

On the parity of the Wiener index

Stephan Wagner

*Department of Mathematical Sciences, Stellenbosch University, Stellenbosch 7602,
South Africa*

Hua Wang

*Department of Mathematics, University of Florida, Gainesville, FL, 32611, United
States of America*

Abstract

It is a known fact that the Wiener index (i.e. the sum of all distances between pairs of vertices in a graph) of a tree with an odd number of vertices is always even. In this paper, we consider the distribution of the Wiener index and the related tree parameter “internal path length” modulo 2 by means of a generating functions approach as well as by constructing bijections for plane trees.

Key words. Wiener index, internal path length, trees, parity, bijections

1 Introduction and Preliminaries

The modeling and prediction of physico-chemical and biological properties of molecules is an important field of research in biochemistry. The selection of molecular structure descriptors (so-called topological indices) has continuously been an active research area. One of the first such indices was proposed by H. Wiener 60 years ago [14]; the properties of the Wiener index have been studied quite extensively in both the mathematical and chemical literature.

Since molecular structures are often acyclic due to natural restrictions on the valencies of the atoms, the Wiener index of trees has received much attention, [4] gives a nice survey on important results. It has been mentioned repeatedly in mathematical [4] and chemical papers [6] that the Wiener index of a tree

Email addresses: swagner@sun.ac.za (Stephan Wagner), hua@math.ufl.edu (Hua Wang).

with an odd number of vertices is always even. In fact, this has been considered as a drawback of the Wiener index, one of the most used topological indices. In the case when a tree has an even number of vertices, it is known that the Wiener index may be either even or odd. Hence it is natural to ask for the proportion of trees with an even/odd Wiener index among all trees with a given even number of vertices. We will prove, for simply generated trees (Section 2), the special case of plane trees (Section 3) and most importantly unrooted trees (Section 4), that there is asymptotically the same number of trees with an even resp. odd Wiener index among all trees with a given even number of vertices.

For a tree T , let $|T|$ be the number of vertices in T , and let $d_T(u, w)$ be the distance between u and w , two vertices in T . Then the Wiener index of T is simply expressed as

$$W(T) = \sum_{\{u,w\} \subseteq V(T)} d_T(u, w) = \frac{1}{2} \sum_{u \in V(T)} \sum_{w \in V(T)} d_T(u, w).$$

We start with an observation that simplifies the generating functions approach.

Lemma 1 Let T be a tree on $n = |T|$ vertices and v be a vertex of T . Then

$$W(T) \equiv (n - 1) \sum_{w \in V(T)} d_T(v, w) \pmod{2}.$$

Proof: For any two vertices u, w of T , note that

$$d_T(u, w) \equiv d_T(v, u) + d_T(v, w) \pmod{2}.$$

Therefore,

$$W(T) \equiv \sum_{\{u,w\} \subseteq V(T)} (d_T(v, u) + d_T(v, w)) = (n - 1) \sum_{w \in V(T)} d_T(v, w) \pmod{2},$$

which proves the lemma. \square

Remark 1 Of course, this also shows the well-known fact that the Wiener index of a tree with an odd number of vertices is always even (see [4]).

2 Simply generated families of trees

A special class of trees for which several kinds of enumeration problems have been successfully treated is the class of so-called *simply generated* trees in

the sense of Meir and Moon (cf. [11]). A simply generated family of trees is determined by a sequence $c_0 = 1, c_1, c_2, \dots$ of weights. The weight of a rooted ordered tree T is then given by

$$c(T) = \prod c_i^{N_i(T)},$$

where $N_i(T)$ is the number of vertices in T with exactly i children. Now one can define a generating function for the total weight of all trees on n vertices via

$$T(x) = \sum_T c(T)x^{|T|},$$

which satisfies a functional equation of the form $T(x) = x\Phi(T(x))$, where $\Phi(t) = \sum_{i=0}^{\infty} c_i t^i$. Special cases include plane trees ($\Phi(t) = \frac{1}{1-t}$), rooted labeled trees ($\Phi(t) = e^t$; in this case, the exponential generating function is used) and s -ary trees ($\Phi(t) = (1+t)^s$); the following theorem provides information on the asymptotic behavior of the coefficients of $T(x)$:

Theorem 2 (Meir/Moon [11]) *Suppose $\Phi(t)$ is an analytic function of t for $|t| < R \leq \infty$, and let $T(x) = \sum_{n \geq 0} t_n x^n$ denote the unique solution of $T(x) = x\Phi(T(x))$ in the neighborhood of $x = 0$. If*

- $c_k \geq 0$ for $k \geq 1$,
- $\gcd(k : c_k > 0) = 1$, and
- $\tau\Phi'(\tau) = \Phi(\tau)$ for some $0 < \tau < R$, then

$$t_n \sim a\rho^{-n}n^{-3/2}$$

as $n \rightarrow \infty$, where $\rho = \tau/\Phi(\tau)$ and $a = (\Phi(\tau)/(2\pi\Phi''(\tau)))^{1/2}$.

Now, let $D(T)$ be the internal path length (also known as total height, cf. [13]) of a rooted tree T , i.e. the sum of the distances from the root r :

$$D(T) = \sum_{w \in V(T)} d_T(r, w).$$

Then our lemma merely states that

$$W(T) \equiv (|T| - 1)D(T) \pmod{2}.$$

If T_1, T_2, \dots, T_k are the branches of T , we have the simple recurrences (cf. [5])

$$|T| = 1 + \sum_{i=1}^k |T_i|$$

and

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

Introducing the bivariate generating function

$$T(x, y) := \sum_T c(T) x^{|T|} y^{D(T)},$$

these recurrences translate to a functional equation for T :

$$T(x, y) = x\Phi(T(xy, y)).$$

We are interested in the subsets of trees with even and odd internal path length respectively. Their generating functions are given by

$$\frac{1}{2}(T(x, 1) + T(x, -1)) \quad \text{and} \quad \frac{1}{2}(T(x, 1) - T(x, -1)).$$

Hence, we have to investigate $S(x) := T(x, -1)$, which satisfies

$$S(x) = x\Phi(S(-x)) \quad \text{and} \quad S(-x) = -x\Phi(S(x)). \quad (1)$$

Furthermore, let Φ satisfy the conditions of Theorem 2. We know that the radius of convergence of $T(x) = T(x, 1)$ is ρ , and since all coefficients of $T(x, y)$ are nonnegative, the radius of convergence of $S(x) = T(x, -1)$ is certainly $\leq \rho$. However, we will show that S is holomorphic within a larger circle. To this end, we assume the contrary, i.e. that $S(x)$ has a singularity ρ_S with $|\rho_S| = \rho$. Since $T(\rho) = \tau$ exists (and is finite), $S(\rho_S)$ and $S(-\rho_S)$ have to be finite as well (making use of the fact that all coefficients of $T(x, y)$ are nonnegative once again). If $\sigma = S(\rho_S)$ and $\bar{\sigma} = S(-\rho_S)$, we have

$$\sigma = \rho_S \Phi(\bar{\sigma}), \quad \bar{\sigma} = -\rho_S \Phi(\sigma),$$

and the Jacobian determinant

$$\begin{vmatrix} 1 & -\rho_S \Phi'(\bar{\sigma}) \\ \rho_S \Phi'(\sigma) & 1 \end{vmatrix} = 1 + \rho_S^2 \Phi'(\sigma) \Phi'(\bar{\sigma}) \quad (2)$$

has to be 0. Otherwise, there would be a unique analytic continuation of $S(x)$ at the point ρ_S by the implicit function theorem. However, the latter is impossible: from our conditions, it is clear that $\rho_S \neq \rho$ and thus

$$|\sigma| < \tau \quad \text{and} \quad |\bar{\sigma}| < \tau$$

by the triangle inequality. Furthermore, since Φ' has only nonnegative coefficients,

