Combinatorics, Probability and Computing (2008) 17, 537–548.
doi:10.1017/S0963548308009176 Printed in the United Kingdom

The Weight and Hopcount of the Shortest Path in the Complete Graph with Exponential Weights

GERARD HOOGHIEMSTRA and PIET VAN MIEGHEM

Delft University of Technology, Electrical Engineering, Mathematics and Computer Science, PO Box 5031, 2600 GA Delft, The Netherlands
(e-mail: G.Hooghiemstra@ewi.tudelft.nl, P.VanMieghem@ewi.tudelft.nl)

Received 20 September 2007; revised 1 April 2008; first published online 16 June 2008

Both the hopcount H_N (the number of links) and the weight W_N (the sum of the weights on links) of the shortest path between two arbitrary nodes in the complete graph K_N with i.i.d. exponential link weights is computed. We consider the joint distribution of the pair (H_N, W_N) and derive, after proper scaling, the joint limiting distribution. One of the results is that H_N and W_N, properly scaled, are asymptotically independent.

1. Introduction

Consider the complete graph K_N with N nodes, and where the $N(N-1)/2$ links are equipped with independent, exponentially with parameter 1 distributed random variables. We take two nodes of K_N at random and construct the shortest path between the two nodes. The shortest path minimizes the weight of all paths between the chosen two nodes. The weight of a path is the sum of the weights of its constituent links. This shortest path model appears in epidemic modelling [2, Chapter 3], in telecommunications [9, Chapter 16], in percolation [4] and in combinatorics [8].

We let H_N denote the number of links of this shortest path and let W_N be its weight. The generating functions of H_N and W_N are given by ([10, 6]; see also [9])

$$\mathbb{E}[s^{H_N}] = \frac{N}{N-1}\left(\frac{\Gamma(N+s)}{N!\Gamma(x+1)} - \frac{1}{N}\right), \tag{1.1}$$

and

$$\mathbb{E}[e^{-tW_N}] = \frac{1}{N-1}\sum_{k=1}^{N-1}\prod_{n=1}^{k}\frac{n(N-n)}{t+n(N-n)}, \tag{1.2}$$

respectively.

In this paper we focus on the joint generating function $\mathbb{E}[s^{H_N}e^{-tW_N}]$, and its asymptotic properties. Interestingly, we find that W_N and H_N are asymptotically independent

(Theorem 3.1), and this matches nicely with one of our earlier findings [7], that the hopcount and the end-to-end delay of an Internet path are seemingly uncorrelated. We give two different proofs of the asymptotic behaviour of the scaled random variables W_N and H_N: the first proof is contained in Section 3, the second one in Section 4. The second, non-probabilistic proof is the shorter one. Finally, we compare the asymptotic law of (the scaled) W_N with earlier results of Janson [8].

2. The joint distribution of the weight and the hopcount

Theorem 2.1. *The joint generating function* $\mathbb{E}[s^{H_N}e^{-tW_N}]$ *is given by*

$$\varphi(s,t) = \mathbb{E}[s^{H_N}e^{-tW_N}] = \frac{1}{N-1}\sum_{n=1}^{N-1}\left(\prod_{k=1}^{n}\frac{k(N-k)}{t+k(N-k)}\right)\cdot\frac{\Gamma(n+s)}{n!\Gamma(s)}. \tag{2.1}$$

Proof. The length and the weight of the shortest path between two random nodes is in distribution equal to the same quantities of node 1 and a random node taken from the set $\{2,3,\ldots,N\}$. We denote the label of this random node by Z, which consequently has a uniform distribution over the above-mentioned discrete set of size $N-1$. Conditioning on the end node hence gives

$$\varphi(s,t) = \frac{1}{N-1}\sum_{n=1}^{N-1}\mathbb{E}[s^{H_N}e^{-tW_N}|Z=n+1].$$

In [4, pp. 227–228], a description is given to calculate the weight of the shortest path in the complete graph K_N, by adding nodes one by one according to a pure birth-process with birth rate $\lambda_k = k(N-k)$. Moreover, after the birth of the kth node, the distance of this node to the root (node 1) is determined by attaching this node independently to a uniform recursive tree (URT). From this construction, we find

$$\mathbb{E}[s^{H_N}e^{-tW_N}|Z=n+1] = \prod_{k=1}^{n}\frac{k(N-k)}{t+k(N-k)}\cdot\sum_{l=0}^{n-1}s^{l+1}\frac{\mathbb{E}[X_n^{(l)}]}{n}, \quad n=1,2,\ldots,N-1,$$

where the product $\prod_{k=1}^{n}\frac{\lambda_k}{t+\lambda_k}$ stems from the n different steps in the birth process to reach state n, and where $X_n^{(l)}$ is the number of nodes in the level set l of an URT of size n.

The basic recursion for these level sets is given by

$$\mathbb{E}[X_n^{(l)}] = \sum_{m=l}^{n-1}\frac{\mathbb{E}[X_m^{(l-1)}]}{m},$$

from which the probability generating function follows, as in [10, Lemma 1, p. 19]:

$$\sum_{l=0}^{n-1}s^l\mathbb{E}[X_n^{(l)}] = \frac{\Gamma(n+s)}{(n-1)!\Gamma(s+1)}.$$

Table 1.

