Random walk hitting times in random trees

by

Joubert Oosthuizen

Dissertation presented for the degree of Doctor of Philosophy in Mathematics at Stellenbosch University

Supervisor: Prof. S.G. Wagner

December 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2017

Abstract

The hitting time H_{xy}, between two vertices x and y of a graph, is the average time that the standard simple random walk takes to get from x to y. We start by giving a recursive formula for higher moments of the random time the simple random walk takes to get from x to y, that allows us to extend two well-known results for the hitting time on trees to all higher moments.
The main part of this thesis is concerned with the distribution of the hitting time between two randomly chosen vertices of a random tree. We consider both uniformly random labelled trees and a more general model with vertex weights akin to simply generated trees. We show that the r-th moment of the hitting time is of asymptotic order $n^{3r/2}$ in trees of order n, and we describe the limiting distribution upon normalisation by means of its moments. Moreover we also obtain joint moments with the distance between the two selected vertices and also the first moment of the hitting time variance.
Finally we consider three classes of random increasing trees. In this setup, the root is of special importance, and so we study the hitting time from the root to a random vertex, from a random vertex to the root and between two random vertices. The hitting times, for all three classes of increasing trees, from the root as well as between two random vertices is of order $n \log n$, with a normal limit law. On the other hand the hitting time to the root is only of linear order and converges in the limit to different Dickman distributions for the different classes of increasing trees.

Opsomming

Die treftyd H_{xy}, tussen twee punte x en y in 'n grafiek, is die gemiddelde tyd wat die ewekansige toevallige wandeling vat om te beweeg van x na y. Eerstens gee ons 'n rekursiewe formule vir die hoër momente van die treftyd wat ons in staat stel om twee bekende resultate oor die treftyd uit te brei na alle momente.

Die hoofdeel van hierdie tesis handel oor die verdeling van die treftyd tussen twee lukraak gekose punte in 'n lukrake boom. Ons beskou beide ewekansige lukrake gemerkte bome en 'n meer algemene model met geweegde punte soortgelyk aan eenvoudig gegenereerde bome. Ons wys dat die r-de moment van die treftyd van asimptotiese orde $n^{3r/2}$ is in bome van orde n en ons beskryf die limietverdeling na normalisering deur middel van sy momente. Verder verkry ons ook gesamentlike momente met die afstand tussen die twee gekose punte asook die eerste moment van die treftyd variansie.

Uiteindelike beskou ons drie klasse van lukraak toenemende bome. Die wortel van die boom is hier van spesiale belang en dus beskou ons die treftyd van die wortel na 'n lukrake punt, van 'n lukrake punt na die wortel en ook tussen twee lukrake punte. Die treftye, vir al drie klasse van toenemende bome, van die wortel asook tussen twee lukrake punte is van orde $n \log n$ met 'n normale limietverdeling. Andersyds is die treftyd na die wortel slegs van lineêre orde en konvergeer die verdelings in die limiet, vir die verskillende klasse van toenemende bome, na verskillende Dickman verdelings.

Acknowledgements

The writing of this thesis would certainly not have been possible without the assistance of Prof. Wagner, not only during this study, but also during the years leading up to the writing of this thesis. Furthermore his support allowed me to attend the ANALCO 2017 conference, after submitting [37], where some of the results in this thesis were presented.

I wish to extend my appreciation to:

- Prof S.J. Steel for introducing me to random walks.
- The University of Stellenbosch for hosting my studies and providing financial support in the form of a merit bursary.
- The National Research Foundation (NRF) for financial support (Reference: SFH150712125249).

Contents

Introduction

Random walks on graphs have diverse applications in fields such as computer science, physics and economics. We consider the standard simple random walk on a finite graph, which starts at a vertex x and moves at each step to one of the neighbours, chosen uniformly at random. This is a standard example of a finite Markov chain, and it is well known that every vertex is eventually reached with probability 1. The average time it takes the random walk to reach y from x is called the *hitting time* (or *first passage time*) and is denoted by H_{xy}.

The hitting time is a very natural parameter associated with a random walk and as such, it has been studied quite thoroughly. An influential paper appeared by Tetali [43], relating the hitting time to a weighted sum of effective resistances, obtained by replacing the graph by its underlying electrical network. In Chapter 2 we give an alternate proof of Tetali's formula and extend this proof to give a recursive formula for higher moments of the hitting time.

The first results on the distribution of the hitting time in random trees are due to Moon [31], who obtained the asymptotic behaviour of the first two moments of the hitting time between two randomly selected nodes of a uniformly random labelled tree. In Chapter 3, we refine the combinatorial analysis of Moon further by providing asymptotic formulas for all moments of the hitting time, which are sufficient to characterise the limiting distribution. We further generalise these results to labelled trees with additional vertex weights that depend on the degrees (akin to simply generated trees, but without roots).

Finally, in Chapter 4, we consider rather different classes of random trees, namely *increasing trees*. The root plays a special role, so we study the hitting time to and from the root as well as between two random vertices. Unlike in the previous chapter we were able to obtain explicit limiting distributions. The hitting time between two random vertices as well as the hitting time from the root has a Gaussian limiting distribution for all three classes of increasing trees we consider. On the other hand the hitting time to the root converges to different Dickman distributions for the different classes of increasing trees.

Intuitively speaking, the reason for this phenomenon is the fact that the root of a recursive tree is generally very central and therefore “attracts” a random walk faster than a random vertex, which typically lies closer to the fringes.

Chapter 1

Some preliminary results

The first section introduces important and well-known results on random walks on graphs that we will frequently require in the remaining chapters. The aim of the remaining two sections is to provide a brief introduction to asymptotic methods (mainly singularity analysis) we will require later in the text. The reader can omit these two sections at first reading and refer back to these sections when the need arises.

1.1 Random walks, electrical networks and the Wiener index

Let $G = (V(G), E(G))$ be a *simple, connected graph* with n vertices and m edges. We adopt the convention throughout the text of referring to the set of vertices of the graph as G. We start at a fixed vertex, move with equal probability to one of its neighbours, then move with equal probability to one of the neighbour's neighbours and so on. The random sequence of vertices $(X_n)_{n\in\mathbb{Z}_{\geq 0}}$ obtained in this way is a *simple random walk* on a graph. The *hitting time*

$$H_{xy} = \mathbb{E}_x\left(\min\{n \geq 0 : X_n = y\}\right)$$

is the expected number of steps to visit vertex y for the first time, starting from vertex x. There is a fruitful connection between random walks on graphs and electrical networks that allows one, among other things, to write the hitting time as a weighted sum of *effective resistances*. The books [13] and [2], by Doyle and Snell and Aldous and Fill respectively, are good introductory references to random walks, electrical networks and related connections.
Given a graph, we obtain the underlying electrical network by replacing vertices with nodes and edges with unit resistances. We define the *effective re-*

sistance between vertices x and y as in [43] as the voltage between x and y when a unit current enters x and leaves y. One can think of the effective resistance between x and y as the total resistance of the network between x and y, since it follows from Norton's theorem that the network of resistances can be replaced by an equivalent single resistance. The effective resistance is a *metric*. Non-negativity and symmetry follows from the definition, whereas the triangle equality is proved in [43, Corollary 3]. It is possible to write the effective resistance as a quotient of *determinants* of *Laplacian matrices*. The *Laplacian matrix* of a graph G is defined as

$$L(G) = D(G) - A(G),$$

where D is the *degree matrix* and A the *adjacency matrix* of the graph. In other words L is the $n \times n$ matrix with elements given by

$$l_{ij} = \begin{cases} d(i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise }, \end{cases}$$

for a graph with vertices $1, \ldots, n$ where $d(i)$ denotes the degree of vertex i. If $L(i_1, \ldots, i_k | j_1, \ldots, j_k)$ denotes the $(n-k) \times (n-k)$ matrix obtained from L by removing rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_k$ then the following theorem holds:

Theorem 1.1.1 (see [3, Theorem 4]). *The effective resistance between vertices x and y in a simple connected graph G is*

$$\begin{aligned} R_{xy} &= \frac{\det L(x, y|x, y)}{\det L(x, x)} \\ &= \frac{\det L(x, y|x, y)}{\tau(G)}, \end{aligned} \tag{1.1}$$

where τ is the number of spanning trees in the graph.

Equation (1.1) follows from the well-known Matrix-Tree Theorem, which states that the number of spanning trees in a graph is equal to any cofactor of the Laplacian matrix. It is shown in [7, Proposition 14.1] that $\det L(x, y|x, y)$ is the number of spanning forests of G consisting of two components, one containing x and the other y. Consequently $R_{xy} = d(x, y)$, where $d(x, y)$ is the distance betweem x and y, in a *tree* T (this also follows from the definition of effective resistance, since no current flows along the edges that are not on the (x, y)-path).

Returning to the subject of the hitting time, the following theorem can be proven by cleverly making use of a connection between random walks and electrical networks.

Theorem 1.1.2 (see [43, Theorem 5]). *The hitting time between any two vertices x, y in a finite, simple and connected graph G is*

$$H_{xy} = \frac{1}{2} \sum_{z \in G} d(z)(R_{xy} + R_{zy} - R_{zx}).$$

An immediate consequence of the previous theorem is a simple formula for the *commute time*:

$$C_{xy} = H_{xy} + H_{yx} = 2mR_{xy}. \tag{1.2}$$

If $G = T$ is a tree Theorem 1.1.2 simplifies to

$$\begin{aligned} H_{xy} &= \frac{1}{2} \sum_{z \in T} d(z)(d(x,y) + d(z,y) - d(z,x)) &(1.3)\\ &= \sum_{z \in T} d(z) l(x, z; y), &(1.4) \end{aligned}$$

where $l(x, z; y)$ is the length of intersection of the (x, y)-path and the (z, y)-path. Another easy consequence of Theorem 1.1.2 is:

Corollary 1.1.3 (see [27, Lemma 2]). *For any vertices x, y, z in a finite, simple and connected graph G,*

$$H_{xy} + H_{yz} + H_{zx} = H_{xz} + H_{zy} + H_{yx}.$$

An important consequence of the preceding symmetry property is the following:

Corollary 1.1.4 (see [27, Corollary 2.5]). *We can define a linear preorder on the vertices by $x \leq y$ if and only if $H_{xy} \leq H_{yx}$. Such an ordering can be obtained by fixing any vertex t and ordering the vertices according to the value of $H_{xt} - H_{tx}$.*

Proof. First note that if $H_{xt} - H_{tx} \leq H_{yt} - H_{ty}$ then $H_{xt} + H_{ty} \leq H_{yt} + H_{tx}$ so that $H_{xy} \leq H_{yx}$ by Theorem 1.1.3. The order is reflexive since $H_{xx} = 0$. Furthermore if $H_{xy} \leq H_{yx}$ and $H_{yz} \leq H_{zy}$ then $H_{xy} + H_{yz} \leq H_{yx} + H_{zy} \iff H_{xz} \leq H_{zx}$ by Corollary 1.1.3, implying transitivity. ∎

This ordering is not unique, because of the ties. However by Corollary 1.1.4 we can define an *equivalence relation* on the vertices by putting x and y in the same class if $H_{xy} = H_{yx}$. Now there is a unique ordering of the equivalence classes. Vertices in the highest class are easy to reach but difficult to get out of, while those in the lowest class are difficult to reach but easy to get out of. It is shown in [6] that in trees the vertex (or two adjacent vertices) in the highest class is the *barycenter* of a tree. The barycenter minimises the distance to all other vertices, i.e. it is the vertex achieving $\min_{x\in T}\sum_{y\in T} d(x,y)$. This result is further generalised in [19]. Here the authors defined the sums

$$D(z)=\sum_{v\in T} d(v,z),\ D_\pi(z)=\sum_{v\in T} d(v)d(v,z),\ CC(z)=\sum_{v\in T} H_{zv},$$

$$RC(z)=\sum_{v\in T} H_{vz} \text{ and } RC_\pi(z)=\sum_{v\in T} d(v)H_{vz},$$

respectively called the *centrality*, *weighted centrality*, *cover cost*, *reverse cover cost*, *weighted reverse cover cost* and proved:

Theorem 1.1.5 (see [19, Theorem 3]). *For any tree T, and any vertices $x, y \in T$, the following are equivalent:*

1. $H_{xy} \leq H_{yx}$;
2. $D(x) \geq D(y)$;
3. $D_\pi(x) \geq D_\pi(y)$;
4. $RC_\pi(x) \geq RC_\pi(y)$;
5. $RC(x) \geq RC(y)$;
6. $CC(x) \leq CC(y)$.

Moreover the following lemma can be used to exhibit a connection between partial sums of hitting times and the *Wiener index* of a tree defined as $W(T) := 1/2\sum_{x,y\in T} d(x,y)$, that is the sum of the distances of all pairs of vertices in the tree.

Lemma 1.1.6 (see [19, Lemma 4.1]). *For any vertex x of a tree T with n vertices, we have*

$$\sum_{y\in T} d(y)d(x,y) = 2\sum_{y\in T} d(x,y) - (n-1).$$

The previous lemma together with Theorem 1.1.2 gives:

Theorem 1.1.7 (see [19, Theorem 1]). *Let $W(T)$ be the Wiener index of a tree T. For any vertex $x \in T$, we have*

$$\sum_{y\in T} (H_{xy} + d(x,y)) = \sum_{x\in T}\sum_{y\in T} d(x,y) = 2W(T).$$

1.2 Singularity analysis

We wish to study the asymptotic growth of functions, whose expansion at an isolated singularity $\zeta \in \mathbb{C}\setminus\{0\}$, involves elements of the form

$$\left(1-\frac{z}{\zeta}\right)^{-\alpha}\left(\log\frac{1}{1-\frac{z}{\zeta}}\right)^{\beta}, \tag{1.5}$$

where $\alpha \in \mathbb{C}\setminus\mathbb{Z}_{\leq 0}, \beta \in \mathbb{C}$. We will frequently encounter these types of functions when studying the hitting time between two vertices on classes of random trees. The treatment in this section is based on [17, Chapter VI] and was first introduced in [16]. If $f(z)$ is singular at $z = \zeta$, then by the scaling rule of Taylor expansions $g(z) = f(z\zeta)$ satisfies

$$[z^n]f(z) = \zeta^{-n}[z^n]f(z\zeta) = \zeta^{-n}[z^n]g(z),$$

where $g(z)$ has a singularity at $z = 1$. As a consequence we shall only examine functions that are singular at 1. We will see, under suitable conditions, that functions of the type

$$(1-z)^{-\alpha}\left(\log\frac{1}{1-z}\right)^{\beta} \quad (z \to 1),$$

contribute subexponential terms of the form

$$n^{\alpha-1}(\log n)^{\beta}.$$

Consequently functions of the type in (1.5) contribute terms of the form

$$\zeta^{-n}n^{\alpha-1}(\log n)^{\beta}.$$

This means we only need to consider the singularities closest to the origin, called *dominant singularities*, since the value of terms contributed by singularities further away from the origin decays exponentially. For simplicity we will only consider functions with a single dominant singularity.

Theorem 1.2.1 (see [17, Theorem VI.2])**.** *Let α be an arbitrary complex number in $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$. The coefficient of z^n in the function*

$$f(z) = (1-z)^{-\alpha}\left(\frac{1}{z}\log\frac{1}{1-z}\right)^{\beta}$$

admits for large n a full asymptotic expansion in descending powers of $\log n$,

$$f_n = [z^n]f(z) \underset{n\to\infty}{\sim} \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{\beta}\left[1 + \frac{C_1}{\log n} + \frac{C_2}{\log^2 n} + \dots\right],$$

where $C_k = \binom{\beta}{k}\Gamma(\alpha)\frac{d^k}{ds^k}\frac{1}{\Gamma(s)}\Big|_{s=\alpha}$.

The idea behind the proof of the previous theorem is to estimate the coefficients of $f(z)$ by means of Cauchy's coefficient formula:

$$[z^n]f(z) = \frac{1}{2i\pi}\int_{\gamma} f(z)\frac{dz}{z^{n+1}}.$$

The next step is to find a suitable contour that encircles the origin. One finds that the integrand around parts of this contour is asymptotically irrelevant

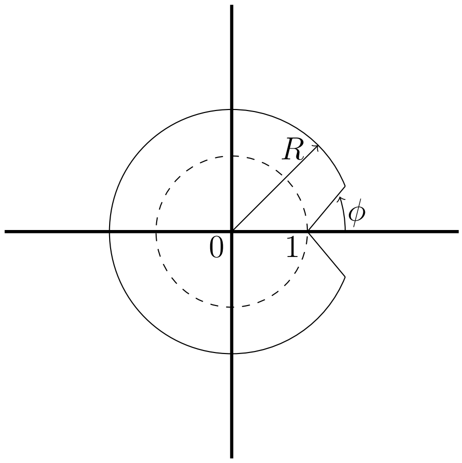

Figure 1.1: $\Delta(\phi, R)$-domain

and one is left to evaluate f_n around a *Hankel contour*, but we refer the reader to [17], to check the details.

Furthermore it is also possible to translate an approximation of a function near a singularity into an asymptotic approximation of its coefficients. Here we require our function to be analytic everywhere in a $\Delta(\phi, R)$ domain, see Figure 1.1, except at 1. The idea behind the proof of the following theorem is similar to that of Theorem 1.2.1 and relies on contour integration of Hankel-type paths.

Theorem 1.2.2 (see [17, Theorem VI.3])**.** *Let $\alpha, \beta \in \mathbb{R}$ and let $f(z)$ be a function that is Δ-analytic.*

1. *Assume that $f(z)$ satisfies in the intersection of a neighbourhood of* 1 *with its Δ-domain the condition*

$$f(z) = \mathcal{O}\left((1-z)^{-\alpha}\left(\log\frac{1}{1-z}\right)^{\beta}\right),$$

   then

$$[z^n]f(z) = \mathcal{O}\left(n^{\alpha-1}(\log n)^{\beta}\right).$$

2. *Assume that $f(z)$ satisfies in the intersection of a neighbourhood of* 1 *with its Δ-domain the condition*

$$f(z) = o\left((1-z)^{-\alpha}\left(\log\frac{1}{1-z}\right)^{\beta}\right),$$

   then

$$[z^n]f(z) = o\left(n^{\alpha-1}(\log n)^{\beta}\right).$$

One immediately obtains the following corollary as a consequence of the second part of Theorem 1.2.2:

Corollary 1.2.3 (see [17, Corollary VI.1]). *Assume that $f(z)$ is Δ-analytic, $\alpha \in \mathbb{R} \setminus \mathbb{Z}_{\leq 0}, \beta \in \mathbb{R}$ and*

$$f(z) \sim (1-z)^{-\alpha} \left(\frac{1}{z} \log \frac{1}{1-z} \right)^{\beta}, \qquad \text{as } z \to 1,\ z \in \Delta.$$

Then, the coefficients of $f(z)$ satisfy

$$[z^n] f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^{\beta}.$$

1.3 Inverse functions

The generating functions of certain families of trees satisfy an equation of the form

$$y(z) = z\phi(y(z)), \tag{1.6}$$

where $\phi(u)$ is analytic at $u = 0$, stemming from the recursive structure of the tree. We will assume $\phi(0) \neq 0$, so that there is a unique solution for $y(z)$. The aim of this section is to obtain information on the coefficients of $y(z)$. We will see that it is possible to find a *singular expansion* of $y(z)$, from which asymptotics for the coefficients of $y(z)$ can be derived by means of singularity analysis. The overview in this chapter is once again based on [17]. Before we can proceed we require the following two analytical inversion lemmas.

Lemma 1.3.1 (see [17, Lemma IV.2]). *Let $\psi(z)$ be analytic at y_0, with $\psi(y_0) = z_0$. Assume that $\psi'(y_0) \neq 0$. Then, for z in some small neighbourhood Ω_0 of z_0, there exists an analytic function $y(z)$ that solves the equation $\psi(y) = z$ and is such that $y(z_0) = y_0$.*

Lemma 1.3.2 (see [17, Lemma IV.3]). *Let $\psi(z)$ be analytic at y_0, with $\psi(y_0) = z_0$. Assume that $\psi'(y_0) = 0$ and $\psi''(y_0) \neq 0$. There exists a small neighbourhood Ω_0 of z_0 such that the following holds: for any fixed direction θ there exist two functions, $y_1(z)$ and $y_2(z)$ defined on*

$$\Omega_0^{\setminus\theta} := \{z \in \Omega_0 |\, \arg(z - z_0) \not\equiv \theta \mod 2\pi, z \neq z_0\}$$

that satisfy $\psi(y(z)) = z$; each is analytic in $\Omega_0^{\setminus\theta}$, has a singularity at the point z_0, and satisfies $\lim_{z\to z_0} y(z) = y_0$.

We can rewrite equation (1.6) as

$$\psi(y(z)) = z \quad \text{where} \quad \psi(u) = \frac{u}{\phi(u)},$$

so that it is an instance of Lemma 1.3.1. Making use of Lemma 1.3.1 it is possible to find the *radius of convergence* of $y(z)$:

Proposition 1.3.3 (see [17, Proposition IV.5]). *Let $\phi(u)$ be a complex function that satisfies the following conditions:*

- *$\phi(u)$ has nonnegative Taylor coefficients and is not of the form $\phi_0 + \phi_1 u$.*
- *$\phi(u)$ is analytic at 0 with radius of convergence $R > 0$ and $\phi(0) \neq 0$.*

Under the condition,

$$\lim_{x \to R^-} \frac{x\phi'(x)}{\phi(x)} > 1,$$

there exists a unique solution $\tau \in (0, R)$ of the characteristic equation,

$$\frac{\tau\phi'(\tau)}{\phi(\tau)} = 1.$$

Then, the formal solution $y(z)$ of the equation $y(z) = z\phi(y(z))$ is analytic at 0 and has radius of convergence

$$\rho = \frac{\tau}{\phi(\tau)} = \frac{1}{\phi'(\tau)}.$$

Since $y(z)$ has nonnegative coefficients it follows from Pringsheim's Theorem (see [17, Theorem IV.6]) that $y(z)$ has a dominant singularity located on its radius of convergence at ρ. The next theorem gives a singular expansion of $y(z)$ near ρ. For the sake of simplicity we will assume

$$\phi(u) = \sum_{j=0}^{\infty} \phi_j u^j$$

is *aperiodic*, which means here that $\gcd\{j : \phi_j \neq 0\} = 1$. (If this is not the case then $y(z)$ has multiple dominant singularities on its radius of convergence (see [17, Discussion VI.17]) and then one has to sum over the contributions of all the singularities located on the radius of convergence when performing a singularity analysis to find the asymptotics of y_n.)

Theorem 1.3.4 (see [17, Theorem VI.6]). *Suppose the conditions on ϕ in Proposition 1.3.3 hold and that there exists a unique positive solution to the characteristic equation, within the open disc of convergence of ϕ at 0, $|z| < R$:*

$$\exists \tau, 0 < \tau < R,\ \phi(\tau) - \tau\phi'(\tau) = 0.$$

Let $y(z)$ be the solution of $y = z\phi(y)$ satisfying $y(0) = 0$. Then, the quantity $\rho = \tau/\phi(\tau)$ is the radius of convergence of $y(z)$ at 0, and the singular expansion of $y(z)$ near ρ is of the form

$$y(z) = \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\sqrt{1 - \frac{z}{\rho}} + \sum_{j \geq 2} (-1)^j d_j \left(1 - \frac{z}{\rho}\right)^{\frac{j}{2}}, \qquad (1.7)$$

with the d_j being computable constants. Assume that, in addition, ϕ is aperiodic. Then one has

$$[z^n]y(z) \sim \sqrt{\frac{\phi(\tau)}{2\phi''(\tau)}} \frac{\rho^{-n}}{\sqrt{\pi n^3}} \left(1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k}\right), \tag{1.8}$$

for a family e_k of computable constants.

Proof. We give an outline of the proof in [17]. Consider the inverse $\psi(u) = \frac{u}{\phi(u)}$ of $y(z)$. The first derivative is

$$\psi'(u) = \frac{1}{\phi(u)} - \frac{u\phi'(u)}{\phi(u)^2} = \frac{1}{\phi(u)}\left(1 - \frac{u\phi'(u)}{\phi(u)}\right),$$

implying $\psi'(\tau) = 0$. Furthermore the second the derivative is

$$\begin{aligned}\psi''(u) &= -\frac{\phi'(u)}{\phi(u)^2} - \frac{\phi'(u) + u\phi''(u)}{\phi(u)^2} + 2\frac{u\phi'(u)^2}{\phi(u)^3} \\ &= 2\frac{\phi'(u)}{\phi(u)^2}\left(\frac{u\phi'(u)}{\phi(u)} - 1\right) - \frac{u\phi''(u)}{\phi(u)^2},\end{aligned}$$

implying $\psi''(\tau) = -\rho\frac{\phi''(\tau)}{\phi(\tau)} \neq 0$, since $\phi_j \neq 0$ for some $j \geq 2$. It follows from Lemma 1.3.2 that there are two solutions to $y(z) = z\phi(y(z))$, which are analytic in a slit neighbourhood of ρ. We now have in the vicinity of $(z, y) = (\rho, \tau)$:

$$\begin{aligned}\rho - z &= \psi(\tau) - \psi(y(z)) \\ &= \psi(\tau) - \left(\psi(\tau) + \frac{\psi''(\tau)}{2}(y(z)-\tau)^2 + \frac{\psi'''(\tau)}{6}(y(z)-\tau)^3 + \ldots\right) \\ &= \rho\frac{\phi''(\tau)}{2\phi(\tau)}(y(z)-\tau)^2\left(1 + c_1(y(z)-\tau) + c_2(y(z)-\tau)^2 + \ldots\right),\end{aligned}$$

implying

$$\begin{aligned}(y(z)-\tau)^2 &= \frac{2\phi(\tau)}{\phi''(\tau)}\left(1 - \frac{z}{\rho}\right) - (y(z)-\tau)^2\left(c_1(y(z)-\tau) + \ldots\right) \\ &= \frac{2\phi(\tau)}{\phi''(\tau)}\left(1 - \frac{z}{\rho}\right)\left(1 - c_1(y(z)-\tau) + \mathcal{O}\left((y(z)-\tau)^2\right)\right),\end{aligned}$$

by iterating. It follows from the binomial series of $\sqrt{1 - c_1(y(z)-\tau)}$ that

$$y(z) = \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}\left(1 - \frac{z}{\rho}\right)}\left(1 - \frac{c_1}{2}(y(z)-\tau) + \mathcal{O}\left((y(z)-\tau)^2\right)\right), \tag{1.9}$$

where we chose the determination with a $-\surd$ since $y(z)$ increases to τ^- as $z \to \rho^-$. The singular expansion in (1.7) follows by iteratively substituting $y(z)$ into the right-hand side of (1.9). Finally it can be shown if ϕ is aperiodic then $y(z)$ has a unique singularity on its radius of convergence so that the asymptotic equivalence in (1.8) is obtained from singularity analysis of (1.7). ∎

Example 1. The exponential generating function of rooted labelled (Cayley) trees satisfies the equation $Y(z) = ze^{Y(z)}$. We have $\phi(u) = e^u$, $\tau = 1$ and $\rho = \frac{1}{e}$ so that $Y(z) \sim 1 - \sqrt{2}\sqrt{1-ez}$ and

$$\left(\frac{1}{1-Y(z)}\right)^m \sim \left(\frac{1}{2(1-ez)}\right)^{\frac{m}{2}}, m \in \mathbb{Z}_{\geq 1}. \tag{1.10}$$

Chapter 2

Higher moments of hitting times on graphs

The aim of this chapter is to find a recursive formula for higher moments of hitting times on graphs. Doing so we extend two well known results for hitting times on trees, namely that the hitting time between two vertices in a tree is an integer and is maximised by the endpoints of the path, to all higher moments.

2.1 A recursive formula

We start by giving an alternative proof of Theorem 1.1.2. Let G be a graph with vertices $1, \ldots, n$. W.l.o.g. fix the vertex 1. Conditioning on the first step of the random walk gives

$$H_{i1} = 1 + \frac{1}{d(i)} \sum_{j \sim i} H_{j1} \quad \text{for} \quad i = 2, \ldots, n,$$

where the sum is over all neighbouring vertices j of i, while $H_{11} = 0$. Let L be the *Laplacian matrix* of G and L_1 the matrix obtained from L by replacing the first row of L by the first row of the identity matrix. We can express the *linear system* of hitting times H_{i1} $(i = 1, \ldots, n)$ as

$$L_1 H = \begin{bmatrix} 0 \\ d(2) \\ \vdots \\ d(n) \end{bmatrix} \text{ where } H = \begin{bmatrix} H_{11} \\ H_{21} \\ \vdots \\ H_{n1} \end{bmatrix}.$$

It follows from the Matrix-Tree Theorem that $\det(L_1) = \tau > 0$, where τ is the number of *spanning trees* in the graph. Hence the linear system of hitting times has a unique solution. Cramer's rule yields

$$H_{i1} = \frac{\det(L_1^i)}{\tau}$$

where L_1^i is the matrix formed from L_1 by replacing the i^{th} column of L_1 by

$$\begin{bmatrix} 0 & d(2) & \dots & d(n) \end{bmatrix}^T.$$

By cofactor expansion along the i^{th} column of L_1^i $(i = 2, \dots, n)$ we obtain

$$\begin{aligned} \det(L_1^i) &= \det\left(L_1^i(1|1)\right) \\ &= \sum_{j=2}^{n} d(j)(-1)^{i+j} \det\left(L(1,j|1,i)\right), \end{aligned} \tag{2.1}$$

where we define $A(i_1, \dots, i_n | j_1, \dots, j_n)$ as the matrix obtained from A after deleting rows $i_1, \dots, i_n$ and columns $j_1, \dots, j_n$.
It follows from [44, Lemma 4] that

$$\begin{aligned} &\det\left(L(1,j|1,i)\right) \\ &= \frac{1}{2}(-1)^{2+i+j} \left(\det\left(L(1,j|1,j)\right) + \det\left(L(1,i|1,i)\right) - \det\left(L(i,j|i,j)\right)\right). \end{aligned} \tag{2.2}$$

Substituting (2.2) into (2.1) gives

$$\det(L_1^i) = \frac{1}{2}\sum_{j=2}^{n} d(j) \left(\det\left(L(1,j|1,j)\right) + \det\left(L(1,i|1,i)\right) - \det\left(L(i,j|i,j)\right)\right).$$

Recalling Theorem 1.1.1, which states that $R_{ij} = \frac{\det(L(i,j|i,j))}{\tau}$, we obtain the formula for the hitting time between two vertices in Theorem 1.1.2:

$$\begin{aligned} H_{i1} &= \frac{1}{2}\sum_{j=2}^{n} d(j) \left(R_{1j} + R_{1i} - R_{ij}\right) \\ &= \frac{1}{2}\sum_{j=1}^{n} d(j) \left(R_{1j} + R_{1i} - R_{ij}\right). \end{aligned}$$

Define the higher moments of the hitting time as $H_{ij}^{(m)} := \mathbb{E}\left(T_{i,j}^m\right)$, where $H_{ij}^{(0)} = 1$. Here $T_{i,j}$ is the number of steps the random walk takes to visit vertex j for the first time from vertex i. We are now ready to state the main theorem of this chapter.

