Size and Weight of Shortest Path Trees with Exponential Link Weights

Remco van der Hofstad *   Gerard Hooghiemstra †   Piet Van Mieghem†

Februari 22, 2005

Abstract

We derive the distribution of the number of links and the average weight for the shortest path tree (SPT) rooted at an arbitrary node to m uniformly chosen nodes in the complete graph of size N with i.i.d. exponential link weights. We rely on the fact that the full shortest path tree to all destinations (i.e., $m = N - 1$) is a uniform recursive tree to derive a recursion for the generating function of the number of links of the SPT, and solve this recursion exactly.

The explicit form of the generating function allows us to compute the expectation and variance of the size of the subtree for all m. We also obtain exact expressions for the average weight of the subtree.

1   Introduction

In a graph with weighted links, the shortest path between two nodes is the path for which the sum of the link weights is minimal. The shortest path tree (SPT) rooted at an arbitrary node to m uniformly chosen nodes is the union of the shortest paths from that node to all m different nodes in the graph. In this article, the SPT in the complete graph with exponential link weights is analyzed. We denote by $H_N(m)$ the number of links in the SPT rooted at an arbitrary node to m other uniformly chosen nodes; $H_N(m)$ also equals the size (or the number of nodes) of that subtree minus 1 node. We denote by $W_N(m)$ the weight of the SPT rooted at an arbitrary node to m other uniformly chosen nodes, i.e. $W_N(m)$ is the sum of all the link weights of the subtree spanned by the root and the m uniform nodes.

Our interest was triggered by the hopcount problem [25] in the Internet: "What is the distribution of the number of hops (or traversed routers) between two arbitrary nodes in the Internet?". As initial model, we concentrated on the class of random graphs $G_p(N)$ [2] with uniformly or exponentially distributed link weights where p is the probability that there is a link between two arbitrary nodes and this probability is independent of the existence of a link between any other node pair. We denote this class of graphs by RGU. In [13, 25], we have rephrased the shortest path problem between two arbitrary nodes in the complete graph $(p = 1)$ with exponential link weights to a Markov discovery

---

process which starts the path searching process at the source and which is a continuous time Markov chain with N states. Each state n represents the n already discovered nodes (including the source node). If at some stage in the Markov discovery process n nodes are discovered, then the next node is reached with rate $\lambda_{n,n+1} = n(N-n)$, which is the transition rate in the continuous time Markov chain. Since the discovering of nodes at each stage only increases n, the Markov discovery process is a pure birth process with birth rate $n(N-n)$. We call τ_n the inter-attachement time between the inclusion of the n-th and $(n+1)$-th node to the SPT for $n = 1, \ldots, N-1$. The inter-attachement time τ_n is exponentially distributed with parameter $n(N-n)$ as follows from the theory of Markov processes. By the memoryless property of the exponential distribution, the new node is added uniformly to an already discovered node. Hence, the resulting SPT to all nodes, i.e. $m = N-1$, is exactly a uniform recursive tree (URT). A URT of size N is a random tree rooted at some source node and where at each stage a new node is attached uniformly to one of the existing nodes until the total number of nodes is equal to N.

It is known [19] that the hopcount of the shortest path between two arbitrary nodes in a complete graph with exponential link weights is equal to the depth $D_{1,N}$ from the source or root to an arbitrary node in a URT of size N. In [13], we have extended this result asymptotically for large N to $G_p(N)$ for $p = p_N < 1$: the law of the hopcount of the shortest path in $G_p(N)$ with exponentially distributed link weights is close (as $N \to \infty$) to the law of $D_{1,N}$. This law for RGU has been proved in [13] under the condition[1] that $Np_N/(\log N)^3 \to \infty$. In other words, for large N, our generalization shows that the hopcount of the shortest path in the class RGU is, to first order, *independent* of $p = p_N > \frac{(\log N)^3}{N}$. In this section, we will first review results on the shortest path and then proceed with the extension to the SPT rooted at an arbitrary node to m uniform nodes.

1.1 The shortest path

For a URT, the hopcount $H_N = H_N(1)$, which is defined as the number of links of the shortest path from one arbitrary node to another arbitrary node, can be exactly determined (see e.g., [19] and [25]). The generating function of the number of hops in the URT from the root to a uniform node (different from the root) equals

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

from which the average hopcount

$$\mathbb{E}[H_N] = \frac{N}{N-1}\sum_{n=2}^{N}\frac{1}{n},$$

is immediate. The average weight of the longest shortest path (LSP), i.e. the largest weight along the shortest paths between the root and any other node, in the complete graph with N nodes and with i.i.d. exponential link weights with mean 1 was first determined by Janson [12] and in [14] for the RGU with the condition on $p = p_N$ that $Np_N/(\log N)^3 \to \infty$ for $N \to \infty$ as

$$\mathbb{E}[W_{\text{LSP}}] = \frac{2}{N}\sum_{n=1}^{N-1}\frac{1}{n} = \frac{2}{N}\log N + O\left(N^{-1}\right).$$

---

[1]Computer simulations confirm the limit law even when $Np_N \to \infty$ at a slower rate and we believe that the law holds for any $p > p_c \sim \frac{\log N}{N}$ where p_c is the disconnectivity threshold in $G_p(N)$.

By the memoryless property of the exponential distribution, the laws of the birth process and of the URT are independent [13]. The generating function of the weight of a shortest path to a uniform destination, defined as $\varphi_{W_N(1)}(t) = \mathbb{E}\left[e^{-tW_N(1)}\right]$, follows as

$$
\begin{aligned}
\varphi_{W_N(1)}(t) &= \sum_{k=1}^{N-1} \mathbb{E}\left[e^{-tW_k}\right] \mathbb{P}\left[\text{endnode is } k\text{-th attached node in URT}\right] \\
&= \frac{1}{N-1} \sum_{k=1}^{N-1} \prod_{n=1}^{k} \frac{n(N-n)}{t + n(N-n)},
\end{aligned}
\tag{2}
$$

where W_k denotes the weight of the path to the kth attached node and $W_k = \sum_{n=1}^{k} \tau_n$. As mentioned above, the inter-attachment time τ_n is an exponential random variable with parameter $n(N-n)$ and τ_n is independent from τ_k for all $k \neq n$. The average length, obtained from (2), is equal to

$$
\mathbb{E}[W_N(1)] = \frac{1}{N-1} \sum_{n=1}^{N-1} \frac{1}{n},
\tag{3}
$$

which is about half the maximum weight of the shortest path which is equal to $\mathbb{E}[W_{\text{LSP}}] = \frac{2}{N} \sum_{n=1}^{N-1} \frac{1}{n}$.

Previous to Janson's work [12] in 1999, the average weight $\mathbb{E}[W_N(1)]$ had been studied around 1985 by various authors (e.g. [8, 16]). Only asymptotic bounds for large N such as $c_1 \frac{\log N}{N} \leq \mathbb{E}[W_N(1)] \leq c_2 \frac{\log N}{N}$ for positive real numbers c_1 and c_2 which appears in [16] had then been published. In the past decade, URTs have received considerable attention [23], but the problem of finding the size of the subtree spanned by the root and m uniform nodes was still open.

1.2 The shortest path tree (SPT)

The SPT rooted at an arbitrary node to m uniformly chosen nodes in the complete graph with exponential link weights seems a reasonable model for the *multicast structure* of the Internet (see e.g. [26],[11]). While unicast is a one-to-one communication, multicast is a one (or many)-to-many communication mode where a single message destined for m different receivers is sent once by the sender and copied at each branch point of the multicast tree. Hence, in multicast, a message travels only once over links from a sender to m receivers, in contrast to unicast where that message is sent m times to each individual receiver (as today in emails). Clearly, when dealing with large groups of destinations, multicast consumes less network resources and is more efficient than unicast. Multicast is expected to be used in direct marketing, movie-distribution, pay-TV, automatic update of software releases and many other services beside the already known applications such as video conference, multicast file transfer, tele-classing and electronic games. Due to the promising potential of multicast, an accurate modeling of multicast tree properties is a main motivation for the analysis of the SPT in this article.

The knowledge of $H_N(m)$ is important to the estimation of the efficiency of multicast over unicast in the Internet. The ratio $\frac{\mathbb{E}[H_N(m)]}{m\mathbb{E}[H_N(1)]}$ is regarded as a measure for the multicast efficiency. From Internet measurements, Chuang and Sirbu [4] observed that the multicast efficiency $\frac{\mathbb{E}[H_N(m)]}{m\mathbb{E}[H_N(1)]}$ decreases as a power of m, and they estimated the exponent by -0.2. Philips *et al.* [20] have dubbed this behavior "the Chuang-Sirbu law". In [20] and [26], the multicast efficiency $\frac{\mathbb{E}[H_N(m)]}{m\mathbb{E}[H_N(1)]}$ was studied in more detail using random graph models. The average number of links of the SPT $g_N(m) = \mathbb{E}[H_N(m)]$ in the

complete graph with exponential link weights to m arbitrary nodes derived in [26] is specified below by (8). The proof of $g_N(m)$ in [26] uses inclusion/exclusion on the number of links, and writes

$$g_N(m) = \sum_{i=1}^{m} \binom{m}{i} (-1)^{i-1} \mathbb{E}\left[X_i^{(N)}\right], \qquad (4)$$

where $X_i^{(N)}$ denotes the number of joint links of the shortest paths from the root to i uniform and different locations. The inclusion/exclusion formula (4) holds for any graph, and was also used in [26] to compute $g_N(m)$ on a regular tree. In this paper, we will use a similar inclusion/exclusion relation to compute the expected weight of the tree $\mathbb{E}[W_N(m)]$ spanned by m uniform and different locations. Moreover, we will give the exact law for $H_N(m)$ (Theorem 2.1) which completes our previous work [26] where only $\mathbb{E}[H_N(m)]$ was derived.

