J. Chem. Inf. Comput. Sci. **1995,** *35,* 547–550

The Szeged Index and an Analogy with the Wiener Index

Padmakar V. Khadikar, Narayan V. Deshpande, and Prabhakar P. Kale

Devy Ahilya University, Indore (M.P.) 452007, India

Andrey Dobrynin

Institute of Mathematics, Russian Academy of Sciences, Siberian Branch, Novosibirsk 630090, Russia

Ivan Gutman*,† and Gyula Dömötör

Institute of Physical Chemistry, Attila Jozsef University, P.O. Box 105, H-6701 Szeged, Hungary

Received December 23, 1994®

A novel molecular-graph-based structural descriptor, referred to as the Szeged index, Sz, is put forward. The Szeged index is based on distances in the molecular graph but is not of the same type as the Wiener index, W. For acyclic systems Sz and W coincide. It is known for some time that if G and G′ are two catacondensed benzenoid systems with an equal number of hexagons, then $W(G) \equiv W(G')$ (mod 8). We now show that $\mathrm{Sz}(G) \equiv \mathrm{Sz}(G')$ (mod 8).

INTRODUCTION

One of the oldest and most thoroughly examined molecular-graph-based structural descriptors of organic molecules is the Wiener index or Wiener number, W.[1–7] This quantity is equal to the sum of distances between all pairs of vertices of the respective molecular graph.

If $\mathrm{d}(u,v|G)$ is the distance[8] of the vertices u and v of the graph G (i.e., the number of edges in the shortest path that connects u and v), and $\mathbf{V}(G)$ is the vertex set of G, then

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

For acyclic molecular graphs Wiener[1] discovered a remarkably simple method for the calculation of W.[1,4,6,9] Let e be an edge of an acyclic molecular graph G (= a tree). Let $n_1(e|G)$ and $n_2(e|G)$ be the numbers of vertices of G lying on two sides of the edge e. Then,

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

Here and later $\mathbf{E}(G)$ denotes the edge set of the graph G.

The quantities $n_1(e|G)$ and $n_2(e|G)$ can be defined in a more formal manner as follows: Let e be an edge of a graph G (which may contain cycles but may also be acyclic) connecting the vertices u and v. Define two sets $N_1(e|G)$ and $N_2(e|G)$ as

$$N_1(e|G) = \{x|x \in \mathbf{V}(G),\ \mathrm{d}(x,u|G) < \mathrm{d}(x,v|G)\}$$

$$N_2(e|G) = \{x|x \in \mathbf{V}(G),\ \mathrm{d}(x,v|G) < \mathrm{d}(x,u|G)\}$$

Then $n_i(e) = n_i(e|G)$ is just the cardinality (= number of elements) of the set $N_i(e|G)$, $i = 1, 2$. In other words: $n_1(e|G)$ is the number of vertices closer to u than to v, and $n_2(e|G)$ is the number of vertices closer to v than to u; vertices equidistant to u and v are not counted.

The above definition has the advantage of being applicable to an arbitrary edge of an arbitrary graph. (Notice that in our definition it is not required that the graph G is connected; in chemical applications, however, one encounters only connected graphs.) In view of this, it readily comes to mind to examine the right-hand side of (2) for cyclic graphs. This, on the other hand, is tantamount to the introduction of a novel topological index $\mathrm{Sz} = \mathrm{Sz}(G)$,[10] defined as

$$\mathrm{Sz}(G) = \sum_{e\in\mathbf{E}(G)} n_1(e|G)n_2(e|G) \tag{3}$$

The right-hand side of (3), although formally identical to the right-hand side of (2), differs in the interpretation of n_1 and n_2. This generalization was conceived at the Attila Jozsef University in Szeged, and we propose it to be called the *Szeged index* and to be denoted by *Sz*.[10] The basic properties of the Szeged index were recently established,[11,12] and Sz was found to be endowed with interesting and mathematically appealing features. In this paper we point out a further property of Sz, revealing the existence of a deep-lying analogy between the Szeged and the Wiener index.

We mention in passing that recently several other generalizations of the Wiener index were put forward.[13–18]

STATEMENT OF THE MAIN RESULT

Some time ago the following regularity was demonstrated[19] to hold for the Wiener index of catacondensed benzenoid hydrocarbons.[20] If G_h and G'_h are catacondensed benzenoid systems, both possessing h hexagons, then the difference $W(G_h) - W(G'_h)$ is divisible by 8. In a more formal mathematical notation this property of the Wiener index is written as $W(G_h) \equiv W(G'_h)$ (mod 8).

