

The Wiener index of random trees

RALPH NEININGER¹
School of Computer Science
McGill University
3480 University Street
Montreal, H3A 2K6
Canada

September 19, 2001

Abstract

The Wiener index is analyzed for random recursive trees and random binary search trees in the uniform probabilistic models. We obtain the expectations, asymptotics for the variances, and limit laws for this parameter. The limit distributions are characterized as the projections of bivariate measures that satisfy certain fixed-point equations. Covariances, asymptotic correlations, and bivariate limit laws for the Wiener index and the internal path length are given.

AMS subject classifications. Primary: 60F05, 05C12; secondary: 05C05, 68Q25.

Key words. Wiener index, weak convergence, distance (in a graph), random binary search tree, random recursive tree, contraction method, bivariate limit law.

1 Introduction and results

The Wiener index of a connected graph is defined as the sum of the distances between all unordered pairs of vertices of the graph, where the distance between two vertices is the minimum number of edges connecting them in the graph. The index was introduced by the chemist H. Wiener about 50 years ago to demonstrate correlations between physico-chemical properties of organic compounds and the index of their molecular graphs. Therefore, the graphs of most practical interest have natural restrictions on their degrees corresponding to the valences of the atoms and are typically trees or have hexagonal or pentagonal cycles. See as general references the monographs of Trinajstić [21] and Gutman and Polansky [11].

While much effort has been made to calculate the Wiener index of various graphs effectively, recently the chemical relevance of taking averages over certain classes of trees or graphs was also noted. Dobrynin and Gutman [8] propose using a normalized Wiener index, where the normalization is with respect to an average of the Wiener index over some set of (molecular) graphs. This is appropriate when molecules of a series under consideration have different sizes (numbers of vertices in their molecular graphs), but the property examined is independent of molecular size and should be correlated to the Wiener index. Dobrynin and Gutman state, “Until now, however, normalized

¹Research supported by NSERC grant A3450 and the Deutsche Forschungsgemeinschaft.

Wiener indices have not been used, because the values $\langle W \rangle$ were not available. [8]" (Here, $\langle W \rangle$ denotes an appropriate average of the Wiener index.)

In combinatorics, mathematicians occasionally considered averages of the Wiener index for combinatorial classes of trees such as simply generated families of trees. In particular, exact and asymptotic formulæ for ordinary ordered trees, rooted labeled trees, and rooted binary trees are given in Entringer, Meir, Moon, and Székely [9].

In this note we give a probabilistic analysis for the Wiener index of random recursive trees and random binary search trees. A tree with n vertices labeled $1, \dots, n$ is a recursive tree if the vertex with label 1 is distinguished as the root, and for each k , $2 \leq k \leq n$, the labels of the vertices in the path from the root to the vertex labeled k form an increasing sequence. There exist $(n - 1)!$ recursive trees with n vertices, and a *random* recursive tree with n vertices is one chosen with equal probability from this set. For reference see the survey of Smythe and Mahmoud [20]. A binary search tree is a data structure built up from an ordered set of distinct numbers. The first number becomes the root of the tree. Then the numbers are successively inserted recursively; each number is compared with the root. If it is smaller than the root, it goes to the left subtree, otherwise to the right subtree. There this procedure is recursively iterated until we reach an empty subtree, where the number is inserted. A *random* binary search tree with n vertices is one built up from an equiprobable permutation of the numbers $1, \dots, n$. For reference see Knuth [13].

For both tree models we obtain the expectations, asymptotic expressions for the variances, and limit laws for the Wiener index as well as the covariances and correlations with the internal path length, and a bivariate limit law for the Wiener index and the internal path length. The internal path length of a rooted tree is defined as the sum of the distances between all vertices and the root.

The analysis is based on the recursive structure of the trees and makes use of the contraction method, originally introduced for the probabilistic analysis of recursive algorithms; see the survey of Rösler and Rüschemdorf [19]. The characteristic of the Wiener index problem from the point of view of the contraction method is that in the natural recurrence we have a certain dependence between the quantities which can usually not be handled by the method. We resolve this problem by considering the bivariate recurrence for the joint distribution of the Wiener index and the internal path length, where the unpleasant dependence can be somehow hidden, resulting in the applicability of a general transfer theorem of Neininger [16]. From the bivariate asymptotic results we then extract the asymptotic relations for the Wiener index.

