Size and Weight of Shortest Path Trees with Exponential Link Weights

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

September 17, 2003

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, as well as to prove a central limit theorem for m fixed. We also obtain exact expressions for the average weight of the subtree.

1 Introduction

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

\[ W_{\text{Steiner},N}(m) = (1 + o(1)) \frac{m}{N} \log \frac{N}{m+1}, \]  

(1)

for large N and m = o(N). The Steiner tree is the minimal weight tree that connects m + 1 given nodes in a graph. Unfortunately, the computation of a Steiner tree in a graph with a specific link weight structure is a hard NP-complete problem [15]. Steiner tree problems arise in many applications. For example, the design of optimal networks attempts to use the least possible number of links between nodes. In simulations on current PC’s, the Steiner tree problem is limited to small graphs of size N around a few tens [22]. Since the exact computation of large Steiner trees is unfeasible, accurate heuristics for Steiner trees receive considerable attention. From a practical point of view it often suffices that the approximation is close enough, so that the Steiner tree itself needs not to be computed. There are many heuristics for Steiner trees and we refer to the book by Winter et al. [14] and the more recent review by Du et al. [6]. A number of heuristics is based on the shortest path tree (SPT).

In this article, we provide an exact analysis of the related SPT problem in the complete graph with i.i.d. exponential link weights with mean 1. The exact law of the average weight for the SPT spanned by m uniform nodes is presented which complements the asymptotic result (1).

∗rhofstad@win.tue.nl, Department of Mathematics and Computer Science, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands.

†G.Hooghiemstra@ewi.tudelft.nl and P.VanMieghem@ewi.tudelft.nl, Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands.
1.1 Previous work

Previously, we have investigated properties of the SPT in the class of random graphs $G_p(N)$ \cite{2} with uniformly or exponentially distributed link weights where $p$ is the probability that links are present. We denote this class of graphs by RGU. In \cite{20} and \cite{11}, we have shown that the problem of finding the shortest path between two arbitrary nodes can be rephrased as a birth process with rate asymptotically equal to $pn(N-n)$, where $n$ equals the number of already discovered nodes.

The resulting SPT rooted at an arbitrary node in the class RGU is asymptotically equal to a uniform recursive tree (URT). In the complete graph with exponential link weights, the SPT is exactly a URT. By the memoryless property of the exponential distribution, the laws of the birth process and of the URT are independent. URTs have received considerable attention \cite{18}, but the problem of finding the size of the subtree spanned by $m$ uniform nodes was still open.

For a URT, the hopcount $H_N$, which is defined as the number of links or hops from one arbitrary node to another arbitrary node, can be exactly determined (see e.g., \cite{16} and \cite{20}). The generating function of the number of hops in the URT from the root to a uniform node (different from the root) equals

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

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 is determined in \cite{12} as

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

We will now focus on the complete graph with exponential weights where $p = 1$. We denote by $W_N(m)$ the weight of the SPT to $m$ uniformly chosen nodes; $W_N(m)$ is the sum of all the link weights of the subtree spanned by the $m$ uniform nodes. The generating function of the weight of a shortest path to a uniform destination, defined as $\varphi_{W_N(1)}(t) = \mathbb{E}[e^{-tW_N(1)}]$, follows from \cite{12} as

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

where $W_k$ denotes the weight of the path to the $k$th attached node and $W_k = \sum_{n=1}^{k} E_n$, with $E_n$ an exponential random variable with parameter $n(N-n)$ and $E_n$ is independent from $E_k$ for all $k \neq n$. Indeed, the law $W_k$ equals the time that the birth process attaches the $k$th node. Hence, $W_k$ is the sum of the inter-attachment times $\tau_n$ between the inclusion the $n$-th and $(n+1)$-th node to the SPT for $n = 1, \ldots, k-1$. That inter-attachment time $\tau_n$ is exponential with parameter $n(N-n)$. The average length, obtained from (3), is equal to

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

Let us denote by $H_N(m)$ the number of links in the SPT to $m$ uniformly chosen nodes. The average number of links of the SPT $g_N(m) = E[H_N(m)]$ in the complete graph with exponential link weights to $m$ arbitrary nodes [21] is specified below by (8). The proof of (8) in [21] uses inclusion/exclusion on the number of links, and writes

$$g_N(m) = \sum_{i=1}^{m} \left( \begin{array}{c} m \nonumber \end{array} \right) (-1)^{i-1} E\left[ X_i^{(N)} \right],$$

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 (5) holds for any graph, and was also used in [21] 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 spanned by $m$ uniform and different locations.