Although the class RGU seems adequate to model *multicast*, that graph is not a good model for *the topology* of the Internet. From measurements, Faloutsos *et al.* [6] observed that the degree distribution of Internet nodes obeys a power-law. On the other hand as shown in [15], the degree distribution of a uniform node in the complete graph with exponential weights, counting only those links that are used along a shortest path, has generating function equal to $\varphi_{H_N}(z)$, and is thus close to a Poisson random variable with expectation equal to $\log N$. Since the URT is asymptotically also the SPT in random graphs $G_p(N)$ with i.i.d. exponential or uniform link weights and with the link density $p = p_N$ such that $\frac{Np_N}{(\log N)^3} \to \infty$ as shown in [13], we believe that the presented results based on the complete graph are also applicable to the class RGU. The interest of this generalization lies in the fact that Ad-Hoc wireless networks [10] and certain peer-to-peer networks are well modelled by random graphs.

Recently, Bollobás *et al.* [3] have computed the asymptotic weight of the Steiner tree spanning $m+1$ nodes in the complete graph with N nodes and with i.i.d. exponential link weights with mean 1 as

$$W_{\text{Steiner},N}(m) = (1 + o(1)) \frac{m}{N} \log \frac{N}{m+1}, \qquad (5)$$

for large N and $m = o(N)$. The Steiner tree is the minimal weight tree that connects $m+1$ given nodes in a graph. Steiner tree problems arise in many applications [17]. The computation of the Steiner tree is a NP-complete problem [18]: on current PCs, the exact computation of the Steiner tree problem is limited to small graphs of size $N < 50$ [27]. From a practical point of view it often suffices that the approximation is close enough, so that it is not necessary to compute the Steiner tree itself. There exist many heuristics (see e.g. Winter *et al.* [17]) for the Steiner tree and many of them employ the SPT. As a by-product of the results in this paper, we will complement the results of Bollobás *et al.* [3] and investigate the relation between the Steiner tree and the SPT.

2 Main results

In this section, we present our main results. We start by identifying the distribution of $H_N(m)$.

Theorem 2.1 *For all $N \geq 1$ and $1 \leq m \leq N-1$, the probability generating function $\varphi_{H_N(m)}(z) = \mathbb{E}\left[z^{H_N(m)}\right]$ equals*

$$\varphi_{H_N(m)}(z) = \frac{m!(N-1-m)!}{((N-1)!)^2} \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} \frac{\Gamma(N+kz)}{\Gamma(1+kz)}. \qquad (6)$$

Consequently,

$$\mathbb{P}\left[H_N(m) = j\right] = \frac{m!(-1)^{N-(j+1)}S_N^{(j+1)}\mathcal{S}_j^{(m)}}{(N-1)!\binom{N-1}{m}}, \tag{7}$$

where $S_N^{(j+1)}$ and $\mathcal{S}_j^{(m)}$ denote the Stirling numbers of first and second kind [1, 24.1.4 and 24.1.4].

Based on the properties of the URT, we derive a recursion relation for the generating function $\varphi_{H_N(m)}(z)$ of the number of links $H_N(m)$ in Proposition 3.1 below. We solve this recursion exactly to obtain (6). Figure 1 plots the probability density function of $H_{50}(m)$ for different values of m.
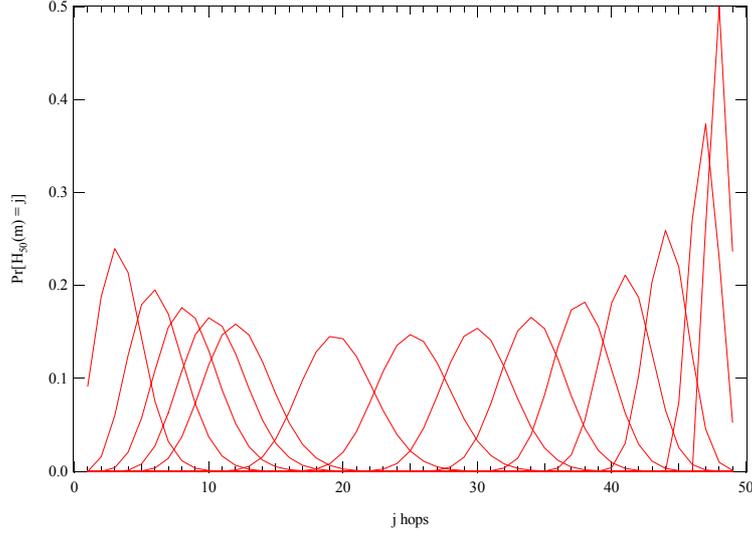

Figure 1: The pdf of $H_{50}(m)$ for $m = 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 47$.

Even though (7) completely determines the law of $H_N(m)$, the Stirling numbers complicate the computation of properties of the law of $H_N(m)$. By differentiating $\varphi_{H_N(m)}(z)$, we have computed the expectation and variance of $H_N(m)$ as follows:

Corollary 2.2 *For all $N \geq 1$ and $1 \leq m \leq N - 1$,*

$$g_N(m) = \mathbb{E}\left[H_N(m)\right] = \frac{mN}{N-m}\sum_{k=m+1}^{N}\frac{1}{k}, \tag{8}$$

and

$$\text{Var}\left[H_N(m)\right] = \frac{N-1+m}{N+1-m}g_N(m) - \frac{g_N^2(m)}{N+1-m} - \frac{m^2N^2}{(N-m)(N+1-m)}\sum_{k=m+1}^{N}\frac{1}{k^2}. \tag{9}$$

For $N = 1000$, Figure 2 illustrates the typical behavior for large N of the expectation $g_N(m)$ and the standard deviation $\sigma_N(m) = \sqrt{\text{Var}\left[H_N(m)\right]}$ of $H_N(m)$ for all values of m. Note that $H_N(N-1) = N - 1$, so that $\text{Var}\left[H_N(N-1)\right] = 0$. Since $g_N(m)$ can be rewritten in terms of the digamma function $\psi(x)$ as $g_N(m) = mN\left(\frac{\psi(N)-\psi(m)}{N-m}\right) - 1$, using [1, 6.1.38] we find the asymptotic law for $N \to \infty$ and any m,

$$g_N(m) = \frac{mN}{N-m}\log\frac{N}{m} - \frac{1}{2} + O\left(\frac{N+m}{mN}\right)$$

From (9), we observe that for all m,

$$\operatorname{Var}\left[H_N(m)\right] \leq \frac{N-1+m}{N+1-m} g_N(m) \leq \frac{2N}{N-m} g_N(m) = \frac{2 g_N^2(m)}{m \sum_{k=m+1}^{N} \frac{1}{k}} = o(g_N^2(m)),$$

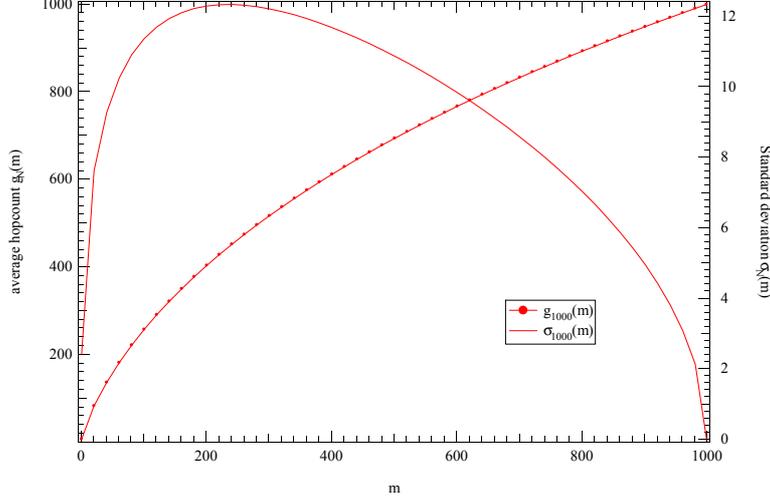which implies that $\frac{|H_N(m) - g_N(m)|}{g_N(m)} = o(1)$ with probability converging to 1 for all m.

Figure 2: The average number of hops $g_N(m)$ (left axis) and the corresponding standard deviation $\sigma_N(m)$ (right axis) as a function of m for $N = 1000$.

Our final result on $H_N(m)$ is more precise. We denote by

$$X_N(m) = \frac{H_N(m) - g_N(m)}{\sigma_N(m)}, \tag{10}$$

the standardized size $H_N(m)$ of the subtree. Figure 1 suggests that $H_N(m)$ is asymptotically normal.

Theorem 2.3 *For all $m = o(\sqrt{N})$, $X_N(m)$ converges in distribution to a standard normal random variable when $N \to \infty$, i.e., $X_N(m) \xrightarrow{d} \mathcal{N}(0,1)$.*

We believe that the convergence towards a normal random variable is true for more values of m, as formulated in the following conjecture:

Conjecture 2.4 *For all m such that $N - m \to \infty$, $X_N(m)$ converges to a standard normal random variable when $N \to \infty$.*

Corollary 2.2 shows that $\sigma_N(m) \to \infty$ precisely when $N - m \to \infty$. Moreover, by the Chebychev inequality, the variable $X_N(m)$ is tight, so that it remains to identify the limit to be Gaussian.