Theorem 2.1.1. *The higher moments of the hitting time between any two vertices i, j, in a simple, finite and connected graph G, is given recursively by*

$$H_{ij}^{(m)} = \frac{1}{2}\sum_{k\in G} d(k)(R_{ij} + R_{kj} - R_{ik}) \sum_{l=0}^{m-1} \binom{m}{l} (-1)^{m-1-l} H_{kj}^{(l)}.$$

Proof. As before we work on a graph with vertices $1, \ldots, n$ and w.l.o.g. fix 1 and find the higher moments of the hitting time to 1. We start by showing

$$d(i)H_{i1}^{(m)} - \sum_{j\sim i} H_{j1}^{(m)} = d(i)\left(\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{i1}^{(k)}\right) \tag{2.3}$$

for $i \neq 1$. Let $T'_{i,1}$ be the remaining number of steps to visit 1 after the first step from i. Conditioning on the first step of the random walk gives

$$\begin{aligned} H_{i1}^{(m)} &= \mathbb{E}\left(T_{i,1}^m\right) \\ &= \mathbb{E}\left(\left(1 + T'_{i,1}\right)^m\right) \\ &= \sum_{r=0}^{m}\binom{m}{r}\frac{1}{d(i)}\sum_{j\sim i} H_{j1}^{(r)}, \end{aligned}$$

which we can rewrite as

$$d(i)H_{i1}^{(m)} - \sum_{j\sim i} H_{j1}^{(m)} = d(i) + \sum_{j\sim i}\left(\binom{m}{1}H_{j1} + \ldots + \binom{m}{m-1}H_{j1}^{(m-1)}\right). \tag{2.4}$$

Our strategy now is to make use of the previous equation to iteratively substitute in $\sum_{j\sim i} H_{j1}^{(k)}$, starting from $k = m-1$ up until $k = 2$ until (2.4) reduces to

$$d(i)H_{i1}^{(m)} - \sum_{j\sim i} H_{j1}^{(m)} = d(i) + C\sum_{j\sim i} H_{j1} + R_1. \tag{2.5}$$

For example substituting in $\binom{m}{m-1}\sum_{j\sim i} H_{j1}^{(m-1)}$ into the right-hand side of (2.4), we get an expression i.t.o. $\sum_{j\sim i} H_{j1}^{(k)}$ for $k = 1, \ldots, m-2$, where each $\sum_{j\sim i} H_{j1}^{(k)}$ has coefficient

$$\binom{m}{k} - \binom{m-1}{k}\binom{m}{m-1}.$$

We start by finding C in (2.5) by means of induction. Suppose that the coefficient of $\sum_{j\sim i} H_{j1}^{(k)}$ $(k = 1, \ldots, l-1)$ is

$$\binom{m}{k}\left(1 - (m-k) + \frac{(m-k)(m-k-1)}{2!} - \ldots\right.$$

$$\left. + (-1)^{(m-l)} \frac{(m-k)(m-k-1)\dots(m-k-(m-l-1))}{(m-l)!} \right)$$
$$= \binom{m}{k} (-1)^{m-l} \binom{m-k-1}{m-l},$$

if the highest moment in the expression obtained from (2.4) is $\sum_{j\sim i} H_{j1}^{(l-1)}$, for $l-1 = m-1, \dots, 2$. The base case holds since if the highest moment in the expression obtained from (2.4) is $m-1$ then the coefficient of $\sum_{j\sim i} H_{j1}^{(k)}$ $(k = 1, \dots, m-1)$ is $\binom{m}{k}$. We need to show if the highest moment in the expression obtained from (2.4) is $\sum_{j\sim i} H_{j1}^{(l-2)}$, then the coefficient of $\sum_{j\sim i} H_{j1}^{(k)}$ $(k = 1, \dots, l-2)$ is

$$\binom{m}{k} (-1)^{m-(l-1)} \binom{m-k-1}{m-(l-1)}.$$

After substituting in

$$\sum_{j\sim i} H_{j1}^{(l-1)} = d(i)H_{i1}^{(l-1)} - d(i) - \binom{l-1}{1} \sum_{j\sim i} H_{j1} - \dots - \binom{l-1}{l-2} \sum_{j\sim i} H_{j1}^{(l-2)}$$

our proof by induction will be complete if we can show that

$$(-1)^{m-l} \binom{m}{k} \binom{m-k-1}{m-l} - (-1)^{m-l} \binom{l-1}{k} \binom{m}{l-1} \binom{m-l}{m-l}$$
$$= (-1)^{m-(l-1)} \binom{m}{k} \binom{m-k-1}{m-(l-1)}.$$

This is indeed the case since

$$(-1)^{m-l} \binom{m}{k} \binom{m-k-1}{m-l} - (-1)^{m-l} \binom{l-1}{k} \binom{m}{l-1} \binom{m-l}{m-l}$$
$$= (-1)^{m-l} \left(\frac{m!(m-l+1) - m!(m-k)}{k!(m-k)(m-l+1)!(l-k-1)!} \right)$$
$$= (-1)^{m-(l-1)} \frac{m!(l-1-k)}{k!(m-k)(m-l+1)!(l-k-1)!}$$
$$= (-1)^{m-(l-1)} \binom{m}{k} \binom{m-k-1}{m-(l-1)}.$$

Setting $k = 1$ and $l = 2$ it follows from (2.4) that

$$C = (-1)^{m-2} \binom{m-2}{m-2} \binom{m}{1} = m(-1)^m.$$

Finally we need to find R_1. Note that we obtained an extra term

$$\begin{aligned}
&\binom{m}{k}\left(1-\frac{m-k}{1!}+\ldots+(-1)^{m-1-k}\frac{(m-k)\ldots 2}{(m-1-k)!}\right)\left(d(i)H_{i1}^{(k)}-d(i)\right)\\
&=\binom{m}{k}\left(\sum_{i=0}^{m-1-k}\binom{m-k}{i}(-1)^i\right)\left(d(i)H_{i1}^{(k)}-d(i)\right)\\
&=\binom{m}{k}(-1)^{m-1-k}\left(d(i)H_{i1}^{(k)}-d(i)\right),
\end{aligned}$$

for each $\sum_{j\sim i} H_{j1}^{(k)}$ $(k=2,\ldots,m-1)$ we substituted in. Hence

$$R_1=\sum_{k=2}^{m-1}\binom{m}{k}(-1)^{m-1-k}\left(d(i)H_{i1}^{(k)}-d(i)\right).$$

Note that

$$\begin{aligned}
d(i)\sum_{k=2}^{m-1}\binom{m}{k}(-1)^{m-k}&=d(i)(-1)^m\sum_{k=2}^{m-1}\binom{m}{k}(-1)^k\\
&=d(i)(-1)^m\left((-1)^{m-1}-1+\binom{m}{1}\right)
\end{aligned}$$

and recall

$$\sum_{j\sim i} H_{j1}=d(i)\left(H_{i1}-1\right).$$

Consequently it follows from (2.5) that

$$\begin{aligned}
&d(i)H_{i1}^{(m)}-\sum_{j\sim i}H_{j1}^{(m)}\\
&=d(i)\left(1-1+(-1)^{m+1}+(-1)^m\binom{m}{1}\right)\\
&\quad+m(-1)^m d(i)\left(H_{i1}-1\right)+d(i)\sum_{k=2}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{i1}^{(k)}\\
&=d(i)\left((-1)^{m+1}+\sum_{k=1}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{i1}^{(k)}\right),
\end{aligned}$$

which is what we wanted to show since this is equivalent to (2.3). We can write (2.3) as a linear system of hitting times as we did at the start of this chapter:

$$L_1 H^{(m)} = \begin{bmatrix} 0 \\ d(2)\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{21}^{(k)} \\ \vdots \\ d(n)\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{n1}^{(k)} \end{bmatrix}$$

where

$$H^{(m)} = \begin{bmatrix} H_{11}^{(m)} \\ H_{21}^{(m)} \\ \vdots \\ H_{n1}^{(m)} \end{bmatrix}.$$

Applying Cramer's rule and working exactly as we did at the start of the chapter yields

$$H_{i1}^{(m)} = \frac{1}{2}\sum_{j=1}^{n} d(j)\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{j1}^{(k)}\left(R_{i1}+R_{j1}-R_{ij}\right),$$

which completes the proof of the theorem since the vertex 1 was chosen w.l.o.g. ■

Example 2. Let $S = S_{n+1}$ be the $n+1$-vertex star with central vertex c and adjacent leaves $1,\dots n$. Fix a leaf, say 1. We will find the exponential generating functions of $H_{c1}^{(m)}$ and $H_{l1}^{(m)}$ for $m \geq 1$, where l is any fixed leaf of S. We may assume $l \neq 1$, since $H_{11}^{(m)} = 0$ for $m \geq 1$. It follows that

$$\begin{aligned} H_{c1}^{(m)} &= \frac{1}{2}d(c)\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{c1}^{(k)}\left(d(c,1)+d(c,1)-d(c,c)\right) \\ &\quad + \frac{1}{2}\sum_{j=2}^{n} d(j)\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{j1}^{(k)}\left(d(c,1)+d(j,1)-d(c,j)\right) \\ &= n\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{c1}^{(k)} + (n-1)\sum_{k=0}^{m-1}\binom{m}{k}(-1)^{m-1-k}H_{l1}^{(k)}, \end{aligned}$$

therefore

$$\begin{aligned} &(1-n)H_{c1}^{(m)} - (n-1)H_{c1}^{(m)} \\ &= n\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-1-k}H_{c1}^{(k)} + (n-1)\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-1-k}H_{l1}^{(k)} \end{aligned} \tag{2.6}$$

for $m \geq 1$. Note that

$$d(l,1) + d(j,1) - d(l,j) = \begin{cases} 4 & \text{if } j = l \\ 2 & \text{if } j \neq l \end{cases}$$

for a leaf j, therefore

$$H_{l1}^{(m)} = n \sum_{k=0}^{m-1} \binom{m}{k} (-1)^{m-1-k} H_{c1}^{(k)} + n \sum_{k=0}^{m-1} \binom{m}{k} (-1)^{m-1-k} H_{l1}^{(k)}.$$

Hence

$$\begin{aligned} &(1-n)H_{l1}^{(m)} - nH_{c1}^{(m)} \\ &= n \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-1-k} H_{c1}^{(k)} + n \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-1-k} H_{l1}^{(k)} \end{aligned} \tag{2.7}$$

for $m \geq 1$. Let

$$H_c(x) = \sum_{m=0}^{\infty} H_{c1}^{(m)} \frac{x^m}{m!} \quad \text{and} \quad H_l(x) = \sum_{m=0}^{\infty} H_{l1}^{(m)} \frac{x^m}{m!}$$

be the exponential generating functions of $H_{c1}^{(m)}$ and $H_{l1}^{(m)}$. We can multiply each term in (2.6) by $\frac{x^m}{m!}$ and sum over both sides for $m \geq 1$. For the case $m = 0$ we have to add 1 to the right-hand side of (2.6) so that it is equal to the left-hand side. Since

$$\sum_{m=0}^{\infty} a_m \frac{x^m}{m!} \sum_{m=0}^{\infty} b_m \frac{x^m}{m!} = \sum_{m=0}^{\infty} \sum_{r=0}^{m} \binom{m}{r} a_r b_{m-r} \frac{x^m}{m!},$$

it follows from (2.6) that

$$(1-n)H_c(x) - (n-1)H_l(x) = n\left(-e^{-x}H_c(x)\right) + (n-1)\left(-e^{-x}H_l(x)\right) + 1. \tag{2.8}$$

Similarly it follows from (2.7) that

$$(1-n)H_l(x) - nH_c(x) = n\left(-e^{-x}H_c(x)\right) + n\left(-e^{-x}H_l(x)\right) + 1. \tag{2.9}$$

Hence

$$(1-n)H_c(x) - (n - 1 + e^{-x})H_l(x) = (1-n)H_l(x) - nH_c(x),$$

therefore

$$H_c(x) = e^{-x}H_l(x). \tag{2.10}$$

Substituting (2.10) into (2.8) and solving the equation gives

$$H_c(x) = \frac{e^{-x}}{ne^{-2x} - (n-1)},$$

and therefore

$$H_l(x) = \frac{1}{ne^{-2x} - (n-1)}.$$

Example 3. We find the variance of the random time it takes to move from one endpoint to the other on the path of length n on vertices $0, 1, \dots, n$. It is well known that $H_{in} = n^2 - i^2$ for $i = 0, 1, \dots, n$. Now

$$\begin{aligned}
\mathrm{Var}_{0,n} &:= H_{0n}^{(2)} - (H_{0n})^2 \\
&= \frac{1}{2}\sum_{j=0}^{n} d(j)\,(-1 + 2H_{jn})\,(d(0,n) + d(j,n) - d(j,0)) - n^4 \\
&= \frac{1}{2}\sum_{j=0}^{n} d(j)\left(-1 + 2(n^2 - j^2)\right)(2n - 2j) - n^4 \\
&= n\left(-1 + 2n^2\right) + 2\sum_{j=1}^{n-1}\left(-n + j + 2n^3 - 2j^2 n - 2jn^2 + 2j^3\right) - n^4 \\
&= \frac{2}{3}\left(n^4 - n^2\right).
\end{aligned}$$

2.2 Some results on hitting times in trees

It is known for a simple random walk that the hitting times between two vertices of a tree are positive integers (see [4]) and more generally if there is a unique path in a graph then the hitting time between the endpoints of this path is an integer (see [9, Corollary 2.5 and 2.6]). The same result holds for the second moment and hence the variance. We obtain the following generalisation of these results from Theorem 2.1.1:

Corollary 2.2.1. *If there exists a unique path between two vertices in a graph then all higher moments of the hitting time between these two vertices are positive integers. In particular all higher moments of hitting times are positive integers on trees.*

The second result we will generalise to all higher moments is that the hitting time between two vertices in a tree is maximised at the endpoints of a path (see [8, Corollary]). We start by showing this result holds for the variance. Before we can proceed we require the following lemma, where we define $T_{x:y}$ as the subtree, of a tree T with adjacent vertices x and y, rooted at x after removing the edge xy.

Lemma 2.2.2. *The maximum value of the weighted reverse cover cost*

$$RC_\pi(r;T) := \sum_{v\in T} d(v) H_{vr}$$

among all trees of order $n \geq 2$*, rooted at a vertex* r*, is* $\frac{4}{3}n^3 - 4n^2 + \frac{11}{3}n - 1$*, and it is attained by a path, rooted at one of its ends.*

Proof. First note that $D(r) := \sum_{v \in T} d(v, r)$ can only be maximised when r is a leaf. The contrary cannot hold, since then $d(r) \geq 2$ so that $|T_{r_1:r}| < |T_{r:r_1}|$ for one of the neighbours r_1 of r, which implies $D(r_1) > D(r)$. It now follows from Theorem 1.1.5 that $RC_\pi(r; T)$ is maximised when r is a leaf. We proceed by induction on n. The statement holds trivially for $n = 2$. Let T be a tree of order $n > 2$. If t is the neighbour of r, then

$$\begin{aligned} RC_\pi(r;T) &= \sum_{v \in T \setminus \{r\}} d(v) H_{vr} \\ &= \sum_{v \in T \setminus \{r\}} d(v) \left(H_{vt} + H_{tr} \right) \\ &= RC_\pi(t; T \setminus \{r\}) + H_{tr} + 2(n-2)H_{tr} \\ &= RC_\pi(t; T \setminus \{r\}) + (2n-3)^2. \end{aligned}$$

The result follows from our induction hypothesis. ∎

Theorem 2.2.3. *The hitting time variance in a tree of size n is maximised by the endpoints of the path.*

Proof. Let $T_{k:l}$ be the component of $T - \{e\}$ that contains k, where k and l are adjacent vertices with edge $e = kl$. Let T' be the induced subtree on the vertices $T_{k:l} \cup \{l\}$. Now the *hitting time variance* from k to l satisfies

$$\begin{aligned} \mathrm{Var}_{k,l} &:= H_{kl}^{(2)} - (H_{kl})^2 \\ &= \frac{1}{2} \sum_{j=1}^{n} d(j) \left(-1 + 2H_{jl} \right) \left(d(k,l) + d(j,l) - d(k,j) \right) \\ &\quad - \left[\frac{1}{2} \sum_{j=1}^{n} d(j) (d(k,l) + d(j,l) - d(k,j)) \right]^2 \\ &= - \sum_{i \in T_{k:l}} d(i) + 2 \sum_{i \in T_{k:l}} d(i) H_{il} - \left(\sum_{i \in T_{k:l}} d(i) \right)^2 \\ &= - \left(2|T_{k:l}| - 1 \right) + 2 \sum_{i \in T_{k:l}} d(i) H_{il} - \left(2|T_{k:l}| - 1 \right)^2 \\ &= 2|T_{k:l}| - 4|T_{k:l}|^2 + 2 \sum_{i \in T'} d(i) H_{il}. \end{aligned} \tag{2.11}$$

It follows from Lemma 2.2.2 that $\mathrm{Var}_{k,l}$ is maximised when $T_{k:l}$ is a path. Since the random times to move along the edges of the path are independent we have

$$\mathrm{Var}_{1,n} = \mathrm{Var}_{1,2} + \mathrm{Var}_{2,3} + \ldots + \mathrm{Var}_{n-1,n}.$$

The result follows since each summand is non-negative and yields the maximum possible variance for a tree of size $|T_{i:i+1}| = i$ $(i = 1, \ldots, n-1)$. ∎

Proposition 2.2.4. *The hitting time variance between two vertices x, y in a tree is a multiple of* 8.

Proof. It is shown in [19] that the weighted reverse cover cost of a tree T, with n vertices, satisfies the identity

$$RC_{\pi}(l) := \sum_{i\in T} d(i)H_{il} = 4(n-1)D(l) + (n-1) - 4W(T),$$

where $D_T(l)$ is defined as in the first chapter as the sum of the distances from the other vertices in T to l and $W(T)$ is the Wiener index of T. Hence we can rewrite $\sum_{i\in T'} d(i)H_{il}$ in (2.11) as

$$\sum_{i\in T'} d(i)H_{il} = 4|T^*|\left(D_{T^*}(l) + |T^*|\right) + |T^*| - 4\left(W(T^*) + D_{T^*}(l) + |T^*|\right),$$

where $T^* := T_{k:l}$ as above. Simplifying (2.11) gives

$$\begin{aligned} \mathrm{Var}_{k,l} &= 2|T^*| - 4|T^*|^2 + 8|T^*|\left(D_{T^*}(k) + |T^*|\right) + 2|T^*| - 8\left(W(T^*) + D_{T^*}(k) + |T^*|\right) \\ &= 8\left(\binom{|T^*|}{2} + |T^*|D_{T^*}(k) - D_{T^*}(k) - W(T^*)\right). \end{aligned} \tag{2.12}$$

The result now follows from a simple induction on $d(x,y)$ (if $d(x,y)=0$ the hitting time variance is 0 and for the induction step we can use equation (2.12)).

∎

Finally we show that the endpoints of the path are maximal for all higher moments of the hitting time on trees. In this case we will be able to prove the result without making use of Theorem 2.1.1. Before proceeding we require the following lemma.

Lemma 2.2.5. *Let $(v_0, v_1, \ldots, v_n)$ be the vertices of a path of length n. Let $\tau_n := T_{v_{n-1},v_n}$ be the random time it takes to move along the last edge on the path. The function*

$$f_a(n) := \mathbb{E}\left((\tau_n + 1)^a\right),$$

with initial value $f_a(0) := 0$, satisfies the inequality

$$f_a\left(\sum_{i=1}^{l} n_i\right) \geq \sum_{i=1}^{l} f_a(n_i)$$

for $a \in \mathbb{N}$ and every l-tuple of positive integers $n_1, \ldots, n_l$.

Proof. Notice that

$$\tau_n = \begin{cases} 1 & \text{with probability } 1/2 \\ 1 + \tau_{n-1} + \theta_n & \text{with probability } 1/2, \end{cases}$$

where

$$\theta_n \overset{d}{=} \tau_n$$

and θ_n and τ_{n-1} are independent. Hence

$$\begin{aligned} f_a(n) &= 2^{a-1} + \frac{1}{2}\mathbb{E}\left((\tau_{n-1} + \theta_n + 2)^a\right) \\ &= 2^{a-1} + \frac{1}{2}\sum_{b=0}^{a}\binom{a}{b}\mathbb{E}\left((\tau_{n-1}+1)^b\right)\mathbb{E}\left((\theta_n+1)^{a-b}\right) \\ &= 2^{a-1} + \frac{1}{2}\sum_{b=0}^{a}\binom{a}{b} f_b(n-1) f_{a-b}(n), \end{aligned}$$

implying

$$f_a(n) = 2^a + \sum_{b=1}^{a}\binom{a}{b} f_b(n-1) f_{a-b}(n). \tag{2.13}$$

Moreover $f_0(n) = 1$ for all non-negative integers n and $f_a(0) = 0$ for $a \in \mathbb{N}$. We first obtain

$$\begin{aligned} f_a(n) &= 2^a + f_a(n-1) + \sum_{b=1}^{a-1}\binom{a}{b} f_b(n-1) f_{a-b}(n) \\ &\geq 2^a + f_a(n-1). \end{aligned} \tag{2.14}$$

We proceed to show by two-dimensional induction on a and n that

$$f_a(n) \geq f_a(n_1) + f_a(n_2),$$

if $n_1 + n_2 = n$ and $a \geq 1$. The statement holds for $a = 1$ (and any n) since $f_1(n) = 2n$ (this follows from equation (1.2) since only one step is required to move from the end vertex in a path to its neighbouring vertex). The induction result is trivial for $n_1 = 0$ or $n_2 = 0$. In particular the other base case for $n = 1$, $a \in \mathbb{N}$ holds. Now

$$\begin{aligned} f_a(n) &= 2^a + f_a(n-1) + \sum_{b=1}^{a-1}\binom{a}{b} f_b(n-1) f_{a-b}(n) \\ &\geq 2^a + 2^a + f_a(n-2) + \sum_{b=1}^{a-1}\binom{a}{b} f_b(n-1) f_{a-b}(n), \end{aligned}$$

by (2.14). Furthermore it follows from the induction hypothesis that

$$f_b(n-1) \geq f_b(n-2) \geq f_b(n_1-1) + f_b(n_2-1)$$

and

$$f_{a-b}(n) \geq f_{a-b}(n_1) + f_{a-b}(n_2).$$

Hence

$$\begin{aligned} f_a(n) \geq\ & 2^a + 2^a + f_a(n_1-1) + f_a(n_2-1) \\ & + \sum_{b=1}^{a-1} \binom{a}{b} \left(f_b(n_1-1) f_{a-b}(n_1) + f_b(n_2-1) f_{a-b}(n_2) \right), \end{aligned}$$

which completes the proof by induction. The statement of the lemma now follows by induction on l and iterating.

■

Theorem 2.2.6. *All higher moments of the hitting time between two vertices in a tree are maximised by the endpoints of a path.*

Proof. Let x, y be two vertices in a tree connected by the path $(x = x_0, x_1, \ldots x_k = y)$. Since

$$T_{x,y} = \sum_{j=1}^{k} T_{x_{j-1},x_j}$$

and the random times to move along the edges are independent, it follows from the multinomial theorem that

$$\mathbb{E}\left(T_{x,y}^r\right) = \sum_{a_1+\ldots+a_k=r} \binom{r}{a_1, a_2, \ldots a_k} \prod_{j=1}^{k} \mathbb{E}\left(T_{x_{j-1},x_j}^{a_j}\right).$$

The number of terms is maximised on a path so we need only show, for adjacent vertices u and v, that $\mathbb{E}\left(T_{u,v}^{a_j}\right)$ is maximised by a path, where there is a fixed number of vertices, say n, closer to or equal to u. We proceed to show $\mathbb{E}\left(T_{u,v}^r\right)$ is maximal on a path by two-dimensional induction on n and r. If $r = 1$ and $n \in \mathbb{N}$ the result holds trivially, since $\mathbb{E}\left(T_{u,v}\right)$ remains constant irrespective of the shape of the tree attached to u. If $r \in \mathbb{N}$ and $n = 1$ the result is also trivial. Suppose the result holds for $n \in \mathbb{N}$, $k < r$ and $r \in \mathbb{N}$, $m < n$. Let the other neighbours of u in a tree T be $z_1, \ldots, z_l$ (see Figure 2.1).

Conditioning on the first step of the random walk gives

$$\mathbb{E}\left(T_{u,v}^r\right) = \frac{1}{l+1} \left(1 + \sum_{j=1}^{l} \mathbb{E}\left(\left(T_{z_j,u} + T_{u,v} + 1 \right)^r \right) \right). \tag{2.15}$$

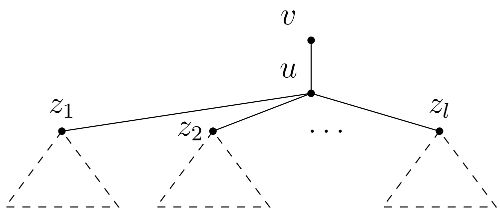

Figure 2.1: Neighbours of u in T.

Hence

$$(l+1)\mathbb{E}\left(T_{u,v}^r\right) = 1 + l\mathbb{E}\left(T_{u,v}^r\right) + \sum_{j=1}^{l} \sum_{\substack{b_1+b_2+b_3=r \\ b_1 \neq r}} \binom{r}{b_1, b_2, b_3} \mathbb{E}\left(T_{u,v}^{b_1}\right) \mathbb{E}\left(T_{z_j,u}^{b_2}\right),$$

implying

$$\mathbb{E}\left(T_{u,v}^r\right) = 1 + \sum_{j=1}^{l} \sum_{\substack{b_1+b_2+b_3=r \\ b_1 \neq r}} \binom{r}{b_1, b_2, b_3} \mathbb{E}\left(T_{u,v}^{b_1}\right) \mathbb{E}\left(T_{z_j,u}^{b_2}\right).$$

It follows from the induction hypothesis that $\mathbb{E}\left(T_{u,v}^{b_1}\right)$ is maximal for a path and that $\mathbb{E}\left(T_{z_j,u}^{b_2}\right)$ is maximal if the j-th branch is a path for $j = 1, \dots, l$. Our proof will be complete if we can show that $\mathbb{E}\left(T_{u,v}^r\right)$ is maximised when $l = 1$. Since we can also write (2.15) as

$$\mathbb{E}\left(T_{u,v}^r\right) = 1 + \sum_{j=1}^{l} \sum_{s=1}^{r} \binom{r}{s} \mathbb{E}\left(T_{u,v}^{r-s}\right) \mathbb{E}\left(\left(T_{z_j,u}+1\right)^s\right),$$

we are left to show $\sum_{j=1}^{l} \mathbb{E}\left(\left(T_{z_j,u}+1\right)^s\right)$ is maximised when $l = 1$, but this is just the statement of Lemma 2.2.5, so we are done.

$\blacksquare$

Chapter 3

Hitting times in random labelled trees

In this chapter we study the distribution of the hitting time between two randomly chosen vertices of a random labelled tree. We consider uniformly random labelled trees in the first section and a more general model with vertex weights, akin to simply generated trees, in the next section. Our results are obtained by means of singularity analysis of generating functions.

3.1 Labelled trees

A *labelled tree* T is a tree where the vertices bear distinct integer labels $1, 2, \dots, |V(T)|$. Let $\mathcal{T} = \cup_{n\geq 1}\mathcal{T}_n$ be the class of non-rooted labelled trees, where $\mathcal{T}_n$ is the class of all such trees of size n. It is well known that $|\mathcal{T}_n| = n^{n-2}$. Before stating the main theorem of this section we develop some intuition behind it. First recall the definition of the Wiener index of a tree T:

$$W(T) = \frac{1}{2}\sum_{x\in T}\sum_{y\in T} d(x, y).$$

Janson determined the limiting distribution of the Wiener index of a random labelled tree in [22]. He made use of Aldous' theory of the continuum random tree [1] to show that this limiting distribution can be described in terms of Brownian excursions:

Theorem 3.1.1 (see [22, Theorem 3.1])**.** *Let T_n be a random labelled tree with Wiener index $W(T_n)$. The distribution of the normalised random variable $n^{-5/2}W(T_n)$ converges weakly to a random variable ζ that can be described in* *terms of a normalised Brownian excursion* $e(t), 0 < t < 1$, *as follows:*

$$\xi = \int_0^1 e(t)dt,$$
$$\eta = 4 \iint_{0<s<t<1} \min_{s\le u\le t} e(u) ds dt,$$
$$\zeta = \xi - \eta = 2 \iint_{0<s<t<1} \left(e(s) + e(t) - 2 \min_{s\le u\le t} e(u) \right) ds\, dt.$$

We know from Theorem 1.1.7 that

$$\sum_{x\in T} \sum_{y\in T} H_{xy} = (2n-1)W(T).$$

Now Slutsky's theorem (see [21, Theorem 2.39]) together with Theorem 3.1.1 gives

$$\frac{n^{-7/2}}{2} \sum_{x\in T} \sum_{y\in T} H_{xy} \xrightarrow{d} \zeta.$$

It is therefore reasonable to expect that the r-th moment of the hitting time satisfies the asymptotic equivalence

$$\mathbb{E}(H_{xy}^r) \sim C_r n^{3r/2}$$

for some constant C_r. This is indeed the case and the aim of this section is to prove the following theorem:

Theorem 3.1.2. *Let x and y be two vertices that are selected uniformly at random of a uniformly random labelled tree with n vertices. The r-th moment of the hitting time between x and y satisfies the following asymptotic formula:*

$$\mathbb{E}(H_{xy}^r) \sim C_r n^{3r/2},$$

where

$$C_r = \frac{\sqrt{\pi}}{\Gamma(\frac{1}{3})} \cdot \left(\frac{3}{\sqrt{2}}\right)^r \cdot \frac{\Gamma(r+1)\Gamma(r+\frac{1}{3})}{\Gamma(\frac{3r+1}{2})}.$$

Consequently, the normalised random variable $n^{-3/2}H_{xy}$ converges weakly to a limit law that is characterised by the moment sequence C_r. This limit law has a continuous density on $[0,\infty)$.

This theorem generalises the results obtained by Moon in [31], where he found the first two moments of the hitting time (between two vertices in a labelled tree where the choices are all made uniformly at random). First we require a combinatorial decomposition between two vertices on the endpoints of a path in a tree, in order to find the higher moments of the hitting time between two

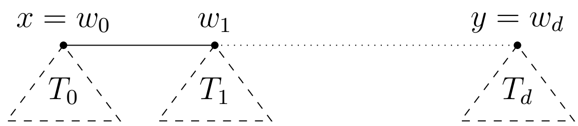

Figure 3.1: Decomposition along the path between x and y.

vertices in a labelled tree. Let $x = w_0, w_1, w_2, \ldots, w_d = y$ be the vertices on this path. When the edges of this path are removed, we are left with $d+1$ components, each of which can be interpreted as a rooted tree. Let T_j be the tree rooted at w_j, see Figure 3.1.

Recall that the hitting time between x and y is given by

$$H_{xy} = \frac{1}{2}\sum_{v\in T} d(v)\left(d(x,y) + d(v,y) - d(v,x)\right).$$

Note that if $v \in T_j$ then

$$\begin{aligned} d(v,y) &= d(w_j,y) + d(v,w_j) \\ &= d(x,y) - d(x,w_j) + d(v,w_j) \end{aligned}$$

and $d(v,x) = d(w_j,x) + d(v,w_j)$. We also have

$$\sum_{v\in T_j} d(v) = \begin{cases} 2|T_j| - 1 & \text{if } j = 0, d \\ 2|T_j| & \text{if } 0 < j < d. \end{cases}$$

Hence

$$\begin{aligned} H_{xy} &= \sum_{j=0}^{d} \left(d(x,y) - d(x,w_j)\right) \sum_{v\in T_j} d(v) \\ &= \sum_{j=0}^{d} 2(d-j)|T_j| - d. \end{aligned}$$

Note that the size of T_d is irrelevant in this formula, since the only vertex the random walk from x to y visits in T_d is the root. Consider the bivariate exponential generating function

$$H(z,u) = \sum_{T,x,y} \frac{1}{|T|!} z^{|T|} u^{H_{xy}}.$$

We can write $H(z,u) = \sum_{d=0}^{\infty} H_d(z,u)$, where

$$H_d(z,u) = \sum_{\substack{T,x,y \\ d(x,y)=d}} \frac{1}{|T|!} z^{|T|} u^{H_{xy}}.$$

If $|T| = n$ we can distribute the labels among the subtrees $T_0, \dots, T_d$ in $\binom{n}{|T_0|, |T_1|, \dots, |T_d|}$ ways. Distributing the labels over the rooted subtrees fixes the labels of the roots of T_0 and T_d so there is no more need to sum over all vertex pairs. Hence

$$\begin{aligned} H_d(z,u) &= \sum_{T_0, T_1, \dots, T_d} \frac{u^{-d}}{n!} \binom{n}{|T_0|, |T_1|, \dots, |T_d|} \prod_{j=0}^{d} z^{|T_j|} u^{2(d-j)|T_j|} \\ &= u^{-d} \sum_{T_0} \frac{z^{|T_0|} u^{2d|T_0|}}{|T_0|!} \sum_{T_1} \frac{z^{|T_1|} u^{2(d-1)|T_1|}}{|T_1|!} \cdots \sum_{T_d} \frac{z^{|T_d|} u^{2(d-d)|T_d|}}{|T_d|!} \\ &= u^{-d} \prod_{j=0}^{d} Y\left(z u^{2(d-j)}\right) \\ &= u^{-d} \prod_{k=0}^{d} Y\left(z u^{2k}\right), \end{aligned} \tag{3.1}$$

where

$$Y(z) = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} z^n$$

is the generating function for rooted labelled trees (there are n^{n-2} labelled trees and a further n possible ways to choose a root). If we choose a tree T uniformly at random from $\mathcal{T}_n$ and two vertices x and y (where x and y may be the same vertex) uniformly at random from T, then we can write the probability that the hitting time, between two random vertices on a random labelled tree, is equal to k as

$$\Pr_{\mathcal{T}_n, x, y}\left(H_{xy} = k\right) = \frac{n![u^k z^n] H(z,u)}{n![z^n] H(z,1)}.$$

We can now derive the factorial moments of the hitting time from the probability generating function

$$\sum_k \Pr_{\mathcal{T}_n, x, y}\left(H_{xy} = k\right) u^k = \frac{[z^n] H(z,u)}{[z^n] H(z,1)}, \tag{3.2}$$

by taking partial derivatives with respect to u and setting $u = 1$.