Our interest in URTs was triggered by the hopcount problem [20] 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 RGU. We have shown that the hopcount of the shortest path in a complete graph ($p = 1$) is equal to the depth $D_{1,N}$ from the source to an arbitrary node in a URT of size $N$. In [11], 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 has been proved in [11] under the condition$^1$ that $Np_N/(\log N)^3 \to \infty$.

In [13], we have investigated the random probability distribution of nodes that are a given number of steps away from the root. By using the special structure of URTs, as we will do in this paper, we have shown that this random distribution is with probability one close to its expected value. This result supplements a theorem by Dobrow and Smythe [5] who show that the total variation distance between the law of $D_{1,N}$ and the Poisson distribution with mean $\lambda_N = E[D_{1,N}]$ is at most $C/\log N$.

The problem of investigating $H_N(m)$ is important to estimate the efficiency of multicast over unicast in the Internet. While unicast is a one-to-one communication, multicast is a one (or many)-to-many communication mode where a single message destined to $m$ different receivers is sent once by the sender and copied at each branch point of the multicast tree with leaves containing receivers. 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 is more efficient. The ratio $E[H_N(m)]/mE[H_N(1)]$ is regarded as a measure for the multicast efficiency. From Internet measurements, Chuang and Sirbu [4] observed that the multicast efficiency $E[H_N(m)]/mE[H_N(1)]$ decreases as a power of $m$, and they estimated the exponent by $-0.2$. Philips et al. [17] have dubbed this behaviour "the Chuang-Sirbu law". In [17] and [21], the multicast efficiency $E[H_N(m)]/mE[H_N(1)]$ was studied in more detail using random graph models. As observed in [21], for the complete graph with exponential weights, $E[H_N(m)]/mE[H_N(1)] = g_N(m)/g_N(1)$ with $g_N(m)$ specified by (8), which on a log-log scale and for $m = o(N)$ looks fairly linear confirming the observed Chuang-Sirbu law.

The complete graph with exponential weights seems a reasonable model for the multicast structure of the Internet. However, it is not a good model for the topology of the Internet. From measurements,

$^1$Computer simulations confirm the limit law even when $Np_N \to \infty$ at a slower rate.
Faloutsos et al. [7] observed that the degree distribution of Internet nodes obeys a power-law. On the other hand, 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 normal random variable with expectation and variance equal to $\log N$.

## 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 $1 \leq m \leq N - 1$,*

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

(6)

*Consequently,*

$$
P[H_N(m) = j] = \frac{m!(-1)^{N-(j+1)} S_N^{(j+1)} S_j^{(m)}}{(N - 1)!(N-1)^m},
$$

(7)

*where $S_N^{(j+1)}$ and $S_j^{(m)}$ 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 $m = 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 47$.](image)

Even though (7) completely determines the law of $H_N(m)$, the Stirling numbers complicate the computation of properties of the law of $H_N(m)$. Using another approach, we have computed the expectation and variance of $H_N(m)$.

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

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

(8)
and

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

For \(N = 1000\), Figure 2 illustrates the typical behavior for large \(N\) of the expectation \(g_N(m)\) and the standard deviation \(\sigma_N(m)\) of \(H_N(m)\) for all values of \(m\). Note that \(H_N(N - 1) = N - 1\), so that \(\text{Var}(H_N(N - 1)) = 0\).

![Figure 2: The average number of hops \(g_N(m)\) (left axis) and the corresponding standard deviation \(\sigma_N(m)\) (right axis) as function of \(m\) for the number of nodes equal to \(N = 1000\).](image)

For our final result on \(H_N(m)\), we denote by

\[
X_N(m) = \frac{H_N(m) - g_N(m)}{\sigma_N(m)},
\]

where \(\sigma_N^2(m) = \text{Var}(H_N(m))\), the standardized size of the subtree \(H_N(m)\). Figure 1 suggests that \(H_N(m)\) is asymptotically normal.

**Theorem 2.3** For all \(m = o(\sqrt{N})\), \(X_N(m)\) converges 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 stated 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 $1 \leq m \leq N - 1$,\n\[ u_N(m) = \sum_{j=1}^{m} \frac{1}{N-j} \sum_{k=j}^{N-1} \frac{1}{k} \] (11)

Corollary 2.6 For all $N \geq 2$,\n\[ u_N(N - 1) = \sum_{n=1}^{N-1} \frac{1}{n^2} \] (12)

Since $u_N(m) \leq u_N(N - 1) < \frac{s^2}{6}$, any SPT in the considered graph has an average weight smaller than 1.645. Moreover, $\mathbb{E}[W_{\text{Steiner},N}(N - 1)] = u_N(N - 1)$.

