Combinatorics, Probability and Computing (2010) 19, 1–19. © Cambridge University Press 2009
doi:10.1017/S0963548309990204

The Longest Minimum-Weight Path in a Complete Graph

LOUIGI ADDARIO-BERRY,[1] NICOLAS BROUTIN[2]
and GÁBOR LUGOSI[3]

[1]Département de Mathématiques et de Statistique, Université de Montréal,
CP 6128, Succ. Centre-ville, Montreal, Quebec, H3C 3J7, Canada
(e-mail: louigi@gmail.com)

[2]Projet Algorithms, INRIA Rocquencourt, 78153 Le Chesnay, France
(e-mail: nicolas.broutin@m4x.org)

[3]Department of Economics, Pompeu Fabra University, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain
(e-mail: lugosi@upf.es)

Received 1 September 2008; revised 6 February 2009; first published online 22 June 2009

We consider the minimum-weight path between any pair of nodes of the n-vertex complete graph in which the weights of the edges are i.i.d. exponentially distributed random variables. We show that the longest of these minimum-weight paths has about $\alpha^\star \log n$ edges, where $\alpha^\star \approx 3.5911$ is the unique solution of the equation $\alpha \log \alpha - \alpha = 1$. This answers a question posed by Janson [8].

1. Introduction

We consider the complete graph K_n on n vertices $[n] := \{1, \ldots, n\}$, augmented with independent exponential mean n (or Exp(n) for short) edge weights $\{X_e : e \in E(K_n)\}$. For any subgraph $H = (V(H), E(H))$ of K_n we write $|H|$ for $|E(H)|$, and let

$$w(H) = \sum_{e \in E(H)} X_e.$$

For $i, j \in [n]$ we let P_{ij} be the minimum-weight path from i to j and write $W_{ij} = w(P_{ij})$ (adopting the convention $P_{ii} = \emptyset$, $w_{ii} = 0$). For any fixed vertex i, $\cup_{j \neq i} P_{ij}$ is a tree, the *shortest path tree* SPT_i rooted at i. For $t \geqslant 0$ and $k \in [n]$, we let $\mathrm{SPT}_i(t)$ be the subtree of SPT_i induced by nodes j with $W_{ij} \leqslant t$.

Let t_k^i be the first time for which $\mathrm{SPT}_i(t)$ contains at least $k+1$ vertices, and write $t_0^i = 0$. Due to the memoryless property of the exponential distribution, for each $k \in [n]$, the location of attachment of the vertex added at time t_k^i is uniform among the vertices of $\mathrm{SPT}_i(t_{k-1}^i)$; in other words, $\mathrm{SPT}_i(t_k^i)$ is distributed like a *random recursive tree* with k vertices. Random recursive

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

trees have been well-studied [13]; in particular, it is known that the depth (number of edges on the path from a uniformly random node to the root) in a recursive tree with n vertices is asymptotic to $\log n$ in probability (see Devroye [3]), and the height (greatest number of edges on any path starting from the root) is asymptotic to $e \log n$ in probability (see Devroye [3] and Pittel [9]).

A variety of authors have studied the *weighted* structure of a single tree SPT_1 (or SPT_i for any i), and of the family of shortest paths $\{P_{ij} : i, j \in [n]\}$. Van der Hofstad, Hooghiemstra and van Mieghem [14, 15] prove that, letting W be the total weight of SPT_1,

$$\frac{(W - \zeta(2))}{\sqrt{n}} \xrightarrow[n\to\infty]{} \mathcal{N}(0, 4\zeta(3))$$

in distribution. Here, ζ denotes the Riemann zeta function $\zeta(s) := \sum_{k\geqslant 1} k^{-s}$. Janson [8] proved the beautiful result that

$$\frac{W_{12}}{\log n} \xrightarrow[n\to\infty]{} 1, \quad \frac{\max_{1\leqslant j\leqslant n} W_{1j}}{\log n} \xrightarrow[n\to\infty]{} 2 \quad \text{and} \quad \frac{\max_{1\leqslant i,j\leqslant n} W_{ij}}{\log n} \xrightarrow[n\to\infty]{} 3$$

in probability, and provided more detailed information about the distribution of W_{12} and of $\max_{1\leqslant j\leqslant n} W_{1j}$. (Janson's results in fact hold for some edge weight distributions other than exponential: see Section 5.) On the other hand, the unweighted structure of shortest path trees has been studied by Hooghiemstra and Van Mieghem [7], who derived the limiting distribution of the pair $(W_{12}, |P_{12}|)$ suitably rescaled.

It is immediate from the aforementioned results for the depth of nodes and the height in random recursive trees that

$$\frac{|P_{12}|}{\log n} \xrightarrow[n\to\infty]{} 1 \quad \text{and} \quad \frac{\max_{1\leqslant j\leqslant n} |P_{1j}|}{\log n} \xrightarrow[n\to\infty]{} e$$

in probability. Furthermore, as noted by Janson [8], the tail bounds for the height of random recursive trees established by Devroye [3] imply that

$$\limsup_{n\to\infty} \frac{\max_{1\leqslant i,j\leqslant n} |P_{ij}|}{\log n} \leqslant \alpha^\star \quad \text{in probability,} \tag{1.1}$$

where $\alpha^\star \approx 3.5911$ is the unique solution of $\alpha \log \alpha - \alpha = 1$. Janson [8] asked whether there is a constant $c \in [e, \alpha^\star]$ such that $\max_{1\leqslant i,j\leqslant n} |P_{ij}|/\log n \to c$ in probability. It is the purpose of this note to answer Janson's question in the affirmative.

Theorem 1.1. *For any function $\omega(n)$ tending to $+\infty$ with n, for some constant $L > 0$, and for all $\delta \in (0, 1)$, for all n sufficiently large, with probability greater than $1 - \delta$,*

$$\alpha^\star \log n - L \log\log n \leqslant \max_{1\leqslant i,j\leqslant n} |P_{ij}| \leqslant \alpha^\star \log n + \omega(n).$$

Furthermore,

$$\alpha^\star \log n - L \log\log n \leqslant \mathbf{E}\Big[\max_{1\leqslant i,j\leqslant n} |P_{ij}|\Big] \leqslant \alpha^\star \log n + L.$$

It follows immediately from the first part of Theorem 1.1 that $\max_{1\leqslant i,j\leqslant n} |P_{ij}|/\log n \to \alpha^\star$ in probability. The easy half of the result is the upper bound, as (1.1) and other results from Devroye [3] will make proving the upper bounds of the theorem a routine exercise. The non-trivial part

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

of the result is the lower bound, which we prove using the second moment method, applied to a suitably defined set of shortest paths with special properties that make them amenable to analysis. Before we get going, a brief remark is in order. We have to work with exponential edge weights in order to exploit the memoryless property of the exponential distribution. In the conclusion (Section 5) we explain how to extend our results to a range of other edge weights that have a finite and positive density at 0, including uniform edge weights.

2. A useful lemma and the upper bound

For $i \in [n]$ and $k \in [n-1]$, let $\tau_k^i = t_{k+1}^i - t_k^i$ denote the kth inter-arrival time. By the memoryless property of the exponential distribution, $\{\tau_k^i, 1 \leqslant k < n\}$ are independent. Furthermore, τ_k^i is the minimum of $k(n-k)$ independent exponential random variables with mean n, hence τ_k^i has an exponential distribution with rate parameter $k(n-k)/n$. Using this fact, the following lemma provides upper bounds on the size of $\mathrm{SPT}_i(t)$.

Lemma 2.1. *For all $t \geqslant 0$ and integers $m \in [n-1]$ with $m \geqslant e^t$,*

$$\mathbf{P}\{|\mathrm{SPT}_i(t)| \geqslant m\} \leqslant 3 \cdot \sqrt{m/e^t} \cdot e^{-m/e^t}.$$

Proof. Observe that, for any $c > 0$,

$$\mathbf{P}\{|\mathrm{SPT}_i(t)| \geqslant m\} = \mathbf{P}\Big\{\sum_{k=1}^{m} \tau_k^i \leqslant t\Big\} = \mathbf{P}\Big\{\prod_{k=1}^{m} \exp\big(-c\tau_k^i\big) \geqslant e^{-ct}\Big\}.$$

By Markov's inequality and the independence of the inter-arrival times $\tau_1^i, \ldots, \tau_{n-1}^i$, we have

$$\mathbf{P}\{|\mathrm{SPT}_i(t)| \geqslant m\} \leqslant e^{ct} \prod_{k=1}^{m} \mathbf{E}\big[\exp\big(-c\tau_k^i\big)\big] \leqslant e^{ct} \prod_{k=1}^{m} \frac{k}{k+c} \leqslant e^{ct} \cdot \frac{\Gamma(c+1)}{m^c},$$

where Γ is the Gamma function. Using the strengthening of Stirling's formula $\Gamma(c+1) \leqslant \sqrt{2\pi c}(c/e)^c e^{1/12}$ (see, *e.g.*, [10]), and setting $c = m/e^t$, we obtain

$$\mathbf{P}\{|\mathrm{SPT}_i(t)| \geqslant m\} \leqslant \sqrt{2\pi} e^{1/12} \cdot \sqrt{c} \cdot e^{-c} \leqslant 3 \cdot \sqrt{c} \cdot e^{-c},$$

as claimed. □

We will also use the following result from Devroye [3], providing an upper bound on the heights of random recursive trees. In the following, $h(T)$ denotes the height of a rooted tree T.

