On the Wiener index of random trees

Stephan Wagner [1]

Department of Mathematical Sciences
Stellenbosch University
Private Bag X1
Matieland 7602
South Africa

Abstract

By a theorem of Janson, the Wiener index of a random tree from a simply generated family of trees converges in distribution to a limit law that can be described in terms of the Brownian excursion. The family of unlabelled trees (rooted or unrooted), which is perhaps the most natural one from a graph-theoretical point of view, since isomorphisms are taken into account, is not covered directly by this theorem though. The aim of this paper is to show how one can prove the same limit law for unlabelled trees by means of generating functions and the method of moments.

Key words: Wiener index, random trees, moments, limit distribution, Brownian excursion

1 Introduction

The *Wiener index*, defined as the sum of all distances between vertices in a connected graph, i.e.,

$$W(G) = \sum_{\{v,w\} \subseteq V(G)} d_G(v,w) = \frac{1}{2} \sum_{v \in V(G)} \sum_{w \in V(G)} d_G(v,w),$$

was introduced by the chemist Harold Wiener in 1947 [14] as a simple parameter that is well correlated to various physico-chemical properties of a molecule (modelled by a graph). Only 30 years later, it was introduced to

Email address: swagner@sun.ac.za (Stephan Wagner).
[1] This material is based upon work supported by the National Research Foundation of South Africa under grant number 70560.

the mathematical literature [2] and given the name *distance* or *total distance*. Clearly the average distance between two randomly chosen vertices is precisely $\binom{n}{2}^{-1} W(G)$.

It was proved by Moon [9] that the average Wiener index of a random unlabelled rooted tree on n vertices is asymptotically $K \cdot n^{5/2}$, where $K \approx 0.5682799594$ is a constant that is related to Otter's tree enumeration constants [11]. A very similar result was obtained by Meir and Moon for rooted labelled trees [8] and more generally by Entringer, Meir, Moon and Székely in [3], where it was shown that the average Wiener index of a random tree on n vertices from a *simply generated family of trees* is always asymptotically of the form $C \cdot n^{5/2}$. Several natural examples are simply generated, for example plane trees, binary trees or rooted labelled trees. They are defined in terms of a weight sequence $1 = w_0, w_1, \ldots$, which is used to define a weight

$$w(T) = \prod_{j \geq 1} w_j^{c_j(T)}$$

for any rooted ordered (i.e., the order of the branches matters) tree T, where $c_j(T)$ is the number of vertices of outdegree j in T. Then it follows that the associated generating function $R(x) = \sum_T w(T) x^{|T|}$ satisfies

$$R(x) = x\Phi(R(x)), \tag{1}$$

where $\Phi(t) = \sum_{j \geq 0} w_j t^j$. Plane trees correspond to $\Phi(t) = \frac{1}{1-t}$, d-ary trees to $\Phi(t) = (1+t)^d$, full d-ary trees (all internal nodes have degree d) to $\Phi(t) = 1 + t^d$, and rooted labelled trees to $\Phi(t) = e^t$. In the last example, one needs to consider exponential generating functions. This construction is essentially equivalent to the probabilistic model of *Galton-Watson trees*.

This connection to a random growth process was exploited by Janson [6] to determine the limit distribution of the Wiener index of a random tree from a simply generated family. His technique, which builds on Aldous' theory of the continuum random tree [1], shows that the limit distribution can be described in terms of Brownian excursions:

Theorem 1. *Let T_n be a random tree of order n from a simply generated family of trees, and denote its Wiener index by $\omega_n = W(T_n)$. Then the distribution of the renormalised random variable $n^{-5/2}\omega_n$ converges to a random variable ζ that can be described in terms of a normalised Brownian excursion $e(t)$,*

$0 < t < 1$, *as follows:*

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

Even though the family of rooted unlabelled trees (where the order of branches is irrelevant, so that isomorphisms are taken into account) does not belong to the class of simply generated families, there are several similarities, and results that hold for simply generated families can therefore be expected to hold for rooted unlabelled trees (also known as *Pólya trees*) and unlabelled trees (without root) as well. This is in contrast to families such as recursive trees, which have a "flatter" shape that only yields a Wiener index of asymptotic order $n^2 \log n$, see [10]. It is mentioned at the end of Janson's paper that his results are very likely to hold for (rooted or unrooted) unlabelled trees as well. The aim of this paper is to provide a formal proof of this fact, which needs quite some effort even though it is "heuristically clear".

The probabilistic techniques employed by Janson cannot be applied directly, since Pólya trees do not stem from a growth process; hence we follow a different approach that makes use of generating functions and moments. The generating function for the number of rooted unlabelled trees satisfies the well-known equation

$$R(x) = \exp\left(\sum_{j\ge1} \frac{1}{j} R(x^j)\right), \tag{2}$$

which is of a similar shape as (1). Indeed, in the asymptotic analysis, it turns out that the terms corresponding to $j \ge 2$ in (2) are essentially irrelevant for the asymptotic behaviour. We also make use of this argument, which will be detailed later, in the proof of our main result.