N	simulations	$\rho(W_N, H_N)$	r.h.s. of (2.4)
10	0.35	0.286814	0.487984
20	0.36	0.324335	0.427821
30	0.36	0.334539	0.401511
50	0.35	0.339257	0.374380
100	0.35	0.336218	0.345057
150	0.34	0.331327	0.330801
200	0.34	0.326995	0.321695
400	0.32	0.314861	0.302515

Together this yields

$$\begin{aligned}\mathbb{E}[s^{H_N} e^{-tW_N} | Z = n+1] &= \prod_{k=1}^{n} \frac{k(N-k)}{t+k(N-k)} \cdot \sum_{l=0}^{n-1} s^{l+1} \frac{\mathbb{E}[X_n^{(l)}]}{n} \\ &= \prod_{k=1}^{n} \frac{k(N-k)}{t+k(N-k)} \cdot \frac{\Gamma(n+s)}{n!\Gamma(s)},\end{aligned}$$

and hence (2.1). □

Obviously, putting $s = 1$ in (2.1) yields (1.2). On the other hand, using the identity

$$\sum_{j=n}^{m} \frac{(a+j)!}{(b+j)!} = \frac{1}{a+1-b} \left\{ \frac{(a+m+1)!}{(b+m)!} - \frac{(a+n)!}{(b+n-1)!} \right\}, \tag{2.2}$$

we find that

$$\sum_{n=1}^{N-1} \frac{\Gamma(n+s)}{n!\Gamma(s)} = \frac{1}{\Gamma(s)} \sum_{n=1}^{N-1} \frac{(n+s-1)!}{n!} = \frac{1}{\Gamma(s+1)} \left\{ \frac{\Gamma(s+N)}{(N-1)!} - \Gamma(s+1) \right\},$$

and hence that

$$\varphi(s,0) = \mathbb{E}[s^{H_N}] = \frac{1}{N-1} \left\{ \frac{\Gamma(s+N)}{\Gamma(s+1)(N-1)!} - 1 \right\},$$

which is, indeed, (1.1).

As shown in the Appendix, the expectation of the product $H_N W_N$ is

$$\mathbb{E}[H_N W_N] = \frac{1}{N-1} \left(\left(\sum_{k=1}^{N-1} \frac{1}{k} \right)^2 - \sum_{k=1}^{N-1} \frac{1}{k} + \sum_{k=1}^{N-1} \frac{1}{k^2} \right). \tag{2.3}$$

For large N, we observe that $\mathbb{E}[H_N W_N] = \frac{\log^2 N - \log N + O(1)}{N}$. The asymptotics of the correlation coefficient $\rho(W_N, H_N)$, derived in the Appendix, is

$$\rho(W_N, H_N) = \frac{\pi\sqrt{2}}{6\sqrt{\ln N}} + o\big((\ln N)^{-1}\big), \tag{2.4}$$

which clearly tends to zero for $N \to \infty$. Table 1 compares the different expressions that we obtained for $\rho(W_N, H_N)$.

3. Limiting behaviour

Our main goal in this section is the limiting behaviour of the joint distribution of W_N and H_N, after proper scaling. Since our probabilistic method can be explained best by first analysing the asymptotic properties of the marginal distribution of W_N, after proper scaling, we start with the latter. This introduces only a small amount of additional work. Thereafter, we compare this marginal with a result of [11], where the limit of $NW_N - \ln N$ was computed by Laplace inversion. We also include a short derivation which shows that the random variable H_N is in the domain of attraction of the normal distribution, *i.e.*, $(\ln N)^{-1/2}(H_N - \ln N)$ converges to a standard normal random variable.

The idea is to condition on the random destination node Z introduced in Section 2. Let $A_N = Z - 1$, be uniformly distributed over the set $\{1, 2, \ldots, N-1\}$; then

$$W_N = \tau_1 + \cdots + \tau_{A_N},$$

where $\tau_1, \tau_2, \ldots$ is a sequence of independent exponentially distributed random variables with τ_k having parameter $\lambda_k = k(N-k)$, and where A_N is independent of this sequence. Indeed, with this interpretation

$$\mathbb{E}[e^{-tW_N}] = \mathbb{E}[e^{-t\sum_{k=1}^{A_N}\tau_k}] = \frac{1}{N-1}\sum_{k=1}^{N-1}\mathbb{E}[e^{-t(\tau_1+\cdots+\tau_k)}] = \frac{1}{N-1}\sum_{k=1}^{N-1}\prod_{i=1}^{k}\mathbb{E}[e^{-t\tau_i}],$$

which equals the right-hand side of (1.2), since

$$\mathbb{E}[e^{-t\tau_i}] = \frac{\lambda_i}{t+\lambda_i} = \frac{i(N-i)}{t+i(n-i)}.$$