We give the proofs in section two. Section three contains some remarks and states open problems connected to the present note. In the rest of this section we fix some notation and state our results.

We denote by H_n the n -th harmonic number $H_n = \sum_{i=1}^n 1/i$, by A^t the transposed of a vector or matrix A , by $x_n \sim y_n$ asymptotic equivalence of sequences of real numbers, i.e., $x_n/y_n \rightarrow 1$ for $n \rightarrow \infty$. By $X_n \xrightarrow{\mathcal{L}} X$ convergence in distribution of the random variates X_n to X is denoted, $\mathcal{L}(X)$ denotes the distribution of X ; \mathbb{E} , Var , Cov , and Cor denote the mean, variance, covariance and correlation. By \mathcal{M}_2 we abbreviate the space of all bivariate, centered probability measures with finite second moments. Finally, we set $\mathcal{E}(U) := U \ln(U) + (1 - U) \ln(1 - U)$.

Theorem 1.1 Let (W_n, P_n) denote the vector of the Wiener index and the internal path length of a random binary search tree with n vertices. Then we have

$$\begin{aligned}
\mathbb{E} W_n &= 2n^2 H_n - 6n^2 + 8nH_n - 10n + 6H_n, & (1) \\
\text{Var}(W_n) &\sim \frac{20 - 2\pi^2}{3} n^4, \\
\text{Cov}(W_n, P_n) &\sim \frac{20 - 2\pi^2}{3} n^3, \\
\text{Cor}(W_n, P_n) &\sim \sqrt{\frac{20 - 2\pi^2}{21 - 2\pi^2}} = 0.4548\dots, \\
\left(\frac{W_n - \mathbb{E} W_n}{n^2}, \frac{P_n - \mathbb{E} P_n}{n} \right) &\xrightarrow{\mathcal{L}} (W, P), & (2)
\end{aligned}$$

where $\mathcal{L}(W, P)$ is the unique fixed-point of the map $T : \mathcal{M}_2 \rightarrow \mathcal{M}_2$ given for $\nu \in \mathcal{M}_2$ by

$$T(\nu) := \mathcal{L} \left(\begin{bmatrix} U^2 & U(1-U) \\ 0 & U \end{bmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} + \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{bmatrix} \begin{pmatrix} Z'_1 \\ Z'_2 \end{pmatrix} + \begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix} \right),$$

with

$$\begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix} := \begin{pmatrix} 6U(1-U) + 2\mathcal{E}(U) \\ 1 + 2\mathcal{E}(U) \end{pmatrix},$$

where $(Z_1, Z_2), (Z'_1, Z'_2), U$ are independent with $\mathcal{L}(Z_1, Z_2) = \mathcal{L}(Z'_1, Z'_2) = \nu$ and U uniform $[0, 1]$ distributed.

The expression (1) was already given in [12]. For the random recursive tree we have:

Theorem 1.2 Let (W_n, P_n) denote the vector of the Wiener index and the internal path length of a random recursive tree with n vertices. Then we have

$$\begin{aligned}
\mathbb{E} W_n &= n^2 H_n - 2n^2 + nH_n, \\
\text{Var}(W_n) &\sim \frac{31 - 3\pi^2}{18} n^4, \\
\text{Cov}(W_n, P_n) &\sim \frac{21 - 2\pi^2}{12} n^3, \\
\text{Cor}(W_n, P_n) &\sim \frac{\sqrt{3}(21 - 2\pi^2)}{2\sqrt{(12 - \pi^2)(31 - 3\pi^2)}} = 0.6342\dots, \\
\left(\frac{W_n - \mathbb{E} W_n}{n^2}, \frac{P_n - \mathbb{E} P_n}{n} \right) &\xrightarrow{\mathcal{L}} (W, P),
\end{aligned}$$

where $\mathcal{L}(W, P)$ is the unique fixed-point of the map T defined in Theorem 1.1 with (b_1^*, b_2^*) there replaced by

$$\begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix} := \begin{pmatrix} 3U(1-U) + \mathcal{E}(U) \\ U + \mathcal{E}(U) \end{pmatrix}.$$

