communications in mathematical
match
and in computer chemistry

no. 35, March 1997

ISSN 0340-6253

MATCDY (35) 91-103 (1997)

RELATIONS BETWEEN WIENER AND SZEGED INDICES OF MONOCYCLIC MOLECULES

Ivan GUTMAN[a,b], Ljiljana POPOVIĆ[b], Padmakar V. KHADIKAR[c], Sneha KARMARKAR[d], Shoba JOSHI[e] and Manorama MANDLOI[e]

[a] *Institute of Physical Chemistry, Atilla József University, P. O. Box 105, H-6701 Szeged, Hungary*

[b] *Faculty of Science, University of Kragujevac, P. O. Box 60, YU-34000 Kragujevac, Yugoslavia*

[c] *Faculty of Sciences, D.A.V.V., 3 Khatipura, Indore 452007, India*

[d] *Chemistry Department, Islamia Karimia College, Chawani Chowk, Indore 452001, India*

[e] *Department of Chemistry, Holkar College of Excellence, A.B. Road, Indore 452001, India*

(received: January 1996)

Abstract

The recently introduced Szeged index (Sz) is defined in such a manner that for acyclic molecules, it coincides with the Wiener index (W). The indices Sz and W of cyclic molecules differ and, in the general case, Sz exceeds W. Otherwise, the relation between Sz and W of cyclic molecules is not known. In order to get some information about this relation we examined monocyclic molecules. In this case certain nontrivial mathematical connections between Sz and W are shown to hold. Besides, a linear correlation between the two indices is established.

Introduction

One of the present authors introduced recently [1] a novel topological index which eventually was named [2, 3] *the Szeged index* and denoted by Sz. For acyclic molecules the Szeged index is, by definition, equal to the Wiener index (W). For cyclic molecules Sz and W differ and not much is known on their relation. Exceptionally, the cyclic

graphs for which $Sz = W$ were fully characterized [4, 5], and it was shown [6] that for all connected graphs, $Sz \geq W$. In addition to this, several analogies were established in the behavior of Sz and W of benzenoid systems [2, 7, 8] and of some other related polycyclic bipartite graphs [9].

In order to get some information on the relation between Sz and W in cyclic molecules, we decided to start with the examination of the simplest such case, namely with the monocyclic systems. The results obtained along these lines are reported in this paper.

In order to be able to present our findings, we first need to specify our notation and terminology, and, in particular, to define Sz and W.

Let G be the usual, hydrogen-atom depleted, graph representation of the molecule under consideration [10, 11]. Hence, G is a connected graph without directed and multiple edges and without loops. By $\mathbf{V}(G)$ and $\mathbf{E}(G)$ we denote the vertex and edge sets, respectively, of G. If e is an edge of G, connecting the vertices u and v, then we write $e = uv$. The number of vertices of G is denoted by $|G|$.

The distance between the vertices of G is defined as usual [12]: the distance $d(v, w|G)$, between two vertices v and w of G is equal to the length of a shortest path, connecting these vertices.

Definition of the Wiener index

The *Wiener index* of a graph G is just the sum of distances between all pairs of vertices of G:

$$W(G) = \frac{1}{2} \sum_{v \in \mathbf{V}(G)} \sum_{w \in \mathbf{V}(G)} d(v, w|G) = \frac{1}{2} \sum_{v \in \mathbf{V}(G)} d(v|G) \qquad (1)$$

where $d(v|G)$ is called the *distance number* of the vertex v, and is defined as [10]

$$d(v|G) = \sum_{w \in \mathbf{V}(G)} d(v, w|G)$$

Definition of the Szeged index

Let $e = uv \in \mathbf{E}(G)$. Then we define two subsets of the vertex set of G as follows:

$$\mathbf{N}_1(e|G) = \{\ x \in \mathbf{V}(G) \mid d(x, u|G) < d(x, v|G)\ \}$$

$$\mathbf{N}_2(e|G) = \{\ x \in \mathbf{V}(G) \mid d(x, u|G) > d(x, v|G)\ \}$$

The numbers of elements of $\mathbf{N}_1(e|G)$ and $\mathbf{N}_2(e|G)$ are denoted by $n_1(e|G)$ and $n_2(e|G)$, respectively. Thus $n_1(e|G)$ counts the vertices of G, lying closer to the vertex u than to the vertex v. The meaning of $n_2(e|G)$ is analogous. The vertices equidistant from both ends of the edge uv belong neither to $\mathbf{N}_1(e|G)$ nor to $\mathbf{N}_2(e|G)$.

