The Weight of the Shortest Path Tree

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

December 22, 2005

Abstract

The minimal weight of the shortest path tree in a complete graph with independent and exponential (mean 1) random link weights, is shown to converge to a Gaussian distribution. We prove a conditional central limit theorem and show that the condition holds with probability converging to 1.

1 Introduction

Consider the complete graph K_{N+1}, with $N+1$ nodes and $\frac{1}{2}N(N+1)$ links. To each link (or edge) we independently assign an exponentially distributed weight with mean 1. The shortest path between two nodes is that path whose sum of its links weights is minimal. (Each of these shortest paths is a.s. unique.) The shortest path tree (SPT) is the union of the N shortest paths from a root (e.g. node 1) to all other nodes in the graph. In this paper we consider the total weight W_N of the SPT rooted at node 1 to all other nodes in the complete graph. In [7, 9], we have rephrased the shortest path problem between two arbitrary nodes in the complete graph 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 $N+1$ 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 $\lambda_n = n\,(N+1-n)$, which is the transition rate in the continuous-time Markov chain. Since the discovery of nodes at each stage only increases n, the Markov discovery process is a pure birth process with birth rate $n\,(N+1-n)$. We call τ_n the inter-attachment time between the inclusion of the n^{th} and $(n+1)^{\text{st}}$ node to the SPT for $n=1,\ldots,N$. The inter-attachment time τ_n is exponentially distributed with parameter λ_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 is exactly a uniform recursive tree (URT). A URT of size $N+1$ 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 $N+1$.

*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.

The average of the weight W_N of the SPT equals

$$\mathbb{E}[W_N] = \sum_{k=1}^{N} \frac{1}{k^2}, \tag{1}$$

and the variance is

$$\text{var}[W_N] = \frac{4}{(N+1)} \sum_{k=1}^{N} \frac{1}{k^3} + 4 \sum_{j=1}^{N} \frac{1}{j^3} \sum_{k=1}^{j} \frac{1}{k} - 5 \sum_{j=1}^{N} \frac{1}{j^4}. \tag{2}$$

The result for the mean (1) has been found first in [8], but it is rederived in Section 2.1 because the method is considerably simpler. The derivation for the variance (2) is in Section 2.2 while many appearing sums are computed in the Appendix. The asymptotic form of the average weight is immediate from (1) as

$$\mathbb{E}[W_N] = \zeta(2) + O\left(\frac{1}{N}\right), \tag{3}$$

while the corresponding result for the variance, derived in Section 2.2, is

$$\text{var}[W_N] = \frac{4\zeta(3)}{N} + o\left(\frac{1}{N}\right). \tag{4}$$

The third and main result in this paper is that we show that the scaled weight of the SPT tends to a Gaussian. In particular,

$$\sqrt{N}\left(W_N - \zeta(2)\right) \xrightarrow{d} N\left(0, \sigma^2\right),$$

where $\sigma^2 = \sigma^2_{\text{SPT}} = 4\zeta(3) \simeq 4.80823$. A related result for the minimum spanning tree (MST) is worth mentioning. The average weight of the minimum spanning tree W_{MST} in the complete graph with exponential with mean 1 (or uniform on $[0,1]$) link weights has been computed earlier by Frieze [3]. For large N, Frieze showed that

$$\mathbb{E}[W_{\text{MST}}] \to \zeta(3).$$

Janson [5] extended Frieze's result by proving that the scaled weight of the MST tends to a Gaussian,

$$\sqrt{N}\left(W_{\text{MST}} - \zeta(3)\right) \xrightarrow{d} N\left(0, \sigma^2_{\text{MST}}\right),$$

where

$$\sigma^2_{\text{MST}} = 2\zeta(4) - 2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \frac{(i+k-1)!k^k(i+j)^{i-2}j}{i!k!(i+j+k)^{i+k+2}} \simeq 1.6857.$$

The triple sum was exactly computed by Wästlund [10] resulting in

$$\sigma^2_{\text{MST}} = 6\zeta(4) - 4\zeta(3).$$

2 The weight of the shortest path tree

From the Markov discovery process briefly explained in Section 1, the discovery time to the k^{th} discovered node from the root equals

$$v_k = \sum_{n=1}^{k} \tau_n, \tag{5}$$

where the inter-attachment times $\tau_1, \tau_2, \cdots, \tau_k$ are independent, exponentially distributed random variables with parameter $\lambda_n = n(N+1-n)$, $1 \leq n \leq k$. An arbitrary uniform recursive tree consisting of $N+1$ nodes and with the root labeled by zero can be represented as

$$(0 \longleftarrow 1)(N_2 \longleftarrow 2)\ldots(N_N \longleftarrow N) \tag{6}$$

where $(N_j \longleftarrow j)$ means that the j^{th} discovered node is attached to node $N_j \in \{0,\ldots,j-1\}$. Hence, N_j is the predecessor of j and this relation is indicated by $\longleftarrow$. The weight W_N of an arbitrary SPT from the root 0 to all N other nodes is with (5) and $v_0 = 0$ and $N_1 = 0$,

$$W_N = \sum_{j=1}^{N}(v_j - v_{N_j}) = \sum_{j=1}^{N}\sum_{n=N_j+1}^{j}\tau_n.$$

In the URT, the integer N_j, $1 \leq j \leq N$, are independent and uniformly distributed over the interval $\{0,\ldots,j-1\}$. It is more convenient to use a discrete uniform random variable on $\{1,\ldots,j\}$ which we define as $A_j = N_j + 1$. We rewrite

$$W_N = \sum_{j=1}^{N}\sum_{n=A_j}^{j}\tau_n = \sum_{j=1}^{N}\sum_{n=1}^{j}\mathbf{1}_{\{A_j\leq n\}}\tau_n = \sum_{n=1}^{N}\tau_n\left(\sum_{j=n}^{N}\mathbf{1}_{\{A_j\leq n\}}\right).$$

The set $\{A_j\}_{1\leq j\leq N}$ are independent random variables with $\mathbb{P}[A_j = k] = \frac{1}{j}$ for $k \in \{1,2,\ldots,j\}$. In addition, we define for $n \in \{1,\ldots,N\}$ the random variables

$$B_n = \sum_{j=n}^{N}\mathbf{1}_{\{A_j\leq n\}}, \tag{7}$$

to obtain

$$W_N = \sum_{n=1}^{N}B_n\tau_n. \tag{8}$$

The N random variables $B_1, B_2, \ldots, B_N$ are dependent. The mean of the random variable B_n follows from (7) as

$$\mathbb{E}[B_n] = \sum_{j=n}^{N}\mathbb{E}\left[\mathbf{1}_{\{A_j\leq n\}}\right] = \sum_{j=n}^{N}\mathbb{P}[A_j \leq n] = \sum_{j=n}^{N}\frac{n}{j}. \tag{9}$$

The variance $\text{var}[B_n]$ and covariances $\text{cov}[B_n, B_m]$ are given in Lemma 1 below.

2.1 The average weight of the SPT

It is immediate from (8) and the independence of the $A_1, A_2, \ldots, A_N$ from the inter-attachment times $\tau_1, \tau_2, \ldots, \tau_N$ that

$$\mathbb{E}[W_N] = \sum_{n=1}^{N}\mathbb{E}[B_n]\,\mathbb{E}[\tau_n] = \sum_{n=1}^{N}\sum_{j=n}^{N}\frac{n}{j}\frac{1}{n(N+1-n)} = \sum_{n=1}^{N}\frac{1}{(N+1-n)}\sum_{j=n}^{N}\frac{1}{j} = \sum_{j=1}^{N}\frac{1}{j}\sum_{k=N+1-j}^{N}\frac{1}{k},$$

which is, by the equality (31) below, equal to (1).