We now turn to the average weight $u_N(m) = \mathbb{E}[W_N(m)]$. Our main result is the following theorem:

Theorem 2.5 *For all $1 \leq m \leq N-1$, the average weight $u_N(m) = \mathbb{E}[W_N(m)]$ equals*

$$u_N(m) = \sum_{j=1}^{m} \frac{1}{N-j} \sum_{k=j}^{N-1} \frac{1}{k}, \tag{11}$$

where $u_N(1)$ is computed in (3).

Corollary 2.6 *For all $N \geq 2$,*

$$u_N(N-1) = \sum_{n=1}^{N-1} \frac{1}{n^2}. \tag{12}$$

Since $u_N(m) \leq u_N(N-1) < \frac{\pi^2}{6} = \zeta(2)$ where $\zeta(z)$ is the Riemann Zeta function, any SPT in the considered graph has an average weight smaller than $\zeta(2) = 1.645$. Frieze [7] has shown that the average weight of the minimum spanning tree in the same considered graph is $\zeta(3) < \zeta(2)$. Since $\mathbb{E}[W_{\text{Steiner},N}(N-1)]$ is increasing in N, there holds for any N that $\mathbb{E}[W_{\text{Steiner},N}(N-1)] \leq \zeta(3)$ and the ratio $\frac{\zeta(2)}{\zeta(3)} = 1.368$ indicates that the use of the SPT (computationally easy) never performs on average more than 37% worse than the optimal Steiner tree (computationally unfeasible). In a broader context and referring to the recent concept of "the Prize of Anarchy", which is broadly explained in [22], the SPT used in a communication network is related to the Nash equilibrium, while the Steiner tree gives the (hardly achievable) global optimum.

Corollary 2.7 *For all $m = O(N^a)$ with $a < 1$,*

$$u_N(m) = \frac{m}{N} \log \frac{N}{m+1} + O\left(N^{2(a-1)} \log N\right) \tag{13}$$

The proof of (13) (which we omit) follows from (11) by using the asymptotic expansion of the digamma-function [1, 6.3.38].

We now use Theorem 2.5 together with the main result (5) in [3] to compare the weight of the SPT to that of the Steiner tree:

Corollary 2.8 *For all $m = o(N)$,*

$$\mathbb{E}[W_{\text{Steiner},N}(m)] = (1 + o(1)) \frac{m}{N} \log \frac{N}{m+1}. \tag{14}$$

Consequently, for all $m = o(N)$ and all $\epsilon > 0$,

$$\mathbb{P}\left(|W_{\text{Steiner},N}(m) - W_N(m)| > \epsilon u_N(m)\right) \to 0, \quad and \quad \mathbb{P}\left(|W_N(m) - u_N(m)| > \epsilon u_N(m)\right) \to 0. \tag{15}$$

Corollary 2.8 shows that the complete graph with exponential weights is an example for which the Steiner tree and the SPT perform asymptotically equally well, at least when $m = o(N)$. Note that when $m = N - 1$, this is not the case as explained above.

3 The recursion for $\varphi_{H_N(m)}(z)$

In this section we derive a recursion for the probability generating function $\varphi_{H_N(m)}(z)$ of the number of links $H_N(m)$ in the SPT to m uniformly chosen nodes. We also discuss a slight generalization of this recursion and derive the recursion for $g_N(m) = \mathbb{E}[H_N(m)] = \varphi'_{H_N(m)}(1)$.

Proposition 3.1 *For $N > 1$ and all $1 \leq m \leq N - 1$,*

$$\varphi_{H_N(m)}(z) = \frac{(N-m-1)(N-1+mz)}{(N-1)^2} \varphi_{H_{N-1}(m)}(z) + \frac{m^2 z}{(N-1)^2} \varphi_{H_{N-1}(m-1)}(z). \tag{16}$$
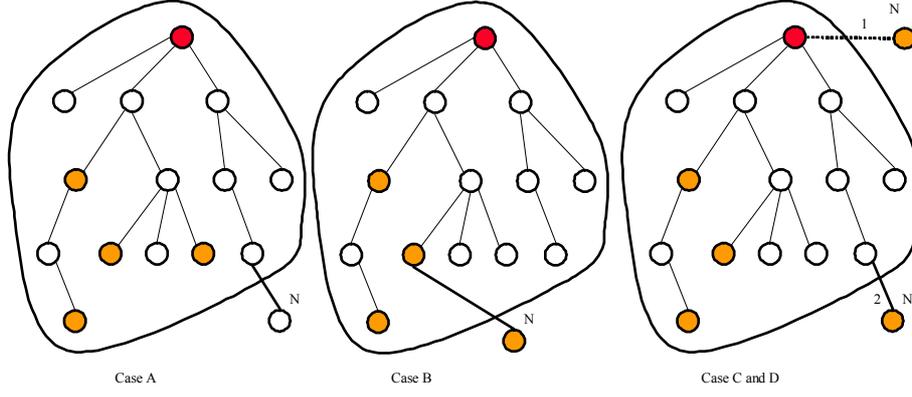
Figure 3: The several possible cases in which the Nth node can be attached uniformly to the URT of size $N-1$. The root is dark shaded while the m member nodes are lightly shaded.

Proof. To prove (16), we use the recursive growth of URTs: a URT of size N is a URT of size $N-1$, where we add an additional link to a uniformly chosen node.

Let $H_N(m)$ denote the number of links in the subtree spanned by the root and the m uniform nodes. In order to obtain a recursion for $H_N(m)$ we distinguish between the m uniformly chosen nodes all being in the URT of size $N-1$ or not. The probability that they all belong to the tree of size $N-1$ is equal to $1 - \frac{m}{N-1}$ (case A in Figure 3). If they all belong to the URT of size $N-1$, then we have that $H_N(m) = H_{N-1}(m)$. Thus, we obtain

$$\varphi_{H_N(m)}(z) = \left(1 - \frac{m}{N-1}\right)\varphi_{H_{N-1}(m)}(z) + \frac{m}{N-1}\mathbb{E}\left(z^{1+L_{N-1}(m)}\right), \tag{17}$$

where $L_{N-1}(m)$ is the number of links in the subtree of the URT of size $N-1$ spanned by $m-1$ uniform nodes together with the ancestor of the added N^{th} node. We complete the proof by investigating the generating function of $L_{N-1}(m)$. Again, there are two cases. In the first case (B in Figure 3), the ancestor of the added N^{th} node is one of the $m-1$ previous nodes (which can only happen if it is unequal to the root), else we get one of the cases C and D in Figure 3. The probability of the first event equals $\frac{m-1}{N-1}$, the probability of the latter equals $1 - \frac{m-1}{N-1}$. If the ancestor of the added N^{th} node is one of the $m-1$ previous nodes, then the number of links equals $H_{N-1}(m-1)$, otherwise the generating function of the number of additional links equals

$$\frac{N-m-1}{N-m}\varphi_{H_{N-1}(m)}(z) + \frac{1}{N-1}\varphi_{H_{N-1}(m-1)}(z).$$

The first contribution comes from the case where the ancestor of the added N^{th} node is not the root, and the second where it is equal to the root. Therefore,

$$
\begin{aligned}
\mathbb{E}\left[z^{L_{N-1}(m)}\right] &= \frac{m-1}{N-1}\varphi_{H_{N-1}(m)}(z) + \frac{N-m}{N-1}\left(\frac{N-m-1}{N-m}\varphi_{H_{N-1}(m)}(z) + \frac{1}{N-m}\varphi_{H_{N-1}(m-1)}(z)\right) \\
&= \frac{m}{N-1}\varphi_{H_{N-1}(m-1)}(z) + \frac{N-m-1}{N-1}\varphi_{H_{N-1}(m)}(z).
\end{aligned} \tag{18}
$$

Substitution of (18) into (17) leads to (16). $\square$

Since $g_N(m) = \mathbb{E}[H_N(m)] = \varphi'_{H_N(m)}(1)$, we obtain from (16) the recursion for $g_N(m)$,

$$g_N(m) = \left(1 - \frac{m^2}{(N-1)^2}\right)g_{N-1}(m) + \frac{m^2}{(N-1)^2}g_{N-1}(m-1) + \frac{m}{N-1}. \tag{19}$$

8

Denote by T the set of links of the SPT from the root to m uniformly chosen nodes. If each link e in the graph is, independently of the i.i.d. exponential link weights, specified by an additional (i.i.d.) link measure $r = r(e)$, then the generating function $\varphi_{R_N(m)}(z)$ of $R_N(m) = \sum_{e \in T} r(e)$ is

$$\varphi_{R_N(m)}(z) = \mathbb{E}\left[z^{R_N(m)}\right] = \mathbb{E}\left[z^{\sum_{e \in T} r(e)}\right] = \mathbb{E}\left[(\varphi_r(z))^{H_N(m)}\right] = \varphi_{H_N(m)}(\varphi_r(z)), \qquad (20)$$

where $\varphi_r(z) = \mathbb{E}[z^r]$ is the generating function of the i.i.d. common link value r. Examples of additional link measures are the monetary usage cost of a link, the total number of lost packets, the physical length of the link, etc. The role and impact of assigning additional properties $r(e)$ (independent of the link weights) to link e in a network are currently important issues of a quality of service (QoS) architecture in the Internet [28].