The *Szeged index* of the graph G is defined as

$$Sz(G) = \sum_{e \in \mathbf{E}(G)} n_1(e|G)\, n_2(e|G) \tag{2}$$

When misunderstanding is avoided, we write $d(v,w)$, $d(v)$, $\mathbf{N}_i(e)$, $n_i(e)$, W and Sz instead of $d(v,w|G)$, $d(v|G)$, $\mathbf{N}_i(e|G)$, $n_i(e|G)$, $W(G)$ and $Sz(G)$, respectively.

Although Eqs. (1) and (2) look quite dissimilar, due to a classical result by Wiener [13], if G is acyclic, then W and Sz coincide. Wiener, namely, deduced the following identity [13]:

Lemma 1. If G is a tree, then

$$W = \sum_{e \in \mathbf{E}(G)} n_1(e)\, n_2(e) \tag{3}$$

Indeed, formula (3) served as a motivation for considering the graph invariant Sz [1], since Sz is nothing else than the right-hand side of Eq. (3) extended to all connected graphs. The (quite simple) proof of Lemma 1 was reproduced in the literature many times, for example in [1, 10]). In a fully analogous manner [14] one arrives at

Lemma 2. Let G be a tree and v its vertex. For each $e = uv \in \mathbf{E}(G)$, we label the sets $\mathbf{N}_1(e)$ and $\mathbf{N}_2(e)$ so that $v \in \mathbf{N}_1(e)$. Then

$$d(v) = \sum_{e \in \mathbf{E}(G)} n_2(e)$$

In view of Lemma 1 it is purposeful to examine the invariant Sz only in the case of graphs that contain circuits. As already mentioned, in this work we consider the simplest case of this kind – monocyclic graphs.

Correlations between Wiener and Szeged indices of monocyclic molecules

In order to examine the correlation between the Wiener and the Szeged indices of monocyclic molecules we designed sets of corresponding molecular graphs, each containing 100 elements.

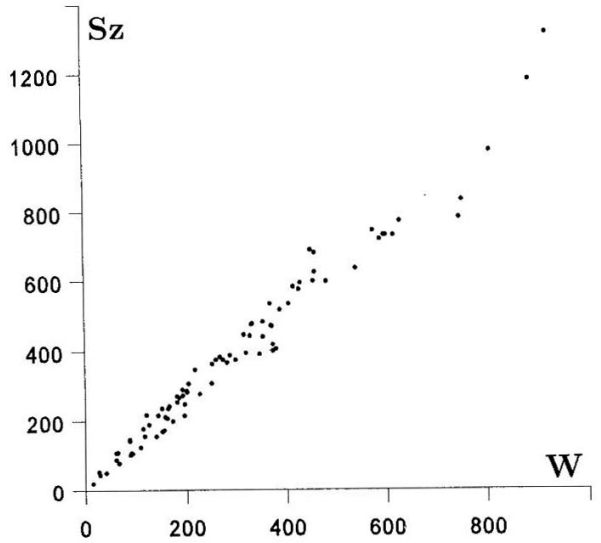


Fig. 1. Szeged index vs. Wiener index for a random sample of monocyclic molecular graphs; for details see text

These sets were chosen so as to contain 5, 20, 50, 20 and 5 graphs corresponding to monocyclic molecules with ring sizes 4, 5, 6, 7 and 8, respectively. Only mono–, di– and three–substituted systems were chosen, and their ratio was set to 1 : 3 : 1. The substituents (all possible alkyl groups up to five carbon atoms) were chosen uniformly by random. The positions of the substituents were also chosen by random, but taking

care that at most two substituents may be attached to each atom of the ring.

Two sets of molecular graphs were constructed in the above described manner. In the following we refer to them as to samples 1 and 2. It is believed that by this selection procedure the samples are more–or–less representative for monocyclic molecules usually encountered in chemistry.

In Fig. 1 the Szeged indices from sample 1 are plotted versus the Wiener indices. The analogous plot for sample 2 looks exactly the same and is therefore not shown.