2.2 The variance of the weight of the SPT

To compute the variance of W_N, we use the formula

$$\mathrm{var}\,[W_N] = \mathrm{var}\,[\mathbb{E}[W_N|B_1,\ldots,B_N]] + \mathbb{E}[\mathrm{var}\,[W_N|B_1,\ldots,B_N]]. \tag{10}$$

Since for an exponential random variable τ_n with parameter $\lambda_n = n\,(N+1-n)$, the expectation equals $1/\lambda_n$ and the variance $1/\lambda_n^2$, we have

$$\mathbb{E}[W_N|B_1,\ldots,B_N] = \mathbb{E}\left[\sum_{n=1}^{N} B_n\tau_n|B_1,\ldots,B_N\right] = \sum_{n=1}^{N}\lambda_n^{-1}B_n, \tag{11}$$

$$\mathrm{var}\,[W_N|B_1,\ldots,B_N] = \mathrm{var}\left[\sum_{n=1}^{N} B_n\tau_n|B_1,\ldots,B_N\right] = \sum_{n=1}^{N}\lambda_n^{-2}B_n^2. \tag{12}$$

Combining (10), (11) and (12),

$$\mathrm{var}\,[W_N] = \mathrm{var}\left[\sum_{n=1}^{N}\lambda_n^{-1}B_n\right] + \sum_{n=1}^{N}\lambda_n^{-2}\mathbb{E}[B_n^2]. \tag{13}$$

To proceed, we need expressions for the covariance of B_n and B_m, which are computed in the following lemma:

Lemma 1 *For every $n, m \geq 1$,*

(i)

$$var\,[B_n] = \sum_{j=n}^{N}\left(\frac{n}{j} - \frac{n^2}{j^2}\right), \tag{14}$$

(ii)

$$cov\,[B_n,B_m] = \sum_{j=m}^{N}\frac{n}{j}\left(1-\frac{m}{j}\right), \qquad n \leq m. \tag{15}$$

Proof. The proof of (i) follows from that of (ii) with $n = m$.

(ii) The bilinearity of the covariance yields, for $n \leq m$,

$$\mathrm{cov}\,[B_n,B_m] = \mathrm{cov}\left[\sum_{i=n}^{N}\mathbf{1}_{\{A_i\leq n\}},\sum_{j=m}^{N}\mathbf{1}_{\{A_j\leq m\}}\right] = \sum_{i=n}^{N}\sum_{j=m}^{N}\mathrm{cov}\left[\mathbf{1}_{\{A_i\leq n\}},\mathbf{1}_{\{A_j\leq m\}}\right].$$

Since A_i and A_j are independent for $i \neq j$, we have that $\mathrm{cov}\left[\mathbf{1}_{\{A_i\leq n\}},\mathbf{1}_{\{A_j\leq m\}}\right] = 0$, for $i \neq j$, such that

$$\sum_{i=n}^{N}\sum_{j=m}^{N}\mathrm{cov}\left[\mathbf{1}_{\{A_i\leq n\}},\mathbf{1}_{\{A_j\leq m\}}\right] = \sum_{j=m}^{N}\mathrm{cov}\left[\mathbf{1}_{\{A_j\leq n\}},\mathbf{1}_{\{A_j\leq m\}}\right].$$

With $\mathrm{cov}\left[\mathbf{1}_{\{A_j\leq n\}},\mathbf{1}_{\{A_j\leq m\}}\right] = \mathbb{E}\left[\mathbf{1}_{\{A_j\leq n\}}\mathbf{1}_{\{A_j\leq m\}}\right] - \mathbb{P}\,[A_j\leq n]\,\mathbb{P}\,[A_j\leq m]$ and $\mathbf{1}_{\{A_j\leq n\}}\mathbf{1}_{\{A_j\leq m\}} = \mathbf{1}_{\{A_j\leq \min(n,m)\}} = \mathbf{1}_{\{A_j\leq n\}}$ for $n \leq m$, we obtain

$$\sum_{j=m}^{N}\mathrm{cov}\left[\mathbf{1}_{\{A_j\leq n\}},\mathbf{1}_{\{A_j\leq m\}}\right] = \sum_{j=m}^{N}\frac{n}{j}\left(1-\frac{m}{j}\right). \qquad \square$$

Applying Lemma 1 to the right side of (13) gives

$$\operatorname{var}[W_N] = \operatorname{var}\left[\sum_{n=1}^{N} \lambda_n^{-1} B_n\right] + \sum_{n=1}^{N} \lambda_n^{-2} \mathbb{E}[B_n^2] = 2 \sum_{n \le m} \lambda_n^{-1} \lambda_m^{-1} \operatorname{cov}[B_n, B_m] + \sum_{n=1}^{N} \lambda_n^{-2} (\mathbb{E}[B_n])^2$$

$$= 2 \sum_{n \le m} \lambda_n^{-1} \lambda_m^{-1} \sum_{j=m}^{N} \frac{n}{j}\left(1 - \frac{m}{j}\right) + \sum_{n=1}^{N} \lambda_n^{-2} \left(\sum_{j=n}^{N} \frac{n}{j}\right)^2 = T_2(N) + T_1(N).$$

where the sum $T_1(N)$ is defined as

$$T_1(N) = \sum_{n=1}^{N} \frac{1}{(N+1-n)^2} \left(\sum_{j=n}^{N} \frac{1}{j}\right)^2 = 2 \sum_{n=1}^{N} \frac{1}{n^3} \sum_{k=1}^{n} \frac{1}{k} - \left(\sum_{k=1}^{N} \frac{1}{k^2}\right)^2, \tag{16}$$

and the last equality is proved in the Appendix (see (46) below), while

$$T_2(N) = 2 \sum_{n=1}^{N} \sum_{m=n}^{N} \frac{1}{(N+1-n)\, m\, (N+1-m)} \sum_{j=m}^{N} \frac{1}{j}\left(1 - \frac{m}{j}\right) \tag{17}$$

$$= \frac{4}{N+1} \sum_{k=1}^{N} \frac{1}{k^3} - 5 \sum_{k=1}^{N} \frac{1}{k^4} + 2 \sum_{n=1}^{N} \frac{1}{n^3} \sum_{k=1}^{n} \frac{1}{k} + \left(\sum_{k=1}^{N} \frac{1}{k^2}\right)^2,$$

where the last equality is proved in the Appendix (see (44) below). Summing $T_1(N)$ and $T_2(N)$ gives the explicit form (2) of the variance for W_N.

We next investigate the asymtotics of the variance of W_N for large N. We write the sum of the last two terms in (2) by

$$Q(N) = 4 \sum_{j=1}^{N} \frac{1}{j^3} \sum_{k=1}^{j} \frac{1}{k} - 5 \sum_{j=1}^{N} \frac{1}{j^4}. \tag{18}$$

Then, for large N,

$$Q(N) - Q(N-1) = \frac{4}{N^3} \sum_{k=1}^{N} \frac{1}{k} - \frac{5}{N^4} = O\left(\frac{\log N}{N^3}\right),$$

and, by summation, $Q(N) = Q + O\left(\frac{\log N}{N^2}\right)$, where the limit $Q = \lim_{N\to\infty} Q(N)$ exists, by (18). It follows from [1, Corollary 4, main theorem] that

$$Q = 4 \sum_{j=1}^{\infty} \frac{1}{j^3} \sum_{k=1}^{j} \frac{1}{k} - 5 \sum_{j=1}^{\infty} \frac{1}{j^4} = 0, \tag{19}$$

so that

$$Q(N) = O\left(\frac{\log N}{N^2}\right).$$

Hence, asymptotically, we arrive at (4).