4 Solution of the recursion for $\varphi_{H_N(m)}(z)$

In this section, the recursion for $\varphi_{H_N(m)}(z)$ is solved and used to prove the other properties of $H_N(m)$.
Proof of Theorem 2.1: By iterating the recursion (16) for small values of m, the computations in Appendix A suggest the solution (6) for (16). It is indeed verified that (6) satisfies (16). This proves (6) of Theorem 2.1.

Using [1, 24.1.3.B], the Taylor expansion around $z = 0$ equals

$$
\begin{aligned}
\varphi_{H_N(m)}(z) &= \frac{m!N(N-1-m)!}{(N-1)!} \sum_{k=0}^{m} \binom{m}{k}(-1)^{m-k}\left(\frac{\Gamma(N+kz)}{N!\Gamma(1+kz)} - \frac{1}{N}\right) \\
&= \frac{m!N(N-1-m)!}{(N-1)!} \sum_{k=0}^{m} \binom{m}{k}(-1)^{m-k} \sum_{j=1}^{N-1} \frac{(-1)^{N-(j+1)}S_N^{(j+1)}}{N!}k^j z^j \\
&= \frac{m!N(N-1-m)!}{(N-1)!} \sum_{j=1}^{N-1} \frac{(-1)^{N-(j+1)}S_N^{(j+1)}}{N!}\left(\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}k^j\right)z^j.
\end{aligned}
$$

where $S_N^{(k)}$ denotes the Stirling Numbers of the first kind [1, 24.1.3]. Using the definition of Stirling Numbers of the second kind [1, 24.1.4.C],

$$m!\mathcal{S}_j^{(m)} = \sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}k^j,$$

for which $\mathcal{S}_j^{(m)} = 0$ if $j < m$, gives

$$\varphi_{H_N(m)}(z) = \frac{(m!)^2(N-1-m)!}{((N-1)!)^2}\sum_{j=1}^{N-1}(-1)^{N-(j+1)}S_N^{(j+1)}\mathcal{S}_j^{(m)}z^j.$$

This proves (7), and completes the proof of Theorem 2.1. $\square$
Proof of Corollary 2.2: The expectation and variance of $H_N(m)$ will not be obtained using the explicit probabilities (7), but by rewriting (6) as

$$
\begin{aligned}
\varphi_{H_N(m)}(z) &= \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\partial_t^{N-1}\left[t^{N-1+kz}\right]_{t=1} \\
&= \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}(-1)^m\partial_t^{N-1}\left[t^{N-1}(1-t^z)^m\right]_{t=1}. \qquad (21)
\end{aligned}
$$

Indeed, we can write

$$
\begin{aligned}
\mathbb{E}[H_N(m)] &= \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}(-1)^m\partial_z\partial_t^{N-1}\left[t^{N-1}(1-t^z)^m\right]_{t=z=1} \\
&= \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}m(-1)^{m-1}\partial_t^{N-1}\left[t^N\log t(1-t)^{m-1}\right]_{t=1},\\
\mathbb{E}[H_N(m)\,(H_N(m)-1)] &= \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}(-1)^m\partial_z^2\partial_t^{N-1}\left[t^{N-1}(1-t^z)^m\right]_{t=z=1} \\
&= \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}m(-1)^{m-1} \\
&\quad \times\partial_t^{N-1}\left[t^N\log^2 t(1-t)^{m-2}\left[-(m-1)t+(1-t)\right]\right]_{t=1}. \tag{22}
\end{aligned}
$$

We will start with the former. Using $\partial_t^i(1-t)^j|_{t=1}=j!(-1)^j\delta_{i,j}$ and Leibniz rule, we find

$$
\mathbb{E}[H_N(m)] = \frac{\Gamma(m+1)\Gamma(N-m)}{\Gamma^2(N)}m!\binom{N-1}{m-1}\partial_t^{N-m}\left[t^N\log t\right]_{t=1}.
$$

Since

$$
\partial_t^k[t^n\log t]_{t=1} = \frac{n!}{(n-k)!}\sum_{j=n-k+1}^{n}\frac{1}{j},
$$

we obtain expression (8) for $\mathbb{E}[H_N(m)]$.

We now extend the above computation to $\mathbb{E}[H_N(m)(H_N(m)-1)]$. Using

$$
\partial_t^k[t^n\log^2 t]_{t=1} = 2\frac{n!}{(n-k)!}\sum_{i=n-k+1}^{n}\sum_{j=i+1}^{n}\frac{1}{ij} = \frac{n!}{(n-k)!}\left[\left(\sum_{i=n-k+1}^{n}\frac{1}{i}\right)^2-\sum_{i=n-k+1}^{n}\frac{1}{i^2}\right],
$$

we obtain,

$$
\begin{aligned}
m(m-1)(-1)^{m-2}\partial_t^{N-1}\left[t^{N+1}\log^2 t(1-t)^{m-2}\right]_{t=1} &= \binom{N-1}{m-2}m(m-1)(m-2)!\partial_t^{N-m+1}\left[t^{N+1}\log^2 t\right]_{t=1} \\
&= (N+1)!\binom{N-1}{m-2}\left[\left(\sum_{k=m+1}^{N+1}\frac{1}{k}\right)^2-\sum_{k=m+1}^{N+1}\frac{1}{k^2}\right].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
m(-1)^{m-1}\partial_t^{N-1}\left[t^N\log^2 t(1-t)^{m-1}\right]_{t=1} &= \binom{N-1}{m-1}m(m-1)!\partial_t^{N-m}\left[t^N\log^2 t\right]_{t=1} \\
&= N!\binom{N-1}{m-1}\left[\left(\sum_{k=m+1}^{N}\frac{1}{k}\right)^2-\sum_{k=m+1}^{N}\frac{1}{k^2}\right].
\end{aligned}
$$

Substitution into (22) leads to

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

From $g_N(m)=\mathbb{E}[H_N(m)]$ and $\mathrm{Var}\,(H_N(m))=\mathbb{E}[H_N(m)(H_N(m)-1)]+g_N(m)-g_N^2(m)$, we obtain (9). This completes the proof of Corollary 2.2. $\square$

Remark: The expectation and variance of the quantity $R_N(m)=\sum_{i=1}^{H_N(m)}r_i$, can be found by conditioning on $H_N(m)$. Denote by $\mu=\mathbb{E}[r]$ and by $\sigma^2=\mathrm{Var}(r)$, then

$$
\mathbb{E}[R_N(m)]=\mu\mathbb{E}[H_N(m)], \qquad \text{and} \qquad \mathrm{Var}(R_N(m))=\sigma^2\mathbb{E}[H_N(m)]+\mu^2\mathrm{Var}(H_N(m)). \tag{23}
$$

10

We close this section with the proof of Theorem 2.3:

Proof of Theorem 2.3: Observe from (9) that $\frac{\sigma_N(m)}{\sqrt{g_N(m)}} \to 1$ as long as $m/N \to 0$. Therefore, it is sufficient to prove that

$$\frac{H_N(m) - g_N(m)}{\sqrt{g_N(m)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Denote by ψ_N the moment generating function of this quantity, then

$$\psi_N(t) = e^{-t\sqrt{g_N(m)}} \cdot \mathbb{E}\left[\exp\left\{\frac{tH_N(m)}{\sqrt{g_N(m)}}\right\}\right] = e^{-t\sqrt{g_N(m)}}\varphi_{H_N(m)}\left(e^{\frac{t}{\sqrt{g_N(m)}}}\right). \qquad (24)$$

Since $\exp(t^2/2)$ is the moment generating function of a standard normal random variable, it suffices to show that $\lim_{N\to\infty} \psi_N(t) = \exp(t^2/2)$. For large N, using [1, 6.1.47] we can write

$$\frac{\Gamma(N+kz)}{\Gamma(N)} = N^{kz}\left(1 + \frac{f_k(z)}{N}\right), \quad 1 \le k \le m,$$

where $f_k(z) = kz(kz-1)/2 + o(1)$ and as long as $k^3/N^2 \le m^3/N^2 \to 0$. From this, we obtain

$$\left|\varphi_{H_N(m)}(z) - \frac{m!(N-1-m)!}{(N-1)!}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\frac{N^{kz}}{\Gamma(1+kz)}\right| \le \frac{m!(N-1-m)!}{(N-1)!}\sum_{k=0}^{m}\binom{m}{k}\frac{f_k(z)N^{kz-1}}{\Gamma(1+kz)}. \qquad (25)$$

Denote the upper bound on the right-hand side of (25) by $e_N(m)$, then for $N \to \infty$,

$$\varphi_{H_N(m)}(z) = \frac{m!(N-1-m)!}{(N-1)!}\sum_{k=0}^{m}\binom{m}{k}\frac{(-1)^{m-k}}{\Gamma(1+kz)}N^{kz} + O(e_N(m)).$$

We use Hankel's contour integral [1, 6.1.4] to write

$$\frac{1}{\Gamma(1+kz)} = \frac{i}{2\pi}\int_C (-u)^{-(1+kz)}e^{-u}du, \qquad (26)$$

and obtain

$$\begin{aligned}
\varphi_{H_N(m)}(z) &= \frac{m!(N-1-m)!}{(N-1)!}\sum_{k=0}^{m}\binom{m}{k}\left(N^{kz}\right)(-1)^{m-k}\frac{i}{2\pi}\int_C(-u)^{-1-kz}e^{-u}du + O(e_N(m)) \\
&= \frac{m!(N-1-m)!}{(N-1)!}N^{mz}\frac{i}{2\pi}\int_C(-u)^{-mz-1}\left(1 - \left(\frac{-u}{N}\right)^z\right)^m e^{-u}du + O(e_N(m)). \qquad (27)
\end{aligned}$$