Corollary 2.7 For all $m = O(N^a)$ with $a < 1$,
\[ u_N(m) = \frac{m}{N} \log \frac{N}{m+1} + O\left(N^{2(a-1)} \log N\right) \] (13)

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

Corollary 2.8 For all $m = o(N)$,
\[ \mathbb{E}[W_{\text{Steiner},N}(m)] = (1 + o(1)) \frac{m}{N} \log \frac{N}{m+1} \] (14)

Consequently, for all $m = o(N)$,
\[ P(|W_{\text{Steiner},N}(m) - W_N(m)| > \epsilon u_N(m)) \rightarrow 0, \quad \text{and} \quad P(|W_N(m) - u_N(m)| > \epsilon u_N(m)) \rightarrow 0. \] (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 $m = o(N)$. The asymptotics of the expected value in (14) is not proved by Bollobás et al. [3], even though it may follow from the proof in [3].

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 $\frac{np_N}{(\log N)^3} \rightarrow \infty$ as shown in [11], we believe that the presented results are also applicable to this class of graphs. The interest of this generalization lies in the fact that Ad-Hoc wireless networks [9] and certain peer-to-peer networks are well modelled by random graphs.

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 $1 \leq m \leq N - 1$,\n\[ \varphi_{H_N(m)}(z) = \frac{(N-m-1)(N-1+mz)}{(N-1)^2} \varphi_{H_{N-1}(m)}(z) + \frac{m^2 z}{(N-1)^2} \varphi_{H_{N-1}(m-1)}(z). \] (16)
Figure 3: The several possible cases in which the $N^{th}$ 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

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

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 $H_{N-1}(m-1)$, else the generating function of the number of additional links equals

$$
\mathbb{E}_z L_{N-1}(m) = \frac{N-m}{N-m} \varphi_{H_{N-1}(m)}(z) + \frac{1}{N-1} \varphi_{H_{N-1}(m-1)}(z).
$$

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,

$$
\mathbb{E}_z L_{N-1}(m) = \frac{m-1}{N-1} \varphi_{H_{N-1}(m)}(z) + \frac{N-m}{N-1} \left(\frac{N-m-1}{N-m} \varphi_{H_{N-1}(m)}(z) + \frac{1}{N-m} \varphi_{H_{N-1}(m-1)}(z)\right) + \frac{m}{N-1} \varphi_{H_{N-1}(m-1)}(z).
$$

Substitution of (18) into (17) leads (16). □

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 value \( r = r(e) \), then the generating function \( \varphi_{R_N(m)}(z) \) of \( R_N(m) = \sum_{e \in T} r(e) \) satisfies the recursion

\[
\varphi_{R_N(m)}(z) = \frac{(N - m - 1)(N - 1 + m^2 \varphi_r(z))}{(N - 1)^2} \varphi_{R_{N-1}(m)}(z) + \frac{m^2 \varphi_r(z)}{(N - 1)^2} \varphi_{R_{N-1}(m-1)}(z),
\]

where \( \varphi_r(z) = \mathbb{E}[z^r] \) is the generating function of the i.i.d. common link value \( r \). Examples are the monetary usage cost of a link, the total number of lost packets, the physical length of the link, etc.

The proof of (19) is straightforward, the major changes are that equation (17) is replaced by

\[
\varphi_{R_N(m)}(z) = \left(1 - \frac{m}{N - 1}\right) \varphi_{R_{N-1}(m)}(z) + \frac{m}{N - 1} \mathbb{E}[z^{Q_{N-1}(m)+r}]
\]

\[
= \left(1 - \frac{m}{N - 1}\right) \varphi_{R_{N-1}(m)}(z) + \frac{m \varphi_r(z)}{N - 1} \mathbb{E}[z^{Q_{N-1}(m)}],
\]

where \( Q_{N-1}(m) \) is the corresponding sum 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, and where the second equality follows from the independence of \( r \) and \( Q_{N-1}(m) \). Furthermore, we have to replace (18) by

\[
\mathbb{E}[z^{Q_{N-1}(m)}] = \frac{m}{N - 1} \varphi_{R_{N-1}(m-1)}(z) + \frac{N - m - 1}{N - 1} \varphi_{R_{N-1}(m)}(z),
\]

which completes the proof of (19). Clearly,

\[
\varphi_{R_N(m)}(z) = \varphi_{H_N}(\varphi_r(z)).
\]

Since \( g_N(m) = \mathbb{E}[H_N(m)] = \varphi_{H_N}(1) \) we obtain the recursion for \( g_N(m) \),

\[
g_N(m) = \left(1 - \frac{m^2}{(N - 1)^2}\right) g_{N-1}(m) + \frac{m^2}{(N - 1)^2} g_{N-1}(m - 1) + \frac{m}{N - 1}.
\]