2 Auxiliary quantities, recursions, generating functions

Let us first consider a recursive procedure to determine the Wiener index of a tree. For this purpose, we need an auxiliary quantity $P(T)$, known as the *total height* or *internal path length* of a rooted tree T, which is defined as the sum of all distances to the root of T. Suppose that $T_1, T_2, \ldots, T_k$ are the branches of T. Then it is easy to see that

$$P(T) = \sum_{j=1}^{k} (P(T_j) + |T_j|) = |T| - 1 + \sum_{j=1}^{k} P(T_j). \tag{3}$$

A limit theorem for the distribution of P was proven by Takács [12] for rooted unlabelled trees (for unrooted trees, P obviously does not make sense). Here we re-prove and refine his result to obtain our main theorem. The Wiener index of T is now given by

$$W(T) = P(T) + \sum_{j=1}^{k} W(T_j) + \sum_{1 \le j_1 < j_2 \le k} |T_{j_1}|(P(T_{j_2}) + |T_{j_2}|).$$

As in Janson's paper [6], it will be advantageous to consider another auxiliary quantity, namely

$$Q(T) = |T|P(T) - W(T). \tag{4}$$

A more natural definition of Q can be given as follows: let r be the root of T, and let $u \wedge v$ denote the lowest common ancestor of two vertices u and v, which is the last vertex that the paths from r to v and w have in common. Then

$$Q(T) = \sum_{v \in V(G)} \sum_{w \in V(G)} d_T(r, v \wedge w).$$

The relation between W and Q follows easily upon noticing that $d_T(r, v) + d_T(r, w) = d_T(v, w) + 2d_T(r, v \wedge w)$. Q satisfies a somewhat simpler recursion, namely

$$Q(T) = \sum_{j=1}^{k} \left(Q(T_j) + |T_j|^2 \right).$$

If the joint distribution of P and Q is known, then the distribution of W follows automatically. Now let us translate these recursions to the world of generating functions. If $G(x, u) = \sum_T x^{|T|} u^{P(T)}$ is the bivariate generating function where the second variable marks P, then we obtain from (3) that

$$G(x, u) = x \exp \left(\sum_{j \ge 1} \frac{1}{j} G(u^j x^j, u^j) \right),$$

which extends (2). If one wants to include Q as well, the relations become more complicated, and it becomes necessary to introduce two more auxiliary variables: in the following, we work with the generating function

$$G(x, y, u, v) = \sum_T x^{|T|} y^{|T|^2} u^{P(T)} v^{Q(T)},$$

where the sum is taken over all rooted (unlabelled) trees T. It is no longer possible to obtain a functional equation for $G(x, y, u, v)$, but if we set $y = 1$, we obtain the following functional equation from the recursions for P and Q:

$$G(x,1,u,v) = x \exp\left(\sum_{j\geq 1} \frac{1}{j} G(u^j x^j, v^j, u^j, v^j)\right). \tag{5}$$

This will be sufficient to determine the asymptotic behaviour of all joint moments of P and Q, which will then in turn lead to the desired limit law. In the following section, we show how this goal can be achieved for random rooted trees with the aid of the functional equation (5). The step to unrooted trees is then performed by means of Otter's well-known dissimilarity theorem, see Section 4.

3 Random rooted trees

In order to study the moments of P and Q, we introduce operators $\Phi_x, \Phi_y, \Phi_u, \Phi_v$ as follows: for a function $A(x,y,u,v)$, we define

$$(\Phi_x A)(x,y,z,u) = x \cdot \frac{\partial}{\partial x} A(x,y,z,u).$$

Φ_y, Φ_u, Φ_v are defined analogously. Note that these operators commute. Now we can write

$$(\Phi_u^r \Phi_v^s G)(x,1,1,1) = \sum_T P(T)^r Q(T)^s x^{|T|},$$

which means that the coefficient of x^n in $(\Phi_u^r \Phi_v^s G)(x,1,1,1)$, divided by the coefficient of x^n in $G(x,1,1,1)$, is precisely the joint moment $\mathbb{E}(P(T_n)^r Q(T_n)^s)$ for random rooted trees of order n. In order to determine relations for generating functions of the form $(\Phi_u^r \Phi_v^s G)(x,1,1,1)$, which will then lead to asymptotic formulas, we differentiate (5) with respect to u and v to obtain

$$(\Phi_u G)(x,1,u,v) = G(x,1,u,v) \sum_{j\geq 1} \Big((\Phi_x G)(u^j x^j, v^j, u^j, v^j) + (\Phi_u G)(u^j x^j, v^j, u^j, v^j) \Big) \tag{6}$$

and

$$(\Phi_v G)(x,1,u,v) = G(x,1,u,v) \sum_{j\geq 1} \Big((\Phi_y G)(u^j x^j, v^j, u^j, v^j) + (\Phi_v G)(u^j x^j, v^j, u^j, v^j) \Big). \tag{7}$$

Remark 2. The following notational convention should be mentioned: in an expression such as $(\Phi_x G)(u^j x^j, v^j, u^j, v^j)$, the operator Φ_x is applied to G before $u^j x^j, \ldots$ are plugged in; if for instance $A(x) = x^2$, then $(\Phi_x A)(x^2) = 2x^4$, while $\Phi_x(A(x^2)) = 4x^4$.