Proposition 3.1.3. *Let x and y be two randomly selected vertices of a uniformly random labelled tree with n vertices. The expected hitting time between x and y satisfies the asymptotic equivalence*

$$\mathbb{E}_{\mathcal{T}_n, x, y}\left(H_{xy}\right) \sim \sqrt{\frac{\pi}{2}} n^{3/2}.$$

Proof. It follows from the product rule that

$$\begin{aligned}\partial_u H(z,u) &= \sum_{d=0}^{\infty}\left(-du^{-d-1}\prod_{k=0}^{d} Y\left(zu^{2k}\right) + u^{-d}\partial_u \prod_{k=0}^{d} Y\left(zu^{2k}\right)\right)\\ &= \sum_{d=0}^{\infty}\left(-du^{-d-1}\prod_{k=0}^{d} Y\left(zu^{2k}\right) + u^{-d}\prod_{k=0}^{d} Y\left(zu^{2k}\right) z\sum_{l=1}^{d}\frac{2lu^{2l-1}Y'\left(zu^{2l}\right)}{Y\left(zu^{2l}\right)}\right),\end{aligned} \tag{3.3}$$

so that

$$\partial_u H(z,u)\Big|_{u=1} = \sum_{d=0}^{\infty}\left(-dY(z)^{d+1} + d(d+1)zY'(z)Y(z)^d\right). \tag{3.4}$$

Since

$$\left(\frac{1}{1-Y(z)}\right)^2 = \sum_{d=0}^{\infty}(d+1)Y(z)^d, \tag{3.5}$$

we have

$$\begin{aligned}-\sum_{d=0}^{\infty} dY(z)^{d+1} &= -Y(z)\left(\sum_{d=0}^{\infty}(d+1)Y(z)^d - \sum_{d=0}^{\infty}Y(z)^d\right)\\ &= -\frac{Y(z)}{(1-Y(z))^2} + \frac{Y(z)}{1-Y(z)}\\ &\sim -\frac{1}{2(1-ez)} - \frac{1}{\sqrt{2}\sqrt{1-ez}} + \frac{1}{\sqrt{2}\sqrt{1-ez}} - 1\\ &\sim -\frac{1}{2(1-ez)},\end{aligned} \tag{3.6}$$

where the asymptotic equivalence follows from (1.10). We also have

$$\begin{aligned}\left(\frac{1}{1-Y(z)}\right)^3 &= \sum_{d=0}^{\infty}\frac{(d+1)(d+2)}{2}Y(z)^d\\ &= \frac{1}{2}\left(\sum_{d=0}^{\infty}d(d+1)Y(z)^d + 2\sum_{d=0}^{\infty}(d+1)Y(z)^d\right),\end{aligned}$$

so that

$$\frac{1}{2}\sum_{d=0}^{\infty}d(d+1)Y(z)^d = \left(\frac{1}{1-Y(z)}\right)^3 - \left(\frac{1}{1-Y(z)}\right)^2. \tag{3.7}$$

Now

$$\sum_{d=0}^{\infty}d(d+1)zY'(z)Y(z)^d = 2zY'(z)\left(\left(\frac{1}{1-Y(z)}\right)^3 - \left(\frac{1}{1-Y(z)}\right)^2\right)$$

$$\sim \frac{2ze}{\sqrt{2(1-ez)}}\left(\left(\frac{1}{2(1-ez)}\right)^{3/2}-\frac{1}{2(1-ez)}\right)$$
$$\sim \frac{1}{2(1-ez)^2}, \tag{3.8}$$

where the asymptotic equivalence again follows from (1.10). Adding equations (3.6) and (3.8) gives

$$\partial_u H(z,u)\Big|_{u=1} \sim \frac{1}{2(1-ez)^2}.$$

It follows from (3.2) that

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(H_{xy}\right) = \frac{[z^n]\partial_u H(z,u)\Big|_{u=1}}{[z^n]H(z,1)},$$

where

$$[z^n]H(z,1) = \frac{n^n}{n!}$$

since there are n^{n-2} labelled trees and n^2 choices of vertices x and y. Applying Stirling's approximation yields

$$\frac{1}{[z^n]H(z,1)} \sim \sqrt{2\pi}n^{1/2}e^{-n}. \tag{3.9}$$

On the other hand it follows from singularity analysis that

$$[z^n]\partial_u H(z,u)\Big|_{u=1} \sim \frac{ne^n}{2},$$

so that

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(H_{xy}\right) \sim \sqrt{\frac{\pi}{2}}n^{3/2}.$$

■

We proceed to find the variance.

Proposition 3.1.4. *Let x and y be two randomly selected vertices of a uniformly random labelled tree with n vertices. The variance of the hitting time between x and y satisfies the asymptotic equivalence*

$$\operatorname{Var}_{\mathcal{T}_n,x,y}\left(H_{xy}\right) \sim \left(\frac{32}{15}-\frac{\pi}{2}\right)n^3.$$

Proof. Taking the partial derivative with respect to u of (3.3) yields

$$
\begin{aligned}
&\partial_u^2 H(z,u)\\
&= \partial_u \sum_{d=0}^{\infty} \left(-du^{-d-1} \prod_{k=0}^{d} Y\left(zu^{2k}\right) + u^{-d} \prod_{k=0}^{d} Y\left(zu^{2k}\right) z \sum_{l=1}^{d} \frac{2lu^{2l-1}Y'\left(zu^{2l}\right)}{Y\left(zu^{2l}\right)} \right)\\
&= \sum_{d=0}^{\infty} \left(d(d+1)u^{-d-2} \prod_{k=0}^{d} Y\left(zu^{2k}\right) - du^{-d-1} \prod_{k=0}^{d} Y\left(zu^{2k}\right) z \sum_{l=1}^{d} \frac{2lu^{2l-1}Y'\left(zu^{2l}\right)}{Y\left(zu^{2l}\right)} \right)\\
&\quad + \sum_{d=0}^{\infty} \left(-du^{-d-1} \prod_{k=0}^{d} Y\left(zu^{2k}\right) + u^{-d} \prod_{k=0}^{d} Y\left(zu^{2k}\right) z \sum_{l=1}^{d} \frac{2lu^{2l-1}Y'\left(zu^{2l}\right)}{Y\left(zu^{2l}\right)} \right)\\
&\quad \cdot \left(z \sum_{k=1}^{d} \frac{2ku^{2k-1}Y'\left(zu^{2k}\right)}{Y\left(zu^{2k}\right)} \right) + \sum_{d=0}^{\infty} u^{-d} \prod_{k=0}^{d} Y\left(zu^{2k}\right)\\
&\quad \cdot \sum_{k=0}^{d} \frac{Y\left(zu^{2k}\right)\left[Y'\left(zu^{2k}\right)2kz(2k-1)u^{2k-2} + (2kzu^{2k-1})^2 Y''\left(zu^{2k}\right)\right] - \left[Y'\left(zu^{2k}\right)(2kzu^{2k-1})\right]^2}{Y\left(zu^{2k}\right)^2}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\partial_u^2 H(z,u)\Big|_{u=1}\\
&= \sum_{d=0}^{\infty} \left(-d(-d-1)Y(z)^{d+1} - d^2(d+1)Y(z)^d zY'(z) \right)\\
&\quad + \sum_{d=0}^{\infty} d(d+1)z\frac{Y'(z)}{Y(z)} \left(-dY(z)^{d+1} + d(d+1)Y(z)^{d+1}\frac{zY'(z)}{Y(z)} \right)\\
&\quad + \sum_{d=0}^{\infty} Y(z)^{d+1} \sum_{k=0}^{d} \frac{Y(z)\left[Y'(z)2k(2k-1)z + (2kz)^2 Y''(z)\right] - (2kzY'(z))^2}{Y(z)^2}\\
&= \sum_{d=0}^{\infty} \left(-d(-d-1)Y(z)^{d+1} - 2d^2(d+1)zY'(z)Y(z)^d + d^2(d+1)^2 z^2 Y'(z)^2 Y(z)^{d-1} \right)\\
&\quad + \sum_{d=0}^{\infty} \frac{2}{3} Y(z)^{d-1} d(d+1)(2d+1) \left[Y(z)Y'(z)z + z^2 \left(Y(z)Y''(z) - Y'(z)^2 \right) \right]\\
&\quad - \sum_{d=0}^{\infty} d(d+1)zY'(z)Y(z)^d. \qquad (3.10)
\end{aligned}
$$

We evaluate each term of (3.10) asymptotically:

$$
\sum_{d=0}^{\infty} d(d+1)Y(z)^{d+1} = 2Y(z) \left(\left(\frac{1}{1-Y(z)} \right)^3 - \left(\frac{1}{1-Y(z)} \right)^2 \right)
$$

$$
\begin{aligned}
&\sim 2\left(1-\sqrt{2(1-ez)}\right)\left(\frac{1}{(2(1-ez))^{3/2}}-\frac{1}{2(1-ez)}\right)\\
&\sim \frac{1}{\sqrt{2}}\frac{1}{(1-ez)^{3/2}}, \qquad (3.11)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{d=0}^{\infty} d^2(d+1)Y(z)^d zY'(z) \qquad (3.12)\\
&= 2zY(z)Y'(z)\frac{1+2Y(z)}{(1-Y(z))^4}\\
&\sim \frac{3}{2\sqrt{2}}(1-ez)^{-5/2}, \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{d=0}^{\infty} d^2(d+1)^2 z^2 Y'(z)^2 Y(z)^{d-1}\\
&= 4z^2Y'(z)^2\frac{1+4Y(z)+Y(z)^2}{(1-Y(z))^5}\\
&\sim \frac{3}{\sqrt{2}}(1-ez)^{-7/2}, \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{d=0}^{\infty}\frac{2}{3}Y(z)^{d-1}d(d+1)(2d+1)\left[Y(z)Y'(z)z+z^2\left(Y(z)Y''(z)-Y'(z)^2\right)\right]\\
&=\frac{4(1+Y(z))}{(1-Y(z))^4}\left[Y'(z)z+\frac{e^2}{2\sqrt{2}}z^2(1-ez)^{-3/2}\right]\\
&\sim \frac{1}{\sqrt{2}}(1-ez)^{-7/2}, \qquad (3.15)
\end{aligned}
$$

$$
\sum_{d=0}^{\infty} d(d+1)zY'(z)Y(z)^d \sim \frac{1}{2(1-ez)^2}, \qquad (3.16)
$$

as in (3.8). It follows after adding equations (3.11)-(3.16) that

$$
\partial_u^2 H(z,u)\Big|_{u=1} \sim \frac{4}{\sqrt{2}}(1-ez)^{-7/2}. \qquad (3.17)
$$

Singularity analysis of (3.17) yields

$$
[z^n]\partial_u^2 H(z,u)\Big|_{u=1} \sim \frac{4e^n}{\sqrt{2}\Gamma(7/2)}n^{5/2}.
$$

Now

$$\begin{aligned}\mathbb{E}_{\mathcal{T}_n,x,y}\left(H_{xy}(H_{xy}-1)\right) &= \frac{[z^n]\partial_u^2 H(z,u)\Big|_{u=1}}{[z^n]H(z,1)} \\ &\sim \sqrt{2\pi}n^{1/2}e^{-n}\cdot\frac{32}{15\sqrt{2\pi}}e^n n^{5/2} \\ &= \frac{32}{15}n^3, \end{aligned} \tag{3.18}$$

so that

$$\mathrm{Var}_{\mathcal{T}_n,x,y}\left(H_{xy}\right) \sim \left(\frac{32}{15}-\frac{\pi}{2}\right)n^3.$$

■

We require some auxiliary calculations to find the higher moments in general. In the following, it will be shown that $\mathbb{E}\left(H_{xy}^r\right)$ is of order $n^{3r/2}$ for all r. It will be somewhat more convenient to work with the generating function for $H_{xy}+d(x,y)$ though, since it does not contain the additional factor u^{-d}:

$$\begin{aligned}\tilde{H}(z,u) &= \sum_{T,x,y}\frac{1}{|T|!}z^{|T|}u^{H_{xy}+d(x,y)} \\ &= \sum_{d=0}^{\infty}\prod_{k=0}^{d}Y\left(zu^{2k}\right).\end{aligned}$$

The r-th derivative with respect to u yields the r-th factorial moment. The asymptotic behaviour of this r-th derivative is determined in the following lemma:

Lemma 3.1.5. *Around the dominant singularity $\frac{1}{e}$, we have*

$$\begin{aligned}&\partial_u^r\tilde{H}(z,u)\Big|_{u=1} \\ &\quad\sim 2^{-(r+1)/2}(1-ez)^{-\frac{3r+1}{2}} \\ &\qquad\sum_{m=1}^{r}\sum_{h_1+\ldots+h_m=r}\binom{r}{h_1,h_2,\ldots,h_m}(r+m)! \\ &\qquad\prod_{i=1}^{m}(2h_i-3)!!\prod_{i=1}^{m}\frac{1}{h_1+\ldots+h_i+i}.\end{aligned} \tag{3.19}$$

Proof. The general Leibniz rule gives us

$$\partial_u^r\prod_{k=0}^{d}Y\left(zu^{2k}\right)\Big|_{u=1}$$

$$= \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m \geq 1 \\ h_1+\ldots+h_m=r}} \binom{r}{h_1,\ldots,h_m} \sum_{1\leq j_1<j_2<\ldots<j_m\leq d} Y(z)^{d+1-m} \prod_{i=1}^{m} \left[\partial_u^{h_i} Y\left(zu^{2j_i}\right)\right]\Big|_{u=1}. \tag{3.20}$$

It can be verified by induction on h that $\partial_u^h Y\left(zu^{2j}\right)\Big|_{u=1} \sim (2jz)^h Y^{(h)}(z)$. Now

$$\begin{aligned} \partial_u^h Y\left(zu^{2j}\right)\Big|_{u=1} &\sim (2jz)^h Y^{(h)}(z) \\ &\sim (2jz)^h e^h (2h-3)!!(2(1-ez))^{1/2-h} \\ &\sim \sqrt{2}(2h-3)!!j^h(1-ez)^{1/2-h}, \end{aligned} \tag{3.21}$$

so that

$$\begin{aligned} &\partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1} \\ &\sim \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m \geq 1 \\ h_1+\ldots+h_m=r}} \binom{r}{h_1,h_2,\ldots,h_m} \sum_{1\leq j_1<j_2<\ldots<j_m\leq d} Y(z)^{d+1-m}(1-ez)^{\frac{m}{2}-r} \\ &\cdot 2^{\frac{m}{2}} \prod_{i=1}^{m} \left[(2h_i-3)!!{j_i}^{h_i}\right]. \end{aligned}$$

We proceed by showing by induction on m that

$$\sum_{1\leq j_1<j_2<\ldots<j_m\leq d} \prod_{i=1}^{m} {j_i}^{h_i} \tag{3.22}$$

$$= \frac{d^{h_1+h_2+\ldots+h_m+m}}{(h_1+1)(h_1+h_2+2)\ldots(h_1+h_2+\ldots+h_m+m)} + \mathcal{O}\left(d^{h_1+h_2+\ldots+h_m+m-1}\right). \tag{3.23}$$

The case $m=1$ follows from Faulhaber's formula, which states that

$$\sum_{i=1}^{n} i^p = \frac{(n+1)^{p+1}}{p+1} + \sum_{k=1}^{p} \frac{B_k}{p-k+1}\binom{p}{k}(n+1)^{p+1-k}, \tag{3.24}$$

where B_k denotes the k^{th} Bernoulli number. Suppose

$$\sum_{1\leq j_1<j_2<\ldots<j_{m-1}\leq d} \prod_{i=1}^{m-1} {j_i}^{h_i}$$

$$= \frac{d^{h_1+h_2+\ldots+h_{m-1}+m-1}}{(h_1+1)(h_1+h_2+2)\ldots(h_1+h_2+\ldots+h_{m-1}+m-1)} + \mathcal{O}\left(d^{h_1+h_2+\ldots+h_{m-1}+m-2}\right).$$

Now

$$\begin{aligned}
&\sum_{1\le j_1<j_2<\ldots<j_m\le d}\prod_{i=1}^{m} {j_i}^{h_i} \\
&= \sum_{j_m=1}^{d} j_m^{h_m}\sum_{j_{m-1}=1}^{j_m-1}\cdots\sum_{j_1=1}^{j_2-1}\prod_{i=1}^{m-1} {j_i}^{h_i} \\
&= \sum_{j_m=1}^{d} j_m^{h_m}\frac{(j_m-1)^{h_1+h_2+\ldots+h_{m-1}+m-1}}{(h_1+1)(h_1+h_2+2)\ldots(h_1+h_2+\ldots+h_{m-1}+m-1)} \\
&\quad + \sum_{j_m=1}^{d} j_m^{h_m}\mathcal{O}\left((j_m-1)^{h_1+h_2+\ldots+h_{m-1}+m-2}\right) \\
&= \sum_{j_m=1}^{d}\left[\frac{{j_m}^{h_1+h_2+\ldots+h_m+m-1}}{(h_1+1)\ldots(h_1+\ldots+h_{m-1}+m-1)} + \mathcal{O}\left({j_m}^{h_1+\ldots+h_m+m-2}\right)\right] \\
&= \frac{d^{h_1+h_2+\ldots+h_m+m}}{(h_1+1)\ldots(h_1+h_2+\ldots+h_m+m)} + \mathcal{O}\left(d^{h_1+h_2+\ldots+h_m+m-1}\right),
\end{aligned}$$

where the last equation follows from the binomial theorem and (3.24). Hence

$$\begin{aligned}
&\sum_{d=0}^{\infty}\partial_u^r\prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1} \\
&\sim \sum_{m=1}^{r}\sum_{h_1+\ldots+h_m=r}\binom{r}{h_1,h_2,\ldots,h_m}\sum_{d=m-1}^{\infty} d^{r+m}Y(z)^{d+1-m} \\
&\quad\cdot\prod_{i=1}^{m}(2h_i-3)!!\prod_{i=1}^{m}\frac{1}{h_1+\ldots+h_i+i}(1-ez)^{\frac{m}{2}-r}2^{\frac{m}{2}},
\end{aligned}\tag{3.25}$$

where the sum over d starts from $m-1$ since the sum $\sum_{1\le j_1<j_2<\ldots<j_{m-1}\le d}$ vanishes for $d=0,1,\ldots m-2$. Note that

$$\begin{aligned}
\left(\frac{1}{1-Y(z)}\right)^{r+m+1} &= \sum_{d=0}^{\infty}\binom{r+m+d}{d}Y(z)^d \\
&= \sum_{d=0}^{\infty}\left(\frac{1}{(r+m)!}d^{r+m} + \mathcal{O}\left(d^{r+m-1}\right)\right)Y(z)^d.
\end{aligned}$$

The asymptotic equivalence

$$\left(\frac{1}{1-Y(z)}\right)^{r+m+1} \sim \sum_{d=0}^{\infty}\frac{d^{r+m}}{(r+m)!}Y(z)^d$$

follows inductively from

$$\begin{aligned}\sum_{d=0}^{\infty} d^{r+m-1} Y(z)^d &\sim \frac{(r+m-1)!}{(1-Y(z))^{r+m}} \\ &\sim \frac{(r+m-1)!}{2(1-ez)^{\frac{r+m}{2}}}.\end{aligned}$$

In particular

$$\begin{aligned}\sum_{d=m-1}^{\infty} d^{r+m} Y(z)^{d+1-m} &\sim \sum_{d=0}^{\infty} d^{r+m} Y(z)^{d+1-m} \\ &\sim \sum_{d=0}^{\infty} d^{r+m} Y(z)^{d} \\ &\sim \frac{(r+m)!}{(1-Y(z))^{r+m+1}},\end{aligned} \tag{3.26}$$

Substituting (3.26) into (3.25) gives

$$\begin{aligned}&\sum_{d=0}^{\infty} \partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\bigg|_{u=1} \\ &\sim 2^{-(r+1)/2}(1-ez)^{-\frac{3r+1}{2}} \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m \geq 1 \\ h_1+\ldots+h_m=r}} \binom{r}{h_1, h_2, \ldots, h_m} (r+m)! \\ &\cdot \prod_{i=1}^{m} (2h_i - 3)!! \prod_{i=1}^{m} \frac{1}{h_1 + \ldots + h_i + i}.\end{aligned}$$

■

After some further manipulations, the double sum simplifies considerably:
Lemma 3.1.6. *Around the dominant singularity $\frac{1}{e}$, we have*

$$\begin{aligned}&\partial_u^r \tilde{H}(z,u)\bigg|_{u=1} \\ &\sim 2^{-(r+1)/2} r!(3r-2)!!!(1-ez)^{-(3r+1)/2},\end{aligned} \tag{3.27}$$

where the triple factorial $n!!!$ is defined recursively as

$$n!!! = \begin{cases} n & 0 < n \leq 3, \\ n \cdot (n-3)!!! & n > 3. \end{cases}$$

Proof. We can write

$$\prod_{i=1}^{m} \frac{1}{h_1 + \ldots + h_i + i} = \int_0^1 \int_0^{x_m} \cdots \int_0^{x_2} x_1^{h_1} x_2^{h_2} \ldots x_m^{h_m} dx_1 dx_2 \ldots dx_m$$
$$= \int \cdots \int_{0 \le x_1 \le \ldots \le x_m \le 1} x_1^{h_1} x_2^{h_2} \ldots x_m^{h_m} dx_1 dx_2 \ldots dx_m.$$

Now

$$\sum_{m=1}^{r} \sum_{\substack{h_1, h_2, \ldots, h_m \ge 1 \\ h_1 + \ldots + h_m = r}} r!(r+m)! \prod_{i=1}^{m} \frac{(2h_i - 3)!!}{h_i!} \prod_{i=1}^{m} \frac{1}{h_1 + \ldots + h_i + i}$$
$$= \sum_{m=1}^{r} r!(r+m)! [u^r] \int \cdots \int_{0 \le x_1 \le \ldots \le x_m \le 1} \prod_{i=1}^{m} \left(\sum_{h_i = 1}^{\infty} \frac{(2h_i - 3)!!}{h_i!} u^{h_i} x_i^{h_i} \right) dx_1 \ldots dx_m$$
$$= \sum_{m=1}^{r} r!(r+m)! [u^r] \int \cdots \int_{0 \le x_1 \le \ldots \le x_m \le 1} \prod_{i=1}^{m} \left(1 - \sqrt{1 - 2ux_i}\right) dx_1 \ldots dx_m$$
$$= \sum_{m=1}^{r} r!(r+m)! [u^r] \frac{1}{m!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{m} \left(1 - \sqrt{1 - 2ux_i}\right) dx_1 \ldots dx_m \tag{3.28}$$
$$= \sum_{m=1}^{r} r!(r+m)! [u^r] \frac{1}{m!} \left(\int_0^1 \left(1 - \sqrt{1 - 2ux}\right) dx \right)^m$$
$$= \sum_{m=1}^{r} \frac{r!(r+m)!}{m!} [u^r] \left(\frac{3u - 1 + (1 - 2u)^{3/2}}{3u} \right)^m, \tag{3.29}$$

where (3.28) follows from symmetry since

$$\int_0^1 \cdots \int_0^1 \prod_{i=1}^{m} \left(1 - \sqrt{1 - 2ux_i}\right) dx_1 \ldots dx_m$$
$$= \sum_{\pi \in S_m} \int \cdots \int_{0 \le \pi(x_1) \le \ldots \le \pi(x_m) \le 1} \prod_{i=1}^{m} \left(1 - \sqrt{1 - 2ux_i}\right) d\pi(x_1) \ldots d\pi(x_m)$$
$$= m! \int \cdots \int_{0 \le x_1 \le \ldots \le x_m \le 1} \prod_{i=1}^{m} \left(1 - \sqrt{1 - 2ux_i}\right) dx_1 \ldots dx_m,$$

where S_m is the set of all permutations of $\{1, 2, \ldots, m\}$. Now

$$\sum_{m=1}^{r} \frac{(r+m)!}{m!} x^m = \sum_{m=1}^{\infty} \frac{(r+m)!}{m!} x^m - \sum_{m=r+1}^{\infty} \frac{(r+m)!}{m!} x^m$$

$$= r!\left((1-x)^{-r-1} - 1\right) - \sum_{m=r+1}^{\infty} \frac{(r+m)!}{m!} x^m,$$

so that

$$\sum_{d=0}^{\infty} \partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1}$$
$$\sim 2^{-(r+1)/2}(1-ez)^{-\frac{3r+1}{2}} r!^2 \left[u^r\right] \left(\left(\frac{1-(1-2u)^{3/2}}{3u}\right)^{-r-1} - 1\right).$$

Finally

$$\begin{aligned}
&\left[u^r\right]\left(\frac{1-(1-2u)^{3/2}}{3u}\right)^{-r-1} \\
&= \frac{1}{2\pi i}\oint_{\mathcal{C}} u^{-r-1}\left(\frac{1-(1-2u)^{3/2}}{3u}\right)^{-r-1} du \\
&= \frac{1}{2\pi i}\oint_{\mathcal{C}} \left(\frac{1-(1-2u)^{3/2}}{3}\right)^{-r-1} du \\
&= \frac{1}{2\pi i}\oint_{\mathcal{C}} t^{-r-1}(1-3t)^{-1/3} dt \quad \left(\text{set } t = \frac{1-(1-2u)^{3/2}}{3}\right) \\
&= \left[t^r\right](1-3t)^{-1/3} \\
&= \frac{(3r-2)!!!}{r!},
\end{aligned}$$

so that

$$\begin{aligned}
\partial_u^r \tilde{H}(z,u)\Big|_{u=1} &= \partial_u^r \sum_{d=0}^{\infty}\prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1} \\
&\sim 2^{-(r+1)/2}(1-ez)^{-\frac{3r+1}{2}} r!(3r-2)!!!. \qquad (3.30)
\end{aligned}$$

■

Stirling's approximation in (3.9) together with singularity analysis of (3.30) gives

Lemma 3.1.7. *The r-th factorial moment of $H_{xy} + d(x,y)$ and thus also the r-th moment of $H_{xy} + d(x,y)$, are asymptotically given by*

$$\begin{aligned}
\mathbb{E}\left((H_{xy} + d(x,y))^r\right) &\sim \mathbb{E}\left((H_{xy} + d(x,y))^{\underline{r}}\right) \\
&= \frac{[z^n]\partial_u^r \tilde{H}(z,u)\Big|_{u=1}}{[z^n]\tilde{H}(z,1)}
\end{aligned}$$

$$\sim \frac{\sqrt{\pi} r!(3r-2)!!!}{2^{\frac{r}{2}} \Gamma\left(\frac{3r+1}{2}\right)} n^{\frac{3r}{2}}$$

$$= \sqrt{\pi} \frac{\left(\frac{3}{\sqrt{2}}\right)^r \Gamma(r+1)\Gamma\left(r+\frac{1}{3}\right)}{\Gamma\left(\frac{1}{3}\right)\Gamma\left(\frac{3r+1}{2}\right)} n^{\frac{3r}{2}}.$$

Now Theorem 3.1.2 follows by induction on r after noticing that

$$\begin{aligned} \mathbb{E}\left((H_{xy} + d(x,y))^r\right) &= \sum_{k=0}^{r} \binom{r}{k} \mathbb{E}\left(H_{xy}^k d(x,y)^{r-k}\right) \\ &= \mathbb{E}\left(H_{xy}^r\right) + \mathcal{O}\left(\sum_{k=0}^{r-1} \binom{r}{k} \mathbb{E}\left(H_{xy}^k\right) n^{r-k}\right). \end{aligned}$$

It only remains to show that the moment sequence C_r defines a limiting distribution with a continuous density on $[0, \infty)$. We shall make use of the fact that a random variable with absolutely integrable characteristic function has a continuous density on $\mathbb{R}$ (see [18, Chapter 13.7, Theorem 14]). We consider the moment generating function associated with the moment sequence C_r, i.e.,

$$C(t) = \sum_{r=0}^{\infty} \frac{C_r}{r!} t^r.$$

We symmetrise it as follows: let Z be a random variable whose moment generating function is $C(t)$ and whose characteristic function is thus $C(it)$. An auxiliary random variable Y that is equal to Z with probability $\frac{1}{2}$ and $-Z$ otherwise has characteristic function

$$\begin{aligned} \frac{C(it) + C(-it)}{2} &= \frac{\sqrt{\pi}}{\Gamma(\frac{1}{3})} \sum_{r=0}^{\infty} \left(\frac{3t}{\sqrt{2}}\right)^r \cdot \frac{\Gamma(r+\frac{1}{3})}{\Gamma(\frac{3r+1}{2})} \cos\left(\frac{\pi r}{2}\right) \\ &= \sum_{r=0}^{\infty} \frac{\left(\frac{2}{3}\right)^{\overline{r}}}{\left(\frac{5}{6}\right)^{\overline{r}} \left(\frac{1}{2}\right)^{\overline{r}}} \left(-\frac{2t^2}{3}\right)^r \\ &= {}_2F_2\left(1, \tfrac{2}{3}; \tfrac{5}{6}, \tfrac{1}{2}; -\tfrac{2t^2}{3}\right), \end{aligned} \tag{3.31}$$

where ${}_2F_2$ denotes a generalised hypergeometric function. This is $O(|t|^{-4/3})$ as $t \to \pm\infty$ (see [23] or [10, Chapter 16.11(ii)]), so it represents an absolutely integrable function. Hence the random variable Y has a continuous density on $\mathbb{R}$, and Z has a continuous density on $[0, \infty)$.

It is possible to obtain the joint moments of the hitting time H_{xy} and the distance $d(x, y)$ between the two randomly selected vertices. One considers the modified generating function

$$H^{(s)}(z, u) = \sum_{T,x,y} \frac{1}{|T|!} z^{|T|} d(x,y)^s u^{H_{xy}}$$

$$= \sum_{d \geq 0} d^s H_d(z, u).$$

Now the r-th derivative with respect to u yields the joint moment $\mathbb{E}(d(x,y)^s H_{xy}^r)$ since

$$\sum_{d=0}^{\infty} \sum_{k=0}^{\infty} d^s \Pr_{\mathcal{T}_n,x,y} \left(d(x,y) = d, H_{xy} = k\right) u^k = \frac{[z^n] \sum_{d=0}^{\infty} d^s H_d(z,u)}{[z^n] H(z,1)},$$

so that

$$\begin{aligned} &\mathbb{E}_{\mathcal{T}_n,x,y} \left(d(x,y)^s H_{xy}^r\right) \\ &\sim \mathbb{E}_{\mathcal{T}_n,x,y} \left(d(x,y)^s H_{xy}(H_{xy} - 1) \ldots (H_{xy} - r + 1)\right) \\ &= \frac{[z^n] \partial_u^r H^{(s)}(z,u)\Big|_{u=1}}{[z^n] H(z,1)}. \end{aligned} \tag{3.32}$$

The statement of our next theorem is as follows, where the proof mostly follows in the same way as the proof of Theorem 3.1.2.