Theorem 2.2 ([3], Theorem 9). *Let T_m be a random recursive tree with $m \geqslant 2$ nodes. Then, for $x > 1$,*

$$\mathbf{P}\{h(T_m) \geqslant x \log m\} \leqslant e^{x-1} m^{x - x\log x}.$$

The upper bounds in Theorem 1.1 follow immediately from Theorem 2.2.

Proof of Theorem 1.1, upper bounds. For any $t \geqslant 0$, by a union bound we have, for $n > 2$,

$$\mathbf{P}\Big\{\max_{1 \leqslant i,j \leqslant n} |P_{ij}| \geqslant \alpha^\star \log n + t\Big\} \leqslant n\mathbf{P}\Big\{\max_{1 \leqslant j \leqslant n} |P_{1j}| \geqslant \alpha^\star \log n + t\Big\}.$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

The event on the right-hand side above holds if SPT_1 has height at least $\alpha^\star \log n + t$. Since SPT_1 is distributed like a random recursive tree on n nodes, to further bound the probability on the right-hand side, we may use Theorem 2.2 with m replaced by n and $x = \alpha^\star + t/\log n$. In order to simplify the obtained expression, observe also that, by the definition of $\alpha^\star \geqslant e$, for $t \geqslant 0$,

$$x \log x - x = \left(\alpha^\star + \frac{t}{\log n}\right) \log\left(\alpha^\star + \frac{t}{\log n}\right) - \left(\alpha^\star + \frac{t}{\log n}\right) \geqslant 1 + \frac{t}{\log n}.$$

We obtain that

$$\mathbf{P}\Big\{ \max_{1 \leqslant j \leqslant n} |P_{1j}| \geqslant \alpha^\star \log n + t \Big\} \leqslant e^x n^{x - x \log x} \leqslant e^{\alpha^\star + t/\log n} n^{-1 + t/\log n},$$

and therefore

$$\mathbf{P}\Big\{ \max_{1 \leqslant i,j \leqslant n} |P_{ij}| \geqslant \alpha^\star \log n + t \Big\} \leqslant e^{\alpha^\star + t/\log n} e^{-t}. \tag{2.1}$$

Choosing $t = t(n) \to \infty$ proves the upper bound in probability. Now, using the bound (2.1), we have, for $n \geqslant 8$,

$$\begin{aligned}
\mathbf{E}\Big[\max_{1 \leqslant i,j \leqslant n} |P_{ij}| \Big] &\leqslant \alpha^\star \log n + 1 + \sum_{\ell=1}^{\infty} 2^\ell \mathbf{P}\Big\{ \max_{1 \leqslant i,j \leqslant n} |P_{ij}| \geqslant \alpha^\star \log n + 2^{\ell-1} \Big\} \\
&\leqslant \alpha^\star \log n + 1 + e^{\alpha^\star} \cdot \sum_{\ell=1}^{\infty} 2^\ell \exp(-2^{\ell-2}) \\
&= \alpha^\star \log n + O(1),
\end{aligned}$$

proving the upper bound in expectation. □

3. Towards the lower bound

Fix a path $P = v_1, \ldots, v_{k+1}$, $k \leqslant 12 \log n$ (where 12 is somewhat arbitrary, the key point being that it is bigger than $3\alpha^\star$). It is easily seen that for $w > 0$, letting $\mathrm{Po}(w)$ denote a Poisson mean w random variable, we have

$$\mathbf{P}\{w(P) \leqslant wn\} = \mathbf{P}\{\mathrm{Po}(w) \geqslant k\} = \sum_{i=k}^{\infty} \frac{w^k e^{-w}}{k!}.$$

When $w = o(1)$, the above sum is dominated by its first term and $e^{-w} = 1 - o(1)$, so we have

$$\mathbf{P}\{w(P) \leqslant wn\} = (1 + o(1)) \frac{w^k}{k!}. \tag{3.1}$$

We next show that, *given* that P has small weight, it is very likely to be the minimum-weight path between its end points. More precisely, let $P^\star(v, w)$ be the minimum-weight path in K_n between two vertices v and w. Then we have the following.

Lemma 3.1. *Fix $c > 2$. For all n sufficiently large, for all $\epsilon > c \log\log n / \log n$ and any path $P = v_1, \ldots, v_{k+1}$ in K_n with $1 \leqslant k \leqslant 12 \log n$, we have*

$$\mathbf{P}\{P \neq P^\star(v_1, v_{k+1}) \mid w(P) \leqslant (1 - \epsilon) \log n\} \leqslant \frac{13k^2}{n^\epsilon}.$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

The restriction $c > 2$ is not necessary in the above lemma, but the upper bound becomes trivial when $\epsilon \leqslant 2 \log\log n / \log n$.

Proof. Let $\mathrm{SPT}_v(t)$ denote the shortest path tree started from a vertex v and stopped at time t. For $k \geqslant 2$, if $P \neq P^\star(v_1, v_{k+1})$ then it must be the case that, for some $i = 1, \ldots, k$, in $K_n \setminus E(P)$, $\mathrm{SPT}_{v_i}((1-\epsilon)\log n)$ contains one of $v_{i+1}, \ldots, v_{k+1}$.

Allowing connections using the edges along the path P only increases the probability of this event. Let K_n^P be K_n where the edge weights X_e, $e \in E(P)$, along the path P have been replaced by independent copies. Let A_i be the event that, in K_n^P, $\mathrm{SPT}_{v_i}((1-\epsilon)\log n)$ contains one of $v_{i+1}, \ldots, v_{k+1}$ for some $i = 1, \ldots, k$. The remark above then implies by the union bound that

$$\mathbf{P}\{P \neq P^\star(v_1, v_{k+1}) \mid w(P) \leqslant (1-\epsilon)\log n\} \leqslant \sum_{i=1}^{k} \mathbf{P}\{A_i\}. \tag{3.2}$$

Let N_i denote the number of nodes in $\mathrm{SPT}_{v_i}((1-\epsilon)\log n)$ in K_n^P. Observe that, conditioning on N_i, the probability that A_i does *not* occur is

$$\begin{aligned} 1 - \mathbf{P}\{A_i \mid N_i\} &= \left(\prod_{j=1}^{N_i}\left(1 - \frac{k-i+1}{n-j}\right)\right) \cdot \mathbf{1}\big[n - N_i > k - i + 1\big] \\ &\geqslant \max\left(1 - \frac{(k-i+1)N_i}{n-N_i}, 0\right), \end{aligned} \tag{3.3}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. Let $E_i(0)$ be the event that $N_i \leqslant 2n^{1-\epsilon}$, for $j = 1, \ldots, \lfloor \log_2(n^\epsilon/2k) \rfloor - 1$, let $E_i(j)$ be the event that $2^j n^{1-\epsilon} < N_i \leqslant 2^{j+1} n^{1-\epsilon}$, and let $E_i(\lfloor \log_2(n^\epsilon/2k) \rfloor)$ be the event that $N_i > 2^{\lfloor \log_2(n^\epsilon/2k) \rfloor} n^{1-\epsilon}$. Expressing $\mathbf{P}\{A_i\}$ as a sum of conditional probabilities, we have

$$\begin{aligned} \mathbf{P}\{A_i\} &= \sum_{j=0}^{\lfloor \log_2(n^\epsilon/2k) \rfloor} \mathbf{P}\{A_i \mid E_i(j)\} \cdot \mathbf{P}\{E_i(j)\} \\ &\leqslant \sum_{j=0}^{\lfloor \log_2(n^\epsilon/2k) \rfloor} \mathbf{P}\{A_i \mid E_i(j)\} \cdot 3\sqrt{2^j} e^{-2^j}, \end{aligned} \tag{3.4}$$

by Lemma 2.1. Using (3.3) to bound $\mathbf{P}\{A_i \mid E_i(j)\}$,

$$\begin{aligned} \mathbf{P}\{A_i\} &\leqslant \sum_{j=0}^{\lfloor \log_2(n^\epsilon/2k) \rfloor - 1} \frac{(k-i+1)2^{j+1} n^{1-\epsilon}}{n - 2^{j+1} n^{1-\epsilon}} \cdot \frac{3\sqrt{2^j}}{e^{2^j}} + \frac{3\sqrt{2^{\lfloor \log_2(n^\epsilon/2k) \rfloor}}}{e^{2^{\lfloor \log_2(n^\epsilon/2k) \rfloor}}} \\ &\leqslant \frac{6(k-i+1)}{n^\epsilon(1-1/2k)} \sum_{j=0}^{\infty} \frac{2^{3j/2}}{e^{2^j}} + \frac{3\sqrt{2^{\lfloor \log_2(n^\epsilon/2k) \rfloor}}}{e^{2^{\lfloor \log_2(n^\epsilon/2k) \rfloor}}} \\ &\leqslant \frac{12(k-i+1)}{n^\epsilon} + n^{-\epsilon}, \end{aligned} \tag{3.5}$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

where we have used the fact that $\sum_{j\geqslant 0} 2^{3j/2}e^{-2^j} \leqslant 1$ and the last inequality holds for all $k \leqslant 12\log n$ and n sufficiently large. It follows that, since $k \geqslant 1$,

$$\sum_{i=1}^{k} \mathbf{P}\{A_i\} \leqslant \frac{1}{n^\epsilon}\left(12\binom{k+1}{2} + k\right) \leqslant \frac{13k^2}{n^\epsilon},$$

which, together with (3.2), completes the proof. □

The following fact follows immediately *from the proof* of Lemma 3.1 (there are just fewer terms in the sum (3.4)); it will be useful later.