The main result of this paper is the proof that a fully analogous divisibility rule holds also for the Szeged index:

If G_h and G'_h are catacondensed benzenoid systems, both possessing h hexagons, $h \geq 1$, then irrespective of any further detail in the structure of G_h and G'_h

† On leave from the Faculty of Science, University of Kragujevac, Yugoslavia.

® Abstract published in *Advance ACS Abstracts,* March 15, 1995.

0095-2338/95/1635-0547$09.00/0

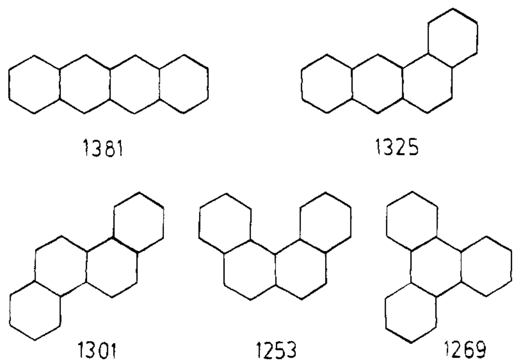

Figure 1. The five distinct catacondensed benzenoid systems with four hexagons and the respective Szeged indices; observe that 1381 = 1253 + 16 × 8, 1325 = 1253 + 9 × 8, 1301 = 1253 + 6 × 8, 1269 = 1253 + 2 × 8. The respective Wiener indices are 569, 553, 545, 529, and 513; observe that 569 = 513 + 7 × 8, 553 = 513 + 5 × 8, 545 = 513 + 4 × 8, 529 = 513 + 2 × 8.

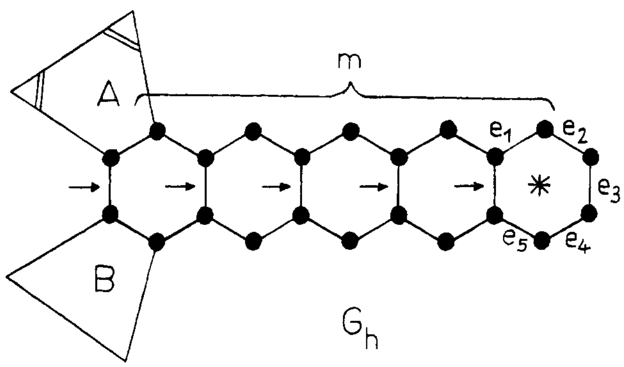

Figure 2. The structure of a general catacondensed benzenoid system; for details see text.

and irrespective of the actual numerical value of $\mathrm{Sz}(G_h)$ and $\mathrm{Sz}(G'_h)$, the difference $\mathrm{Sz}(G_h) - \mathrm{Sz}(G'_h)$ is divisible by eight, i.e.

$$\mathrm{Sz}(G_h) \equiv \mathrm{Sz}(G'_h) \pmod 8 \tag{4}$$

An example illustrating the above modulo 8 rule is given in Figure 1.

PROOF OF THE MAIN RESULT

We will often use the fact that if G is a connected bipartite graph, then for any of its edges e, $n_1(e|G) + n_2(e|G)$ = number of vertices of G.[11] The molecular graphs of benzenoid hydrocarbons are bipartite and connected.

There is only one catacondensed benzenoid hydrocarbon with one six-membered ring (benzene) and with two six-membered rings (naphthalene). Therefore (4) holds in a trivial manner for $h = 1$ and $h = 2$. By direct calculation we easily check that (4) is obeyed also for $h = 3$: here Sz(anthracene) = 656, Sz(phenanthrene) = 632, Sz(anthracene) − Sz(phenanthrene) = 24 = 3 × 8. The data presented in Figure 1 show that the modulo 8 rule holds also for $h = 4$.

In Figure 2 is depicted the structure of a general catacondensed benzenoid system G_h. Here and later the subscript h indicates the total number of hexagons.

The hexagon of G_h, marked by an asterisk, is terminal; every catacondensed system (provided $h \geq 2$) possesses at least two terminal hexagons. The starred terminal hexagon is assumed to belong to a linear segment of length m, $m \geq 2$.

The branches A and B may but need not exist. The fragments A and B are catacondensed systems themselves, possessing h_A and h_B hexagons, respectively, $h_A, h_B \geq 0$. The fragments A and B are present in G_h only if h_A and h_B, respectively, are nonzero. Evidently, $h_A + h_B + m = h$.