$$|\Phi'(\sigma)| < \Phi'(\tau) \quad \text{and} \quad |\Phi'(\bar{\sigma})| < \Phi'(\tau),$$

and it follows that

$$\left| \rho_S^2 \Phi'(\sigma) \Phi'(\bar{\sigma}) \right| < \rho^2 \Phi'(\tau)^2 = \left(\frac{\tau \Phi'(\tau)}{\Phi(\tau)} \right)^2 = 1,$$

which makes it impossible that the Jacobian in (2) vanishes. Therefore, the function S is holomorphic in a circle of radius $> \rho$, which implies that its radius of convergence is $> \rho$, i.e.

$$[x^n]S(x) = [x^n]T(x, -1) = o(\rho^{-n}).$$

This implies that

$$[x^n] \left(\frac{1}{2} (T(x, 1) + T(x, -1)) \right) = \frac{1}{2} [x^n]T(x, 1) + o(\rho^{-n}) \sim \frac{a}{2} \rho^{-n} n^{-3/2}$$

and

$$[x^n] \left(\frac{1}{2} (T(x, 1) - T(x, -1)) \right) = \frac{1}{2} [x^n]T(x, 1) + o(\rho^{-n}) \sim \frac{a}{2} \rho^{-n} n^{-3/2},$$

and so we have the following theorem:

Theorem 3 *Let a simply generated family that satisfies the conditions of Theorem 2 be given. Then the total weight of all trees with n vertices and even internal path length is asymptotically the same as the total weight of all trees with odd internal path. Furthermore, the same statement holds for trees with an even number n of vertices and the Wiener index.*

Remark 2 Using k -th roots of unity instead of ± 1 , the proof can easily be generalized to the distribution of the internal path length modulo an arbitrary integer $k > 1$.

For example, let us consider the case of binary trees ($\Phi(t) = (1+t)^2$). We obtain the equation

$$S(x) = x(1 + S(-x))^2 = x \left(1 - x(1 + S(x))^2 \right)^2.$$

It is well known (see [1,2]) that the smallest singularity of a function that is given by such a polynomial equation of the form $F(S, x) = 0$ can be found by computing the common zeros of $F(S, x) = 0$ and $F_S(S, x) = 0$ (the only other potential singularities are zeros of the coefficient of the highest power of S in F ; but the only such zero is $x = 0$ in our case). Making use of Gröbner bases and the power of a computer algebra system, this leads to a single polynomial equation for x :

$$256x^4 - 288x^2 - 27 = 0.$$

This shows that the singularities of smallest modulus are $\pm \frac{3i}{4} \sqrt{\frac{2}{\sqrt{3}} - 1}$, i.e. the radius of convergence is $\tilde{\rho} = \frac{3}{4} \sqrt{\frac{2}{\sqrt{3}} - 1} = 0.2949899199$, and we get

$$[x^n]S(x) = O \left(n^{-3/2} \tilde{\rho}^{-n} \right).$$

Using standard singularity analysis, this could even be refined, see [8] and [9]. It is well known that the number of binary trees with n vertices is the *Catalan number*

$$\frac{1}{n+1} \binom{2n}{n},$$

which grows asymptotically like $\frac{1}{\sqrt{\pi n^3}} 4^n$ —note that $\tilde{\rho} > \frac{1}{4}$. Thus, the number of binary trees with a given number n of vertices and even (odd) internal path length is

$$\frac{1}{2(n+1)} \binom{2n}{n} + O(n^{-3/2} \tilde{\rho}^{-n}).$$

If n is even, this also gives the number of binary trees with even (odd) Wiener index.