We now follow the interpretation of [5, example on p. 118]. Define

$$Z_k = (N-k)\cdot\tau_k. \tag{3.1}$$

Then Z_k has an exponential distribution with parameter k. We claim that for each sequence $M_N \to \infty$, satisfying $M_N = o(N)$,

$$\sum_{k=1}^{M_N}(N\tau_k) - \ln M_N = \sum_{k=1}^{M_N}\left(\frac{NZ_k}{N-k}\right) - \ln M_N = (1+o(1))\sum_{k=1}^{M_N}Z_k - \ln M_N \xrightarrow{d} V, \tag{3.2}$$

where V denotes a Gumbel random variable, *i.e.*, a random variable with distribution function

$$\mathbb{P}\big[V \leqslant t\big] = \Lambda(t) = \exp(-e^{-t}).$$

Indeed (3.2) follows from the classical extreme value theorem (see, *e.g.*, [3]) for independent exponential random variables $\xi_1, \xi_2, \ldots, \xi_M$ with mean 1, for which the spacings

$$\xi_{(i)} - \xi_{(i-1)}, \quad i = 1, \ldots, M \quad (\xi_0 = 0)$$

are exponentially distributed with parameter $M - i + 1$, as follows:

$$\sum_{k=1}^{M} Z_k \stackrel{d}{=} \sum_{k=1}^{M}(\xi_{(i)} - \xi_{(i-1)}) = \xi_{(M)} = \max_{1\leqslant i\leqslant M}\xi_i.$$

This proves (3.2) since, by the mentioned extreme value limit theorem, $\xi_{(M_N)} - \ln M_N \xrightarrow{d} V$.

Writing $\tau_k^* = \tau_k - \mathbb{E}[\tau_k]$, we get, for any $0 < \delta < 1$,

$$\begin{aligned}\lim_{N\to\infty} \mathbb{P}(NW_N - \ln N \leqslant t) &= \lim_{N\to\infty} \mathbb{P}\left(N\sum_{k=1}^{A_N}\tau_k - \ln N \leqslant t\right)\\ &= \lim_{N\to\infty} \mathbb{P}\left(N\sum_{k=1}^{A_N}\tau_k^* - \ln N + N\sum_{k=1}^{A_N}\mathbb{E}[\tau_k] \leqslant t\right)\\ &= \lim_{N\to\infty} \mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{A_N}\mathbb{E}[\tau_k] \leqslant t\right), \end{aligned} \tag{3.3}$$

where the replacement of $N\sum_{k=1}^{A_N}\tau_k^*$ by $N\sum_{k=1}^{N^\delta}\tau_k^*$ in the last line is justified by Chebyshev inequality [9, p. 88] because, using conditioning on A_N,

$$\mathrm{Var}\left(N\sum_{k=N^\delta}^{A_N}\tau_k^*\right) = N^2\mathbb{E}\left[\sum_{k=N^\delta}^{A_N}\frac{1}{k^2(N-k)^2}\right] = \frac{N^2}{N-1}\left[\sum_{n=1}^{N-1}\sum_{k=N^\delta}^{n}\frac{1}{k^2(N-k)^2}\right] \to 0,$$

where the convergence to 0 follows by

$$\begin{aligned}\sum_{n=1}^{N-1}\sum_{k=N^\delta}^{n}\frac{1}{k^2(N-k)^2} &= \sum_{n=N^\delta}^{N-1}\sum_{k=N^\delta}^{n}\frac{1}{k^2(N-k)^2} = \sum_{k=N^\delta}^{N-1}\frac{1}{k^2(N-k)^2}\sum_{n=k}^{N-1}1\\ &= \sum_{k=N^\delta}^{N-1}\frac{1}{k^2(N-k)} = \frac{1}{N^2}\sum_{k=N^\delta}^{N-1}\frac{1}{k} + \frac{1}{N^2}\sum_{k=N^\delta}^{N-1}\frac{1}{(N-k)} + \frac{1}{N}\sum_{k=N^\delta}^{N-1}\frac{1}{k^2} = o\big(N^{-1}\big).\end{aligned}$$

By the law of total probability,

$$\begin{aligned} r_N &= \mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{A_N}\mathbb{E}[\tau_k] \leqslant t\right)\\ &= \sum_{j=1}^{N-1}\mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{A_N}\mathbb{E}[\tau_k] \leqslant t \,\middle|\, A_N = j\right)\Pr\big[A_N = j\big]\\ &= \frac{1}{N-1}\sum_{j=1}^{N-1}\mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{j}\mathbb{E}[\tau_k] \leqslant t\right). \end{aligned}$$

Replacing the latter sum by an integral gives

$$\begin{aligned} r_N &= \frac{1}{N-1}\int_0^{N-1}\mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{u}\mathbb{E}[\tau_k] \leqslant t\right)du + \Delta(N)\\ &= \frac{N}{N-1}\int_0^{\frac{N-1}{N}}\mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{\alpha N}\mathbb{E}[\tau_k] \leqslant t\right)d\alpha + \Delta(N). \end{aligned}$$