Corollary 3.2. *Fix $c > 2$. For all n sufficiently large, for all $\epsilon > c\log\log n/\log n$ and any path $P = v_1, \dots, v_{k+1}$ in K_n with $1 \leqslant k \leqslant 12\log n$, we have*

$$\mathbf{P}\{P \neq P^\star(v_1, v_{k+1}) \mid (1-2\epsilon)\log n \leqslant w(P) \leqslant (1-\epsilon)\log n\} \leqslant \frac{13k^2}{n^\epsilon}.$$

It follows from Lemma 3.1, at least intuitively, that to get a lower bound of k on the length of the longest minimum-weight path, we can instead bound from below the probability that there is *some* path with k edges of weight $(1-\epsilon)\log n$. Given a positive integer k and real $\epsilon > 0$, let $\mathcal{P}_{k,\epsilon} = \mathcal{P}_{k,\epsilon}(n)$ be the set of paths with k edges and weight at most $(1-\epsilon)\log n$ in K_n. As a first step, we remark that by (3.1),

$$\mathbf{E}|\mathcal{P}_{k,\epsilon}| \sim n^{k+1}\left(\frac{(1-\epsilon)\log n}{n}\right)^k \frac{1}{k!} \sim \frac{n}{\sqrt{2\pi k}}\left(\frac{e\log n}{k}\right)^k (1-\epsilon)^k. \tag{3.6}$$

Let α_ϵ be the solution of

$$\alpha\log\alpha - \alpha(1+\log(1-\epsilon)) = 1, \tag{3.7}$$

so $\alpha_\epsilon < \alpha^\star$ and $\alpha_\epsilon \to \alpha^\star > e$ as $\epsilon \to 0$. When $t = k - \alpha_\epsilon\log n = o(\sqrt{\log n})$, we then have

$$\mathbf{E}|\mathcal{P}_{k,\epsilon}| = \frac{(1+o(1))}{\sqrt{2\pi\alpha_\epsilon\log n}}\left(\frac{e}{\alpha_\epsilon}\right)^t (1-\epsilon)^t.$$

Letting

$$k_\epsilon = \alpha_\epsilon\log n - \beta_\epsilon\log\log n \quad \text{where} \quad \beta_\epsilon = \frac{1}{2\big(1+\log\alpha_\epsilon - \log(1-\epsilon)\big)}, \tag{3.8}$$

the following estimate for $\mathbf{E}|\mathcal{P}_{k,\epsilon}|$, for k close to k_ϵ.

Lemma 3.3. *Let $\alpha_\epsilon, \beta_\epsilon$ and k_ϵ be defined as in* (3.7) *and* (3.8) *above. Given ϵ with $0 < \epsilon < 1/2$, there is a positive constant c_ϵ such that $\mathbf{E}|\mathcal{P}_{k_\epsilon,\epsilon}| = (1+o(1))c_\epsilon$ and, for $|t| = o(\sqrt{\log n})$,*

$$\mathbf{E}|\mathcal{P}_{k_\epsilon+t,\epsilon}| = (1+o(1))c_\epsilon \cdot \left(\frac{e}{\alpha_\epsilon}\right)^t (1-\epsilon)^t.$$

In particular, it follows that if $t = t(n) \to -\infty$ then $\mathbf{E}|\mathcal{P}_{k_\epsilon+t,\epsilon}(n)| \to \infty$. To derive a lower bound on the probability that there exists such a path, we use the second moment method, and now introduce the version of it we require. Given any random set $\mathcal{S}$ of paths in K_n and two paths P

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

and Q, let $q_{\mathcal{S}}(P,Q)$ be the probability that P and Q are both in $\mathcal{S}$, and let

$$\Delta(\mathcal{S}) = \sum_{P,Q} q_{\mathcal{S}}(P,Q), \tag{3.9}$$

where the sum is over pairs P,Q of distinct but intersecting paths in K_n. By Corollary 4.3.4 of Alon, Spencer and Erdős [1], we then have

$$\mathbf{P}\{|\mathcal{S}| = 0\} \leqslant \frac{1}{\mathbf{E}|\mathcal{S}|} + \frac{\Delta(\mathcal{S})}{(\mathbf{E}|\mathcal{S}|)^2}. \tag{3.10}$$

Given the preceding discussion, a natural choice for the set $\mathcal{S}$ would be $\mathcal{P}_{k_\epsilon+t,\epsilon}$, for some $t = t(n)$ tending to $-\infty$ with n. Unfortunately, for this choice of $\mathcal{S}$ and for the values of t we wish to consider, the quantity $\Delta(\mathcal{S})$ is too large for (3.10) to yield a useful bound. However, it is both useful and instructive to proceed as though this were our choice of $\mathcal{S}$, and see how far we can get.

Given paths P and Q in K_n, let

$$q_\epsilon(P,Q) = \mathbf{P}\{w(P) \leqslant (1-\epsilon)\log n, w(Q) \leqslant (1-\epsilon)\log n\}.$$

Also, for $t \in \mathbb{R}$, let

$$\Delta_t = \Delta_t(n,\epsilon) = \sum_{P,Q} q_\epsilon(P,Q), \tag{3.11}$$

where the sum is over pairs P,Q of distinct but intersecting paths with $\lceil k_\epsilon + t\rceil$ edges in K_n. (So, Δ_t is just $\Delta(\mathcal{S})$ when $\mathcal{S}$ is the set of paths in $\mathcal{P}_{\lceil k_\epsilon+t\rceil,\epsilon}$.)

3.1. Light intersecting paths

In order to bound Δ_t, we first decompose the sum in (3.11). For a given real number t and integers i,j with $1 \leqslant j \leqslant i < \lceil k_\epsilon + t\rceil$, let $\Delta_{t,i,j}$ be the sum of $q_\epsilon(P,Q)$ over paths P,Q with $\lceil k_\epsilon + t\rceil$ edges, such that P and Q share i edges, and these i edges form precisely j connected components. Then, we have

$$\Delta_t = \sum_{1\leqslant j\leqslant i<\lceil k_\epsilon+t\rceil} \Delta_{t,i,j}. \tag{3.12}$$

We now need to consider (a) the probability that two paths of k edges (for k near to k_ϵ) sharing i edges both have weight at most $(1-\epsilon)\log n$, and (b) the number of such pairs of paths. The following lemma bounds the former probability. Counting the number of terms of a given sum $\Delta_{t,i,j}$ is the subject of Section 3.2.

Lemma 3.4. *Given two paths P,Q in K_n, each consisting of k edges, i of which are common to P and Q, and any $s \geqslant 0$, we have*

$$\mathbf{P}\{w(P) \leqslant sn, w(Q) \leqslant sn\} \leqslant 4^{k-i} \cdot \frac{s^{2k-i}}{(2k-i)!}.$$

Proof. To make the formulas clearer, introduce $w'(P) = w(P)/n$ and $w'(Q) = w(Q)/n$. This corresponds to the case of exponential edge weights with mean 1 instead of the exponential mean n for the individual edge weights. Thus, we need to evaluate $\mathbf{P}\{w'(P) \leqslant s, w'(Q) \leqslant s\}$. Observe that the sum of ℓ exponential random variables is has a Gamma$(1,\ell)$ distribution, with

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

density function $f_\ell(t) = t^{\ell-1}e^{-t}/(\ell-1)!$ and distribution function $F_\ell(t)$. So, conditioning on the aggregated weight of the i shared edges, we see that

$$\mathbf{P}\{w'(P) \leqslant s, w'(Q) \leqslant s\} = \int_0^s f_i(t)F_{k-i}(t)^2 dt.$$

However, for $t \geqslant 0$, $f_\ell(t) \leqslant t^{\ell-1}/(\ell-1)!$ and $F_\ell(t) \leqslant t^\ell/\ell!$, which implies that

$$\mathbf{P}\{w'(P) \leqslant s, w'(Q) \leqslant s\} \leqslant \int_0^s \frac{t^{i-1}}{(i-1)!}\frac{(s-t)^{2(k-i)}}{(k-i)!^2}dt = \frac{s^{2k-i}}{(2k-i)!}\binom{2(k-i)}{k-i}.$$

Using the classical bound for the central binomial coefficients $\binom{2n}{n} \leqslant 4^n$ completes the proof. □

3.2. The number of intersecting pairs

What is actually needed is to count the number of *pairs* (P, Q), where P and Q are two paths containing k edges such that $P \cap Q$ has i edges in j connected components. More precisely, $P \cap Q$ is a graph composed of j disjoint paths, with i edges in total. Our aim is to analyse shortest paths in the graph, and hence it suffices to consider *self-avoiding* paths (that do not intersect themselves even at vertices). Let $N_{k,i,j}$ denote the number of such pairs.

Lemma 3.5. *The number of pairs of self-avoiding paths (P, Q) of length k, such that $P \cap Q$ contains i edges in j connected components, satisfies*

$$N_{k,i,j} \leqslant n^{2k+2-i-j}(2k^3)^j.$$

Proof. Observe first that if $k < i + 2j - 2$ then $N_{k,i,j} = 0$, since for a given path P, if Q shares i edges in j connected components with P then Q has at least $i + 2j - 2$ edges. We now assume that $k \geqslant i + 2j - 2$.