### 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 readily 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

\[
\varphi_{H_N}(m)(z) = \frac{m!N(N - 1 - m)!}{(N - 1)!} \sum_{k=0}^m \binom{m}{k} \frac{(-1)^{m-k} \Gamma(N + k z)}{N! \Gamma(1 + k z)} - \frac{1}{N},
\]

\[
= \frac{m!N(N - 1 - m)!}{(N - 1)!} \sum_{k=0}^m \binom{m}{k} (-1)^{m-k} \sum_{j=1}^{N-1} \frac{(-1)^{N-j+1} S_{N}^{(j+1)} k^j z^j}{N!}
\]

\[
= \frac{m!N(N - 1 - m)!}{(N - 1)!} \sum_{j=1}^{N-1} (-1)^{N-j+1} S_{N}^{(j+1)} \frac{1}{N!} \left( \sum_{k=0}^m \binom{m}{k} (-1)^{m-k} k^j \right) z^j.
\]
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],

\[
ml_S^{(m)}(j) = \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} k^j,
\]

for which \( S_j^{(m)} = 0 \) if \( j < m \), gives

\[
\varphi_{H_N(m)}(z) = \frac{(m!)^2 (N-1-m)!}{((N-1)!)^2} \sum_{j=1}^{N-1} (-1)^{N-(j+1)} S_{N-1}^{(j+1)} S_j^{(m)} z^j.
\]

This proves (7), and completes the proof of Theorem 2.1. □

**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

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

Indeed,

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

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

We will start with the former. Using \( \partial_t^j (1-t)^j \big|_{t=1} = j! (-1)^j \delta_{i,j} \) 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 \( \mathbb{E}[H_N(m) \cdot (H_N(m) - 1)] \). 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,

\[
m(m-1)(-1)^{m-2} \partial_t^{N-1} \left[ t^{N+1} \log^2(t)(1-t)^{m-2} \right]_{t=1} = \left( \frac{N-1}{m-2} \right) m(m-1)(m-2)! \partial_t^{N-m+1} \left[ t^{N+1} \log^2(t) \right]_{t=1} = (N+1)! \left( \frac{N-1}{m-2} \right) \left[ \left( \frac{N+1}{m-1} \right)^2 - \sum_{k=m+1}^{N} \frac{1}{k^2} \right].
\]
Denote the upper bound on the right-hand side of (26) by 

\[ m(-1)^{m-1} \partial^N [t^N \log^2 t(1-t)^{m-1}]_{t=1} = \binom{N-1}{m-1} m(m-1)! \partial^N [t^N \log^2 t]_{t=1} \]

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

Substitution into (23) leads to 

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

From \( g_N(m) = \mathbb{E}[H_N(m)] \) and \( \text{Var}(H_N(m)) = \mathbb{E}[H_N(m)(H_N(m)-1)] + g_N(m) - g_N^2(m) \), 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. Denote by \( \mu = \mathbb{E}[r] \) and by \( \sigma^2 = \text{Var}(r) \), then 

\[ \mathbb{E}[R_N(m)] = \mu \mathbb{E}[H_N(m)], \quad \text{and} \quad \text{Var}(R_N(m)) = \sigma^2 \mathbb{E}[H_N(m)] + \mu^2 \text{Var}(H_N(m)). \]

We close this section with the 

**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 

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

Denote by \( \psi_N(t) \) the moment generating function of this quantity, then 

\[ \psi_N(t) = e^{-t\sqrt{g_N(m)}} \cdot \mathbb{E} \left[ \exp \left( \frac{t H_N(m)}{\sqrt{g_N(m)}} \right) \right] = e^{-t\sqrt{g_N(m)}} \varphi_H(m) \left( \frac{-i}{\sqrt{g_N(m)}} \right). \]

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 

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

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

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

Denote the upper bound on the right-hand side of (26) by \( e_N(m) \), then for \( N \to \infty \), 

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

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

\[ \frac{1}{\Gamma(1+kz)} = \frac{i}{2\pi} \int_C (-u)^{-(1+kz)} e^{-u} du, \]
and obtain

\[ \varphi_{H_N(m)}(z) = \frac{m!(N-1-m)!}{(N-1)!} \sum_{k=0}^{m} \binom{m}{k} (N^k) (-1)^{m-k} \frac{i}{2\pi} \int_C (-u)^{-1-kz} e^{-u} du + O(e_N(m)) \]

Further, because \( m \rightarrow N \) as \( N \rightarrow \infty \), we have

\[ \varphi_{H_N(m)}(z) \rightarrow 1 \]

We substitute (28) into (25) and compute the appearing factors separately. First, rewrite

\[ e^{-i \sqrt{g_N(m)} \frac{t}{N}} \frac{(N-1-m)!}{(N-1)!} = \frac{(N-1-m)!}{(N-1)!} \cdot N^m \cdot e^{-i \sqrt{g_N(m)} \frac{t}{N} (e^{\sqrt{g_N(m)}} - 1)} \]