Taking the limit $N \to \infty$ yields

$$\lim_{N\to\infty}\mathbb{P}(NW_N - \ln N \leqslant t) = \lim_{N\to\infty}\int_0^1\mathbb{P}\left(N\sum_{k=1}^{N^\delta}\tau_k^* - \ln N + N\sum_{k=1}^{\alpha N}\mathbb{E}[\tau_k] \leqslant t\right)d\alpha, \tag{3.4}$$

provided that $\lim_{N\to\infty} \Delta(N) = 0$. Before proceeding with the right-hand side of (3.4) we prove that $\lim_{N\to\infty} \Delta(N) = 0$. Indeed, the random variable A_N has a uniform distribution on $\{1, 2, \ldots, N-1\}$, so that A_N/N converges in distribution to a uniform $(0,1)$ random variable. Hence, for any bounded and continuous function g on $(0,1)$, we have

$$\int_0^1 g(\alpha)\,d\alpha = \lim_{N\to\infty} \sum_{n=1}^{N-1} \frac{1}{N-1} g\left(\frac{n}{N-1}\right) = \lim_{N\to\infty} \sum_{n=1}^{N-1} g\left(\frac{A_N}{N-1}\right) \mathbb{P}(A_N = n). \tag{3.5}$$

Equality (3.5) on its own does not justify the conclusion $\lim_{N\to\infty} \Delta(N) = 0$. However, it *does* justify this conclusion whenever, for $\alpha \in (0,1)$, the function

$$g_N(\alpha) = \mathbb{P}\left(N \sum_{k=1}^{N^\delta} \tau_k^* - \ln N + N \sum_{k=1}^{\alpha N} \mathbb{E}[\tau_k] \leqslant t\right)$$

converges pointwise to some bounded and continuous function $g(\alpha)$, and this is demonstrated implicitly in the further steps of the proof below, with $g(\alpha) = \Lambda\big(t - \ln \frac{\alpha}{1-\alpha}\big)$. Therefore, we are allowed to proceed with the right-hand side of (3.4).

By the dominated convergence theorem [9, p. 100], bounding the involved probability by 1,

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{P}(NW_N - \ln N \leqslant t) &= \int_0^1 \lim_{N\to\infty} \mathbb{P}\left(N \sum_{k=1}^{N^\delta} \tau_k^* - \ln N + N \sum_{k=1}^{\alpha N} \mathbb{E}[\tau_k] \leqslant t\right) d\alpha \\
&= \int_0^1 \lim_{N\to\infty} g_N(\alpha)\,d\alpha.
\end{aligned}$$

Since $\tau_k^* = \tau_k - \mathbb{E}[\tau_k]$,

$$\begin{aligned}
N \sum_{k=1}^{N^\delta} \mathbb{E}[\tau_k] = \sum_{k=1}^{N^\delta} \frac{N}{k(N-k)} &= \left(\sum_{k=1}^{N^\delta} \frac{1}{k}\right)\big(1 + O(N^{\delta-1})\big) \\
&= \delta \ln N + O(N^{\delta-1} \ln N),
\end{aligned}$$

and

$$N \sum_{k=1}^{\alpha N} \mathbb{E}[\tau_k] = \sum_{k=1}^{\alpha N} \left(\frac{1}{k} + \frac{1}{N-k}\right) = \ln \alpha N - \ln(1-\alpha)N = \ln \frac{\alpha}{1-\alpha}.$$

We obtain

$$N \sum_{k=1}^{N^\delta} \tau_k^* - \ln N + N \sum_{k=1}^{\alpha N} \mathbb{E}[\tau_k] = N \sum_{k=1}^{N^\delta} \tau_k - \delta \ln N + O(N^{\delta-1} \ln N) - \ln N + \ln \frac{\alpha}{1-\alpha}.$$

Applying (3.2), with $M_N = N^\delta = o(N)$, finally gives for each fixed t

$$\lim_{N\to\infty} g_N(\alpha) = \lim_{N\to\infty} \mathbb{P}\left(N \sum_{k=1}^{N^\delta} \tau_k^* - \ln N + N \sum_{k=1}^{\alpha N} \mathbb{E}[\tau_k] \leqslant t\right) = \mathbb{P}\left(V + \ln \frac{\alpha}{1-\alpha} \leqslant t\right) = g(\alpha),$$

and hence

$$\lim_{N\to\infty} \mathbb{P}(NW_N - \ln N \leqslant t) = \int_0^1 g(\alpha)\,d\alpha = \int_0^1 \Lambda\left(t - \ln \frac{\alpha}{1-\alpha}\right) d\alpha. \tag{3.6}$$

In [11], the limit of $NW_N - \ln N$ was derived by first computing the limit of the scaled transform (1.2) and then applying the inversion theorem for transforms. We verify that our result (3.6) is identical to [11, (10)]. Integration by parts and subsequently, a change of the variable $u = \frac{e^{-t}}{1-\alpha}$, yields

$$\begin{aligned}\int_0^1 \Lambda\left(t - \ln\frac{\alpha}{1-\alpha}\right) d\alpha &= e^{-t}\int_0^1 \frac{\alpha}{(1-\alpha)^2}\exp\left\{\frac{-\alpha e^{-t}}{(1-\alpha)}\right\} d\alpha \\ &= e^{-t}\int_{e^{-t}}^{\infty} ue^t(ue^t-1)\exp\{-e^{-t}(ue^t-1)\}\,\frac{e^{-t}}{u}du \\ &= e^{e^{-t}}\int_{e^{-t}}^{\infty}\frac{u-e^{-t}}{u}e^{-u}\,du = 1 - e^{-t}e^{e^{-t}}\int_{e^{-t}}^{\infty}\frac{e^{-u}}{u}\,du.\end{aligned}$$