By means of simple induction, we can deduce the following relations for higher derivatives:

Lemma 3. *For $r \geq 1, s \geq 0$, we have*

$$(\Phi_u^r \Phi_v^s G)(x,1,u,v) = \sum_{k=0}^{r-1} \sum_{\ell=0}^{s} \binom{r-1}{k} \binom{s}{\ell} (\Phi_u^k \Phi_v^\ell G)(x,1,u,v) \sum_{j \geq 1} j^{r+s-k-\ell-1} \Big(((\Phi_x + \Phi_u)^{r-k} (\Phi_y + \Phi_v)^{s-\ell} G)(u^j x^j, v^j, u^j, v^j) \Big) \quad (8)$$

Similarly, for $r \geq 0, s \geq 1$, we have

$$(\Phi_u^r \Phi_v^s G)(x,1,u,v) = \sum_{k=0}^{r} \sum_{\ell=0}^{s-1} \binom{r}{k} \binom{s-1}{\ell} (\Phi_u^k \Phi_v^\ell G)(x,1,u,v) \sum_{j \geq 1} j^{r+s-k-\ell-1} \Big(((\Phi_x + \Phi_u)^{r-k} (\Phi_y + \Phi_v)^{s-\ell} G)(u^j x^j, v^j, u^j, v^j) \Big). \quad (9)$$

In particular, we obtain functional equations for the functions $(\Phi_u^r \Phi_v^s G)(x,1,1,1)$ for arbitrary r and s. In order to derive asymptotic information, we first have to review some known properties of $G(x,1,1,1)$, see [5]:

Lemma 4. *The generating function $R(x) = G(x,1,1,1)$ of the family of rooted trees has a square-root singularity at $\rho \approx 0.33832185$, where it can be expanded into a series*

$$R(x) = 1 - c_1 \sqrt{1 - x/\rho} + c_2 (1 - x/\rho) + \dots,$$

and can otherwise be continued analytically to a circle of radius $\rho' > \rho$ (with a branch cut at ρ).

Now we show that this also holds for $(\Phi_u^r \Phi_v^s G)(x,1,1,1)$ $(r, s \geq 0)$, except for the shape of the expansion around the singularity ρ:

Proposition 5. *For $r, s \geq 0$ and $r+s \geq 1$, the function $R_{r,s}(x) = (\Phi_u^r \Phi_v^s G)(x,1,1,1)$ has an expansion of the form*

$$R_{r,s}(x) = b_{r,s} (1 - x/\rho)^{-(3r+5s-1)/2} + O\Big((1 - x/\rho)^{-(3r+5s-2)/2} \Big)$$

around ρ and can otherwise be continued analytically to a circle of radius $\rho' > \rho$ (with a branch cut at ρ). The coefficients $b_{r,s}$ can be determined from the recursion

$$b_{r,s} = \frac{1}{c_1} \Bigg(\frac{3r+5s-4}{2} \cdot r b_{r-1,s} + \frac{(3r+5s-4)(3r+5s-6)}{4} \cdot s b_{r,s-1} + \frac{1}{2} \sum_{\substack{k=0 \\ 0<k+\ell<r+s}}^{r} \sum_{\ell=0}^{s} \binom{r}{k} \binom{s}{\ell} b_{k,\ell} b_{r-k,s-\ell} \Bigg) \quad (10)$$

with initial values $b_{1,0} = \frac{1}{2}$ *and* $b_{0,1} = \frac{1}{4}$ *(and the convention that* $0 \cdot b_{r,-1} = 0 \cdot b_{-1,s} = 0$*).*

Proof. Let us start with a few more well-known facts on the function $R(x)$ (that are also essential for the proof of Lemma 4): from (2), one obtains

$$R(x) = -W\left(-x \exp\left(\sum_{j\geq 2} \frac{1}{j} R(x^j)\right)\right),$$

where W denotes the Lambert W-function, which is known to have a branch cut at $-1/e$ (with $W(-1/e) = 1$). The function $\sum_{j\geq 2} \frac{1}{j} R(x^j)$ is analytic for $|x| < \sqrt{\rho}$, which can be seen as follows (cf. again [5]): if

$$R(x) = \sum_{n\geq 1} r_n x^n,$$

then

$$\sum_{j\geq 2} \frac{1}{j} R(x^j) \leq \sum_{n\geq 1} r_n \sum_{j\geq 2} x^{jn} = \sum_{n\geq 1} r_n \frac{x^{2n}}{1 - x^n} \leq \frac{1}{1-x} R(x^2) < \infty$$

for $x < \sqrt{\rho}$, which shows that the radius of convergence is $\sqrt{\rho}$. In the following, we will frequently make use of the fact that similar sums represent analytic functions within a larger circle without mentioning it explicitly every time. This shows that ρ is also the unique positive solution to the equation

$$x \exp\left(\sum_{j\geq 2} \frac{1}{j} R(\rho^j)\right) = \frac{1}{e}.$$

Since the coefficients of R are all positive, there are no other solutions inside a larger circle of radius ρ' if ρ' is chosen sufficiently small. This also implies that $R(x) \neq 1$ inside this circle (except for the point $x = \rho$).

Now we proceed with the proof of the proposition by induction on $r + s$. For $r = 1$ and $s = 0$, we have, in view of (6),

$$R_{1,0}(x) = R(x) \sum_{j\geq 1} \left(x^j R'(x^j) + R_{1,0}(x^j)\right).$$

Solving for $R_{1,0}(x)$ yields

$$R_{1,0}(x) = \frac{R(x)}{1 - R(x)} \left(xR'(x) + \sum_{j\geq 2} \left(x^j R'(x^j) + R_{1,0}(x^j)\right)\right).$$