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

\[ e^{-i \sqrt{g_N(m)} \frac{t}{N}} N^m \left( e^{\sqrt{g_N(m)}} - 1 \right) = \exp \left( -t \sqrt{g_N(m)} + m \log N \left( e^{\sqrt{g_N(m)}} - 1 \right) \right) \]

\[ = \exp \left( -t \sqrt{g_N(m)} + g_N(m) \left( \frac{t}{\sqrt{g_N(m)}} + \frac{t^2}{2g_N(m)} + O \left( \frac{t^3}{g_N^2(m)} \right) \right) \right) \]

\[ = e^{\frac{t^2}{4}} \left( 1 + O \left( \frac{1}{\sqrt{g_N(m)}} \right) \right). \]

For \( m^2/N \to 0 \),

\[ \frac{(N-1)!}{(N-m-1)!N^m} = \left( 1 - \frac{1}{N} \right) \left( 1 - \frac{2}{N} \right) \ldots \left( 1 - \frac{m}{N} \right) \to 1, \]

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

\[ \frac{m! \frac{i}{2\pi} \int_C (-u)^{-m \sigma_{N(m)}} \left( 1 - \frac{u}{N} \right) e^{\sigma_{N(m)} t} N^m \left( e^{\sqrt{g_N(m)}} - 1 \right) e^{-u} du}{1} \]

Hence, for \( m^2/N \to 0 \), the first term of \( \psi(t) \to e^{t^2/2} \). We proceed by showing that for \( m^2/N \to 0 \) the error term \( e^{\sqrt{g_N(m)}} e_N(m) \) will vanish. Indeed, again using Newton’s binomium and (27)

\[ e_N(m) \leq \frac{m!(N-1-m)!}{(N-1)!} \max_{1 \leq k \leq m} f_k(z) \sum_{k=0}^{m} \binom{m}{k} N^{k-1} \frac{1}{\Gamma(1+kz)} \]

\[ = \frac{m!(N-1-m)!}{(N-1)!} N^{m-1} \max_{1 \leq k \leq m} f_k(z) \frac{i}{2\pi} \int_C (-u)^{-mz-1} \left( 1 + \left( \frac{-u}{N} \right)^z \right) e^{-u} du. \]

As in the treatment of the main term, the expression

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

as \( N \to \infty \). Hence, \( e^{\sqrt{g_N(m)}} e_N(m) \to 0 \) whenever \( N^{-1} \max_{1 \leq k \leq m} f_k(z) \to 0 \), which is satisfied when \( m^2/N \to 0 \). This proves Theorem 2.3. \( \square \)
5 The expected weight \( u_N(m) = \mathbb{E}[W_N(m)] \) of the SPT

In this section we will prove Theorem 2.5. The proof consists of several steps which will be 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 [20], the discovery process of the nodes in this SPT is pure birth process with birth rate \( \lambda_n = n(N - n) \).

Introduce for \( 1 \leq i \leq N - 1 \), the random variables \( Y_i^{(N)} \) as the sum of the link weights in the SPT to \( i \) uniformly chosen nodes. Obviously, \( \mathbb{E}[Y_i^{(N)}] = u_N(1) = \mathbb{E}[W_N(1)] \) which is given by (4). By a similar argument as in [21, Theorem 3], the average weight \( u_N(m) \) to \( m > 1 \) users can be obtained by inclusion/exclusion

\[
u_N(m) = \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \mathbb{E}[Y_i^{(N)}]. \tag{29}\]

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

\[
\mathbb{E}[Y_i^{(N)}] = \sum_{j=1}^{N} \sum_{l=1}^{j-1} \frac{1}{(N-l)} q_N(j, i) \tag{30}
\]

where, as we will show below,

\[
q_N(j, i) = \mathbb{P}(j^{\text{th}} \text{ node in URT is father of } i \text{ uniform nodes})
= \sum_{n=1}^{N-j+1} \frac{(j-1)(N-j)!((N-n-1)!}{(N-1)!((N-j-n+1)!} \frac{n!(N-i-1)!}{(n-i)!(N-1)!} \left[ 1 - \frac{1}{i} \right]. \tag{31}
\]

Putting together (29), (30) and (31) yields an explicit expression for \( u_N(m) \), from which (11) via (??) will follow after simplification.

