A Topological Index Based on Distances of Edges of Molecular Graphs

Ernesto Estrada[†] and Ivan Gutman*[,‡,§]

Departamento de Diseño de Fármacos, Centro de Bioactivos Químicos, Universidad Central de Las Villas, Santa Clara 54830, Villa Clara, Cuba, and Institute of Physical Chemistry, Attila József University, P.O. Box 105, H-6701 Szeged, Hungary

A novel molecular topological index *MTI(E)* is proposed, based on edge-distances in molecular graphs. Edge-distances in a graph are equal to distances between vertices of the respective line graph. A simple relation is found between edge-distances and the distances between the vertices that are incident to the respective edges. *MTI(E)* is defined in analogy to the graph theoretical invariant of Schultz (Schultz, H. P. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227−228). It has integer values and is easy to compute. *MTI(E)* is found to be well correlated with the heats of formation of alkanes.

INTRODUCTION

In two previous papers[1,2] we conceived a whole class of novel topological indices based on the edge-adjacency matrix of molecular graphs, i.e., on the vertex-adjacency matrix of the respective line graphs. We believe that research along this direction will lead to the development of useful new procedures for the study of quantitative structure−property relationships (QSPR).

The first line-graph-based topological index introduced and investigated by us was the "*edge connectivity index*",[1] the edge-analog of the classical Randić connectivity index.[3] This quantity was eventually recognized[2] to be equal to the Randić index of the line graph. The edge connectivity index was shown to be applicable for the prediction of molar volumes and refractivities of alkanes.[1] It was successfully extended to molecules containing heteroatoms and correlated with molar volumes of several classes of organic compounds.[4]

Encouraged with this initial success we examined[2] the edge-analog of the Wiener index.[5−7] It was found,[2] however, that in the case of alkanes this "new" topological index differs from the standard Wiener index only by a constant, and, consequently, reflects the very same features of molecular structure.

In this work we show that the redundancy observed in the case of the vertex- and edge-versions of the Wiener index does not occur in the case of other distance-based topological indices. In particular, we show this for the molecular topological index of Schultz.[8,9]

Before defining the edge-version of the Schultz index, we wish to note a few facts concerning the usage of edge-adjacencies and line graphs in chemistry.

Let $G = (V,E)$ be a (molecular) graph, with $V = \{v_1,v_2,...,v_n\}$ and $E = \{e_1,e_2,...,e_m\}$ being the vertex- and edge-sets of G, respectively. Then the vertex-adjacency matrix $\mathbf{A}^v = \mathbf{A}^v(G)$ of the graph G is a square matrix of order n, whose (ij)-entry is defined as

$(\mathbf{A}^v)_{ij} = 1$ if the vertices v_i and v_j are adjacent

0 otherwise

The edge-adjacency matrix $\mathbf{A}^e = \mathbf{A}^e(G)$ is a square matrix of order m, defined analogously:

$(\mathbf{A}^e)_{ij} = 1$ if the edges e_i and e_j are incident

0 otherwise

It is straightforward to realize that $\mathbf{A}^e(G) = \mathbf{A}^v(L(G))$, where $L(G)$ stands for the line graph of the graph G.

The vertex-adjacency matrix is a mathematical object almost ubiquitously encountered in chemical graph theory.[10−13] The edge-adjacency matrix has, until recently, found very little application. It is, nevertheless, considered and explicitly defined in many treatises (e.g., on pp 180−181 of ref 10 and pp 32−33 of ref 11). Line graphs are used to visualize the interactions of localized orbitals (see refs 14 and 15, and pp 164−166 of ref 12) and within the Clar aromatic sextext theory.[16,17]

Whereas the vertex-distance-concept is the basis of a large number of topological indices[6,7,18,19] (and is also extensively studied by mathematicians[20]), the closely analogous edge-distances are rarely encountered; we define them in the subsequent section.

EDGE-DISTANCES IN MOLECULAR GRAPHS

Let $G = (V,E)$ be a molecular graph; recall that such a graph is necessarily connected.[11,12] Let $e_1,e_2,...,e_r$ be distinct edges of G, such that e_i and e_{i+1} are incident (i.e., share a common vertex), $i = 1,2,...,r - 1$. Denote by v_{i-1} and v_i the end-vertices of the edge e_i. Let no two among the vertices $v_0,v_1,...,v_r$ coincide. Then

$$\pi = [v_0,e_1,v_1,e_2,v_2,...,e_r,v_r] \qquad (1)$$

is a path in G, connecting v_0 with v_r and having length r.

Definition 1a. The length of a shortest path that connects two vertices is the distance between these vertices.[20]

In the same time the path π can be viewed as connecting the edges e_1 and e_r. This suggests the following.