The last sum is analytic for $|x| < \sqrt{\rho}$, which proves that $R_{1,0}(x)$ is indeed analytic inside a circle of radius ρ' for sufficiently small ρ', except for the singularity at ρ (and the associated branch cut). The asymptotic expansion of R around ρ given in Lemma 4 now shows that

$$R_{1,0}(x) = \frac{1}{2(1-x/\rho)} + O\left((1-x/\rho)^{-1/2}\right),$$

i.e., our statement holds, and $b_{1,0} = \frac{1}{2}$. Noting that $(\Phi_y G)(x,y,u,v) = (\Phi_x^2 G)(x,y,u,v)$ by definition of G, one proves in a similar way that

$$R_{0,1}(x) = \frac{1}{4(1-x/\rho)^2} + O\left((1-x/\rho)^{-3/2}\right).$$

For the induction step, we apply Lemma 3: for arbitrary $r \geq 1$ and $s \geq 0$, we obtain from (8) (making use of the identity $(\Phi_y G)(x,y,u,v) = (\Phi_x^2 G)(x,y,u,v)$ again)

$$\begin{aligned} R_{r,s}(x) = \sum_{k=0}^{r-1} \sum_{\ell=0}^{s} \binom{r-1}{k} \binom{s}{\ell} R_{k,\ell}(x) \sum_{j \geq 1} j^{r+s-k-\ell-1} \\ \sum_{h_1=0}^{r-k} \sum_{h_2=0}^{s-\ell} \binom{r-k}{h_1} \binom{s-\ell}{h_2} (\Phi_x^{h_1+2h_2} R_{r-k-h_1,s-\ell-h_2})(x^j). \quad (11) \end{aligned}$$

One of the terms on the right hand side of this equation (corresponding to $k = \ell = h_1 = h_2 = 0$ and $j = 1$) is $R(x)R_{r,s}(x)$. We solve the equation for $R_{r,s}(x)$ to obtain

$$\begin{aligned} R_{r,s}(x) = (1-R(x))^{-1} \Bigg(\sum_{k=0}^{r-1} \sum_{\ell=0}^{s} \sum_{h_1=0}^{r-k} \sum_{h_2=0}^{s-\ell} {\sum_{j\geq 1}}^{*} j^{r+s-k-\ell-1} \\ \binom{r-1}{k} \binom{s}{\ell} \binom{r-k}{h_1} \binom{s-\ell}{h_2} R_{k,\ell}(x) (\Phi_x^{h_1+2h_2} R_{r-k-h_1,s-\ell-h_2})(x^j) \Bigg), \quad (12) \end{aligned}$$

where $\sum^*$ indicates that the summand corresponding to $k = \ell = h_1 = h_2 = 0$ and $j = 1$ is left out. It remains to identify the terms inside the bracket whose asymptotic order at the singularity ρ is highest: by the induction hypothesis,

- the terms corresponding to $j \geq 2$ are $O\Big((1-x/\rho)^{-(3r+5s-4)/2}\Big)$,
- for $j = 1$, we have

$$\begin{aligned} &R_{k,\ell}(x)(\Phi_x^{h_1+2h_2} R_{r-k-h_1,s-\ell-h_2})(x) \\ &\quad = b_{k,\ell} b_{r-k-h_1,s-\ell-h_2} \left(\frac{3(r-k-h_1) + 5(s-\ell-h_2) - 1}{2} \right)^{\overline{h_1+2h_2}} \\ &\qquad \cdot (1-x/\rho)^{-(3r+5s-h_1-h_2-2)/2} + O\Big((1-x/\rho)^{-(3r+5s-h_1-h_2-3)/2}\Big) \end{aligned}$$

if $(k,\ell) \neq (0,0)$, and

$$
\begin{aligned}
&R_{0,0}(x)(\Phi_x^{h_1+2h_2} R_{r-h_1,s-h_2})(x) \\
&\qquad = b_{r-h_1,s-h_2} \left(\frac{3(r-h_1)+5(s-h_2)-1}{2} \right)^{\overline{h_1+2h_2}} \\
&\qquad\quad \cdot (1-x/\rho)^{-(3r+5s-h_1-h_2-1)/2} + O\Big((1-x/\rho)^{-(3r+5s-h_1-h_2-2)/2} \Big)
\end{aligned}
$$

otherwise.

Therefore, the relevant terms are those corresponding to $j=1$ and either $(k,\ell,h_1,h_2) = (0,0,1,0)$ or $(k,\ell,h_1,h_2) = (0,0,0,1)$ or $h_1 = h_2 = 0$ ($(k,\ell) \neq (0,0)$ arbitrary), and we can conclude that

$$
R_{r,s}(x) = b_{r,s}(1-x/\rho)^{-(3r+5s-1)/2} + O\left((1-x/\rho)^{-(3r+5s-2)/2} \right)
$$

with

$$
\begin{aligned}
b_{r,s} = \frac{1}{c_1} \Bigg(&\frac{3r+5s-4}{2} \cdot r b_{r-1,s} + \frac{(3r+5s-4)(3r+5s-6)}{4} \cdot s b_{r,s-1} \\
&+ \sum_{\substack{k=0 \\ k+\ell>0}}^{r-1} \sum_{\ell=0}^{s} \binom{r-1}{k} \binom{s}{\ell} b_{k,\ell} b_{r-k,s-\ell} \Bigg).
\end{aligned} \tag{13}
$$

Similarly, (9) yields, for $r \geq 0$ and $s \geq 1$,

$$
\begin{aligned}
b_{r,s} = \frac{1}{c_1} \Bigg(&\frac{3r+5s-4}{2} \cdot r b_{r-1,s} + \frac{(3r+5s-4)(3r+5s-6)}{4} \cdot s b_{r,s-1} \\
&+ \sum_{\substack{k=0 \\ k+\ell>0}}^{r} \sum_{\ell=0}^{s-1} \binom{r}{k} \binom{s-1}{\ell} b_{k,\ell} b_{r-k,s-\ell} \Bigg).
\end{aligned} \tag{14}
$$