The linear segment of G_h possesses $4m + 2$ vertices. The branches A and B possess additional $4h_A$ and $4h_B$ vertices which do not belong to the linear segment. Thus the number of vertices of G_h is $4m + 2 + 4h_A + 4h_B = 4h + 2$.

By deleting from G_h the hexagon marked by an asterisk we obtain a catacondensed benzenoid system with $h - 1$ hexagons, which we denote by G_{h-1}. It is clear that every catacondensed benzenoid system G_h can be constructed from a pertinently chosen G_{h-1}.

Among the edges of G_h we distinguish five which do not belong to G_{h-1}; these are labeled by e_1, e_2, e_3, e_4, and e_5 (see Figure 2). We denote by **J** the set of those edges of G_h which are indicated by arrows in Figure 2. All other edges of G_h form the set **K**. Hence,

$$\mathbf{E}(G_h) = \mathbf{J} \cup \mathbf{K} \cup \{e_1, e_2, e_3, e_4, e_5\} \quad \text{and} \quad \mathbf{E}(G_{h-1}) = \mathbf{J} \cup \mathbf{K} \tag{5}$$

With the notation thus introduced we are ready to start to demonstrate the validity of the modulo 8 rule (4).

First consider the edge e_1 and denote its two vertices by u and v, so that u is incident to e_2, see Figure 2. Then $N_1(e_1|G_h)$ consists of only three vertices, namely of the vertex u and of the two vertices which are the endpoints of the edge e_3, see Figure 2. This means that $n_1(e_1|G_h) = 3$ and, consequently, $n_2(e_1|G_h) = (4h + 2) - 3 = 4h - 1$.

The same is true for the edges e_2, e_4, and e_5, so that we have

$$n_1(e_i|G_h) = 3 \quad \text{and} \quad n_2(e_i|G_h) = 4h - 1 \quad \text{for } i = 1, 2, 4, 5 \tag{6}$$

It is easy to see that for $e \in \{e_3\} \cup \mathbf{J}$

$$n_1(e|G_h) = 2m + 1 + 4h_A \quad \text{and} \quad n_2(e|G_h) = 2m + 1 + 4h_B \tag{7}$$

This, in particular, means that for $e \in \mathbf{J}$

$$n_1(e|G_h) = n_1(e|G_{h-1}) + 2 \quad \text{and} \quad n_2(e|G_h) = n_2(e|G_{h-1}) + 2 \tag{8}$$

In a similar manner we have for $e \in \mathbf{K}$

$$n_1(e|G_h) = n_1(e|G_{h-1}) \quad \text{and} \quad n_2(e|G_h) = n_2(e|G_{h-1}) + 4 \tag{9}$$

Using (5)−(9) we can deduce a recurrence relation for the Szeged index of G_h

$$\mathrm{Sz}(G_h) = \sum_{e\in \mathbf{E}(G_h)} n_1(e|G_h)n_2(e|G_h) = \sum_{e\in\mathbf{J}} n_1(e|G_h)n_2(e|G_h) + \sum_{e\in\mathbf{K}} n_1(e|G_h)n_2(e|G_h) + \sum_{i=1}^{5} n_1(e_i|G_h)n_2(e_i|G_h) = \sum_{e\in\mathbf{J}}[n_1(e|G_{h-1}) + 2][n_2(e|G_{h-1}) + 2] + \sum_{e\in\mathbf{K}} n_1(e|G_{h-1})[n_2(e|G_{h-1}) + 4] + 4 \times [3 \times (4h - 1)] + (2m + 1 + 4h_A)(2m + 1 + 4h_B)$$

Now, because $n_1(e|G_{h-1}) + n_2(e|G_{h-1})$ = number of vertices of $G_{h-1} = 4h - 2$, and m = number of elements of the set **J**

$$\sum_{e\in\mathbf{J}}[n_1(e|G_{h-1}) + 2][n_2(e|G_{h-1}) + 2] = \sum_{e\in\mathbf{J}} n_1(e|G_{h-1})n_2(e|G_{h-1}) + 2m(4h - 2) + 4m \quad (10)$$