2 Analysis

We use the recursive structure of the trees under consideration in order to setup a recurrence for the Wiener index. We start with the analysis of the binary search tree. Let I_n and $J_n = n - 1 - I_n$ denote the cardinalities of the left and right subtree of the root of a binary search tree containing n vertices. We denote by $(W_{I_n}, P_{I_n}), (W'_{J_n}, P'_{J_n})$ the pairs of the Wiener index and the internal path length in the left and right subtree of the root respectively. Thus by direct enumeration we obtain the recurrence

$$W_n = W_{I_n} + W'_{J_n} + b_n,$$

where

$$b_n = (P_{I_n} + P'_{J_n} + n - 1) + J_n P_{I_n} + I_n P'_{J_n} + 2I_n J_n.$$

While such a decomposition is known to be inferior to other representations from the practical point of view of figuring out W_n for a given tree, see [6, section 4], the recursive decomposition is the basis of the present approach. In the *random* binary search tree W_n and P_n become random variables. It is known that the cardinality of the left subtree I_n is uniformly distributed over $\{0, \dots, n - 1\}$ and that, conditioned on this cardinality I_n , the left and right subtree have the distributions of random binary search trees of cardinalities I_n and J_n respectively and are (stochastically) independent of each other. This implies that with two sequences $(W_n, P_n), (W'_n, P'_n)$ of pairs of Wiener indices and internal path lengths in random binary search trees such that $(W_n, P_n), (W'_n, P'_n)$ and I_n are independent we obtain the distributional recurrence

$$W_n \stackrel{\mathcal{D}}{=} W_{I_n} + W'_{J_n} + b_n, \quad n \geq 1, \quad (3)$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution of the left and right hand side. Note that we have $W_0 = 0$.

We proceed in three steps: First we derive, using (3), the mean of W_n as outlined in [12]. In the second step we setup a distributional recurrence for the vector (W_n, P_n) and apply the contraction method in form of a general transfer theorem as given in [16], which is restated for the readers convenience. This leads to the limit law and the characterization of the limit in form of a fixed-point equation. The reason why we cannot apply the method directly to the recurrence (3) is that conditioned on I_n the quantities b_n, W_{I_n}, W'_{J_n} are dependent, where independence would be essential for the application of the method. This dependence is caused by the dependence of the Wiener index and the internal path length in each subtree. In the last step we derive the second (mixed) moments from the bivariate fixed-point equation.

Proof of Theorem 1.1. *Expectation:* Denote $\alpha_n := \mathbb{E} W_n$ and $\gamma_n := \mathbb{E} P_n$. We have $\alpha_0 = 0$ and recursively by conditioning on I_n in (3)

$$\alpha_n = \beta_n + \frac{2}{n} \sum_{k=0}^{n-1} \alpha_k, \quad n \geq 1, \quad (4)$$

with

$$\beta_n := \mathbb{E} b_n = \frac{1}{n} \sum_{k=0}^{n-1} (\gamma_k + \gamma_{n-1-k} + n - 1 + 2k(\gamma_{n-1-k} + n - 1 - k)).$$

With the well-known formula $\gamma_n = 2(n+1)H_n - 4n$, $n \geq 0$, we obtain from identities for harmonic numbers that

$$\beta_n = \left(\frac{2}{3}H_n - \frac{14}{9} \right) (n+1)(n+2) + \frac{7}{3}(n+1) + \frac{2}{3}, \quad n \geq 1. \quad (5)$$

Note that the recurrence (4) initialized by $\alpha_0 = 0$ can be resolved to, Lemma 1 in [12],

$$\alpha_n = \beta_n + 2(n+1) \sum_{k=1}^{n-1} \frac{\beta_k}{(k+1)(k+2)}, \quad n \geq 1.$$

Plugging in the expression for β_n leads to the formula for $\mathbb{E}W_n$ given in (1).

