AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **37** (2007), Pages 187–203

On the average Wiener index of degree-restricted trees*

Stephan G. Wagner

Department of Mathematics
Graz University of Technology
Steyrergasse 30, A-8010 Graz
Austria
wagner@finanz.math.tu-graz.ac.at

Abstract

The Wiener index, defined as the total sum of distances in a graph, is one of the most popular graph-theoretical indices. Its average value has been determined for several classes of trees, giving an asymptotics of the form $Kn^{5/2}$ for some K, where n is the number of vertices. In this note, it is shown how the method can be extended to trees with restricted degrees. Particular emphasis is placed on chemical trees — trees with maximal degree ≤ 4 — since the Wiener index is of interest in theoretical chemistry.

1 Introduction

The Wiener index of a graph G, named after the chemist Harold Wiener [18], who considered it in connection with paraffin boiling points, is given by

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

where d_G denotes the distance in G. Besides its purely graph-theoretic value, the Wiener index has interesting applications in chemistry. We quote [2], which gives an extensive summary on the various works, and refer to [16] for further information on the chemical applications.

The average behaviour of the Wiener index was first studied by Entringer et al. [5], who considered so-called simply generated families of trees (introduced by Meir and

* This work was supported by Austrian Science Fund project no. S-8307-MAT.

Moon, cf. [9]). They were able to prove that the average Wiener index is asymptotically $Kn^{5/2}$, where the constant K depends on the specific family of trees. In more recent articles, Neininger [11] studied recursive and binary search trees, and Janson [7] determined moments of the Wiener index of random rooted trees. Dobrynin and Gutman [3] calculated numerical values for the average Wiener index of trees and chemical trees of small order by direct computer calculation.

The average Wiener index of a tree (taking isomorphies into account) has been determined, in a different context, in a paper of Moon [10]—it is given asymptotically by $0.56828n^{5/2}$.

The aim of this note is to extend the cited results to trees with restricted degree, especially chemical trees. In fact, the enumeration method for chemical trees is older than the result of Otter for arbitrary trees and goes back to Cayley [1] and Pólya [13].

For a given set $\mathcal{D} \subseteq \mathbb{N}$, we denote by $\mathcal{F}_\mathcal{D}$ the family of rooted trees with the property that the outdegree of every vertex lies in $\mathcal{D}_0 = \mathcal{D} \cup \{0\}$, and we denote by $\tilde{\mathcal{F}}_\mathcal{D}$ the family of trees with the property that all degrees lie in the set $\tilde{\mathcal{D}} = \{d+1 : d \in \mathcal{D}_0\}$. In a chemical context, the sets $\mathcal{D} = \{1,2,3\}$ (corresponding to the so-called "chemical trees", i.e. all degrees are ≤ 4), $\mathcal{D} = \{1,2\}$ and $\mathcal{D} = \{3\}$ are of particular interest. Picking a tree T_n with n vertices uniformly at random from $\mathcal{F}_\mathcal{D}$ or $\tilde{\mathcal{F}}_\mathcal{D}$, we are interested in the average Wiener index.

Following the method of Entringer et al. [5], we consider an auxiliary value $D(T)$, denoting the sum of the distances of all vertices from the root. This is also known as the *total height* [14] of the tree T. Other names for $D(T)$ are *total path length* or *internal path length*. It will help us to derive functional equations for generating functions yielding to asymptotic formulas.

The main results of this paper are the following:

Theorem 1 *Let $\mathcal{D} \subseteq \mathbb{N}$ be an arbitrary subset of the positive integers such that $\mathcal{D} \neq \{1\}$ and $\gcd(d : d \in \mathcal{D}) = 1$. Then the average total height $D(T_n)$ of a random tree $T_n \in \mathcal{F}_\mathcal{D}$ with n vertices is asymptotically $2Kn^{3/2}$, the average Wiener index is asymptotically $Kn^{5/2}$, where K depends on $\mathcal{D}$.*

Theorem 2 *Let $\mathcal{D} \subset \mathbb{N}$ be a subset of the positive integers as in Theorem 1. Then the average Wiener index of a tree $T_n \in \tilde{\mathcal{F}}_\mathcal{D}$ chosen uniformly at random is asymptotically $Kn^{5/2}$, with the same K as in Theorem 1.*

It will be shown in Section 5 how K can be (numerically) computed. Finally, there are similar formulas for the variances and the covariance of $D(T)$ and $W(T)$:

Theorem 3 *Let T_n be a tree with n vertices chosen uniformly at random from $\mathcal{F}_\mathcal{D}$.*

Then we have, for the variance of $D(T_n)$ and $W(T_n)$ and the covariance of the two,

$$\mathrm{Var}(D(T_n)) \sim \frac{4K^2(10-3\pi)}{3\pi} n^3,$$
$$\mathrm{Cov}(D(T_n), W(T_n)) \sim \frac{2K^2(16-5\pi)}{5\pi} n^4,$$
$$\mathrm{Var}(W(T_n)) \sim \frac{K^2(16-5\pi)}{5\pi} n^5.$$

Also, if $\tilde{T}_n$ is a tree with n vertices chosen uniformly at random from $\tilde{\mathcal{F}}_{\mathcal{D}}$, we have

$$\mathrm{Var}(W(\tilde{T}_n)) \sim \frac{K^2(16-5\pi)}{5\pi} n^5,$$

where the value of K is again the same as in Theorem 1.

