2}\log t)$$
$$+\frac{1}{t}\left((1-(a+bk))\widehat{D}_k + (c+d(k-1))\widehat{D}_{k-1} + \sum_{j=1}^{j_1-1} q_j(e+f(k-j))\widehat{D}_{k-j}\right). \tag{9}$$

Now define the sequence $(d_0, d_1, ..., d_k, ...)$ by $d_0 = 0$ and for $k \geq 1$

$$d_k(a+bk) = (1-\alpha)p_k + (c+d(k-1))d_{k-1} + \sum_{j=1}^{k-1}(e+f(k-j))q_j d_{k-j}. \tag{10}$$

Since $a > 0$, this system of equations clearly has a unique solution.
We now prove that for small $k$, $\overline{D}_k(t) \approx d_k t$ as $t \to \infty$.

**Lemma 1.** *There exists an absolute constant $M > 0$ such that for $k = 1, 2, \dots,$*

$$|\overline{D}_k(t) - td_k| \leq M^k k! t^{1/2} \log t.$$

*Proof.* Let $\Delta_k(t) = \overline{D}_k(t) - td_k$. It follows from (9) and (10) that

$$\Delta_k(t) = \Delta_k(t-1)\left(1 - \frac{a+bk-1}{t}\right) + O(t^{-1/2}\log t) +$$
$$t^{-1}\left((c+d(k-1))\Delta_{k-1}(t-1) + \sum_{j=1}^{k-1}(e+f(k-j))q_j\Delta_{k-j}(t-1)\right). \tag{11}$$

Let $L$ denote the hidden constant in $O(t^{-1/2}\log t)$ and assume inductively that $\Delta_\kappa(\tau) \leq M^\kappa \kappa! \tau^{1/2} \log \tau$ for $\tau < t$ and $\kappa < \omega(\tau)$. Also assume that $t, k$ are sufficiently large (we can adjust $M$ to deal with small values of $t, k$). Then (11) implies that for $M$ large,

$$|\Delta_k(t)| \leq M^k k!(t-1)^{1/2}\log t + Lt^{-1/2}\log t$$
$$+\frac{\log t}{t^{1/2}}\left((c+d(k-1))M^{k-1}(k-1)! + \sum_{j=1}^{k-1}(e+f(k-j))q_j M^{k-j}(k-j)!\right)$$
$$\leq M^k k! t^{1/2}\log t\left(1-\frac{1}{t}\right) + Lt^{-1/2}\log t + 2M^{k-1}k!t^{-1/2}\log t$$
$$\leq M^k k! t^{1/2}\log t$$

This completes the induction. □

**Analysis of (10)**

Re-writing (10) we see that for $k \geq j_0$, $d_k$ satisfies

$$d_k = d_{k-1}\frac{c + d(k-1)}{a + bk} + \sum_{j=1}^{j_1} d_{k-j} q_j \frac{e + f(k-j)}{a + bk}, \tag{12}$$

which is a linear difference equation with rational coefficients [12]. The general solution for $d_k$ is a power law, i.e. there are constants $x_L, x_U$ such that $Ak^{-x_L} \leq d_k \leq Bk^{-x_U}$. This is established in Lemma 2.
In the cases where $j_1 = 1$ (a new vertex generates a single edge) or $f = 0$ (old initial vertices are chosen u.a.r) a direct solution to (12) can easily be found. In general however, when $d > 0$ or $d = 0$ and $\{q_j\}$ is non-periodic, we use classical results on the solution of Laplace's difference equation, (of which (10) is an example) given in [12].

**A general power law bound for $d_k$**

**Lemma 2.** *Let $p_j = 0, j \geq j_0$ and $q_j = 0, j > j_1$.*

**(i)** *For all $k \geq j_0$, $d_k > 0$.*
**(ii)** *then, for $k \geq j_0 + j_1$, there exist constants $C, D > 0$ such that*

$$Ck^{-(1+d+f\mu_q)/b} \leq d_k \leq Dk^{-(1+d+f\mu_q)/bj_1}.$$

*Proof.* Let $I$ be the first index such that $p_I > 0$, so that, from (10), $d_I > 0$. As it is impossible for both $c$ and $d$ to be zero, the coefficient of $d_{k-1}$ in (10) is non-zero and thus $d_k > 0$ for $k \geq I$.
For $k \geq j_0$ the recurrence (10) satisfies (12), that is

$$d_k = d_{k-1}\frac{c + d(k-1)}{a + bk} + \sum_{j=1}^{j_1} d_{k-j} q_j \frac{e + f(k-j)}{a + bk}.$$