Limit law: We also have

$$P_n \stackrel{\mathcal{D}}{=} P_{I_n} + P'_{J_n} + n - 1, \quad n \geq 1,$$

and obtain as well a distributional recurrence for the bivariate quantities (W_n, P_n) :

$$\begin{pmatrix} W_n \\ P_n \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W_{I_n} \\ P_{I_n} \end{pmatrix} + \begin{bmatrix} 1 & n - J_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W'_{J_n} \\ P'_{J_n} \end{pmatrix} + \begin{pmatrix} 2I_n J_n + n - 1 \\ n - 1 \end{pmatrix}.$$

The rescaled quantities $X_0 := 0$ and

$$X_n := \left(\frac{W_n - \mathbb{E}W_n}{n^2}, \frac{P_n - \mathbb{E}P_n}{n} \right)^t, \quad n \geq 1,$$

and the analogously defined X'_n satisfy the recurrence

$$X_n \stackrel{\mathcal{D}}{=} A_1^{(n)} X_{I_n} + A_2^{(n)} X'_{J_n} + b^{(n)}, \quad n \geq 1, \quad (6)$$

where

$$\begin{aligned} A_1^{(n)} &= \begin{bmatrix} 1/n^2 & 0 \\ 0 & 1/n \end{bmatrix} \begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_n^2 & 0 \\ 0 & I_n \end{bmatrix} \\ &= \begin{bmatrix} (I_n/n)^2 & I_n(n - I_n)/n^2 \\ 0 & I_n/n \end{bmatrix}, \\ A_2^{(n)} &= \begin{bmatrix} (J_n/n)^2 & J_n(n - J_n)/n^2 \\ 0 & J_n/n \end{bmatrix}, \end{aligned}$$

and $b^{(n)} = (b_1^{(n)}, b_2^{(n)})$ with

$$\begin{aligned} \begin{pmatrix} b_1^{(n)} \\ b_2^{(n)} \end{pmatrix} &= \begin{bmatrix} 1/n^2 & 0 \\ 0 & 1/n \end{bmatrix} \left(\begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \alpha_{I_n} \\ \gamma_{I_n} \end{pmatrix} + \begin{bmatrix} 1 & n - J_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \alpha_{J_n} \\ \gamma_{J_n} \end{pmatrix} \right. \\ &\quad \left. - \begin{pmatrix} \alpha_n \\ \gamma_n \end{pmatrix} + \begin{pmatrix} 2I_n J_n + n - 1 \\ n - 1 \end{pmatrix} \right). \end{aligned}$$

For the asymptotic analysis of $b^{(n)}$ we plug in expansions for α_n and γ_n , obtained from expanding $H_n = \ln(n) + \gamma + o(1)$ for $n \rightarrow \infty$, where γ denotes Euler's constant,

$$\begin{aligned} \alpha_n &= 2n^2 \ln(n) + (2\gamma - 6)n^2 + o(n^2), \\ \gamma_n &= 2n \ln(n) + (2\gamma - 4)n + o(n). \end{aligned}$$

After cancellation of some terms this yields with the convention $x \ln(x) := 0$ for $x = 0$,

$$\begin{aligned} b_1^{(n)} &= \frac{1}{n^2} \left(2I_n^2 \ln \left(\frac{I_n}{n} \right) + 2J_n^2 \ln \left(\frac{J_n}{n} \right) + 2I_n J_n \ln \left(\frac{I_n}{n} \right) \right. \\ &\quad \left. + 2I_n J_n \ln \left(\frac{J_n}{n} \right) + 6I_n J_n \right) + o(1), \\ b_2^{(n)} &= \frac{1}{n} \left(2I_n \ln \left(\frac{I_n}{n} \right) + 2J_n \ln \left(\frac{J_n}{n} \right) + n \right) + o(1), \end{aligned}$$

where the $o(1)$ s are random but the convergences hold uniformly. We may take the freedom to model all quantities on a joint probability space such that $I_n/n \rightarrow U$ for a uniform $[0, 1]$ distributed random variate, where the convergence holds almost surely and thus in L_2 . Then, by dominated convergence, we obtain the following convergences in L_2 :

$$A_1^{(n)} \rightarrow A_1^* := \begin{bmatrix} U^2 & U(1-U) \\ 0 & U \end{bmatrix}, \quad (7)$$

$$A_2^{(n)} \rightarrow A_2^* := \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{bmatrix},$$

$$b^{(n)} \rightarrow b^* := \begin{pmatrix} 2U \ln(U) + 2(1-U) \ln(1-U) + 6U(1-U) \\ 2U \ln(U) + 2(1-U) \ln(1-U) + 1 \end{pmatrix}. \quad (8)$$