Of course, it is not surprising that the asymptotic behaviour is essentially the same as in the known case of simply generated families of trees [5]. However, apart from an exact knowledge of the involved constants, the functional equations of Section 3 also allow exact calculations of the expected values without generating all trees of a certain family, which is of use in chemical applications (cf. [3]).

2 Preliminaries

For an element α of a permutation group A, let $j_k(\alpha)$ be the number of k-cycles in the decomposition of α into disjoint cycles. The *cycle index* of A, denoted by $Z(A)$, is defined as the following polynomial in $s_1, s_2, \ldots$ (cf. [6]):

$$Z(A) = |A|^{-1} \sum_{\alpha \in A} \prod_{k=1}^{n} s_k^{j_k(\alpha)}.$$

We write $Z(A, f(z))$ for the cycle index $Z(A)$ with $f(z^l)$ substituted for the variable s_l. If $T_{\mathcal{G}}(z)$ and $T_{\mathcal{G}_k}(z)$ are the counting series for two classes $\mathcal{G}$, $\mathcal{G}_k$ of rooted trees, where $\mathcal{G}_k$ is constructed by attaching a collection of k trees from the family $\mathcal{G}$ to a common root (ignoring the order), we have, using Pólya's theorem [6],

$$T_{\mathcal{G}_k}(z) = z \cdot Z(S_k, T_{\mathcal{G}}(z)), \tag{2}$$

where S_k denotes the symmetric group. Additionally, we define $Z(S_0, f(z)) = 1$ and $Z(S_k, f(z)) = 0$ for $k < 0$. This gives us, for example, the functional equation for the counting series $T_3(z)$ of rooted trees with maximal outdegree ≤ 3:

$$T_3(z) = z \cdot \sum_{k=0}^{3} Z(S_k, T_3(z)).$$

In Section 5, we will make use of the following property of the cycle indices of symmetric groups:

Lemma 4 The cycle index $Z(S_k)$ of the symmetric group S_k satisfies the equation

$$\frac{\partial}{\partial s_l} Z(S_k) = \frac{1}{l} Z(S_{k-l}).$$

Proof: From [6], we know that the cycle index of S_k has the explicit representation

$$Z(S_k) = \frac{1}{k!} \sum_{(j)} h(j) \prod_{r=1}^{k} s_r^{j_r},$$

where the sum runs over all partitions $(j) = (j_1, \ldots, j_k)$ of k (j_r denotes the number of parts equal to r) and $h(j)$ is given by

$$h(j) = \frac{k!}{\prod_{r=1}^{k} r^{j_r} j_r!}.$$

There is an obvious bijection between the partitions of k which contain l and the partitions of $k-l$. For a partition (j) of k that contains l, let (j') be the partition of $k-l$ which results from replacing j_l by $j_l - 1$. Then it is easy to see that

$$h(j') = \frac{(k-l)! l j_l h(j)}{k!}.$$

This shows that

$$\begin{aligned} \frac{\partial}{\partial s_l} Z(S_k) &= \frac{1}{k!} \sum_{(j)} \frac{j_l h(j)}{s_l} \prod_{r=1}^{k} s_r^{j_r} \\ &= \frac{1}{(k-l)!} \sum_{(j')} \frac{h(j')}{l} \prod_{r=1}^{k} s_r^{j'_r} = \frac{1}{l} Z(S_{k-l}). \end{aligned}$$

□

Corollary 5

$$\frac{d}{dz} Z(S_k, f(z)) = \sum_{l=1}^{k} z^{l-1} f'(z^l) Z(S_{k-l}, f(z)).$$

Proof: This follows trivially upon application of the chain rule. □

3 Functional equations for the total height and Wiener index

The auxiliary value $D(T)$ can be calculated recursively from the branches $T_1, \ldots, T_k$ of T [5], viz.

$$D(T) = \sum_{i=1}^{k} D(T_i) + |T| - 1, \tag{3}$$

where $|T|$ is the size (number of vertices) of T. Now we have to translate this recursive property into a functional equation. Again, we suppose that the branches come from a certain family $\mathcal{G}$, and denote the corresponding generating function for $D(T)$ by

$$D_{\mathcal{G}}(z) = \sum_{T \in \mathcal{G}} D(T) z^{|T|}.$$

Let $\mathcal{G}_k$ be defined as before and define $D_{\mathcal{G}_k}(z)$ analogously. There is an obvious bijection between the elements of $\mathcal{G}_{k-j}$ and the elements of $\mathcal{G}_k$ which contain a certain tree $T \in \mathcal{G}$ at least j times as a branch. Therefore, if $g_{k,n}$ denotes the number of trees of size n in $\mathcal{G}_k$, the branch B appears

$$\sum_{j=1}^{k} g_{k-j,n-j|B|}$$

times in all rooted trees of size n belonging to $\mathcal{G}_k$. Together with (3), this gives us

$$\begin{aligned} D_{\mathcal{G}_k}(z) &= \sum_{B \in \mathcal{G}} D(B) \sum_{j=1}^{k} \sum_{n \geq 1} g_{k-j,n-j|B|} z^n + z T'_{\mathcal{G}_k}(z) - T_{\mathcal{G}_k}(z) \\ &= z \sum_{j=1}^{k} D_{\mathcal{G}}(z^j) Z(S_{k-j}, T_{\mathcal{G}}(z)) + z T'_{\mathcal{G}_k}(z) - T_{\mathcal{G}_k}(z). \end{aligned} \tag{4}$$