The Sz– and W–values were calculated [15] and analysed by a standard statistical procedure. The correlation is clearly linear (see Fig. 1) and may be presented in the form

$$Sz = a\ W + \ b \tag{4}$$

The following results were obtained:

sample 1:

$a = 1.23 \pm\ 0.02\ \ ,\ \ b = 21.5 \pm\ 7.3$

standard error: 42.1 , correlation coefficient: 0.986

sample 2:

$a = 1.26 \pm\ 0.02\ \ ,\ \ b = 10.3 \pm\ 7.0$

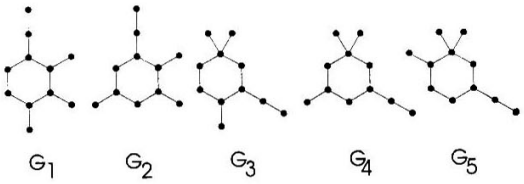
standard error: 38.4 , correlation coefficient: 0.988

It is seen that the constant term b in Eq. (4) to a great extend depends on the sample considered, and its calculated value is thus not reliable. On the other hand, the slope a of the regression line is practically sample independent. In our case this slope is very close to $1.25 = 5/4$, which means that Sz exceeds W by not less than 25%.

Anyway, the correlation between Sz and W is not very good, but these two topological indices certainly cannot be considered as unrelated. In the subsequent section we establish mathematical expressions for Sz and W of monocyclic systems, the form of which will provide a rationalization for the observed correlation.

At this point we wish to mention that neither W nor Sz have a high isomer–distinguishing power. (This property of W is, of course, long known [16, 17].) There exist numerous families of molecular graphs, possessing either equal W, or equal

Sz, or both equal W and Sz. For example, the molecular graphs $G_1 - G_5$ all have $W = 146$ and $Sz = 234$.



The structure–dependence of the Wiener and Szeged indices of monocyclic molecules

In this section we formulate, in the most general case, the dependence of W and Sz on the structure of monocyclic molecules. The general form of the molecular graph of a monocyclic molecule is depicted in Fig. 2. We denote this graph by $C_n[T_1, T_2, \ldots, T_n]$.

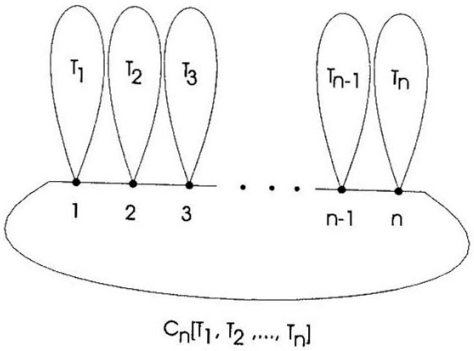


Fig. 2. The structure of the graph $C_n[T_1, T_2, \ldots, T_n]$ representing monocyclic

molecules; for details see text

The graph $C_n[T_1, T_2, \ldots, T_n]$ is constructed as follows. Let C_n be the circuit on n vertices. Label its vertices consecutively by $v_1, v_2, \ldots, v_n$. Let T_i be a tree and r_i its distinguished vertex, $i = 1, 2, \ldots, n$. As usual, T_i is referred to as a rooted tree, and r_i is its root.

Then $C_n[T_1, T_2, \ldots, T_n]$ is obtained by identifying the root r_i of T_i with the vertex v_i of C_n , for all $i = 1, 2, \ldots, n$, see Fig. 2. Therefore,

$$| C_n[T_1, T_2, \ldots, T_n] | = \sum_{i=1}^{n} |T_i| \tag{5}$$

Evidently, every connected unicyclic graph can be viewed as having the form $C_n[T_1, T_2, \ldots, T_n]$; the size of its unique circuit is n.

The rooted tree T_i will be said to form a branch of $C_n[T_1, T_2, \ldots, T_n]$, attached at the i-th vertex of the circuit. This branch possesses $|T_i| - 1$ vertices. Consequently, if T_i consists only of the root r_i (namely, if $|T_i| = 1$), then $C_n[T_1, T_2, \ldots, T_n]$ has no branch at position i.

Define an auxiliary function $\delta(n)$, such that $\delta(n) = 0$ if n if an even number, and $\delta(n) = 1$ if n is odd.

THEOREM. For the sake of brevity denote $C_n[T_1, T_2, \ldots, T_n]$ by M. Then the Wiener and Szeged indices of M conform to the following formulas:

$$W(G) = \sum_{i=1}^{n} W(T_i) + \sum_{i=1}^{n} (|M| - |T_i|)\, d(r_i|T_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} |T_i|\, |T_j|\, d(v_i, v_j|C_n) \tag{6}$$

$$Sz(G) = \sum_{i=1}^{n} W(T_i) + \sum_{i=1}^{n} (|M| - |T_i|)\, d(r_i|T_i) + \sum_{i=1}^{n} \sum_{j=1}^{n} |T_i|\, |T_j|\, d(v_i, v_j|C_n) - \delta(n) \sum_{i<j} |T_i|\, |T_j| \tag{7}$$

Recall that $|M|$ is given by Eq. (5).