The situation is now tailored for the application of Theorem 4.1 in Neininger [16], which says: Let a sequence (X_n) satisfy a recurrence as (6) such that for some A_1^*, A_2^*, b^* we have

$$(A_1^{(n)}, A_2^{(n)}, b^{(n)}) \rightarrow (A_1^*, A_2^*, b^*) \text{ in } L_2, \quad n \rightarrow \infty, \quad (9)$$

$$\mathbb{E} \left\| (A_1^*)^t A_1^* \right\|_{\text{op}} + \mathbb{E} \left\| (A_2^*)^t A_2^* \right\|_{\text{op}} < 1, \quad (10)$$

$$\mathbb{E} \left[\mathbf{1}_{\{I_n \leq \ell\} \cup \{I_n = n\}} \left\| (A_1^{(n)})^t A_1^{(n)} \right\|_{\text{op}} \right] \rightarrow 0, \quad (11)$$

$$\mathbb{E} \left[\mathbf{1}_{\{J_n \leq \ell\} \cup \{J_n = n\}} \left\| (A_2^{(n)})^t A_2^{(n)} \right\|_{\text{op}} \right] \rightarrow 0$$

for all $\ell \in \mathbb{N}$ and $n \rightarrow \infty$, where $\|A\|_{\text{op}} := \sup_{\|x\|=1} \|Ax\|$ denotes the spectral radius of a matrix A and $\mathbf{1}_B$ the indicator function of a set B . Then (X_n) converges in distribution and with second moments to a distribution $\mathcal{L}(X)$, which is the unique fixed-point of the map $T : \mathcal{M}_2 \rightarrow \mathcal{M}_2$ given by $T(\nu) := \mathcal{L}(A_1^* Z + A_2^* Z' + b^*)$, where (A_1^*, A_2^*, b^*) , Z, Z' are independent with $\mathcal{L}(Z) = \mathcal{L}(Z') = \nu$.

Thus the verification of (9)–(11) implies the limit law (2) with the characterization of $\mathcal{L}(W, P)$ as given in the theorem. Condition (9) is satisfied by (7) and (8). Condition (11) holds since $\|(A_r^{(n)})^t A_r^{(n)}\|_{\text{op}}$ are deterministically bounded, $r = 1, 2$, and

$$\mathbb{P}(\{I_n \leq \ell\} \cup \{I_n = n\}) = \mathbb{P}(\{J_n \leq \ell\} \cup \{J_n = n\}) \leq \frac{\ell + 1}{n} \rightarrow 0$$

for all $\ell \in \mathbb{N}$ and $n \rightarrow \infty$.

It remains to check condition (10): Solving the characteristic equation for $(A_1^*)^t A_1^*$ we obtain that the eigenvalue $\lambda(U)$ of $(A_1^*)^t A_1^*$ being larger in absolute value is given by

$$\lambda(U) = U^2 \left(\frac{1 + U^2 + (1-U)^2}{2} + \sqrt{\frac{(1 + U^2 + (1-U)^2)^2}{4} - U^2} \right).$$

This implies, since $(A_1^*)^t A_1^*$ and $(A_2^*)^t A_2^*$ are identically distributed, that

$$\begin{aligned}\mathbb{E} \|(A_1^*)^t A_1^*\|_{\text{op}} + \mathbb{E} \|(A_2^*)^t A_2^*\|_{\text{op}} &= 2 \mathbb{E} \lambda(U) \\ &= \frac{3}{10} + \frac{29}{60} \sqrt{2} + \frac{1}{4} \ln(\sqrt{2} - 1) \\ &< 1.\end{aligned}$$

Thus condition (10) is fulfilled and the limit law (2) holds.

Second (mixed) moments: Let $(W, P), (W', P'), U$ be independent with $\mathcal{L}(W, P) = \mathcal{L}(W', P')$ being the fixed-point of T in \mathcal{M}_2 and U uniform $[0, 1]$ distributed. Thus we have $\mathbb{E}W = \mathbb{E}P = 0$ and

$$\begin{pmatrix} W \\ P \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} U^2 W + U(1-U)(P + P') + (1-U)^2 W' + 6U(1-U) + 2\mathcal{E}(U) \\ UP + (1-U)P' + 1 + 2\mathcal{E}(U) \end{pmatrix}.$$