Similarly, we introduce generating functions for the Wiener index:

$$W_{\mathcal{G}}(z) = \sum_{T \in \mathcal{G}} W(T) z^{|T|},$$

and $W_{\mathcal{G}_k}(z)$ is defined analogously. Now, we use the following recursive relation from [5], which relates the Wiener index of a rooted tree T to the Wiener indices of its branches $T_1, \ldots, T_k$:

$$W(T) = D(T) + \sum_{i=1}^{k} W(T_i) + \sum_{i \neq j} \Big(D(T_i) + |T_i| \Big) |T_j|, \tag{5}$$

where the last sum goes over all $k(k-1)$ pairs of different branches. Now, we have to determine the number of times the pair $(B_1, B_2) \in \mathcal{G}^2$ appears in trees with n vertices belonging to $\mathcal{G}_k$. By the same argument that was applied before, this number is given by

$$\sum_{j=1}^{k-1} \sum_{i=1}^{k-j} g_{k-j-i,n-j|B_1|-i|B_2|}$$

if B_1 and B_2 are distinct elements of $\mathcal{G}$. If, on the other hand, $B_1 = B_2 = B$ are equal, the number is

$$\sum_{j=1}^{k} j(j-1) \Big(g_{k-j,n-j|B|} - g_{k-j-1,n-(j+1)|B|} \Big) = \sum_{j=1}^{k} 2(j-1) g_{k-j,n-j|B|}.$$

Together with (5), this yields

$$\begin{aligned} W_{\mathcal{G}_k}(z) &= D_{\mathcal{G}_k}(z) + \sum_{B\in\mathcal{G}} W(B) \sum_{j=1}^{k} \sum_{n\geq 1} g_{k-j,n-j|B|} z^n \\ &+ \sum_{B_1\in\mathcal{G}} \sum_{B_2\in\mathcal{G}} \Big(D(B_1) + |B_1|\Big)|B_2| \sum_{j=1}^{k-1} \sum_{i=1}^{k-j} \sum_{n\geq 1} g_{k-j-i,n-j|B_1|-i|B_2|} z^n \\ &+ \sum_{B\in\mathcal{G}} \Big(D(B) + |B|\Big)|B| \sum_{j=1}^{k} \sum_{n\geq 1} (j-1) g_{k-j,n-j|B|} z^n \end{aligned}$$

or

$$\begin{aligned} W_{\mathcal{G}_k}(z) &= D_{\mathcal{G}_k}(z) + z \sum_{j=1}^{k} W_{\mathcal{G}}(z^j) Z(S_{k-j}, T_{\mathcal{G}}(z)) \\ &+ z \sum_{j=1}^{k-1} \sum_{i=1}^{k-j} \Big(D_{\mathcal{G}}(z^j) + z^j T'_{\mathcal{G}}(z^j)\Big) \cdot z^i T'_{\mathcal{G}}(z^i) Z(S_{k-j-i}, T_{\mathcal{G}}(z)) \qquad (6) \\ &+ z \sum_{j=1}^{k} (j-1) z^j \Big(D'_{\mathcal{G}}(z^j) + T'_{\mathcal{G}}(z^j) + z^j T''_{\mathcal{G}}(z^j)\Big) Z(S_{k-j}, T_{\mathcal{G}}(z)). \end{aligned}$$

These functional equations (and combinations of them for different values of k) enable us to calculate the average Wiener indices for various sorts of degree-restricted rooted trees. For the study of unrooted trees, however, we need yet another tool. In particular, we want to determine the average Wiener index of trees with maximal degree ≤ 4, also known as chemical trees (cf. [3]).

By a theorem of Otter (cf. [6]), the number of different representations of a tree as a rooted tree equals 1 plus the number of representations as a pair of two unequal rooted trees (the order being irrelevant), with their roots joined by an edge. Thus, for counting the trees in $\tilde{\mathcal{F}}_{\mathcal{D}}$, one has to take

- rooted trees with $k \in \tilde{\mathcal{D}}$ branches from $\mathcal{F}_{\mathcal{D}}$

minus

- pairs of unequal rooted trees from $\mathcal{F}_{\mathcal{D}}$, joined by an edge.

If $T_{\mathcal{D}}$ and $\tilde{T}_{\mathcal{D}}$ are the respective generating functions for the number of trees in $\mathcal{F}_{\mathcal{D}}$ and $\tilde{\mathcal{F}}_{\mathcal{D}}$, this means that

$$\tilde{T}_{\mathcal{D}}(z) = z + z \sum_{k\in\mathcal{D}_0} Z(S_{k+1}, T_{\mathcal{D}}(z)) - \frac{1}{2}\Big(T_{\mathcal{D}}^2(z) - T_{\mathcal{D}}(z^2)\Big). \qquad (7)$$

The first summand, corresponding to the tree with only a single vertex, can be included or not, since it makes no real difference. The generating function for the

Wiener index of trees from $\tilde{\mathcal{F}}_{\mathcal{D}}$ is also a difference of the respective generating functions for the two possibilities of representing a tree from $\tilde{\mathcal{F}}_{\mathcal{D}}$ which were given above. If we denote it by $\tilde{W}_{\mathcal{D}}(z) = \tilde{W}^{(1)}_{\mathcal{D}}(z) - \tilde{W}^{(2)}_{\mathcal{D}}(z)$, the first summand is given by equation (8), which is easily deduced from (4) and (6).

