Discrete scale-free distributions and associated limit theorems

2004 J. Phys. A: Math. Gen. 37 L635

(http://iopscience.iop.org/0305-4470/37/48/L01)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **37** (2004) L635–L642 PII: S0305-4470(04)86534-X

LETTER TO THE EDITOR

Discrete scale-free distributions and associated limit theorems

K I Hopcraft, E Jakeman and J O Matthews

Applied Mathematics Division, School of Mathematical Science, University of Nottingham, Nottingham NG7 2RD, UK

Received 21 September 2004, in final form 22 October 2004
Published 17 November 2004
Online at stacks.iop.org/JPhysA/37/L635
doi:10.1088/0305-4470/37/48/L01

Abstract

Consideration is given to the convergence properties of sums of identical, independently distributed random variables drawn from a class of discrete distributions with power-law tails, which are relevant to scale-free networks. Different limiting distributions, and rates of convergence to these limits, are identified and depend on the index of the tail. For indices $\geqslant 2$, the topology evolves to a random Poisson network, but the rate of convergence can be extraordinarily slow and unlikely to be yet evident for the current size of the WWW for example. It is shown that treating discrete scale-free behaviour with continuum or mean-field approximations can lead to incorrect results.

PACS numbers: 02.50.Fz, 05.45.Df, 05.56.+b, 89.75.Fb

Scale-free random networks are an important genre of complex behaviour: by which is meant correlated, nonlinear, collective phenomena, in large multi-component systems. A measure of the topological connectivity of a network is the 'order distribution', e.g. [1]. This is a discrete probability distribution for the number of vertices or links into, or out of, nodes comprising the system. Diverse naturally occurring, man-made and social networks have order distributions with power-law asymptotes, giving rise to the notion of discrete, in contrast to continuous, random self-similar behaviour. This letter addresses this increasingly important aspect of complexity by first identifying and discussing the properties of the discrete analogue of the Lévy-stable densities. However, the discrete-stable class does not include the majority of naturally occurring networks. This prompts analysing a class of distributions that can model such networks, and investigating how their properties differ from the discrete-stable class. Properties of the discrete distributions behave differently from their continuous analogues and, as a consequence, formulations of network behaviour founded on a continuum or 'mean field' are inappropriate in specified regimes.

Aspects of complexity are frequently described by fluctuations with self-similar characteristics, for which the class of Lévy-stable densities [2] can provide a model. The epithet 'stable' derives from the property that sums of independent, identically distributed (iid) random variables are distributed like their summands to within a scaling factor. The best-known example is the sum of N iid Gaussian random variables of zero mean and variance σ^2, which also is exactly a Gaussian variable with variance $N\sigma^2$. But the Gaussian is also an attractor for sums of random variables that have finite variance—this being the central limit theorem of classical statistics [3]. The Lévy-stable class is broader however, with other members sharing the property that their probability densities have power-law tails with index falling within the range -3 to -1, so that the variance does not exist. These distributions are the attractor for fluctuations with infinite variance [4].

The discrete-stable distributions are defined through their generating function [5–7]:

$$q(s) = \sum_{n=0}^{\infty} (1-s)^n P(n) = \exp(-As^\nu)$$

where $P(n)$ is the probability distribution for occurrences of the random integer variable $n \geqslant 0$, A is a positive, real constant which acts as scale factor and $0 < \nu \leqslant 1$. The probability distributions are derived from the generating function by differentiation:

$$P(n) = \frac{(-1)^n}{n!} \left.\frac{\mathrm{d}^n q(s)}{\mathrm{d}s^n}\right|_{s=1}$$

and if $n \gg 1$, $P(n) \sim 1/n^{1+\nu}$ provided that $0 < \nu < 1$, giving power-law asymptote with index in the range $-2 < -(1+\nu) < -1$; hence the mean of these distributions does not exist. The stability property results from the fact that the generating function for a sum of N iid random variables is $q(s)^N$, which for the stable law gives the same distribution with scaling factor $A' \to N^{1/\nu}A$. The case $\nu = 1$ corresponds to the Poisson distribution, where the scaling factor can then be interpreted as the mean. The Poisson distribution therefore has the same significance for discrete-stable distributions as does the Gaussian for the continuous Lévy-stable densities.

There are numerous instances where the discrete distribution has power-law asymptote but with index falling in the range -3 to -2 [1], which is outside the discrete-stable regime of validity. Such distributions are quantitatively different from the discrete-stable distributions since the mean exists, but the variance does not. It is to be expected that sums of iid random variables with power-law index outside the regime for the discrete-stable regime should have the Poisson distribution as attractor. The principal concern hereon is to verify this expectation, quantify the rate of convergence to the limit and discuss the implications.