3 Central limit theorem for W_N

In this section we prove a central limit theorem for W_N. We use the symbol $\xrightarrow{d}$ to denote convergence in distribution and the symbol $\xrightarrow{P}$ for convergence in probability. We denote by $\sigma(W_N)$, the standard deviation of W_N, so $\sigma^2(W_N) = \text{var}[W_N]$. We denote by $\mathcal{N}(0,1)$ a random variable with standard normal distribution. The main result proved in this section is the following central limit theorem for W_N:

Theorem 2 *As $N \to \infty$,*

$$\frac{W_N - \mathbb{E}[W_N]}{\sigma(W_N)} \xrightarrow{d} \mathcal{N}(0,1).$$

We start with an outline of the proof. We wish to prove that W_N is asymptotically normal, in the sense that $\sqrt{N}(W_N - \mathbb{E}[W_N])$ has an asymptotic normal distribution. We first define

$$s_N^2 = \sum_{j=1}^{N} \frac{B_j^2}{j^2(N+1-j)^2}. \tag{20}$$

We note that s_N^2 is a random variable, and we sometimes make this explicit by writing $s_N^2(\omega)$, where ω is an element of the probability space.

We split $W_N = X_N + Y_N$, where

$$X_N = \sum_{j=1}^{N} \Big(\tau_j - \frac{1}{j(N+1-j)}\Big) B_j, \qquad \text{and} \qquad Y_N = \sum_{j=1}^{N} \frac{B_j}{j(N+1-j)}. \tag{21}$$

Our strategy is to prove the following steps:

1. Define an event $\mathcal{A}_N$ such that: (a) $\mathcal{A}_N$ is measurable with respect to the σ-algebra generated by $\{A_j\}_{j=1}^N$, (b) $\mathbb{P}(\mathcal{A}_N^c) \le N^{-\delta}$, and (c) Uniformly for $\omega \in \mathcal{A}_N$, we have $Ns_N^2(\omega) - \sigma_{1,N}^2 = o(1)$, where
$$\sigma_{1,N}^2 = N \sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} = NT_1(N).$$
Consecutively, we show that $\sigma_1^2 = \lim_{N\to\infty} \sigma_{1,N}^2$ exists.

2. Prove the central limit theorem for $\sqrt{N}X_N$ with variance Ns_N^2, conditionally on $\{A_j\}_{j=1}^N$, when $\{A_j\}_{j=1}^N$ is such that $\mathcal{A}_N$ holds. More precisely, we will show that uniformly on $\mathcal{A}_N$,
$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2 N s_N^2/2} + o(1),$$
where $\mathbb{E}_A$ is the conditional expectation given $\{A_j\}_{j=1}^N$.

3. Prove that $\sqrt{N}(Y_N - \mathbb{E}[Y_N])$ converges in distribution to a normal random variable with variance $\sigma_2^2 = \lim_{N\to\infty} NT_2(N)$.

Together, these steps prove Theorem 2. Indeed, we compute, using that $\mathcal{A}_N$ is measurable with respect to the sigma-algebra generated by $\{A_j\}_{j=1}^N$,

$$\phi(t) = \mathbb{E}[e^{it\sqrt{N}W_N}] = \mathbb{E}[e^{it\sqrt{N}W_N}\mathbf{1}_{\mathcal{A}_N}] + O(\mathbb{P}(\mathcal{A}_N^c)) = \mathbb{E}\Big[\mathbb{E}_A[e^{it\sqrt{N}W_N}]\mathbf{1}_{\mathcal{A}_N}\Big] + O(N^{-\delta}).$$

We split $W_N = X_N + Y_N$, and use that Y_N is measurable with respect to $\{A_j\}_{j=1}^N$ to arrive at

$$\phi(t) = \mathbb{E}\Big[e^{it\sqrt{N}Y_N}\mathbb{E}_A[e^{it\sqrt{N}X_N}]\mathbf{1}_{\mathcal{A}_N}\Big] + O(N^{-\delta}).$$

According to Step 2, uniformly on $\mathcal{A}_N$,

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2 N s_N^2/2} + o(1),$$

and, according to Step 1, uniformly on $\mathcal{A}_N$,

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2\sigma_{1,N}^2/2} + o(1).$$

Therefore, using that $\mathbb{E}[W_N] = \mathbb{E}[Y_N]$,

$$\mathbb{E}[e^{it\sqrt{N}(W_N-\mathbb{E}[W_N])}] = \mathbb{E}[e^{it\sqrt{N}(W_N-\mathbb{E}[W_N])}\mathbf{1}_{\mathcal{A}_N}] + o(1) = e^{-t^2\sigma_{1,N}^2/2}\mathbb{E}[e^{it\sqrt{N}(Y_N-\mathbb{E}[Y_N])}\mathbf{1}_{\mathcal{A}_N}] + o(1),$$

again by Step 1. Now, by Step 3, we have that

$$\mathbb{E}[e^{it\sqrt{N}(Y_N-\mathbb{E}[Y_N])}\mathbf{1}_{\mathcal{A}_N}] = \mathbb{E}[e^{it\sqrt{N}(Y_N-\mathbb{E}[Y_N])}] + o(1) = e^{-t^2\sigma_2^2/2} + o(1),$$

so that

$$\mathbb{E}[e^{it\sqrt{N}(W_N-\mathbb{E}[W_N])}] = e^{-t^2\sigma^2/2} + o(1),$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2$, and $\sigma_1^2 = \lim_{N\to\infty}\sigma_{1,N}^2$.

We now turn to the details of the proof. We will prove Steps 1-3 in Sections 3.1–3.3, respectively.

3.1 Step 1: The good event and convergence in probability of Ns_N^2

Fix $a \in (0,1)$ and an integer n_0, and define

$$\mathcal{A}_N = \mathcal{B}_N \cap \mathcal{C}_N,$$

where

$$\mathcal{B}_N = \bigcap_{j=N^a}^{N-N^a} \{|B_j - \mathbb{E}[B_j]| \le \epsilon_N \mathbb{E}[B_j]\}, \tag{22}$$

and

$$\mathcal{C}_N = \bigcap_{j=1}^{N^a} \{B_j \le \max(2n_0, j)\log N\}, \tag{23}$$

with

$$\epsilon_N = N^{-a/3}.$$

Later we will see that in fact we need n_0 large and $a > \frac{3}{4}$. On the event $\mathcal{B}_N$, with all random variables B_j, with $N^a \le j \le N - N^a$, are close to their respective mean $\mathbb{E}[B_j]$; on the event $\mathcal{C}_N$, we have a logarithmic bound on the random variables B_j, with $1 \le j \le N^a$.

We will show two lemmas. The first shows that $\mathcal{A}_N$ occurs with high probability, while the second proves that Ns_N^2 is close to a constant on $\mathcal{A}_N$. Together the lemmas imply the claims in step 1.

Lemma 3 *Fix $a \in (\frac{1}{2}, 1)$ and n_0 sufficiently large. Then, for N sufficiently large,*

$$\mathbb{P}(\mathcal{A}_N^c) \le N^{-(2-a)},$$

so that we can take $\delta = 2 - a > 1$ in Step 1.