3 Plane trees

In this section, we consider the special case of *plane trees* (i.e. $\Phi(t) = \frac{1}{1-t}$), where the equation (1) can be solved explicitly:

$$S(x) = \frac{x}{1 + x \cdot \frac{1}{1-S(x)}} = \frac{x(1-S(x))}{1+x-S(x)}$$

or equivalently

$$S(x)^2 - (1+2x)S(x) + x = 0.$$

The solution of this quadratic equation is given by

$$S(x) = \frac{1+2x - \sqrt{1+4x^2}}{2}.$$

Therefore, the number of plane trees with even/odd internal path length (and thus Wiener index in the case of an even number of vertices) can be written in terms of *Catalan numbers*:

- Trees with n vertices and even internal path length:

$$\begin{cases} \frac{1}{2n} \binom{2n-2}{n-1} + (-1)^{n/2} \cdot \frac{1}{n} \binom{n-2}{n/2-1} & n \text{ even} \\ \frac{1}{2n} \binom{2n-2}{n-1} & n \text{ odd, } n > 1 \end{cases}$$

- Trees with n vertices and odd internal path length:

$$\begin{cases} \frac{1}{2n} \binom{2n-2}{n-1} - (-1)^{n/2} \cdot \frac{1}{n} \binom{n-2}{n/2-1} & n \text{ even} \\ \frac{1}{2n} \binom{2n-2}{n-1} & n \text{ odd, } n > 1 \end{cases}$$

The only exceptional case is $n = 1$, where the only existing tree has internal path length 0. This is exactly the same distribution that has also been observed for plane trees with an even resp. odd number of leaves—we refer the interested reader to [3,7]. For this result, bijective proofs have been given as well as purely algebraic proofs, and so our task is now to prove our observation by means of bijections as well.

In the case of an odd number of vertices, it is easy to achieve our goal, and we mention three possible bijections. If the number of vertices is even, the situation is more intricate, and only the third method (which is based on an idea due to Chen, Shapiro and Yang [3]) works.

- (1) The simplest bijection can be described as follows: let r be the root of a plane tree T and v be its leftmost child. Furthermore, let T_v be the tree rooted at v and $T_r = T \setminus T_v$ the rest of the tree, rooted at r . Interchanging T_r and T_v is clearly an involution on the set of plane trees. Furthermore, the distance of $|T_v|$ vertices from the root decreases by 1, whereas the distance of $|T_r|$ vertices increases by 1. Since $|T_r| + |T_v| = |T|$ is odd, so is $|T_r| - |T_v|$, which implies that the parity of the internal path length changes.
- (2) The bijection used by Eu, Liu and Yeh in [7] is similar: either T_r or T_v has an odd number of vertices. Set $G(T) = T_r$ if $|T_r|$ is odd and $G(T) = T_v$ otherwise. Then, there is a k such that $G^{k+1}(T)$ is just a single vertex. Now, apply the operation described in (1) to the tree $G^k(T)$. This involution on the set of plane trees changes the parity of the number of leaves as well as the parity of the internal path length.
- (3) The third construction, a modified version of that described by Chen, Shapiro and Yang in [3], is more involved, but it turns out to be applicable to the case of an even number of vertices as well. It depends on the notion of *legal* and *illegal* vertices: a (non-root) vertex is called legal if it is a leaf and the leftmost child of an internal vertex or if it is an internal vertex and not the leftmost child. A plane tree is called legal if all its vertices (excluding the root) are legal (otherwise, the tree is called illegal). Obviously, any legal tree can be constructed by attaching a leaf as the leftmost child to every vertex of some given plane tree, which implies that all legal trees have an even number of vertices and that the number of legal trees with $2n$ vertices is exactly $\frac{1}{n} \binom{2n-2}{n-1}$. Furthermore, if T' is a legal tree with $2n$ vertices that is constructed from a tree T with n vertices in this way, then $D(T') = 2D(T) + n$, since the distance from the root of a vertex w that is attached to a vertex v is $(1 + \text{distance of } v \text{ from the root})$. This means that the parity of the internal path length of a legal tree only depends on its size.