Using the independence and $\mathbb{E}W = \mathbb{E}P = 0$ this yields

$$\begin{aligned}\mathbb{E}[WP] &= \mathbb{E}[U^3 WP] + \mathbb{E}[(1-U)^3 W' P'] \\ &\quad + \mathbb{E}[U^2(1-U)P^2] + \mathbb{E}[U(1-U)^2(P')^2] \\ &\quad + \mathbb{E}[(6U(1-U) + 2\mathcal{E}(U))(1 + 2\mathcal{E}(U))].\end{aligned}$$

Thus, with $\mathbb{E}P^2 = 7 - 2\pi^2/3$, we obtain

$$\text{Cov}(W, P) = \mathbb{E}[WP] = \frac{20 - 2\pi^2}{3}.$$

With this we can resolve for $\mathbb{E}W^2$. We have

$$\begin{aligned}\mathbb{E}W^2 &= \mathbb{E}[U^4 W^2] + \mathbb{E}[(1-U)^4 (W')^2] \\ &\quad + 2\mathbb{E}[U^3(1-U)WP] + 2\mathbb{E}[U(1-U)^3 W' P'] \\ &\quad + \mathbb{E}[U^2(1-U)^2(P + P')^2] + \mathbb{E}[(6U(1-U) + 2\mathcal{E}(U))^2],\end{aligned}$$

which implies

$$\mathbb{E}W^2 = \frac{20 - 2\pi^2}{3}.$$

This yields the correlation of the components W, P of the fixed-point:

$$\text{Cor}(W, P) = \sqrt{\frac{20 - 2\pi^2}{21 - 2\pi^2}}.$$

Since the convergence $((W_n - \mathbb{E}W_n)/n^2, (P_n - \mathbb{E}P_n)/n) \rightarrow (W, P)$ holds with second moments, this implies $\text{Var}(W_n) \sim n^4 \text{Var}(W)$, $\text{Cov}(W_n, P_n) \sim n^3 \text{Cov}(W, P)$ and $\text{Cor}(W_n, P_n) \sim \text{Cor}(W, P)$, which finishes the proof. \blacksquare

For recursive trees the proof given above can easily be adjusted:

Proof of Theorem 1.2. (*Sketch.*) Consider the random recursive tree of n vertices labeled $1, \dots, n$, $n \geq 2$. We decompose the tree into the subtree rooted at the vertex labeled 2 and the

rest of the tree, which is then rooted at the root of the original tree. It is known, see Smythe and Mahmoud [20] and Dobrow and Fill [5], that the cardinality I_n of the subtree rooted at the vertex labeled 2 is uniformly distributed on $\{1, \dots, n-1\}$ for $n \geq 2$ and that conditioned on its cardinality this subtree and the rest of the tree are random recursive trees of cardinalities I_n and $J_n := n - I_n$ respectively being independent of each other. Thus with two sequences $(W_n, P_n), (W'_n, P'_n)$ of pairs of Wiener indices and internal path lengths in random recursive trees such that $(W_n, P_n), (W'_n, P'_n)$ and I_n are independent, we obtain by enumeration

$$W_n \stackrel{\mathcal{D}}{=} W_{I_n} + W'_{J_n} + b_n, \quad n \geq 2, \quad (12)$$

with

$$b_n = J_n P_{I_n} + I_n P'_{J_n} + I_n J_n$$

and $(W_0, P_0) = (W_1, P_1) = (0, 0)$. Since $P_n \stackrel{\mathcal{D}}{=} P_{I_n} + P'_{J_n} + I_n$, $n \geq 2$, we have

$$\begin{pmatrix} W_n \\ P_n \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W_{I_n} \\ P_{I_n} \end{pmatrix} + \begin{bmatrix} 1 & n - J_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W'_{J_n} \\ P'_{J_n} \end{pmatrix} + \begin{pmatrix} I_n J_n \\ I_n \end{pmatrix}.$$

For the calculation of the expectation note that for $\alpha_n := \mathbb{E} W_n$ we have $\alpha_0 = \alpha_1 = 0$ and, by (12) and conditioning on I_n ,

$$\alpha_n = \beta_n + \frac{2}{n-1} \sum_{k=2}^{n-1} \alpha_k, \quad n \geq 2, \quad (13)$$

where β_n can be calculated as in the binary search tree case to be

$$\beta_n = \frac{1}{3}n(n+1)H_n - \frac{4}{9}n(n+1) + \frac{1}{3}n, \quad n \geq 2.$$

Recurrence (13), initialized by $\alpha_0 = \alpha_1 = 0$, can be resolved to

$$\alpha_n = \beta_n + 2n \sum_{k=2}^{n-1} \frac{\beta_k}{k(k+1)}, \quad n \geq 2.$$

This leads to the expression for $\mathbb{E} W_n$ stated.