Proof. We use Boole's inequality to obtain,

$$\mathbb{P}(\mathcal{A}_N^c) \le \sum_{j=1}^{N^a} \mathbb{P}(B_j > \max(2n_0, j)\log N) + \sum_{j=N^a}^{N-N^a} \mathbb{P}(|B_j - \mathbb{E}[B_j]| \ge \epsilon_N \mathbb{E}[B_j]).$$

Note that B_j is the sum of independent indicators, and, therefore, by the estimate of Janson [6] and with $0 < \epsilon < 1$,

$$\mathbb{P}(|B_j - \mathbb{E}[B_j]| \ge \epsilon \mathbb{E}[B_j]) \le 2e^{-\frac{3}{8}\epsilon^2 \mathbb{E}[B_j]},$$

where $\mathbb{E}[B_j]$ is given in (9) which we bound as

$$j \log\left(\frac{N}{j}\right) = j \int_j^N \frac{1}{x} dx \le \mathbb{E}[B_j] \le j \int_{j-1}^N \frac{1}{x} dx = j \log\left(\frac{N}{j-1}\right).$$

Therefore, we have that

$$\mathbb{P}(|B_j - \mathbb{E}[B_j]| \ge \epsilon \mathbb{E}[B_j]) \le 2e^{-\frac{3}{8}\epsilon^2 j \log \frac{N}{j}},$$

which is $o(N^{-2})$ for all $n_0 \le j \le N^a$ and n_0 sufficiently large. On the other hand, for $j \le n_0$,

$$\mathbb{P}(B_j \ge 2n_0 \log N) \le \mathbb{P}(|B_j - \mathbb{E}[B_j]| \ge n_0 \mathbb{E}[B_j]) \le N^{-2},$$

again for n_0 sufficiently large. Hence,

$$\sum_{j=1}^{N^a} \mathbb{P}(B_j > \max(2n_0, j)\log N) \le N^a \cdot N^{-2} = N^{-2+a}, \tag{24}$$

for n_0 sufficiently large.

We complete the argument as follows. For $N^a \le j \le N - N^a$, we have that

$$\mathbb{P}(|B_j - \mathbb{E}[B_j]| \ge \epsilon \mathbb{E}[B_j]) \le 2e^{-\frac{3}{8}\epsilon^2 j \log \frac{N}{j}} \le 2e^{-\frac{3}{16}\epsilon^2 N^a}, \tag{25}$$

since, uniformly for all j such that $N^a \le j \le N - N^a$, we have $j \log \frac{N}{j} \ge \frac{1}{2} N^a$. Indeed, this follows since $f_N(j) = j \log \frac{N}{j}$ is first increasing and then decreasing. Therefore, uniformly for $N^a \le j \le N - N^a$,

$$f_N(j) \ge \min\left(f_N(N^a), f_N(N - N^a)\right). \tag{26}$$

We note that, for N sufficiently large and $a \in (\frac{1}{2}, 1)$ fixed,

$$f_N(N^a) = (1-a) N^a \log N \ge \frac{1}{2} N^a,$$

and, using that $\log(1-x) \le -x$, $0 < x < 1$,

$$f_N(N - N^a) \ge -(N - N^a)\log(1 - N^{a-1}) \ge (N - N^a) N^{a-1} \ge \frac{1}{2} N^a.$$

For $\epsilon_N = N^{-a/3}$, we can use (25), to obtain, again for N sufficiently large,

$$\sum_{j=N^a}^{N-N^a} \mathbb{P}(|B_j - \mathbb{E}[B_j]| \ge \epsilon_N \mathbb{E}[B_j]) \le 2Ne^{-\frac{3}{16}N^{\frac{a}{3}}} = o(N^{-2+a}). \tag{27}$$

Combining the bounds (24) and (27) we obtain the statement in the lemma. □

Recall that $Ns_N^2 = N \sum_{j=1}^N \frac{B_j^2}{j^2(N+1-j)^2}$. We next investigate Ns_N^2 on the event $\mathcal{A}_N$:

Lemma 4 *For $N \to \infty$, and uniformly on the event $\mathcal{A}_N$,*

$$Ns_N^2(\omega) - N\sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} = o(1).$$

Proof. From (7), it follows that $B_j \le N+1-j$. For N sufficiently large,

$$N\sum_{j=N-N^a}^{N} \frac{B_j^2}{j^2(N+1-j)^2} \le N\sum_{j=N-N^a}^{N} \frac{1}{j^2} \le \frac{2N^a}{N-N^a} = O(N^{a-1}),$$

for any $a<1$. Therefore, we have that

$$Ns_N^2(\omega) = N\sum_{j=1}^{N-N^a} \frac{B_j^2}{j^2(N+1-j)^2} + o(1).$$

On the event $\mathcal{C}_N$, for N sufficiently large,

$$N\sum_{j=1}^{N^a} \frac{B_j^2}{j^2(N+1-j)^2} \le 2N\sum_{j=1}^{N^a} \frac{(\max(2n_0,j))^2(\log N)^2}{j^2N^2} \le O\big(N^{-1+a}(\log N)^2\big) = o(1),$$

so that, on $\mathcal{C}_N$,

$$Ns_N^2(\omega) = N\sum_{j=N^a}^{N-N^a} \frac{B_j^2}{j^2(N+1-j)^2} + o(1).$$

On $\mathcal{B}_N$, and for $N^a \le j \le N-N^a$, we can sandwich $(1-\epsilon_N)^2(\mathbb{E}[B_j])^2 \le B_j^2 \le (1+\epsilon_N)^2(\mathbb{E}[B_j])^2$, so with probability at least $1-O(N^{-(2-a)})$, we find,

$$Ns_N^2(\omega) = (1+O(\epsilon_N))^2 N\sum_{j=N^a}^{N-N^a} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} + o(1).$$

Similar estimates as above yield that

$$N\sum_{j=1}^{N^a} \frac{\mathbb{E}[B_j]^2}{j^2(N+1-j)^2} = o(1), \qquad N\sum_{j=N-N^a}^{N} \frac{\mathbb{E}[B_j]^2}{j^2(N+1-j)^2} = o(1).$$

This completes the proof of the lemma. □

The argument of convergence in probability of $Ns_N^2(\omega)$ is complete when we prove that

$$N\sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} \to \sigma_1^2.$$

For this, we note that

$$N\sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} = NT_1(N),$$

and from (47), we find that $\sigma_1^2 = 2\zeta(2)$.

3.2 Step 2: Conditional central limit theorem for X_N

In this section, we compute $\mathbb{E}_A[e^{it\sqrt{N}X_N}]$, where $\{A_j\}_{j=1}^N$ is such that $\mathcal{A}_N$ holds. For this, we note that, for any random variable X with finite third moment, we have that

$$\phi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu - t^2\sigma^2/2 + O(|t|^3 m_3)}, \tag{28}$$

where $\mu = \mathbb{E}[X]$, $\sigma^2 = \mathrm{var}(X)$ and $m_3 = \mathbb{E}[|X|^3]$. The independence of the τ_j, conditionally on $\{A_j\}_{j=1}^N$, gives that

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = \prod_{j=1}^N \mathbb{E}_A[e^{it\sqrt{N}\left(\tau_j - \frac{1}{j(N+1-j)}\right)B_j}].$$

By (28), and since $B_1, \ldots, B_N$ are measurable with respect to the $\sigma-$algebra spanned by the random variables $A_1, A_2, \ldots, A_N$, we obtain that

$$\mathbb{E}_A[e^{it\sqrt{N}\left(\tau_j - \frac{1}{j(N+1-j)}\right)B_j}] = \exp\Big[-t^2 N \frac{B_j^2}{2j^2(N+1-j)^2} + O\Big(|t|^3 N^{3/2} \frac{B_j^3}{j^3(N+1-j)^3}\Big)\Big].$$