Proof of Theorem 2.5: We start with a proof of (30). We have already seen that the additional time (weight) between the attachment of node \( l \) and node \( l + 1 \) equals \( \frac{1}{n(N-l)} \). This explains the factor \( \sum_{l=1}^{j-1} \frac{1}{n(N-l)} \) and hence the weight of the common nodes if the \( j^{\text{th}} \) node is the father. We proceed with the proof of (31). By the law of total probability, we can write

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

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

\[
\mathbb{P}(j^{\text{th}} \text{ node in URT is father of } i \text{ uniform nodes, } |T_j^{(N)}| = n)
\]
equals
\[
\mathbb{P}(|T_j^{(N)}| = n) \cdot \mathbb{P}(i \text{ uniform nodes are in } T_j^{(N)}| |T_j^{(N)}| = n) \cdot [1 - p_n(i)],
\]

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 \( \sum_{l=1}^{j-1} \frac{1}{n(N-l)} \) equals 0. Since we choose the uniform nodes unequal to the root of the URT, we have that

\[
\mathbb{P}(i \text{ uniform nodes are in } T_j^{(N)}| |T_j^{(N)}| = n) = \frac{n!}{(N-1)!} \frac{(N-i-1)!}{(n-i)!(N-1)!}.
\]
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
\[
\mathbb{P}(|T_j^{(N)}| = n) = \frac{(N - j)!}{n - 1} \frac{(n - 1)!}{j} \frac{(j - 1)!}{j} \cdots \frac{(N - n - 1)!}{(N - 1)} = \frac{(j - 1)(N - j)!}{(N - 1)!} \frac{(N - n - 1)!}{(N - j - n + 1)!}.
\]

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 on 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 [13, 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
\[
\{Z_N^{(k)}\}_{k \geq 0} \overset{d}{=} \{U_N^{(k-1)} + V_N^{(k)}\}_{k \geq 0},
\]
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 understand 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 one has size \( N_1 \) and the second tree has size \( N - 1 - N_1 \). Given \( N_1 \), which has a discrete uniform distribution on the numbers \( \{1, 2, \ldots, N - 1\} \), the two trees are independent. Thus,
\[
p_N(i) = \frac{1}{N - 1} \sum_{k=1}^{N-1} \mathbb{P}(i \text{ uniform nodes in tree rooted at } 2 \text{ of size } k) + \frac{1}{N - 1} \sum_{k=1}^{N-1} \mathbb{P}(i \text{ uniform nodes are outside tree rooted at } 1 \text{ of size } N - k, \text{ share link})
\]
\[
= \frac{1}{N - 1} \sum_{k=1}^{N-1} \frac{k(k - 1) \cdots (k - i + 1)}{N(N - 1) \cdots (N - i + 1)} + \frac{1}{N - 1} \sum_{k=1}^{N-1} p_k(i) \frac{k(k - 1) \cdots (k - i + 1)}{N(N - 1) \cdots (N - i + 1)}.
\]

Defining \( \alpha_N(i) = N(N - 1) \cdots (N - i + 1)p_N(i) \), we have that
\[
\alpha_N(i) = \frac{1}{N - 1} \sum_{k=1}^{N-1} k(k - 1) \cdots (k - i + 1) + \frac{1}{N - 1} \sum_{k=1}^{N-1} \alpha_k(i).
\]

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

Together with the fact that \( \alpha_i(i) = 0 \), we end up with
\[
\alpha_N(i) = \sum_{j=1}^{N-i} \frac{(N - 1 - j)!}{(N - i - j)!} = \sum_{l=1}^{N-2} \frac{l!}{(l-i)!} = \frac{1}{i} \frac{(N - 1)!}{(N - i - 1)!}.
\]
Thus,
\[ p_N(i) = \frac{(N - i)!}{N!} \alpha_N(i) = \frac{1}{i} \frac{(N - i)}{N}. \]

Combining all terms, we end up with (31). Substitution into (29) now yields
\[
u_N(m) = \mu_N(1) + \sum_{i=2}^{m} \binom{m}{i} (-1)^{i+1} \sum_{j=1}^{N-1} \sum_{l=1}^{i+1} \frac{1}{l(N-l)} \times \sum_{n=1}^{N-j+1} (j-1)(N-j)! \frac{(N-n-1)! n!(N-i-1)!}{[(N-1)]^2 (N-j-n+1)!} \left[ 1 - \frac{1}{i} \frac{(n-i)}{n} \right].
\]

(33)

The above four-fold sum can be reordered as
\[
u_N(m) = \mu_N(1) + \sum_{i=2}^{m} \binom{m}{i} (-1)^{i+1} \sum_{n=1}^{N-1} (N-n-1)! n!(N-i-1)! \left[ 1 - \frac{1}{i} \frac{(n-i)}{n} \right] \times \sum_{l=1}^{N-n+1} \frac{1}{l(N-l)} \sum_{j=l+1}^{N-n} \frac{(j-1)(N-j)!}{(N-j-n+1)!}.
\]