We proceed with the asymptotic analysis of the generating function

$$\mathbb{E}[s^{Y_N}] = \frac{\Gamma(N+s)}{N!\,\Gamma(s+1)}. \tag{3.7}$$

The random variable Y_N has the same asymptotic properties as the random variable H_N, as can be seen by comparing (3.7) with (1.1). It is straightforward to compute the expectation $\mu_N = \mathbb{E}[Y_N]$ and the standard deviation $\sigma_N = \sqrt{\mathrm{Var}(Y_N)}$:

$$\begin{aligned}\mu_N &= \ln N + \gamma - 1 + O\left(\frac{\ln N}{N}\right), \\ \sigma_N &= \sqrt{\ln N + \gamma - \frac{\pi^2}{6} + O\left(\frac{\ln N}{N}\right)}.\end{aligned}$$

The random variable Y_N (and hence H_N) is asymptotic normal (the generating function (3.7) is close to that of a Poisson random variable), *i.e.*,

$$\lim_{N\to\infty}\mathbb{P}\left(\frac{H_N-\mu_N}{\sigma_N}\leqslant y\right) = \lim_{N\to\infty}\mathbb{P}\left(\frac{Y_N-\mu_N}{\sigma_N}\leqslant y\right) = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-z^2/2}\,dz.$$

The joint limiting behaviour of W_N and H_N is given in the next theorem.

Theorem 3.1.

$$\lim_{N\to\infty}\mathbb{P}\left(NW_N - \ln N \leqslant t, \frac{H_N-\mu_N}{\sigma_N}\leqslant y\right) = \Phi(y)\cdot\int_0^1 \Lambda\left(t-\ln\frac{\alpha}{1-\alpha}\right)d\alpha. \tag{3.8}$$

In particular, it follows that W_N and H_N are asymptotically independent.

Proof. Again we observe from (1.2) that, conditionally on $A_N = n$, the random variables H_N and W_N are independent, where

$$\begin{aligned}\left(W_N|A_N = n\right) &\overset{d}{=} \tau_1 + \tau_2 + \cdots + \tau_n, \\ \left(H_N|A_N = n\right) &\overset{d}{=} Y_n,\end{aligned}$$

and where $\{\tau_n\}_{n=1}^{N-1}$ is defined as before and Y_N is independent from the sequence $\{\tau_n\}_{n=1}^{N-1}$, having generating function (3.7) with $N = n$. Parallel to the derivation which leads to (3.6):

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}\left(N \sum_{k=1}^{A_N} \tau_k - \ln N \leqslant t, \frac{Y_{A_N} - \mu_N}{\sigma_N} \leqslant y \right) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N-1} \mathbb{P}\left(N \sum_{k=1}^{n} \tau_k - \ln N \leqslant t, \frac{Y_n - \mu_N}{\sigma_N} \leqslant y | A_N = n \right) \mathbb{P}(A_N = n) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N-1} \mathbb{P}\left(N \sum_{k=1}^{n} \tau_k - \ln N \leqslant t | A_N = n \right) \cdot \mathbb{P}\left(\frac{Y_n - \mu_N}{\sigma_N} \leqslant y | A_N = n \right) \mathbb{P}(A_N = n) \\
&= \lim_{N\to\infty} \int_0^1 \mathbb{P}\left(N \sum_{k=1}^{\alpha N} \tau_k - \ln N \leqslant t \right) \mathbb{P}\left(\frac{Y_{\alpha N} - \mu_N}{\sigma_N} \leqslant y \right) d\alpha. \qquad (3.9)
\end{aligned}$$

As said before, the first factor in the integrand on the right-hand side of (3.9) has been treated in the derivation leading to (3.6). For the second factor we write:

$$\mathbb{P}\left(\frac{Y_{\alpha N} - \mu_N}{\sigma_N} \leqslant y \right) = \mathbb{P}\left(\frac{Y_{\alpha N} - \mu_{\alpha N}}{\sigma_{\alpha N}} \leqslant y \frac{\sigma_N}{\sigma_{\alpha N}} - \frac{\mu_{\alpha N} - \mu_N}{\sigma_{\alpha N}} \right) \to \Phi(y),$$

since, for $\alpha \in (0, 1)$ fixed,

$$\frac{\sigma_N}{\sigma_{\alpha N}} \to 1, \qquad \frac{\mu_{\alpha N} - \mu_N}{\sigma_{\alpha N}} \to 0.$$

The interchange of limit and integral is again justified by the Lebesgue theorem (dominated convergence), the integrand being dominated by 1, since it is a probability. □