The results summarized in the Theorem were previously communicated [14], together with a detailed proof. In view of this, in this paper, in the last two sections, we indicate only the less routine steps of the proof of Eqs. (6) and (7). Before that, however, we point out a few noteworthy consequences of our Theorem.

Relations between Wiener and Szeged indices of monocyclic molecules

The inspection of Eqs. (6) and (7) reveals a striking similarity in the structure-dependence of the Wiener and Szeged indices of monocyclic molecules. Indeed, these two (otherwise quite complicated) expressions differ only in a multiplier $\frac{1}{2}$ in the third term, and in the term proportional to $\delta(n)$, that is present in Eq. (7), but not in Eq. (6). In the case of alternant systems (for which n is even and $\delta(n) = 0$), the similarity between the right-hand sides of (6) and (7) is even more pronounced.

In view of the above, by combining Eqs. (6) and (7) we arrive at

$$Sz(M) - W(M) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} |T_i| \, |T_j| \, d(v_i, v_j | C_n) - \delta(n) \sum_{i<j} |T_i| \, |T_j| \qquad (8)$$

and

$$2\,W(M) - Sz(M) = \sum_{i=1}^{n} W(T_i) + \sum_{i=1}^{n} (|M| - |T_i|) \, d(r_i | T_i) + \delta(n) \sum_{i<j} |T_i| \, |T_j| \qquad (9)$$

Because Sz is never smaller than W [6], the right-hand side of Eq. (8) cannot be negative. According to the previous characterization [5] of graphs for which the Szeged and Wiener indices coincide, the right-hand side of (8) is equal to zero whenever $n = 3$, and only then. In other words, $Sz(M) - W(M) = 0$ if, and only if, M is the molecular graph of a cyclopropane derivative.

It is easy to see that the right-hand side of Eq. (9) is also non-negative. In fact, the right-hand side of (9) is equal to zero only if n is even and $|T_1| = |T_2| = \cdots = |T_n| = 1$, i.e., if $M = C_n$, i.e., if M is the molecular graph of cyclobutane or cyclohexane or cyclooctane etc.

These observations imply

COROLLARY 1. If M is a unicyclic graph (not necessarily a molecular graph), then

$$W(M) \leq Sz(M) \leq 2\,W(M)$$

Equality between $Sz(M)$ and $W(M)$ occurs if, and only if, the size of the circuit of M is 3. Equality between $Sz(M)$ and $2\,W(M)$ occurs if, and only if, $M = C_n$ and $n = 4, 6, 8, \ldots$.

The following results are straightforward special cases of Eq. (8). The considerations leading to Corollaries 2 and 4, although elementary, require lengthy calculations and are therefore skipped.

COROLLARY 2. Let M be the graph of the form $C_n[T_1, T_2, \ldots, T_n]$, such that $|T_1| > 1$ and $|T_2| = |T_3| = \cdots = |T_n| = 1$. In other words, we require that the molecule has only one side group attached. Then for n being even-valued

$$Sz(M) - W(M) = \frac{1}{8} n^2[n + 2\ (|T_1| - 1)]$$

whereas for odd n

$$Sz(M) - W(M) = \frac{1}{8} (n - 1)\ (n - 3)\ [n + 2\ (|T_1| - 1)]$$

COROLLARY 3. Suppose that for isomeric monocyclic molecules, specified in Corollary 2, both the ring size (n) and the number of atoms in the unique side group attached to it ($|T_1| - 1$) are fixed. Then an exact linear relation exists between the Wiener and the Szeged indices. The slope of the Sz/W-line is unity.

COROLLARY 4. Let M be the graph of the form $C_n[T_1, T_2, \ldots, T_n]$, such that $|T_1| > 1$, $|T_2| > 1$, $|T_3| = \cdots = |T_n| = 1$. In other words, we require that there are two side groups attached to the ring of M, and that these are attached to adjacent atoms. Let n be odd. Then

$$Sz(G) - W(G) = \frac{1}{8} (n - 1)\ (n - 3)\ [n + 2\ (|T_1| + |T_2| - 2)]$$

COROLLARY 5. Suppose that for the monocyclic molecules specified in Corollary 4, both the ring size (n, odd) and the total number of atoms in the two side groups ($|T_1| + |T_2| - 2$) are fixed. Then an exact linear relation exists between the Wiener and the Szeged indices. The slope of the Sz/W-line is unity.