We substitute (27) into (24) and compute the appearing factors separately. First, rewrite

$$e^{-t\sqrt{g_N(m)}} \cdot \frac{(N-1-m)!}{(N-1)!}N^{me^{\frac{t}{\sqrt{g_N(m)}}}} = \frac{(N-1-m)!}{(N-1)!} \cdot N^m \cdot e^{-t\sqrt{g_N(m)}}N^{m(e^{\frac{t}{\sqrt{g_N(m)}}}-1)}.$$

Further, because $g_N(m) \to \infty$ as $N \to \infty$, we have

$$\begin{aligned}
e^{-t\sqrt{g_N(m)}}N^{m\left(e^{\frac{t}{\sqrt{g_N(m)}}}-1\right)} &= \exp\left(-t\sqrt{g_N(m)} + m\log N\left(e^{\frac{t}{\sqrt{g_N(m)}}} - 1\right)\right) \\
&= \exp\left(-t\sqrt{g_N(m)} + g_N(m)\left(\frac{t}{\sqrt{g_N(m)}} + \frac{t^2}{2g_N(m)} + O\left(\frac{t^3}{g_N^{3/2}(m)}\right)\right)\right) \\
&= e^{\frac{t^2}{2}}\left(1 + O\left(\frac{1}{\sqrt{g_N(m)}}\right)\right).
\end{aligned}$$

11

For $m = o\left(\sqrt{N}\right)$,

$$\frac{(N-1)!}{(N-m-1)!N^m} = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\dots\left(1 - \frac{m}{N}\right) \to 1,$$

and since $e^{\frac{t}{\sqrt{g_N(m)}}} \to 1$,

$$m!\frac{i}{2\pi}\int_C (-u)^{-me^{\frac{t}{\sqrt{g_N(m)}}}-1}\left(1 - \left(\frac{-u}{N}\right)^{e^{\frac{t}{\sqrt{g_N(m)}}}}\right)^m e^{-u}du \to 1.$$

Hence, for $m = o\left(\sqrt{N}\right)$, the first term of $\psi_N(t) \to e^{t^2/2}$. We proceed by showing that for $m^2/N \to 0$ the error term $e^{\frac{t}{\sqrt{g_N(m)}}}e_N(m)$ will vanish. Indeed, again using Newton's binomium and (26)

$$\begin{aligned}
e_N(m) &\leq \frac{m!(N-1-m)!}{(N-1)!}\max_{1 \leq k \leq m} f_k(z) \sum_{k=0}^{m}\binom{m}{k}\frac{N^{kz-1}}{\Gamma(1+kz)} \\
&= \frac{m!(N-1-m)!}{(N-1)!}N^{mz-1}\max_{1 \leq k \leq m} f_k(z)\frac{i}{2\pi}\int_C(-u)^{-mz-1}\left(1 + \left(\frac{-u}{N}\right)^z\right)^m e^{-u}du.
\end{aligned}$$

As in the treatment of the main term for $m = o\left(\sqrt{N}\right)$, the expression

$$\frac{e^{\frac{t}{\sqrt{g_N(m)}}}e_N(m)}{N^{-1}\max_{1 \leq k \leq m} f_k(z)} \to \exp(t^2/2),$$

as $N \to \infty$. Hence, $e^{\frac{t}{\sqrt{g_N(m)}}}e_N(m) \to 0$ whenever $N^{-1}\max_{1 \leq k \leq m} f_k(z) \to 0$, which is again satisfied when $m = o\left(\sqrt{N}\right)$. This proves Theorem 2.3. $\qquad\square$

5   The expected weight $u_N(m) = \mathbb{E}\left[W_N(m)\right]$ of the SPT

In this section we prove Theorem 2.5. The proof consists of several steps which are sketched first.

As explained in Section 1.1, the SPT rooted at an arbitrary node to m uniformly chosen nodes in the complete graph, with *exponential* weights is a URT. As shown in [25], the discovery process of the nodes in this SPT is pure birth process with birth rate $\lambda_{n,n+1} = n(N-n)$.

Introduce for $1 \leq i \leq N-1$, the random variables $Y_i^{(N)}$ as the sum of the link weights in the SPT that i uniformly chosen nodes have in common. Obviously, $\mathbb{E}[Y_1^{(N)}] = u_N(1) = \mathbb{E}[W_N(1)]$ which is given by (3). By a similar argument as in [26, Theorem 3], the average weight $u_N(m)$ to $m > 1$ users can be obtained by inclusion/exclusion as

$$u_N(m) = \sum_{i=1}^{m}\binom{m}{i}(-1)^{i+1}\mathbb{E}[Y_i^{(N)}]. \tag{28}$$

We define the *father of multiple nodes* as the (unique) oldest common ancestor of these nodes. We prove below that

$$\mathbb{E}[Y_i^{(N)}] = \sum_{j=1}^{N}\sum_{l=1}^{j-1}\frac{1}{l(N-l)}q_N(j,i) \tag{29}$$

where

$$q_N(j,i) = \mathbb{P}(j^{\text{th}} \text{ node in URT is father of } i \text{ uniform nodes})$$
$$= \sum_{n=1}^{N-j+1} \frac{(j-1)(N-j)!(N-n-1)!}{(N-1)!(N-j-n+1)!} \frac{n!(N-i-1)!}{(n-i)!(N-1)!} \left[1 - \frac{1}{i}\frac{(n-i)}{n}\right]. \qquad (30)$$

Putting together (28), (29) and (30) yields an explicit expression for $u_N(m)$, from which (11) will follow after simplification.

Proof of Theorem 2.5: We start with a proof of (29). We have already seen that the additional time (weight) between the attachment of node l and node $l+1$ equals $\frac{1}{l(N-l)}$. This explains the factor $\sum_{l=1}^{j-1}\frac{1}{l(N-l)}$ and, hence, the sum of the link weights in the SPT that i uniform nodes have common nodes if the jth node is the father. We proceed with the proof of (30). By the law of total probability, we can write

$$q_N(j,i) = \sum_{n=1}^{N-j+1} \mathbb{P}(j^{\text{th}} \text{ node in URT is father of } i \text{ uniform nodes}, |\mathcal{T}_j^{(N)}| = n),$$

where $|\mathcal{T}_j^{(N)}|$ is the size of the subtree rooted at j. The probability

$$\mathbb{P}(j^{\text{th}} \text{ node in URT is father of } i \text{ uniform nodes}, |\mathcal{T}_j^{(N)}| = n)$$

equals

$$\mathbb{P}(|\mathcal{T}_j^{(N)}| = n) \cdot \mathbb{P}(i \text{ uniform nodes are in } \mathcal{T}_j^{(N)} \mid |\mathcal{T}_j^{(N)}| = n) \cdot [1 - p_n(i)],$$

where $p_n(i)$ is the probability that the paths from the root to i uniform nodes in a URT of size n share a common link. Each of the above three factors can be computed. We start with the second, which is the simplest one. Note that we may assume for $i > 1$ that $j > 1$, since the contribution from $j = 1$ in $\sum_{l=1}^{j-1}\frac{1}{l(N-l)}$ equals 0. Since we choose the uniform nodes unequal to the root of the URT, we have that

$$\mathbb{P}(i \text{ uniform nodes are in } \mathcal{T}_j^{(N)} \mid |\mathcal{T}_j^{(N)}| = n) = \frac{\binom{n}{i}}{\binom{N-1}{i}} = \frac{n!(N-i-1)!}{(n-i)!(N-1)!}.$$

We proceed with the first factor. We need to attach $n-1$ nodes to the tree rooted at j, and $N-j-n+1$ to the other $j-1$ nodes which leads to

$$\mathbb{P}(|\mathcal{T}_j^{(N)}| = n) = \binom{N-j}{n-1}\frac{(n-1)!(j-1)j\cdots(N-n-1)}{j(j+1)\cdots(N-1)} = \frac{(j-1)(N-j)!(N-n-1)!}{(N-1)!(N-j-n+1)!}.$$

We complete the determination of $q_N(j,i)$ by computing $p_N(i)$. In a URT of size N we call the set of nodes at distance k from the root as the level k set and denote its size by $U_N^{(k)}$. The following *basic property* of a URT, proved in [15, Lemma 2.1], will be used: Let $\{U_N^{(k)}\}_{k,N\geq 0}$ and $\{V_N^{(k)}\}_{k,N\geq 0}$ be two independent copies of the sizes of the level sets of two sequences of independent recursive trees. Then the vector $\{Z_N^{(k)}\}_{k\geq 0}$ of the size of the level sets in the URT of size N obeys

$$\{Z_N^{(k)}\}_{k\geq 0} \stackrel{d}{=} \{U_{N_1}^{(k-1)} + V_{N-N_1}^{(k)}\}_{k\geq 0}, \qquad (31)$$

where on the right-hand side N_1 is uniformly distributed over the set $\{1, 2, \ldots, N-1\}$, independently of $\{U_N^{(k)}\}$ and $\{V_N^{(k)}\}$. The above property can intuitively be understood as follows. The recursive tree of size N can be divided into two subtrees, namely, the tree of nodes which are in hops closest to

node 1 (the root), and the ones that are closest to node 2 (the first one that is attached to the root). The basic property states that both the tree connected to node 2 and the tree of nodes connected to the root form a URT. The first tree has size N_1 and the second tree has size $N - 1 - N_1$, where N_1 is uniform over $\{1, 2, \ldots, N-1\}$. Given N_1, the two trees are independent. Thus,

$$
\begin{aligned}
p_N(i) &= \frac{1}{N-1} \sum_{k=1}^{N-1} \mathbb{P}(i \text{ uniform nodes in tree rooted at 2 of size } k) \\
&+ \frac{1}{N-1} \sum_{k=1}^{N-1} \mathbb{P}(i \text{ uniform nodes are outside tree rooted at 1 of size } N - k, \text{ share link}) \\
&= \frac{1}{N-1} \sum_{k=i}^{N-1} \frac{k(k-1)\cdots(k-i+1)}{N(N-1)\cdots(N-i+1)} + \frac{1}{N-1} \sum_{k=1}^{N-1} p_k(i) \frac{k(k-1)\cdots(k-i+1)}{N(N-1)\cdots(N-i+1)}.
\end{aligned}
$$