We first focus on the choice of P with the edges of $P \cap Q$ distinguished. First fix a path P, self-avoiding, as an ordered sequence of $k+1$ vertices. There are $\binom{n}{k+1} \cdot (k+1)!$ such choices. (In fact, we are double-counting here, as the reversed sequence would yield the same path P; we can afford to ignore this fact in obtaining our upper bound.) We next choose the sequence of sizes of the j parts of $P \cap Q$ as they appear along P. Since $P \cap Q$ contains i edges, it suffices to split i into j ordered parts; there are $\binom{i+j}{j}$ possibilities for this partition. Now that we have the sequence of sizes of the parts of $P \cap Q$, it remains to choose the positions in P where these edges appear. We can count the number of such choices as follows: once we have removed $P \cap Q$, there remains a sequence of $k - i$ edges of P. The portions of $P \cap Q$ can be inserted at any of the $k - i + 1$ separating positions (the extremities are included), and hence there are $\binom{k-i+1}{j}$ choices. So in the end, fixing the path P together with the edges of $P \cap Q$ can be done in

$$\binom{n}{k+1}(k+1)! \cdot \binom{i+j}{j} \cdot \binom{k-i+1}{j} \tag{3.13}$$

distinct ways.

It now remains to choose the second path Q, also self-avoiding, so that it intersects P at the distinguished edges. We first choose the order in which the j parts of $P \cap Q$ appear in Q: there are $j!$ possible choices. The intersection $P \cap Q$ contains $i + j$ vertices and we need $k + 1 - i + j$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

other vertices to complete Q. It is possible that some vertices of $Q \setminus (P \cap Q)$ are in P, so we can choose an ordered sequence of these vertices in $\binom{n-i-j}{k+1-i-j}(k+1-i-j)!$ different ways. (In fact, not all such sequences yield valid choices for $Q \setminus P$. For example, $Q \setminus P$ should not contain two consecutive vertices from P in the same part, or $Q \cap P$ will not be what we claimed. However, we only seek an upper bound, and so can afford to ignore this issue.) Finally, we need to choose how the j (now ordered) parts of $P \cap Q$ interlace with these $(k+1-i-j)$ extra vertices. The extra vertices define $k-i-j+1$ intervals (with the extremities): there are $\binom{k-i-j+1}{j}$ ways to choose j of them (and insert the parts of $P \cap Q$ at these spots). It follows by this argument that, for a fixed path P with distinguished edges forming j parts, there are *at most*

$$j! \cdot \binom{n-i-j}{k+1-i-j}(k+1-i-j)! \cdot \binom{k-i-j+1}{j} \tag{3.14}$$

possible choices for the path Q.

Now, the desired number $N_{k,i,j}$ of pairs (P,Q) of paths of length k such that $P \cap Q$ contains i edges in j connected parts is *at most* the product of the numbers appearing in (3.13) and (3.14). The desired bound follows by routine bounding using the inequality $\binom{n}{k} \leqslant n^k/k!$. □

3.3. Paths intersecting at least twice

Lemmas 3.4 and 3.5 yield more than sufficient control over all $\Delta_{t,i,j}$ with $j \geqslant 2$, which we quantify in the following lemma.

Lemma 3.6. *There is an $\epsilon_0 > 0$ such that for all $0 < \epsilon \leqslant \epsilon_0$, if $k = \alpha_\epsilon \log n + o(\sqrt{\log n})$, then, for all n sufficiently large,*

$$\frac{\sum_{2 \leqslant j \leqslant i < k} \Delta_{t,i,j}}{\mathbf{E}\big[|\mathcal{P}_{k,\epsilon}|\big]^2} \leqslant n^{-0.95}.$$

Proof. Combining the bound on the number of intersecting paths (Lemma 3.5) with Lemma 3.4 with $s = (1-\epsilon)\log n/n$, we obtain

$$\begin{aligned}
\Delta_{t,i,j} &:= N_{k,i,j} \cdot \mathbf{P}\{w(P) \leqslant sn, w(Q) \leqslant sn\} \\
&\leqslant n^{2-j}(2k^3)^j((1-\epsilon)\log n)^{2k-i}\frac{4^{k-i}}{(2k-i)!} \\
&\leqslant n^{2-j}(2k^3)^j\left(\frac{(1-\epsilon)e\log n}{2k-i}\right)^{2k-i}4^{k-i},
\end{aligned}$$

since, for $\ell \geqslant 1$, we have $\ell! \geqslant (\ell/e)^\ell$. We have assumed that $j \geqslant 2$, so for n large enough, $(2k^3)^j n^{2-j} \leqslant 4k^6$. It follows that

$$\sum_{2 \leqslant j \leqslant i < k} \Delta_{t,i,j} \leqslant 4k^8 \max_{2 \leqslant i < k}\left\{4^{k-i}\left(\frac{e(1-\epsilon)\log n}{(2k-i)}\right)^{2k-i}\right\}.$$

Using the estimate (3.6) for $\mathbf{E}|\mathcal{P}_{k,\epsilon}|$, it follows that, for n large enough,

$$\frac{\sum_{2 \leqslant j \leqslant i < k} \Delta_{t,i,j}}{\mathbf{E}\big[|\mathcal{P}_{k,\epsilon}|\big]^2} \leqslant \frac{7\pi k^9}{n^2} \max_{2 \leqslant i < k}\left\{\left(\frac{4e(1-\epsilon)\log n}{2k-i}\right)^{-i}\left(\frac{2k}{2k-i}\right)^{2k}\right\}. \tag{3.15}$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

We bound the right-hand side of (3.15) by allowing i to take real values. We write $k = \gamma \log n$ and $x = \beta \log n$ (so $0 \leqslant \beta \leqslant \gamma$); we then see that

$$\left(\frac{4e(1-\epsilon)\log n}{2k-x}\right)^{-x}\left(\frac{2k}{2k-x}\right)^{2k} = e^{g(\beta)\log n},$$

where

$$g(\beta) := \beta \log\left(\frac{2\gamma-\beta}{4e(1-\epsilon)}\right) + 2\gamma \log\left(\frac{2\gamma}{2\gamma-\beta}\right).$$

For γ close to $\alpha^\star$, the function g is suitably approximated by $g^\star$ defined by

$$g^\star(\beta) := \beta \log\left(\frac{2\alpha^\star-\beta}{4e}\right) + 2\alpha^\star \log\left(\frac{2\alpha^\star}{2\alpha^\star-\beta}\right).$$

By differentiation, we see that $g^\star(\beta)$ is maximized on $[0,\alpha^\star]$ by taking $\beta = 2\alpha^\star - 4$, at which point $g^\star(\beta) \approx 1.02$. For any δ, we may ensure that $|\gamma - \alpha^\star| \leqslant \delta$ for n large by choosing ϵ sufficiently small. Since $g(\beta) = g_\gamma(\beta)$ is bounded and continuous in both γ and β away from $\beta \in [2\gamma, \infty)$, for all ϵ sufficiently small and all n sufficiently large we have

$$|g(\beta) - g^\star(\beta)| \leqslant 0.01,$$

for all $k = \alpha_\epsilon \log n + o(\sqrt{\log n})$ and all $\beta \in [0,\gamma]$. It follows that $g(\beta) \leqslant 1.04$ for all $\beta \in [0,\gamma]$, which, combined with (3.15), yields that

$$\frac{\sum_{2\leqslant j\leqslant i<k} \Delta_{t,i,j}}{\mathbf{E}\left[|\mathcal{P}_{k,\epsilon}|\right]^2} \leqslant \frac{7\pi k^9}{n^{0.96}} = O\left(\frac{\log^9 n}{n^{0.96}}\right).$$

This proves the lemma. □

4. Dealing with paths intersecting only once

Unfortunately, for paths intersecting only once (*i.e.*, when $j = 1$), for some values of i the quantity $\Delta_{t,i,1}$ is too large for us to apply the straightforward approach used above. (Note that this is not an artefact of our upper bounds: Lemmas 3.4 and 3.5 are essentially tight up to logarithmic factors. Also, two shortest paths between two pairs of vertices typically have one connected component in common, so this class of paths is the main problem.) The most natural and naive way to deal with this complication is simply to throw away all pairs of shortest paths P and Q whose intersection is connected, and try to bound the probability that one of the remaining minimum-weight paths is 'long' (in the same sense as above). This is essentially our approach. However, in order to keep a handle on the conditioning imposed in doing so, it is useful to proceed 'from the other direction': building a set of shortest paths with special properties that guarantee that (a) no pair of such paths has an intersection which is connected, yet (b) the set contains a minimum-weight path with about $\alpha^\star \log n$ edges.

In order to describe this set, we first need to introduce a few concepts. We say that paths P and P' *intersect once* if $P \cap P'$ has only one connected component containing at least one edge. (There may be other components of $P \cap P'$ which are isolated vertices.) We say that a path P is a *local optimum* if, for all paths P' with $|P| = |P'|$ that intersect P once, we have $w(P') > w(P)$. From now on, let $\mathcal{O}_k$ denote the set of paths with k edges that are local optima, let $\mathcal{O}_{k,\epsilon}$ denote

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

the set of elements of $\mathcal{O}_k$ with weight at most $(1-\epsilon)\log n$. Also, let $\mathcal{P}^\star = \mathcal{P}(n) = \{P_{ij} : 1 \leqslant i < j \leqslant n\}$ denote the set of shortest paths in K_n. Note that $\mathcal{O}_{k,\epsilon} \subseteq \mathcal{P}_{k,\epsilon}$.