Then we scale as in the binary search tree case and obtain a recurrence for the scaled quantities. Again the corresponding coefficients $A_1^{(n)}, A_2^{(n)}$ converge to the A_1^*, A_2^* as defined in (7). Using the expansions $\mathbb{E} W_n = n^2 \ln(n) + (\gamma - 2)n^2 + o(n^2)$ and $\mathbb{E} P_n = n \ln(n) + (\gamma - 1)n + o(1)$ we obtain the limit (b_1^*, b_2^*) of the adjusted $b^{(n)}$ as given in Theorem 1.2. The conditions (9)–(11) are satisfied and the (mixed) second moments can be extracted from the fixed-point equation as shown for the binary search tree case. ■

3 Remarks

We conclude with some remarks and open problems:

Internal path length. Our limit laws give in the second components as well the limit law for the internal path length of the random trees under consideration. These were obtained before in

Régnier [17] and Mahmoud [14] showing that suitably scaled versions of the internal path length form a martingale and were rederived with characterizations of the limit distributions in Rösler [18] and Dobrow and Fill [5], where Rösler used the approach we were following in the present paper, and Dobrow and Fill used the method of moments.

Random split trees. Devroye [3] introduced a model of a random tree, the random split tree, which for special choices of its parameters covers many common random tree models. Two important parameters of this class are the *branch factor* $b \in \mathbb{N}$ and the *splitter* V , which is a random variable in $[0, 1]$. For a subclass of this tree class the internal path length P_n satisfies

$$\mathbb{E} P_n = \frac{1}{\mu} n \ln(n) + c_p n + o(n)$$

with $\mu := b \mathbb{E}[V \ln(1/V)]$ and some constant $c_p \in \mathbb{R}$, see Neininger [15, Chapter 2]. The present approach may be applicable for the analysis of the Wiener index of trees in this subclass. However, a necessary relation would be the expansion

$$\mathbb{E} W_n = \frac{1}{\mu} n^2 \ln(n) + c_w n^2 + o(n^2),$$

with

$$c_w = c_p - \frac{b \mathbb{E} V^2}{1 - b \mathbb{E} V^2}.$$

Candidates could be random quadtrees, random m -ary search trees or random median-of- $(2k + 1)$ search trees, where the techniques developed in Flajolet, Labelle, Lafortest, and Salvy [10], Chern and Hwang [2], and Chern and Hwang [1] respectively could prove useful.

Simply generated families of trees. As shown in Entringer, Meir, Moon, and Székely [9] the mean of the Wiener index of some simply generated families of trees is of the order $n^{5/2}$, where we obtained for the recursive and binary search trees $n^2 \ln(n)$. This behavior is in accordance with the similar different scalings for the average depth of a vertex or the internal path length being of the orders \sqrt{n} versus $\ln(n)$ and $n^{3/2}$ versus $n \ln(n)$ respectively.

Digital tree structures. Trees based on bit comparisons such as digital search trees, tries or Patricia tries exhibit under various models of randomness the presence of periodic functions in the moments of many parameters. Does the Wiener index follow this pattern? Do we have asymptotic normality for the Wiener index in these structures?

Properties of limit distributions. The limit distributions in our theorems are given implicitly as the fixed-points of the maps T in \mathcal{M}_2 . From the fixed-point equations higher moments may successively be pumped out. What are further properties of these distributions? Could one approximate their distribution functions or other characteristics?

Independent toll functions. In Hwang and Neininger [12] recursions of the type (3) have been analyzed for various types of *toll functions* b_n . There, dependence between b_n and I_n is allowed, but a general assumption is that, conditioned on I_n , the quantities W_{I_n} , W'_{J_n} and b_n are independent. Clearly this is not satisfied for the Wiener index recurrence (3) caused by the dependence with the

internal path length. In the case of conditionally independent b_n the growth order $n^2 \ln(n)$ for $\mathbb{E} b_n$, cf. (5), would imply a variance of the order $n^4 \ln^2(n)$, see [12]. Thus the additional dependence of b_n in the Wiener index example causes the reduced order n^4 for the variance.