Defining $\alpha_N(i) = N(N-1)\cdots(N-i+1)p_N(i)$, we have that

$$
\alpha_N(i) = \frac{1}{N-1} \sum_{k=i}^{N-1} \frac{k!}{(k-i)!} + \frac{1}{N-1} \sum_{k=1}^{N-1} \alpha_k(i).
$$

Subtraction yields

$$
(N-1)\alpha_N(i) - (N-2)\alpha_{N-1}(i) = \frac{(N-1)!}{(N-1-i)!} + \alpha_{N-1}(i),
$$

so that

$$
\alpha_N(i) = \frac{(N-2)!}{(N-1-i)!} + \alpha_{N-1}(i).
$$

Iteration yields

$$
\alpha_N(i) = \sum_{j=1}^{l} \frac{(N-1-j)!}{(N-i-j)!} + \alpha_{N-l}(i).
$$

Together with the fact that $\alpha_i(i) = 0$, we end up with

$$
\alpha_N(i) = \sum_{j=1}^{N-i} \frac{(N-1-j)!}{(N-i-j)!} = \sum_{l=i-1}^{N-2} \frac{l!}{(l+1-i)!} = \frac{1}{i}\frac{(N-1)!}{(N-i-1)!}.
$$

Thus,

$$
p_N(i) = \frac{(N-i)!}{N!}\alpha_N(i) = \frac{N-i}{iN}.
$$

Combining all terms, proves (30).

We now proceed to prove (11). Substitution into (28) now yields

$$
\begin{aligned}
u_N(m) &= mu_N(1) + \sum_{i=2}^{m} \binom{m}{i}(-1)^{i+1} \sum_{j=1}^{N}\sum_{l=1}^{j-1} \frac{1}{l(N-l)} \\
&\times \sum_{n=1}^{N-j+1} \frac{(j-1)(N-j)!(N-n-1)!}{[(N-1)!]^2(N-j-n+1)!} \frac{n!(N-i-1)!}{(n-i)!}\left[1 - \frac{1}{i}\frac{(n-i)}{n}\right]. \quad (32)
\end{aligned}
$$

The above four-fold sum can be reordered as

$$
\begin{aligned}
u_N(m) &= mu_N(1) + \sum_{i=2}^{m} \binom{m}{i}(-1)^{i+1} \sum_{n=1}^{N-1} \frac{(N-n-1)!n!(N-i-1)!}{[(N-1)!]^2(n-i)!}\left[1 - \frac{1}{i}\frac{(n-i)}{n}\right] \\
&\times \sum_{l=1}^{N-n} \frac{1}{l(N-l)} \sum_{j=l+1}^{N-n+1} \frac{(j-1)(N-j)!}{(N-j-n+1)!}. \quad (33)
\end{aligned}
$$

14

The last sum equals

$$\sum_{j=l+1}^{N-n+1} \frac{(j-1)(N-j)!}{(N-j-n+1)!} = \frac{1}{n(n+1)} \frac{(N-l)!}{(N-l-n)!} [(N-n)+nl].$$

Using [26, Lemma 10]

$$\sum_{n=1}^{N} \frac{(K-n)!}{(N-n)!} \frac{1}{n} = \frac{K!}{N!} [\psi(K+1) - \psi(K-N+1)], \tag{34}$$

we can next compute the l-sum and obtain

$$\sum_{l=1}^{N-n} \frac{1}{l(N-l)} \sum_{j=l+1}^{N-n+1} \frac{(j-1)(N-j)!}{(N-j-n+1)!} = \frac{1}{n(n+1)} \frac{(N-1)!}{(N-n-1)!} [\psi(N) - \psi(n) + 1].$$

Since

$$\sum_{n=1}^{N-1} \frac{\psi(N) - \psi(n) + 1}{n(n+1)} = \sum_{n=1}^{N-1} \frac{1}{n}, \tag{35}$$

the term $mu_N(1) = m\mathbb{E}[W_N(1)]$ given in (3) equals the contribution due to $i=1$ in (33). Substitution and changing the order of summation leads to

$$u_N(m) = \sum_{n=1}^{N-1} \frac{(n-1)!}{(N-1)!} \frac{\psi(N) - \psi(n) + 1}{n+1} v_N(m) \tag{36}$$

where

$$v_N(m) = \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!} \left[1 - \frac{n-i}{in}\right],$$

With $\binom{m}{i} - \binom{m-1}{i} = \binom{m}{i}\frac{i}{m}$ and $i \cdot \left[1 - \frac{n-i}{in}\right] = n - \frac{(n+1)(n-i)}{n}$, and defining

$$\Delta v_N(m) = v_N(m) - v_N(m-1),$$

we find that

$$m\Delta v_N(m) = \sum_{i=1}^{m} m \left\{\binom{m}{i} - \binom{m-1}{i}\right\} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!} \left[1 - \frac{n-i}{in}\right]$$

$$= n \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!} - \frac{n+1}{n} \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i-1)!}$$

Using an instance of Vandermonde's convolution formula [21, p. 8],

$$\sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!} = -(N-n-1)! \binom{N-m-1}{n} + \frac{(N-1)!}{n!}$$

and similarly

$$\sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i-1)!} = -(N-n)! \binom{N-m-1}{n-1} + \frac{(N-1)!}{(n-1)!}$$

we arrive at

$$m\Delta v_N(m) = (N-n-1)! \binom{N-m-1}{n} \left(\frac{N+nm-n}{N-m-n}\right) - \frac{(N-1)!}{n!}.$$

15

Relation (36) shows that the difference $\Delta u_N(m) = u_N(m) - u_N(m-1)$ is directly written in terms of the difference $\Delta v_N(m)$,

$$
\begin{aligned}
m\Delta u_N(m) &= \sum_{n=1}^{N-1} \frac{(n-1)!}{(N-1)!} \frac{\psi(N)-\psi(n)+1}{n+1} \left[(N-n-1)! \binom{N-m-1}{n} \left(\frac{N+nm-n}{N-m-n} \right) - \frac{(N-1)!}{n!} \right] \\
&= \frac{(N-m-1)!}{(N-1)!} \sum_{n=1}^{N-1} \frac{\psi(N)-\psi(n)+1}{n(n+1)} \left[\frac{(N-n-1)!(N+mn-n)}{(N-m-n)!} \right] - \sum_{n=1}^{N-1} \frac{1}{n}.
\end{aligned}
$$

Using the identity (39) proved in Appendix B yields

$$
m\Delta u_N(m) = \frac{N\psi(N) - m\psi(m)}{N-m} - \sum_{n=1}^{N-1} \frac{1}{n} = \frac{m}{N-m} \sum_{k=m}^{N-1} \frac{1}{k},
$$

or $\Delta u_N(m) = \frac{1}{N-m} \sum_{k=m}^{N-1} \frac{1}{k}$ which is equivalent to (11). This completes the proof of Theorem 2.5. $\square$

6   Proof of Corollaries 2.6 and 2.8

We will first prove (12) by induction on (11). We also have a second proof by computing the sums in (36), which we will omit here.

The induction is initiated by $N = 2$, for which both sides give 1. The inductive step from N to $N+1$ uses the induction hyptothesis as follows:

$$
\begin{aligned}
\sum_{k=1}^{N} \frac{1}{k^2} &= \sum_{k=1}^{N-1} \frac{1}{k^2} + \frac{1}{N^2} = \sum_{k=1}^{N-1} \frac{1}{N-k} \sum_{n=k}^{N-1} \frac{1}{n} + \frac{1}{N^2} = \sum_{l=2}^{N} \frac{1}{N-l+1} \sum_{n=l-1}^{N-1} \frac{1}{n} + \frac{1}{N^2} \\
&= \sum_{l=1}^{N} \frac{1}{N-l+1} \sum_{n=l}^{N} \frac{1}{n} - \frac{2}{N} \sum_{n=1}^{N-1} \frac{1}{n} + \sum_{l=1}^{N-1} \frac{1}{l} \frac{1}{N-l} = \sum_{k=1}^{N} \frac{1}{N+1-k} \sum_{n=k}^{N} \frac{1}{n} \\
&= u_{N+1}(N)
\end{aligned}
$$

since $\sum_{l=1}^{N-1} \frac{1}{l} \frac{1}{N-l} = \frac{2}{N} \sum_{n=1}^{N-1} \frac{1}{n}$. This proves (12). $\square$

We proceed by proving the statements in Corollary 2.8. For (14), we note that the upper bound follows from the bound $W_{\mathrm{Steiner},N}(m) \leq W_N(m)$ and (11). The lower bound follows immediately from (5), since $W_{\mathrm{Steiner},N}(m) \geq \frac{m}{N} \log \frac{N}{m+1}(1 + o(1))$ with probability converging to one.