An instructive example. To motivate our next definitions, consider the unlikely (impossible) but instructive event that we find a path $P = (v_1, \ldots, v_{k+1})$, all of whose edges have weight exactly 1, and consider a path Q of length k intersecting P in exactly $i < k$ consecutive edges, say $Q \cap P = (v_1, \ldots, v_{i+1})$, for example. Let Q_1 (respectively Q_2) be the component of $Q \setminus P$ containing v_1 (respectively v_{i+1}). In order that $w(Q) \leqslant k$, then certainly we must have both $w(Q_1) \leqslant k - i$ and $w(Q_2) \leqslant k - i$. On the other hand, one of Q_1 and Q_2 has at least $(k-i)/2$ edges. It follows that if $w(Q) \leqslant k$, then in K_n^P either $\mathrm{SPT}_{v_1}(k-i)$ or $\mathrm{SPT}_{v_{i+1}}(k-i)$ has height at least $(k-i)/2$.

This observation is key to our approach. Roughly speaking, we wish to consider long minimum-weight paths with the special property that none of the shortest path trees leaving them are 'too tall'. This property will guarantee that the paths are local optima (Lemma 4.5), and we can then use the second moment method to prove concentration for the number of such paths. However, the idea suggested by the above sketch gives a little too much away in terms of what 'too tall' means; in particular, it is not careful enough about the interplay between the contributions of Q_1 and Q_2 to the length of Q. When we formalize our idea, we will have to deal with this interplay to make the details work out.

4.1. The fluctuations of conditioned partial sums

In the above example, there was a second simplification: the edge weights have the extremely desirable property that, for every subpath P' of P, $w(P')$ is *exactly* $\mathbf{E}\big[w(P') \mid w(P) = k\big]$, which makes the conditions we need to impose on the shortest path trees leaving P very easy to state. It is, of course, too much to ask for paths whose edge weights are as well behaved as in the above example. In general, the weights of subpaths will fluctuate considerably from their conditional expected value given $w(P)$. Our tool for controlling the size of these fluctuations is the *law of the iterated logarithm*; we will use the version found in Rogers and Williams [11]. (These bounds are in fact stronger than we require; however, we did not see a substantially simpler proof of a simpler but sufficient result; furthermore, Corollary 4.3 is perhaps of interest independent of its role in the current work.)

Theorem 4.1 ([11], Corollary 16.5). *Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with finite variance, and let $S_n = \sum_{j=1}^{j} Y_j$. Then*

$$\mathbf{P}\left\{ \limsup_{n\to\infty} \frac{S_n - n\mathbf{E}Y_1}{\sqrt{\mathbf{Var}[Y_1] 2n \log\log n}} = 1 \right\} = 1.$$

We then immediately have the following finite statement, more useful for our purposes.

Corollary 4.2. *Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with finite variance, and let $S_n = \sum_{j=1}^{j} Y_j$. For all $\delta > 0$ there is a constant $C > 0$ depending on δ and the distribution of Y_1 (we write $C(\delta, Y_1)$ for short) such that*

$$\mathbf{P}\left\{ \sup_{n\geqslant 1} \left| \frac{S_n - n\mathbf{E}Y_1}{\sqrt{\mathbf{Var}[Y_1] 2n \log\log n}} \right| \leqslant C \right\} \geqslant 1 - \delta.$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

(To avoid annoying technicalities in our formulae, in Corollary 4.2 and hereafter, when we write $\log\log n$ we mean $\max(\log\log n, 1)$.) At this point, the fact that we are considering exponential random variables comes in very handy, as it allows us to apply Corollary 4.2 conditional upon the value of S_n. More precisely, we have the following result.

Corollary 4.3. *Suppose Y_1 is an exponential random variable. Then, for all $\delta > 0$, there exists $C' = C'(\delta)$ such that, for all n sufficiently large,*

$$\mathbf{P}\left\{\forall k \in [n],\ \left|\frac{S_k - S_n(k/n)}{\sqrt{\mathbf{Var}[Y_1 \mid S_n] 2k\log\log k}}\right| \leqslant C' \;\middle|\; S_n\right\} \geqslant 1-\delta. \tag{4.1}$$

We emphasize that the probability in (4.1) is a *random variable*, measurable with respect to S_n; the content of the lemma is that this random variable is *deterministically* at least $1-\delta$. In particular, if we let $\mathcal{E}$ be the event in (4.1) (so (4.1) bounds $\mathbf{P}\{\mathcal{E} \mid S_n\}$), then $\mathbf{P}\{\mathcal{E}\} = \mathbf{E}_{S_n}\left[\mathbf{P}\{\mathcal{E} \mid S_n\}\right] \geqslant 1-\delta$ and more strongly, for any event $\mathcal{E}'$ which is measurable with respect to S_n,

$$\mathbf{P}\{\mathcal{E} \mid \mathcal{E}'\} = \mathbf{E}_{S_n}\left[\mathbf{P}\{\mathcal{E} \mid S_n\} \mid \mathcal{E}'\right] \geqslant 1-\delta.$$

Proof of Corollary 4.3. Fix $\delta > 0$ and n. For $i = 1, \ldots, n$, let $E_i = S_i/S_n$. Then the set $\{E_1, \ldots, E_n\}$ is independent of S_n and distributed like n independent $[0,1]$-uniform random variables (for a proof, see Chapter 8 of [12]). It follows in particular that E_1, the minimum of n independent uniforms, has a Beta$(1, n-1)$ distribution and $\mathbf{Var}[Y_1 \mid S_n] = (S_n/n)^2(n-1)/(n+1) \geqslant (S_n/n)^2/3$, for $n \geqslant 2$. Using this estimate of the conditional variance, (4.1) may be bounded as

$$\begin{aligned}
&\mathbf{P}\left\{\forall k \in [n],\ \left|\frac{S_k - S_n(k/n)}{\sqrt{\mathbf{Var}[Y_1 \mid S_n] 2k\log\log k}}\right| \leqslant C' \;\middle|\; S_n\right\} \\
&\quad\geqslant \mathbf{P}\left\{\forall k \in [n],\ \left|\frac{n\frac{S_k}{S_n} - k}{\sqrt{2k\log\log k}}\right| \leqslant \sqrt{3}C' \;\middle|\; S_n\right\} \\
&\quad= \mathbf{P}\left\{\forall k \in [n],\ \left|\frac{n\frac{S_k}{S_n} - k}{\sqrt{2k\log\log k}}\right| \leqslant \sqrt{3}C'\right\},
\end{aligned}$$

where we used the independence of the S_k/S_n of S_n. Thus, proving (4.1) reduces to finding C' such that the *unconditioned* statement

$$\mathbf{P}\left\{\sup_{1\leqslant k\leqslant n}\left|\frac{(n/S_n)S_k - k}{\sqrt{2k\log\log k}}\right| \leqslant C'\right\} \geqslant 1-\delta \tag{4.2}$$

holds. By choosing C' large enough we can certainly ensure that this holds for small n; shortly we will take advantage of this fact. By Corollary 4.2 we have that

$$\mathbf{P}\left\{\sup_{1\leqslant k\leqslant n}\left|\frac{S_k - k}{\sqrt{2k\log\log k}}\right| \leqslant C_1\right\} \geqslant 1-\frac{\delta}{2}, \tag{4.3}$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

where $C_1 = C(\delta/2, Y_1)$ as in the statement of Corollary 4.2. If, on the other hand, there exists $k \in \{1, \dots, n\}$ such that

$$\left| \frac{(n/S_n)S_k - k}{\sqrt{2k \log\log k}} \right| \geqslant C_1 + c,$$

for some $c > 0$, then either the event whose probability is bounded in (4.3) must fail to hold, or by the triangle inequality we must have

$$\left| \frac{(n/S_n - 1)S_k}{\sqrt{2k \log\log k}} \right| \geqslant c,$$

for some k, or more simply $|S_n - n| \geqslant c\sqrt{2k \log\log k} \cdot S_n/S_k$. For this to occur, we must either have $|S_n - n| \geqslant \sqrt{cn}$ or $S_k \geqslant \sqrt{c(2k \log\log k)} \cdot (\sqrt{n} - \sqrt{c})$. It follows that, for n large enough that $\sqrt{n} - \sqrt{c} \geqslant \sqrt{n/2}$ (and recalling that $\log\log k \geqslant 1$ for all k by convention), we have

$$\begin{aligned} \mathbf{P}\left\{ \sup_{1 \leqslant k \leqslant n} \left| \frac{(n/S_n - 1)S_k}{\sqrt{2k \log\log k}} \right| \geqslant c \right\} &\leqslant \sum_{k=1}^{n} \mathbf{P}\{S_k \geqslant k\sqrt{c}\} + \mathbf{P}\{|S_n - n| \geqslant \sqrt{cn}\} \\ &\leqslant \sum_{k=1}^{n} e^{-kH(\sqrt{c})} + \frac{1}{c}, \end{aligned} \tag{4.4}$$

where we have bounded the first term using a large deviations bound for sums of exponential random variables [2] with $H(x) := x - 1 - \log x$, and the second term by Chebyshev's inequality. We now choose c large enough that the right-hand side of (4.4) above is at most $\delta/2$ for all n. Combining (4.4) with (4.3) and taking $C' = C_1 + c$ (or slightly larger if necessary, to deal with small n) then establishes (4.2) and completes the proof. □