Therefore,

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2 N s_N^2/2} e^{O(|t|^3 v_N)},$$

where

$$v_N = N^{3/2} \sum_{j=1}^N \frac{B_j^3}{j^3(N+1-j)^3}.$$

We finally show that $v_N = o(1)$ on $\mathcal{A}_N$. First, we note that, since $B_j \leq N+1-j$ and $a < 1$,

$$N^{3/2} \sum_{j=N-N^a}^N \frac{B_j^3}{j^3(N+1-j)^3} \leq N^{3/2} \sum_{j=N-N^a}^N \frac{1}{j^3} \leq N^{\frac{3}{2}-3+a} \leq N^{-\frac{1}{2}}.$$

When $j \leq N - N^a$, we can use the bounds provided by $\mathcal{A}_N$. We start with the contribution due to $j \leq N^a$, for which we can bound on $\mathcal{C}_N$, for sufficiently large N, and some constant C depending on n_0 and a,

$$N^{3/2} \sum_{j=1}^{N^a} \frac{B_j^3}{j^3(N+1-j)^3} \leq N^{3/2} \sum_{j=1}^{N^a} \frac{(\max(2n_0, j))^3 (\log N)^3}{j^3(N+1-j)^3} \leq C(\log N)^3 N^{\frac{3}{2}-3+a} \leq N^{-\frac{1}{2}}.$$

Finally, for $N^a \leq j \leq N - N^a$, we obtain on $\mathcal{B}_N$, using $\mathbb{E}[B_j] \leq j \log N$,

$$\begin{aligned} N^{3/2} \sum_{j=N^a}^{N-N^a} \frac{B_j^3}{j^3(N+1-j)^3} &\leq (1+\epsilon_N)^3 N^{3/2} \sum_{j=N^a}^{N-N^a} \frac{(\mathbb{E}[B_j])^3}{j^3(N+1-j)^3} \\ &\leq (1+\epsilon_N)^3 N^{3/2} \sum_{j=N^a}^{N-N^a} \frac{(\log N)^3}{(N+1-j)^3} \\ &\leq (1+\epsilon_N)^3 (\log N)^3 N^{3/2} N^{-2a} \leq N^{-\eta}, \end{aligned}$$

for any $\eta < 2a - \frac{3}{2}$, and we note that $\eta > 0$ when $a > \frac{3}{4}$. This completes the proof that $v_N \leq N^{-\eta}$ when $a > \frac{3}{4}$, and that, for $\{A_j\}_{j=1}^N$ such that $\mathcal{A}_N$ holds,

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2 N s_N^2/2} e^{O(N^{-\eta})} = e^{-t^2 N s_N^2/2} + o(1).$$

3.3 Step 3: The central limit theorem for Y_N

We again use convergence of characteristic functions to that of a normal random variable with mean 0. We rewrite

$$\sqrt{N}\big(Y_N - \mathbb{E}[Y_N]\big) = \sqrt{N}\sum_{j=1}^{N}\frac{1}{j(N+1-j)}\sum_{k=j}^{N}\big(\mathbf{1}_{\{A_k\le j\}} - \frac{j}{k}\big) = \sum_{k=1}^{N}(Y_{k,N} - \mathbb{E}[Y_{k,N}]),$$

where

$$Y_{k,N} = \sqrt{N}\sum_{j=1}^{k}\frac{\mathbf{1}_{\{A_k\le j\}}}{j(N+1-j)}. \tag{29}$$

The summands $Y_{1,N},\dots,Y_{N,N}$ are independent. We wish to show that $\sqrt{N}(Y_N - \mathbb{E}[Y_N])$ is asymptotically normal with asymptotic variance $N\mathrm{var}(Y_N)$. From the independence of the summands,

$$\mathbb{E}[e^{it\sqrt{N}(Y_N-\mathbb{E}[Y_N])}] = \prod_{k=1}^{N}\mathbb{E}[e^{it(Y_{k,N}-\mathbb{E}[Y_{k,N}])}].$$

Then, we note that, for N sufficiently large and using that $\frac{1}{j(N+1-j)} = \frac{1}{N+1}(\frac{1}{j}+\frac{1}{N+1-j})$,

$$|Y_{k,N}| \le \sqrt{N}\sum_{j=1}^{k}\frac{1}{j(N+1-j)} \le \sqrt{N}\sum_{j=1}^{N}\frac{1}{j(N+1-j)} \le \frac{3\log N}{\sqrt{N}}. \tag{30}$$

Therefore, we have that, for N sufficiently large and $t>0$,

$$\mathbb{E}[e^{it(Y_{k,N}-\mathbb{E}[Y_{k,N}])}] = \exp\big\{-(t^2/2)\mathrm{var}(Y_{k,N}) + O(|t|^3 m_{k,N})\big\},$$

where $m_{k,N} = \mathbb{E}[|Y_{k,N} - \mathbb{E}Y_{k,N}|^3]$ denotes the absolute third central moment. By (30), we have that $|Y_{k,N} - \mathbb{E}Y_{k,N}| \le \frac{3\log N}{\sqrt{N}}$, so that

$$m_{k,N} \le \frac{3\log N}{\sqrt{N}}\mathrm{var}(Y_{k,N}).$$

Hence

$$\begin{aligned}\mathbb{E}[e^{it\sqrt{N}(Y_N-\mathbb{E}[Y_N])}] &= \prod_{k=1}^{N}\mathbb{E}[e^{it(Y_{k,N}-\mathbb{E}[Y_{k,N}])}] = \prod_{k=1}^{N}e^{(-t^2\mathrm{var}(Y_{k,N})/2+O(|t|^3 m_{k,N})}\\ &= e^{-t^2\sigma_{2,N}^2/2}e^{O(|t|^3\sigma_{2,N}^2N^{-1/2}\log N)}.\end{aligned}$$

This completes the proof because

$$\sigma_{2,N}^2 = \sum_{k=1}^{N}\mathrm{var}(Y_{k,N}) = N\mathrm{var}(Y_N) = N\mathrm{var}\left(\sum_{j=1}^{N}\lambda_j^{-1}B_j\right) = NT_2(N) \to \sigma_2 = 4\zeta(3) - 2\zeta(2),$$

as shown by (45) in the appendix.

Appendix

In Section A we prove a couple of identities formulated as lemmas. Lemma 5 to Lemma 10 are all proven in an identical way by taking differences. We therefore leave out some of the details. We will denote the partial sums in these identities by $C(N), D(N),\dots$, instead of $C_N, D_N,\dots$, in order to distinguish them from the standard notation for random variables. In Section B, we apply these identities to obtain asymptotic expressions for the variance of X_N and Y_N.

A Identities

Lemma 5 *For all $N \geq 1$,*

$$C(N) = \sum_{j=1}^{N} \frac{1}{j} \sum_{k=N+1-j}^{N} \frac{1}{k} = \sum_{k=1}^{N} \frac{1}{k^2}. \tag{31}$$

The identity (31) was proved in [8] by induction. Earlier Coppersmith and Sorkin [2] have proved (31) also by induction. We give a new and simpler proof.

Proof: Clearly, $C(1) = 1$ and

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

Summing both sides from $N = 2$ to $N = M$, using $C(1) = 1$ and relabeling $M \to N$ then leads to the right hand side in (31). □

A related sum which we will need is

$$D(N) = \sum_{j=1}^{N} \frac{1}{j} \sum_{k=1}^{j} \frac{1}{k} = \frac{1}{2} \left(\sum_{k=1}^{N} \frac{1}{k} \right)^2 + \frac{1}{2} \sum_{k=1}^{N} \frac{1}{k^2}. \tag{32}$$

Relation (32) is straightforward by symmetry.