COROLLARY 6. For molecules specified in Corollary 4, for which the ring size is even, the difference $Sz - W$ depends not only on n and $|T_1| + |T_2| - 2$, but also on the product $|T_1|\,|T_2|$. Therefore the linear relation between Sz and W is restricted to systems in which each side group has a fixed number of carbon-atoms.

Corollary 6 can be generalized by observing that the right-hand side of Eq. (8) depends only on the number of vertices of T_1, T_2 ,..., T_n and on the positions to which these are attached to the circuit, but does not depend on any structural detail of the respective rooted trees:

COROLLARY 7. For graphs of the form $C_n[T_1, T_2, \ldots, T_n]$ in which the size of the circuit and the number of vertices in each branch are fixed, and in which the distribution of the branches is fixed, the Wiener and the Szeged indices are linearly related. The slope of the Sz/W-line is unity.

In contrast to Eq. (8), the right-hand side of Eq. (9) does depend on the structure of the rooted trees T_i, $i = 1, 2, \ldots, n$ (via the terms $W(T_i)$ and $d(r_i|T_i)$), but is independent of the order in which these trees are attached to the circuit. Therefore Eq. (9) implies the following conclusion.

COROLLARY 8. For graphs of the form $C_n[T_1, T_2, \ldots, T_n]$ in which the size of the circuit and the structure of each branch (including its size) are fixed, but in which the distribution of the branches varies, the Wiener and the Szeged indices are linearly related. The slope of the Sz/W-line is two.

The above listed corollaries of the Theorem can be viewed as a kind of rationalization of the observed linear correlation between Sz and W. They show that in some families of monocyclic molecules there exist exact linear relations between Sz and W. The respective slopes are sometimes 1, sometimes 2. These relations are more-or-less violated by other monocyclic systems, resulting in a linear correlation in the case of the general set of monocyclic molecules. The respective slope is between 1 and 2 and is found to be nearly 1.25.

Proof of formula (6)

The vertex pairs of $M = C_n[T_1, T_2, \ldots, T_n]$ can be divided into following groups:

(a) both vertices belong to the tree T_i , $i = 1, 2, \ldots, n$;

(b) one vertex belongs to T_i, the other to T_j, $1 \leq i \leq j \leq n$.

The contributions of vertex pairs (a) to $W(M)$ are equal to

$$\frac{1}{2} \sum_{v \in \mathbf{V}(T_i)} \sum_{w \in \mathbf{V}(T_i)} d(v,w|T_i)$$

whereas the contributions of the vertex pairs (b) satisfy

$$\sum_{v \in \mathbf{V}(T_i)} \sum_{w \in \mathbf{V}(T_j)} [d(v,r_i|T_i) + (v_i,v_j|C_n) + d(w,r_j|T_j)]$$
$$= |T_j|\, d(r_i|T_i) + |T_i|\, |T_j|\, d(v_i,v_j|C_n) + |T_i|\, d(r_j|T_j)$$

Consequently,

$$W(M) = \sum_{i=1}^{n} W(T_i) + \sum_{i<j} [\, |T_j|\, d(r_i|T_i) + |T_i|\, |T_j|\, d(v_i,v_j|C_n) + |T_i|\, d(r_j|T_j)\,] \quad (10)$$

In addition to this, by using Eq. (5) we get

$$\sum_{i<j} (\, |T_j|\, d(r_i|T_i) + |T_i|\, d(r_j|T_j)\,) = [\sum_{j=1}^{n} |T_j|\,] [\sum_{i=1}^{n} d(r_i|T_i)\,] - \sum_{i=1}^{n} |T_i|\, d(r_i|T_i)$$
$$= \sum_{i=1}^{n} (|M| - |T_i|\,)\, d(r_i|T_i) \quad (11)$$

Substituting (11) back into (10) one arrives at formula (6).

Proof of formula (7)

This time it is purposeful to divide the edges of $M = C_n[T_1, T_2, \ldots, T_n]$ into the following groups:

(a) the edges belonging to the tree T_i, $i = 1, 2, \ldots n$;

(b) the edges belonging to the circuit C_n .