Motivated by Corollary 4.3, we say that a path $P = (v_1, \dots, v_{k+1})$ in K_n is *C-legal* (for a given $C > 0$), and write $P \in \mathcal{L}_C = \mathcal{L}_C(n)$, if for all $i = 1, \dots, k$,

$$\left| \frac{k}{w(P)} \cdot \sum_{j=1}^{i} X_{v_j v_{j+1}} - i \right| \leqslant C\sqrt{2i \log\log i}, \tag{4.5}$$

and

$$\left| \frac{k}{w(P)} \cdot \sum_{j=k-i+1}^{k} X_{v_j v_{j+1}} - i \right| \leqslant C\sqrt{2i \log\log i}. \tag{4.6}$$

We need both conditions as we wish to ensure the path is well-behaved 'from both ends'. Although these conditions only restrict the weight of subpaths of P starting from an end, they can be combined to control the behaviour of other subpaths. More precisely, let

$$m(i,k) = \min\{i - 1, k + 1 - i\} \quad \text{and} \quad s(i,j,k) = \max\{m(i,k), m(j,k)\}. \tag{4.7}$$

Intuitively, $s(i,j,k)$ is the furthest distance from either i or j to one of the ends of the path. We then have the following.

Lemma 4.4. *If $P = (v_1, \dots, v_{k+1})$ is C-legal, then, for all $1 \leqslant i < j \leqslant k + 1$,*

$$\left| \frac{k}{w(P)} \cdot \sum_{m=i}^{j-1} X_{v_m v_{m+1}} - (j - i) \right| \leqslant 2C \cdot \sqrt{2s(i,j,k) \log\log(s(i,j,k))}. \tag{4.8}$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

Proof. Write $P(i)$ (respectively $P(i,j)$) for the path $(v_1,\dots,v_i)$ (respectively $(v_i,\dots,v_j)$), then apply (4.5) and (4.6) to $P(i)$ or $P \setminus P(i)$ according to whether $i \leqslant (k+1)/2$ or $i > (k+1)/2$, and to $P(j)$ or $P \setminus P(j)$, similarly. Finally, use one of the equalities

$$P(i,j) = P(j) \setminus P(i-1) = (P \setminus P(i-1)) \setminus (P \setminus P(j)) = P \setminus (P(i-1) \cup (P \setminus P(j))),$$

together with the triangle inequality, to prove the result. □

4.2. Bounding the heights of shortest path trees

We have now introduced all the technical apparatus we will need to control the path P itself. Returning to our example from page 11, we recall that our main goal is to bound the *heights of the shortest path trees leaving* P; Lemma 4.4 tells us just how tall we can allow them to be. We say that P is *C-bonsai* if:

($\star$) for all $i = 1,\dots,k+1$, and all integers ℓ with $\ell \geqslant m(i,k)/40$, in K_n^P,

$$\mathrm{SPT}_{v_i}(\ell + 2C\sqrt{500\ell \log\log \ell}) \text{ has height less than } \left(\frac{9k}{10w(P)}\right) \cdot \ell.$$

We denote the set of C-bonsai paths in K_n by $\mathcal{B}_C = \mathcal{B}_C(n)$.

Lemma 4.5. *For any $C \geqslant 0$, if $w(P) \leqslant k \leqslant 4w(P)$ and P is both C-legal and C-bonsai, i.e., $P \in \mathcal{L}_C \cap \mathcal{B}_C$, then P is a local optimum.*

Proof of Lemma 4.5. Let $P = (v_1,\dots,v_{k+1}) \in \mathcal{L}_C \cap \mathcal{B}_C$. Our aim is to show that, for any integers i,j with $1 \leqslant i < j \leqslant k+1$ and any path Q of length k whose intersection with P is $(v_i,\dots,v_j)$, we have $w(Q) > w(P)$. By symmetry we may assume that $i < (k+1)/2$ (so $m(i,k) = i-1$) and that $m(i,k) \leqslant m(j,k)$ (so $s(i,j,k) = m(j,k)$).

Let Q_i (respectively Q_j) be the component of $Q \setminus P$ containing v_i (respectively v_j). We must have $|Q_i| + |Q_j| = k - (j-i) \geqslant m(i,k) + m(j,k)$. Let $k_i = (10/9)(w(P)/k)|Q_i|$ and define k_j similarly. If $|Q_i| \geqslant km(i,k)/(40w(P))$, then, since P is C-bonsai and $w(P)/k \geqslant 1/4$,

$$w(Q_i) \geqslant k_i + 2C\sqrt{500k_i \log\log k_i} \geqslant k_i + 2C\sqrt{m(i,k)\log\log m(i,k)}.$$

(The constant 500 ensures that the second inequality holds.) Similarly, if

$$|Q_j| \geqslant (m(j,k)/40)k/w(P),$$

then

$$w(Q_j) \geqslant k_j + 2C\sqrt{500k_j \log\log k_j} \geqslant k_j + 2C\sqrt{m(j,k)\log\log m(j,k)}.$$

If both these conditions occur, then by the definitions of k_i and k_j we have

$$w(Q_i) + w(Q_j) \geqslant \frac{w(P)}{k}(|Q_i| + |Q_j|) + 2C\sqrt{s(i,j,k)\log\log s(i,j,k)}.$$

Since $w(P) \leqslant k$, combining the preceding equation with the lower bound on $w(P \cap Q)$ given by Lemma 4.4 completes the proof in this case.

On the other hand, if $|Q_j| \leqslant km(j,k)/(40w(P)) \leqslant m(j,k)/10$, then $|Q_i| \geqslant m(i,k)$, and

$$|Q_i| \geqslant (k-(j-i)) - \frac{m(j,k)}{10} \geqslant \frac{9(k-(j-i))}{10} \geqslant \frac{9m(j,k)}{10}.$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

Since P is C-bonsai we have $w(Q_i) \geqslant k_i + 2C\sqrt{500k_i \log\log k_i}$. Moreover, in this case, $k_i \geqslant (w(P)/k)(k-(j-i))$ and $k_i \geqslant m(j,k)/4 = s(i,j,k)/4$, the preceding equation implies that

$$w(Q_i) \geqslant \frac{w(P)}{k}(k-(j-i)) + 2C\sqrt{s(i,j,k)\log\log s(i,j,k)},$$

which, combined with Lemma 4.4 and the fact that $w(P) \leqslant k$, completes the proof in this case. The case $|Q_i| \leqslant (m(i,k)/40)(k/w(P))$ is similar to but easier than the case $|Q_j| \leqslant (m(i,k)/40)(k/w(P))$. □

We are now in a position to fully define the special set of paths we wish to consider. Given $\epsilon > 0$ and a positive integer k, let $\mathcal{U}_{\epsilon,k,C} = \mathcal{U}_{\epsilon,k,C}(n)$ be the set of paths P of length k in K_n for which the following events hold:

$$\begin{aligned} A_1 &:= \{(1-2\epsilon)\log n \leqslant w(P) \leqslant (1-\epsilon)\log n\}, \\ A_2 &:= \{P \in \mathcal{P}^\star\}, \\ A_3 &:= \{P \in \mathcal{L}_C \cap \mathcal{B}_C\}. \end{aligned} \tag{4.9}$$

We remark that if $\epsilon > 0$ is small enough (in particular, any $\epsilon = o(1)$ suffices), $C > 0$ is fixed, $t = t(n) = o(\sqrt{\log n})$, and $P \in \mathcal{U}_{\epsilon,\lceil k_\epsilon + t\rceil,C}$ (where k_ϵ was defined in (3.8)) then for n large enough, $w(P) \leqslant k \leqslant 4w(P)$, so, by Lemma 4.5, P is a local optimum. The following lemma bounds the expected size of $\mathcal{U}_{\epsilon,k,C}$.