(34)

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

Using [21, Lemma 10]
\[
\sum_{n=1}^{N} \frac{(K-n)!}{(N-n)!} \frac{1}{n} = \frac{K!}{N!} [\psi(K+1) - \psi(K - N + 1)],
\]

(35)

we can next sum the \( l \)-sum and obtain
\[
\sum_{l=1}^{N-n} \frac{1}{l(N-l)} \sum_{j=l+1}^{N-n+1} \frac{(j-1)(N-j)!}{(N-j-n+1)!} = \frac{1}{n+1} \frac{(N-1)!}{(N-n-1)!} [\psi(N) - \psi(n) + 1].
\]

Since
\[
\sum_{n=1}^{N-1} \frac{\psi(N) - \psi(n) + 1}{n(n+1)} = \sum_{n=1}^{N-1} \frac{1}{n}.
\]

(36)

the term \( \mu_N(1) \) equals the contribution due to \( i = 1 \). Substitution leads to
\[
\nu_N(m) = \sum_{n=1}^{N-1} \frac{(n-1)!}{(N-1)!} \frac{\psi(N) - \psi(n) + 1}{n+1} \nu_N(m)
\]

(37)

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

With \( \binom{m}{i} - \binom{m-1}{i} = \binom{m}{i} \frac{1}{m} \) and \( i \cdot \frac{1}{m} \left[ 1 - \frac{n-1}{in} \right] = n - \frac{(n+1)(n-i)}{m} \), and defining \( \Delta \nu_N(m) = \nu_N(m) - \nu_N(m-1) \), we find that
\[
m \Delta \nu_N(m) = \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!} \left[ 1 - \frac{n-i}{in} \right] = n \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!} - \frac{n+1}{n} \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \frac{(N-i-1)!}{(n-i)!}.
\]
Using an instance of Vandermonde’s convolution formula [10, 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
\[
m \Delta v_N(m) = (N-n-1)! \left( \frac{N-m-1}{n} \right) \left( \frac{N+nm-n}{N-m-n} \right) - \frac{(N-1)!}{n!}.
\]
Relation (37) shows that the difference \( \Delta u_N(m) = u_N(m) - u_N(m-1) \) is directly written in terms of the difference \( \Delta v_N(m) \),
\[
m \Delta u_N(m) = \sum_{n=1}^{N-1} \frac{(N-n-1)!}{(N-1)!} \left[ \psi(N) - \psi(n+1) \right] \frac{(N-m-1)}{n+1} \left( \frac{N-m-1}{n} \right) \left( \frac{N+nm-n}{N-m-n} \right) - \frac{(N-1)!}{n!}
\]
\[
= \frac{(N-m-1)!}{(N-1)!} \sum_{n=1}^{N-1} \frac{\psi(N) - \psi(n+1)}{n(n+1)} \left[ \frac{(N-n-1)!}{(N-m-n)!} \right] \frac{(N-m-1)}{n+1} \left( \frac{N-m-1}{n} \right) \left( \frac{N+nm-n}{N-m-n} \right) - \frac{(N-1)!}{n!}
\]
Using the identity (40) proved in the Appendix B yields
\[
m \Delta u_N(m) = \frac{N! \psi(N) - \psi(N+1)}{N-m} - \sum_{n=1}^{N-1} \frac{1}{n} = \frac{m}{N-m} \sum_{k=m}^{N-1} \frac{1}{k},
\]
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.

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 (37), which we will omit here.

The induction is initiated by \( N = 2 \), for which both sides give 1. The inductive step from \( N \) to \( N+1 \) is as follows:
\[
\sum_{k=1}^{N} \frac{1}{k^2} = \sum_{k=1}^{N-1} \frac{1}{k^2} + \frac{1}{N^2} = \sum_{k=1}^{N-1} \frac{1}{k^2} + \frac{1}{N^2} - \frac{1}{N} \sum_{n=1}^{N-1} \frac{1}{n^2} = \sum_{l=1}^{N} \frac{1}{N-l+1} - \sum_{n=l-1}^{N-1} \frac{1}{n^2}
\]
\[
= \sum_{l=1}^{N} \frac{1}{N-l+1} \sum_{n=l-1}^{N-1} \frac{1}{n} - \sum_{n=1}^{N} \frac{1}{n^2} = \sum_{k=1}^{N-1} \frac{1}{N+1-k} \sum_{n=k}^{N} \frac{1}{n}
\]
since \( \sum_{l=1}^{N} \frac{1}{N-l+1} = \frac{2}{N} \sum_{n=1}^{N-1} \frac{1}{n} \). This proves (12).