Thus, we only have to define a parity-reversing involution on the set of illegal trees. To this end, we proceed as in [3] and perform a depth search for the first illegal vertex: starting at the root, we choose the rightmost

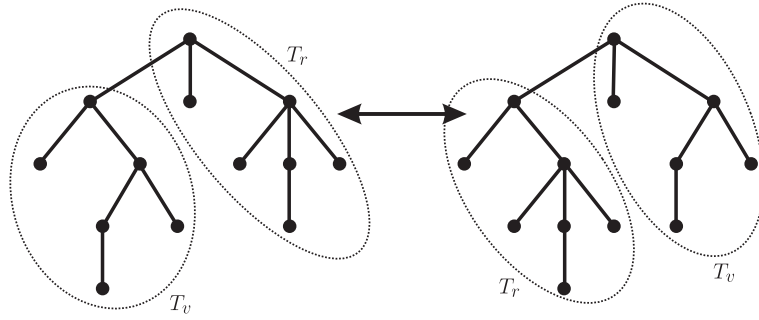


Fig. 1. Parity-reversing bijection (1)

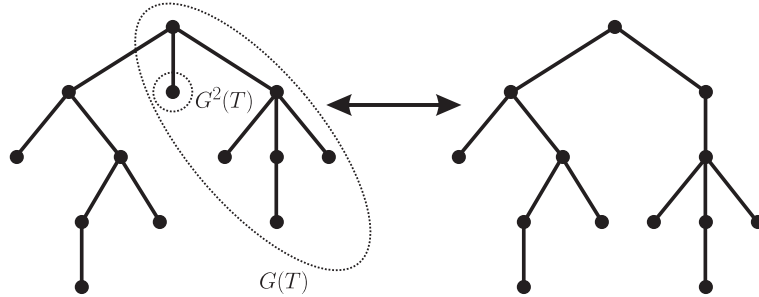


Fig. 2. Parity-reversing bijection (2)

illegal branch and iterate this process until we reach an illegal vertex (since every illegal tree contains at least one illegal vertex, this is always the case). Now, we have to distinguish two cases:

- If the illegal vertex v is a leaf, then it is not the leftmost child of its parent vertex w . Let u_1, u_2, \dots, u_k ($k \geq 1$) be the siblings of v that are to the left of v , and let T_{u_i} be subtree rooted at u_i . Now, we remove v , attach a new vertex v' to u_k in the leftmost position, and move the subtrees $T_{u_1}, T_{u_2}, \dots, T_{u_{k-1}}$ from w to v' . By this movement, the distances of the vertices in $T_{u_1}, T_{u_2}, \dots, T_{u_{k-1}}$ from the root change by 2, so their parities remain the same. The only other change in the internal path length results from replacing v by v' , which changes the distance from the root by exactly 1. Therefore, the parity of the internal path length is reversed by our procedure. Finally, note that through this process, u_k becomes the illegal vertex that is found by our depth search.
- If the illegal vertex is an internal node, we just reverse the procedure described in the first case.

The first bijection is illustrated in Figure 1, the second in Figure 2, and the third one in Figure 3. Let us end this section with a question that suggests itself now:

Problem 1 Is there a simple bijection between the set of plane trees with an even resp. odd number of leaves and those with an even resp. odd internal path length?

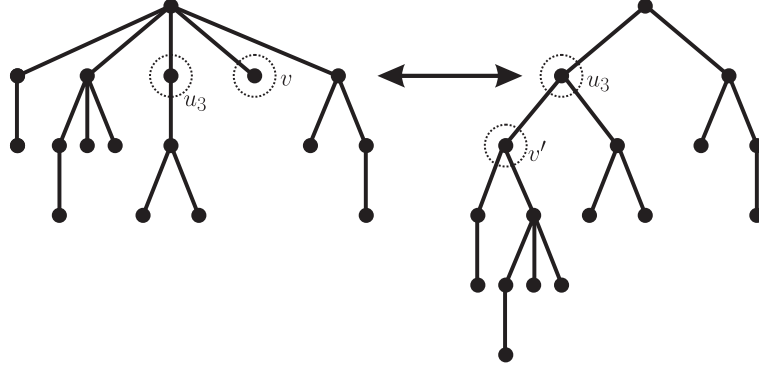


Fig. 3. Parity-reversing bijection (3)