The double sums might look different, but they are not, and they can both be easily rewritten as

$$
\begin{aligned}
\sum_{\substack{k=0 \\ k+\ell>0}}^{r-1} \sum_{\ell=0}^{s} \binom{r-1}{k} \binom{s}{\ell} b_{k,\ell} b_{r-k,s-\ell} &= \sum_{\substack{k=0 \\ k+\ell>0}}^{r} \sum_{\ell=0}^{s-1} \binom{r}{k} \binom{s-1}{\ell} b_{k,\ell} b_{r-k,s-\ell} \\
&= \frac{1}{2} \sum_{\substack{k=0 \\ 0<k+\ell<r+s}}^{r} \sum_{\ell=0}^{s} \binom{r}{k} \binom{s}{\ell} b_{k,\ell} b_{r-k,s-\ell},
\end{aligned}
$$

which completes the proof of our proposition. ■

With Proposition 5 at hand, we can apply singularity analysis [4, Theorem VI.4] to the functions $R_{r,s}(x)$ and obtain

$$[x^n]R_{r,s}(x) = \frac{b_{r,s}}{\Gamma((3r+5s-1)/2)} \cdot n^{(3r+5s-3)/2} \cdot \rho^{-n}\Big(1 + O(n^{-1/2})\Big).$$

Dividing by the number of rooted trees, which is

$$[x^n]R(x) = \frac{-c_1}{\Gamma(-1/2)} n^{-3/2} \cdot \rho^{-n}\Big(1 + O(n^{-1/2})\Big),$$

we obtain the mixed moment $\mathbb{E}(P(T_n)^r Q(T_n)^s)$ for a random rooted tree T_n of order n:

$$\mathbb{E}(P(T_n)^r Q(T_n)^s) = \frac{[x^n]R_{r,s}(x)}{[x^n]R(x)} = \frac{2\sqrt{\pi} b_{r,s}}{c_1 \Gamma((3r+5s-1)/2)} \cdot n^{(3r+5s)/2}\Big(1+O(n^{-1/2})\Big).$$

Finally, define the numbers $\omega_{r,s}$ (which turn out to be integers) by

$$\omega_{r,s} = \frac{c_1^{r+s-1} 2^{2r+3s-1}}{r!s!} b_{r,s},$$

so that, after a few manipulations,

$$\mathbb{E}(P(T_n)^r Q(T_n)^s) = \frac{\sqrt{\pi} r!s!\omega_{r,s}}{2^{(5r+7s-4)/2}\Gamma((3r+5s-1)/2)} \cdot \left(\frac{\sqrt{2}}{c_1}\right)^{r+s} n^{(3r+5s)/2}\Big(1+O(n^{-1/2})\Big).$$

The recursion for $b_{r,s}$ in Proposition 5 becomes

$$\omega_{r,s} = 2(3r+5s-4)\omega_{r-1,s} + 2(3r+5s-4)(3r+5s-6)\omega_{r,s-1} + \sum_{\substack{k=0 \\ 0<k+\ell<r+s}}^{r} \sum_{\ell=0}^{s} \omega_{k,\ell}\omega_{r-k,s-\ell}, \tag{15}$$

with initial values $\omega_{0,1} = \omega_{1,0} = 1$ and the convention that $\omega_{r,s} = 0$ if $r < 0$ or $s < 0$. As shown by Janson in [6], the fraction

$$\frac{\sqrt{\pi} r!s!\omega_{r,s}}{2^{(5r+7s-4)/2}\Gamma((3r+5s-1)/2)} \tag{16}$$

is precisely the mixed moment $\mathbb{E}(\xi^r \eta^s)$, where ξ and η are defined as in Theorem 1. In view of (4), this also shows that the mixed moments including the Wiener index satisfy

$$\mathbb{E}\left(P(T_n)^r Q(T_n)^s W(T_n)^t\right) = \mathbb{E}\left(\xi^r \eta^s (\xi - \eta)^t\right) \cdot \left(\frac{\sqrt{2}}{c_1}\right)^{r+s+t} n^{(3r+5s+5t)/2}\Big(1+O(n^{-1/2})\Big).$$

The moments of $\zeta = \xi - \eta$ grow slowly enough to characterise the distribution. Indeed, since

$$\mathbb{E}(\zeta^k) \le \mathbb{E}(\xi^k) = \frac{k!\sqrt{\pi}\omega_{k,0}}{2^{(5k-4)/2}\Gamma((3k-1)/2)}$$

and

$$\omega_{k,0} \sim \frac{1}{2\pi} \cdot 6^k (k-1)!,$$

see [6], the moment generating function $\mathbb{E}(e^{t\zeta})$ of ζ converges for all t, so that convergence of moments implies convergence in distribution (see [4, Theorem C.2]). Hence we have our first main theorem:

Theorem 6. *The normalised Wiener index $n^{-5/2} W(T_n)$ of a random rooted 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 1. Furthermore, all moments converge:*

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

As expected, the behaviour of the Wiener index of random rooted unlabelled trees is exactly the same as for simply generated trees. In the following section, we show that this remains true if random unlabelled trees without root are considered.