Let $y = 1 + d + f\mu_q$, then

$$\frac{c + d(k-1)}{a + bk} + \sum_{j=1}^{j_1} q_j \frac{e + f(k-j)}{a + bk} = 1 - y/(a + bk) \geq 0$$

and thus

$$\left(1 - \frac{y}{a + bk}\right) \min\{d_{k-1}, ..., d_{k-j_1}\} \leq d_k \leq \left(1 - \frac{y}{a + bk}\right) \max\{d_{k-1}, ..., d_{k-j_1}\}. \tag{13}$$

It follows that

$$d_{j_0} \prod_{j=j_0}^{k} \left(1 - \frac{y}{a + bj}\right) \leq d_k \leq \prod_{s=0}^{\lfloor (k-(j_0+j_1))/j_1 \rfloor} \left(1 - \frac{y}{a + b(k - sj_1)}\right). \tag{14}$$

The LHS is clear. For the RHS note that $d_k \leq 1$ (as can be seen by using induction and the upper bound in (13)). When iterating $d_j$ backwards on the RHS, we must make at least $\lfloor (k-(j_0+j_1))/l \rfloor$ iterations, and at least one value of $j$ falls in each interval $[k-(s+1)l, k-sj_1)$. For that value of $j$ we upper bound $(1-y/(a+bj))$ by $(1-y/(a+b(k-sj_1))$.
Now consider the product in the LHS of (14).

$$\log\left(\prod_{j=j_0}^{k}\left(1-\frac{y}{a+bj}\right)\right) = \sum_{j=j_0}^{k}\left(-\frac{y}{a+bj}-\frac{1}{2}\left(\frac{y}{a+bj}\right)^2-\cdots\right)$$
$$= O(1) - \sum_{j=j_0}^{k}\frac{y}{a+bj}.$$

This justifies the lower bound of the lemma and the upper bound follows similarly for the upper bound of (14). □

### The case $j_1 = 1$

When $q_1 = 1, p_j = 0, j \geq j_0 = \Theta(1)$ the general value of $d_k$, $k \geq j_0$ can be found directly, by iterating the recurrence (10). Thus

$$\begin{aligned} d_k &= \frac{1}{a+bk}\left(d_{k-1}\left((a-1)+b(k-1)\right)\right) \\ &= d_{k-1}\left(1-\frac{1+b}{a+bk}\right) \\ &= d_{j_0}\prod_{j=j_0}^{k}\left(1-\frac{1+b}{a+jb}\right). \end{aligned}$$

Thus, for some constant $C$,

$$d_k \sim C(a+bk)^{-x}$$

where

$$x = 1+\frac{1}{b} = 1+\frac{2}{\alpha(1-\delta)+(1-\alpha)(1-\beta)+\alpha(1-\gamma)}.$$

### The case $f = 0$

The case $(f=0)$ arises in two ways. Firstly if $\alpha = 0$ so that a new vertex is added at each step. Secondly, if $\alpha \neq 0$ but $\delta = 1$ so that the initial vertex of an OLD choice is sampled u.a.r.
We first prove that for a sufficiently large absolute constant $A > 0$ and for all sufficiently large $k$, that

$$\frac{d_k}{d_{k-1}} = 1 - \frac{1+d}{a+dk} + \frac{\xi(k)}{k^2} \tag{15}$$

where $|\xi(k)| \leq A$.
We use induction and re-write (10) as

$$\frac{d_k}{d_{k-1}} = \frac{c + d(k-1)}{a + dk} + \sum_{j=1}^{j_1} \frac{e}{a + dk} \prod_{t=k-j}^{k-2} \frac{d_t}{d_{t-1}}. \tag{16}$$