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

For the first statement in (15), we again use that \( W_{\text{Steiner},N}(m) \leq W_N(m) \), so that \(|W_{\text{Steiner},N}(m) - W_N(m)| = W_N(m) - W_{\text{Steiner},N}(m)\). Since both have the same asymptotics by (14), we have that
\[ W_{\text{Steiner},N}(m) - W_N(m) \] 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
\[
\frac{W_N(m)}{mN \log \frac{N}{m+1}} = \frac{W_{\text{Steiner},N}(m)}{mN \log \frac{N}{m+1}} + \frac{W_N(m) - W_{\text{Steiner},N}(m)}{mN \log \frac{N}{m+1}}.
\]
The first term converges to 1 in probability and the second term to 0 in probability, so that \( \frac{W_N(m)}{m \log \frac{N}{m+1}} \) converges to 1 in probability.

**Acknowledgements.** We thank Milena Janic for performing simulations to check our computations and Roelof Koekoek for useful discussions.

**References**


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 (2). 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 \( \varphi_{H_N(N-1)}(z) = E \left[ z^{H_N(N-1)} \right] \) because \( H_N(N-1) = N-1 \).

For \( m = 2 \), using (2), the recursion becomes
\[
\varphi_{H_N(2)}(z) = \frac{(N-3)(N-1+2z)}{(N-1)^2}\varphi_{H_N-1(2)}(z) + \frac{4z}{(N-1)(N-2)} \left( \frac{\Gamma(z+N-1)}{\Gamma(z+1)} - \frac{1}{N-1} \right),
\]
with initial value \( \varphi_{H_N(2)}(z) = z^2 \). Iteration yields that \( \varphi_{H_N(2)}(z) \) equals
\[
\prod_{j=1}^{p}(N-2-j) \prod_{j=1}^{p}(N-j+2z) \varphi_{H_{N-p}}(2)(z)
\]
\begin{align*}
+4z \sum_{k=1}^{p} & \prod_{j=3}^{k+1}(N-j)(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)}{\Gamma(z+1) - \Gamma(z+1)} - \frac{1}{N-k} \right). 
\end{align*}{
With $\varphi_{H_3}(2)(z) = z^2$, which is reached for $N - p = 3$ or $p = N - 3$, the iteration ends with result

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

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 [13, Lemma A.4]

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

we obtain after some manipulations

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

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

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

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

Using (39), 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

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

18
and with \( \varphi_{H_4}(z) = z^3 \), which is reached for \( N - p = 4 \) or \( p = N - 4 \), the iteration ends with result

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

We rewrite the above as for \( m = 2 \) as

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

Again using (38), we finally find

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

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

## B Proof of the identity (40).

**Lemma B.1**

\[
\sum_{n=1}^{N-1} \psi(n) - \frac{\psi(n)}{n} + 1 = \frac{(N - n - 1)! (N + mn - n)}{(N - m - n)!} = \frac{(N - 1)!}{(N - m)!} (N \psi(N) - m \psi(m)) . \quad (40)
\]

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

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

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

to express the left-hand side (LS) of (40) as the sum \( LS_1 + LS_2 \). With the definition of the digamma function we obtain

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

and similarly

\[
LS_2 = (N - m + 1) \sum_{n=1}^{N-1} \frac{1}{n + 1} \frac{(N - n - 1)!}{(N - n - m)!} + (N - m + 1) \sum_{k=1}^{N-1} \frac{1}{k} \sum_{n=1}^{N-1} \frac{1}{n + 1} \frac{(N - n - 1)!}{(N - n - m)!} \\
- (N - m + 1) \sum_{n=1}^{N-1} \sum_{k=1}^{N-1} \frac{1}{k n + 1} \frac{(N - n - 1)!}{(N - n - m)!} ,
\]
From (35) 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
\[ \sum_{n=1}^{N} \frac{(K-n)!}{(N-n)!} \left( \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{k} \right) = -\frac{1}{2} \frac{K!}{N!} \left( \sum_{k=1}^{K} \frac{1}{k^2} - \sum_{l=1}^{K-N} \frac{1}{l^2} \right) + \frac{1}{2} \frac{K!}{N!} [\psi(K+1) - \psi(K-N+1)]^2 \]

which follows from [19, pp. 252] after some manipulations, we have
\[ \sum_{n=1}^{N-1} \sum_{k=1}^{n-1} \frac{1}{kn} \frac{(N-n-1)!}{(N-n-m)!} = -\frac{1}{2} \frac{(N-1)!}{(N-m)!} \left( \sum_{k=m}^{N} \frac{1}{k^2} \right) + \frac{1}{2} \frac{(N-1)!}{(N-m)!} [\psi(N) - \psi(m)]^2. \]

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 (40).