4 Random unrooted trees

Recall Otter's dissimilarity theorem [11,5], which states that the number of edge orbits of a tree is always one less than the number of vertex orbits, unless the tree has a symmetry edge (in which case the numbers are the same). Hence if $R(x)$ and $\tilde{R}(x)$ are the generating functions for rooted and unrooted unlabelled trees respectively, one obtains

$$\tilde{R}(x) = R(x) - \frac{1}{2}\left(R(x)^2 - R(x^2)\right),$$

see again [11,5]. The first summand counts rooted trees (and thus unrooted trees weighted by the number of their vertex orbits), the second summand edge-rooted trees (unrooted trees weighted by the number of their edge orbits), and the last one takes trees with a symmetry edge into account. For our purposes, we need a version of this identity for the generating functions

$$W_r(x) = \sum_T W(T)^r x^{|T|}, \qquad \tilde{W}_r(x) = \sum_T W(T)^r x^{|T|}$$

in which the sums are taken over all rooted resp. unrooted unlabelled trees T. It is easy to express W_r in terms of the functions $R_{r,s}$ from the previous section:

$$W_r(x) = \sum_{p=0}^{r} (-1)^{r-p} \binom{r}{p} \Phi_x^p R_{p,r-p}(x), \tag{17}$$

since $W(T) = |T|P(T) - Q(T)$ for all trees T. In order to write $\tilde{W}_r$ in terms of these functions, we first need a formula for the Wiener index of an edge-rooted tree: suppose that T is obtained by joining two rooted trees T_1 and T_2 by an edge. Then it is easily verified that

$$\begin{aligned} W(T) &= W(T_1) + W(T_2) + |T_1|P(T_2) + |T_2|P(T_1) + |T_1||T_2| \\ &= (P(T_1) + P(T_2))|T| - Q(T_1) - Q(T_2) + |T_1||T_2|. \end{aligned}$$

Hence we obtain, by way of Otter's dissimilarity theorem, that

$$\tilde{W}_r(x) = W_r(x) - B_r^{(1)}(x) + B_r^{(2)}(x),$$

where

$$\begin{aligned} B_r^{(1)}(x) = \frac{1}{2} \sum_{k_1+k_2+k_3+k_4+k_5=r} (-1)^{k_3+k_4} \binom{r}{k_1, k_2, k_3, k_4, k_5} \\ \varPhi_x^{k_1+k_2} \left(\left(\varPhi_x^{k_5} R_{k_1,k_3}(x) \right) \left(\varPhi_x^{k_5} R_{k_2,k_4}(x) \right) \right) \end{aligned} \tag{18}$$

and

$$B_r^{(2)}(x) = \frac{1}{2} \sum_{k_1+k_2+k_3} (-1)^{k_2} 2^{2k_1+k_2} \left(\varPhi_x^{k_1+2k_3} R_{k_1,k_2} \right) (x^2).$$

For the following asymptotic analysis, $B_r^{(2)}$ (which corresponds to trees with a symmetry edge) is negligible, since its radius of convergence is $\sqrt{\rho} > \rho$. It remains to deal with W_r and $B_r^{(1)}$: in the equation (17) for W_r, plug in (11) for $R_{p,r-p}$ to obtain

$$\begin{aligned} W_r = \sum_{p=1}^{r} (-1)^{r-p} \binom{r}{p} \varPhi_x^p \Bigg(\sum_{k=0}^{p-1} \sum_{\ell=0}^{r-p} \binom{p-1}{k} \binom{r-p}{\ell} R_{k,\ell}(x) \sum_{j \geq 1} j^{r-k-\ell-1} \\ \sum_{h_1=0}^{p-k} \sum_{h_2=0}^{r-p-\ell} \binom{p-k}{h_1} \binom{r-p-\ell}{h_2} (\varPhi_x^{h_1+2h_2} R_{p-k-h_1,r-p-\ell-h_2})(x^j) \Bigg) \\ + (-1)^r \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} R_{0,\ell} \sum_{j \geq 1} j^{r-\ell-1} \sum_{h=0}^{r-\ell} \binom{r-\ell}{h} (\varPhi_x^{2h} R_{0,r-\ell-h})(x^j). \end{aligned}$$

This formula looks long and messy, but in fact many of its terms are not actually needed or cancel. Note first that all terms with $j \geq 2$, $h_1 + h_2 > 2$ ($h > 2$ in the second sum) or $h_1 + h_2 = 2$, $(k, \ell) \neq 0$ ($h = 2$, $\ell \neq 0$ in the second sum) are $O\Big((1 - x/\rho)^{-(5r-4)/2}\Big)$. A priori, the terms with $h_1 + h_2 = 2$ and $(k, \ell) = (0, 0)$ ($h = 2$, $\ell = 0$) would be of higher order, but it turns out that the essential parts cancel:

$$\begin{aligned}\sum_{p=0}^{r}(-1)^{r-p}\binom{r}{p}\Phi_x^p\left(R_{0,0}(x)\sum_{h_1+h_2=2}\binom{p}{h_1}\binom{r-p}{h_2}\Phi_x^{h_1+2h_2}R_{p-h_1,r-p-h_2}\right)\\ = R_{0,0}(x)\sum_{p=0}^{r}(-1)^{r-p}\binom{r}{p}\sum_{h_1+h_2=2}\binom{p}{h_1}\binom{r-p}{h_2}\Phi_x^{p+h_1+2h_2}R_{p-h_1,r-p-h_2}\\ + O\Big((1-x/\rho)^{-(5r-4)/2}\Big)\\ = O\Big((1-x/\rho)^{-(5r-4)/2}\Big),\end{aligned}$$

which follows easily upon noticing that the coefficient of $\Phi_x^{a+4}R_{a,r-a-2}$ is

$$\sum_{h_1=0}^{2}(-1)^{r-a-h_1}\binom{r}{a+h_1}\binom{a+h_1}{h_1}\binom{r-a-h_1}{2-h_1} = \frac{(-1)^{r-a}r!}{2a!(r-a-2)!}\sum_{h_1=0}^{2}(-1)^{h_1}\binom{2}{h_1} = 0.$$