Lemma 4.6. *There exist $C > 0$ and $\delta > 0$ such that, for any fixed $c > 2$ and any ϵ with $\delta > \epsilon > c\log\log n/\log n$, for any function $t(n) = o(\sqrt{\log n})$, and for all n sufficiently large,*

$$\mathbf{E}|\mathcal{U}_{\epsilon,\lceil k_\epsilon+t\rceil,C}| \geqslant \delta \mathbf{E}|\mathcal{P}_{\lceil k_\epsilon+t\rceil,\epsilon}|.$$

Proof. Recall that $h(T)$ denotes the (unweighted) height of a rooted tree T. Fix $\epsilon > 0$ as above, and a path P of length $k = \lceil k_\epsilon + t\rceil$. It suffices to prove that for any such path P, $\mathbf{P}\{P \in \mathcal{U}_{\epsilon,\lceil k_\epsilon+t\rceil,C}\} \geqslant \delta \mathbf{P}\{P \in \mathcal{P}_{\lceil k_\epsilon+t\rceil,\epsilon}\}$ for a suitable δ. Define the following conditional probability:

$$\mathbf{P}_\epsilon\{\,\cdot\,\} := \mathbf{P}\{\,\cdot \mid (1-2\epsilon)\log n \leqslant w(P) \leqslant (1-\epsilon)\log n\}.$$

Then, by the definition (4.9), we have

$$\begin{aligned} \mathbf{P}\{P \in \mathcal{U}_{\epsilon,\lceil k_\epsilon+t\rceil,C}\} &= \mathbf{P}\{A_1, A_2, A_3\} \\ &= \mathbf{P}_\epsilon\{A_2, A_3\} \cdot \mathbf{P}\{(1-2\epsilon)\log n \leqslant w(P) \leqslant (1-\epsilon)\log n\} \\ &= \mathbf{P}_\epsilon\{A_2, A_3\} \cdot (1+o(1))\mathbf{P}\{w(P) \leqslant (1-\epsilon)\log n\} \\ &\qquad \text{(by (3.1))} \\ &= \mathbf{P}_\epsilon\{A_2, A_3\} \cdot (1+o(1))\mathbf{P}\{P \in \mathcal{P}_{\lceil k_\epsilon+t\rceil,\epsilon}\}. \end{aligned}$$

So, to prove the lemma it suffices to show that $\mathbf{P}_\epsilon\{A_2, A_3\}$ is bounded away from zero. To do so we first write

$$\begin{aligned} \mathbf{P}_\epsilon\{A_2, A_3\} &= \mathbf{P}_\epsilon\{P \in \mathcal{P}^\star, P \in \mathcal{L}_C, P \in \mathcal{B}_C\} \\ &\geqslant \mathbf{P}_\epsilon\{P \in \mathcal{B}_C, P \in \mathcal{L}_C\} - \mathbf{P}_\epsilon\{P \notin \mathcal{P}^\star\}. \end{aligned}$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

For ϵ as above, $\mathbf{P}_\epsilon\{P \notin \mathcal{P}^\star\} \to 0$ as $n \to \infty$ by Corollary 3.2, so we need only prove that $\mathbf{P}_\epsilon\{P \in \mathcal{B}_C, P \in \mathcal{L}_C\}$ is bounded away from zero. To do so, rather than considering the set $\mathcal{B}_C$ we consider a slightly different collection, which we denote $\mathcal{B}_C^\epsilon$. We let $P \in \mathcal{B}_C^\epsilon$ if ($\star$) on page 14 holds with the term $w(P)$ replaced by $(1-\epsilon)\log n$. Given that $w(P) \leqslant (1-\epsilon)\log n$, we have $\mathcal{B}_C^\epsilon \subseteq \mathcal{B}_C$, so $\mathbf{P}_\epsilon\{P \in \mathcal{B}_C, P \in \mathcal{L}_C\} \geqslant \mathbf{P}_\epsilon\{P \in \mathcal{B}_C^\epsilon, P \in \mathcal{L}_C\}$. But the event $P \in \mathcal{B}_C^\epsilon$ is independent from the events $P \in \mathcal{L}_C$ and

$$A_1 = \{(1-2\epsilon)\log n \leqslant w(P) \leqslant (1-\epsilon)\log n\},$$

since they are determined by disjoint sets of edges. Thus, we have

$$\mathbf{P}_\epsilon\{P \in \mathcal{B}_C, P \in \mathcal{L}_C\} \geqslant \mathbf{P}\{P \in \mathcal{B}_C^\epsilon\}\mathbf{P}_\epsilon\{P \in \mathcal{L}_C\}.$$

Now choose $C = C'(1/2)$, where $C'(1/2)$ is as in Corollary 4.3. By that corollary we then have $\mathbf{P}_\epsilon\{P \in \mathcal{L}_C\} \geqslant 1/2$, so

$$\mathbf{P}_\epsilon\{P \in \mathcal{B}_C, P \in \mathcal{L}_C\} \geqslant \frac{1}{2}\mathbf{P}\{P \in \mathcal{B}_C^\epsilon\}.$$

To complete the proof, it thus suffices to prove that $\mathbf{P}\{P \in \mathcal{B}_C^\epsilon\}$ is bounded below by a constant.

We assume ϵ is small enough and n is large enough (say $n \geqslant n_0$ for some n_0) that

$$\frac{(1-\epsilon)\log n}{\lceil k_\epsilon + t\rceil} \leqslant \frac{1}{3.5}. \tag{4.10}$$

(Note that this is possible by the value of $\alpha^\star$. We also remark that the preceding equation is the only place we use the hypothesis that ϵ is 'small'.) To further simplify our lives, we now consider a subset of $\mathcal{B}_C^\epsilon$ that we denote $\bigcap_{i=1}^{k+1}\mathcal{B}_{C,i}$, where $\mathcal{B}_{C,i}$ is the set of all paths $Q = (v_1, \dots, v_{k+1})$ in K_n for which, for all integers ℓ with $\ell \geqslant m(i,k)/40$ (recall that $m(i,k)$ is defined in (4.7)), in K_n^Q,

$$h\big(\mathrm{SPT}_{v_i}\big(\ell + 2C\sqrt{500\ell\log\log\ell}\big)\big) < 3.15\ell. \tag{4.11}$$

The set $\bigcap_{i=1}^{k+1}\mathcal{B}_{C,i}$ is indeed a subset of $\mathcal{B}_C^\epsilon$, since $9k/(10(1-\epsilon)\log n) \geqslant 3.15$ by (4.10). The events $P \in \mathcal{B}_{C,i}$, $i = 1, \dots, k+1$ are all increasing in the weight values, so by the FKG inequality [6, 1, 5, 4], we have

$$\mathbf{P}\{P \in \mathcal{B}_C^\epsilon\} \geqslant \prod_{i=1}^{k+1}\mathbf{P}\{P \in \mathcal{B}_{C,i}\} \geqslant \prod_{i=1}^{\lfloor (k+1)/2\rfloor}\mathbf{P}\{P \in \mathcal{B}_{C,i}\}^2, \tag{4.12}$$

where the second inequality takes advantage of the symmetry. For small ℓ, the probability that (4.11) occurs is bounded away from zero: it is at least the probability that all $n-1$ edges leaving v_i have weight at least $\ell + 2C\sqrt{500\ell\log\log\ell}$, and every edge weight has an exponential distribution with mean n. Thus, choose an integer ℓ_0 large enough that the following four conditions are satisfied for $\ell \geqslant \ell_0$:

$$\begin{cases} \ell + 2C\sqrt{500\ell\log\log\ell} \leqslant 1.05\ell, \\ 1 - e^{-\ell/10}/2 \geqslant \exp(-e^{-\ell/10}), \end{cases} \quad \text{and} \quad \begin{cases} e^{3-\ell/8} & \leqslant e^{-\ell/10}/4, \\ 3e^{\ell/40}\exp(-e^{\ell/20}) & \leqslant e^{-\ell/10}/4. \end{cases}$$

Then there exists $\alpha_0 > 0$ such that, for $\ell \leqslant \ell_0$, the event in (4.11) occurs with probability at least α_0. Furthermore, for each ℓ the event in (4.11) is increasing in the weight values, so the FKG

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

inequality yields

$$\mathbf{P}\{P \in \mathcal{B}_{C,i}\} \geqslant \prod_{\ell=\lceil m(i,k)/40\rceil}^{\infty} \mathbf{P}\{h\big(\mathrm{SPT}_i\big(\ell + 2C\sqrt{500\ell \log\log \ell}\big)\big) < 3.15\ell\}$$
$$\geqslant \alpha_0^{\xi_i} \cdot \prod_{\ell \geqslant \ell_0 \vee m(i,k)/40} \mathbf{P}\{h\big(\mathrm{SPT}_i(1.05\ell)\big) < 3.15\ell\}, \tag{4.13}$$

where we have written $\xi_i = \max\{\ell_0 - m(i,k)/40, 0\}$. We now turn to bounding these remaining probabilities for $\ell \geqslant \ell_0$. By Lemma 2.1, and our choice for ℓ_0,

$$\mathbf{P}\{|\mathrm{SPT}_i(1.05\ell)| \geqslant e^{1.1\ell}\} \leqslant 3e^{\ell/40} \exp\big(-e^{\ell/40}\big) \leqslant \frac{e^{-\ell/10}}{4}.$$

Furthermore, with the following values and our choice for ℓ_0, the bound of Theorem 2.2 yields

$$\mathbf{P}\{h\big(\mathrm{SPT}_i(1.05\ell)\big) \geqslant 3.15\ell \mid |\mathrm{SPT}_i(1.05\ell)| < e^{1.1\ell}\} \leqslant \frac{e^{-\ell/10}}{4}.$$

It follows from the two preceding equations that, for $\ell \geqslant \ell_0$,

$$\mathbf{P}\{h\big(\mathrm{SPT}_i(1.05\ell)\big) < 3.15\ell\} \geqslant \left(1 - \frac{e^{-\ell/10}}{4}\right)^2 \geqslant 1 - \frac{e^{-\ell/10}}{2} \geqslant e^{-e^{-\ell/10}}.$$

Combining this inequality with (4.13) yields

$$\mathbf{P}\{P \in \mathcal{B}_{C,i}\} \geqslant \alpha_0^{\xi_i} \exp\left(-\sum_{\ell=\lceil m(i,k)/40\rceil}^{\infty} e^{-\ell/10}\right) \geqslant \alpha_0^{\xi_i} \exp\big(-26e^{-m(i,k)/400}\big).$$