Now use induction to write

$$\prod_{t=k-j}^{k-2} \frac{d_t}{d_{t-1}} = 1 - \frac{j-1}{a + dk} + \frac{\xi'(j,k)}{k^2} \tag{17}$$

where $|\xi'(j,k)| \leq Aj$.
Substituting (17) into (16) gives

$$\frac{d_k}{d_{k-1}} = \frac{c + d(k-1)}{a + dk} + \frac{e}{a + dk} - \frac{e\mu_q(d+1)}{(a + dk)^2} + \frac{\xi''(k)}{(a + dk)k^2}$$

where $|\xi''(k)| \leq 2A\mu_q$.
Equation (15) follows immediately and on iterating this we see that

$$d_k \sim Ck^{-\left(1+\frac{1}{d}\right)}.$$

## 3 Analysis of the general undirected model

### Linear difference equations with rational coefficients: The method of Laplace

This section summarizes Chapter XV (pages 478-503) of *The Calculus of Finite Differences* by I. M. Milne-Thomson [12].
The equation (12) is an example of a linear difference equation with rational coefficients. It can equivalently be written as,

$$d_k(a + bk) - d_{k-1}(c + d(k-1)) - \sum_{j=1}^{k-1} d_{k-j} q_j (e + f(k-j)) = 0. \tag{18}$$

Laplace's difference equation is the name given to the equation whose coefficients are linear functions of a real variable $w$ and an integer $l$. The general form of the homogeneous equation is

$$\sum_{j=0}^{l} [A_{l-j}(w + l - j) + B_{l-j}] u(w + l - j) = 0. \tag{19}$$

Thus (18) is a special case of (19) with $l = j_1$, $w = k - j_1$.
A method of solving difference equations with rational coefficients in general, and equation (19) in particular is to use the substitution

$$u(w) = \oint_C t^{w-1} v(t) dt.$$

The function $v(t)$ is obtained as the solution of the differential equation (22), given below, and $C$ is a suitable contour of integration.
Let

$$\phi_1(t) = A_l t^l + A_{l-1} t^{l-1} + \cdots + A_1 t + A_0 \tag{20}$$
$$\phi_0(t) = B_l t^l + B_{l-1} t^{l-1} + \cdots + B_1 t + B_0, \tag{21}$$

where $\phi_1(t)$ is the *characteristic equation*. The differential equation referred to, is

$$t\phi_1(t)\frac{dv(t)}{dt} - \phi_0(t)v(t) = 0. \tag{22}$$

The general method of solution requires (19) to be of the *Normal type*, namely:

**N(i)** Both $A_l$ and $A_0$ are non-zero.
**N(ii)** The differential equation (22) satisfied by $v(t)$ is of the Fuchsian type.

Let the roots of the characteristic equation be $a_1, ..., a_l$ (with repetition). The condition that $v(t)$ is of the Fuchsian type, requires that $\phi_0(t)/\phi_1(t)$ can be expressed as a convergent power series of $t$ for some $t > 0$. Thus either the roots $a_1, ..., a_l$ of the characteristic equation must be distinct, or if $a$ is repeated $\nu$ times, then $a$ is a root of $\phi_0(t)$ at least $\nu - 1$ times.
Assuming the roots are distinct,

$$\begin{aligned} \frac{v'(t)}{v(t)} &= \frac{\phi_0(t)}{t\phi_1(t)} \\ &= \frac{-\alpha_0}{t} + \frac{\beta_1}{t - a_1} + \cdots + \frac{\beta_l}{t - a_l}, \end{aligned} \tag{23}$$

and $\phi_0(t)/\phi_1(t)$ has the required series expansion. The general solution is

$$v^*(t) = t^{-\alpha_0}(t - a_1)^{\beta_1}...(t - a_l)^{\beta_l}.$$

As long as there are no repeated roots, a system of fundamental solutions $(u_j(w),\ j = 1, ..., l)$ is given by

$$u_j(w) = \frac{1}{2\pi i}\oint_{C_j} t^{w-1-\alpha_0}(t - a_1)^{\beta_1}\cdots(t - a_l)^{\beta_l}dt,$$

where $C_j$ is a contour containing 0 and $a_j$ but excluding the other roots. If $\beta_j$ is integer the contour integral is replaced by the integral from 0 to $a_j$.
A specific solution for $u_j(w)$, valid for $\Re(w) > \alpha_0$, can be obtained as

$$u_j(w) = (a_j)^w \sum_{m=0}^{\infty} C_m B\left(\frac{w - \alpha_0 + \theta - 1}{\theta}, \beta_j + m + 1\right)$$

where $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$.