4. The asymptotic PGF and PDF

Theorem 3.1 can also be proved by inverting the Laplace transform (2.1). Following a procedure similar to [9, pp. 518–520], we write

$$t + k(N-k) = \left(\sqrt{\left(\frac{N}{2}\right)^2 + t} + \frac{N}{2} - k \right) \left(\sqrt{\left(\frac{N}{2}\right)^2 + t} - \left(\frac{N}{2} - k\right) \right),$$

and define $y = \sqrt{\left(\frac{N}{2}\right)^2 + t}$. Then,

$$\begin{aligned}
\prod_{k=1}^{n} \frac{k(N-k)}{t + k(N-k)} &= \frac{n!(N-1)!}{(N-n-1)!} \prod_{k=1}^{n} \frac{1}{(y + \frac{N}{2} - k)} \prod_{k=1}^{n} \frac{1}{(y - \frac{N}{2} + k)} \\
&= \frac{n!(N-1)!}{(N-n-1)!} \frac{\Gamma(y + \frac{N}{2} - n)}{\Gamma(y + \frac{N}{2})} \frac{\Gamma(y - \frac{N}{2} + 1)}{\Gamma(y - \frac{N}{2} + n + 1)},
\end{aligned}$$

and, substituted in (2.1), yields

$$\varphi(s, t) = \frac{(N-2)!\Gamma(y - \frac{N}{2} + 1)}{\Gamma(y + \frac{N}{2})\Gamma(s)} \sum_{n=1}^{N-1} \frac{\Gamma(n+s)}{(N-n-1)!} \frac{\Gamma(y + \frac{N}{2} - n)}{\Gamma(y - \frac{N}{2} + n + 1)}.$$

For large N and $|t| < N$, we have that $y = \sqrt{(\frac{N}{2})^2 + t} \sim \frac{N}{2} + \frac{t}{N}$, such that

$$\varphi(s,t) \sim \frac{(N-2)!\Gamma(\frac{t}{N}+1)}{\Gamma(N+\frac{t}{N})\Gamma(s)} \sum_{n=1}^{N-1} \frac{\Gamma(n+s)}{(N-n-1)!} \frac{\Gamma(N+\frac{t}{N}-n)}{\Gamma(\frac{t}{N}+n+1)}.$$

We now introduce the scaling $t = Nx$, where $|x| < 1$,

$$\varphi(s,Nx) \sim \frac{(N-2)!\Gamma(x+1)}{\Gamma(N+x)\Gamma(s)} \sum_{n=1}^{N-1} \frac{\Gamma(n+s)}{(N-n-1)!} \frac{\Gamma(x+N-n)}{\Gamma(x+1+n)}. \tag{4.1}$$

Following an approach analogous to that of [11], the sum

$$S = \sum_{n=1}^{N-1} \frac{\Gamma(n+s)}{(N-n-1)!} \frac{\Gamma(N+x-n)}{\Gamma(x+1+n)}$$

can be transformed into

$$S = \frac{\Gamma(N+s+x)}{\Gamma(x+1-s)\Gamma(N-1)} \sum_{k=0}^{\infty} \frac{\Gamma(x+1-s+k)}{k!(1+x+k)} \frac{\Gamma(N+x+k)}{\Gamma(N+2x+1+k)}.$$

For large N and fixed s and x, and using (as in [11]) Gauss's series for the hypergeometric series [1, 15.1.20], the asymptotic order of the sum S scales as

$$S = \frac{\Gamma(N+s+x)N^{-x-1}}{\Gamma(N-1)} \frac{\Gamma(1+x)\Gamma(s-x)}{\Gamma(1+s)} \left(1 - O\left(\frac{1}{N}\right)\right). \tag{4.2}$$

Substitution of (4.2) into (4.1), leads, for large N, fixed s and $|x| < 1$, to

$$\varphi(s,Nx) \sim N^{s-1-x} \frac{\Gamma^2(1+x)\Gamma(s-x)}{\Gamma(s)\Gamma(1+s)} \left(1 - O\left(\frac{1}{N}\right)\right).$$

This result suggests considering the scaling

$$\mathbb{E}\left[s^{\frac{H_N - \mathbb{E}[H_N]}{a_N}} e^{-xNW_N}\right] = s^{-\frac{\mathbb{E}[H_N]}{a_N}} \varphi\left(s^{\frac{1}{a_N}}, xN\right)$$

for large N, where $\mathbb{E}[H_N] \sim \mu_N \sim \ln N$ and where a_N will be determined to have a finite limit for $N \to \infty$. With this scaling, we have

$$s^{-\frac{\mathbb{E}[H_N]}{a_N}} \varphi\left(s^{\frac{1}{a_N}}, xN\right) \sim N^{-x} s^{-\frac{\mathbb{E}[H_N]}{a_N}} N^{s^{\frac{1}{a_N}}-1} \frac{\Gamma^2(1+x)\Gamma\left(s^{\frac{1}{a_N}} - x\right)}{\Gamma\left(s^{\frac{1}{a_N}}\right)\Gamma\left(1+s^{\frac{1}{a_N}}\right)} \left(1 - O\left(\frac{1}{N}\right)\right).$$