This leaves us with $j=1$ and $h_0+h_1\le 1$. However, the terms with $(h_0,h_1)=(0,0)$ ($h=0$ in the second sum) cancel identically with the summand $k_5=0$ in formula (18) for $B_r^{(1)}$:

$$\begin{aligned}\sum_{p=1}^{r}(-1)^{r-p}\binom{r}{p}\Phi_x^p\left(\sum_{k=0}^{p-1}\sum_{\ell=0}^{r-p}\binom{p-1}{k}\binom{r-p}{\ell}R_{k,\ell}(x)R_{p-k,r-p-\ell}(x)\right)\\ +(-1)^r\sum_{\ell=0}^{r-1}\binom{r-1}{\ell}R_{0,\ell}R_{0,r-\ell}(x)\\ = \frac{1}{2}\sum_{k_1+k_2+k_3+k_4=r}(-1)^{k_3+k_4}\binom{r}{k_1,k_2,k_3,k_4}\\ \Phi_x^{k_1+k_2}\left(R_{k_1,k_3}(x)R_{k_2,k_4}(x)\right),\end{aligned}$$

which follows easily upon comparing coefficients of $\Phi_x^p R_{k,\ell}(x)R_{p-k,r-p-\ell}(x)$.

Since the terms corresponding to $k_5\ge 2$ in (18) are also $O\Big((1-x/\rho)^{-(5r-4)/2}\Big)$, the only relevant parts that remain are

$$\begin{aligned}\tilde{W}_r(x) = \sum_{p=1}^{r}(-1)^{r-p}\binom{r}{p}\Phi_x^p\left(\sum_{k=0}^{p-1}\sum_{\ell=0}^{r-p}\binom{p-1}{k}\binom{r-p}{\ell}R_{k,\ell}(x)\right.\\ \left.\Big((p-k)\Phi_x R_{p-k-1,r-p-\ell}(x)+(r-p-\ell)\Phi_x^2 R_{p-k,r-p-\ell-1}(x)\Big)\right)\\ +(-1)^r\sum_{\ell=0}^{r-1}\binom{r-1}{l}(r-\ell)R_{0,\ell}(x)\Phi_x^2 R_{0,r-\ell-1}(x)\\ -\frac{r}{2}\sum_{k_1+k_2+k_3+k_4=r-1}\binom{r-1}{k_1,k_2,k_3,k_4}\Phi_x^{k_1+k_2}\left(\Phi_x R_{k_1,k_3}(x)\Phi_x R_{k_2,k_4}(x)\right)\\ +O\Big((1-x/\rho)^{-(5r-4)/2}\Big).\end{aligned}$$

All remaining summands are of order $(1-x/\rho)^{-(5r-3)/2}$, except possibly for those with $(k,\ell)=(0,0)$ ($\ell=0$ in the second sum). However, all terms of higher order cancel (as in the case $h_1+h_2=2$ before), so that we can indeed conclude that

$$\tilde{W}_r(x)=\tilde{\lambda}_r(1-x/\rho)^{-(5r-3)/2}+O\Big((1-x/\rho)^{-(5r-4)/2}\Big),$$

where (with $b_{0,0}=-c_1$ and thus $\omega_{0,0}=-\frac{1}{2}$, so that (16) remains correct)

$$\begin{aligned}\tilde{\lambda}_r=&\sum_{p=1}^{r}(-1)^{r-p}\binom{r}{p}\left(\frac{5r-2p-3}{2}\right)^{\overline{p}}\left(\sum_{k=0}^{p-1}\sum_{\ell=0}^{r-p}\binom{p-1}{k}\binom{r-p}{\ell}b_{k,\ell}(x)\right.\\&\quad\Bigg(\frac{(p-k)(5r-2p-3k-5\ell-4)b_{p-k-1,r-p-\ell}}{2}\\&\quad\left.+\frac{(r-p-\ell)(5r-2p-3k-5\ell-4)(5r-2p-3k-5\ell-6)b_{p-k,r-p-\ell-1}}{4}\Bigg)\right)\\&+(-1)^r\sum_{\ell=0}^{r-1}\binom{r-1}{l}\frac{(r-\ell)(5r-5\ell-4)(5r-5\ell-6)b_{0,\ell}b_{0,r-\ell-1}}{4}\\&-\frac{r}{2}\sum_{k_1+k_2+k_3+k_4=r-1}\binom{r-1}{k_1,k_2,k_3,k_4}\left(\frac{5r-2k_1-2k_2-3}{2}\right)^{\overline{k_1+k_2}}\\&\quad\frac{(3k_1+5k_3-1)(3k_2+5k_4-1)b_{k_1,k_3}b_{k_2,k_4}}{4}\end{aligned}$$