Theorem 3.1.8. *Let x and y be two randomly selected vertices of a random labelled tree with n vertices. The joint moments of the distance $d(x,y)$ and the hitting time H_{xy} between x and y satisfy the asymptotic formula*

$$\mathbb{E}(d(x,y)^s H_{xy}^r) \sim C_{r,s} n^{3r/2+s/2},$$

where

$$C_{r,s} = \frac{\sqrt{\pi}(-3)^{r+s} r!(r+s)!}{2^{(r+3s)/2} \Gamma(\frac{3r+s+1}{2})} \sum_{k=0}^{s} (-1)^k \binom{s}{k} \binom{(2k-1)/3}{r+s}.$$

Proof. Working exactly as before we obtain a slightly modified version of (3.25):

$$\begin{aligned} &\sum_{d=0}^{\infty} d^s \partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\Bigg|_{u=1} \\ &\sim \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m \geq 1 \\ h_1+\ldots+h_m=r}} \binom{r}{h_1, h_2, \ldots, h_m} \sum_{d=m-1}^{\infty} d^{r+m+s} Y(z)^{d+1-m} \\ &\cdot \prod_{i=1}^{m} (2h_i - 3)!! \prod_{i=1}^{m} \frac{1}{h_1 + \ldots + h_i + i} (1 - ez)^{\frac{m}{2} - r} 2^{\frac{m}{2}}. \end{aligned}$$

Recall from (3.26) that

$$\sum_{d=m-1}^{\infty} d^{r+m+s} Y(z)^{d+1-m} \sim \sum_{d=0}^{\infty} d^{r+m+s} Y(z)^d \sim \frac{(r+m+s)!}{(1 - Y(z))^{r+m+s+1}}.$$

Since $1 - Y(z) \sim \sqrt{2(1-ez)}$ we have

$$\sum_{d=0}^{\infty} d^s \partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1}$$
$$\sim 2^{-\frac{r+s+1}{2}}(1-ez)^{-\frac{3r+s+1}{2}} \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m \geq 1 \\ h_1+\ldots+h_m=r}} (r+m+s)! \binom{r}{h_1, h_2, \ldots, h_m}$$
$$\cdot \prod_{i=1}^{m}(2h_i-3)!! \prod_{i=1}^{m} \frac{1}{h_1+\ldots+h_i+i}.$$

It follows as before that

$$\sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m \geq 1 \\ h_1+\ldots+h_m=r}} r!(r+m+s)! \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_i!} \prod_{i=1}^{m} \frac{1}{h_1+\ldots+h_i+i}$$
$$= \sum_{m=1}^{r} \frac{r!(r+m+s)!}{m!} [u^r] \left(\frac{3u-1+(1-2u)^{3/2}}{3u}\right)^m.$$

Now
$$\sum_{m=1}^{\infty} \frac{(r+m+s)!}{m!} x^m = (r+s)!\left((1-x)^{-r-s-1}-1\right),$$
so that

$$\sum_{d=0}^{\infty} d^s \partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1}$$
$$\sim 2^{-\frac{r+s+1}{2}}(1-ez)^{-\frac{3r+s+1}{2}} r!(r+s)! [u^r] \left(\left(\frac{1-(1-2u)^{3/2}}{3u}\right)^{-r-s-1} - 1\right).$$

Finally

$$[u^r] \left(\frac{1-(1-2u)^{3/2}}{3u}\right)^{-r-s-1}$$
$$= \frac{1}{2\pi i} \oint_{\mathcal{C}} u^{-r-1} \left(\frac{1-(1-2u)^{3/2}}{3u}\right)^{-r-s-1} du$$
$$= \frac{1}{2\pi i} \oint_{\mathcal{C}} \left(\frac{3u}{1-(1-2u)^{3/2}}\right)^{s} \left(\frac{1-(1-2u)^{3/2}}{3}\right)^{-r-1} du$$
$$= \frac{1}{2\pi i} \oint_{\mathcal{C}} t^{-r-s-1} \left(\frac{1}{2}\left(1-(1-3t)^{2/3}\right)\right)^{s} (1-3t)^{-1/3} dt \ \left(\text{set } t = \frac{1-(1-2u)^{3/2}}{3}\right)$$

$$= \left(\frac{1}{2}\right)^s [t^{r+s}] \left(1-(1-3t)^{2/3}\right)^s (1-3t)^{-1/3}$$
$$= \left(\frac{1}{2}\right)^s [t^{r+s}] \sum_{k=0}^{s}(-1)^k \binom{s}{k} (1-3t)^{\frac{2k-1}{3}}$$
$$= \left(\frac{1}{2}\right)^s [t^{r+s}] \sum_{k=0}^{s}(-1)^k \binom{s}{k} \sum_{n=0}^{\infty} \binom{\frac{2k-1}{3}}{n} (-3t)^n$$
$$= \left(\frac{1}{2}\right)^s \sum_{k=0}^{s}(-1)^k \binom{s}{k} (-3)^{r+s} \binom{\frac{2k-1}{3}}{r+s},$$

so that

$$\sum_{d=0}^{\infty} d^s \partial_u^r \prod_{k=0}^{d} Y\left(zu^{2k}\right)\Big|_{u=1}$$
$$\sim 2^{-\frac{r+3s+1}{2}} (1-ez)^{-\frac{3r+s+1}{2}} r!(r+s)!(-3)^{r+s} \sum_{k=0}^{s}(-1)^k \binom{s}{k} \binom{\frac{2k-1}{3}}{r+s}. \tag{3.33}$$

Singularity analysis of (3.33) yields

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(d(x,y)^s H_{xy}^r\right)$$
$$\sim \frac{[z^n]\sum_{d=0}^{\infty} d^s \partial_u^r H_d(z,u)\Big|_{u=1}}{[z^n]H(z,1)}$$
$$\sim \frac{\sqrt{\pi} r!(r+s)!(-3)^{r+s}\sum_{k=0}^{s}(-1)^k \binom{s}{k}\binom{\frac{2k-1}{3}}{r+s}}{2^{\frac{r+3s}{2}}\Gamma\left(\frac{3r+s+1}{2}\right)} n^{\frac{3r+s}{2}}.$$

■

For the special case of $r = s = 1$ we have

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(d(x,y)H_{xy}\right)$$
$$\sim \frac{\sqrt{\pi}2!(-3)^2\sum_{k=0}^{1}(-1)^k \binom{1}{k}\binom{\frac{2k-1}{3}}{2}}{2^{\frac{4}{2}}\Gamma\left(\frac{5}{2}\right)} n^2$$
$$= \frac{9\sqrt{\pi}\,(1/3)}{(2)\left(\frac{3}{4}\sqrt{\pi}\right)} n^2$$
$$= 2n^2.$$

As before

$$\mathbb{E}\left(H_{xy}\right) \sim \sqrt{\frac{\pi}{2}} n^{\frac{3}{2}} \quad \text{and} \quad \mathrm{Var}\left(H_{xy}\right) \sim \left(\frac{32}{15} - \frac{\pi}{2}\right) n^3$$

whereas setting $r = 0$ gives

$$\mathbb{E}\left(d(x,y)\right) \sim \sqrt{\frac{\pi}{2}} n^{\frac{1}{2}} \quad \text{and} \quad \mathrm{Var}\left(d(x,y)\right) \sim \left(2 - \frac{\pi}{2}\right) n$$

so that

$$\mathrm{cov}\left(d(x,y), H_{xy}\right) \sim \left(2 - \frac{\pi}{2}\right) n^2$$

and

$$\begin{aligned} \mathrm{corr}\left(d(x,y), H_{xy}\right) &\sim \frac{\left(2 - \frac{\pi}{2}\right) n^2}{\sqrt{(2 - \frac{\pi}{2})(\frac{32}{15} - \frac{\pi}{2}) n^4}} \\ &\approx 0.873486. \end{aligned} \tag{3.34}$$

We conclude this section by finding the expected hitting time variance, defined in Chapter 2. Recall equation (2.12) which states that

$$\mathrm{Var}_{k,l} = 8\left(\binom{|T|}{2} + |T| D_T(k) - D_T(k) - W(T)\right)$$

for adjacent vertices k, l, where T is the tree rooted at k after removing edge kl. Let

$$P(T) = \left(\binom{|T|}{2} + |T| D_T(k) - D_T(k) - W(T)\right)$$

and

$$Q(T) = |T| D_T(k) - W(T).$$

Furthermore let $d_T(k, i \wedge j)$ be the distance from the root k to the common ancestor, furthest away from the root, of vertices i and j. Notice

$$\begin{aligned} Q(T) &= \sum_{i \in T} \sum_{j \in T} \frac{1}{2}\left(d(k,i) + d(k,j) - d(i,j)\right) \\ &= \sum_{i \in T} \sum_{j \in T} d(k, i \wedge j), \end{aligned}$$

since $d(k,i) + d(k,j) = d(i,j) + 2d(k, i \wedge j)$. Hence $Q(T)$ satisfies the recursion

$$Q(T) = \sum_{\tau \prec T} \left(Q(\tau) + |\tau|^2\right),$$

with $Q(T) = 0$ if $|T| = 1$, where the sum is over all the root subtrees of T. Now

$$P(T) = \binom{|T|}{2} + Q(T) - D_T(k),$$

where

$$D_T(k) = \sum_{\tau \prec T} (D_\tau(k_r) + |\tau|)$$

and k_r denotes the root of subtree τ. It follows that

$$\begin{aligned}
P(T) &= \sum_{\tau \prec T} \left(Q(\tau) - D_\tau(k_r) + |\tau|^2 - |\tau|\right) + \binom{|T|}{2} \\
&= \sum_{\tau \prec T} \left(Q(\tau) - D_\tau(k_r) + 2\binom{|\tau|}{2}\right) + \binom{|T|}{2} \\
&= \sum_{\tau \prec T} \left(P(\tau) + \binom{|\tau|+1}{2}\right) - \sum_{\tau \prec T} |\tau| + \binom{|T|}{2} \\
&= \sum_{\tau \prec T} \left(P(\tau) + \binom{|\tau|+1}{2}\right) + \binom{|T|-1}{2}.
\end{aligned}$$

Now $R(T) := P(T) + \binom{|T|+1}{2}$ satisfies the recursion

$$\begin{aligned}
R(T) &= \sum_{\tau \prec T} \left(R(\tau) - \binom{|\tau|+1}{2} + \binom{|\tau|+1}{2}\right) + \binom{|T|-1}{2} + \binom{|T|+1}{2} \\
&= \sum_{\tau \prec T} R(\tau) + |T|^2 - |T| + 1.
\end{aligned}$$

More generally for any two vertices x and y in a tree we let $x = w_0, w_1, w_2, \ldots, w_d = y$ be the (xy)-path (recall Figure 3.1). Since the random times to move along the edges are independent we have

$$\mathrm{Var}_{x,y} = \sum_{j=0}^{d-1} \mathrm{Var}_{w_j, w_{j+1}} = 8 \sum_{j=0}^{d-1} \left(R(T_{w_j:w_{j+1}}) - \binom{|T_{w_j:w_{j+1}}|+1}{2}\right).$$

Adding the recursions for R along the path from x to y gives

$$\sum_{j=0}^{d-1} R(T_{w_j:w_{j+1}}) = \sum_{i=0}^{d-1} (d-i) \left(\sum_{\tau \prec T_i} R(\tau) + \left(\sum_{j=0}^{i} |T_j|\right)^2 - \sum_{j=0}^{i} |T_j| + 1\right).$$

Hence

$$\begin{aligned}
\mathrm{Var}_{x,y} = 8 \sum_{i=0}^{d-1} (d-i) \left(\sum_{\tau \prec T_i} R(\tau) + 1\right) &+ 8 \sum_{i=0}^{d-1} \frac{2(d-i)-1}{2} \left(\sum_{j=0}^{i} |T_j|\right)^2 \\
&- 8 \sum_{i=0}^{d-1} \frac{2(d-i)+1}{2} \sum_{j=0}^{i} |T_j|. \qquad (3.35)
\end{aligned}$$

We are now ready to prove the following theorem:

Theorem 3.1.9. *Let x and y be two randomly selected vertices of a uniformly random labelled tree with n vertices. The expected hitting time variance between x and y satisfies the asymptotic equivalence*

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(Var_{x,y}\right) \sim \frac{32}{15}n^3.$$

Proof. We will evaluate the three main terms in (3.35) separately and sum the asymptotically relevant contributions. We know that (if $d=0$, $\mathrm{Var}_{x,y}$ vanishes)

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(\mathrm{Var}_{x,y}\right) = \frac{[z^n]\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d}\mathrm{Var}_{x,y}\prod_{j=0}^{d}\frac{z^{|T_j|}}{|T_j!|}}{n^n/n!}. \tag{3.36}$$

First notice

$$\begin{aligned}
R(z) &:= \sum_{T_i} R(T_i)\frac{z^{|T_i|}}{|T_i|!} \\
&= \sum_{T_i}\sum_{\tau\prec T_i} R(\tau)\frac{z^{|T_i|}}{|T_i|!} + z^2Y''(z) + Y(z) \\
&= z\left(\sum_{k=1}^{\infty} k\frac{Y(x)^{k-1}}{k!}\right)R(z) + z^2Y''(z) + Y(z) \\
&= zR(z)e^{Y(z)} + z^2Y''(z) + Y(z),
\end{aligned}$$

implying

$$R(z) = \frac{z^2Y''(z)+Y(z)}{1-Y(z)}.$$

Now, where we factor out the sum

$$\sum_{T_d}\frac{z^{|T_d|}}{|T_d|!}$$

in (3.37), we have

$$\begin{aligned}
&8\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d}\sum_{i=0}^{d-1}(d-i)\left(\sum_{\tau\prec T_i}R(\tau)+1\right)\prod_{j=0}^{d}\frac{z^{|T_j|}}{|T_j!|} \\
&= 8Y(z)\sum_{d=1}^{\infty}\left(\sum_{i=0}^{d-1}(d-i)Y(z)^{d-1}\left(\frac{z^2Y''(z)+Y(z)}{1-Y(z)} - z^2Y''(z) - Y(z)\right)\right. \\
&\qquad\left. + \frac{d(d+1)}{2}Y(z)^d\right)
\end{aligned} \tag{3.37}$$

$$
\begin{aligned}
&= 8Y(z)\sum_{d=1}^{\infty}\frac{d(d+1)}{2}\left(Y(z)^{d-1}\left(\frac{z^2Y''(z)+Y(z)}{1-Y(z)}-z^2Y''(z)-Y(z)\right)+Y(z)^d\right)\\
&\sim \frac{4z^2Y''(z)}{1-Y(z)}\sum_{d=1}^{\infty}d^2Y(z)^d\\
&\sim \frac{4}{2\sqrt{2}}(1-ez)^{-3/2}\frac{2!}{(1-Y(z))^4}\\
&\sim \frac{1}{\sqrt{2}}(1-ez)^{-7/2}, \qquad (3.38)
\end{aligned}
$$

whereas

$$
\begin{aligned}
&8\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d}\sum_{i=0}^{d-1}\frac{2(d-i)-1}{2}\left(\sum_{j=0}^{i}|T_j|\right)^2\prod_{j=0}^{d}\frac{z^{|T_j|}}{|T_j!|}\\
&= 8Y(z)\sum_{d=1}^{\infty}\sum_{i=0}^{d-1}\frac{2(d-i)-1}{2}\left(z(Y'(z)+zY''(z))Y(z)^{d-1}(i+1)+z^2Y'(z)^2Y(z)^{d-2}i(i+1)\right)\\
&= 4z(Y'(z)+zY''(z))\sum_{d=1}^{\infty}Y(z)^d\left(2\sum_{i=0}^{d-1}((d-1)i-i^2+d)-\sum_{i=0}^{d-1}(i+1)\right)\\
&\quad + 4z^2Y'(z)^2\sum_{d=1}^{\infty}Y(z)^{d-1}\left(2\sum_{i=0}^{d-1}((d-1)i^2-i^3+di)-\sum_{i=0}^{d-1}i(i+1)\right)\\
&= 4z(Y'(z)+zY''(z))\sum_{d=1}^{\infty}Y(z)^d\frac{2d^3+3d^2+d}{6}\\
&\quad + 4z^2Y'(z)^2\sum_{d=1}^{\infty}Y(z)^{d-1}\frac{d^4-d^2}{6}\\
&\sim \frac{4}{3}z^2Y''(z)\sum_{d=1}^{\infty}Y(z)^d d^3+\frac{2}{3}z^2Y'(z)^2\sum_{d=1}^{\infty}Y(z)^d d^4\\
&\sim \frac{4}{3}\cdot\frac{1}{2\sqrt{2}}(1-ez)^{-3/2}\frac{3!}{(1-Y(z))^4}+\frac{2}{3}\cdot\frac{1}{2(1-ez)}\frac{4!}{(1-Y(z))^5}\\
&\sim \frac{3}{\sqrt{2}}(1-ez)^{-7/2} \qquad (3.39)
\end{aligned}
$$

and finally

$$
\begin{aligned}
&-8\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d}\sum_{i=0}^{d-1}\frac{2(d-i)+1}{2}\sum_{j=0}^{i}|T_j|\prod_{j=0}^{d}\frac{z^{|T_j|}}{|T_j!|}\\
&= -4Y(z)zY'(z)\sum_{d=1}^{\infty}Y(z)^{d-1}\sum_{i=0}^{d-1}(2(d-i)+1)(i+1)
\end{aligned}
$$

$$= -4zY'(z)\sum_{d=1}^{\infty}\frac{2d^3+9d^2+7d}{6}Y(z)^d$$
$$\sim \sqrt{2}(1-ez)^{-5/2}. \tag{3.40}$$

Singularity analysis of the sum of (3.38), (3.39) and (3.40) gives

$$[z^n]\sum_{d=1}^{\infty}\sum_{T_0,\ldots,T_d} Var_{x,y}\prod_{j=0}^{d}\frac{z^{|T_j|}}{|T_j!|} \sim \frac{4}{\sqrt{2}}\frac{e^n}{\Gamma(7/2)}n^{5/2}.$$

Since

$$\frac{n!}{n^n} \sim \sqrt{2\pi n}e^{-n}$$

the result follows from (3.36). ∎

3.2 Weighted labelled trees

Recall that $\mathcal{T}$ denotes the class of labelled trees, where the number of trees with n vertices, $|\mathcal{T}_n|$, is equal to n^{n-2}. In this section, we describe how Theorem 3.1.2 is generalised to labelled trees with vertex weights based on degrees. The resulting trees are similar to simply generated trees (which are in turn essentially equivalent to Galton-Watson trees), see [15] for a general reference. The main difference is that our trees do not have a root. We assign a weight

$$w(T) = \prod_{v\in T} w(d(v))$$

to $T \in \mathcal{T}$ and the probability distribution

$$\mathrm{Pr}_{\mathcal{T}_n}(T) = \frac{w(T)}{\sum_{T\in\mathcal{T}_n} w(T)} \tag{3.41}$$

to $T \in \mathcal{T}_n$.

Example 4. Consider the class of labelled trees with degrees bounded by a constant D. For $v \in T$ we have

$$w(d(v)) = \begin{cases} 1 & \text{if } d(v) \le D \\ 0 & \text{if } d(v) > D, \end{cases}$$

so that $W(T) = 0$ for any labelled tree with a vertex with degree greater than D.

Let

$$HW(z,u)=\sum_{T,x,y}\frac{w(T)}{|T|!}z^{|T|}u^{H_{xy}}$$

be the bivariate exponential generating function where u marks the hitting time between vertices x and y, chosen uniformly at random and z marks the size of the tree. Now $HW(z,u)=\sum_{d=0}^{\infty}HW_d(z,u)$, where

$$HW_d(z,u)=\sum_{\substack{T,x,y\\ d(x,y)=d}}\frac{w(T)}{|T|!}z^{|T|}u^{H_{xy}}.$$

Note that

$$HW_0(z,u)=\sum_{T_0}\frac{w(T_0)}{|T_0|!}z^{|T_0|},$$

where the sum is over all rooted labelled trees. For $d\geq 1$ the sum over all weighted labelled trees decomposes into a sum over rooted weighted labelled trees, similar to (3.1):

$$\begin{aligned}
HW_d(z,u)&=\sum_{T_0,T_1,\dots,T_d}\frac{u^{-d}}{n!}\binom{n}{|T_0|,|T_1|,\dots,|T_d|}\prod_{j=0}^{d}w(T_j)z^{|T_j|}u^{2(d-j)|T_j|}\\
&=u^{-d}\prod_{j=0}^{d}\sum_{T_j}\frac{w(T_j)z^{|T_j|}u^{2(d-j)|T_j|}}{|T_j|!}\\
&=u^{-d}\prod_{k=0}^{d}\sum_{T_k}\frac{w(T_k)z^{|T_k|}u^{2k|T_k|}}{|T_k|!}\\
&=u^{-d}Y(z)Y\left(zu^{2d}\right)\prod_{k=1}^{d-1}\tilde{Y}\left(zu^{2k}\right),
\end{aligned}$$

where all the vertices, in the trees T_0 and T_d rooted at the endpoints of the path, have weight equal to $w(d^+(v)+1)$, where $d^+(v)$ denotes the out-degree of a vertex. On the other hand the roots of the trees $T_1,\dots,T_{d-1}$ have weight equal to $w(d^+(v)+2)$, whereas the remaining vertices in these trees have weight equal to $w(d^+(v)+1)$. We slightly abuse notation in this section, since

$$w(T_i)=w(d(r)+1)\prod_{v\neq r}w(d(v))\;\text{ if }\;i=0,d$$

and

$$w(T_i)=w(d(r)+2)\prod_{v\neq r}w(d(v))$$

if $i = 1, \dots, d-1$, where r is the root of T_i. It then follows from the multiplicative structure of the weights, together with the recursive structure of the tree that

$$\begin{aligned} Y(z) &= z \sum_{d^+(v)=0}^{\infty} \frac{w(d^+(v)+1)}{d^+(v)!} Y(z)^{d^+(v)} \\ &= z\phi(Y(z)) \end{aligned}$$

and

$$\begin{aligned} \tilde{Y}(z) &= z \sum_{d^+(v)=0}^{\infty} \frac{w(d^+(v)+2)}{d^+(v)!} Y(z)^{d^+(v)} \\ &= z\tilde{\phi}(Y(z)). \end{aligned}$$

We suppose ϕ is aperiodic and satisfies the conditions in Proposition 1.3.3. Furthermore

$$\tilde{Y}(z) = z\tilde{\phi}(Y(z)) = z\phi'(Y(z)),$$

since

$$\phi'(Y(z)) = \sum_{d-1=0}^{\infty} \frac{w(d+1)}{(d-1)!} Y(z)^{d-1}.$$

On the other hand

$$Y'(z) = \phi(Y(z)) + z\phi'(Y(z))Y'(z) = \frac{Y(z)}{z} + z\phi'(Y(z))Y'(z)$$

so that

$$\tilde{Y}(z) = 1 - \frac{Y(z)}{zY'(z)}.$$

We know from Theorem 1.3.4 that $Y(z) \sim \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\sqrt{1 - z\phi'(\tau)}$, where τ satisfies the conditions in Proposition 1.3.3. The aim of this section is to prove the following generalisation of Theorem 3.1.2:

Theorem 3.2.1. *Let x and y be two vertices that are selected uniformly at random of a random labelled tree with n vertices, chosen according to the probability distribution in (3.41). The r-th moment of the hitting time between x and y satisfies the following asymptotic formula:*

$$\mathbb{E}(H_{xy}^r) \sim C_r \left(\frac{1}{\tau} \sqrt{\frac{\phi(\tau)}{\phi''(\tau)}} \right)^r n^{3r/2},$$

where

$$C_r = \frac{\sqrt{\pi}}{\Gamma(\frac{1}{3})} \cdot \left(\frac{3}{\sqrt{2}} \right)^r \cdot \frac{\Gamma(r+1)\Gamma(r+\frac{1}{3})}{\Gamma(\frac{3r+1}{2})}$$

and

$$\phi(t) = \sum_{d=0}^{\infty} \frac{w(d+1)}{d!} t^d.$$

Consequently, the normalised random variable

$$\left(\frac{1}{\tau} \sqrt{\frac{\phi(\tau)}{\phi''(\tau)}} \right)^{-1} n^{-3/2} H_{xy}$$

converges weakly to a limit law that is characterised by the moment sequence C_r*. This is the same limit law as in Theorem 3.1.2.*

We start by finding the first moment.

Proposition 3.2.2. *Let x and y be two randomly selected vertices of a weighted random labelled tree with n vertices. The expected hitting time between x and y satisfies the asymptotic equivalence*

$$\mathbb{E}_{\mathcal{T}_n,x,y}\left(H_{xy}\right) \sim \sqrt{\frac{\pi}{2}} \left(\frac{1}{\tau} \sqrt{\frac{\phi(\tau)}{\phi''(\tau)}} \right) n^{3/2}.$$

Proof. We start by recalling

$$\mathbb{E}\left(H_{xy}\right) = \frac{[z^n]\partial_u HW(z,u)\Big|_{u=1}}{[z^n]HW(z,1)}. \tag{3.42}$$

The term $HW_0(z,u)$ is asymptotically irrelevant so that

$$\begin{aligned} HW(z,1) &\sim \sum_{d=1}^{\infty} HW_d(z,1) \\ &\sim \sum_{d=1}^{\infty} Y(z)^2 \left(1 - \frac{Y(z)}{zY'(z)}\right)^{d-1} \\ &= Y(z)^2 \sum_{d=0}^{\infty} \left(1 - \frac{Y(z)}{zY'(z)}\right)^{d} \\ &= Y(z)^2 \frac{zY'(z)}{Y(z)} \\ &= zY'(z)Y(z) \\ &\sim \tau \sqrt{\frac{\phi(\tau)}{2\phi''(\tau)}} \left(1 - z\phi'(\tau)\right)^{-1/2}. \end{aligned} \tag{3.43}$$

Singularity analysis of (3.43) yields

$$[z^n]HW(z,1) \sim \tau \sqrt{\frac{\phi(\tau)}{2\phi''(\tau)\pi}} \phi'(\tau)^n n^{-1/2}. \tag{3.44}$$

We proceed to find $[z^n]\partial_u HW(z,u)\Big|_{u=1}$. Note that

$$\partial_u HW(z,u)\Big|_{u=1} = \partial_u \sum_{d=0}^{\infty} u^{-d} Y(z) Y\left(zu^{2d}\right) \prod_{k=1}^{d-1} \left(1 - \frac{Y(zu^{2k})}{zu^{2k} Y'(zu^{2k})}\right)\Bigg|_{u=1}, \tag{3.45}$$

since the partial derivative of the first term in the sum (the $d = 0$ term) vanishes.

Now

$$\begin{aligned}
&\partial_u HW(z,u)\Big|_{u=1} \\
&= \partial_u \sum_{d=0}^{\infty} u^{-d} Y(z) Y\left(zu^{2d}\right) \prod_{k=1}^{d-1} \left(1 - \frac{Y(zu^{2k})}{zu^{2k} Y'(zu^{2k})}\right)\Bigg|_{u=1} \\
&= \sum_{d=0}^{\infty} \left(-dY(z)^2 + 2dzY(z)Y'(z)\right) \left(1 - \frac{Y(z)}{zY'(z)}\right)^{d-1} \\
&\quad + \sum_{d=0}^{\infty} Y(z)^2 \left(1 - \frac{Y(z)}{zY'(z)}\right)^{d-1} \sum_{l=1}^{d-1} \frac{-2lz^2Y'(z)^2 + Y(z)(2lzY'(z) + 2lz^2Y''(z))}{zY'(z)\left(zY'(z) - Y(z)\right)} \\
&= \sum_{d=0}^{\infty} \left(1 - \frac{Y(z)}{zY'(z)}\right)^{d-1} \Big[-dY(z)^2 + 2dzY(z)Y'(z) - \tfrac{d(d-1)z^2Y'(z)^2Y(z)^2}{zY'(z)(zY'(z)-Y(z))} \\
&\quad + \tfrac{Y(z)^2}{zY'(z)(zY'(z)-Y(z))} \left(d(d-1)zY(z)Y'(z) + d(d-1)z^2Y(z)Y''(z)\right) \Big] \\
&\sim \left(-Y(z)^2 + 2zY(z)Y'(z)\right) \frac{zY'(z)}{zY'(z) - Y(z)} \left(1 - \left(1 - \frac{Y(z)}{zY'(z)}\right)\right)^{-2} \\
&\quad - \frac{(zY'(z))^2Y(z)^2}{(zY'(z) - Y(z))^2} \left[2\left(1 - \left(1 - \tfrac{Y(z)}{zY'(z)}\right)\right)^{-3} - \left(1 - \left(1 - \tfrac{Y(z)}{zY'(z)}\right)\right)^{-2}\right] \\
&\quad + \frac{z^2Y'(z)^2Y(z)^3}{zY'(z)\left(zY'(z) - Y(z)\right)^2} \left[2\left(1 - \left(1 - \tfrac{Y(z)}{zY'(z)}\right)\right)^{-3} - \left(1 - \left(1 - \tfrac{Y(z)}{zY'(z)}\right)\right)^{-2}\right] \\
&\quad + \frac{z^2Y''(z)Y(z)^3}{\left(zY'(z) - Y(z)\right)^2} \left[2\left(1 - \left(1 - \tfrac{Y(z)}{zY'(z)}\right)\right)^{-3} - \left(1 - \left(1 - \tfrac{Y(z)}{zY'(z)}\right)\right)^{-2}\right] \\
&\sim -\frac{\left(zY'(z)\right)^3}{zY'(z) - Y(z)} + \frac{2\left(zY'(z)\right)^4}{Y(z)\left(zY'(z) - Y(z)\right)} - \frac{2\left(zY'(z)\right)^5}{Y(z)\left(zY'(z) - Y(z)\right)^2} \\
&\quad + \frac{2\left(zY'(z)\right)^4}{\left(zY'(z) - Y(z)\right)^2} + \frac{2z^2Y''(z)\left(zY'(z)\right)^3}{\left(zY'(z) - Y(z)\right)^2} \\
&\sim -\left(zY'(z)\right)^2 + \frac{2}{\tau}\left(zY'(z)\right)^3 - \frac{2}{\tau}\left(zY'(z)\right)^3 + 2\left(zY'(z)\right)^2 + 2z^2Y''(z)\left(zY'(z)\right) \\
&\sim 2z^2Y''(z)\left(zY'(z)\right)
\end{aligned}$$

$$\sim 2z^2 \left(\frac{1}{4}\phi'(\tau)^2\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\left(1 - z\phi'(\tau)\right)^{-3/2}\right)\left(\frac{1}{2}\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\left(1 - z\phi'(\tau)\right)^{-1/2}\right)$$
$$\sim \frac{\phi(\tau)}{2\phi''(\tau)}\left(1 - z\phi'(\tau)\right)^{-2}. \tag{3.46}$$

Singularity analysis of (3.46) yields

$$[z^n]\partial_u HW(z,u)\Big|_{u=1} \sim \frac{\phi(\tau)}{2\phi''(\tau)}\phi'(\tau)^n n. \tag{3.47}$$

It follows from (3.42),(3.44) and (3.47) that

$$\mathbb{E}\left(H_{xy}\right) \sim \frac{1}{\tau}\cdot\sqrt{\frac{\pi\phi(\tau)}{2\phi''(\tau)}}n^{3/2}. \tag{3.48}$$

■

We proceed to find the higher moments. It follows as in (3.45) that

$$\partial_u^r HW(z,u)\Big|_{u=1} = \partial_u^r \sum_{d=0}^{\infty} u^{-d} Y(z) Y(zu^{2d}) \prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right)\Big|_{u=1}.$$

Applying the general Leibniz rule yields

$$\partial_u^r HW(z,u)\Big|_{u=1} = \sum_{d=0}^{\infty} Y(z) \sum_{l=0}^{r} \binom{r}{l} \partial_u^l \left(\prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right)\right) \partial_u^{r-l} Y\left(zu^{2d}\right)\Big|_{u=1}. \tag{3.49}$$

We require the following proposition to simplify (3.49).