$$\begin{aligned}\tilde{W}^{(1)}_{\mathcal{D}}(z) = \sum_{k\in\tilde{D}} \Bigg(& z\sum_{j=1}^{k} D_{\mathcal{D}}(z^j) Z(S_{k-j}, T_{\mathcal{D}}(z)) + z\left(\frac{d}{dz} z \cdot Z(S_k, T_{\mathcal{D}}(z))\right) \\ & - z \cdot Z(S_k, T_{\mathcal{D}}(z)) + z\sum_{j=1}^{k} W_{\mathcal{D}}(z^j) Z(S_{k-j}, T_{\mathcal{D}}(z)) \\ & + z\sum_{j=1}^{k-1}\sum_{i=1}^{k-j}\Big(D_{\mathcal{D}}(z^j) + z^j T'_{\mathcal{D}}(z^j)\Big)\cdot z^i T'_{\mathcal{D}}(z^i) Z(S_{k-j-i}, T_{\mathcal{D}}(z)) \\ & + z\sum_{j=1}^{k}(j-1)z^j\Big(D'_{\mathcal{D}}(z^j) + T'_{\mathcal{D}}(z^j) + z^j T''_{\mathcal{D}}(z^j)\Big) Z(S_{k-j}, T_{\mathcal{D}}(z))\Bigg),\end{aligned} \tag{8}$$

On the other hand, if two rooted trees T_1 and T_2 are joined by an edge, the Wiener index of the resulting tree T is given by

$$W(T) = W(T_1) + W(T_2) + D(T_1)|T_2| + D(T_2)|T_1| + |T_1||T_2|.$$

Therefore, we obtain

$$\begin{aligned}\tilde{W}^{(2)}_{\mathcal{D}}(z) &= \frac{1}{2}\sum_{T_1\in\mathcal{F}_{\mathcal{D}}}\sum_{T_2\in\mathcal{F}_{\mathcal{D}}}\Big(W(T_1) + W(T_2) + D(T_1)|T_2| + D(T_2)|T_1| + |T_1||T_2|\Big) z^{|T_1|+|T_2|} \\ &\quad - \frac{1}{2}\sum_{T\in\mathcal{F}_{\mathcal{D}}}\Big(2W(T) + 2D(T)|T| + |T|^2\Big) z^{2|T|} \\ &= \frac{1}{2}\Big(2W_{\mathcal{D}}(z)T_{\mathcal{D}}(z) + 2D_{\mathcal{D}}(z)\cdot zT'_{\mathcal{D}}(z) + z^2 T'_{\mathcal{D}}(z)^2 \\ &\quad - 2W_{\mathcal{D}}(z^2) - 2z^2 D'_{\mathcal{D}}(z^2) - z^2(z^2 T''_{\mathcal{D}}(z^2) + T'_{\mathcal{D}}(z^2))\Big).\end{aligned} \tag{9}$$

4 Wiener index of trees and chemical trees

Equations (4), (6), (8) and (9) enable us to calculate the exact average Wiener index of all trees of size n from a certain family $\mathcal{F}$ with degree restrictions for considerably high n. As an example, we calculate the average Wiener index of all chemical trees (i.e. maximal degree ≤ 4) up to $n = 100$. We have to start with the generating function T_3 for $\mathcal{F}_3$, the class of rooted trees with maximal outdegree ≤ 3, whose functional equation is given by

$$T_3(z) = z\cdot\sum_{k=0}^{3} Z(S_k, T_3(z)).$$

Then, the generating function for the number of trees with degree ≤ 4 is given by

$$\tilde{T}_3(z) = z\sum_{k=0}^{4} Z(S_k, T_3(z)) - \frac{1}{2}\Big(T_3^2(z) - T_3(z^2)\Big).$$

From (4), we know that the corresponding generating function for $D(T)$ satisfies

$$D_3(z) = z\sum_{k=1}^{3}\sum_{j=1}^{k} D_3(z^j)Z(S_{k-j}, T_3(z)) + zT_3'(z) - T_3(z).$$

Analogously, from (6), we obtain

$$\begin{aligned} W_3(z) = D_3(z) + \sum_{k=1}^{3}\Bigg(& z\sum_{j=1}^{k} W_3(z^j)Z(S_{k-j}, T_3(z)) \\ & + z\sum_{j=1}^{k-1}\sum_{i=1}^{k-j}\Big(D_3(z^j) + z^jT_3'(z^j)\Big)\cdot z^iT_3'(z^i)Z(S_{k-j-i}, T_3(z)) \\ & + z\sum_{j=1}^{k}(j-1)z^j\Big(D_3'(z^j) + T_3'(z^j) + z^jT_3''(z^j)\Big)Z(S_{k-j}, T_3(z))\Bigg). \end{aligned}$$

$\tilde{W}_3$, the generating function for the sum of the Wiener indices of all trees with maximal degree ≤ 4, is then given by (8) and (9). Easy computer calculations yield Table 1; up to $n = 20$, the values were given in [3] by direct computation; $\tilde{t}_{4,n}$ denotes the number of trees of size n with maximal degree ≤ 4, $\tilde{w}_{4,n}$ the total of their Wiener indices.

n	$\tilde{t}_{4,n}$	$\tilde{w}_{4,n}$	$\tilde{w}_{4,n}/\tilde{t}_{4,n}$
1	1	0	0
2	1	1	1
3	1	4	4
4	2	19	9.5
5	3	54	18
6	5	155	31
7	9	432	48
8	18	1252	69.56
9	35	3384	96.69
10	75	9714	129.52
20	366319	310884129	848.67
50	$1.11774 \cdot 10^{18}$	$1.05659 \cdot 10^{22}$	9452.93
100	$5.92107 \cdot 10^{39}$	$3.34957 \cdot 10^{44}$	56570.38

Table 1: Some numerical values for chemical trees.

n	w_n	$\tilde{w}_n$	w_n/t_n	$\tilde{w}_n/\tilde{t}_n$
1	0	0	0	0
2	1	1	1	1
3	8	4	4	4
4	38	19	9.5	9.5
5	164	54	18.22222	18
6	609	180	30.45	30
7	2256	508	47	46.18182
8	7815	1533	67.95652	66.65217
9	26892	4332	94.02797	92.17021
10	90146	13041	125.37691	123.02830
20	10319401978	655274837	804.55470	796.13984
50	$3.73537 \cdot 10^{24}$	$9.20871 \cdot 10^{22}$	8768.95009	8732.57790
100	$2.66359 \cdot 10^{48}$	$3.25933 \cdot 10^{46}$	51836.59972	51724.32112

Table 2: Some numerical values for trees.