The variable $\theta > 1$ measures the angular separation, about the origin, of the root $a_j$ from the other roots in the transformation $a_j z^{1/\theta} = t$ used to expand the transformed integral about $z = 1$ and obtain the above solution.
Now using the fact that $\Gamma(x) \sim \sqrt{2\pi} e^{-x} x^{x-1/2}$, as $w \to \infty$,

$$u_j(w) \sim C_j a_j^w w^{-(1+\beta_j)}(1 + O(1/w)). \tag{24}$$

**Application of the technique**

For convenience we let $l = j_1$ for the rest of this section.
Considering the equation (18) we see that

$$\phi_1(y) = y^l - \left( \frac{d + q_1 f}{b} y^{l-1} + \frac{q_2 f}{b} y^{l-2} + \cdots + \frac{q_l f}{b} \right)$$
$$\phi_0(y) = \frac{a}{b} y^l - \left( \frac{c + q_1 e}{b} y^{l-1} + \frac{q_2 e}{b} y^{l-2} + \cdots + \frac{q_l e}{b} \right).$$

We assume that $f > 0$ so that N(i) is satisfied. Let the roots of the characteristic equation be ordered in decreasing size so that $|a_1| \geq |a_2| \geq \cdots \geq |a_l|$. Because of the SOLUTION CONDITIONS we see from Lemma 3, given below, that

$$a_1 = 1$$

and all other roots are either negative or complex and satisfy $|a| < 1$. Considering the partial fraction expansion (23) we see that

$$\phi_0(0) = -\alpha_0 \phi_1(0),$$

so that $\alpha_0 = -e/f$. Also

$$\phi_0(1) = \beta_1 \psi(1),$$

where $\psi(y) = \phi_1(y)/(y-1)$ is given by

$$\psi(z) = z^{l-1} + (1 - \alpha_1) z^{l-2} + (1 - \alpha_1 - \alpha_2) z^{l-3} + \cdots + (1 - \alpha_1 - \cdots - \alpha_{l-2}) z + (1 - \alpha_1 - \cdots - \alpha_{l-1}), \tag{25}$$

and where

$$\alpha_1 = \frac{d + q_1 f}{b}, \qquad \alpha_2 = \frac{q_2 f}{b}, \ldots, \alpha_l = \frac{q_l f}{b}.$$

Now $\phi_0(1) = 1/b$ as $c + d = a - 1$ and $\psi(1) = (d + f\mu_q)/b$. Thus $\beta_1 = 1/(d + f\mu_q)$. The other $\beta_j$ require detailed knowledge of the roots of $\phi_1(t)$ and are not relevant to the asymptotic solution.
The solutions $u_j(w)$ are valid for $\Re(w) > \alpha_0 = -e/f$ which includes all $k \geq 0$. Thus considering the root $a_1 = 1$ we see that

$$u_1(k) = C k^{-(1+\beta_1)} \left( 1 + O\left( \frac{1}{k} \right) \right)$$

where $\beta_1 = \phi_0(1)/\psi(1) = \frac{1}{d+f\mu_q}$, and $1+\beta_1$ is the parameter $x$ of our degree sequence.
For $j \geq 2$, we use (24), giving

$$u_j(k) \to (a_j)^k k^{-(1+\beta_j)} \to 0,$$

faster than $o(1/k)$, if $|a_j| < 1$.
The specific solution for the sequence $(d_1, d_2, ..., d_k, ...)$ is

$$d_k = b_1 u_1(k) + \cdots + b_l u_l(k),$$

where $u_1(w), ..., u_l(w)$ are the fundamental solutions corresponding to the roots $a_1, ..., a_l$. We note that $b_1 \neq 0$. Indeed from Lemma 2, we know $d_k$ obeys a power law, whereas if $b_1 = 0$, then $d_k$ would decay exponentially as $|a_2|^k$.
Thus the error in the approximation of $d_k$ is $O(1/k)$ from the non-asymptotic expansion of $u_1(w)$, and we conclude

$$d_k = Ck^{-\left(1+\frac{1}{d+f\mu_q}\right)}\left(1+O\left(\tfrac{1}{k}\right)\right).$$

In the case where $\phi_1(t)$ has other solutions $|a_j| = 1,\ j = 2, ..., j', j' \leq l$, then the asymptotic solution $d_k$ will be a linear combination of $k$-th powers of these roots.