For the first statement in (15), we again use that $W_{\mathrm{Steiner},N}(m) \leq W_N(m)$, so that $|W_{\mathrm{Steiner},N}(m) - W_N(m)| = W_N(m) - W_{\mathrm{Steiner},N}(m)$. Since both have asymptotically the same expectation by (14), we have that $[W_{\mathrm{Steiner},N}(m) - W_N(m)]/u_N(m)$ converges to 0 in L^1, which in turn implies the convergence in probability in the first part of (15). The second statement follows since

$$
\frac{W_N(m)}{\frac{m}{N} \log \frac{N}{m+1}} = \frac{W_{\mathrm{Steiner},N}(m)}{\frac{m}{N} \log \frac{N}{m+1}} + \frac{W_N(m) - W_{\mathrm{Steiner},N}(m)}{\frac{m}{N} \log \frac{N}{m+1}}.
$$

The first term converges to 1 in probability by the main result (5) in [3] and the second term to 0 in probability, so that $\frac{W_N(m)}{\frac{m}{N} \log \frac{N}{m+1}}$ converges to 1 in probability. $\square$

A Solving the recursion for $\varphi_{H_N(m)}(z)$

In this section, we solve the recursion relation (16) for $\varphi_{H_N(m)}(z)$ for $m = 1, 2$ and 3, which suggests (6). For $m = 1$, relation (6) equals (1). For the other extreme $m = N-1$ and with $\varphi_{H_N(N)}(z) = 0$, the recursion (16) reduces to $\varphi_{H_N(N-1)}(z) = z\varphi_{H_{N-1}(N-2)}(z)$ with solution $\varphi_{H_N(N-1)}(z) = z^{N-1}$. This result, needed for the initial values below, directly follows from $\varphi_{H_N(N-1)}(z) = \mathbb{E}\left[z^{H_N(N-1)}\right]$ because $H_N(N-1) = N-1$.

For $m = 2$, using (1), the recursion becomes

$$\varphi_{H_N(2)}(z) = \frac{(N-3)(N-1+2z)}{(N-1)^2}\varphi_{H_{N-1}(2)}(z) + \frac{4z}{(N-1)(N-2)}\left(\frac{\Gamma(z+N-1)}{(N-1)!\Gamma(z+1)} - \frac{1}{N-1}\right),$$

with initial value $\varphi_{H_3(2)}(z) = z^2$. Iteration yields that $\varphi_{H_N(2)}(z)$ equals

$$\frac{\prod_{j=1}^{p}(N-2-j)}{\prod_{j=1}^{p}(N-j)^2}\prod_{j=1}^{p}(N-j+2z)\varphi_{H_{N-p}(2)}(z)$$

$$+4z\sum_{k=1}^{p}\frac{\prod_{j=3}^{k+1}(N-j)}{(N-k)(N-k-1)\prod_{j=1}^{k-1}(N-j)^2}\prod_{j=1}^{k-1}(N-j+2z)\left(\frac{\Gamma(z+N-k)}{(N-k)!\Gamma(z+1)} - \frac{1}{N-k}\right).$$

With $\varphi_{H_3(2)}(z) = z^2$, which is reached for $N-p = 3$ or $p = N-3$, the iteration ends with result

$$\varphi_{H_N(2)}(z) = \frac{\prod_{j=1}^{N-3}(N-2-j)}{\prod_{j=1}^{N-3}(N-j)^2}\prod_{j=1}^{N-3}(N-j+2z)z^2$$

$$+4z\sum_{k=1}^{N-3}\frac{\prod_{j=3}^{k+1}(N-j)}{(N-k)(N-k-1)\prod_{j=1}^{k-1}(N-j)^2}\prod_{j=1}^{k-1}(N-j+2z)\left(\frac{\Gamma(z+N-k)}{(N-k)!\Gamma(z+1)} - \frac{1}{N-k}\right)$$

$$= \frac{4(N-3)!}{((N-1)!)^2}z^2\frac{\Gamma(N+2z)}{\Gamma(3+2z)} + 4z\frac{(N-3)!\Gamma(N+2z)}{((N-1)!)^2}\sum_{s=3}^{N-1}\frac{(s-1)!}{\Gamma(s+1+2z)}\left(\frac{\Gamma(z+s)}{(s-1)!\Gamma(z+1)} - 1\right).$$

Furthermore,

$$\sum_{s=3}^{N-1}\frac{(s-1)!}{\Gamma(s+1+2z)}\left(\frac{\Gamma(z+s)}{(s-1)!\Gamma(z+1)} - 1\right) = \frac{1}{\Gamma(z+1)}\sum_{s=3}^{N-1}\frac{\Gamma(z+s)}{\Gamma(s+1+2z)} - \sum_{s=3}^{N-1}\frac{\Gamma(s)}{\Gamma(s+1+2z)},$$

and with [15, Lemma A.4]

$$\sum_{k=a}^{b}\frac{\Gamma(k+x)}{\Gamma(k+y)} = \frac{1}{1+x-y}\left(\frac{\Gamma(1+b+x)}{\Gamma(b+y)} - \frac{\Gamma(a+x)}{\Gamma(a-1+y)}\right), \tag{37}$$

we obtain after some manipulations

$$\varphi_{H_2(N)}(z) = \frac{2}{(N-1)!(N-1)(N-2)}\left(\frac{\Gamma(N+2z)}{\Gamma(1+2z)} - 2\frac{\Gamma(N+z)}{\Gamma(z+1)} + \Gamma(N)\right). \tag{38}$$

This proves (6) for $m = 2$.

For $m = 3$, we obtain from (16) that

$$\varphi_{H_N(3)}(z) = \frac{(N-4)(N-1+3z)}{(N-1)^2}\varphi_{H_{N-1}(3)}(z) + \frac{9z}{(N-1)^2}\varphi_{H_{N-1}(2)}(z).$$

17

Using (38), we obtain

$$
\varphi_{H_N(3)}(z) = \frac{(N-4)(N-1+3z)}{(N-1)^2}\varphi_{H_{N-1}(3)}(z)
$$
$$
+\frac{18z}{(N-1)^2(N-2)(N-3)}\left(\frac{\Gamma(N-1+2z)}{(N-2)!\Gamma(1+2z)}-2\frac{\Gamma(N-1+z)}{(N-2)!\Gamma(z+1)}+1\right),
$$

with initial value $\varphi_{H_4(3)}(z) = z^3$. Similarly as for $m = 2$, we observe by iteration that

$$
\varphi_{H_N(3)}(z) = \frac{\prod_{j=1}^p(N-3-j)}{\prod_{j=1}^p(N-j)^2}\prod_{j=1}^p(N-j+3z)\varphi_{H_{N-p}(3)}(z)
$$
$$
+18z\sum_{k=1}^p\frac{\prod_{j=1}^{k-1}(N-3-j)}{(N-k)^2(N-k-1)(N-k-2)\prod_{j=1}^{k-1}(N-j)^2}\prod_{j=1}^{k-1}(N-j+3z)
$$
$$
\times\left(\frac{\Gamma(N-k+2z)}{(N-k-1)!\Gamma(1+2z)}-2\frac{\Gamma(N-k+z)}{(N-k-1)!\Gamma(z+1)}+1\right),
$$

and with $\varphi_{H_4(3)}(z) = z^3$, which is reached for $N - p = 4$ or $p = N - 4$, the iteration ends with result

$$
\varphi_{H_N(3)}(z) = \frac{\prod_{j=1}^{N-4}(N-3-j)}{\prod_{j=1}^{N-4}(N-j)^2}\prod_{j=1}^{N-4}(N-j+3z)z^3
$$
$$
+18z\sum_{k=1}^{N-4}\frac{\prod_{j=1}^{k-1}(N-3-j)}{(N-k)^2(N-k-1)(N-k-2)\prod_{j=1}^{k-1}(N-j)^2}\prod_{j=1}^{k-1}(N-j+3z)
$$
$$
\times\left(\frac{\Gamma(N-k+2z)}{(N-k-1)!\Gamma(1+2z)}-2\frac{\Gamma(N-k+z)}{(N-k-1)!\Gamma(z+1)}+1\right).
$$

We rewrite the above as for $m = 2$ as

$$
\varphi_{H_N(3)}(z) = \frac{(3!)^2(N-4)!}{((N-1)!)^2}\frac{\Gamma(N+3z)}{\Gamma(4+3z)}z^3 + 18z\frac{(N-4)!\Gamma(N+3z)}{((N-1)!)^2}\sum_{s=4}^{N-1}\frac{(s-1)!}{\Gamma(s+1+3z)}
$$
$$
\times\left(\frac{\Gamma(s+2z)}{(s-1)!\Gamma(1+2z)}-2\frac{\Gamma(s+z)}{(s-1)!\Gamma(z+1)}+1\right).
$$

Again using (37), we finally find

$$
\varphi_{H_N(3)}(z) = \frac{18(N-4)!}{((N-1)!)^2}\left[\frac{\Gamma(N+3z)}{\Gamma(4+3z)}\left(2z^3+\frac{\Gamma(4+2z)}{\Gamma(1+2z)}-\frac{\Gamma(4+z)}{\Gamma(1+z)}+2\right)-\frac{\Gamma(N+2z)}{\Gamma(1+2z)}+\frac{\Gamma(N+z)}{\Gamma(1+z)}-\frac{(N-1)!}{3}\right]
$$
$$
= \frac{3!}{(N-1)!(N-1)(N-2)(N-3)}\left[\frac{\Gamma(N+3z)}{\Gamma(1+3z)}-3\frac{\Gamma(N+2z)}{\Gamma(1+2z)}+3\frac{\Gamma(N+z)}{\Gamma(1+z)}-\Gamma(N)\right],
$$

which proves (6) for $m = 3$. This derivation suggests the general formula (6).