A model for a probability distribution which has power-law asymptote, but is not of the discrete-stable class, is

$$P(n; m+\nu, a) = \frac{1}{\zeta(m+\nu, a)(n+a)^{m+\nu}} \tag{1}$$

where $0 < \nu < 1$, m is a positive integer, a is real and positive and the generalized Riemann zeta function [8],

$$\zeta(m+\nu, a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^{m+\nu}}$$

provides the normalization. Selecting $a = 1$ defines the Riemann zeta function $\zeta(m+\nu)$. The model (1), whilst not the most general form for a scale-free distribution, nevertheless encapsulates the essential properties required for analysing limiting distributions. The generating function equivalent to the distributions (1) is

$$q(s; m+\nu, a) = \sum_{n=0}^{\infty} P(n; m+\nu, a)(1-s)^n$$

which can be written in terms of the Lerch transcendent [9]

$$\Phi(1-s, m+\nu, a) = \zeta(m+\nu, a) q(s; m+\nu, a) = \sum_{n=0}^{\infty} \frac{(1-s)^n}{(n+a)^{m+\nu}}.$$

The properties of this function for small values of s determine the large n asymptotics of distributions that describe sums of iid random variables. There are surprisingly few instances documenting results for the form of Φ in this regime. Expressions obtained here are non-trivial and display subtle behaviour (insofar as the singularities of the function are concerned) as the index $m+\nu$ passes through integer values. It is the nature of these singularities that ultimately influence the asymptotic behaviour of the distribution and corresponding convergence to the limiting distribution.

Two key properties of the Lerch transcendent that can be utilized to reveal its small s asymptotics are the observation that

$$D^m((1-s)^a \Phi(1-s, m+\nu, a)) = (-1)^m (1-s)^a \Phi(1-s, \nu, a) \tag{2}$$

where the differential operator $D \equiv (1-s)\, \mathrm{d}/\mathrm{d}s$ and m is a positive integer, coupled with the limit [9]

$$\lim_{s\to 0} s^{1-\nu} \Phi(1-s, \nu, a) = \Gamma(1-\nu) \tag{3}$$

valid when $0 < \nu < 1$.

The first instance to consider is the convergence to the discrete-stable attractor of a distribution described by equation (1), with $m = 1$, giving index $1+\nu$ where $0 < \nu < 1$. Consider the generating function

$$q(s; 1+\nu, a) = \frac{1}{\zeta(1+\nu, a)} \sum_{n=0}^{\infty} \frac{(1-s)^n}{(a+n)^{1+\nu}} = \frac{\Phi(1-s, 1+\nu, a)}{\zeta(1+\nu, a)}$$

whose associated distribution has a power-law tail with index in the regime of the discrete-stable class. Evaluating equation (2) for $m = 1$ gives,

$$\frac{\mathrm{d}}{\mathrm{d}s}((1-s)^a \Phi(1-s, 1+\nu, a)) = -(1-s)^{a-1} \Phi(1-s, \nu, a), \tag{4}$$

whereupon taking the limit $s \to 0$, applying equation (3) to the right-hand side of (4) and then integrating obtains

$$(1-s)^a \Phi(1-s, 1+\nu, a) \approx -\frac{\Gamma(1-\nu)}{\nu} s^{\nu} {}_2F_1(\nu, 1-a, 1+\nu; s) + \zeta(1+\nu, a).$$

Here ${}_2F_1$ is the hypergeometric function [8], and $\zeta(1+\nu, a)$ the integration constant. Hence, an expansion for the generating function that is accurate to order s is

$$q(s, 1+\nu, a) \equiv \frac{\Phi(1-s, 1+\nu, a)}{\zeta(1+\nu, a)} \approx 1 - \frac{\Gamma(1-\nu)}{\nu \zeta(1+\nu, a)} s^{\nu} + as + O(s^{1+\nu}). \tag{5}$$

The generating function (5) is not differentiable at $s = 0$, indicating the mean of the distribution does not exist, as can be verified directly from equation (1). This generating function can also be used to investigate the convergence for sums of iid random variables. Consider the sum of N independent random variables having distribution given by equation (1). The generating function for the resultant is

$$Q_{1+\nu}(s) \equiv q(s, 1+\nu, a)^N$$

and the asymptotic properties of this are obtained from the behaviour of the generating function near $s = 0$ in the limit $N \to \infty$. Upon taking the logarithm of equation (5) and rescaling