Furthermore,

$$\begin{aligned} \ln\left(s^{-\frac{\mathbb{E}[H_N]}{a_N}} N^{s^{\frac{1}{a_N}}-1}\right) &\sim \left(s^{\frac{1}{a_N}} - 1\right) \ln N - \frac{\mathbb{E}[H_N]}{a_N} \ln s \\ &\sim \ln N \left(e^{\frac{\ln s}{a_N}} - \frac{\ln s}{a_N} - 1\right) = \ln N \left(\frac{\ln^2 s}{2a_N^2} + O\left(a_N^{-3}\right)\right), \end{aligned}$$

which tends to a finite limit provided $a_N^2 \sim \ln N$. Hence, if we choose $a_N = \sigma_N \sim \sqrt{\ln N}$, then, for any finite complex number $s \neq 0$, we arrive at

$$s^{-\frac{\mathbb{E}[H_N]}{a_N}} \varphi\left(s^{\frac{1}{a_N}}, xN\right) \sim N^{-x}\Gamma(x+1)e^{\frac{\ln^2 s}{2}}\Gamma(1+x)\Gamma(1-x)\left(1-O\left(\frac{1}{N}\right)\right)$$
$$= N^{-x}e^{\frac{\ln^2 s}{2}}\Gamma(x+1)\frac{\pi x}{\sin \pi x}\left(1-O\left(\frac{1}{N}\right)\right),$$

from which

$$\lim_{N\to\infty} \mathbb{E}\left[s^{\frac{H_N-\mathbb{E}[H_N]}{\sigma_{H_N}}} e^{-x(NW_N-\ln N)}\right] = e^{\frac{\ln^2 s}{2}}\Gamma(x+1)\frac{\pi x}{\sin \pi x}. \tag{4.3}$$

This again shows that the normalized (continuous) random variables

$$\frac{H_N - \mathbb{E}[H_N]}{\sigma_{H_N}} \quad \text{and} \quad NW_N - \ln N$$

are asymptotically independent. After replacing $s \to e^{-y}$, the inverse Laplace transform then yields

$$\lim_{N\to\infty} \Pr\left[\frac{H_N - \mathbb{E}[H_N]}{\sigma_{H_N}} \leqslant t\right] = \Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-z^2/2}\,dz$$

and, as shown in [11],

$$\lim_{N\to\infty} \Pr\left[NW_N - \ln N \leqslant t\right] = 1 - e^{-t}e^{e^{-t}}\int_{e^{-t}}^{\infty}\frac{e^{-u}}{u}du$$
$$= \int_0^1 \Lambda\left(t - \ln\frac{\alpha}{1-\alpha}\right)d\alpha.$$

The latter integral is a mixture of the Gumbel distribution.

5. Discussion

Janson was the first to compute the asymptotics of $NW_N - \ln N$ in [8], where he gave a short proof that $NW_N - \ln N$ converges in distribution to the convolution of the Gumbel distribution with the logistic distribution ($L(x) = e^x/(1+e^x)$). In our notation, Janson proves that, in asymptotic-L_2 sense, the distribution of $N\sum_{k=1}^{A_N}\tau_k$ is equivalent to that of

$$\sum_{k=1}^{\infty}\frac{1}{k}(\xi_k - 1) + \ln\frac{A_N}{N-A_N} + \ln N + \gamma,$$

where $\{\xi_k\}$ is a sequence of i.i.d. exponentially distributed random variables with mean 1.

Using probability generating functions (PGFs), Janson then recognizes that

$$\sum_{k=1}^{\infty}\frac{1}{k}(\xi_k - 1) + \gamma$$

has a Gumbel distribution and that the logistic distribution, which is the limit of $\ln(A_N/(N-A_N))$ is the difference of two independent Gumbel random variables [8, Theorem 5],

$$NW_N - \ln N \xrightarrow{d} V_1 + V_2 - V_3, \tag{5.1}$$

where V_1, V_2 and V_3 are independent Gumbel-distributed random variables. Since

$$\frac{\pi x}{\sin \pi x} = \Gamma(1+x)\Gamma(1-x),$$

and $\Gamma(1+x)$ is the PGF of a Gumbel random variable, relation (4.3) in the second proof leads to the same nice interpretation of (5.1).

In our (independent) first proof (Section 3) we were able to identify why the Gumbel distribution appears. This is explained by writing the *deterministic* sum: $\sum_{k=1}^{M} Z_k$ as a maximum of i.i.d. exponentially distributed random variables, see below (3.2). However, NW_N is a *random* sum of the variables $Z_1, Z_2, \ldots$, and consequently, by conditioning, we obtain as the end result a mixture of the Gumbel distribution: see (3.6).

For the second and third Gumbel random variable we have no better explanation than that apparently $\ln(U/(1-U))$, the limit of $\ln \frac{A_N/N}{1-A_N/N}$, is the difference of two independent Gumbels or equivalently, that the quotient of two independent Exp(1) random variables is equal in distribution to $U/(1-U)$. Both distributions are equal to a shifted Pareto distribution:

$$\mathbb{P}(\xi_1/\xi_2 > t) = \mathbb{P}(U/(1-U) > t) = (1+t)^{-1}, \quad t > 0.$$

Acknowledgements

We would like to thank Huijuan Wang for simulating the correlation coefficient $\rho(W_N, H_N)$.