Things are somewhat easier in the case of ordinary trees. If $\mathcal{D} = \mathbb{N}$, the functional equations reduce to

$$D(z) = T(z) \sum_{j \geq 1} D(z^j) + zT'(z) - T(z),$$

$$\begin{aligned} W(z) &= D(z) + T(z) \sum_{j \geq 1} W(z^j) + \sum_{j \geq 1} \sum_{i \geq 1} \Big(D(z^j) + z^j T'(z^j) \Big) \cdot z^i T'(z^i) \cdot T(z) \\ &\quad + \sum_{j \geq 1} (j-1) z^j \Big(D'(z^j) + T'(z^j) + z^j T''(z^j) \Big) \cdot T(z), \\ \tilde{W}(z) &= W(z) - \frac{1}{2} \Big(2W(z)T(z) + 2D(z) \cdot zT'(z) + z^2 T'(z)^2 \\ &\quad - 2W(z^2) - 2z^2 D'(z^2) - z^2 (z^2 T''(z^2) + T'(z^2)) \Big). \end{aligned}$$

These equations are also given in Moon's paper [10]. Numerical values are provided in Table 2.

5 Asymptotic analysis

Now we are going to prove the main asymptotic results as stated in the introduction, in particular the following fairly general theorem:

Theorem 1 *Let $\mathcal{D} \subseteq \mathbb{N}$ be an arbitrary subset of the positive integers such that $\mathcal{D} \neq \{1\}$ and $\gcd(d : d \in \mathcal{D}) = 1$. Then the average total height $D(T_n)$ of a tree*

$T_n \in \mathcal{F}_{\mathcal{D}}$ with n vertices chosen uniformly at random is asymptotically $2Kn^{3/2}$, the average Wiener index is asymptotically $Kn^{5/2}$, where K is given by

$$K = \frac{\sqrt{\pi}}{2\alpha b \rho^{3/2}}$$

and α, b and ρ are defined as follows:

- *ρ is the radius of convergence of $T_{\mathcal{D}}(z)$,*
- *The expansion of $T_{\mathcal{D}}(z)$ around ρ is given by*

$$T_{\mathcal{D}}(z) = t_0 - b\sqrt{\rho - z} + O(\rho - z), \tag{10}$$

- *$\alpha = \sum_{k\in\mathcal{D}} Z(S_{k-2}, T_{\mathcal{D}}(z))|_{z=\rho}$.*

Remark: If $\mathcal{D} = \mathbb{N}$, we have $\alpha = \frac{1}{\rho} = 2.95576528\ldots$, $\rho = 0.33832185\ldots$ and $b = 2.68112814\ldots$, the constants given by Otter [12].

Proof: We fix $\mathcal{D}$ and use the abbreviations T, D, W for $T_{\mathcal{D}}$, $D_{\mathcal{D}}$, $W_{\mathcal{D}}$. We start with the equation

$$T(z) = z \sum_{k\in\mathcal{D}_0} Z(S_k, T(z)). \tag{11}$$

The gcd-condition for $\mathcal{D}$ ensures that all but finitely many coefficients of T are positive. Following [6, pp. 208–214], one can prove that T has positive radius of convergence $1 > \rho \geq 0.33832\ldots$ (the lower bound being given by the case $\mathcal{D} = \mathbb{N}$), that T converges at $z = \rho$ and that ρ is the only singularity on the circle of convergence. Furthermore, T has an expansion of the form (10) around ρ, giving an asymptotic formula for the number $t_{\mathcal{D},n}$ of trees of size n in $\mathcal{F}_{\mathcal{D}}$:

$$t_{\mathcal{D},n} \sim \frac{b}{2\sqrt{\pi}} \rho^{-n+1/2} n^{-3/2}.$$

The values of ρ, t_0 and b can be determined numerically. Differentiating (11) yields, by Corollary 5,

$$\begin{aligned} T'(z) &= \frac{T(z)}{z} + z \sum_{k\in\mathcal{D}} \sum_{l=1}^{k} z^{l-1} T'(z^l) Z(S_{k-l}, T(z)) \\ &= \frac{T(z)}{z} + zT'(z) \sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z)) + \sum_{k\in\mathcal{D}} \sum_{l=2}^{k} z^l T'(z^l) Z(S_{k-l}, T(z)) \end{aligned}$$

and thus

$$T'(z)\left(1 - z\sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z))\right) = \frac{T(z)}{z} + \sum_{k\in\mathcal{D}} \sum_{l=2}^{k} z^l T'(z^l) Z(S_{k-l}, T(z)). \tag{12}$$

We set

$$\beta := \sum_{k\in\mathcal{D}} \sum_{l=2}^{k} z^l T'(z^l) Z(S_{k-l}, T(z))\Big|_{z=\rho}.$$