Lemma 6 *For all $N \geq 1$,*

$$F(N) = \sum_{k=1}^{N} \frac{1}{k} \sum_{j=N+1-k}^{N} \frac{1}{j^2} = \sum_{j=1}^{N} \frac{1}{j^2} \sum_{k=N+1-j}^{N} \frac{1}{k} = 2 \sum_{k=1}^{N} \frac{1}{k^3} - \sum_{k=1}^{N} \frac{1}{k^2} \sum_{j=1}^{k} \frac{1}{j}. \tag{33}$$

Proof: The first equality follows after reversion of the summations. Parallel to the proof of Lemma 5, the second equality is derived as

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

Writing $\frac{1}{N^2} \sum_{k=1}^{N} \frac{1}{k} = \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{N^3}$, and using $\frac{1}{j(N-j)} = \frac{1}{N}(\frac{1}{j} + \frac{1}{N-j})$ on the second summand, we find that

$$F(N) - F(N-1) = \frac{1}{N^3} - \frac{1}{N^2} \sum_{j=1}^{N-1} \frac{1}{j}.$$

As in the proof of Lemma 5 this leads to the quoted result by iteration from $F(1) = 1$. □

The next lemma states a somewhat more involved identity:

Lemma 7 *For all $N \geq 1$,*

$$G(N) = \sum_{k=1}^{N} \frac{1}{k} \left(\sum_{m=N+1-k}^{N} \frac{1}{m} \right)^2 = \sum_{k=1}^{N} \frac{1}{k^2} \sum_{m=1}^{k} \frac{1}{m}. \tag{34}$$

Proof:

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

Now, using identity (32),

$$\sum_{k=1}^{N-1} \frac{1}{(N-k)} \sum_{m=N+1-k}^{N-1} \frac{1}{m} = \sum_{j=1}^{N-1} \frac{1}{j} \sum_{m=j+1}^{N-1} \frac{1}{m} = D(N-1) - \sum_{k=1}^{N-1} \frac{1}{j^2} = \frac{1}{2} \left(\sum_{n=1}^{N-1} \frac{1}{n} \right)^2 - \frac{1}{2} \sum_{n=1}^{N-1} \frac{1}{n^2},$$

and the partial fractions result:

$$\frac{1}{k(N-k)^2} = \frac{1}{N^2 k} + \frac{1}{N^2 (N-k)} + \frac{1}{N (N-k)^2}, \tag{35}$$

we arrive at

$$G(N) - G(N-1) = \frac{1}{N} \left(\sum_{m=1}^{N} \frac{1}{m} \right)^2 - \frac{1}{N} \left(\sum_{n=1}^{N-1} \frac{1}{n} \right)^2 - \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} = \frac{1}{N^2} \sum_{m=1}^{N} \frac{1}{m}.$$

As before we obtain the result by iteration. □

Lemma 8 *For all $N \geq 1$,*

$$L(N) = \sum_{k=1}^{N} \sum_{j=1}^{k} \frac{1}{kj} \sum_{n=N+1-j}^{N} \frac{1}{n} = \sum_{k=1}^{N} \frac{1}{k^3}. \tag{36}$$

Proof: After a tedious, but straightforward, computation we get

$$L(N) - L(N-1) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{j} \sum_{n=N+1-j}^{N} \frac{1}{n} - \frac{1}{N} \sum_{k=1}^{N-1} \frac{1}{k} \sum_{j=N-k}^{N-1} \frac{1}{j}.$$

With identity (31),

$$L(N) - L(N-1) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{j^2} - \frac{1}{N} \sum_{j=1}^{N-1} \frac{1}{j^2} = \frac{1}{N^3}.$$

This yields the proof. □

Lemma 9 *For all $N \geq 1$,*

$$R(N) = 2\sum_{k=1}^{N}\sum_{n=1}^{k}\frac{1}{kn}\sum_{j=N+1-n}^{N}\frac{1}{j^2} = 5\sum_{k=1}^{N}\frac{1}{k^4} - 2\sum_{n=1}^{N}\frac{1}{n^3}\sum_{k=1}^{n}\frac{1}{k} - \left(\sum_{k=1}^{N}\frac{1}{k^2}\right)^2. \tag{37}$$

Proof: Once more we compute the difference

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

Using the partial fraction result (35) on the last sum

$$\begin{aligned} R(N) - R(N-1) &= \frac{2}{N}\sum_{k=1}^{N}\frac{1}{k}\sum_{j=N+1-k}^{N}\frac{1}{j^2} - \frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k}\sum_{m=N-k}^{N-1}\frac{1}{m} - \frac{2}{N}\sum_{k=1}^{N-1}\frac{1}{k}\sum_{m=N-k}^{N-1}\frac{1}{m^2} \\ &= \frac{2}{N^2}\sum_{j=1}^{N}\frac{1}{j^2} + \frac{2}{N}\sum_{k=1}^{N-1}\frac{1}{k}\left(\sum_{j=N+1-k}^{N}\frac{1}{j^2} - \sum_{m=N-k}^{N-1}\frac{1}{m^2}\right) - \frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k}\sum_{m=N-k}^{N-1}\frac{1}{m} \\ &= \frac{2}{N^2}\sum_{j=1}^{N}\frac{1}{j^2} - \frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k}\sum_{m=N-k}^{N-1}\frac{1}{m} + \frac{2}{N}\sum_{k=1}^{N-1}\frac{1}{k}\left(\frac{1}{N^2} - \frac{1}{(N-k)^2}\right) \\ &= \frac{2}{N^2}\sum_{j=1}^{N}\frac{1}{j^2} - \frac{2}{N^2}\sum_{j=1}^{N-1}\frac{1}{j^2} + \frac{2}{N^3}\sum_{k=1}^{N-1}\frac{1}{k} - \frac{2}{N}\sum_{k=1}^{N-1}\frac{1}{k}\frac{1}{(N-k)^2}, \end{aligned}$$

where we have used (31). Using (35) to replace the last sum on the right side, we obtain,

$$R(N) - R(N-1) = \frac{2}{N^4} - \frac{2}{N^3}\sum_{k=1}^{N-1}\frac{1}{k} - \frac{2}{N^2}\sum_{m=1}^{N-1}\frac{1}{m^2} = \frac{6}{N^4} - \frac{2}{N^3}\sum_{k=1}^{N}\frac{1}{k} - \frac{2}{N^2}\sum_{m=1}^{N}\frac{1}{m^2}. \tag{38}$$

Summing both sides, and using $R(1) = 2$,

$$R(N) = 6\sum_{k=1}^{N}\frac{1}{k^4} - 2\sum_{n=1}^{N}\frac{1}{n^3}\sum_{k=1}^{n}\frac{1}{k} - 2\sum_{k=1}^{N}\frac{1}{k^2}\sum_{m=1}^{k}\frac{1}{m^2}.$$

Finally by an argument parallel to (32),

$$\sum_{k=1}^{N}\frac{1}{k^2}\sum_{m=1}^{k}\frac{1}{m^2} = \frac{1}{2}\left(\sum_{k=1}^{N}\frac{1}{k^2}\right)^2 + \frac{1}{2}\sum_{k=1}^{N}\frac{1}{k^4}. \tag{39}$$

Together, this yields the proof. □

Lemma 10 *For all $N \geq 1$,*

$$T(N) = \sum_{k=1}^{N}\frac{1}{k^2}\left(\sum_{m=N+1-k}^{N}\frac{1}{m}\right)^2 = 2\sum_{k=1}^{N}\frac{1}{k^3}\sum_{n=1}^{k}\frac{1}{n} - \left(\sum_{k=1}^{N}\frac{1}{k^2}\right)^2. \tag{40}$$