Appendix: Derivation of the joint expectation $\mathbb{E}[H_N W_N]$

From (2.1), we obtain by differentiation

$$\mathbb{E}[H_N W_N] = -\frac{\partial^2 \varphi(s,t)}{\partial t \partial s}\Big|_{s=1,t=0} = \frac{1}{N-1}\sum_{n=1}^{N-1}\left(\sum_{k=1}^{n}\frac{1}{k(N-k)} \cdot \sum_{j=1}^{n}\frac{1}{j}\right). \tag{A.1}$$

This expression can be simplified to

$$\mathbb{E}[H_N W_N] = \frac{1}{N-1}\left(2\sum_{k=1}^{N-1}\frac{1}{k}\sum_{j=1}^{k}\frac{1}{j} - \sum_{k=1}^{N-1}\frac{1}{k}\right).$$

In addition,

$$\begin{aligned}\left(\sum_{k=1}^{N-1}\frac{1}{k}\right)^2 &= \sum_{k=1}^{N-1}\frac{1}{k}\sum_{j=1}^{N-1}\frac{1}{j} = \sum_{k=1}^{N-1}\frac{1}{k}\left(\sum_{j=1}^{k}\frac{1}{j} + \sum_{j=k}^{N-1}\frac{1}{j} - \frac{1}{k}\right)\\ &= 2\sum_{k=1}^{N-1}\frac{1}{k}\sum_{j=1}^{k}\frac{1}{j} - \sum_{k=1}^{N-1}\frac{1}{k^2},\end{aligned}$$

whose use leads to (2.3).

The linear correlation coefficient is defined as

$$\rho(W_N, H_N) = \frac{\mathbb{E}[H_N W_N] - \mathbb{E}[W_N]\mathbb{E}[H_N]}{\sigma_{W_N}\sigma_{H_N}}. \tag{A.2}$$

Using (see [9, p. 360])

$$E[W_N] = \frac{1}{N-1}\sum_{n=1}^{N-1}\frac{1}{n}, \qquad \mathrm{Var}(W_N) = \frac{3}{N(N-1)}\sum_{n=1}^{N-1}\frac{1}{n^2} - \frac{\left(\sum_{n=1}^{N-1}\frac{1}{n}\right)^2}{(N-1)^2 N}, \tag{A.3}$$

and

$$E[H_N] = \frac{N}{N-1}\sum_{l=2}^{N}\frac{1}{l}, \qquad \mathrm{Var}(H_N) = \frac{N}{N-1}\sum_{l=1}^{N}\frac{1}{l} - \frac{N}{N-1}\sum_{l=1}^{N}\frac{1}{l^2},$$

we obtain

$$\mathbb{E}[H_N W_N] - \mathbb{E}[W_N]\mathbb{E}[H_N] = \frac{1}{N-1}\sum_{k=1}^{N-1}\frac{1}{k^2} - \left(\frac{1}{N-1}\sum_{n=1}^{N-1}\frac{1}{n}\right)^2,$$

which is, for large N,

$$\mathbb{E}[H_N W_N] - \mathbb{E}[W_N]\mathbb{E}[H_N] = \frac{1}{N}\frac{\pi^2}{6} - \frac{\ln^2 N}{N^2} + O\left(\frac{1}{N^2}\right),$$

and

$$\mathrm{Var}(W_N)\mathrm{Var}(H_N) = \frac{\frac{\pi^2}{2}\ln N}{N^2} - \frac{\frac{\pi^4}{12}}{N^2} + O\left(\frac{\ln N}{N^3}\right).$$

Introducing these asymptotics in (A.2) leads to (2.4).

References

[1] Abramowitz, M. and Stegun, I. A. (1968) *Handbook of Mathematical Functions*, Dover.

[2] Daley, D. J. and Gani, J. (1999) *Epidemic Modelling: An Introduction*, Studies in Mathematical Biology, Cambridge University Press.

[3] Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn, Krieger.

[4] van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2001) First-passage percolation on the random graph. *Probab. Engrg Inform. Sci.* (PEIS) **15** 225–237.

[5] van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2002) The flooding time in random graphs. *Extremes* **5** 111–129.

[6] van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2006) The size and the weight of the shortest path trees with exponential link weights. *Combin. Probab. Comput.* **16** 1–26.

[7] Hooghiemstra, G. and Van Mieghem, P. (2001) Delay distributions on fixed internet paths. Delft University of Technology, report 20011020. `www.nas.ewi.tudelft.nl/people/Piet/TUDelftreports`

[8] Janson, S. (1999) One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.* **8** 347–361.

[9] Van Mieghem, P. (2006) *Performance Analysis of Communications Systems and Networks*, Cambridge University Press.

[10] Van Mieghem, P., Hooghiemstra, G. and van der Hofstad, R. (2000) A scaling law for the hopcount. Report, TU-Delft.

[11] Van Mieghem, P. and Tang, S. (2008) Weight of the shortest path to the first encountered peer in a peer group of size *m*. *Probab. Engrg Inform. Sci.* (PEIS) **22**, 37–52.