### Roots of the characteristic equation

**Lemma 3.** *Let $\alpha_1 = (d+q_1 f)/(d+f)$ and for $2 \leq j \leq l$, let $\alpha_j = q_j f/(d+f)$, and let*

$$\phi_1(z) = z^l - \alpha_1 z^{l-1} - \alpha_2 z^{l-2} - \cdots - \alpha_l.$$

*Provided $\alpha_1 > 0$ or $\{q_j\}$ is not periodic, then the solutions of $\phi_1(z) = 0$ are*

**i)** *An un-repeated root at $z = 1$,*
**ii)** *$l-1$ other (possibly repeated) roots $\lambda$ satisfying $|\lambda| < 1$.*

*Proof.* We note the following (see Pólya & Szegő [13] p106 16,17). A polynomial $f(z)$ of the form

$$f(z) = z^n - p_1 z^{n-1} - p_2 z^{n-2} - \cdots - p_{n-1} z - p_n,$$

where $p_i \geq 0,\ i = 1, ..., n$ and $p_1 + \cdots + p_n > 0$ has just one positive zero $\zeta$. All other zeroes $z_0$ of $f(z)$ satisfy $|z_0| \leq \zeta$.
Now $\alpha_i \geq 0$ and $\sum \alpha_i = 1$, and so $\phi_1(1) = 0$ and all other zeros, $z_0$, of $\phi_1(z)$ satisfy $|z_0| \leq 1$.
Let $\psi(z) = \phi_1(z)/(z-1)$ be as in (25). Now $\psi(1)$ is given by

$$1 + (1-\alpha_1) + (1-\alpha_1-\alpha_2) + \cdots + (1-\alpha_1-\cdots-\alpha_{l-1}) = \frac{d+f\mu_q}{d+f}, \quad (26)$$

and thus $\psi(1) \neq 0$, so that $z = 1$ is not a repeated root of $\phi_1$.
Let $z$ satisfy $\phi_1(z) = 0$, $|z| = 1$, $z \neq 1$, and let $w = 1/z$; then $\phi_1(z) = 0$ is equivalent to $h(w) = 1$, where

$$h(w) = \alpha_1 w + \alpha_2 w^2 + \cdots + \alpha_l w^l.$$

Suppose there exists $w \neq 1$, on the unit circle satisfying $h(w) = 1$. Let $T = \{w, w^2, ..., w^l\}$ then all elements of $T$ are points on the unit circle. As $w \neq 1$, $\Re(w) < 1$ and $\Re(w^j) \leq 1$, $j = 2, ..., l$.
Now, by S(i), either $\alpha_1 > 0$ or $\alpha_1 = 0$ but $\{q_j\}$ is not periodic.
If $\alpha_1 > 0$, then

$$\sum \alpha_j \Re(w^j) \leq \alpha_1 \Re(w) + \alpha_2 + \cdots \alpha_l < 1,$$

and the conclusion, that $h(w) \neq 1$ follows.
Suppose $\alpha_1 = 0$. If $1 \notin T$, then the real part of $w^j$ satisfies $\Re(w^j) < 1$, contradicting $h(w) = 1$. If $1 \in T$ then $w = e^{2\pi i/m}$ for some integer $m > 1$. However, as $\{q_j\}$ is not periodic, the conclusion that $h(w) < 1$ follows as before. $\square$

## References


1. M. Adler and M. Mitzenmacher, *Toward Compressing Web Graphs*, To appear in the 2001 Data Compression Conference.
2. R. Albert, A. Barabasi and H. Jeong. *Diameter of the world wide web.* Nature 401:103-131 (1999) see also `http://xxx.lanl.gov/abs/cond-mat/9907038`
3. B. Bollobás, O. Riordan and J. Spencer, *The degree sequence of a scale free random graph process*, to appear.
4. B. Bollobás and O. Riordan, *The diameter of a scale free random graph*, to appear.
5. A. Broder, R. Kumar, F.Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener. *Graph structure in the web.* `http://gatekeeper.dec.com/pub/DEC/SRC/publications/stata/www9.htm`
6. C. Cooper, A. M. Frieze. *General models of web graphs.* Submitted to Random Structures and Algorithms (2001).
7. E. Drinea, M. Enachescu and M. Mitzenmacher, *Variations on random graph models for the web.*
8. W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Jornal of the American Statistical Association.
9. S.Janson, T.Łuczak and A.Ruciński, *Random Graphs*, John Wiley and Sons, 2000.
10. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal. *The web as a graph.* `www.almaden.ibm.com`
11. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal. *Stochastic models for the web graph.* `www.almaden.ibm.com`
12. I. M. Milne-Thomson. *The Calculus of Finite Differences.* Macmillian, London (1951).
13. G. Pólya and G. Szegő. *Problems and Theorems in Analysis I.* Springer, Berlin (1970).