Proposition 3.2.3. *The r-th partial derivative of the bivariate exponential generating function $HW(z,u)$, evaluated at $u = 1$, satisfies the asymptotic equivalence*

$$\partial_u^r HW(z,u)\Big|_{u=1} \sim \sum_{d=0}^{\infty} Y(z) \partial_u^r \left(\prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right)\right) Y\left(zu^{2d}\right)\Big|_{u=1}.$$

Proof. If $l = 0$ then

$$\sum_{d=0}^{\infty} Y(z) \prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right) \partial_u^r Y\left(zu^{2d}\right)\Big|_{u=1}$$
$$\sim \sum_{d=0}^{\infty} Y(z) \left(\tilde{Y}(z)\right)^{d-1} (2dz)^r Y^{(r)}(z)$$

$$\sim \sum_{d=0}^{\infty} \tau \tilde{Y}(z)^d d^r \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}(2r-3)!!\,(1-z\phi'(\tau))^{1/2-r} \tag{3.50}$$

$$\sim \tau \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}} \frac{r!}{(1-\tilde{Y}(z))^{r+1}}(2r-3)!!\,(1-z\phi'(\tau))^{1/2-r} \tag{3.51}$$

$$\sim \frac{\tau}{(2\tau)^{r+1}} \left(\frac{2\phi(\tau)}{\phi''(\tau)}\right)^{\frac{r+2}{2}} r!(2r-3)!!\,(1-z\phi'(\tau))^{-3r/2}, \tag{3.52}$$

where (3.50) and (3.51) follows from (3.21) and (3.26) respectively, since

$$\frac{1}{1-\tilde{Y}(z)} \sim \frac{1}{2\tau}\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\,(1-z\phi'(\tau))^{-1/2}.$$

If $l = 1, \ldots, r-1$, then as in the derivation of (3.20) we have,

$$\sum_{d=0}^{\infty} Y(z) \sum_{l=0}^{r} \binom{r}{l} \partial_u^l \left(\prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right) \right) \partial_u^{r-l} Y\left(zu^{2d}\right) \Big|_{u=1}$$

$$\sim \sum_{d=0}^{\infty} Y(z) \sum_{l=0}^{r} \binom{r}{l} \sum_{m=1}^{l} \sum_{h_1+\ldots+h_m=l} \binom{l}{h_1,\ldots,h_m} \sum_{1\le j_1<j_2<\ldots<j_m\le d-1} \tilde{Y}(z)^{d-1-m}$$

$$\cdot \prod_{i=1}^{m} \left[\partial_u^{h_i} \tilde{Y}\left(zu^{2j_i}\right)\right] \Big|_{u=1} d^{r-l} \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}(2(r-l)-3)!!\,(1-z\phi'(\tau))^{1/2-(r-l)}.$$

It follows as in (3.21) that

$$\partial_u^{h_i} \tilde{Y}\left(zu^{2j_i}\right) \Big|_{u=1} \sim (2j_i z)^{h_i} \tau \sqrt{\frac{\phi''(\tau)}{\phi(\tau)}} \phi'(\tau)^{h_i} (2h_i-3)!!\,(2(1-z\phi'(\tau)))^{1/2-h_i}$$

whereas application of the binomial theorem, as in (3.23), gives

$$\begin{aligned}
&\sum_{1\le j_1<j_2<\ldots<j_m\le d-1} \prod_{i=1}^{m} {j_i}^{h_i} \\
&= \frac{(d-1)^{h_1+h_2+\ldots+h_m+m}}{(h_1+1)(h_1+h_2+2)\ldots(h_1+h_2+\ldots+h_m+m)} \\
&+ \mathcal{O}\left((d-1)^{h_1+h_2+\ldots+h_m+m-1}\right) \\
&= \frac{d^{h_1+\ldots+h_m+m}}{(h_1+1)\ldots(h_1+h_2+\ldots+h_m+m)} + \mathcal{O}\left(d^{h_1+\ldots+h_m+m-1}\right).
\end{aligned}$$

Hence

$$\sum_{d=0}^{\infty} Y(z) \sum_{l=0}^{r} \binom{r}{l} \partial_u^l \left(\prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right) \right) \partial_u^{r-l} Y\left(zu^{2d}\right) \Big|_{u=1}$$

$$\sim \sum_{d=0}^{\infty} Y(z) \sum_{l=0}^{r} \binom{r}{l} \sum_{m=1}^{l} \sum_{h_1+\ldots+h_m=l} \binom{l}{h_1,\ldots,h_m} \tilde{Y}(z)^{d-1-m} \left(1-z\phi'(\tau)\right)^{\frac{m}{2}-l}$$
$$\cdot 2^{\frac{m}{2}} \left(\tau\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\right)^m \prod_{i=1}^{m} (2h_i-3)!! d^{l+m} \prod_{i=1}^{m} \frac{1}{h_1+\ldots+h_i+i} d^{r-l} \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}$$
$$\cdot (2(r-l)-3)!! \left(1-z\phi'(\tau)\right)^{1/2-(r-l)}$$
$$\sim \sum_{d=0}^{\infty} \tau \sum_{l=0}^{r} \binom{r}{l} \sum_{m=1}^{l} \sum_{h_1+\ldots+h_m=l} \binom{l}{h_1,\ldots,h_m} \tilde{Y}(z)^{d-1-m} \left(1-z\phi'(\tau)\right)^{\frac{m+1}{2}-r} 2^{m/2}$$
$$\cdot \tau \left(\tau\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\right)^{m-1} \prod_{i=1}^{m} (2h_i-3)!! d^{m+r} \prod_{i=1}^{m} \frac{1}{h_1+\ldots+h_i+i} \sqrt{2}(2(r-l)-3)!!$$
$$\sim \tau^2 \sum_{l=0}^{r} \binom{r}{l} \sum_{m=1}^{l} \sum_{h_1+\ldots+h_m=l} \binom{l}{h_1,\ldots,h_m} \frac{(r+m)!}{(1-\tilde{Y}(z))^{r+m+1}} \left(1-z\phi'(\tau)\right)^{\frac{m+1}{2}-r}$$
$$\cdot 2^{\frac{m+1}{2}} \left(\tau\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\right)^{m-1} \prod_{i=1}^{m} (2h_i-3)!! \prod_{i=1}^{m} \frac{1}{h_1+\ldots+h_i+i} (2(r-l)-3)!!$$
$$\sim \tau\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}} \sum_{l=0}^{r} \binom{r}{l} \left(1-z\phi'(\tau)\right)^{-3r/2} 2^{-r/2} \sum_{m=1}^{l} \sum_{h_1+\ldots+h_m=l} \binom{l}{h_1,\ldots,h_m}$$
$$\cdot (r+m)! \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+1} \prod_{i=1}^{m} (2h_i-3)!! \prod_{i=1}^{m} \frac{1}{h_1+\ldots+h_i+i} (2(r-l)-3)!!. \tag{3.53}$$

Finally if $l=r$ then,

$$\sum_{d=0}^{\infty} Y(z)\partial_u^r \left(\prod_{k=1}^{d-1} \tilde{Y}\left(zu^{2k}\right)\right) Y\left(zu^{2d}\right)\Big|_{u=1}$$
$$\sim \sum_{d=0}^{\infty} Y(z)^2 \sum_{m=1}^{r} \sum_{h_1+\ldots+h_m=r} \binom{r}{h_1,\ldots,h_m} \sum_{1\le j_1<\ldots<j_m\le d-1} \tilde{Y}(z)^{d-1-m}$$
$$\cdot \left(1-z\phi'(\tau)\right)^{\frac{m}{2}-r} 2^{\frac{m}{2}} \left(\tau\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\right)^m \prod_{i=1}^{m} j_i^{h_i}(2h_i-3)!!$$
$$\sim \sum_{d=0}^{\infty} \tau^2 \sum_{m=1}^{r} \sum_{h_1+\ldots+h_m=r} \binom{r}{h_1,\ldots,h_m} d^{r+m} \tilde{Y}(z)^{d-1-m} \left(1-z\phi'(\tau)\right)^{\frac{m}{2}-r}$$
$$\cdot 2^{\frac{m}{2}} \left(\tau\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\right)^m \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i}$$

$$\sim \tau^2 \sum_{m=1}^{r} \sum_{h_1+\ldots+h_m=r} \binom{r}{h_1,\ldots,h_m} (r+m)! \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+1} (1-z\phi'(\tau))^{-\frac{3r+1}{2}}$$

$$\cdot\, 2^{-\frac{r+1}{2}} \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i}$$

$$= \tau^2 (1-z\phi'(\tau))^{-\frac{3r+1}{2}} 2^{-\frac{r+1}{2}} \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+1} \sum_{m=1}^{r} \sum_{h_1+\ldots+h_m=r} \binom{r}{h_1,\ldots,h_m}$$

$$\cdot\, (r+m)! \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i}. \tag{3.54}$$

The proof of this proposition follows after adding together (3.52), (3.53) and (3.54). ∎

Now

$$\partial_u^r HW(z,u)\Big|_{u=1} \sim \tau^2 (1-z\phi'(\tau))^{-\frac{3r+1}{2}} 2^{-\frac{r+1}{2}} \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+1}$$
$$\cdot \sum_{m=1}^{r} \sum_{h_1+\ldots+h_m=r} \binom{r}{h_1,\ldots,h_m} (r+m)! \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i},$$

where as in (3.29) and (3.30),

$$\sum_{m=1}^{r} \sum_{h_1+\ldots+h_m=r} r!(r+m)! \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_i!} \prod_{i=1}^{m} \frac{1}{h_1+\ldots h_i+i}$$
$$= \sum_{m=1}^{r} \frac{r!(r+m)!}{m!} [u^r] \left(\frac{3u-1+(1-2u)^{3/2}}{3u}\right)^m$$
$$= r!^2 \frac{(3r-2)!!}{r!}.$$

Hence

$$\partial_u^r HW(z,u)\Big|_{u=1} \sim \tau^2 \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+1} (1-z\phi'(\tau))^{-\frac{3r+1}{2}} 2^{-\frac{r+1}{2}} r!(3r-2)!!. \tag{3.55}$$

Singularity analysis of (3.55) yields

$$[z^n]\partial_u^r HW(z,u)\Big|_{u=1} \sim \tau^2 \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+1} 2^{-\frac{r+1}{2}} r!(3r-2)!!! \frac{\phi'(\tau)^n}{\Gamma\left(\frac{3r+1}{2}\right)} n^{\frac{3r+1}{2}-1}.$$

Recall that

$$\frac{1}{[z^n]HW(z,1)} \sim \frac{1}{\tau}\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\sqrt{2\pi}n^{1/2}\phi'(\tau)^{-n}. \tag{3.56}$$

Altogether this gives

$$\begin{aligned}\mathbb{E}\left(H_{xy}^r\right) &\sim \mathbb{E}\left(H_{xy}(H_{xy}-1)\dots(H_{xy}-r+1)\right)\\ &= \frac{[z^n]\partial_u^r HW(z,u)\Big|_{u=1}}{[z^n]HW(z,1)}\\ &\sim \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^r \frac{\sqrt{\pi}r!(3r-2)!!!}{2^{\frac{r}{2}}\Gamma\left(\frac{3r+1}{2}\right)}n^{\frac{3r}{2}}.\end{aligned}$$

The final statement about the limiting distribution of the normalised hitting time in Theorem 3.2.1 follows exactly in the same way as in the proof of Theorem 3.1.2.

We proceed to find the mixed moments for weighted trees. Note that

$$\Pr\left(H_{xy}=k, d(x,y)=d\right) = \frac{n![u^k z^n]HW_d(z,u)}{n![z^n]HW(z,1)},$$

so that

$$\sum_k \Pr\left(H_{xy}=k, d(x,y)=d\right)u^k = \frac{[z^n]HW_d(z,u)}{[z^n]HW(z,1)}$$

and

$$\sum_{d=0}^{\infty}\sum_{k=0}^{\infty} d^s \Pr\left(d(x,y)=d, H_{xy}=k\right)u^k = \frac{[z^n]\sum_{d=0}^{\infty} d^s HW_d(z,u)}{[z^n]HW(z,1)}.$$

We obtain the following generalisation of Theorem 3.1.8.

Theorem 3.2.4. *Let x and y be two randomly selected vertices of a random weighted labelled tree with n vertices. The joint moments of the distance $d(x,y)$ and the hitting time H_{xy} between x and y satisfy the asymptotic formula*

$$\mathbb{E}(d(x,y)^s H_{xy}^r) \sim C_{r,s}\left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+s} n^{3r/2+s/2},$$

where

$$C_{r,s} = \frac{\sqrt{\pi}(-3)^{r+s}r!(r+s)!}{2^{(r+3s)/2}\Gamma(\frac{3r+s+1}{2})}\sum_{k=0}^{s}(-1)^k\binom{s}{k}\binom{(2k-1)/3}{r+s}.$$

Proof. Working as before we obtain a modified version of (3.54):

$$\begin{aligned}
&\sum_{d=0}^{\infty} d^s \partial_u^r HW_d(z,u)\Big|_{u=1} \\
&\sim \sum_{d=0}^{\infty} \tau^2 \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m\geq 1\\ h_1+\ldots+h_m=r}} \binom{r}{h_1,\ldots,h_m} d^{r+m+s} \tilde{Y}(z)^{d-1-m} \left(1-z\phi'(\tau)\right)^{\frac{m}{2}-r} \\
&\quad \cdot 2^{\frac{m}{2}} \left(\tau\sqrt{\frac{\phi''(\tau)}{\phi(\tau)}}\right)^m \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i} \\
&\sim \tau^2 \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m\geq 1\\ h_1+\ldots+h_m=r}} \binom{r}{h_1,\ldots,h_m} (r+m+s)! \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+s+1} \\
&\quad \cdot \left(1-z\phi'(\tau)\right)^{-\frac{3r+s+1}{2}} 2^{-\frac{r+s+1}{2}} \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i} \\
&= \tau^2 \left(1-z\phi'(\tau)\right)^{-\frac{3r+s+1}{2}} 2^{-\frac{r+s+1}{2}} \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+s+1} \\
&\quad \cdot \sum_{m=1}^{r} \sum_{\substack{h_1,h_2,\ldots,h_m\geq 1\\ h_1+\ldots+h_m=r}} \binom{r}{h_1,\ldots,h_m} (r+m+s)! \prod_{i=1}^{m} \frac{(2h_i-3)!!}{h_1+\ldots+h_i+i}.
\end{aligned}$$

Working in exactly the same way as we did to derive (3.33) we obtain,

$$\begin{aligned}
&\sum_{d=0}^{\infty} d^s \partial_u^r HW_d(z,u)\Big|_{u=1} \\
&\sim \tau^2 \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+s+1} 2^{-\frac{r+3s+1}{2}} \left(1-z\phi'(\tau)\right)^{-\frac{3r+s+1}{2}} r!(r+s)!(-3)^{r+s} \\
&\quad \cdot \sum_{k=0}^{s} (-1)^k \binom{s}{k} \binom{\frac{2k-1}{3}}{r+s}.
\end{aligned} \tag{3.57}$$

Singularity analysis of (3.57), together with the asymptotic equivalence of $HW(z,1)$ in (3.56), yields

$$\begin{aligned}
&\mathbb{E}\left(d(x,y)^s H_{xy}^r\right) \\
&\sim \mathbb{E}\left(d(x,y)^s H_{xy}(H_{xy}-1)\ldots(H_{xy}-r+1)\right) \\
&\sim \frac{[z^n]\sum_{d=0}^{\infty} d^s \partial_u^r HW_d(z,u)\Big|_{u=1}}{[z^n]HW(z,1)}
\end{aligned}$$

$$\sim \left(\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}\right)^{r+s} \frac{\sqrt{\pi} r!(r+s)!(-3)^{r+s} \sum_{k=0}^{s}(-1)^k \binom{s}{k}\binom{\frac{2k-1}{3}}{r+s}}{2^{\frac{r+3s}{2}}\Gamma\left(\frac{3r+s+1}{2}\right)} n^{\frac{3r+s}{2}}.$$

■

Working as we did to derive (3.34), we obtain

$$\begin{aligned}
\operatorname{corr}\left(d(x,y), H_{xy}\right) &= \frac{\operatorname{cov}\left(d(x,y), H_{xy}\right)}{\sqrt{\operatorname{Var}\left(d(x,y)\right)\operatorname{Var}\left(H_{xy}\right)}} \\
&\sim \frac{\frac{\phi(\tau)}{\phi''(\tau)}\left(2-\frac{\pi}{2}\right)n^2}{\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}\left(2-\frac{\pi}{2}\right)\frac{\phi(\tau)}{\phi''(\tau)}\left(\frac{32}{15}-\frac{\pi}{2}\right)n^4}} \\
&\approx 0.873486,
\end{aligned}$$

so that the asymptotic correlation between the distance between two vertices and their hitting time remains the same in the weighted case. We conclude this section by finding the first moment of the hitting time variance in weighted labelled trees.

Theorem 3.2.5. *Let x and y be two randomly selected vertices of a random weighted labelled tree with n vertices. The expected hitting time variance between x and y satisfies the asymptotic equivalence*

$$\mathbb{E}\left(Var_{x,y}\right) \sim \frac{\phi(\tau)}{\tau^2\phi''(\tau)}\frac{32}{15}n^3.$$

Proof. Recall that

$$\begin{aligned}
\operatorname{Var}_{x,y} = 8\sum_{i=0}^{d-1}(d-i)\left(\sum_{\tau_v \prec T_i} R(\tau_v)+1\right) + 8\sum_{i=0}^{d-1}\frac{2(d-i)-1}{2}\left(\sum_{j=0}^{i}|T_j|\right)^2 \\
-8\sum_{i=0}^{d-1}\frac{2(d-i)+1}{2}\sum_{j=0}^{i}|T_j|.
\end{aligned} \tag{3.58}$$

We need to find

$$\mathbb{E}\left(\operatorname{Var}_{x,y}\right) = \frac{[z^n]\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d} Var_{x,y}\prod_{j=0}^{d}\frac{w(T_j)z^{|T_j|}}{|T_j!|}}{[z^n]\sum_{T,x,y}\frac{w(T)}{|T|!}z^{|T|}}, \tag{3.59}$$

where we already know from (3.44) that

$$[z^n]\sum_{T,x,y}\frac{w(T)}{|T|!}z^{|T|} \sim \tau\sqrt{\frac{\phi(\tau)}{2\phi''(\tau)\pi}}\phi'(\tau)^n n^{-1/2}.$$

Recall

$$Y(z) \sim \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}} \cdot \sqrt{1 - z\phi'(\tau)}$$

and

$$\tilde{Y}(z) = 1 - \frac{Y(z)}{zY'(z)} \sim 1 - \tau\sqrt{\frac{2\phi''(\tau)}{\phi(\tau)}} \cdot \sqrt{1 - z\phi'(\tau)}.$$

Now for the tree T_0, rooted at the starting point of the path,

$$\begin{aligned} R(z) &:= \sum_{T_0} R(T_0)w(T_0)\frac{z^{|T_0|}}{|T_0|!} \\ &= \sum_{T_0}\sum_{\tau \prec T_0} w(T_0)R(\tau)\frac{z^{|T_0|}}{|T_0|!} + z^2Y''(z) + Y(z) \\ &= zR(z)\sum_{d=1}^{\infty} w(d+1)\frac{dY(z)^{d-1}}{d!} + z^2Y''(z) + Y(z) \\ &= zR(z)\phi'(Y(z)) + z^2Y''(z) + Y(z), \end{aligned}$$

implying

$$R(z) = \frac{z^2Y''(z) + Y(z)}{1 - z\phi'(Y(z))},$$

where

$$(1 - z\phi'(Y(z)))^{-1} = \frac{zY'(z)}{Y(z)}$$

as in the start of this section. On the other hand for a tree T_i, not rooted at any of the endpoints of the path,

$$\begin{aligned} \tilde{R}(z) &:= \sum_{T_i} R(T_i)w(T_i)\frac{z^{|T_i|}}{|T_i|!} \\ &= zR(z)\tilde{\phi}'(Y(z)) + z^2\tilde{Y}''(z) + \tilde{Y}(z), \end{aligned}$$

where

$$z\tilde{\phi}'(Y(z)) = \frac{\tilde{Y}'(z) - \tilde{Y}(z)/z}{Y'(z)} \sim \tau\frac{\phi''(\tau)}{\phi(\tau)}.$$

We proceed by evaluating the three main terms in (3.58). Now

$$\begin{aligned} &8\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d}\sum_{i=0}^{d-1}(d-i)\left(\sum_{\tau \prec T_i} R(\tau) + 1\right)\prod_{j=0}^{d}\frac{w(T_j)z^{|T_j|}}{|T_j!|} \\ &= 8Y(z)\sum_{d=1}^{\infty}\left(d\left(\frac{z^2Y''(z) + Y(z)}{1 - \phi'(Y(z))} - z^2Y''(z) - Y(z)\right)\tilde{Y}(z)^{d-1}\right. \end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^{d-1}(d-i)Y(z)\tilde{Y}(z)^{d-2}\left[\tilde{R}(z)-z^2\tilde{Y}''(z)-\tilde{Y}(z)\right]+\sum_{i=0}^{d-1}(d-i)Y(z)\tilde{Y}(z)^{d-1}\Bigg)\\
&\sim 8Y(z)\sum_{d=1}^{\infty}\frac{d^2}{2}Y(z)\tilde{Y}(z)^{d-2}\tilde{R}(z)\\
&\sim 4Y(z)^2 z\tilde{\phi}'(Y(z))\frac{z^2Y''(z)}{1-z\phi'(Y(z))}\sum_{d=1}^{\infty}d^2\tilde{Y}(z)^{d-2}\\
&\sim 4\tau^2\cdot\tau\frac{\phi''(\tau)}{\phi(\tau)}\cdot\frac{1}{2\sqrt{2}}\sqrt{\frac{\phi(\tau)}{\phi''(\tau)}}(1-z\phi'(\tau))^{-3/2}\cdot\frac{1}{\tau}\sqrt{\frac{\phi(\tau)}{2\phi''(\tau)}}(1-z\phi'(\tau))^{-1/2}\\
&\quad\cdot\frac{2!}{\tau^3}\left(\frac{\phi(\tau)}{2\phi''(\tau)}\right)^{3/2}(1-z\phi'(\tau))^{-3/2}\\
&\sim\frac{1}{\sqrt{2}\tau}\left(\frac{\phi(\tau)}{\phi''(\tau)}\right)^{3/2}(1-z\phi'(\tau))^{-7/2},
\end{aligned}
\tag{3.60}
$$

whereas

$$
\begin{aligned}
&8\sum_{d=1}^{\infty}\sum_{T_0,\dots,T_d}\sum_{i=0}^{d-1}\frac{2(d-i)-1}{2}\left(\sum_{j=0}^{i}|T_j|\right)^2\prod_{j=0}^{d}\frac{w(T_j)z^{|T_j|}}{|T_j!|}\\
&=8Y(z)\sum_{d=1}^{\infty}\sum_{i=0}^{d-1}\frac{2(d-i)-1}{2}\Bigg(z(Y'(z)+zY''(z))\tilde{Y}(z)^{d-1}\\
&\quad+iz(\tilde{Y}'(z)+z\tilde{Y}''(z))Y(z)\tilde{Y}(z)^{d-2}\\
&\quad+2iz^2Y'(z)\tilde{Y}'(z)\tilde{Y}(z)^{d-2}+i(i-1)z^2\tilde{Y}'(z)^2Y(z)\tilde{Y}(z)^{d-3}\Bigg)\\
&=Y(z)\sum_{d=1}^{\infty}\Bigg(4d^2z(Y'(z)+zY''(z))\tilde{Y}(z)^{d-1}\\
&\quad+\frac{2}{3}(2d^3-3d^2+d)z(\tilde{Y}'(z)+z\tilde{Y}''(z))Y(z)\tilde{Y}(z)^{d-2}\\
&\quad+\frac{4}{3}(2d^3-3d^2+d)z^2Y'(z)\tilde{Y}'(z)\tilde{Y}(z)^{d-2}\\
&\quad+\frac{2}{3}(d^4-4d^3+5d^2-2d)z^2\tilde{Y}'(z)^2Y(z)\tilde{Y}(z)^{d-3}\Bigg)\\
&\sim Y(z)\sum_{d=1}^{\infty}\left(\frac{4}{3}d^3z^2\tilde{Y}''(z)Y(z)\tilde{Y}(z)^{d-2}+\frac{2}{3}Y(z)d^4\tilde{Y}(z)^{d-3}(z\tilde{Y}'(z))^2\right)\\
&\sim\frac{4}{3}\tau^2\cdot\frac{\tau}{2}\sqrt{\frac{\phi''(\tau)}{2\phi(\tau)}}(1-z\phi'(\tau))^{-3/2}\cdot\frac{3!}{(1-\tilde{Y}(z))^4}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{2}{3}\tau^2\cdot\tau^2\frac{\phi''(\tau)}{2\phi(\tau)}(1-z\phi'(\tau))^{-1}\frac{4!}{(1-\tilde{Y}(z))^5}\\
&\sim\frac{1}{\tau}\left(\frac{\phi(\tau)}{\phi''(\tau)}\right)^{3/2}\left(\frac{1}{\sqrt{2}}+\frac{2}{\sqrt{2}}\right)(1-z\phi'(\tau))^{-7/2}\\
&=\frac{1}{\tau}\left(\frac{\phi(\tau)}{\phi''(\tau)}\right)^{3/2}\frac{3}{\sqrt{2}}(1-z\phi'(\tau))^{-7/2} \qquad (3.61)
\end{aligned}
$$

and finally the term

$$
-8\sum_{d=1}^{\infty}\sum_{T_0,\ldots,T_d}\sum_{i=0}^{d-1}\frac{2(d-i)+1}{2}\sum_{j=0}^{i}|T_j|\prod_{j=0}^{d}\frac{w(T_j)z^{|T_j|}}{|T_j!|}
$$

is not asymptotically relevant, as in the unweighted case. Singularity analysis of the sum of (3.60) and (3.61) gives

$$
[z^n]\sum_{d=1}^{\infty}\sum_{T_0,\ldots,T_d}Var_{x,y}\prod_{j=0}^{d}\frac{w(T_j)z^{|T_j|}}{|T_j!|}\sim\frac{1}{\tau}\left(\frac{\phi(\tau)}{\phi''(\tau)}\right)^{3/2}\cdot\frac{4}{\sqrt{2}}\cdot\frac{\phi'(\tau)^n}{\Gamma(7/2)}n^{5/2}
$$

and the result follows from (3.59). ∎

3.3 A final remark

The family of unlabeled trees (unrooted) are the most natural to consider from a graph-theoretic point of view. Following Janson's result in Theorem 3.1.1 (which also holds in more generality for simply generated trees) Wagner proved a similar limit law in [45] for the Wiener index of unlabelled trees:

Theorem 3.3.1 (see [45, Theorem 7]). *The normalised Wiener index $n^{-5/2}W(T_n)$ of a random unlabelled tree on n vertices converges in distribution to $\frac{\sqrt{2}}{c_1}\cdot\zeta$, where ζ is defined in terms of the Brownian excursion as in Theorem 3.1.1. Furthermore, all moments converge:*

$$
\mathbb{E}\left(W(T_n)^k\right)=\mathbb{E}\left(\zeta^k\right)\cdot\left(\frac{\sqrt{2}}{c_1}\right)^k n^{\frac{5k}{2}}\left(1+\mathcal{O}\left(n^{-1/2}\right)\right).
$$

We did not perform the calculations to find the higher moments of the hitting time in unlabelled trees, which should be a bit trickier since one has to take symmetries into account, but as a consequence of Theorem 1.1.7 we expect the order to remain the same as in the labelled case. We expect that

$$
\mathbb{E}(H_{xy}^r)\sim C_rK^rn^{3r/2},
$$

where C_r is defined as in Theorem 3.1.2 and K is a constant.

Chapter 4

Hitting times in random increasing trees

An *increasing tree* is a rooted labelled tree in which the labels increase along any branch from the root. The first unified treatment of these trees was given in [5] and before this classes of these trees have surfaced under various guises as tree representations of permutations, as data structures in computer science and as probabilistic models in diverse applications. In the first section we consider hitting times in *non-plane* increasing trees, called *recursive trees*. The next section deals with increasing trees where the internal nodes are labelled and their number of children at most d. In the final section we consider hitting times in a generalisation of *plane* increasing trees, called *generalised plane oriented recursive trees*.
Random increasing trees are quite different from the labelled trees we considered in the previous section. These trees are more balanced, i.e. "flatter" than labelled or more generally simply-generated trees; they have average distances of order $\log n$ (see [32]). The root plays a special role in increasing trees so we study the hitting time from the root to a random vertex and vice versa. Unlike the previous chapter where we could only show that the limiting distribution is continuous, we found explicit limiting distributions in this chapter. The (normalised) hitting times from the root to a random vertex as well as the hitting times between two random vertices converges weakly to the standard normal distribution for all classes of increasing trees, whereas the (normalised) hitting times from a random vertex to the root converges weakly to different Dickman distributions for the different classes of increasing trees.

4.1 Recursive trees

A *recursive tree* is a rooted labelled tree with the property that any child node has a larger label than its parent, where the root has label 1. Denote the class of recursive trees of size n by $\mathcal{R}_n$. A recursive tree of size n is obtained by attaching label 2 to the root, 3 to either 1 or 2, 4 to 1 or 2 or 3 and so on. Since we can attach each $k = 2, \ldots, n$ to $k-1$ possible labels, there are $\prod_{k=2}^{n-1} k = (n-1)!$ recursive trees of size n. We assume each new vertex in this process is attached with equal probability to one of the preceding vertices so that each of the $(n-1)!$ recursive trees occurs with equal probability.
Recursive trees have been intensively studied in the literature. Early literature includes the survey by Mahmoud and Smythe [41] and the paper [30]. Neininger obtained asymptotics in [35] for the mean and variance of the Wiener index in a random recursive tree:

$$\mathbb{E}\left(W(T_n)\right) \sim n^2 \log n \quad \text{and} \quad \mathrm{Var}\left(W(T_n)\right) \sim \frac{31 - 3\pi^2}{18} n^4.$$

It is shown in [12] and [28] that the asymptotics for the mean of the *total path length* (the sum of the distances to the root) is $n \log n$. Various results on the *depth* are also known (see [15, Theorem 6.17] and [11, Theorem 3]). We start by finding the *depth* to a vertex v chosen uniformly at random from a random recursive tree of size n.