Proof: As before

$$T\left(N\right)-T\left(N-1\right)=\frac{1}{N^2}\left(\sum_{m=1}^{N}\frac{1}{m}\right)^2+\sum_{k=1}^{N-1}\frac{1}{k^2}\left(\sum_{m=N+1-k}^{N-1}\frac{1}{m}+\frac{1}{N}\right)^2-\sum_{k=1}^{N-1}\frac{1}{k^2}\left(\sum_{m=N+1-k}^{N-1}\frac{1}{m}+\frac{1}{N-k}\right)^2$$

$$=\frac{1}{N^2}\left(\sum_{m=1}^{N}\frac{1}{m}\right)^2+\sum_{k=1}^{N-1}\frac{1}{k^2}\left(\left(\frac{2}{N}-\frac{2}{N-k}\right)\sum_{m=N+1-k}^{N-1}\frac{1}{m}+\frac{1}{N^2}-\frac{1}{\left(N-k\right)^2}\right)$$

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

Now from taking partial fractions, (31) and (32),

$$\sum_{k=1}^{N-1}\frac{1}{k(N-k)}\sum_{m=N+1-k}^{N-1}\frac{1}{m}=\frac{C(N-1)}{N}-\frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k}+\frac{D(N-1)}{N}-\frac{1}{N}\sum_{m=1}^{N-1}\frac{1}{m^2}.$$

We can simplify this using the expressions for $C(N)$ in (31) and $D(N)$ in (32),

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

Using partial fraction expansion again yields

$$\sum_{k=1}^{N-1}\frac{1}{k^2}\frac{1}{\left(N-k\right)^2}=\frac{4}{N^3}\sum_{k=1}^{N-1}\frac{1}{k}+\frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k^2}\tag{41}$$

Combining these results then gives

$$T\left(N\right)-T\left(N-1\right)=\frac{1}{N^2}\left(\sum_{n=1}^{N}\frac{1}{n}\right)^2-\frac{1}{N^2}\left(\sum_{n=1}^{N-1}\frac{1}{n}\right)^2+\frac{4}{N^3}\sum_{k=1}^{N-1}\frac{1}{k}-\frac{1}{N^2}\sum_{n=1}^{N-1}\frac{1}{n^2}+\frac{1}{N^2}\sum_{k=1}^{N-1}\frac{1}{k^2}$$

$$-\frac{4}{N^3}\sum_{k=1}^{N-1}\frac{1}{k}-\frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k^2}$$

$$=\frac{1}{N^2}\left(\sum_{n=1}^{N-1}\frac{1}{n}+\frac{1}{N}\right)^2-\frac{1}{N^2}\left(\sum_{n=1}^{N-1}\frac{1}{n}\right)^2-\frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k^2}$$

$$=\frac{2}{N^3}\sum_{n=1}^{N-1}\frac{1}{n}+\frac{1}{N^4}-\frac{2}{N^2}\sum_{k=1}^{N-1}\frac{1}{k^2}=\frac{2}{N^3}\sum_{n=1}^{N}\frac{1}{n}+\frac{1}{N^4}-\frac{2}{N^2}\sum_{k=1}^{N}\frac{1}{k^2},$$

from which by summing both sides from $N=2$ to $N=M$ (and then $M\to N$ again)

$$T\left(N\right)-1=2\sum_{k=2}^{N}\frac{1}{k^3}\sum_{n=1}^{k}\frac{1}{n}+\sum_{k=2}^{N}\frac{1}{k^4}-2\sum_{k=2}^{N}\frac{1}{k^2}\sum_{n=1}^{k}\frac{1}{n^2}.$$

Using (39) then yields the proof. □

B The asymptotic results for the variances

In this section we use the identities of Section A, in order to compute simplified expressions for $T_2(N)$ and $T_1(N)$. Consequently we use these results for the asymptotic variance of W_N.

The sum $T_2(N)$ (compare (17)) equals

$$T_2(N) = R_1(N) - R_2(N),$$

where

$$R_1(N) = 2\sum_{n=1}^{N}\sum_{m=n}^{N}\frac{1}{(N+1-n)\,m\,(N+1-m)}\sum_{j=m}^{N}\frac{1}{j},$$

and

$$R_2(N) = 2\sum_{n=1}^{N}\sum_{m=n}^{N}\frac{1}{(N+1-n)\,(N+1-m)}\sum_{j=m}^{N}\frac{1}{j^2}.$$

We start with the sum $R_1(N)$, and interchange the sums to obtain

$$\begin{aligned} R_1(N) &= 2\sum_{m=1}^{N}\sum_{j=m}^{N}\frac{1}{j}\sum_{n=1}^{m}\frac{1}{(N+1-n)\,m\,(N+1-m)} \\ &= 2\sum_{k=1}^{N}\frac{1}{k}\sum_{n=1}^{k}\frac{1}{n\,(N+1-n)}\sum_{m=1}^{n}\frac{1}{N+1-m}. \end{aligned}$$

Splitting $(n(N+1-n))^{-1}$ into two parts,

$$R_1(N) = \frac{2}{N+1}\sum_{k=1}^{N}\frac{1}{k}\sum_{n=1}^{k}\frac{1}{n}\sum_{j=N+1-n}^{N}\frac{1}{j} + \frac{2}{N+1}\sum_{k=1}^{N}\frac{1}{k}\sum_{m=N+1-k}^{N}\frac{1}{m}\sum_{j=m}^{N}\frac{1}{j}.$$

The first sum equals $2L(N)/(N+1)$, where $L(N)$ was simplified in Lemma 8.

By the same method that we used in (32) to obtain $D(N)$, we find

$$\sum_{m=N+1-k}^{N}\frac{1}{m}\sum_{j=m}^{N}\frac{1}{j} = \frac{1}{2}\left(\sum_{j=N+1-k}^{N}\frac{1}{j}\right)^2 + \frac{1}{2}\sum_{j=N+1-k}^{N}\frac{1}{j^2}. \tag{42}$$

This implies using (33),

$$\begin{aligned} \sum_{k=1}^{N}\frac{1}{k}\sum_{m=N+1-k}^{N}\frac{1}{m}\sum_{j=m}^{N}\frac{1}{j} &= \frac{1}{2}\sum_{k=1}^{N}\frac{1}{k}\left(\sum_{m=N+1-k}^{N}\frac{1}{m}\right)^2 + \frac{1}{2}\sum_{k=1}^{N}\frac{1}{k}\sum_{m=N+1-k}^{N}\frac{1}{m^2} \\ &= \frac{1}{2}\left(G(N)+F(N)\right) = \sum_{k=1}^{N}\frac{1}{k^3}. \end{aligned}$$

Combining all,

$$R_1(N) = \frac{4}{N+1}\sum_{k=1}^{N}\frac{1}{k^3} = \frac{4\zeta(3)}{N}(1+O(N^{-2})). \tag{43}$$

We now turn to the sum

$$R_2(N) = 2\sum_{n=1}^{N}\sum_{m=n}^{N}\frac{1}{(N+1-n)(N+1-m)}\sum_{j=m}^{N}\frac{1}{j^2}$$
$$= 2\sum_{k=1}^{N}\sum_{m=N+1-k}^{N}\frac{1}{k(N+1-m)}\sum_{j=m}^{N}\frac{1}{j^2} = 2\sum_{k=1}^{N}\sum_{n=1}^{k}\frac{1}{kn}\sum_{j=N+1-n}^{N}\frac{1}{j^2} = R(N),$$

and $R(N)$ was simplified in Lemma 8.