Observe that the multiplicative factor $\alpha_0^{\xi_i} = 1$, but for the values of i such that $m(i,k) \leqslant 40\ell_0$. Observe also that for $i \leqslant (k+1)/2$, we have by definition (4.7) $m(i,k) = i - 1$. So, together with (4.12) the preceding equation yields

$$\mathbf{P}\{P \in \mathcal{B}_C^{\epsilon}\} \geqslant \alpha_0^{80\ell_0^2} \exp\left(-52 \sum_{i=1}^{\lfloor (k+1)/2\rfloor} e^{-(i-1)/400}\right),$$

which completes the proof, since the series on the right-hand side converges. □

With Lemma 4.6 under our belt, we can now complete our proof of Theorem 1.1

Proof of Theorem 1.1, lower bounds. Let C be as in Lemma 4.6, choose $\epsilon = \epsilon(n) = 3\log\log n/\log n$, and choose a function $t = t(n) = o(\sqrt{\log n})$ such that $t(n) \to -\infty$ with n; we will make a more precise choice for $t(n)$ shortly. As noted just before the statement of Lemma 4.6, if $\epsilon = o(1)$, $C > 0$ is fixed, $t = o(\sqrt{\log n})$, and $P \in \mathcal{U}_{\epsilon,\lceil k_\epsilon + t\rceil,C}$ then for n large enough, $w(P) \leqslant k \leqslant 4w(P)$, so, by Lemma 4.5, P is a local optimum. Our choices of ϵ, C, and t satisfy these conditions, so for n large enough we have

$$\mathcal{U}_{\lceil k_\epsilon + t\rceil,\epsilon,C} \subseteq \mathcal{O}_{\lceil k_\epsilon + t\rceil,\epsilon} \subseteq \mathcal{P}_{\lceil k_\epsilon + t\rceil,\epsilon}.$$

By Lemmas 3.3 and 4.6, and the definition of k_ϵ in (3.8), there exists $K > 0$ such that, choosing $t(n) = K\log\log n$, for n large enough we have

$$\mathbf{E}|\mathcal{U}_{\lceil k_\epsilon + t\rceil,\epsilon}| \geqslant \log n.$$

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

Recalling the definition (3.9) of $\Delta(\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon})$, we remark that since $\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon,C} \subseteq \mathcal{O}_{\lceil k_\epsilon+t\rceil,\epsilon}$, the contribution of pairs of paths which intersect once to $\Delta(\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon})$, is zero. Also, since $\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon,C} \subseteq \mathcal{P}_{\lceil k_\epsilon+t\rceil,\epsilon}$, we have

$$\Delta(\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon}) \leqslant \sum_{2\leqslant j\leqslant i<\lceil k_\epsilon+t\rceil} \Delta_{t,i,j},$$

so by Lemma 3.6, for n large enough, we have $\Delta(\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon}) \leqslant n^{-0.95}\mathbf{E}\big[|\mathcal{P}_{\lceil k_\epsilon+t\rceil,\epsilon}|\big]^2$. By our choice of ϵ there exists $K' > 0$ such that $k_\epsilon + t \geqslant \alpha^\star \log n - K' \log\log n$. It follows by the second moment method (3.10) and Lemma 4.6 that

$$\begin{aligned}\mathbf{P}\Big\{\max_{1\leqslant i,j\leqslant n} |P_{ij}| \leqslant \alpha^\star \log n - K' \log\log n\Big\} &\leqslant \mathbf{P}\{|\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon}| = 0\} \\ &< \frac{1}{\mathbf{E}|\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon}|} + \frac{\Delta(\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon})}{\mathbf{E}\big[|\mathcal{U}_{\lceil k_\epsilon+t\rceil,\epsilon}|\big]^2} \\ &\leqslant \frac{1}{\log n} + \frac{n^{-0.95}}{\delta^2} \xrightarrow[n\to\infty]{} 0,\end{aligned}$$

proving the lower bound in probability. We then have from the previous bound

$$\mathbf{E}\Big[\max_{1\leqslant i,j\leqslant n} |P_{ij}|\Big] \geqslant (\alpha^\star \log n - K' \log\log n)\left(1 - \frac{1}{\log n} - \frac{n^{-0.95}}{\delta^2}\right),$$

which proves the lower bound in expectation. □

5. Conclusion

We have shown that the longest minimum-weight path in a complete graph with i.i.d. exponentially distributed edge weights is asymptotically $\alpha^\star \log n$ in probability, where $\alpha^\star \approx 3.5911$ satisfies $\alpha^\star \log \alpha^\star = \alpha^\star + 1$. The difficult part of the proof is the lower bound, as the upper bound follows easily from Devroye's [3] results on the height of a random recursive tree and a simple union bound. The fact that such a simple union bound leads to correct bounds means that the n shortest path trees SPT_i essentially behave as if they were independent.

The lower bound is proved by a second moment argument for a carefully chosen random set of minimum-weight paths, all of length about $\alpha^\star \log n$, that satisfy certain regularity conditions. This careful choice allows us to control the variance of the size of the set of such paths. The proof shows that, with high probability, there exist minimum-weight paths of length about $\alpha^\star \log n$ whose total weight is close to $\log n$. Thus, these extra-long paths have a total weight just like a typical minimum-weight path.

Finally, just as Janson [8] did in proving his results discussed in the Introduction, we can use a standard coupling argument to extend Theorem 1.1 to distributions other than exponential. To explain the coupling, it is useful to divide the edge weights by n, so we consider the Exp(1) edge weights $X'_e = X_e/n$.

Let Z be any non-negative random variable, with distribution function $F(t) = \mathbf{P}\{Z \leqslant t\}$, and let $F^{-1}(t) = \sup\{x : F(x) \leqslant t\}$. Let U be a random variable uniform on $[0,1]$. Then $F^{-1}(U)$ is distributed as Z. We may thus couple the exponential edge weights X'_e to uniform edge weights U_e by setting $X'_e = F_E^{-1}(U_e)$, where $F_E(t)$ is the distribution function of an Exp(1) random

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press

variable. We have $F_E(t) = 1 - e^{-t} = t - t^2/2 + O(t^3)$ as $t \to 0$, and it follows that, for all edges e of K_n, $|X'_e - U_e| = O((X'_e)^2)$, uniformly when $X'_e \leqslant 1/2$.

Of particular use for us is the following consequence: for all edges e with $X'_e \leqslant 12 \log n/n$ (say), we have $|X'_e - U_e| = O(\log^2 n/n^2)$, so for any *path* P with $O(\log n)$ edges and with $w'(P) := \sum_{e \in E(P)} X'_e \leqslant 12 \log n/n$, we have

$$\left| w'(P) - \sum_{e \in E(P)} U_e \right| = O\left(\frac{\log^3}{n^2} \right). \tag{5.1}$$

It then follows easily that the conclusions of Theorem 1.1 hold for uniform $[0, 1]$, as well as exponential, edge weights. (Technically, one must redo all the arguments of the paper keeping (5.1) in mind, and accordingly adjust some of the constants very slightly; we omit this step.)

In fact all we really need is a bound of $o(1/n)$ in (5.1) in order to extend Theorem 1.1 to uniform $[0, 1]$ edge weights. Thus, by another coupling as above, we may extend Theorem 1.1 from uniform edge weights to any other edge weights Y whose distribution function F_Y satisfies $F_Y(t) = ct + o(t/\log^2 t)$ for some constant $c > 0$, as $t \to 0$.

References

[1] Alon, N. and Spencer, J. (2000) *The Probabilistic Method*, 2nd edn, Wiley, New York.

[2] Dembo, A. and Zeitouni, O. (1998) *Large Deviation Techniques and Applications*, 2nd edn, Springer.

[3] Devroye, L. (1987) Branching processes in the analysis of the heights of trees. *Acta Informatica* **24** 277–298.

[4] Fortuin, C. M., Kasteleyn, P. W. and Ginibre, J. (1971) Correlational inequalities for partially ordered sets. *Comm. Math. Phys.* **22** 89–103.

[5] Grimmett, G. R. (1999) *Percolation*, Vol. 321 of *A Series of Comprehensive Studies in Mathematics*, 2nd edn, Springer.

[6] Harris, T. E. (1960) A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* **56** 13–20.

[7] Hooghiemstra, G. and Van Mieghem, P. (2008) The weight and hopcount of the shortest path in the complete graph with exponential weights. *Combin. Probab. Comput.* **17** 537–548.

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

[9] Pittel, B. (1994) Note on the height of recursive trees and m-ary search trees. *Random Struct. Alg.* **5** 337–347.

[10] Robbins, H. (1955) A remark on Stirling's formula. *Amer. Math. Monthly* **62** 26–29.

[11] Rogers, L. C. G. and Williams, D. (2000) *Diffusions, Markov Processes, and Martingales*, Vol. 1, Cambridge University Press.

[12] Shorack, G. R. and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics*, Wiley.

[13] Smythe, R. T. and Mahmoud, H. M. (1995) A survey of recursive trees. *Theoret. Probab. Math. Statist.* **51** 1–27.

[14] van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2006) Size and weight of shortest path trees with exponential link weights. *Combin. Probab. Comput.* **15** 903–926.

[15] van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2007) The weight of the shortest path tree. *Random Struct. Alg.* **30** 359–379.

https://doi.org/10.1017/S0963548309990204 Published online by Cambridge University Press