Lemma 4.1.1. *Let $d(1, v)$ be the distance from the root to a random vertex v in a recursive tree T chosen randomly from $\mathcal{R}_n$. The probability generating function for the depth is*

$$\begin{aligned} R(u) &:= \sum_{k=0}^{n-1} \mathrm{Pr}_{\mathcal{R}_n, v}\left(d(1, v) = k\right) u^k \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{k} \mathrm{Pr}\left(d(1, i) = k\right) u^k \\ &= \frac{1}{n!} \prod_{j=1}^{n-1} (j + u). \end{aligned}$$

Proof. We prove this assertion by induction on n. The result holds trivially for $n = 1$. If $|T| = n$ and $i = 1, \ldots, n-1$ then

$$\begin{aligned} \mathrm{Pr}_{\mathcal{R}_n}\left(d(1, i) = k\right) &= \frac{(n-1)|T_{n-1,i,k}|}{(n-1)!} \\ &= \frac{|T_{n-1,i,k}|}{(n-2)!} \\ &= \mathrm{Pr}_{\mathcal{R}_{n-1}, v}\left(d(1, v) = k\right), \end{aligned}$$

where $|T_{n,i,k}|$ counts the number of recursive trees of size n with $d(1,i) = k$. If $i = n$ it follows from the law of total probability after conditioning on the parent of n that

$$\begin{aligned}\Pr_{\mathcal{R}_n}\left(d(1,n)=k\right) &= \frac{1}{n-1}\sum_{j=1}^{n-1}\Pr_{\mathcal{R}_n}\left(d(1,j)=k-1\right)\\ &= \frac{1}{n-1}\sum_{j=1}^{n-1}\Pr_{\mathcal{R}_{n-1}}\left(d(1,j)=k-1\right).\end{aligned}$$

It follows from the induction hypothesis that

$$\begin{aligned}&\sum_k \Pr_{\mathcal{R}_{n,v}}\left(d(1,v)=k\right)u^k\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_k \Pr_{\mathcal{R}_n}\left(d(1,i)=k\right)u^k\\ &= \frac{1}{n}\sum_{i=1}^{n-1}\sum_k \Pr_{\mathcal{R}_{n-1}}\left(d(1,i)=k\right)u^k + \frac{1}{n}\sum_k \frac{1}{n-1}\sum_{j=1}^{n-1}\Pr_{\mathcal{R}_{n-1}}\left(d(1,j)=k-1\right)u^k\\ &= \frac{n-1}{n}\frac{1}{(n-1)!}\prod_{j=1}^{n-2}(j+u) + \frac{u}{n}\frac{1}{(n-1)!}\prod_{j=1}^{n-2}(j+u)\\ &= \frac{1}{(n-1)!}\prod_{j=1}^{n-2}(j+u)\left(\frac{n-1}{n}+\frac{u}{n}\right)\\ &= \frac{1}{n!}\prod_{j=1}^{n-1}(j+u).\end{aligned}$$

■

Alternatively one may notice

$$\sum_{i=1}^{n}|T_{n,i,k}| = \sum_{i=k+1}^{n}|T_{n,i,k}| = \begin{bmatrix} n \\ k+1 \end{bmatrix},$$

where $\begin{bmatrix} n \\ k \end{bmatrix}$ counts the number of permutations of $1,\ldots,n$ with exactly k cycles and is known as the (unsigned) Stirling number of the first kind (see [20, Chapter 6]). Indeed if $d(1,i) = k$ there is a sequence of recursive subtrees $T_0, T_1, \ldots, T_k$ of T induced by removing the edges between the roots of those trees on the path $(x_0 = 1, x_1, \ldots, x_k = i)$. For each of these $k+1$ subtrees T_j, rooted along the path, there are $(|T_j| - 1)!$ possibilities, corresponding to the $k+1$ cycles in a permutation of $1,\ldots,n$. The statement of Lemma 4.1.1 follows since

$$\Pr_{\mathcal{R}_{n,v}}\left(d(1,v)=k\right) = \frac{\sum_{i=1}^{n}|T_{n,i,k}|}{n!}$$

and

$$\sum_{k=0}^{n-1} \begin{bmatrix} n \\ k+1 \end{bmatrix} u^k = \prod_{j=1}^{n-1}(j+u)$$

for $n \geq 1$. Notice that

$$\frac{1}{n!}\prod_{j=1}^{n-1}(j+u) = \prod_{j=2}^{n}\left(\frac{j-1}{j}+\frac{u}{j}\right),$$

i.e. the depth is the sum of $n-1$ independent Bernoulli random variables with success probability $1/j$:

$$d(1,v) = \sum_{j=2}^{n} X_j, \tag{4.1}$$

$X_j \sim \mathrm{Be}(1/j)$. We can identify these Bernoulli random variables. If $I(A)$ denotes the indicator of event A and $A_{j,k}$ the event that j is on the path from 1 to k then

$$X_j = I(A_{j,v}).$$

Now

$$\begin{aligned}\Pr(A_{j,v}) &= \frac{1}{n}\sum_{i=1}^{n} \Pr\left(j \text{ on path from 1 to } i\right)\\ &= \frac{1}{n} + \frac{1}{n}\sum_{i=j+1}^{n} \Pr\left(j \text{ on path from 1 to } i\right).\end{aligned}$$

We need to show for $i = j+1, \ldots, n$ that

$$\Pr\left(j \text{ on path from 1 to } i\right) = \frac{1}{j}, \tag{4.2}$$

so that

$$\Pr(A_{j,v}) = \frac{1}{n} + \frac{n-j}{nj} = \frac{1}{j}.$$

To prove (4.2) we only need to show

$$\Pr\left(j \text{ on path from 1 to } n\right) = \frac{1}{j},$$

since

$$\Pr_{\mathcal{R}_i}\left(j \text{ on path from 1 to } n\right) = \Pr_{\mathcal{R}_n}\left(j \text{ on path from 1 to } n\right).$$

We need to show the number of recursive trees of size n with the property that j is on the path from the root to n, $r_{n,j}$, is equal to $(n-1)!/j$. We have to insert n in the recursive subtree rooted at j. The size of the subtree rooted at j can be $l = 1, 2, \ldots, n-j$. There are l places to insert n, $\binom{n-j-1}{l-1}$ ways to choose the labels in the subtree rooted at j and $j-1$ possible vertices adjacent to j. Hence

$$\begin{aligned} r_{n,j} &= \sum_{l=1}^{n-j}(l-1)!(n-l-2)!l\binom{n-j-1}{l-1}(j-1) \\ &= (n-j-1)!(j-1)!\sum_{l=1}^{n-j} l\binom{n-2-l}{j-2}. \end{aligned}$$

Finally

$$\sum_{l+1=2}^{n-j+1} l\binom{n-1-(l+1)}{j-2} = \binom{n-1}{j}.$$

Indeed, partition the j-element subsets chosen from $\{1, 2, \ldots, n-1\}$ according to the value of the second smallest number, $l+1$, chosen. There are l remaining possibilities for the smallest number. The remaining $j-2$ elements can be chosen from $\{l+2, l+3, \ldots, n-1\}$. Summing over all values of $l+1$ yields the result so that $r_{n,j} = (n-1)!/j$ as required.

We wish to generalise the probability generating function, obtained in Lemma 4.1.1, by including the hitting time from the root to a random vertex.

Theorem 4.1.2. *Let $d(1, v)$ and H_{1v} be the distance and hitting time from the root to a vertex v chosen uniformly at random in a recursive tree T chosen randomly from $\mathcal{R}_n$. The joint probability generating function for the hitting time and distance from the root is*

$$\begin{aligned} p_n(u, y) &:= \sum_l \sum_k \mathrm{Pr}_{\mathcal{R}_n, v}\left(d(1, v) = k, H_{1v} = l\right) u^l y^k \\ &= \frac{1}{n!}\sum_{T\in\mathcal{R}_n, i\in T} u^{H_{1i}} y^{d(1,i)} \\ &= \frac{1}{n!}\prod_{j=1}^{n-1}\left(n-j+u^{2j-1}y\right). \end{aligned}$$

Proof. We prove this assertion by induction on n. The result holds trivially for $n = 1$. If $T = n$ we can divide T into two recursive subtrees, T_1 and T_2 rooted at 1 and 2, by removing the edge between 1 and 2. We can partition the trees in $\mathcal{R}_n$ according to the size r of T_2. Summing over all possible values of r gives

$$(n-1)! = \sum_{r=1}^{n-1} \binom{n-2}{r-1} (r-1)!(n-r-1)!,$$

since the labels in T_2 can be chosen in $\binom{n-2}{r-1}$ ways. In particular r takes on each value in $\{1, 2, \ldots, n-1\}$ with equal probability $\frac{1}{n-1}$. Let A_r be the event that the subtree rooted at 2 is of cardinality r. Let $B_{v,m}$ be the event that the random vertex v is in the subtree rooted at $m = 1, 2$. Applying the law of total probability twice gives

$$\begin{aligned}
&\Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l) \\
&= \sum_{r=1}^{n-1} \Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r) \Pr_{\mathcal{R}_n}(A_r) \\
&= \frac{1}{n-1} \sum_{r=1}^{n-1} \Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r) \\
&= \frac{1}{n-1} \sum_{r=1}^{n-1} \sum_{m=1}^{2} \Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r, B_{v,m}) \Pr_{\mathcal{R}_n,v}(B_{v,m} | A_r) \\
&= \frac{1}{n-1} \sum_{r=1}^{n-1} \frac{n-r}{n} \Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r, B_{v,1}) \\
&\quad + \frac{1}{n-1} \sum_{r=1}^{n-1} \frac{r}{n} \Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r, B_{v,2}).
\end{aligned}$$

Note that

$$\Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r, B_{v,1}) = \Pr_{\mathcal{R}_{n-r},v}(d(1,v) = k, H_{1v} = 2rk + l), \tag{4.3}$$

since there is a path of length k, $(x_0 = 1, x_1, \ldots, x_k = v)$ from 1 to v with $H_{1v} = H_{1x_1} + H_{x_1x_2} + \ldots + H_{x_{k-1}v}$. Attaching a tree, of size r with r edges, to the root of a tree in $\mathcal{R}_{n-r}$ increases the hitting time between each pair of adjacent vertices along this path by $2r$ (the commute time, $H_{x_ix_{i+1}} + H_{x_{i+1}x_i}$, between two adjacent vertices on this path is equal to two times the number of edges, whilst the hitting time $H_{x_{i+1}x_i}$ remains the same in both cases). Equation (4.3) follows after summing the hitting times along this path.

Similarly, keeping in mind that we have to add the edge between 1 and 2 to the distance, we have

$$\Pr_{\mathcal{R}_n,v}(d(1,v) = k, H_{1v} = l | A_r, B_{v,2}) = \Pr_{\mathcal{R}_r,v}(d(1,v) = k+1, H_{1v} = m),$$

where

$$\begin{aligned} m &= H_{12} + 2(n-r)(d(1,v)-1) + l \\ &= 2(n-r)(k+1) - 1 + l. \end{aligned}$$

Now

$$\begin{aligned} &\sum_l \sum_k \Pr_{\mathcal{R}_{n,v}}\left(d(1,v)=k, H_{1v}=l\right) u^l y^k \\ &= \frac{1}{n-1}\sum_{r=1}^{n-1} \frac{n-r}{n} \sum_l \sum_k \Pr_{\mathcal{R}_{n-r,v}}\left(d(1,v)=k, H_{1v}=l\right) u^{2rk+l} y^k \\ &\quad + \frac{1}{n-1}\sum_{r=1}^{n-1} \frac{r}{n} \sum_l \sum_k \Pr_{\mathcal{R}_{r,v}}\left(d(1,v)=k, H_{1v}=l\right) u^{2(n-r)(k+1)-1+l} y^{k+1} \\ &= \frac{1}{n-1}\sum_{r=1}^{n-1} \frac{n-r}{n} \sum_l \sum_k \Pr_{\mathcal{R}_{n-r,v}}\left(d(1,v)=k, H_{1v}=l\right) u^l \left(yu^{2r}\right)^k \\ &\quad + \frac{1}{n-1}\sum_{r=1}^{n-1} \frac{r}{n} \frac{yu^{2(n-r)}}{u} \sum_l \sum_k \Pr_{\mathcal{R}_{r,v}}\left(d(1,v)=k, H_{1v}=l\right) u^l \left(yu^{2(n-r)}\right)^k \\ &= \frac{1}{n-1}\sum_{r=1}^{n-1} \frac{n-r}{n} \frac{1}{(n-r)!} \prod_{j=1}^{n-r-1} \left(n-r-j+u^{2(j+r)-1}y\right) \\ &\quad + \frac{1}{n-1}\sum_{r=1}^{n-1} \frac{r}{n} \frac{yu^{2(n-r)}}{u} \frac{1}{r!} \prod_{j=1}^{r-1} \left(r-j+u^{2(n-r+j)-1}y\right), \end{aligned} \tag{4.4}$$

where (4.4) follows from the induction hypothesis. Adding the terms in the second summand in reverse order to those in the first summand in (4.4) gives

$$\begin{aligned} &\sum_l \sum_k \Pr_{\mathcal{R}_{n,v}}\left(d(1,v)=k, H_{1v}=l\right) u^l y^k \\ &= \frac{1}{n!}\sum_{r=1}^{n-1} \left(1+u^{2r-1}y\right) \prod_{i=n-r}^{n-2} i \prod_{m=r}^{n-2} \left(n-1-m+u^{2m+1}y\right). \end{aligned}$$

Finally we show by induction on $k \le n-2$ that

$$\begin{aligned} &\frac{1}{n!}\prod_{j=1}^{n-1}\left(n-j+u^{2j-1}y\right) - \frac{1}{n!}\sum_{r=1}^{k}\left(1+u^{2r-1}y\right)\prod_{i=n-r}^{n-2} i \prod_{m=r}^{n-2}\left(n-1-m+u^{2m+1}y\right) \\ &= \frac{(n-k-1)(n-k)\dots(n-2)}{n!} \prod_{m=k}^{n-2}\left(n-1-m+u^{2m+1}y\right). \end{aligned}$$

The result clearly holds for $k = 1$. Suppose it holds for $k = 1, \ldots, n-3$. Now

$$\frac{1}{n!}\prod_{j=1}^{n-1}\left(n-j+u^{2j-1}y\right) - \frac{1}{n!}\sum_{r=1}^{k+1}\left(1+u^{2r-1}y\right)\prod_{i=n-r}^{n-2} i \prod_{m=r}^{n-2}\left(n-1-m+u^{2m+1}y\right)$$
$$= \frac{(n-k-1)(n-k)\ldots(n-2)}{n!}\prod_{m=k}^{n-2}\left(n-1-m+u^{2m+1}y\right)$$
$$- \frac{(n-k-1)(n-k)\ldots(n-2)}{n!}\left(1+u^{2k+1}y\right)\prod_{m=k+1}^{n-2}\left(n-1-m+u^{2m+1}y\right)$$
$$= \frac{(n-k-2)(n-k-1)\ldots(n-2)}{n!}\prod_{m=k+1}^{n-2}\left(n-1-m+u^{2m+1}y\right),$$

completing our proof by induction. Hence

$$\frac{1}{n!}\prod_{j=1}^{n-1}\left(n-j+u^{2j-1}y\right) - \frac{1}{n!}\sum_{r=1}^{n-1}\left(1+u^{2r-1}y\right)\prod_{i=n-r}^{n-2} i \prod_{m=r}^{n-2}\left(n-1-m+u^{2m+1}y\right)$$
$$= 0,$$

completing the proof of our theorem. ■

Notice that

$$\begin{aligned}
\sum_{l} \Pr_{\mathcal{R}_{n,v}}\left(H_{1v} = l\right) u^l &= \frac{1}{n!}\prod_{j=1}^{n-1}\left(n-j+u^{2j-1}\right) \\
&= \prod_{j=2}^{n}\left(\frac{n-j+1}{j} + \frac{u^{2(j-1)-1}}{j}\right) \\
&= \prod_{j=2}^{n}\left(\frac{j-1}{j} + \frac{u^{2(n-j+1)-1}}{j}\right), \qquad (4.5)
\end{aligned}$$

i.e. the hitting time from the root is a weighted sum of $n-1$ independent Bernoulli random variables with success probability $1/j$:

$$H_{1v} = \sum_{j=2}^{n}(2(n-j)+1)X_j,$$

$X_j \sim \mathrm{Be}(1/j)$. Hence

$$\mathbb{E}_{\mathcal{R}_{n,v}}\left(H_{1v}\right) = \sum_{j=2}^{n}\frac{(2(n-j)+1)}{j}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n-1} \frac{2j-1}{n-j+1} \\
&= 2\sum_{j=1}^{n-1} \frac{j}{n-j+1} - \sum_{j=1}^{n-1} \frac{1}{n-j+1} \\
&= 2\left(\sum_{j=1}^{n} H_j - n\right) - (H_n - 1) \\
&= 2\left((n+1)H_{n+1} - (n+1) - n\right) - (H_n - 1) \\
&= (2n+1)H_n - 4n + 1 \\
&\sim 2n\log n
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Var}_{\mathcal{R}_{n},v}\left(H_{1v}\right) \\
&= \sum_{j=2}^{n} (2(n-j)+1)^2 \frac{j-1}{j^2} \\
&= \sum_{j=2}^{n} \frac{(2(n-j)+1)^2}{j} - \sum_{j=2}^{n} \left(\frac{2(n-j)+1}{j}\right)^2 \\
&= 4n^2\left(H_n - 1\right) - 8n(n-1) + 2n(n+1) - 4 + 4n\left(H_n - 1\right) - 4(n-1) \\
&\quad + H_n - 1 - (4n^2 + 4n + 1)\left(H_n^{(2)} - 1\right) + 8n\left(H_n - 1\right) - 4(n-1) + 4\left(H_n - 1\right) \\
&= (2n+1)(2n+5)H_n - (2n+1)^2 H_n^{(2)} - 6n(n+1) \\
&\sim 4n^2 \log n.
\end{aligned}
$$

Theorem 4.1.3. *Let H_{1v} be the hitting time from the root to a random vertex v in a random recursive tree of size n. Then*

$$
H_n^* := \frac{H_{1v} - 2n\log n}{2n\sqrt{\log n}} \xrightarrow{d} N(0,1).
$$

Proof. We need to show the moment generating function

$$
M_n(t) := \mathbb{E}_{\mathcal{R}_{n},v}\left(e^{tH_n^*}\right) \to e^{t^2/2} \text{ as } n \to \infty.
$$

It follows from (4.5) that

$$
\begin{aligned}
M_n(t) &= \mathbb{E}\left(e^{t\left(\frac{H_{1v} - 2n\log n}{2n\sqrt{\log n}}\right)}\right) \\
&= e^{-t\sqrt{\log n}}\,\mathbb{E}\left(e^{\frac{t}{2n\sqrt{\log n}}H_{1v}}\right) \\
&= e^{-t\sqrt{\log n}} \prod_{j=2}^{n} \left(\frac{j-1}{j} + \frac{e^{\frac{t(2(n-j)+1)}{2n\sqrt{\log n}}}}{j}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= e^{-t\sqrt{\log n}} \exp\left(\sum_{j=2}^{n} \log\left(1 + \frac{e^{\frac{t(2(n-j)+1)}{2n\sqrt{\log n}}} - 1}{j}\right)\right) \\
&\sim e^{-t\sqrt{\log n}} \exp\left(\sum_{j=2}^{n} \frac{e^{\frac{t(2(n-j)+1)}{2n\sqrt{\log n}}} - 1}{j}\right) \\
&\sim e^{-t\sqrt{\log n}} \exp\left(\sum_{j=2}^{n} \frac{e^{\frac{t(n-j)}{n\sqrt{\log n}}} - 1}{j}\right) \\
&= e^{-t\sqrt{\log n}} \exp\left(\sum_{k=1}^{\infty} \frac{\left(\frac{t}{n\sqrt{\log n}}\right)^k}{k!} \sum_{j=2}^{n} \frac{(n-j)^k}{j}\right).
\end{aligned}
$$

It follows from the Euler-Maclaurin summation formula that

$$
\sum_{j=2}^{n} \frac{(n-j)^k}{j} = n^k \log n + \mathcal{O}\left(n^k\right),
$$

implying

$$
\exp\left(\sum_{k=1}^{\infty} \frac{\left(\frac{t}{n\sqrt{\log n}}\right)^k}{k!} \sum_{j=2}^{n} \frac{(n-j)^k}{j}\right) = \exp\left(t\sqrt{\log n} + t^2/2 + \mathcal{O}\left(1/\sqrt{\log n}\right)\right)
$$

and hence $M_n(t) \to e^{t^2/2}$.

■

We can derive similar results for the hitting time from a random vertex to the root.

Theorem 4.1.4. *Let $d(v,1)$ and H_{v1} be the distance and hitting time from a vertex v, chosen uniformly at random, to the root, in a recursive tree T chosen randomly from $\mathcal{R}_n$. The joint probability generating function for the hitting time and distance to the root is*

$$
\begin{aligned}
q_n(u,y) &:= \sum_{l} \sum_{k} \Pr_{\mathcal{R}_n,v}\left(d(v,1) = k, H_{v1} = l\right) u^l y^k \\
&= \frac{1}{n!} \sum_{T \in \mathcal{R}_n, i \in T} u^{H_{i1}} y^{d(i,1)} \\
&= \frac{1}{n!} \prod_{j=1}^{n-1} \left(j + u^{2j-1} y\right).
\end{aligned}
$$

We only give an outline of the proof, it mostly follows in the same way as the proof by induction of Theorem 4.1.2.

Proof. As before we partition T into two recursive subtrees, T_1 and T_2. Recall that A_r is the event that the subtree rooted at 2 is of cardinality r and $B_{v,m}$ is the event that v is in the subtree rooted at $m = 1, 2$. In this case

$$\Pr_{\mathcal{R}_{n,v}}\left(d(v,1) = k, H_{v1} = l | A_r, B_{v,1}\right) = \Pr_{\mathcal{R}_{n-r,v}}\left(d(v,1) = k, H_{v1} = l\right)$$

and

$$\Pr_{\mathcal{R}_{n,v}}\left(d(v,1) = k, H_{v1} = l | A_r, B_{v,2}\right) = \Pr_{\mathcal{R}_{r,v}}\left(d(v,1) = k+1, H_{v1} = 2r-1+l\right).$$

Working as we did to derive (4.4) we obtain

$$\begin{aligned}
&\sum_l \sum_k \Pr_{\mathcal{R}_{n,v}}\left(d(v,1) = k, H_{v1} = l\right) u^l y^k \\
&= \frac{1}{n-1}\sum_{r=1}^{n-1}\left(\frac{n-r}{n}\frac{1}{(n-r)!}\prod_{j=1}^{n-r-1}\left(j + u^{2j-1}y\right) + \frac{r}{n} y u^{2r-1}\frac{1}{r!}\prod_{j=1}^{r-1}\left(j + u^{2j-1}y\right)\right).
\end{aligned}$$

Rearranging the terms gives

$$\begin{aligned}
&\sum_l \sum_k \Pr_{\mathcal{R}_{n,v}}\left(d(v,1) = k, H_{v1} = l\right) u^l y^k \\
&= \frac{1}{n!}\sum_{r=1}^{n-1}\left(1 + u^{2(n-r)-1}y\right)\prod_{i=n-r}^{n-2} i \prod_{m=r}^{n-2}\left(n-1-m+u^{2(n-2-m)+1}y\right).
\end{aligned}$$

Finally one can show by induction on $k \leq n-2$ that

$$\begin{aligned}
&\frac{1}{n!}\prod_{j=1}^{n-1}\left(j + u^{2j-1}y\right) - \frac{1}{n!}\sum_{r=1}^{k}\left(1 + u^{2(n-r)-1}y\right)\prod_{i=n-r}^{n-2} i \prod_{m=r}^{n-2}\left(n-1-m+u^{2(n-2-m)+1}y\right) \\
&= \frac{(n-k-1)(n-k)\dots(n-2)}{n!}\prod_{m=k}^{n-2}\left(n-1-m+u^{2(n-2-m)+1}y\right),
\end{aligned}$$

completing the proof. ■

Note that Theorem 4.1.4 is in fact equivalent to Theorem 4.1.2:

$$\begin{aligned}
q_n(u,y) &= \frac{1}{n!}\sum_{T,i} u^{-H_{1i}} y^{d(1,i)}\left(u^{2(n-1)d(1,i)}\right) \\
&= p_n(u^{-1}, u^{2(n-1)}y) \qquad (4.6)
\end{aligned}$$

$$= \frac{1}{n!} \prod_{j=1}^{n-1} \left(j + u^{2j-1} y\right). \tag{4.7}$$

We can write

$$\begin{aligned} \sum_{l} \Pr_{\mathcal{R}_{n,v}} \left(H_{v1} = l\right) u^{l} &= \frac{1}{n!} \prod_{j=1}^{n-1} \left(j + u^{2j-1}\right) \\ &= \prod_{j=2}^{n} \left(\frac{j-1}{j} + \frac{u^{2(j-1)-1}}{j}\right), \end{aligned} \tag{4.8}$$

i.e. the hitting time to the root is a weighted sum of $n-1$ independent Bernoulli random variables with success probability $1/j$:

$$H_{v1} = \sum_{j=2}^{n} (2j-3) X_j,$$

$X_j \sim \mathrm{Be}(1/j)$. Hence

$$\begin{aligned} \mathbb{E}_{\mathcal{R}_{n,v}} \left(H_{v1}\right) &= \sum_{j=2}^{n} \frac{(2j-3)}{j} \\ &= 2(n-1) - 3\left(H_n - 1\right) \\ &\sim 2n \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} \mathrm{Var}_{\mathcal{R}_{n,v}} \left(H_{v1}\right) &= \sum_{j=2}^{n} (2j-3)^2 \frac{j-1}{j^2} \\ &= \sum_{j=2}^{n} \frac{(2j-3)^2}{j} - \sum_{j=2}^{n} \left(\frac{2j-3}{j}\right)^2 \\ &= 4\left(\frac{n(n+1)}{2} - 1\right) - 16(n-1) + 21\left(H_n - 1\right) - 9\left(H_n^{(2)} - 1\right) \\ &= 2n(n-7) + 21 H_n - 9 H_n^{(2)} \\ &\sim 2n^2. \end{aligned}$$

Contrary to the case for the hitting time from the root we show a normalised limiting distribution, of the hitting time to the root, converges to a Dickman distribution. The reader can consult the paper [39] by Penrose and Wade for an overview of the Dickman distribution. The Dickman distribution also appears in a context similar to ours in [24].

Theorem 4.1.5. *Let H_{v1} be the hitting time, from a random vertex v to the root, in a random recursive tree of size n. Then the moment generating function*

$$F_n(t) := \mathbb{E}\left(e^{\frac{tH_{v1}}{2n}}\right) \to \exp\left(\sum_{m=1}^{\infty} \frac{t^m}{m!m}\right) \quad \textit{as } n \to \infty.$$

In particular $(2n)^{-1}H_{v1} \xrightarrow{d} \mathrm{GD}(1)$, *where* $\mathrm{GD}(1)$ *is the Dickman distribution.*

Proof. It follows from (4.8) that

$$\begin{aligned}
\mathbb{E}\left(e^{\frac{tH_{v1}}{2n}}\right) &= \exp\left(\sum_{j=2}^{n} \log\left(1 + \frac{e^{\frac{(2j-3)t}{2n}} - 1}{j}\right)\right) \\
&\sim \exp\left(\sum_{j=2}^{n} \frac{e^{\frac{(2j-3)t}{2n}} - 1}{j}\right) \\
&\sim \exp\left(\sum_{j=1}^{n} \frac{e^{\frac{jt}{n}} - 1}{j}\right) \\
&= \exp\left(\sum_{j=1}^{n} \frac{1}{n} \frac{e^{\frac{jt}{n}} - 1}{j/n}\right) \\
&\sim \exp\left(\int_0^1 \frac{e^{at} - 1}{a} da\right) \\
&= \exp\left(\int_0^t \frac{e^{x} - 1}{x} dx\right) \\
&= \exp\left(\sum_{m=1}^{\infty} \frac{t^m}{m!m}\right). \qquad (4.10)
\end{aligned}$$

This completes the proof since (4.10) is the moment generating function of the Dickman distribution. ∎

Remark 1. We remark that Moon showed in [33] that the *barycentre* (the vertex or two adjacent vertices that minimises the sum of the distances to all other vertices) is typically close to the root of the recursive tree. Recalling the statement of Theorem 1.1.5 it is not at all surprising that $H_{1v} \geq H_{v1}$. Moreover the different limiting distributions reflects that H_{1v} is concentrated while H_{v1} is not.

We require the following theorem to find the limiting distribution between two random vertices in a recursive tree:

Theorem 4.1.6 (see [25, Theorem 6, 7] and the discussion above)**.** *Let $v \wedge w$ be the lowest common ancestor (i.e. the vertex furthest away from the root* *that has both v and w as descendants) of two vertices $v < w$ in a recursive tree of size n with root 1. The distribution of $d(1, v \wedge w)$ is independent of w. Furthermore the limit distribution of the depth of $v \wedge w$, as $v, w \to \infty$, is* Geom(1/2) *and*

$$\mathbb{E}\left(d(1, n-1 \wedge n)\right) = \frac{n-2}{n-1}.$$

Now

$$\mathbb{E}\left(d(1, v \wedge w)\right) = \mathbb{E}\left(d(1, v \wedge v+1)\right) = \frac{v-1}{v} = \mathcal{O}(1).$$

Letting $s = v \wedge w$ gives

$$\begin{aligned} |H_{vw} - H_{1w}| &= |H_{vs} + H_{sw} - H_{1s} - H_{sw}| \\ &= |H_{v1} - H_{s1} - H_{1s}| \\ &\leq H_{v1} + 2(n-1)d(1, s). \end{aligned}$$

Hence

$$\mathbb{E}\left(|H_{vw} - H_{1w}|\right) \leq \mathbb{E}\left(H_{v1}\right) + 2(n-1)\mathbb{E}\left(d(1,s)\right) = \mathcal{O}(n),$$

by Theorem 4.1.6 and (4.9). Markov's inequality states that

$$\Pr\left(|X| \geq t\mathbb{E}\left(|X|\right)\right) \leq \frac{1}{t}.$$

Letting

$$X = \frac{H_{vw} - H_{1w}}{2n\sqrt{\log n}}$$

and $t = (\log n)^{1/3}$ we see that

$$\frac{H_{vw} - H_{1w}}{2n\sqrt{\log n}} \xrightarrow{P} 0.$$

Since

$$\frac{H_{1w} - 2n\log n}{2n\sqrt{\log n}} \xrightarrow{d} N(0,1),$$

Slutsky's theorem (see [21, Theorem 2.39]) gives

$$\frac{H_{vw} - 2n\log n}{2n\sqrt{\log n}} \xrightarrow{d} N(0,1).$$

We state this result as

Theorem 4.1.7. *The normalised hitting time between two random vertices v and w in a random recursive tree converges weakly to the standard normal distribution:*

$$\frac{H_{vw} - 2n\log n}{2n\sqrt{\log n}} \xrightarrow{d} N(0,1).$$

We conclude this section by finding the asymptotic correlation between the hitting time and the distance to and from the root.

Proposition 4.1.8. *There is asymptotically no correlation between the hitting time and distance to the root, that is* $\operatorname{corr}(d(v,1), H_{v1}) \sim \sqrt{\frac{2}{\log n}}$. *On the other hand the hitting time and distance from the root are asymptotically perfectly correlated,* $\operatorname{corr}(d(1,v), H_{1v}) \sim 1$.

Proof. It follows from (4.1) that

$$\mathbb{E}(d(1,v)) = \sum_{j=2}^{n} \frac{1}{j} = H_n - 1,$$

whereas

$$\begin{aligned}\partial_u p_n(u,y) &= \frac{1}{n!} \sum_{j=1}^{n-1} \partial_u \left(n - j + u^{2j-1} y\right) \prod_{i \neq j} \left(n - i + u^{2i-1} y\right) \\ &= \frac{1}{n!} \sum_{j=1}^{n-1} \frac{(2j-1)u^{2j-2} y}{n - j + u^{2j-1} y} \prod_{i=1}^{n-1} \left(n - i + u^{2i-1} y\right).\end{aligned}$$

Now

$$\begin{aligned}\left.\frac{\partial^2}{\partial_y \partial_u} p_n(u,y)\right|_{u=1} &= \frac{1}{n!} \sum_{j=1}^{n-1} \frac{(n-j+y)(2j-1) - (2j-1)y}{(n-j+y)^2} \prod_{i=1}^{n-1} (n-i+y) \\ &+ \frac{1}{n!} \sum_{j=1}^{n-1} \frac{(2j-1)y}{n-j+y} \prod_{i=1}^{n-1} (n-i+y) \sum_{k=1}^{n-1} \frac{1}{n-k+y},\end{aligned}$$

so that

$$\begin{aligned}&\mathbb{E}\left(H_{1v} d(1,v)\right) \\ &= \left.\frac{\partial^2}{\partial_y \partial_u} p_n(u,y)\right|_{u=y=1} \\ &= \sum_{j=1}^{n-1} \left(\frac{(n-j)(2j-1)}{(n-j+1)^2} + \frac{2j-1}{n-j+1}(H_n - 1) \right) \\ &= -\sum_{j=1}^{n-1} \frac{2j-1}{(n-j+1)^2} + H_n \sum_{j=1}^{n-1} \frac{2j-1}{n-j+1} \\ &= -2\left(\sum_{j=1}^{n} H_j^{(2)} - n \right) + H_n^{(2)} - 1 + H_n \left(2 \left(\sum_{j=1}^{n} H_j - n \right) - H_n + 1 \right) \\ &= -2\left((n+1)H_n^{(2)} - H_n - n\right) + H_n^{(2)} - 1 + H_n \left(2\left(nH_n + H_n - 2n\right) - H_n + 1\right)\end{aligned}$$

$$= (2n+1)H_n^2 - 4nH_n + 3H_n + 2n - (2n+1)H_n^{(2)} - 1.$$

Recall

$$\mathbb{E}\left(H_{1v}\right) = (2n+1)H_n - 4n + 1$$

so that

$$\operatorname{cov}\left(d(1,v), H_{1v}\right) = (2n+3)H_n - 2n - (2n+1)H_n^{(2)}.$$

Finally

$$\operatorname{Var}\left(d(1,v)\right) = \sum_{j=2}^{n} \frac{j-1}{j^2} = H_n - H_n^{(2)}$$

and

$$\operatorname{Var}\left(H_{1v}\right) = (2n+1)(2n+5)H_n - (2n+1)^2 H_n^{(2)} - 6n(n+1)$$

implies

$$\operatorname{corr}\left(d(1,v), H_{1v}\right) \sim 1.$$

Similarly

$$\begin{aligned}
\mathbb{E}\left(H_{v1} d(v,1)\right) &= \frac{\partial^2}{\partial_y \partial_u} q_n(u,y)\bigg|_{u=y=1} \\
&= \frac{1}{n!} \sum_{j=1}^{n-1} \left(\frac{(2j-1)j}{(j+1)^2} + \frac{2j-1}{j+1}\left(H_n - 1\right) \right) \\
&= 2nH_n - 3H_n^2 - H_n + 3H_n^{(2)} - 1,
\end{aligned}$$

so that

$$\operatorname{cov}\left(d(v,1), H_{v1}\right) = 2n - 5H_n + 3H_n^{(2)}.$$

Since

$$\operatorname{Var}\left(H_{v1}\right) = 2n(n-7) + 21H_n - 9H_n^{(2)}$$

we have

$$\operatorname{corr}\left(d(v,1), H_{v1}\right) \sim \sqrt{\frac{2}{\log n}}.$$

■

Recall that

$$H_{xy} = H_{xx_1} + H_{x_1x_2} + \ldots + H_{x_{l-1}y}$$

for any path $(x_0 = x, x_2, \ldots, x_l = y)$ in a tree where $H_{x_i,x_{i+1}} = 2|T_{x_i:x_{i+1}}| - 1$ and $T_{x_i:x_{i+1}}$ is the subtree rooted at x_i after the removal of the edge x_ix_{i+1}. Hence the previous proposition implies that for "nearly all" recursive trees the order of the tree is not concentrated in a single branch attached to the root, but is spread out over the remaining branches attached to the root.