B   Proof of the identity (39)

Lemma B.1

$$
\sum_{n=1}^{N-1}\frac{\psi(N)-\psi(n)+1}{n(n+1)}\left[\frac{(N-n-1)!(N+mn-n)}{(N-m-n)!}\right] = \frac{(N-1)!}{(N-m)!}\left(N\psi(N)-m\psi(m)\right). \qquad (39)
$$

Proof: Use $\frac{N+mn-n}{n(n+1)} = \frac{N}{n} - \frac{N-m+1}{n+1}$, and the definitions

$$LS_1 \;=\; N\sum_{n=1}^{N-1}\frac{\psi(N)-\psi(n)+1}{n}\frac{(N-n-1)!}{(N-n-m)!},$$

$$LS_2 \;=\; (N-m+1)\sum_{n=1}^{N-1}\frac{\psi(N)-\psi(n)+1}{n+1}\frac{(N-n-1)!}{(N-n-m)!},$$

to express the left-hand side(LS) of (39) as the sum $LS_1 + LS_2$. With the definition of the digamma function we obtain

$$LS_1 = N\sum_{n=1}^{N-1}\frac{1}{n}\frac{(N-n-1)!}{(N-n-m)!} + N\sum_{k=1}^{N-1}\frac{1}{k}\sum_{n=1}^{N-1}\frac{1}{n}\frac{(N-n-1)!}{(N-n-m)!} - N\sum_{n=1}^{N-1}\sum_{k=1}^{n-1}\frac{1}{k}\frac{1}{n}\frac{(N-n-1)!}{(N-n-m)!},$$

and similarly

$$LS_2 = (N-m+1)\sum_{n=1}^{N-1}\frac{1}{n+1}\frac{(N-n-1)!}{(N-n-m)!} + (N-m+1)\sum_{k=1}^{N-1}\frac{1}{k}\sum_{n=1}^{N-1}\frac{1}{n+1}\frac{(N-n-1)!}{(N-n-m)!}$$
$$-(N-m+1)\sum_{n=1}^{N-1}\sum_{k=1}^{n-1}\frac{1}{k}\frac{1}{n+1}\frac{(N-n-1)!}{(N-n-m)!}.$$

From (34) we obtain

$$\sum_{n=1}^{N-1}\frac{1}{n}\frac{(N-n-1)!}{(N-n-m)!} = \frac{(N-1)!}{(N-m)!}[\psi(N)-\psi(m)],$$

Using

$$\sum_{n=1}^{N}\frac{(K-n)!}{(N-n)!}\left(\frac{1}{n}\sum_{k=1}^{n-1}\frac{1}{k}\right) = -\frac{1}{2}\frac{K!}{N!}\left(\sum_{l=1}^{K}\frac{1}{l^2}-\sum_{l=1}^{K-N}\frac{1}{l^2}\right) + \frac{1}{2}\frac{K!}{N!}\left[\psi(K+1)-\psi(K-N+1)\right]^2$$

which follows from [24, pp. 252] after some manipulations, we have

$$\sum_{n=1}^{N-1}\sum_{k=1}^{n-1}\frac{1}{k}\frac{1}{n}\frac{(N-n-1)!}{(N-n-m)!} = -\frac{1}{2}\frac{(N-1)!}{(N-m)!}\left(\sum_{k=m}^{N}\frac{1}{k^2}\right) + \frac{1}{2}\frac{(N-1)!}{(N-m)!}[\psi(N)-\psi(m)]^2.$$

Substitution yields

$$LS_1 = \frac{N!(1+\psi(N))}{(N-m)!}[\psi(N)-\psi(m)] + \frac{1}{2}\frac{N!}{(N-m)!}\left(\sum_{k=m}^{N}\frac{1}{k^2}\right) - \frac{1}{2}\frac{N!}{(N-m)!}[\psi(N)-\psi(m)]^2.$$

Similarly

$$LS_2 \;=\; \frac{(N-m+1)(N-1)!}{(N-m)!}\psi(m) - \frac{N!\psi(N)}{(N-m)!}[\psi(N+1)-\psi(m)]$$
$$-\frac{1}{2}\frac{N!}{(N-m)!}\left(\sum_{k=m}^{N+1}\frac{1}{k^2}\right) + \frac{1}{2}\frac{N!}{(N-m)!}[\psi(N+1)-\psi(m)]^2.$$

Simplifying the sum $LS = LS_1 + LS_2$ gives the right-hand side of (39).

References

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

[2] B. Bollobás, *Random Graphs*, Cambridge University Press, second edition, 2001.

[3] B. Bollobás, D. Gamarnik, O. Riordan and B. Sudakov, "On the value of a random minimum weight Steiner tree", Combinatorica, Vol. 24, No. 2, pp. 187-207, April 2004.

[4] J. Chuang and M. Sirbu, "Pricing Multicast Communication: A Cost-Based Approach", Proceedings of the INET'98, 1998.

[5] R.P. Dobrow and R.T. Smythe, "Poisson approximations for functionals of random trees", Random Struct. and Alg., vol. 9, pp. 79-92, 1996.

[6] M. Faloutsos, P. Faloutsos and C. Faloutsos, "On power law relations of the Internet Topology", Proceedings of ACM SIGCOM'99, Cambridge Massachusetts, 251-262, 1999.

[7] A. M. Frieze, "On the Value of a random minimum spanning tree problem", Discrete Applied Mathematics, Vol. 10, pp. 47-56, 1985.

[8] A. M. Frieze and G. R. Grimmett, "The shortest-path problem for graphs with random arc-lengths", Discrete Applied Mathematics, Vol. 10, pp. 57-77, 1985.

[9] I.S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, Fifth Edition, 1994.

[10] R. Hekmat and P. Van Mieghem, "Degree Distribution and Hopcount in Wireless Ad-hoc Networks", 11th IEEE International Conference on Networks (ICON 2003), Sydney, Australia, Sept. 28-Oct. 1, 2003.

[11] M. Janic and P. Van Mieghem, "The Gain and Cost of Multicast Routing Trees", International Conference on Systems, Man and Cybernetics (IEEE SMC 2004), October 10-13, The Hague, The Netherlands, 2004.

[12] S. Janson, "One, Two and Three Times $\log n/n$ for Paths in a Complete Graph with Random Weights", Combinatorics, Probability and Computing, vol. 8, pp. 347-361, 1999.

[13] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, "First-passage percolation on the random graph", Probability Engineering Informational Science (PEIS), vol. 15, pp. 225-237, 2001.

[14] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, "The Flooding Time in Random Graphs", Extremes, vol. 5, No. 2, pp. 111-129, 2002.

[15] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, "On the covariance of the level sizes in random recursive trees", Random Structures and Algorithms, vol. 20, pp. 519-539, 2002.

[16] R. Hassin and E. Zemel, "On shortest paths in graphs with random weights", Mathematics of Operations Research, Vol. 10, No. 4, pp. 557-564, 1985.

[17] F.K. Hwang, D.S. Richards and P. Winter, *The Steiner tree problem,* Annals of Discrete Mathematics, **53**, North-Holland Publishing Co., Amsterdam, 1992.

[18] R.M. Karp, "On the computational complexity of combinatorial problems", Networks, vol. 5, pp. 45–68, 1975.

[19] J.W. Moon, "The distance between nodes in recursive trees", London Math. Soc. Lecture Notes Ser., **13**, Cambridge University Press, London, pp. 125-132, 1974

[20] G. Phillips, S. Shenker and H. Tangmunarunkit, "Scaling of multicast trees: Comments on the Chuang-Sirbu scaling law", Proc. ACM Sigcomm 1999.

[21] J. Riordan, *Combinatorial Identities,* Wiley, N. Y., 1968.

[22] S. Robinson, "The Prize of Anarchy ", SIAM News, Vol. 37, No. 5, pp. 1-4, June 2004.

[23] R.T. Smythe and H.M. Mahmoud, "A Survey of Recursive Trees", Theor. Probability and Math. Statist. vol. 51, pp. 1-27, 1995

[24] Srivastava, H. M. and J. Choi, *Series Associated with the Zeta and Related Functions*, Kluwer Academic Publishers, Dordrecht, 2001.

[25] P. Van Mieghem, G. Hooghiemstra and R. van der Hofstad, "A Scaling Law for the Hopcount", Report 2000125 (http://www.nas.ewi.tudelft.nl/people/Piet/teleconference.html).

[26] P. Van Mieghem, G. Hooghiemstra and R. van der Hofstad, "On the Efficiency of Multicast", IEEE/ACM Transactions on Networking, vol. 9, No. 6, pp. 719-732., 2001.

[27] P. Van Mieghem and M. Janic, "Stability of a Multicast Tree", Proceedings IEEE INFOCOM2002, Vol. 2, pp. 1099-1108, 2002.

[28] P. Van Mieghem, F.A. Kuipers, T. Korkmaz, M. Krunz, M. Curado, E. Monteiro, X. Masip-Bruin, J. Solé-Pareta and S. Sánchez-López, 2003, *Quality of Service Routing*, Chapter 3 in Quality of Future Internet Services, EU-COST 263 Final Report, edited by Smirnov et al. in Springer LNCS 2856, pp. 80-117.