This proves convergence of moments: it is well known that the number of unrooted trees on n vertices is asymptotically $\frac{c_1^3}{4\sqrt{\pi}}n^{-5/2}\rho^{-n}(1+O(n^{-1/2}))$ (see [5]), so that singularity analysis yields

$$\mathbb{E}(W(\tilde{T}_n)^r)=\frac{4\sqrt{\pi}\tilde{\lambda}_r}{c_1^3\Gamma((5r-3)/2)}n^{5r/2}(1+O(n^{-1/2}))$$

for the r-th moment of the Wiener index of a random unrooted tree $\tilde{T}_n$. However, it is not yet clear that the moments are (asymptotically) the same as in the unrooted case. On the other hand, we can write

$$W_r(x)=\lambda_r(1-x/\rho)^{-(5r-1)/2}+O\Big((1-x/\rho)^{-(5r-2)/2}\Big)$$

with

$$\lambda_r=\sum_{p=0}^{r}(-1)^{r-p}\binom{r}{p}\left(\frac{5r-2p-1}{2}\right)^{\overline{p}}b_{p,r-p},$$

so that

$$\mathbb{E}(W(T_n)^r)=\frac{2\sqrt{\pi}\lambda_r}{c_1\Gamma((5r-1)/2)}n^{5r/2}(1+O(n^{-1/2}))$$

We are done if we can show that $\tilde{\lambda}_r=\frac{c_1^2\lambda_r}{5r-3}$. The proof of this fact is simple albeit indirect: one can repeat the exact same steps for *labelled* rooted and unrooted trees; in the labelled case, it is trivial that the distribution of the Wiener index has to be the same for rooted and unrooted trees, hence the stated identity between λ_r and $\tilde{\lambda}_r$ must hold. Our main theorem follows immediately:

Theorem 7. *The normalised Wiener index $n^{-5/2}W(T_n)$ of a random unrooted 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 1. Furthermore, all moments converge:*

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

5 Conclusion

Not surprisingly, Janson's limit theorem for the Wiener index of simply generated trees extends to unlabelled (rooted or unrooted) trees. It would be possible to add additional constraints, e.g. prescribing the set of possible vertex degrees as in [13]. However, it seemed reasonable to present only the main case to keep the technicalities at a decent level.

The author has been unable to find a direct proof of the identity $\tilde{\lambda}_r = \frac{c_1^2\lambda_r}{5r-3}$ from the recursion for the coefficients $b_{k,\ell}$ (or $\omega_{k,\ell}$); it would be interesting to see such a proof. One can also obtain another interesting identity for these numbers: note first that the Wiener index of a tree T can be written as

$$W(T) = \sum_{e \in E(T)} n_1(e)n_2(e),$$

where $n_1(e)$ and $n_2(e)$ are the orders of the two connected components of $T \setminus e$. Now consider labelled trees; edge-rooted labelled trees correspond to unordered pairs of rooted labelled trees. Hence we have

$$\frac{1}{2}\sum_{T_1}\sum_{T_2} |T_1||T_2|(W(T_1)+W(T_2)+|T_1|P(T_2)+|T_2|P(T_1)+|T_1||T_2|)^{r-1}\frac{x^{|T_1|+|T_2|}}{(|T_1|+|T_2|)!} = \sum_{T} W(T)^r \frac{x^{|T|}}{|T|!}$$

by the above formula for $W(T)$, where the first double sum is over all rooted labelled trees T_1 and T_2 and the second sum is over all unrooted labelled trees T. If both sides of the identity are translated to generating functions and the moments are computed both ways, one obtains (after some additional manipulations) the curious identity

$$\frac{r}{2}\sum_{p=0}^{r}(-2)^p\left(\frac{5r-2p-1}{2}\right)^{\overline{p}}\omega_{p,r-p}$$
$$=\sum_{p=1}^{r}(-2)^p\left(\frac{5r-2p-1}{2}\right)^{\overline{p}}\sum_{k=0}^{p-1}\sum_{\ell=0}^{r-p}(3k+5\ell-1)(5r-2p-3k-5\ell-4)\omega_{k,\ell}\omega_{p-k-1,r-p-\ell}$$

for all $r \geq 1$. These observations suggest that the array of numbers $\omega_{k,\ell}$ defined by the recursion (15) has several interesting properties that deserve further study (cf. also [7]).

References

[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] R. C. Entringer, D. E. Jackson, and D. A. Snyder. Distance in graphs. *Czechoslovak Math. J.*, 26(101)(2):283–296, 1976.

[3] R. C. Entringer, A. Meir, J. W. Moon, and L. A. Székely. The Wiener index of trees from certain families. *Australas. J. Combin.*, 10:211–224, 1994.

[4] P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[5] F. Harary and E. M. Palmer. *Graphical enumeration.* Academic Press, New York, 1973.

[6] S. Janson. The Wiener index of simply generated random trees. *Random Structures Algorithms*, 22(4):337–358, 2003.

[7] S. Janson. Brownian excursion area, Wright's constants in graph enumeration, and other Brownian areas. *Probab. Surv.*, 4:80–145 (electronic), 2007.

[8] A. Meir and J. W. Moon. The distance between points in random trees. *J. Combinatorial Theory*, 8:99–103, 1970.

[9] J. W. Moon. On the expected distance from the centroid of a tree. *Ars Combin.*, 20(A):263–276, 1985. Tenth British combinatorial conference (Glasgow, 1985).

[10] R. Neininger. The Wiener index of random trees. *Combin. Probab. Comput.*, 11(6):587–597, 2002.

[11] R. Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[12] L. Takács. The asymptotic distribution of the total heights of random rooted trees. *Acta Sci. Math. (Szeged)*, 57(1-4):613–625, 1993.

[13] S. G. Wagner. On the average Wiener index of degree-restricted trees. *Australas. J. Combin.*, 37:187–203, 2007.

[14] H. Wiener. Structural determination of paraffin boiling points. *J. Amer. Chem. Soc.*, 69:17–20, 1947.