Note, on this occasion, that $T(z^l)$ is holomorphic within a larger circle than $T(z)$ if $l > 1$, and that the sum over l can be uniformly bounded by a geometric sum on any compact subset of this larger circle. Furthermore, since it is a well-known fact that

$$\sum_{k\geq 0} Z(S_k, f(z)) = \exp\left(\sum_{m\geq 1} \frac{1}{m} f(z^m)\right),$$

we know that the sum over all $k \in \mathcal{D}$ converges as the sum $\sum_{m\geq 1} \frac{1}{m} T(\rho^m)$ is bounded. This argument will be used quite frequently in the following steps without being mentioned explicitly. Now, expanding around ρ gives us

$$1 - z \sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z)) \sim \frac{2}{b}\left(\frac{t_0}{\rho} + \beta\right)\sqrt{\rho - z}. \tag{13}$$

On the other hand, we have

$$\begin{aligned} \frac{d}{dz}\left(1 - z\sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z))\right) &= -\sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z)) - zT'(z)\sum_{k\in\mathcal{D}} Z(S_{k-2}, T(z)) \\ &\quad - z\sum_{k\in\mathcal{D}}\sum_{l=2}^{k-1} z^{l-1} T'(z^l) T(S_{k-1-l}, T(z)). \end{aligned}$$

The first and the last summand are bounded; therefore, if we set

$$\alpha := \sum_{k\in\mathcal{D}} Z(S_{k-2}, T(z))\Big|_{z=\rho},$$

we obtain

$$\frac{d}{dz}\left(1 - z\sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z))\right) \sim -\frac{\rho b \alpha}{2}(\rho - z)^{-1/2},$$

giving us $\alpha = \frac{2}{b^2\rho}\left(\frac{t_0}{\rho} + \beta\right)$. Next, we turn to the functional equation for $D(z)$:

$$D(z) = zT'(z) - T(z) + zD(z)\sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z)) + z\sum_{k\in\mathcal{D}}\sum_{l=2}^{k} D(z^l) Z(S_{k-l}, T(z)). \tag{14}$$

The last summand is bounded around ρ; note that $D(z)$ has the same radius of convergence as $T(z)$, since $D(T) \leq \frac{|T|(|T|-1)}{2}$ for all trees T; the same argument holds true for the generating function of the Wiener index. Solving for $D(z)$ yields

$$D(z) = \frac{zT'(z) - T(z) + z\sum_{k\in\mathcal{D}}\sum_{l=2}^{k} D(z^l) Z(S_{k-l}, T(z))}{1 - z\sum_{k\in\mathcal{D}} Z(S_{k-1}, T(z))}.$$

Therefore the expansion of $D(z)$ around ρ is given by

$$D(z) \sim \frac{b^2\rho^2}{4(t_0+\beta\rho)}(\rho-z)^{-1} = \frac{1}{2\alpha}(\rho-z)^{-1}, \tag{15}$$

which follows upon combining (10), (13) and the formula for α.

Finally, we consider the function $W(z)$:

$$\begin{aligned} W(z) &= D(z) + zW(z)\sum_{k\in\mathcal{D}} Z(S_{k-1},T(z)) + z\sum_{k\in\mathcal{D}}\sum_{j=2}^{k} W(z^j)Z(S_{k-j},T(z)) \\ &+ z\sum_{k\in\mathcal{D}}\sum_{j=1}^{k-1}\sum_{i=1}^{k-j}\Big(D(z^j)+z^jT'(z^j)\Big)\cdot z^iT'(z^i)Z(S_{k-j-i},T(z)) \\ &+ z\sum_{k\in\mathcal{D}}\sum_{j=1}^{k}(j-1)z^j\Big(D'(z^j)+T'(z^j)+z^jT''(z^j)\Big)Z(S_{k-j},T(z)). \end{aligned} \tag{16}$$

We extract the asymptotically relevant terms to obtain

$$W(z)\left(1 - z\sum_{k\in\mathcal{D}} Z(S_{k-1},T(z))\right) = z^2D(z)T'(z)\sum_{k\in\mathcal{D}} Z(S_{k-2},T(z)) + O((\rho-z)^{-1}).$$

The right hand side of this equation behaves like $\frac{\rho^2 b}{4}(\rho-z)^{-3/2}$, so this yields

$$W(z) \sim \frac{\rho}{4\alpha}(\rho-z)^{-2}. \tag{17}$$

Thus, if $t_{\mathcal{D},n}$, $d_{\mathcal{D},n}$ and $w_{\mathcal{D},n}$ denote the coefficients of $T(z)$, $D(z)$ and $W(z)$ respectively, we have

$$t_{\mathcal{D},n} \sim \frac{b}{2\sqrt{\pi}}\rho^{-n+1/2}n^{-3/2},\ d_{\mathcal{D},n} \sim \frac{1}{2\alpha}\rho^{-n-1},\ w_{\mathcal{D},n} \sim \frac{1}{4\alpha}\rho^{-n-1}n.$$

So the average values of $D(T_n)$ and $W(T_n)$ for $T_n \in \mathcal{F}_{\mathcal{D}}$ are given by

$$\frac{d_{\mathcal{D},n}}{t_{\mathcal{D},n}} \sim \frac{\sqrt{\pi}}{\alpha b\rho^{3/2}}n^{3/2},\ \frac{w_{\mathcal{D},n}}{t_{\mathcal{D},n}} \sim \frac{\sqrt{\pi}}{2\alpha b\rho^{3/2}}n^{5/2},$$

which finally proves the claim. □

In the same manner, we prove our second main theorem:

Theorem 2 *Let $\mathcal{D} \subset \mathbb{N}$ be a subset of the positive integers as in Theorem 1. Then the average Wiener index of a random tree $T_n \in \tilde{\mathcal{F}}_{\mathcal{D}}$ is asymptotically $Kn^{5/2}$, where K is defined as in Theorem 1.*

Proof: We use the abbreviations T, D, W again and write $\tilde{T}$, $\tilde{W}$ for $\tilde{T}_{\mathcal{D}}$, $\tilde{W}_{\mathcal{D}}$. We consider the generating function $\tilde{T}(z)$ first:

$$\tilde{T}(z) = z + z \sum_{k \in \mathcal{D}_0} Z(S_{k+1}, T(z)) - \frac{1}{2}\Big(T^2(z) - T(z^2)\Big). \tag{18}$$

Clearly, $\tilde{T}(z)$ must have the same radius of convergence as T, and ρ is the only singularity of $\tilde{T}(z)$ on the circle of convergence. Thus we have to determine the expansion of $\tilde{T}(z)$ around ρ. First, we differentiate (18):

$$\begin{aligned}\tilde{T}'(z) &= 1 + \sum_{k \in \mathcal{D}_0} Z(S_{k+1}, T(z)) + z \sum_{k \in \mathcal{D}_0} \sum_{l=1}^{k+1} z^{l-1} T'(z^l) Z(S_{k+1-l}, T(z)) \\ &\qquad\qquad - T(z)T'(z) + zT'(z^2) \\ &= 1 + \sum_{k \in \mathcal{D}_0} Z(S_{k+1}, T(z)) + T'(z)\left(z \sum_{k \in \mathcal{D}_0} Z(S_k, T(z)) - T(z)\right) \\ &\quad + z \sum_{k \in \mathcal{D}_0} \sum_{l=2}^{k+1} z^{l-1} T'(z^l) Z(S_{k+1-l}, T(z)) + zT'(z^2) \\ &= 1 + \sum_{k \in \mathcal{D}_0} Z(S_{k+1}, T(z)) + z \sum_{k \in \mathcal{D}} \sum_{l=2}^{k+1} z^{l-1} T'(z^l) Z(S_{k+1-l}, T(z)) + zT'(z^2).\end{aligned}$$

Thus the derivative of $\tilde{T}(z)$ is bounded at $z = \rho$. Differentiating again yields

$$\tilde{T}''(z) = \sum_{k \in \mathcal{D}_0} T'(z) Z(S_k, T(z)) + z \sum_{k \in \mathcal{D}} \sum_{l=2}^{k+1} z^{l-1} T'(z^l) T'(z) Z(S_{k-l}, T(z)) + \ldots,$$

the remaining terms being bounded at $z = \rho$. We find that

$$\tilde{T}''(z) \sim \left(\beta + \frac{t_0}{\rho}\right) T'(z) = \frac{b^2 \alpha \rho}{2} T'(z)$$

around $z = \rho$. This means that $\tilde{T}(z)$ has an expansion of the form

$$\tilde{T}(z) = \tilde{t}_0 + a_1(\rho - z) + \frac{b^3 \alpha \rho}{3} (\rho - z)^{3/2} + O((\rho - z)^2), \tag{19}$$

giving the asymptotic formula for the number $\tilde{t}_{\mathcal{D},n}$ of trees of size n in $\tilde{\mathcal{F}}_{\mathcal{D}}$:

$$t_{\mathcal{D},n} \sim \frac{b^3 \alpha}{4\sqrt{\pi}} \rho^{-n+5/2} n^{-5/2}.$$

We only have to determine the expansion of $\tilde{W}(z)$ now. This function is given by $\tilde{W}(z) = \tilde{W}^{(1)}(z) - \tilde{W}^{(2)}(z)$, where $\tilde{W}^{(1)}$ and $\tilde{W}^{(2)}$ are given by (8) and (9) respectively.

We extract all asymptotically relevant parts and obtain

$$
\begin{aligned}
\tilde{W}^{(1)}(z) &= z(D(z)+W(z))\sum_{k\in\tilde{\mathcal{D}}} Z(S_{k-1},T(z)) \\
&\qquad + z^2T'(z)(D(z)+zT'(z))\sum_{k\in\tilde{\mathcal{D}}} Z(S_{k-2},T) \\
&\qquad + zD(z)\sum_{k\in\tilde{\mathcal{D}}}\sum_{l=2}^{k-1} z^lT'(z^l)Z(S_{k-1-l},T(z)) + O((\rho-z)^{-1/2}) \\
&= z(D(z)+W(z))\sum_{k\in\mathcal{D}_0} Z(S_k,T(z)) \\
&\qquad + z^2T'(z)(D(z)+zT'(z))\sum_{k\in\mathcal{D}} Z(S_{k-1},T) \\
&\qquad + zD(z)\sum_{k\in\mathcal{D}}\sum_{l=2}^{k} z^lT'(z^l)Z(S_{k-l},T(z)) + O((\rho-z)^{-1/2}).
\end{aligned}
\tag{20}
$$

and

$$
\tilde{W}^{(2)}(z) = W(z)T(z) + zT'(z)D(z) + \frac{z^2}{2}T'(z)^2 + O((\rho-z)^{-1/2}). \tag{21}
$$