4 Unrooted trees

Considering rooted unordered trees and unrooted trees has the obvious advantage that isomorphisms are taken into account, but the drawback that the enumeration is somewhat more difficult than in the case of simply generated trees. If $T(x)$ is the generating function for rooted unordered trees, it is well-known (see for instance [10]) that

$$T(x) = x \exp \left(\sum_{n=1}^{\infty} \frac{1}{n} T(x^n) \right).$$

We introduce the bivariate generating function $T(x, y) = \sum_T x^{|T|} y^{D(T)}$ again and get analogously

$$T(x, y) = x \exp \left(\sum_{n=1}^{\infty} \frac{1}{n} T(x^n y^n, y^n) \right).$$

To determine the asymptotics of $T(x) = T(x, 1)$, one notes that $R(x) = \sum_{n=2}^{\infty} \frac{1}{n} T(x^n)$ has a larger radius of convergence than $T(x)$, and so it is holomorphic within a larger circle around the origin. Then, since

$$T(x) = x \exp(T(x) + R(x)),$$

the remaining steps are similar to those for simply generated families of trees (cf. [10] again), and one obtains that the smallest singularity of $T(x)$ is $\rho = 0.3383218569$, and that the coefficients $t_n = [x^n]T(x)$ satisfy

$$t_n \sim a \rho^{-n} n^{-3/2},$$

where $a = 0.4399240126$. Indeed, even more is true: we can use literally the same proof as for simply generated trees to show that $T(x, -1)$ has a larger radius of convergence than $T(x, 1)$, and so there are equally many rooted unordered trees (asymptotically) with an even internal path length as with an

odd internal path length, and the same statement holds again for trees with an even number of vertices and the Wiener index.

Finally, let us consider unrooted trees. The main observation that one makes is due to Otter [12]: the number of representations of a tree as a rooted tree is 1 larger than the number of representations as a pair of distinct rooted trees whose roots are joined by an edge. In terms of generating functions, this yields the equation

$$\tilde{T}(x) = T(x) - \frac{1}{2} (T(x)^2 - T(x^2)),$$

where \tilde{T} denotes the generating function for (unrooted) trees. Now, in order to study the parity of the Wiener index, we have to write the Wiener index of a tree T that results from joining two rooted trees T_1 and T_2 at their roots in terms of T_1 and T_2 : it is easy to see that one has

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

The first summand takes account of all pairs in T_1 , the second summand of all pairs in T_2 , and the remaining three summands of the mixed pairs. Now, we are mainly interested in the case that $|T|$ is even—then, there are two possibilities:

- If $|T_1|$ and $|T_2|$ are even, then we have $W(T_i) \equiv D(T_i) \pmod{2}$, and thus

$$W(T) \equiv D(T_1) + D(T_2) \pmod{2}.$$

- If $|T_1|$ and $|T_2|$ are odd, then we have $W(T_i) \equiv 0 \pmod{2}$, and thus

$$W(T) \equiv D(T_1) + D(T_2) + 1 \pmod{2}.$$

Thus, in order to get the generating function for trees with an even number of vertices and even Wiener index, we have to consider the following two cases for T_1 and T_2 :

- $|T_1|$ and $|T_2|$ are even, and $D(T_1) \equiv D(T_2) \pmod{2}$.
- $|T_1|$ and $|T_2|$ are odd, and $D(T_1) \not\equiv D(T_2) \pmod{2}$.