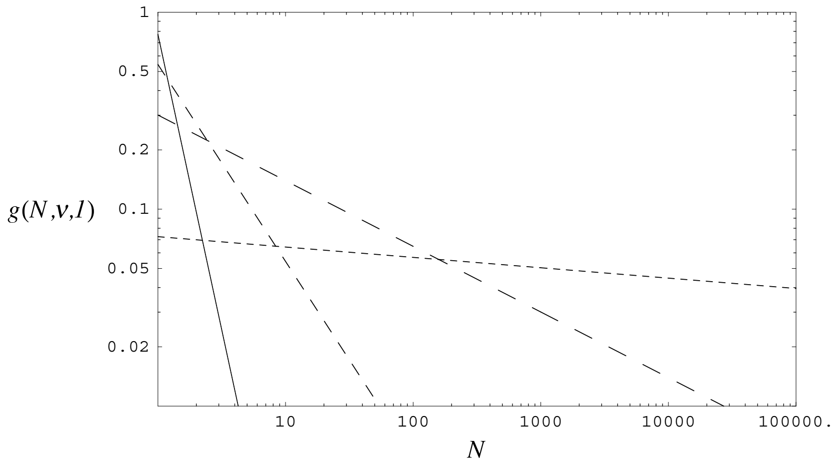

Figure 1. The term $g(N, \nu, 1)$ appearing in the expansion of the generating function given by equation (6), illustrating the convergence to the limiting discrete-stable distribution for sums of N iid random variables with index 1.25 ($\nu = 0.25$) (full), 1.5 ($\nu = 0.5$) (dashed), 1.75 ($\nu = 0.75$) (long-dashed), 1.95 ($\nu = 0.95$) (short-dashed). The parameter $a = 1$.

$s \to (\nu\zeta(1+\nu, a)/N\Gamma(1-\nu))^{1/\nu} s'$:

$$\begin{aligned}\ln(Q_{1+\nu}(s')) &\approx N \ln\left[1 - \frac{\Gamma(1-\nu)}{\nu\zeta(1+\nu, a)} s^{\nu} + as \ldots\right] \\ &\approx -s'^{\nu} + a\left(\frac{\nu\zeta(1+\nu, a)}{\Gamma(1-\nu)}\right)^{1/\nu} N^{1-1/\nu} s' + \cdots \\ &= -s'^{\nu} + g(N, \nu, a)s' + O(s^{1+\nu}) \qquad (6)\end{aligned}$$

which to leading order yields the generating function for the discrete-stable law, with a correction that vanishes like a fractional inverse power of N. Figure 1 depicts $g(N, \nu, a)$ as a function of N for selected of values of ν, illustrating that the convergence towards the stable attractor slows substantially as $\nu \to 1$, i.e. as the index of the distribution approaches 2.

The next case to consider is probability distributions (1) with index $2+\nu$, i.e. $m = 2$, again where $0 < \nu < 1$. Such distributions have finite mean but no variance. The marginal cases $\nu = 0$ and $\nu = 1$ will require separate treatment due to different nature of the singularities of the function at $s = 0$.

Evaluating equation (3) for $m = 2$ and proceeding as before, the following expansion for the generating function is obtained that is accurate to order $s^{1+\nu}$:

$$\begin{aligned}q(s, 2+\nu, a) &\equiv \frac{\Phi(1-s, 2+\nu, a)}{\zeta(2+\nu, a)} \\ &\approx 1 - \left(\frac{\zeta(1+\nu, a)}{\zeta(2+\nu, a)} - a\right) s + \frac{\pi \operatorname{cosec}(\pi\nu)}{\Gamma(2+\nu)\zeta(2+\nu, a)} s^{1+\nu} + O(s^2). \qquad (7)\end{aligned}$$