4.2 d-ary increasing trees

A *d-ary tree* is a rooted plane tree with the property that all of its nodes have either $d \geq 2$ children (internal nodes) or no children (external nodes). Note that a d-ary tree with $dn+1$ nodes has n internal nodes and $(d-1)n+1$ external nodes since the tree has dn edges and each internal node is connected to its children by d edges. An *increasing d-ary tree* is a d-ary tree in which the internal nodes are labelled in such a way that the labels increase along any branch stemming from the root and the external nodes are left unlabelled.
There seems to be more results in the literature on recursive trees than on d-ary increasing trees. Either way results on the *height* and *level* of nodes in these trees appear in [14] and [38] respectively. Furthermore results on the Wiener index and total path length of these trees appear in [34].
We denote the class of d-ary increasing trees with n internal vertices by $\mathcal{D}_n$. The root is labelled 1 and each label $k = 2, \dots n$ can be attached in $(d-1)(k-1)+1$ places (after $k-1$ labels have been attached we can attach the next label in any of the $(d-1)(k-1)+1$ places an external node occupied). Hence

$$\begin{aligned} |\mathcal{D}_n| &= \prod_{k=2}^{n} ((d-1)(k-1)+1) \\ &= \prod_{k=1}^{n-1} (k(d-1)+1) \, . \end{aligned}$$

We mention that if $d = 2$ we get the class of binary increasing trees, which frequently appear in applications. There are $n!$ binary increasing trees of order n. The *rotation correspondence* provides a correspondence between recursive trees of order $n+1$ and binary increasing trees of order n. Order the children of any node in a recursive tree increasingly from left to right according to their label values. Remove the root and all connections to children that are not leftmost children. Connect the siblings and perform a 45 degrees rotation, so that the siblings in the recursive tree are now rightmost children in the binary increasing tree. This correspondence is $1-1$ since the process can be reversed.

Returning to the more general class of d-ary increasing trees we adopt the convention $|\mathcal{D}_0| = 1$ to simplify calculations. Notice that any d-ary increasing tree consists of a root appended to a rooted forest, where each of the d (possibly empty) components is a subtree. The remaining labels can be distributed in $\binom{n-1}{r_1, \dots, r_d}$ ways. Summing over all possible sizes of the subtrees yields

$$D_n := |\mathcal{D}_n| = \sum_{r_1+r_2+\ldots+r_d=n-1} \frac{(n-1)!}{r_1! \ldots r_d!} D_{r_1} \ldots D_{r_d}.$$

Hence the exponential generating function

$$D(x) = \sum_{n=0}^{\infty} D_n \frac{x^n}{n!}$$

satisfies $D'(x) = D(x)^d$ with solution

$$D(x) = (1-(d-1)x)^{-\frac{1}{d-1}}.$$

Note that

$$D_n \sim n! \frac{(d-1)^n}{\Gamma\left(\frac{1}{d-1}\right)} n^{\frac{1}{d-1}-1}$$

by singularity analysis. We start by finding the joint probability generating function of the depth and the hitting time to a random vertex.

Proposition 4.2.1. *The joint probability generating function of the depth $d(1,v)$, of a random vertex v in a random d-ary increasing tree T, and the hitting time H_{1v} satisfies the recursion*

$$\begin{aligned}
E_n(u,y) &:= \sum_l \sum_k \mathrm{Pr}_{\mathcal{D}_n,v}\left(d(1,v)=k, H_{1v}=l\right) u^l y^k \\
&= \frac{1}{n\prod_{k=1}^{n-1}\left(k(d-1)+1\right)} \sum_{T\in\mathcal{D}_n, i\in T} u^{H_{1i}} y^{d(1,i)} \\
&= \frac{1}{n} + \frac{d}{(d-1)nA_n} \sum_{r=1}^{n-1} A_r y u^{2(n-r)-1} E_r\left(u, yu^{2(n-r)}\right) \qquad (4.11)
\end{aligned}$$

for $n \in \mathbb{N}$, where $A_n := \frac{D_n}{(n-1)!(d-1)^n}$.

Proof. Let B_0 and B_r be the events that the random vertex v is the root and in a subtree of size r, attached to the root, respectively. By the law of total probability

$$\begin{aligned}
\mathrm{Pr}_{\mathcal{D}_n,v}\left(d(1,v)=k, H_{1v}=l\right) &= \mathrm{Pr}\left(d(1,v)=k, H_{1v}=l | B_0\right)\mathrm{Pr}(B_0) \\
&\quad + \sum_{r=1}^{n-1} \mathrm{Pr}\left(d(1,v)=k, H_{1v}=l | B_r\right)\mathrm{Pr}(B_r),
\end{aligned} \qquad (4.12)$$

where $\mathrm{Pr}(B_0) = 1/n$. Notice that, for subtrees attached to the root,

$$\begin{aligned}
\mathrm{Pr}(B_r) &= \frac{r(\#\text{ subtrees of size } r \in \mathcal{D}_n)}{nD_n} \\
&= \frac{rd(\#\text{ trees in } \mathcal{D}_n \text{ where fixed subtree has size } r)}{nD_n}
\end{aligned}$$

$$= \frac{rd\binom{n-1}{r} D_r (n-r-1)![x^{n-r-1}]D(x)^{d-1}}{nD_n} \tag{4.13}$$
$$= \frac{rd\binom{n-1}{r} D_r (n-r-1)!(d-1)^{n-r-1}}{nD_n}$$
$$= \frac{d}{(d-1)n}\frac{A_r}{A_n}.$$

Now

$$\Pr_{\mathcal{D}_n,v}\left(d(1,v)=k, H_{1v}=l|B_0\right) = \begin{cases} 1 & \text{if } k=l=0 \\ 0 & \text{otherwise .} \end{cases}$$

On the other hand, relabelling the vertices where necessary, we have

$$\Pr_{\mathcal{D}_n,v}\left(d(1,v)=k, H_{1v}=l|B_r\right) = \Pr_{\mathcal{D}_r,v}\left(d(1,v)=k+1, H_{1v}=m\right),$$

where $m = 2(n-r)(k+1)-1+l$.

The result follows from equation (4.12) by working as in the last part of the proof of Theorem 4.1.2.

■

Note in the previous theorem that

$$A_n = n\binom{n-1+\frac{1}{d-1}}{n} \sim \frac{n^{\frac{1}{d-1}}}{\Gamma\left(\frac{1}{d-1}\right)}, \tag{4.14}$$

where the asymptotic equivalence follows from Stirling's formula. We obtain the following result for the probability generating function to the root.

Proposition 4.2.2. *The joint probability generating function of the depth $d(v,1)$, of a random vertex v in a random d-ary increasing tree T, and the hitting time H_{v1} satisfies the recursion*

$$F_n(u,y) := \sum_l \sum_k \Pr_{\mathcal{D}_n,v}\left(d(v,1)=k, H_{v1}=l\right) u^l y^k$$
$$= \frac{1}{n\prod_{k=1}^{n-1}(k(d-1)+1)} \sum_{T\in\mathcal{D}_n, i\in T} u^{H_{i1}} y^{d(i,1)}$$
$$= \frac{1}{n} + \frac{d}{(d-1)nA_n}\sum_{r=1}^{n-1} A_r y u^{2r-1} F_r(u,y) \tag{4.15}$$

for $n \in \mathbb{N}$, where $A_n := \frac{D_n}{(n-1)!(d-1)^n}$.

The proof of the previous proposition follows in the same manner as the derivation of equation (4.7). Now $q_n := \partial_u F_n(u,1)\Big|_{u=1}$ satisfies the recursion

$$q_n = \frac{d}{(d-1)nA_n}\sum_{r=1}^{n-1} A_r(2r-1+q_r).$$

Hence the generating function $Q(x) := \sum_{n=1}^{\infty} A_n q_n x^n$ satisfies the differential equation

$$xQ'(x) - \frac{d}{d-1}\cdot\frac{x}{1-x}Q(x) = \frac{d}{d-1}\sum_{n=1}^{\infty}\sum_{r=1}^{n-1} A_r(2r-1)x^n.$$

Now

$$\frac{d}{dx}\left((1-x)^{\frac{d}{d-1}}Q(x)\right) = \frac{d}{d-1}(1-x)^{\frac{d}{d-1}}\sum_{n=1}^{\infty}\sum_{r=1}^{n-1} A_r(2r-1)x^{n-1}$$

so that

$$Q(x) = \frac{d}{d-1}(1-x)^{-\frac{d}{d-1}}\int_0^x (1-t)^{\frac{d}{d-1}}\sum_{n=1}^{\infty}\sum_{r=1}^{n-1} A_r(2r-1)t^{n-1}dt. \tag{4.16}$$

Substituting

$$\begin{aligned}
&\sum_{n=1}^{\infty}\sum_{r=1}^{n-1} A_r(2r-1)x^{n-1}\\
&= 2\sum_{n=1}^{\infty}\sum_{r=1}^{n-1}\frac{r^2 D_r}{r!(d-1)^r}x^{n-1} - \sum_{n=1}^{\infty}\sum_{r=1}^{n-1}\frac{rD_r}{r!(d-1)^r}x^{n-1}\\
&= \frac{2x}{1-x}\frac{d}{dx}\left(x\frac{d}{dx}(1-x)^{-\frac{1}{d-1}}\right) - \frac{x}{1-x}\frac{d}{dx}(1-x)^{-\frac{1}{d-1}}\\
&= \frac{x}{1-x}\cdot\frac{1}{d-1}(1-x)^{-\frac{d}{d-1}} + \frac{2d}{(d-1)^2}\frac{x^2}{1-x}(1-x)^{\frac{1-2d}{d-1}}
\end{aligned}$$

into (4.16) gives

$$\begin{aligned}
&Q(x)\\
&= \frac{d}{d-1}(1-x)^{-\frac{d}{d-1}}\int_0^x\left(\frac{t}{(d-1)(1-t)} + \frac{2d}{(d-1)^2}\left(\frac{t}{1-t}\right)^2\right)dt\\
&= \frac{d}{d-1}(1-x)^{-\frac{d}{d-1}}\left[\frac{1}{d-1}(-x-\log(1-x)) + \frac{2d}{(d-1)^2}\left(x+\frac{1}{1-x}+2\log(1-x)\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{2d}{(d-1)^2}\Bigg] \\
&= \frac{d}{d-1}(1-x)^{-\frac{d}{d-1}}\left(\frac{2d}{(d-1)^2}\cdot\frac{1}{1-x} + \frac{d+1}{(d-1)^2}x + \frac{3d+1}{(d-1)^2}\log(1-x) - \frac{2d}{(d-1)^2}\right).
\end{aligned}
\tag{4.17}
$$

Differentiating both sides of

$$
e^{(z+1)\log(1-x)^{-1}} = (1-x)^{-z-1} = \sum_n \binom{z+n}{n} x^n
$$

w.r.t. z yields

$$
-(1-x)^{-z-1}\log(1-x) = \sum_n \binom{z+n}{n}\left(\frac{1}{z+n} + \frac{1}{z+n-1} + \ldots + \frac{1}{z+1}\right)x^n.
\tag{4.18}
$$

Hence

$$
\begin{aligned}
q_n &= \frac{(n-1)!(d-1)^n}{\prod_{k=1}^{n-1}((d-1)k+1)}\cdot\frac{d}{d-1}\Bigg[\frac{d+1}{(d-1)^2}(-1)^{n-1}\binom{-\frac{d}{d-1}}{n-1} + \frac{2d}{(d-1)^2}(-1)^{n-1}\binom{\frac{1-2d}{d-1}}{n-1} \\
&\quad - \frac{3d+1}{(d-1)^2}\binom{\frac{1}{d-1}+n}{n}\left(\frac{1}{\frac{1}{d-1}+n} + \ldots + \frac{1}{\frac{1}{d-1}+1}\right)\Bigg] \\
&= \frac{1}{n\binom{\frac{1}{d-1}+n-1}{n}}\cdot\frac{d}{(d-1)^3}\Bigg[(d+1)\binom{\frac{d}{d-1}+n-2}{n-1} + 2d\binom{\frac{2d-1}{d-1}+n-2}{n-1} \\
&\quad - (3d+1)\binom{\frac{1}{d-1}+n}{n}\left(\frac{1}{\frac{1}{d-1}+n} + \ldots + \frac{1}{\frac{1}{d-1}+1}\right)\Bigg] \\
&= \frac{d(d+1)}{(d-1)^2} + \frac{2d}{d-1}\left(n + \frac{1}{d-1}\right) - \frac{d(3d+1)(1+(d-1)n)}{n(d-1)^2}\sum_{k=1}^{n}\frac{1}{(d-1)k+1} \\
&= \frac{2d}{d-1}n - \frac{d(3d+1)(1+(d-1)n)}{n(d-1)^2}\sum_{k=1}^{n}\frac{1}{(d-1)k+1} + \frac{d(d+3)}{(d-1)^2}.
\end{aligned}
\tag{4.19}
$$

We state this result as:

Proposition 4.2.3. *The expected hitting time from a random vertex v to the root in a random d-ary increasing tree of order n is*

$$
\begin{aligned}
\mathbb{E}\left(H_{v1}\right) &= \frac{2d}{d-1}n - \frac{d(3d+1)(1+(d-1)n)}{n(d-1)^2}\sum_{k=1}^{n}\frac{1}{(d-1)k+1} + \frac{d(d+3)}{(d-1)^2} \\
&= \frac{2d}{d-1}n + \mathcal{O}\left(\log n\right).
\end{aligned}
$$

We require the following lemma to find the expected hitting time from the root.

Lemma 4.2.4. *The expected depth of a random vertex v in a random d-ary increasing tree of order n is*

$$\begin{aligned}\mathbb{E}\left(d(1,v)\right) &= d\sum_{k=1}^{n}\frac{1}{(d-1)k+1} - \frac{d}{d-1} + \frac{d}{n(d-1)}\sum_{k=1}^{n}\frac{1}{(d-1)k+1} \\ &= \frac{d}{d-1}\log n - \frac{d}{d-1}\left(1+\frac{\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)}\right) + o(1).\end{aligned}$$

Proof. It follows from (4.15) that

$$g_n := \partial_y F_n(1,y)\Big|_{y=1} = \mathbb{E}\left(d(1,v)\right)$$

satisfies the recursion

$$nA_ng_n - \frac{d}{d-1}\sum_{r=1}^{n-1}A_rg_r = \frac{d}{d-1}\sum_{r=1}^{n-1}A_r,$$

where $A_n := \frac{D_n}{(n-1)!(d-1)^n}$. Working exactly as we did to derive (4.17) we find

$$\begin{aligned}G(x) &:= \sum_{n=1}^{\infty}A_ng_nx^n \\ &= (1-x)^{-\frac{d}{(d-1)}}\frac{d}{(d-1)^2}\left(-\log(1-x)-x\right). \qquad (4.20)\end{aligned}$$

Equating the coefficients of x^n yields

$$\begin{aligned}\mathbb{E}\left(d(1,v)\right) &= \frac{1}{n\binom{\frac{1}{d-1}+n-1}{n}}\cdot\frac{d}{(d-1)^2}\left[-\binom{\frac{d}{d-1}+n-2}{n-1}\right. \\ &\quad \left. + \binom{\frac{1}{d-1}+n}{n}\left(\frac{1}{\frac{1}{d-1}+n}+\dots+\frac{1}{\frac{1}{d-1}+1}\right)\right] \\ &= -\frac{d}{d-1} + \frac{d(1+(d-1)n)}{n(d-1)}\sum_{k=1}^{n}\frac{1}{(d-1)k+1} \\ &= \frac{d}{d-1}\sum_{k=1}^{n}\frac{1}{k+\frac{1}{d-1}} - \frac{d}{d-1} + \frac{d}{n(d-1)}\sum_{k=1}^{n}\frac{1}{(d-1)k+1}.\end{aligned}$$

It follows from equation (4.18) that

$$\binom{n+\frac{1}{d-1}}{n}\sum_{k=1}^{n}\frac{1}{k+\frac{1}{d-1}} = [x^n](1-x)^{-\frac{1}{d-1}-1}\left(-\log(1-x)\right),$$

where singularity analysis yields

$$\binom{n+\frac{1}{d-1}}{n} = [x^n](1-x)^{-\frac{d}{d-1}} \sim \frac{n^{\frac{d}{d-1}-1}}{\Gamma\left(\frac{d}{d-1}\right)}$$

and

$$[x^n](1-x)^{-\frac{1}{d-1}-1}\left(-\log(1-x)\right) = \frac{n^{\frac{d}{d-1}-1}}{\Gamma\left(\frac{d}{d-1}\right)}\log n\left(1-\frac{\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)\log n}\right)+o(1)$$

so that the result follows. ∎

We have

$$\begin{aligned}
\mathbb{E}\left(H_{1v}\right) &= 2(n-1)\mathbb{E}\left(d(1,v)\right)-\mathbb{E}\left(H_{v1}\right)\\
&= \frac{d(1+(d-1)n)}{n(d-1)}\sum_{k=1}^{n}\frac{1}{(d-1)k+1}\left(2(n-1)+\frac{3d+1}{d-1}\right)-\frac{4dn}{d-1}+\frac{d(d-5)}{(d-1)^2}\\
&\sim \frac{2d\left(1+(d-1)n\right)}{d-1}\sum_{k=1}^{n}\frac{1}{(d-1)k+1}-\frac{4dn}{d-1}\\
&\sim 2dn\left(\frac{\log n}{d-1}-\frac{1}{d-1}\cdot\frac{\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)}\right)-\frac{4dn}{d-1}\\
&\sim \frac{2d}{d-1}n\log n-\frac{2d}{d-1}\left(2+\frac{\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)}\right)n.
\end{aligned}$$

We state this result as:

Proposition 4.2.5. *The expected hitting time from the root to a random vertex v in a random d-ary increasing tree of order n is*

$$\begin{aligned}
\mathbb{E}\left(H_{1v}\right) &= \frac{d(1+(d-1)n)}{n(d-1)}\sum_{k=1}^{n}\frac{1}{(d-1)k+1}\left(2(n-1)+\frac{3d+1}{d-1}\right)-\frac{4dn}{d-1}+\frac{d(d-5)}{(d-1)^2}\\
&\sim \frac{2d}{d-1}n\log n-\frac{2d}{d-1}\left(2+\frac{\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)}\right)n.
\end{aligned}$$

We proceed to find the limiting distribution of the hitting time to the root.

Theorem 4.2.6. *The s-th moment of the hitting time from a random vertex v to the root in a random d-ary increasing tree of order n satisfies the asymptotic formula*

$$\mathbb{E}(H_{v1}^s) \sim C_s n^s,$$

where the coefficients satisfy the recursion

$$C_s = \frac{d}{d-1} \cdot \frac{1}{s} \sum_{k=0}^{s-1} \binom{s}{k} 2^{s-k} C_k, \quad C_0 = 1,$$

with exponential generating function

$$\begin{aligned} C(x) &:= \sum_{s=0}^{\infty} C_s \frac{x^s}{s!} \\ &= \exp\left(\frac{d}{d-1} \sum_{n=1}^{\infty} \frac{(2x)^n}{n!n} \right). \end{aligned}$$

Consequently, the normalised random variable $\frac{H_{v1}}{2n}$ *converges weakly to the generalised Dickman distribution* $\mathrm{GD}\left(\frac{d}{d-1}\right)$.

Proof. Applying the Leibniz product rule to (4.15) yields

$$q_{n,s} = \frac{d}{(d-1)nA_n} \sum_{r=1}^{n-1} A_r \sum_{k=0}^{s} \binom{s}{k} q_{r,k} \prod_{i=1}^{s-k} (2r-i) \tag{4.21}$$

where $A_n := \frac{D_n}{(n-1)!(d-1)^n}$ and $q_{n,s} := \partial_u^s F_n(u,1)\Big|_{u=1}$. It follows from (4.17) that

$$Q_1(x) \underset{x\to 1}{=} \frac{2d^2}{(d-1)^3}(1-x)^{-\frac{2d-1}{d-1}} + \mathcal{O}\left((1-x)^{-\frac{d}{d-1}} \log(1-x)\right)$$

where $Q_s(x) := \sum_{n=1}^{\infty} A_n q_{n,s} x^n$. We suppose

$$Q_k(x) \underset{x\to 1}{=} \frac{L_k}{(1-x)^{\frac{(k+1)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{k(d-1)+1}{d-1}}} \right)$$

for $k = 2, \dots, s-1$ and prove this result by induction for $Q_s(x)$. It follows from (4.21) that

$$\sum_{n=1}^{\infty} nA_n q_{n,s} x^n - \frac{d}{d-1} \sum_{n=1}^{\infty} \sum_{r=1}^{n-1} A_r q_{r,s} x^n = \frac{d}{d-1} \sum_{n=1}^{\infty} \sum_{r=1}^{n-1} A_r \sum_{k=0}^{s-1} \binom{s}{k} q_{r,k} \prod_{i=1}^{s-k} (2r-i) x^n. \tag{4.22}$$

If $k = 2, \dots, s-1$ we have

$$\sum_{n=1}^{\infty} A_n q_{n,k} n^{s-k} x^n \underset{x\to 1}{=} (x\partial_x)^{s-k} \left(\frac{L_k}{(1-x)^{\frac{(k+1)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{k(d-1)+1}{d-1}}} \right) \right)$$

$$= \frac{L_k \prod_{i=k+1}^{s} \frac{i(d-1)+1}{d-1}}{(1-x)^{\frac{(s+1)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{s(d-1)+1}{d-1}}} \right),$$

implying

$$\sum_{n=1}^{\infty} \sum_{r=1}^{n-1} A_r q_{r,k} r^{s-k} x^n = \frac{x}{1-x} \left(\frac{L_k \prod_{i=k+1}^{s} \frac{i(d-1)+1}{d-1}}{(1-x)^{\frac{(s+1)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{s(d-1)+1}{d-1}}} \right) \right).$$

It now follows from (4.22) that

$$Q_s'(x) - \frac{d}{d-1} \cdot \frac{Q_s(x)}{1-x} = \frac{d}{d-1} \sum_{k=0}^{s-1} 2^{s-k} \binom{s}{k} \frac{L_k \prod_{i=k+1}^{s} \frac{i(d-1)+1}{d-1}}{(1-x)^{\frac{(s+2)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{(s+1)(d-1)+1}{d-1}}} \right)$$

implying

$$\frac{d}{dx}\left((1-x)^{\frac{d}{(d-1)}} Q_s(x) \right) = \frac{d}{d-1} \sum_{k=0}^{s-1} 2^{s-k} \binom{s}{k} \frac{L_k \prod_{i=k+1}^{s} \frac{i(d-1)+1}{d-1}}{(1-x)^{s+1}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{s}} \right).$$

Now

$$\begin{aligned}
&Q_s(x) \\
&= (1-x)^{-\frac{d}{(d-1)}} \frac{d}{d-1} \sum_{k=0}^{s-1} 2^{s-k} \binom{s}{k} \frac{L_k \prod_{i=k+1}^{s} \frac{i(d-1)+1}{d-1}}{s(1-x)^{s}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{s-1}} \right) \\
&= \frac{d}{s(d-1)} \frac{\sum_{k=0}^{s-1} 2^{s-k} \binom{s}{k} L_k \prod_{i=k+1}^{s} \frac{i(d-1)+1}{d-1}}{(1-x)^{\frac{(s+1)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{s(d-1)+1}{d-1}}} \right) \\
&= \frac{L_s}{(1-x)^{\frac{(s+1)(d-1)+1}{d-1}}} + \mathcal{O}\left(\frac{\log(1-x)}{(1-x)^{\frac{s(d-1)+1}{d-1}}} \right),
\end{aligned}$$

completing the induction proof. Singularity analysis of $Q_s(x)$ gives

$$\begin{aligned}
\mathbb{E}(H_{v1}^s) &\sim q_{n,s} \\
&\sim \frac{\Gamma\left(\frac{1}{d-1}\right)}{n^{\frac{1}{d-1}}} \cdot \frac{L_s}{\Gamma\left(\frac{(s+1)(d-1)+1}{d-1}\right)} n^{\frac{s(d-1)+1}{d-1}} \\
&= \frac{L_s}{\prod_{i=0}^{s} \frac{i(d-1)+1}{d-1}} n^s,
\end{aligned}$$

where the coefficients $C_j := \frac{L_j}{\prod_{i=0}^{j} \frac{i(d-1)+1}{d-1}}$ satisfy the recursion

$$C_s = \frac{d}{d-1} \cdot \frac{1}{s} \sum_{k=0}^{s-1} \binom{s}{k} 2^{s-k} C_k, \quad C_0 = 1.$$

The exponential generating function $C(x) := \sum_{n=0}^{\infty} C_n \frac{x^n}{n!}$ satisfies

$$xC'(x) + \frac{d}{d-1} C(x) = \frac{d}{d-1} C(x) e^{2x}$$

so that

$$\frac{C'(x)}{C(x)} = \frac{d}{d-1} \left(\frac{e^{2x} - 1}{x} \right)$$

and hence since $C(0) = 1$,

$$\begin{aligned} C(x) &= \exp\left(\frac{d}{d-1} \int_0^x \frac{e^{2t} - 1}{t} dt \right) \\ &= \exp\left(\frac{d}{d-1} \sum_{n=1}^{\infty} \frac{(2x)^n}{n!n} \right). \end{aligned}$$

Finally the coefficients

$$K_s := \frac{C_s}{2^s} = \frac{d}{d-1} \cdot \frac{1}{s} \sum_{k=0}^{s-1} \binom{s}{k} K_k$$

have the exponential generating function

$$\begin{aligned} K(x) &= \exp\left(\frac{d}{d-1} \sum_{n=1}^{\infty} \frac{x^n}{n!n} \right) \\ &= \exp\left(\frac{d}{d-1} \int_0^x \frac{e^t - 1}{t} dt \right), \end{aligned}$$

equivalent to the moment generating function of the generalised Dickman distribution GD $\left(\frac{d}{d-1}\right)$ (see [39]). ■

Before we find the limiting distribution of the hitting time from the root, we find the covariances of the hitting time from and to the root with the depth respectively. Our strategy is to stay on the level of generating functions this time. Let

$$F(x, u, y) = \sum_{n=1}^{\infty} A_n F_n(u, y) x^n \tag{4.23}$$

be the generating function where x marks the joint probability distribution in (4.15). We set

$$F'(x,u,y) = \frac{d}{dx}F(x,u,y),$$
$$F(x) = F(x,u,y)\big|_{u=y=1}$$

and

$$F(x,u) = F(x,u,y)\big|_{y=1}.$$

Rearranging the terms in (4.15) we find

$$\sum_{n=1}^{\infty} nA_nF_n(u,y)x^n = \sum_{n=1}^{\infty} A_nx^n + \frac{d}{d-1}\sum_{n=1}^{\infty}\sum_{r=1}^{n-1} yu^{2r-1}A_rF_r(u,y)x^n,$$

so that

$$\begin{aligned} xF'(x,u,y) &= x\frac{d}{dx}(1-x)^{-\frac{1}{d-1}} + \frac{yd}{u(d-1)}\cdot\frac{x}{1-x}F(xu^2,u,y) \\ &= \frac{x}{d-1}(1-x)^{-\frac{d}{d-1}} + \frac{yd}{u(d-1)}\cdot\frac{x}{1-x}F(xu^2,u,y). \end{aligned} \tag{4.24}$$

Now

$$x\partial_y F'(x,u,y)\big|_{y=1} = \frac{dx}{u(d-1)(1-x)}F(xu^2,u,1) + \frac{dx}{u(d-1)(1-x)}\partial_y F(xu^2,u,y)\big|_{y=1}$$

implies

$$\begin{aligned} &x\partial_u\left(\partial_y F'(x,u,y)\big|_{y=1}\right)\Bigg|_{u=1} \\ &= -\frac{dx}{u^2(d-1)(1-x)}F(xu^2,u)\Bigg|_{u=1} + \frac{dx}{u(d-1)(1-x)}\left(\partial_u F(x,u) + 2uxF'(x,u)\right)\Bigg|_{u=1} \\ &\quad - \frac{dx}{u^2(d-1)(1-x)}\partial_y F(xu^2,u,y)\big|_{y=1}\Bigg|_{u=1} + \frac{dx}{u(d-1)(1-x)}\Bigg(\partial_u\left(\partial_y F(x,u,y)\big|_{y=1}\right) \\ &\quad + 2ux\partial_x\left(\partial_y F(x,u,y)\big|_{y=1}\right)\Bigg)\Bigg|_{u=1}. \end{aligned} \tag{4.25}$$

Setting

$$F_{yu}(x) = \partial_u\left(\partial_y F(x,u,y)\big|_{y=1}\right)\Bigg|_{u=1}$$

we obtain from equation (4.25)

$$F'_{yu}(x) - \frac{d}{(d-1)(1-x)}F_{yu}(x) = -\frac{d\left(F(x)-F_u(x)\right)}{(d-1)(1-x)} + \frac{2dx}{(d-1)(1-x)}F'(x) - \frac{d}{(d-1)(1-x)}F_y(x) + \frac{2dx}{(d-1)(1-x)}F'_y(x),$$

where

$$F'(x) = \frac{d}{dx}\left(\frac{x}{d-1}(1-x)^{-\frac{d}{d-1}}\right) = \frac{1}{d-1}(1-x)^{-\frac{d}{d-1}} + \frac{dx}{(d-1)^2}(1-x)^{-\frac{d}{d-1}-1},$$