In addition to this,

$$\sum_{e\in\mathbf{K}} n_1(e|G_{h-1}) = \sum_{e\in\mathbf{E}(G_{h-1})} n_1(e|G_{h-1}) - \sum_{e\in\mathbf{J}}(2m - 1 + 4h_A) = \sum_{e\in\mathbf{E}(G_{h-1})} n_1(e|G_{h-1}) - m(2m - 1 + 4h_A) \quad (11)$$

$$2m + 1 + 4h_A = 2h + 1 + 2(h_A - h_B) \quad (12)$$

and

$$2m + 1 + 4h_B = 2h + 1 - 2(h_A - h_B) \quad (13)$$

Equations 11–13 originate from (7) and from $h_A + h_B + m = h$, respectively. Bearing in mind (10)–(13) the above expression for $\mathrm{Sz}(G_h)$ is transformed into

$$\mathrm{Sz}(G_h) = \mathrm{Sz}(G_{h-1}) + 4 \sum_{e\in\mathbf{E}(G_{h-1})} n_1(e|G_{h-1}) + 8m(1 - h_A + h_B) + 4[h + m - (h_A - h_B)^2] + (4h^2 + 48h - 11) \quad (14)$$

Suppose that there is another catacondensed benzenoid system G'_h for which in full analogy to (14) we have

$$\mathrm{Sz}(G'_h) = \mathrm{Sz}(G'_{h-1}) + 4 \sum_{e\in\mathbf{E}(G'_{h-1})} n_1(e|G'_{h-1}) + 8m'(1 - h'_A + h'_B) + 4[h + m' - (h'_A - h'_B)^2] + (4h^2 + 48h - 11) \quad (15)$$

By subtracting (15) from (14) we get

$$[\mathrm{Sz}(G_h) - \mathrm{Sz}(G'_h)] = T_1 + 4T_2(h-1) + 8T_3 + 4T_4 \quad (16)$$

where

$$T_1 = \mathrm{Sz}(G_{h-1}) - \mathrm{Sz}(G'_{h-1})$$

$$T_2(h - 1) = \sum_{e\in\mathbf{E}(G_{h-1})} n_1(e|G_{h-1}) - \sum_{e\in\mathbf{E}(G'_{h-1})} n_1(e|G'_{h-1})$$

$$T_3 = m(1 - h_A + h_B) - m'(1 - h'_A + h'_B)$$

$$T_4 = [h + m - (h_A - h_B)^2] - [h + m' - (h'_A - h'_B)^2]$$

The expression (16) enables us to prove (4) by induction on the number h of hexagons.

We already know that the modulo 8 rule is valid for $h \le 4$. Assume that it is satisfied by all catacondensed systems with $h - 1$ hexagons. This means that we assume that T_1 is divisible by 8. In order to accomplish the proof we have to demonstrate that from this assumption it follows that the left-hand side of (16) is divisible by 8. This will be the case if the remaining three terms, occurring on the right-hand side of (16), are all divisible by 8.

The term $8T_3$ is obviously divisible by 8. It is elementary to verify that because of $h = m + h_A + h_B$, the expression $h + m - (h_A - h_B)^2$ is always even and, consequently, T_4 is even, i.e., $4T_4$ is divisible by 8.

It remains to show that $T_2(h - 1)$ is an even number. Because this property must be independent of h, we examine $T_2(h)$ and decompose it using (5), (6), (7), and (9):

$$T_2(h) = \sum_{e\in\mathbf{E}(G_h)} n_1(e|G_h) - \sum_{e\in\mathbf{E}(G'_h)} n_1(e|G'_h) = [\sum_{e\in\mathbf{J}}[n_1(e|G_{h-1}) + 2] - \sum_{e\in\mathbf{J}'}[n_1(e|G'_{h-1}) + 2]] + [\sum_{e\in\mathbf{K}} n_1(e|G_{h-1}) - \sum_{e\in\mathbf{K}'} n_1(e|G'_{h-1})] + [4 \times 3 + (2m + 1 + 4h_A)] - [4 \times 3 + (2m' + 1 + 4h'_A)]$$

$$= T_2(h - 1) + 4(m + h_A - m' - h'_A)$$

By induction on h we now immediately show that $T_2(h)$ is not only an even number, but that it is divisible by 4.

By this we demonstrated that the second, third, and fourth terms on the right-hand side of (16) are divisible by 8. If the first term is divisible too (which is just the inductive hypothesis), then the left-hand side of (16) is divisible by 8.

This completes the proof of the modulo 8 formula (4).