Note that we can express the generating functions for the four possible classes of rooted trees (according to the parities of the number of vertices and the internal path length) in terms of $T(x, y)$:

- Rooted trees with an even number of vertices and even internal path length:

$$T_{00}(x) = \frac{1}{4} (T(x, 1) + T(x, -1) + T(-x, 1) + T(-x, -1)).$$

- Rooted trees with an even number of vertices and odd internal path length:

$$T_{01}(x) = \frac{1}{4} (T(x, 1) - T(x, -1) + T(-x, 1) - T(-x, -1)).$$

- Rooted trees with an odd number of vertices and even internal path length:

$$T_{10}(x) = \frac{1}{4} (T(x, 1) + T(x, -1) - T(-x, 1) - T(-x, -1)).$$

- Rooted trees with an odd number of vertices and odd internal path length:

$$T_{11}(x) = \frac{1}{4} (T(x, 1) - T(x, -1) - T(-x, 1) + T(-x, -1)).$$

Putting all these considerations together, we get the generating function for trees with an even number of vertices and even Wiener index:

$$\tilde{T}_{00}(x) = T_{00}(x) - \frac{1}{2} (T_{00}(x)^2 - T_{00}(x^2)) - \frac{1}{2} (T_{01}(x)^2 - T_{01}(x^2)) - T_{10}(x)T_{11}(x).$$

Upon simplification, this is equal to

$$\tilde{T}_{00}(x) = \frac{1}{4} (\tilde{T}(x) + \tilde{T}(-x)) + Q(x),$$

where

$$Q(x) = \frac{1}{4} (T(x, -1) + T(-x, -1) - T(x, -1)T(-x, -1) + T(-x^2, 1))$$

has a larger radius of convergence than $\tilde{T}(x)$. It follows that

Theorem 4 *The Wiener index is asymptotically equidistributed modulo 2 among trees with an even number of vertices.*

In the following table, we give the number of trees with n vertices and even resp. odd Wiener index for some small values of n :

Number of vertices	2	4	6	8	10	12	14	16	18
Even Wiener index	0	1	2	12	49	278	1559	9674	61814
Odd Wiener index	1	1	4	11	57	273	1600	9646	62053
Total	1	2	6	23	106	551	3159	19320	123867

References

- [1] E. A. Bender, Asymptotic methods in enumeration, SIAM Rev. 16 (1974) 485–515.

- [2] E. R. Canfield, Remarks on an asymptotic method in combinatorics, *J. Combin. Theory Ser. A* 37 (3) (1984) 348–352.
- [3] W. Y. C. Chen, L. W. Shapiro, L. L. M. Yang, Parity reversing involutions on plane trees and 2-Motzkin paths, *European J. Combin.* 27 (2) (2006) 283–289.
- [4] A. A. Dobrynin, R. Entinger, I. Gutman, Wiener index of trees: theory and applications, *Acta Appl. Math.* 66 (3) (2001) 211–249.
- [5] R. C. Entinger, A. Meir, J. W. Moon, L. A. Székely, The Wiener index of trees from certain families, *Australas. J. Combin.* 10 (1994) 211–224.
- [6] E. Estrada, O. Ivanciuc, I. Gutman, A. Gutierrez, L. Rodriguez, Extended wiener indices. a new set of descriptors for quantitative structure-property studies, *New J. Chem.* 22 (1998) 819–822.
- [7] S.-P. Eu, S.-C. Liu, Y.-N. Yeh, Odd or even on plane trees, *Discrete Math.* 281 (1-3) (2004) 189–196.
- [8] P. Flajolet, A. M. Odlyzko, Singularity analysis of generating functions, *SIAM J. Discrete Math.* 3 (1990) 216–240.
- [9] P. Flajolet, R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, to appear.
- [10] F. Harary, E. M. Palmer, *Graphical enumeration*, Academic Press, New York, 1973.
- [11] A. Meir, J. W. Moon, On the altitude of nodes in random trees, *Canad. J. Math.* 30 (5) (1978) 997–1015.
- [12] R. Otter, The number of trees, *Ann. of Math.* (2) 49 (1948) 583–599.
- [13] J. Riordan, N. J. A. Sloane, The enumeration of rooted trees by total height, *J. Austral. Math. Soc.* 10 (1969) 278–282.
- [14] H. Wiener, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.* 69 (1947) 17–20.