Random distance in a tree. A quantity related to the Wiener index is the distance of two distinct vertices chosen uniformly at random in a tree. While the Wiener index is the sum of all the distances these quantities match (up to scaling) only in the mean since the Wiener index is a sum of dependent distances. This is similar to the depth of a typical vertex being asymptotically normal distributed and the internal path length, the sum of all depths, being asymptotically non-normal distributed. For the distribution of distances of vertices with fixed labels in the random recursive tree and related results see Dobrow [4] and the references therein.

Molecular graphs. Not much is known about the stochastic behavior of the Wiener index for subclasses of graphs, which are more closely related to its chemical origin. For a discussions of random models for *hexagonal chains* and *chemical trees* as they arise as graph representations of certain molecules, see Dobrynin and Gutman [7, 8] and the references therein.

Acknowledgment

I thank Hsien-Kuei Hwang for pointing the Wiener index out to me and Andrey A. Dobrynin for making the papers [6, 7, 8] available to me.

References

- [1] Chern, H.-H. and Hwang, H.-K. (2001) Transitional behaviors of the average cost of Quicksort with median-of- $(2t + 1)$. *Algorithmica* **29** 44–69.
- [2] Chern, H.-H. and Hwang, H.-K. (2001) Phase changes in random m -ary search trees and generalized quicksort. *Random Structures Algorithms* **19** to appear.
- [3] Devroye, L. (1999) Universal limit laws for depths in random trees. *SIAM Journal on Computing* **28** 409–432.
- [4] Dobrow, R. P. (1996) On the distribution of distances in recursive trees. *J. Appl. Probab.* **33** 749–757.
- [5] Dobrow, R. P. and Fill, J. A. (1999) Total path length for random recursive trees. *Combin. Probab. Comput.* **8** 317–333.
- [6] Dobrynin, A. A., Entringer, R. and Gutman, I. (2001) Wiener index of trees: theory and applications. *Acta Appl. Math.* **66** 211–249.
- [7] Dobrynin, A. A. and Gutman, I. (1999) The average Wiener index of hexagonal chains. *Comput. Chem.* **23** 571–576.
- [8] Dobrynin, A. A. and Gutman, I. (1999) The average Wiener index of trees and chemical trees. *J. Chem. Inf. Comput. Sci.* **39** 679–683.

- [9] Entringer, R. C., Meir, A., Moon, J. W. and Székely, L. A. (1994) The Wiener index of trees from certain families. *Australas. J. Combin.* **10** 211–224.
- [10] Flajolet, P., Labelle, G., Laforest, L. and Salvy, B. (1995) Hypergeometrics and the cost structure of quadrees. *Random Structures Algorithms* **7** 117–144.
- [11] Gutman, I. and Polansky, O. E. (1986) *Mathematical concepts in organic chemistry*. Springer-Verlag, Berlin.
- [12] Hwang, H.-K. and Neininger, R. (2001) Phase change of limit laws in the quicksort recurrence under varying toll functions. Preprint.
- [13] Knuth, D. E. (1998) *The art of computer programming. Vol. 3: sorting and searching. 2nd ed.* Addison-Wesley.
- [14] Mahmoud, H. M. (1991) Limiting distributions for path lengths in recursive trees. *Probab. Engrg. Inform. Sci.* **5** 53–59.
- [15] Neininger, R. (1999) *Limit Laws for Random Recursive Structures and Algorithms*. Ph.D. Dissertation, Institut für Mathematische Stochastik, Universität Freiburg.
- [16] Neininger, R. (2001) On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms* **19** to appear.
- [17] Régnier, M. (1989) A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23** 335–343.
- [18] Rösler, U. (1991) A limit theorem for “Quicksort”. *RAIRO Inform. Théor. Appl.* **25** 85–100.
- [19] Rösler, U. and Rüschemdorf, L. (2001) The contraction method for recursive algorithms. *Algorithmica* **29** 3–33.
- [20] Smythe, R. T. and Mahmoud, H. M. (1994) A survey of recursive trees. *Teor. Ľmovĭr. Mat. Stat.* **51** 1–29.
- [21] Trinajstić, N. (1992) *Chemical graph theory*. CRC Press, Boca Raton, FL.