CONCLUDING REMARKS

Although the modulo 8 rules for the Szeged and for the Wiener indices are fully analogous, they could be deduced by means of two completely different proof techniques. The reason for this is to be sought in the different concepts on which the two indices are constructed. Whereas the Wiener index is defined so as to involve properties of pairs of vertices that are at arbitrarily large distances, see (1), the Szeged index deals with properties of pairs of vertices at unit distance, see (3). Whereas with the Wiener index we count edges (between pairs of vertices), with the Szeged index we count vertices (lying closer to one end of an edge than to its other end, and vice versa). Whereas the Wiener index is globally defined and its decomposition into local contributions is a difficult and not fully resolved task, the Szeged index is (formally) locally defined and its local contributions are immediately recognized.

In view of these conceptual differences it is somewhat surprising that there is any similarity in the behavior of W and Sz. The result outlined in this paper shows that, nevertheless, the analogy between the two indices is sometimes quite deep. In some other cases, however, the analogy between W and Sz is completely lacking.

Work on the elucidation of the relations between the Szeged and the Wiener index, which ultimately should enable the application of the Szeged index to structure–property and structure–activity studies, is in progress.

REFERENCES AND NOTES

(1) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.

(2) There exist a large number of reviews and books dealing with the Wiener number and its chemical applications; we mention only the most recent ones[3–7] in which references to earlier works can be found.

(3) Mihalić, Z.; Veljan, D.; Amić, D.; Nikolić, S.; Plavšić, D.; Trinajstić, N. The Distance Matrix in Chemistry. *J. Math. Chem.* **1992**, *11*, 223–258.

(4) Gutman, I.; Yeh, Y. N.; Lee, S. L.; Luo, Y. L. Some Recent Results in the Theory of the Wiener Number. *Indian J. Chem.* **1993**, *32A*, 651–661.

(5) Rouvray, D. H. In *Applications of Mathematical Concepts to Chemistry*; Trinajstić, N., Ed.; Horwood: Chichester, 1985.

(6) Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer-Verlag: Berlin, 1986.

(7) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed.; CRC Press: Boca Raton, FL, 1992.

(8) Buckley, F.; Harary, F. *Distance in Graphs*; Addison-Wesley: Reading, MA, 1990.

(9) Gutman, I. A new Method for the Calculation of the Wiener Number of Acyclic Molecules. *J. Mol. Struct. (Theochem)* **1993**, *285*, 137–142.

(10) In Hungarian "sz" is pronounced as "s" in English. For instance, the Hungarian spelling of "September", "software", "sex", and "story" is "szeptember", "szoftver", "szex", and "sztori", whereas the pronunciation (as well as the meaning) are essentially the same as in English.

(11) Gutman, I. A Formula for the Wiener Number of Trees and Its Extension to Graphs Containing Cycles. *Graph Theory Notes New York* **1994**, *27*, 9–15.

(12) Dobrynin, A.; Gutman, I. Solving a Problem Connected with Distances in Graphs. *Graph Theory Notes New York* **1995**, *28*, in press.

(13) Lukovits, I. Wiener Indices and Partition Coefficients of Unsaturated Hydrocarbons. *Quant. Struct. Act. Relat.* **1990**, *9*, 227–231.

(14) Lukovits, I. Correlation between Components of the Wiener Index and Partition Coefficients of Hydrocarbons. *Int. J. Quantum Chem. Quantum. Biol. Symp.* **1992**, *19*, 217–223.

(15) Randić, M. Novel Molecular Descriptor for Structure–Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478–483.

(16) Mohar, B.; Babić, D.; Trinajstić, N. A Novel Definition of the Wiener Index of Trees. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 153–154.

(17) Randić, M.; Guo, X. F.; Oxley, T.; Krishnapriyan, H. Wiener Matrix–Source of Novel Graph Invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.

(18) Klein, D. J.; Lukovits, I.; Gutman, I. On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50–52.

(19) Gutman, I. Wiener Numbers of Benzenoid Hydrocarbons: Two Theorems. *Chem. Phys. Lett.* **1987**, *136*, 134–136.

(20) Recall that a catacondensed benzenoid hydrocarbon is a benzenoid hydrocarbon in which no carbon atoms belong to three six-membered rings. All catacondensed benzenoid hydrocarbons possessing an equal number (= h) of six-membered rings are isomers; they all have the same formula: $C_{4h+2}H_{2h+4}$. For more details on this matter, see: Gutman, I.; Cyvin, S. J. *Introduction to the Theory of Benzenoid Hydrocarbons*; Springer-Verlag: Berlin, 1989.

CI940139R