Differentiating (7) with respect to s and setting $s = 0$ gives the mean:

$$\langle n\rangle = \frac{\zeta(1+\nu, a)}{\zeta(2+\nu, a)} - a$$

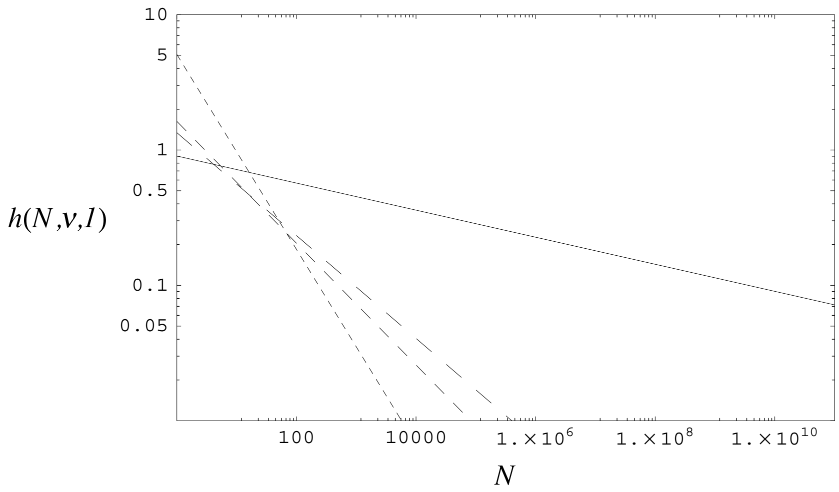

Figure 2. The term $h(N, \nu, 1)$ appearing in the expansion of the generating function given by equation (8), illustrating the convergence to the limiting Poisson distribution for sums of N iid random variables with index 2.1 ($\nu = 0.1$) (full), 2.25 ($\nu = 0.25$) (long-dashed), 2.5 ($\nu = 0.5$) (dashed), 2.75 ($\nu = 0.75$) (short-dashed). Again, the convergence is slowest as the index approaches 2. The parameter $a = 1$.

and again, this can be verified directly from equation (1). The second derivative of (7) at $s = 0$ formally obtains the second factorial moment $\langle n(n-1)\rangle$, but this does not exist since $0 < \nu < 1$.

The limiting distribution for the sum of N iid random variables can be found as before with the aid of equation (7), but here the relevant scaling factor is now $s \to s'/\langle n\rangle N$ which obtains

$$\ln(Q_{2+\nu}(s)) \approx -s'\left[1 - \frac{\pi\,\mathrm{cosec}(\pi\nu)}{\Gamma(2+\nu)\zeta(2+\nu)\langle n\rangle^{1+\nu}N^{\nu}}s'^{\nu} + \cdots\right].$$

$$\approx -s'(1 - h(N,\nu,a)s'^{\nu}) + O(s^2) \qquad (8)$$

This shows that the limiting distribution is the Poisson with a correction that vanishes as an inverse fractional power of N: note however that the value of the exponent differs from the previous case. Figure 2 shows $h(N, \nu, a)$ for different values of ν, and it can be seen that as $\nu \to 0$, i.e. as the index of the distribution approaches 2, the convergence to the Poisson slows substantially. Clearly, 2 is a marginal value for the index of the distribution.

This different behaviour occurs because of the nature of the singularity of the Lerch transcendent at $s = 0$ which changes for integer values of ν. This is revealed with greatest economy and transparency on setting $a = 1$ in the marginal case $m = 2$, $\nu = 0$, for then:

$$\frac{\mathrm{d}}{\mathrm{d}s}((1-s)\Phi(1-s,2,1)) = \frac{\ln s}{1-s}$$

and integrating this obtains the exact result:

$$\Phi(1-s,2,1) = \frac{\pi^2}{6}\frac{1}{1-s} + \frac{1}{1-s}\int_0^s \frac{\ln s'}{1-s'}\,\mathrm{d}s'$$

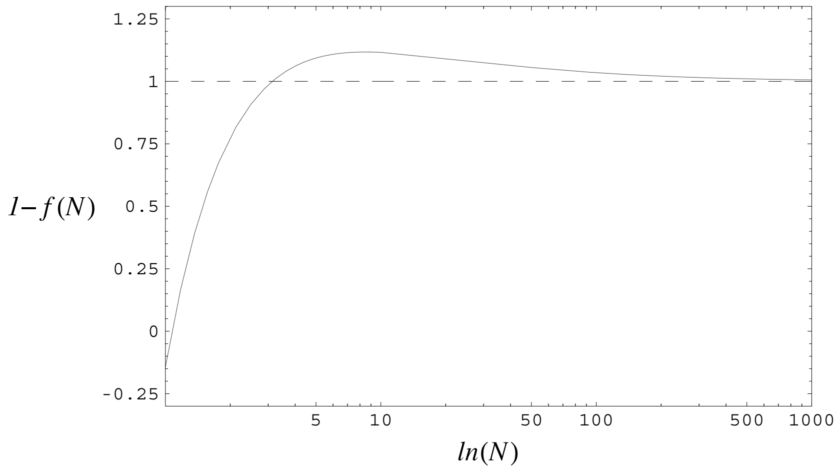

Figure 3. Illustrating the very slow convergence to the Poisson distribution for sums of N iid random variables with index 2. Note the abscissa is a logarithmic scale for the logarithm of N.