Together we find

$$T_2(N) = R_1(N) - R_2(N) = R_1(N) - R(N)$$
$$= \frac{4}{N+1}\sum_{k=1}^{N}\frac{1}{k^3} - 5\sum_{k=1}^{N}\frac{1}{k^4} + 2\sum_{n=1}^{N}\frac{1}{n^3}\sum_{k=1}^{n}\frac{1}{k} + \left(\sum_{k=1}^{N}\frac{1}{k^2}\right)^2. \tag{44}$$

From (38) we find that $R(N) - R(N-1) = \frac{-2\zeta(2)}{N^2} + O(\frac{\log N}{N^3})$, so that, by summation,

$$R(N) = R + \frac{2\zeta(2)}{N} + O\left(\frac{\log N}{N^2}\right),$$

where

$$R = 5\sum_{k=1}^{\infty}\frac{1}{k^4} - 2\sum_{n=1}^{\infty}\frac{1}{n^3}\sum_{k=1}^{n}\frac{1}{k} - \left(\sum_{k=1}^{\infty}\frac{1}{k^2}\right)^2 = \frac{5}{2}\zeta(4) - \zeta(2)^2 = 0,$$

The first equality follows by (19), while the second follows by [4, (9.542)1]. Thus, we obtain the asymptotics

$$T_2(N) = R_1(N) - R(N) = \frac{4\zeta(3) - 2\zeta(2)}{N} + O\left(\frac{\log N}{N^2}\right). \tag{45}$$

We finally turn to the second sum $T_1(N)$ (see (16)), which sum is equal to the sum $T(N)$ displayed in Lemma 10. Therefore,

$$T_1(N) = \sum_{n=1}^{N}\frac{1}{(N+1-n)^2}\left(\sum_{j=n}^{N}\frac{1}{j}\right)^2 = 2\sum_{k=1}^{N}\frac{1}{k^3}\sum_{n=1}^{k}\frac{1}{n} - \left(\sum_{k=1}^{N}\frac{1}{k^2}\right)^2. \tag{46}$$

From the proof of Lemma 10, the difference

$$T_1(N) - T_1(N-1) = \frac{2}{N^3}\sum_{n=1}^{N}\frac{1}{n} + \frac{1}{N^4} - \frac{2}{N^2}\sum_{k=1}^{N}\frac{1}{k^2} = -\frac{2}{N^2}\sum_{k=1}^{N}\frac{1}{k^2} + O\left(\frac{\log N}{N^3}\right),$$

which shows, by summation that for large N, $T_1(N)$ behaves asymptotically as

$$T_1(N) = T_1 + \frac{2\zeta(2)}{N} + O\left(\frac{\log N}{N^2}\right), \tag{47}$$

where we write

$$T_1 = 2\sum_{k=1}^{\infty}\frac{1}{k^3}\sum_{n=1}^{k}\frac{1}{n} - \left(\sum_{k=1}^{\infty}\frac{1}{k^2}\right)^2 = \frac{5}{2}\zeta(4) - \zeta^2(2), \tag{48}$$

and, again, the second equality follows by (19), while Equation [4, (9.542)1] implies that $T_1 = 0$.

Remark. We note that $T_1 = 0$ and $Q = 0$ can also be proved directly from the first equality in (46) without resorting to [1, Corollary 4, main theorem]. We split $T_1(N)$ as

$$T_1(N) = \sum_{n=1}^{N/2} \frac{1}{(N+1-n)^2} \left(\sum_{j=n}^{N} \frac{1}{j} \right)^2 + \sum_{n=1+N/2}^{N} \frac{1}{(N+1-n)^2} \left(\sum_{j=n}^{N} \frac{1}{j} \right)^2$$

For N sufficiently large, the first sum is bounded as

$$\sum_{n=1}^{N/2} \frac{1}{(N+1-n)^2} \left(\sum_{j=n}^{N} \frac{1}{j} \right)^2 < \frac{3\log^2 N}{N},$$

while the second sum is, assuming that $N/2$ is an integer,

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

For $m \leq N/2$, we bound the sum between brackets as

$$\sum_{j=N+1-m}^{N} \frac{1}{j} < \int_{N-m}^{N} \frac{dx}{x} = -\log\left(1 - \frac{m}{N}\right) = \frac{m}{N} + O\left(\frac{m^2}{N^2}\right)$$

Hence, we obtain

$$\sum_{n=1+N/2}^{N} \frac{1}{(N+1-n)^2} \left(\sum_{j=n}^{N} \frac{1}{j} \right)^2 < \sum_{m=1}^{N/2} \frac{1}{m^2} \left(\frac{m}{N}\right)^2 + O\left(\frac{1}{N^3} \sum_{m=1}^{N/2} m \right) = O\left(\frac{1}{N}\right).$$

Combining both estimates shows that $T_1(N) \to 0$ as $N \to \infty$, hence $T_1 = 0$.

From (17), we find that $T_2 = \lim_{N\to\infty} T_2(N)$ is

$$T_2 = -5 \sum_{k=1}^{\infty} \frac{1}{k^4} + 2 \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{k=1}^{n} \frac{1}{k} + \left(\sum_{k=1}^{\infty} \frac{1}{k^2} \right)^2. \tag{49}$$

Equation (48) shows that

$$T_2 = 2 \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{k=1}^{n} \frac{1}{k} + \zeta^2(2) - 5\zeta(4) = -T_1 = 0.$$

The definition (18) of $Q(N)$ together with (2) gives

$$Q(N) = T_1(N) + T_2(N) + O\left(N^{-1}\right)$$

from which $Q = \lim_{N\to\infty} Q(N)$ follows as

$$Q = T_1 + T_2 = 0$$

This result proves (19) independently from Borwein's paper [1, Corollary 4, main theorem].

Acknowledgements

The work of RvdH was supported in part by the Netherlands Organisation for Scientific Research (NWO). The authors are grateful to a referee for many remarks that improve the presentation in the paper, and, in particular, for outlining the remark in Appendix B.

References

[1] D. Borwein and J.M. Borwein. On an intriguing integral and some series related to $\zeta(4)$. *Proc. Amer. Math. Soc.*, 123(4): 1191-1198, 1995.

[2] D. Coppersmith and G. B. Sorkin. constructive bounds and exact expectations for the random assignment problem. *Random Structures and Algorithms*, Vol 15, No. 2, 113-144, 1999.

[3] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, 10:47–56, 1985.

[4] I. S. Gradsteyn and I. M. Ryzhik. Tables of Integrals, Series and Products (Fifth Edition). Academic Press, London, 1994.

[5] S. Janson. The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Structures and Algorithms*, 7(4):337–356, 1995.

[6] S. Janson. On concentration of probability. Contemporary Combinatorics, ed. B. Bollobás, Bolyai Soc. Math. Stud. **10**, János Bolyai Mathematical Society, Budapest, 289-301, 2002.

[7] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. First passage percolation on the random graph. *Probability in the Engineering and Informational Sciences (PEIS)*, 15:225–237, 2001.

[8] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Size and weight of shortest path trees with exponential link weights. *Combinatorics, Probability and Computing*, to appear 2006.

[9] P. Van Mieghem, G. Hooghiemstra, and R. van der Hofstad. A scaling law for the hopcount in internet. *Delft University of Technology, report 2000125*, 2000.

[10] J. Wästlund. Evaluation of Janson's constant for the variance in the random minimum spanning tree problem. *Linköping studies in Mathematics, No. 7, Series Editor: Bengt Ove Turesson, www.ep.liu.se/ea/lsm/2005/007*, 2005.