Now, we make use of equations (11) and (12). Some algebraic manipulations then lead us to

$$
\begin{aligned}
\tilde{W}(z) &= (D(z)+W(z))T(z) - W(z)T(z) + zT'(z)D(z)\left(z\sum_{k\in\mathcal{D}} Z(S_{k-1},T) - 1\right) \\
&\qquad + \frac{z^2}{2}T'(z)^2 + z^2T'(z)^2\left(z\sum_{k\in\mathcal{D}} Z(S_{k-1},T) - 1\right) \\
&\qquad + zD(z)\sum_{k\in\mathcal{D}}\sum_{l=2}^{k} z^lT'(z^l)Z(S_{k-l},T(z)) + O((\rho-z)^{-1/2}) \\
&= D(z)T(z) + \frac{z^2}{2}T'(z)^2 \\
&\qquad - (D(z)+zT'(z))\left(T(z) + z\sum_{k\in\mathcal{D}}\sum_{l=2}^{k} z^lT'(z^l)Z(S_{k-l},T(z))\right) \\
&\qquad + zD(z)\cdot\beta + O((\rho-z)^{-1/2}) \\
&= D(z)\cdot t_0 + \frac{z^2}{2}T'(z)^2 - (D(z)+zT'(z))(t_0+\rho\beta) + D(z)\cdot\rho\beta + O((\rho-z)^{-1/2}) \\
&= \frac{z^2}{2}T'(z)^2 + O((\rho-z)^{-1/2}).
\end{aligned}
$$

Therefore, the expansion of $\tilde{W}$ around ρ is given by

$$
\tilde{W}(z) \sim \frac{\rho^2b^2}{8}(\rho-z)^{-1}, \tag{22}
$$

giving us an asymptotic formula for the coefficients of $\tilde{W}(z)$:

$$\tilde{w}_{\mathcal{D},n} \sim \frac{b^2}{8}\rho^{-n+1}.$$

Dividing by $\tilde{t}_{\mathcal{D},n}$ finally yields the theorem. □

As a conclusion, we give numerical values of K for $\mathcal{D} = \{1, \ldots, M\}$ in some special cases:

M	$K(M)$
2	0.7842482154
3	0.6418839467
4	0.5962854459
5	0.5790571390
10	0.5683583008
∞	0.5682799594

Table 3: Some numerical values of K.

Remark: The theorem still holds—mutatis mutandis—when the gcd-condition for $\mathcal{D}$ is violated. In this case, there are several singularities of equal behaviour on the circle of convergence. If, for example, $\mathcal{D} = \{3\}$ (in this case, $\tilde{\mathcal{F}}_{\mathcal{D}}$ corresponds to saturated hydrocarbons), there are only trees in $\mathcal{F}_{\mathcal{D}}$ with a number of vertices $n \equiv 1 \mod 3$, and their average Wiener index is asymptotically $0.3705918694n^{5/2}$.

The proof of Theorem 3 follows essentially along the same lines. However, the involved functional equations are rather long and tedious; therefore, no details are given here for the sake of brevity. The interested reader is referred to [17]. For instance, in the case $\mathcal{D} = \mathbb{N}$, the recurrence

$$\begin{aligned} D(T)^2 &= \left(\sum_{i=1}^{k} D(T_i) + |T| - 1\right)^2 = 2D(T)(|T|-1) - (|T|-1)^2 + \left(\sum_{i=1}^{k} D(T_i)\right)^2 \\ &= 2D(T)(|T|-1) - (|T|-1)^2 + \sum_{i=1}^{k} D(T_i)^2 + \sum_{i\neq j} D(T_i)D(T_j) \end{aligned}$$

yields the functional equation

$$\begin{aligned} D_2(z) &= 2zD'(z) - 2D(z) - z^2T''(z) + zT'(z) - T(z) \\ &+ \left(\sum_{i\geq 1}\sum_{j\geq 1} D(z^i)D(z^j) + \sum_{i\geq 1} iD_2(z^i)\right)T(z) \end{aligned} \tag{23}$$

for the generating function

$$D_2(z) := \sum_{T} D(T)^2 z^{|T|}.$$

Acknowledgment

The author would like to thank an anonymous referee whose suggestions helped to improve this paper.

References

[1] A. Cayley, A theorem on trees, *Quart. J. Math. Oxford Ser.* 23 (1889), 376–378.

[2] A. A. Dobrynin, R. Entringer and I. Gutman, Wiener index of trees: theory and applications, *Acta Appl. Math.* 66(3) (2001), 211–249.

[3] A. A. Dobrynin and I. Gutman, The Average Wiener Index of Trees and Chemical Trees, *J. Chem. Inf. Comput. Sci.* 39 (1999), 679–683.

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

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

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

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

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

[9] A. Meir and J. W. Moon, On the altitude of nodes in random trees, *Canad. J. Math.* 30(5) (1978), 997–1015.

[10] J. W. Moon, On the expected distance from the centroid of a tree, *Ars Combin.* 20(A) (1985), 263–276.

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

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

[13] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* 68 (1937), 145–254.

[14] J. Riordan and N. J. A. Sloane, The enumeration of rooted trees by total height, *J. Austral. Math. Soc.* 10 (1969), 278–282.

[15] N. J. A. Sloane and S. Plouffe, *The encyclopedia of integer sequences*, Academic Press Inc., San Diego, 1995. Online edition available at `http://www.research.att.com/~njas/sequences`.

[16] N. Trinajstić, *Chemical graph theory*, CRC Press, Boca Raton, FL., 1992.

[17] S. Wagner, *Graph-theoretical enumeration and digital expansions—an analytic approach*, PhD thesis, Graz University of Technology, 2006. Available at `http://finanz.math.tugraz.at/~wagner`.

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

(Received 27 Jan 2006)