$F_u(x) = Q(x)$ in (4.17) and

$$F_y(x) = \frac{d}{(d-1)^2}(1-x)^{-\frac{d}{(d-1)}}\left(-\log(1-x)-x\right).$$

Hence

$$\begin{aligned}
&\frac{d}{dx}\left((1-x)^{\frac{d}{d-1}}F_{yu}(x)\right)\\
&= \frac{d}{d-1}(1-x)^{\frac{d}{d-1}}\Bigg[-\frac{x}{(1-x)(d-1)}(1-x)^{-\frac{d}{d-1}} + \frac{Q(x)}{1-x}\\
&\quad + \frac{2x}{1-x}\left(\frac{1}{d-1}(1-x)^{-\frac{d}{d-1}} + \frac{dx}{(d-1)^2}(1-x)^{-\frac{d}{d-1}-1}\right)\\
&\quad - \frac{d}{(d-1)^2}(1-x)^{-\frac{d}{d-1}-1}\left(-\log(1-x)-x\right)\\
&\quad + \frac{2dx}{(1-x)(d-1)^2}\left(\frac{x(1-x)^{-\frac{d}{d-1}}}{1-x} + \frac{d(1-x)^{-\frac{d}{d-1}-1}(-x-\log(1-x))}{d-1}\right)\Bigg]\\
&= \frac{d}{d-1}\Bigg[\frac{x}{1-x}\left(-\frac{1}{d-1} + \frac{d(d+1)}{(d-1)^3} + \frac{2}{d-1} + \frac{d}{(d-1)^2}\right)\\
&\quad + \frac{1}{(1-x)^2}\left(\frac{2d^2}{(d-1)^3} - \frac{2d^2x^2}{(d-1)^3} + \frac{2dx^2}{(d-1)^2} + \frac{2dx^2}{(d-1)^2}\right)\\
&\quad + \frac{\log(1-x)}{1-x}\left(\frac{d(3d+1)}{(d-1)^3} + \frac{d}{(d-1)^2}\right) - \frac{2d^2}{(1-x)(d-1)^3} - \frac{2d^2x\log(1-x)}{(d-1)^3(1-x)^2}\Bigg],
\end{aligned}$$

so that

$$\begin{aligned}
F_{yu}(x) = \frac{d}{d-1}(1-x)^{-\frac{d}{d-1}}\Bigg[&\frac{2d^2}{(d-1)^3}\left(\frac{-2\log(1-x)+(x-1)\log^2(1-x)-2}{2(1-x)}+1\right)\\
&+ \frac{2d^2x}{(d-1)^3(1-x)} + \left(\frac{4d}{(d-1)^2} - \frac{2d^2}{(d-1)^3}\right)\left(x+(1-x)^{-1}+2\log(1-x)-1\right)
\end{aligned}$$

$$-\frac{1}{2}\left(\frac{d(3d+1)}{(d-1)^3}+\frac{d}{(d-1)^2}\right)\log^2(1-x)$$
$$\left.-(x+\log(1-x))\left(\frac{1}{d-1}+\frac{d(d+1)}{(d-1)^3}+\frac{d}{(d-1)^2}\right)+\frac{2d^2}{(d-1)^3}\log(1-x)\right]$$
$$\underset{x\to 1}{=}\frac{2d^3}{(d-1)^4}(1-x)^{-\frac{d}{d-1}-1}(-\log(1-x))+\frac{2d^2(d-2)}{(d-1)^4}(1-x)^{-\frac{d}{d-1}-1}$$
$$+\mathcal{O}\left(\log^2(1-x)(1-x)^{-\frac{d}{d-1}}\right). \tag{4.26}$$

Singularity analysis of (4.26) gives

$$[x^n]F_{yu}(x)\sim\frac{2d^3}{\Gamma\left(\frac{d}{d-1}+1\right)(d-1)^4}n^{\frac{d}{d-1}}\log n$$
$$+\left(\frac{2d^2(d-2)}{(d-1)^4}-\frac{2d^3\Gamma'\left(\frac{d}{d-1}+1\right)}{\Gamma\left(\frac{d}{d-1}+1\right)(d-1)^4}\right)\frac{n^{\frac{d}{d-1}}}{\Gamma\left(\frac{d}{d-1}+1\right)}.$$

Recalling that

$$A_n=\frac{n^{\frac{1}{d-1}}}{\Gamma\left(\frac{1}{d-1}\right)}\left(1+\mathcal{O}(1/n)\right),$$

it follows that

$$\mathbb{E}\left(H_{v1}d(v,1)\right)\sim\frac{2d^2}{(d-1)^2}n\log n+\left(\frac{2d(d-2)}{(d-1)^2}-\frac{2d^2\Gamma'\left(\frac{d}{d-1}+1\right)}{\Gamma\left(\frac{d}{d-1}+1\right)(d-1)^2}\right)n$$

and as a consequence of Proposition 4.2.3 and Lemma 4.2.4 we have

$$\operatorname{cov}\left(d(v,1),H_{v1}\right)\sim\left(\frac{2d(d-2)}{(d-1)^2}+\frac{2d^2}{(d-1)^2}\right)n=\frac{4d}{d-1}n.$$

We state this result as:

Proposition 4.2.7. *The covariance of the hitting time from a random vertex v to the root and the distance from v to the root, in a random d-ary increasing tree of order n, is*

$$\operatorname{cov}\left(d(v,1),H_{v1}\right)\sim\frac{4d}{d-1}n.$$

We now require the second moment of the depth to find

$$\mathbb{E}\left(H_{1v}d(v,1)\right)=2(n-1)\mathbb{E}\left(d(v,1)^2\right)-\mathbb{E}\left(H_{v1}d(v,1)\right). \tag{4.27}$$

Taking the partial derivative w.r.t. y twice in (4.24) and setting $u=y=1$ gives

$$xF'_{yy}(x)-\frac{dx}{(d-1)(1-x)}F_{yy}(x)=\frac{2dx}{(d-1)(1-x)}F_y(x).$$

Solving for $F_{yy}(x)$ yields

$$\begin{aligned} F_{yy}(x) &= \frac{2d^2}{(d-1)^3}(1-x)^{-\frac{d}{d-1}} \int_0^x (1-t)^{-1}(-\log(1-t)-t)dt \\ &= \frac{2d^2}{(d-1)^3}(1-x)^{-\frac{d}{d-1}} \left(\frac{1}{2}\log^2(1-x)+\log(1-x)+x\right). \end{aligned} \tag{4.28}$$

Singularity analysis of (4.28) gives

$$\begin{aligned} [x^n]F_{yy}(x) \sim & \frac{d^2}{\Gamma\left(\frac{d}{d-1}\right)(d-1)^3} n^{\frac{1}{d-1}} \log^2 n \left(1 - \frac{2\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)} \cdot \frac{1}{\log n}\right) \\ & - \frac{2d^2}{\Gamma\left(\frac{d}{d-1}\right)(d-1)^3} n^{\frac{1}{d-1}} \log n + \mathcal{O}\left(n^{\frac{1}{d-1}}\right), \end{aligned}$$

implying

$$\begin{aligned} & \mathbb{E}\left(d(v,1)(d(v,1)-1)\right) \\ & = \frac{d^2}{(d-1)^2}\log^2 n \left(1 - \frac{2\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)} \cdot \frac{1}{\log n}\right) - \frac{2d^2}{(d-1)^2}\log n + \mathcal{O}(1). \end{aligned}$$

It now follows from Lemma 4.2.4 that

$$\mathrm{Var}\left(d(v,1)\right) \sim \frac{d}{d-1}\log n. \tag{4.29}$$

We also obtain, from (4.27)

$$\begin{aligned} \mathbb{E}\left(H_{1v}d(v,1)\right) = & \frac{2d^2}{(d-1)^2} n\log^2 n + 2n\log n \left(-\frac{2d^2\Gamma'\left(\frac{d}{d-1}\right)}{(d-1)^2\Gamma\left(\frac{d}{d-1}\right)} - \frac{3d^2}{(d-1)^2} + \frac{d}{d-1}\right) \\ & + o(n\log n) \end{aligned}$$

and since

$$\begin{aligned} & \mathbb{E}\left(H_{1v}\right)\mathbb{E}\left(d(v,1)\right) \\ & = \frac{2d^2}{(d-1)^2} n\log^2 n - \frac{2d^2}{(d-1)^2}\left(3 + \frac{2\Gamma'\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{d}{d-1}\right)}\right) n\log n + o(n\log n) \end{aligned}$$

we have $\mathrm{cov}\left(d(v,1), H_{1v}\right) \sim \frac{2d}{d-1} n\log n$. We state this result as:

Proposition 4.2.8. *The covariance of the hitting time from the root to a random vertex v and the distance of the root to v, in a random d-ary increasing tree of order n, is*

$$\mathrm{cov}\left(d(v,1), H_{1v}\right) \sim \frac{2d}{d-1} n\log n.$$

Our next step is to find the limiting distribution of the hitting time from the root. We start by finding the limiting distribution of the depth. We can solve for $F(x,y)$ by setting $u=1$ in (4.24):

$$\begin{aligned} F(x,y) &= \frac{(1-x)^{-\frac{dy}{d-1}}}{d-1}\int_0^x (1-t)^{\frac{d(y-1)}{d-1}}dt \\ &= \frac{(1-x)^{-\frac{dy}{d-1}}-(1-x)^{-\frac{1}{d-1}}}{dy-1}. \end{aligned}$$

It follows from singularity analysis that

$$[x^n](1-x)^{-\frac{1}{d-1}} = \mathcal{O}\left(\frac{1}{\sqrt{n}}[x^n](1-x)^{-\frac{dy}{d-1}}\right),$$

for y sufficiently close to 1. Hence

$$[x^n]F(x,y) = \frac{1}{dy-1}[x^n](1-x)^{-\frac{dy}{d-1}}\left(1+\mathcal{O}\left(n^{-\frac{1}{2}}\right)\right)$$

and it follows from [17, Theorem IX.11] that

$$\frac{d(1,v)-\frac{d}{d-1}\log n}{\sqrt{\frac{d}{d-1}\log n}} \xrightarrow{d} N(0,1).$$

In particular

$$\frac{2nd(1,v)-\frac{2d}{d-1}n\log n}{2n\sqrt{\frac{d}{d-1}\log n}} \xrightarrow{d} N(0,1). \tag{4.30}$$

Now

$$\frac{H_{1v}-\frac{2d}{d-1}n\log n}{2n\sqrt{\frac{d}{d-1}\log n}} = \frac{2nd(1,v)-\frac{2d}{d-1}n\log n}{2n\sqrt{\frac{d}{d-1}\log n}} - \frac{H_{v1}+2d(1,v)}{2\sqrt{\frac{d}{d-1}n\log n}},$$

where it follows from Proposition 4.2.3 that

$$-\frac{H_{v1}+2d(1,v)}{2\sqrt{\frac{d}{d-1}}n\sqrt{\log n}} \xrightarrow{P} 0.$$

Finally it follows from Slutsky's theorem (see [21, Theorem 2.39]) that

$$\frac{H_{1v}-\frac{2d}{d-1}n\log n}{2n\sqrt{\frac{d}{d-1}\log n}} \xrightarrow{d} N(0,1).$$

We state this result as:

Theorem 4.2.9. *The normalised hitting time from the root to a random vertex v, in a random d-ary increasing tree of size n, converges weakly to the standard normal distribution:*

$$\frac{H_{1v} - \frac{2d}{d-1} n \log n}{2n\sqrt{\frac{d}{d-1}\log n}} \xrightarrow{d} N(0,1).$$

We also have:

Theorem 4.2.10. *The normalised hitting time between two random vertices v and w in a random d-ary increasing tree of size n converges weakly to the standard normal distribution:*

$$\frac{H_{vw} - \frac{2d}{d-1} n \log n}{2n\sqrt{\frac{d}{d-1}\log n}} \xrightarrow{d} N(0,1).$$

Proof. The proof follows in the same way as that of Theorem 4.1.7 since a similar result as Theorem 4.1.6 (see [25, Theorem 7]) holds for d-ary increasing trees. ■

4.3 Generalised plane oriented recursive trees

A *plane oriented recursive tree*, also called a PORT, is a rooted plane tree in which the internal nodes are labelled in such a way that the labels increase along any branch stemming from the root. Results on the depth and path length of PORTs are known and appear in [29] and [40].
We denote the class of PORTs by

$$\mathcal{P} = \cup_{n\geq 1}\mathcal{P}_n,$$

where $\mathcal{P}_n$ is the class of PORTs of order n, $|\mathcal{P}_0| = 0$ and $|\mathcal{P}_1| = 1$. Notice that $P_n := |\mathcal{P}_n| = (2n-3)!!$ for $n \geq 2$. Indeed each vertex with label $k = 2, \dots, n$ can be attached in $2k-3$ places: We can attach k to $k-1$ possible vertices, where the sum of the out-degrees of these vertices is $k-2$. Moreover if a vertex has out-degree $l \geq 0$, a new vertex can be attached to the vertex in $l+1$ ways. Hence there are $k-1+k-2 = 2k-3$ places to insert k. Multiplying over all values of k gives us the result for P_n. The result for P_n can alternatively be derived from the exponential generating function

$$P(x) := \sum_{n=1}^{\infty} P_n \frac{x^n}{n!}$$

after noticing

$$P_n = \sum_{j=1}^{n-1} \sum_{\substack{r_1+\ldots+r_j=n-1 \\ r_1,\ldots,r_j \geq 1}} \frac{(n-1)!}{r_1!\ldots r_j!} P_{r_1}\ldots P_{r_j} \tag{4.31}$$

for $n \geq 2$, which holds since any PORT of size n consists of a root appended to a rooted forest, where each component is a PORT. Now

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{P_n}{(n-1)!} x^{n-1} &= 1 + \sum_{n=2}^{\infty} \sum_{j=1}^{n-1} \sum_{\substack{r_1+\ldots+r_j=n-1 \\ r_1,\ldots,r_j \geq 1}} \frac{1}{r_1!\ldots r_j!} P_{r_1}\ldots P_{r_j} x^{n-1} \\ &= 1 + \sum_{n=1}^{\infty} \sum_{j=1}^{n} \sum_{\substack{r_1+\ldots+r_j=n \\ r_1,\ldots,r_j \geq 1}} \frac{1}{r_1!\ldots r_j!} P_{r_1}\ldots P_{r_j} x^{n},\end{aligned}$$

implying

$$P'(x) = \frac{1}{1-P(x)}.$$

Solving for $P(x)$ gives

$$\int_0^x P'(t)(1-P(t))dt = x, \tag{4.32}$$

so that

$$\begin{aligned}P(x) &= 1 - \sqrt{1-2x} \\ &= \sum_{n=1}^{\infty} \frac{(2n-3)!!}{n!} x^n.\end{aligned}$$

Notice that

$$(2n-3)!! = n!2^{1-n}C_n,$$

where C_n is the shifted Catalan number

$$C_n := \frac{1}{n}\binom{2n-2}{n-1}.$$

More generally we can assign a weight

$$w(T) = \prod_{v\in T} \phi_{d^+(v)} = \prod_{j\geq 0} \phi_j^{D_j(T)},$$

for $T \in \mathcal{P}$, where $d^+(v)$ denotes the out-degree of vertex v and $D_j(T)$ the number of vertices in T with j children. We now have the probability distribution

$$\pi_n(T) = \frac{w(T)}{\sum_{T\in\mathcal{P}_n} w(T)}$$

on the class of PORTs of order n. A *generalised plane oriented recursive tree*, also called a GPORT, is a weighted PORT, where the vertices have the weight sequence

$$\phi_j = \binom{\alpha + j - 1}{j}.$$

We denote the class of GPORTs by

$$\mathcal{G} = \cup_{n \geq 1} \mathcal{G}_n,$$

where $\mathcal{G}_n$ is the class of GPORTs of order n. The case $\alpha = 1$ assigns a weight of 1 to each tree so that the class of GPORTs reduces to the class of PORTs in this case. Let

$$G_n = \sum_{T \in \mathcal{P}_n} \prod_{j \geq 1} \binom{\alpha + j - 1}{j}^{D_j(T)}$$

be the sum of the weights of GPORTs of order n, with associated exponential generating function

$$G(x) := \sum_{n=1}^{\infty} G_n \frac{x^n}{n!}.$$

It follows as in (4.31) that

$$G_n = \sum_{j=1}^{n-1} \binom{\alpha + j - 1}{j} \sum_{\substack{r_1 + \ldots + r_j = n-1 \\ r_1, \ldots, r_j \geq 1}} \frac{(n-1)!}{r_1! \ldots r_j!} G_{r_1} \ldots G_{r_j} \tag{4.33}$$

for $n \geq 2$, implying

$$\begin{aligned}
G'(x) &= 1 + \sum_{n=2}^{\infty} \sum_{j=1}^{n-1} \binom{\alpha + j - 1}{j} \sum_{\substack{r_1 + \ldots + r_j = n-1 \\ r_1, \ldots, r_j \geq 1}} \frac{1}{r_1! \ldots r_j!} G_{r_1} \ldots G_{r_j} x^{n-1} \\
&= 1 + \sum_{n=1}^{\infty} \sum_{j=1}^{n} \binom{\alpha + j - 1}{j} \sum_{\substack{r_1 + \ldots + r_j = n \\ r_1, \ldots, r_j \geq 1}} \frac{1}{r_1! \ldots r_j!} G_{r_1} \ldots G_{r_j} x^{n} \\
&= \sum_{n=0}^{\infty} \binom{\alpha + n - 1}{n} G(x)^n \\
&= (1 - G(x))^{-\alpha}.
\end{aligned}$$

Solving for $G(x)$, as in (4.32) gives,

$$G(x) = 1 - (1 - (\alpha + 1)x)^{\frac{1}{\alpha+1}}. \tag{4.34}$$

In particular the sum of the weights of GPORTs of size n are

$$\begin{aligned} G_n &= n!(-1)^{n-1}(\alpha+1)^n \binom{\frac{1}{\alpha+1}}{n} \\ &= \prod_{k=1}^{n-1} (k(\alpha+1)-1) \\ &\sim -n!\,(\alpha+1)^n \, \frac{n^{-\frac{1}{\alpha+1}-1}}{\Gamma\left(-\frac{1}{\alpha+1}\right)}, \end{aligned} \tag{4.35}$$

where the asymptotic equivalence follows from singularity analysis of (4.34). Our first step is to find a recursion for the joint probability generating function of the depth to a random vertex and the hitting time from the root to a random vertex.

Proposition 4.3.1. *The joint probability generating function of the depth $d(1,v)$, of a random vertex v in a random GPORT T, and the hitting time H_{1v} satisfies the recursion*

$$\begin{aligned} K_n(u,y) &:= \sum_l \sum_k \Pr_{\mathcal{G}_n,v}\left(d(1,v)=k, H_{1v}=l\right) u^l y^k \\ &= \frac{1}{n\prod_{k=1}^{n-1}(k(\alpha+1)-1)} \sum_{T\in\mathcal{G}_n, i\in T} u^{H_{1i}} y^{d(1,i)} \\ &= \frac{1}{n} + \frac{\alpha}{(\alpha+1)nE_n} \sum_{r=1}^{n-1} E_r y u^{2(n-r)-1} K_r\left(u, y u^{2(n-r)}\right) \end{aligned} \tag{4.36}$$

for $n \in \mathbb{N}$, where $E_n := \frac{G_n}{(\alpha+1)^n(n-1)!}$.

Proof. The proof is similar to that of Proposition 4.2.1. Define the events B_0 and B_r, $r = 1, \ldots, n-1$ as in the proof of Proposition 4.2.1. As before $\Pr(B_0) = 1/n$ whereas

$$\Pr\left(B_r\right) = \sum_{k=1}^{n-1} \Pr\left(B_{r,k}\right),$$

where $B_{r,k}$ is the event that the random vertex v is in a subtree of size r, attached to the root with out-degree k.

It follows in a similar manner as in (4.13) that

$$
\begin{aligned}
\Pr(B_{r,k}) &= \frac{\binom{\alpha+k-1}{k} rk \binom{n-1}{r} G_r (n-r-1)! [x^{n-r-1}] G(x)^{k-1}}{nG_n} \\
&= \frac{1}{nG_n} \cdot \binom{\alpha+k-1}{k} rk \frac{(n-1)!}{r!} G_r \\
&\cdot \sum_{\substack{r_1+\ldots+r_{k-1}=n-r-1 \\ r_1,\ldots,r_{k-1}\geq 1}} \frac{\prod_{i=1}^{r_1-1}(i(\alpha+1)-1)\ldots\prod_{i=1}^{r_{k-1}-1}(i(\alpha+1)-1)}{r_1!\ldots r_{k-1}!}.
\end{aligned}
$$

Now

$$
\begin{aligned}
\Pr(B_r) &= \frac{(n-1)! r G_r}{nG_n r!} \sum_{k=0}^{n-2} (k+1) \binom{\alpha+k}{k+1} \\
&\cdot \sum_{\substack{r_1+\ldots+r_k=n-r-1 \\ r_1,\ldots,r_k\geq 1}} \frac{\prod_{i=1}^{r_1-1}(i(\alpha+1)-1)\ldots\prod_{i=1}^{r_k-1}(i(\alpha+1)-1)}{r_1!\ldots r_k!} \\
&= \frac{\alpha(n-1)! r G_r}{nG_n r!} \sum_{k=0}^{\infty} \binom{\alpha+k}{k} \\
&\cdot \sum_{\substack{r_1+\ldots+r_k=n-r-1 \\ r_1,\ldots,r_k\geq 1}} \frac{\prod_{i=1}^{r_1-1}(i(\alpha+1)-1)\ldots\prod_{i=1}^{r_k-1}(i(\alpha+1)-1)}{r_1!\ldots r_k!},
\end{aligned}
$$

where

$$
\begin{aligned}
&\sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \binom{\alpha+k}{k} \sum_{\substack{r_1+\ldots+r_k=m \\ r_1,\ldots,r_k\geq 1}} \frac{\prod_{i=1}^{r_1-1}(i(\alpha+1)-1)\ldots\prod_{i=1}^{r_k-1}(i(\alpha+1)-1)}{r_1!\ldots r_k!} x^m \\
&= \sum_{k=0}^{\infty} \binom{\alpha+k}{k} \left(1-(1-(\alpha+1)x)^{\frac{1}{\alpha+1}}\right)^k \\
&= (1-(\alpha+1)x)^{-1}.
\end{aligned}
$$

Hence

$$
\Pr(B_r) = \frac{\alpha(n-1)! r G_r}{nG_n r!} (\alpha+1)^{n-r-1} = \frac{\alpha E_r}{(\alpha+1) n E_n},
$$

where

$$
E_n := \frac{G_n}{(\alpha+1)^n (n-1)!}.
$$

The final part of the proof of this proposition follows exactly as before. ■

Equivalently one obtains:

Proposition 4.3.2. *The joint probability generating function of the depth $d(1,v)$, of a random vertex v in a random GPORT T, and the hitting time H_{v1} satisfies the recursion*

$$\begin{aligned} L_n(u,y) &:= \sum_l \sum_k \Pr_{\mathcal{G}_n,v}\left(d(1,v)=k, H_{v1}=l\right) u^l y^k \\ &= \frac{1}{n\prod_{k=1}^{n-1}\left(k(\alpha+1)-1\right)} \sum_{T\in\mathcal{G}_n, i\in T} u^{H_{i1}} y^{d(1,i)} \\ &= \frac{1}{n} + \frac{\alpha}{(\alpha+1)nE_n}\sum_{r=1}^{n-1} E_r y u^{2r-1} L_r(u,y) \end{aligned} \tag{4.37}$$

for $n \in \mathbb{N}$, where $E_n := \frac{G_n}{(\alpha+1)^n (n-1)!}$.

It follows from (4.35) that

$$E_n = -n\binom{n-1-\frac{1}{\alpha+1}}{n} \sim -\frac{n^{-\frac{1}{\alpha+1}}}{\Gamma\left(-\frac{1}{\alpha+1}\right)}. \tag{4.38}$$

Hence $E_n = -A_n$ if $\alpha = -d$, where A_n is defined in the previous section as in (4.14). Now the recursions for GPORTs in (4.36) and (4.37) are identical to the recursions for d-ary increasing trees in (4.11) and (4.15) if $\alpha = -d$. Since they have the same initial value they must be identical in this case. Hence all the results in the previous section hold for GPORTs as well if we substitute $\alpha = -d$. We state the main results here.

Proposition 4.3.3. *The expected hitting time from a random vertex v to the root in a random GPORT of order n is*

$$\begin{aligned} \mathbb{E}\left(H_{v1}\right) &= \frac{2\alpha}{\alpha+1}n - \frac{\alpha(3\alpha-1)((\alpha+1)n-1)}{n(\alpha+1)^2}\sum_{k=1}^{n}\frac{1}{k(\alpha+1)-1} + \frac{\alpha(\alpha-3)}{(\alpha+1)^2} \\ &= \frac{2\alpha}{\alpha+1}n + \mathcal{O}\left(\log n\right). \end{aligned}$$

Proposition 4.3.4. *The expected hitting time from the root to a random vertex v in a random GPORT of order n is*

$$\begin{aligned} \mathbb{E}\left(H_{1v}\right) &= \frac{\alpha((\alpha+1)n-1)}{(\alpha+1)n}\sum_{k=1}^{n}\frac{1}{k(\alpha+1)-1}\left(2(n-1)+\frac{3\alpha-1}{\alpha+1}\right) - \frac{4\alpha n}{\alpha+1} + \frac{\alpha(\alpha+5)}{(\alpha+1)^2} \\ &\sim \frac{2\alpha}{\alpha+1}n\log n - \frac{2\alpha}{\alpha+1}\left(2+\frac{\Gamma'\left(\frac{\alpha}{\alpha+1}\right)}{\Gamma\left(\frac{\alpha}{\alpha+1}\right)}\right)n. \end{aligned}$$

Theorem 4.3.5. *The s-th moment of the hitting time from a random vertex v to the root in a random GPORT of order n satisfies the asymptotic formula*

$$\mathbb{E}(H_{v1}^s) \sim C_s n^s,$$

where the coefficients satisfy the recursion

$$C_s = \frac{\alpha}{\alpha+1} \cdot \frac{1}{s} \sum_{k=0}^{s-1} \binom{s}{k} 2^{s-k} C_k, \quad C_0 = 1,$$

with exponential generating function

$$\begin{aligned} C(x) &:= \sum_{s=0}^{\infty} C_s \frac{x^s}{s!} \\ &= \exp\left(\frac{\alpha}{\alpha+1} \sum_{n=1}^{\infty} \frac{(2x)^n}{n!n} \right). \end{aligned}$$

Consequently, the normalised random variable $\frac{H_{v1}}{2n}$ *converges weakly to the generalised Dickman distribution* $\mathrm{GD}\left(\frac{\alpha}{\alpha+1}\right)$.

Theorem 4.3.6. *The normalised hitting time from the root to a random vertex* v*, in a random GPORT of size* n*, converges weakly to the standard normal distribution:*

$$\frac{H_{1v} - \frac{2\alpha}{\alpha+1} n \log n}{2n\sqrt{\frac{\alpha}{\alpha+1}\log n}} \xrightarrow{d} N(0,1).$$

Theorem 4.3.7. *The normalised hitting time between two random vertices* v *and* w *in a random GPORT of size* n *converges weakly to the standard normal distribution:*

$$\frac{H_{vw} - \frac{2\alpha}{\alpha+1} n \log n}{2n\sqrt{\frac{\alpha}{\alpha+1}\log n}} \xrightarrow{d} N(0,1).$$

Bibliography

[1] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

[2] D. Aldous and J. Fill. Reversible Markov chains and random walks on graphs. `www.stat.berkeley.edu/~aldous/RWG/book.html`.

[3] R. B. Bapat. Resistance distance in graphs. *Math. Student*, 68(1-4):87–98, 1999.

[4] R. B. Bapat. On the first passage time of a simple random walk on a tree. *Statist. Probab. Lett.*, 81(10):1552–1558, 2011.

[5] F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992.

[6] A. Beveridge. Centers for random walks on trees. *SIAM J. Discrete Math.*, 23(1):300–318, 2008/09.

[7] N. Biggs. Algebraic potential theory on graphs. *Bull. London Math. Soc.*, 29(6):641–682, 1997.

[8] G. Brightwell and P. Winkler. Extremal cover times for random walks on trees. *J. Graph Theory*, 14(5):547–554, 1990.

[9] H. Chen. The generating functions of hitting times for random walk on trees. *Statist. Probab. Lett.*, 77(15):1574–1579, 2007.

[10] NIST Digital library of mathematical functions. `http://dlmf.nist.gov/`, Release 1.0.10 of 2015-08-07, 2015. Online companion to [36].

[11] R. P. Dobrow. On the distribution of distances in recursive trees. *J. Appl. Probab.*, 33(3):749–757, 1996.

[12] R. P. Dobrow and J. A. Fill. Total path length for random recursive trees. *Combin. Probab. Comput.*, 8(4):317–333, 1999. Random graphs and combinatorial structures (Oberwolfach, 1997).

[13] P. G. Doyle and J. L. Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.

[14] M. Drmota. The height of increasing trees. *Ann. Comb.*, 12(4):373–402, 2009.

[15] M. Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[16] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.

[17] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[18] B. Fristedt and L. Gray. *A modern approach to probability theory*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1997.

[19] A. Georgakopoulos and S. Wagner. Hitting times, cover cost, and the Wiener index of a tree. Journal of Graph Theory, to appear, 2016.

[20] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics*. Addison-Wesley Publishing Company, Reading, MA, second edition, 1994. A foundation for computer science.

[21] D. Hunter. Notes for a graduate-level course in asymptotics for statisticians. `http://sites.stat.psu.edu/~dhunter/asymp/lectures/asymp.pdf`.

[22] S. Janson. The Wiener index of simply generated random trees. *Random Structures Algorithms*, 22(4):337–358, 2003.

[23] S. K. Kim. The asymptotic expansion of a hypergeometric function ${}_2F_2\,(1, \alpha; \rho_1, \rho_2; z)$. *Math. Comp.*, 26:963, 1972.

[24] M. Kuba and A. Panholzer. On weighted path lengths and distances in increasing trees. *Probab. Engrg. Inform. Sci.*, 21(3):419–433, 2007.

[25] M. Kuba and S. Wagner. On the distribution of depths in increasing trees. *Electron. J. Combin.*, 17(1):Research Paper 137, 9, 2010.

[26] G. F. Lawler and V. Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[27] L. Lovász. Random walks on graphs: a survey. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 353–397. János Bolyai Math. Soc., Budapest, 1996.

[28] H. M. Mahmoud. Limiting distributions for path lengths in recursive trees. *Probab. Engrg. Inform. Sci.*, 5(1):53–59, 1991.

[29] H. M. Mahmoud. Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.*, 41(1-2):237–245, 1992. Asymptotic methods in analysis and combinatorics.

[30] A. Meir and J. W. Moon. Packing and covering constants for recursive trees. pages 403–411. Lecture Notes in Math., Vol. 642, 1978.

[31] J. W. Moon. Random walks on random trees. *J. Austral. Math. Soc.*, 15:42–53, 1973.

[32] J. W. Moon. The distance between nodes in recursive trees. pages 125–132. London Math. Soc. Lecture Note Ser., No. 13, 1974.

[33] J. W. Moon. On the centroid of recursive trees. *Australas. J. Combin.*, 25:211–219, 2002.

[34] G. O. Munsonius and L. Rüschendorf. Limit theorems for depths and distances in weighted random b-ary recursive trees. *J. Appl. Probab.*, 48(4):1060–1080, 2011.

[35] R. Neininger. The Wiener index of random trees. *Combin. Probab. Comput.*, 11(6):587–597, 2002.

[36] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, 2010.

[37] J. Oosthuizen and S. Wagner. On the distribution of random walk hitting times in random trees. Proceedings of the ANALCO17 Meeting on Analytic Algorithmics and Combinatorics, Barcelona, January 16-17, 2017.

[38] A. Panholzer and H. Prodinger. Level of nodes in increasing trees revisited. *Random Structures Algorithms*, 31(2):203–226, 2007.

[39] M. D. Penrose and A. R. Wade. Random minimal directed spanning trees and Dickman-type distributions. *Adv. in Appl. Probab.*, 36(3):691–714, 2004.

[40] H. Prodinger. Depth and path length of heap ordered trees. *International journal of computer science 7*, pages 293–299, 1996.

[41] R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.*, (51):1–29, 1994.

[42] F. Spitzer. *Principles of random walk*. Springer-Verlag, New York-Heidelberg, second edition, 1976. Graduate Texts in Mathematics, Vol. 34.

[43] P. Tetali. Random walks and the effective resistance of networks. *J. Theoret. Probab.*, 4(1):101–109, 1991.

[44] E. Teufl and S. Wagner. Determinant identities for Laplace matrices. *Linear Algebra Appl.*, 432(1):441–457, 2010.

[45] S. Wagner. On the Wiener index of random trees. *Discrete Math.*, 312(9):1502–1511, 2012.