where the numerical value for $\zeta(2)$, the integration constant, has been substituted. Expanding the denominator of the integrand enables the integral to be evaluated as an infinite series of elementary functions, namely

$$\int_0^s \frac{\ln s'}{1-s'}\,\mathrm{d}s' = \sum_{n=0}^{\infty} \frac{s^{1+n}}{1+n}\left(\ln s - \frac{1}{1+n}\right).$$

Thus, the expansion of the corresponding generating function about $s = 0$ is readily determined to be

$$q(s;2,1) \equiv \frac{\Phi(1-s,2,1)}{\zeta(2)} \approx 1 + \frac{6}{\pi^2}s\ln s + \left(1-\frac{6}{\pi^2}\right)s + \frac{9}{\pi^2}s^2\ln s + \left(1-\frac{15}{2\pi^2}\right)s^2 + O(s^3\ln s). \quad (9)$$

Note that this function is not differentiable at $s = 0$ by virtue of the logarithmic singularity, implying that the mean of this distribution does not exist.

The convergence properties are found with the aid of equation (9) on scaling according to $s \to \pi^2 s'/6N\ln N$ which obtains

$$\ln(Q_2(s')) \approx N\ln\left[1+\frac{s'}{N\ln N}\ln\left(\frac{\pi^2}{6}\frac{s'}{N\ln N}\right)+\left(\frac{\pi^2}{6}-1\right)\frac{s'}{N\ln N}+\cdots\right]$$
$$\approx -s'\left[1-\frac{1}{\ln N}\left(\frac{\pi^2}{6}-1+\ln\left(\frac{\pi^2}{6}\right)-\ln(\ln N)\right)\right]+\frac{s'\ln s'}{\ln N}+\cdots$$
$$= -s'(1-f(N)) + \frac{s'\ln s'}{\ln N} + O((s'\ln s')^2). \quad (10)$$

This shows that the distribution of the resultant is Poisson, but the convergence is extraordinarily slow. Figure 3 shows $1 - f(N)$ over three decades of $\ln(N)$. This function passes through unity at $N \sim 23$, has a maximum at $N \sim 5023$ before converging to unity. The function is within 5% of its limiting value of unity only for $N > 3 \times 10^{25}$. The second term

Table 1. Summarizing the convergence properties of discrete power-law distributions: $0 < \nu < 1$ throughout.

Index	Limit distribution	Convergence rate
$1 + \nu$	Discrete stable of index $1 + \nu$	$N^{-(1/\nu-1)}$
2	Poisson	$\ln(\ln(N))/\ln(N)$
$2 + \nu$	Poisson	$N^{-\nu}$
$\geqslant 3$	Poisson	N^{-1}

containing the logarithmic singularity indicates that the mean of the resultant diverges for any finite value of N.

The other marginal case $\nu = 1$ can be derived from equation (9) on utilizing equation (2),

$$\frac{\mathrm{d}}{\mathrm{d}s}((1-s)\Phi(1-s,3,1)) = -\Phi(1-s,2,1) = -\zeta(2)q(s;2,1).$$

Performing the integration and inserting the value of the integration constant obtains the following expansion for the generating function about $s = 0$:

$$q(s;3,1) = \frac{\Phi(1-s,3,1)}{\zeta(3)} \approx 1 - \left(\frac{\zeta(2)}{\zeta(3)} - 1\right)s - \frac{1}{2\zeta(3)}s^2\ln s + O(s^2). \tag{11}$$

The first derivative of this function does exist at $s = 0$, and hence so too does the mean of the distribution. However, the function has a higher order logarithmic singularity which implies that it is not twice differentiable at $s = 0$. In general, the expansion of the generating function with integer value m for the index will have a term $\sim s^{m-1}\ln s$ implying that the $(m-1)$th order derivative does not exist at $s = 0$, and so the $(m-1)$th, and higher order factorial moments of the corresponding distribution are undefined.