The contributions to $Sz(M)$ pertaining to the edges of type (a) are

$$\sum_{e \in \mathbf{E}(T_i)} n_1(e|M)\, n_2(e|M) = \sum_{e \in \mathbf{E}(T_i)} [\, n_1(e|T_i) + |M| - |T_i|\,]\, n_2(e|T_i)$$
$$= \sum_{e \in \mathbf{E}(T_i)} n_1(e|T_i)\, n_2(e|T_i) + (|M| - |T_i|) \sum_{e \in \mathbf{E}(T_i)} n_2(e|T_i)$$

and by taking into account Lemmas 1 and 2,

$$\sum_{e \in \mathbf{E}(T_i)} n_1(e|M)\, n_2(e|M) = W(T_i) + (\, |M| - |T_i|\,)\, d(r_i|T_i) \quad (12)$$

In order to calculate the sum of the products $n_1(e|M)\ n_2(e|M)$ over the edges $e \in \mathbf{E}(C_n)$ we need an auxiliary result, which we state without proof [14].

Let, as before, C_n be the n-vertex circuit and $v_1, v_2, \ldots, v_n$ be its vertices. The labeling of the vertices of C_n is specified above (see Fig. 2).

Lemma 3. Let $v_i, v_j \in \mathbf{V}(C_n)$. Let $\gamma(i,j)$ be the number of edges e of C_n, such that $v_i \in \mathbf{N}_1(e)$ and $v_j \in \mathbf{N}_2(e)$. Then $\gamma(i,j) = 2\ d(v_i, v_j) - \delta(n)$. In addition to this,

$$\sum_{e \in \mathbf{E}(C_n)} n_1(e|M)\ n_2(e|M) = \sum_{i<j} \gamma(i,j)\ |T_i|\ |T_j|$$

By means of Lemma 3 we obtain

$$\sum_{e \in \mathbf{E}(C_n)} n_1(e|M)\ n_2(e|M) = \sum_{i=1}^{n} \sum_{j=1}^{n} |T_i|\ |T_j|\ d(v_i, v_j|C_n) - \delta(n) \sum_{i<j} |T_i|\ |T_j| \quad (13)$$

From (12) and (13),

$$Sz(M) = \sum_{i=1}^{n} [\ W(T_i) + (|M| - |T_i|)\ d(r_i|T_i)\ ] + \sum_{i=1}^{n} \sum_{j=1}^{n} |T_i|\ |T_j|\ d(v_i, v_j|C_n) - \delta(n) \sum_{i<j} |T_i|\ |T_j|$$

and formula (7) follows straightforwardly.

Acknowledgement.

One author (I. G.) thanks the Mathematical Institute in Belgrade for financial support of this research.

References

[1] I. Gutman, *Graph Theory Notes New York* **27**, 9 (1994).

[2] P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. Dobrynin, I. Gutman and G. Dömötör, *J. Chem. Inf. Computer Sci.* **35**, 547 (1995).

[3] The Szeged index is named after the Hungarian city Szeged, where it has been conceived. In Hungarian "Sz" is pronounced as "S" in English. For instance, in Hungarian the words "sztár", "szeptember" and "szex" have both the same meanings and pronunciations as the English words "star", "September" and "sex".

[4] A. Dobrynin and I. Gutman, *Publ. Inst. Math. (Beograd)* **56**, 18 (1994).

[5] A. Dobrynin and I. Gutman, *Graph Theory Notes New York* **28**, 21 (1995).

[6] S. Klavžar, A. Rajapakse and I. Gutman, *Appl. Math. Letters*, in press.

[7] I. Gutman, P. V. Khadikar, P. V. Rajput and S. Karmarkar, *J. Serb. Chem. Soc.* **60**, 759 (1995).

[8] A.A.Dobrynin and I.Gutman, *Croat. Chem. Acta* **69**, 000 (1996), in press.

[9] A. Dobrynin, I. Gutman and G. Dömötör, *Appl. Math. Letters* **8**(5), 57 (1995).

[10] I. Gutman and O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer–Verlag, Berlin 1986.

[11] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton 1992.

[12] F. Buckley and F. Harary, *Distance in Graphs*, Addison-Wesley, Redwood 1990.

[13] H. Wiener, *J. Am. Chem. Soc.*, **69**, 17 (1947).

[14] I. Gutman, *Bull. Acad. Serbe. Sci. Arts. (Cl. Math. Natur.)*, in press.

[15] A FORTRAN 77 program for the calculation of Sz and W was written by L. P. and is available upon request.

[16] M. Razinger, J. R. Chretien and J. E. Dubois, *J. Chem. Inf. Comput. Sci.* **25**, 24 (1985).

[17] I. Gutman and L. Šoltés, *Z. Naturforsch.* **46a**, 865 (1991).