The convergence properties in this regime of the index can be found using equation (11) and scaling according to $s \to s'/\langle n\rangle N$:

$$\begin{aligned}\ln(Q_3(s)) &\approx -s' + \frac{1}{2\zeta(3)\langle n\rangle^2 N}s'^2\ln s' \\ &\quad + \frac{1}{\langle n\rangle^2}\left(\left(\frac{3}{2\zeta(3)} + 2 - \frac{3\zeta(2)}{\zeta(3)}\right) + \frac{(\ln N + \ln\langle n\rangle)}{\zeta(3)}\right)\frac{s'^2}{2N} + \cdots \\ &\approx -s'\left(1 - \frac{217.029}{N}s'\ln s'\right) + O(s'^2)\end{aligned} \tag{12}$$

which has the customary N^{-1} convergence to the Poisson attractor, as indeed will all distributions with index $\geqslant 3$. A summary of the limit distributions and rates of convergence to these limits in the different index regimes is given in table 1.

The common emergent feature of these results is the slowing of convergence for sums of random variables to the Poisson attractor as their distributions tend towards $1/n^2$: this being the margin between those stable distributions with infinite mean and the Poisson distribution. This behaviour is especially extreme when considering ensembles at the margin, and is particularly relevant to scale-free networks, examples of which have indices that appear to cluster around 2. Whilst the results presented here conclude that networks with index $\geqslant 2$ inevitably evolve towards a random Poisson network, the size of that system should not be confused with the quantity N appearing in equations (6), (8), (10), (12): rather N refers to the number of independent ensembles that must be assembled. If any of these ensembles are correlated, then the sum required for convergence to the limit will be still larger. Estimates of the size of the Web in August 2003—evaluated in terms of the number of catalogued documents—is $\sim 3 \times 10^9$ [10]. The index of the order distribution depends on whether the number of incoming links are counted, in which case it ranges between 1.94 and 2.1, or if the number of outgoing links are considered, for which the range is 2.38—2.72 [1, 12, 13]. The index for incoming links straddles the marginal value 2 and hence, given the current size of the WWW, any evolution in the network topology will not yet be evident. Although the range of values for the index for outgoing links is larger, the convergence is still sufficiently slow as to make any evolution in the index undetectable at present. Nevertheless, it is the topology of the 'outgoing' network that will be first discerned as the overall size of the web increases.

It is important to stress fundamental differences between scale-free discrete and continuous random variables. The discrete-stable distributions are distinct from the one-sided Lévy-stable densities, the latter having non-zero mode whilst the former have zero mode, quite apart from their differing regimes of power-law behaviour. If, for example, sums of iid random variables with $P(n) \sim 1/n^{2+\nu}$, where $0 < \nu < 1$, were treated as sums of continuous iid random variable x with $p(x) \sim 1/x^{2+\nu}$, their convergence properties would be totally different. The continuous random variable would evolve towards the Lévy-stable density of index $2 + \nu$, for which the variance does not exist, whereas the discrete system would evolve towards a Poisson density for which all moments exist. Hence, treating discrete scale-free networks with continuum approximations can lead to erroneous conclusions with regard to the evolution of the network.

The existence of an outer scale to the distribution modifies these results—the limit being Poisson irrespective of the value of the index since, perforce, the mean exists for any distribution with an outer scale. However this addition is incremental on the current work, and introduces technicalities that obscure the straightforward treatment and conclusions given here, and so will be considered elsewhere.

Acknowledgment

J O Matthews acknowledges the support of the Leverhulme Trust.

References

[1] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[2] Feller W 1970 *An Introduction to Probability Theory and Its Applications* vols I and II 3rd edn (New York: Wiley)
[3] Lévy P 1937 *Theorie de l'Addition des Variables Aléatoires* (Paris: Gauthier-Villars)
[4] Cristoph G and Wolf W 1992 *Convergence Theorems with a Stable Limit Law* (Berlin: Akademie)
[5] Stuetael F W and van Harn K 1979 *Ann. Prob.* **5** 893
[6] Hopcraft K I, Jakeman E and Matthews J O 2002 *J. Phys. A: Math. Gen.* **35** L745
[7] Matthews J O, Hopcraft K I and Jakeman E 2003 *J Phys. A: Math. Gen.* **46** 11585
[8] Abramowitz M and Stegun I A 1972 *Handbook of Mathematical Functions* (New York: Dover)
[9] Gradshteyn I S and Ryzhik M 1994 *Tables of Integrals, Series and Products* 5th edn (London: Academic)
[10] www.searchenginewatch.com/reports/article.php/2156481
[11] Albert R, Jeong H and Barabási A L 1999 *Nature* **401** 130
[12] Kumar R *et al* 1999 *Proc. 9th ACM Symposium on Principles of Database Systems* p 1
[13] Broder A *et al* 2000 *Comput. Netw.* **33** 309