Definition 1b. The length of a shortest path that connects two edges, minus one, is the distance between these edges.

[†] Universidad Central de Las Villas.
[‡] Attila József University.
[§] On leave from the Faculty of Science, University of Kragujevac, 34000 Kragujevac, Yugoslavia.

Clearly, this edge-distance is identical to the corresponding vertex-distance in the line graph.

In what follows, both the vertex- and edge-distances will be denoted by $d(x,y)$, where x and y stand for the respective vertices or edges.

Theorem 1. Let e_p and e_q be a pair of distinct edges in G. If the end-vertices of e_p are v_i and v_j and the end-vertices of e_q are v_k and v_h, then the distance between e_p and e_q can be expressed as

$$d(e_p,e_q) = min\{d(v_i,v_k),d(v_i,v_h),d(v_j,v_k),d(v_j,v_h)\} + 1$$

Proof. Consider a path π of the form (1), such that $e_p = e_1$ and $e_q = e_r$. Let this be a shortest possible path of this kind; its length is r.

Now, because π needs not be a shortest path between the vertices v_0 and v_r, their distance is either r or $r - 1$ or $r - 2$. It cannot be less than $r - 2$ because otherwise we could choose a path π'

$$\pi' = [v_1,e_1,v_0, \text{other vertices}, v_r,e_r,v_{r-1}]$$

whose length would be less than r, contradicting to the assumption that π is minimal.

Similarly, $d(v_1,v_r)$ and $d(v_0,v_{r-1})$ is either $r - 1$ or $r - 2$.

Only for the vertices v_1 and v_{r-1} we know for sure that their distance is equal to $r - 2$. Namely, the existence of the path π guarantees that $d(v_1,v_{r-1})$ does not exceed $r - 2$. If the distance between v_1 and v_{r-1} would be smaller than $r - 2$, then there would exist a path between the edges e_p and e_q, shorter than π, contradicting the assumption that the length of π is minimal.

Anyway, the minimum among the distances $d(v_i,v_k)$, $d(v_i,v_h)$, $d(v_j,v_k)$, and $d(v_j,v_h)$ is $r - 2$. On the other hand, the distance between e_p and e_q is, by definition, equal to $r - 1$.

Theorem 1 follows.

Corollary. If the graph G is acyclic, then

$$d(e_p,e_q) = \frac{1}{4}[d(v_i,v_k) + d(v_i,v_h) + d(v_j,v_k) + d(v_j,v_h)]$$

Proof. In the case of acyclic graphs, paths between any two vertices are unique, and therefore $d(v_i,v_k)$, $d(v_i,v_h)$, $d(v_j,v_k)$, $d(v_j,v_h)$ are equal to r, $r - 1$, $r - 1$, and $r - 2$ (not necessarily in the listed order).

Theorem 1 could be important because it permits one to calculate more realistic distances between edges in a graph by using the real distances between atoms in the molecule. It also shows how to obtain topographic descriptors based on distances between bonds in molecules.

A MOLECULAR TOPOLOGICAL INDEX BASED ON EDGE-DISTANCES

The *molecular topological index* was introduced by Schultz for the characterization of alkanes.[8] We denote it by $MTI(V)$ in order to indicate that it is based on vertex-distances, in contrast to analogous $MTI(E)$ which is based on edge-distances. $MTI(V)$ appears to be an attractive graph-theoretical structure-descriptor because it is integer-valued, easy to compute, and has structural significance.[7−9,18] Following the notation used by Klein et al.,[21] $MTI(V)$ is defined as follows

$$MTI(V) = \sum_{j=1}^{n} s_j \tag{2}$$

where

$$s_j = \sum_{i=1}^{n} \mathbf{v}_i^v[(\mathbf{A}^v)_{ij} + (\mathbf{D}^v)_{ij}] \tag{3}$$

In eq 3 $\mathbf{v}_i^v$ is the degree or valency of the ith vertex of the molecular graph G, $(\mathbf{A}^v)_{ij}$ is the element of the vertex-adjacency matrix, whereas $(\mathbf{D}^v)_{ij}$ is the element of the vertex-distance matrix of the molecular graph.

In Figure 1 the vertex-based molecular topological index of 31 alkanes is plotted versus (the negative values of) the standard heats of formation, taken from ref 22.

The edge-based version of the molecular topological index is now defined in full analogy to eqs 2 and 3

$$MTI(E) = \sum_{j=1}^{n} t_j \tag{4}$$

where

$$t_j = \sum_{i=1}^{n} \mathbf{v}_i^e[(\mathbf{A}^e)_{ij} + (\mathbf{D}^e)_{ij}] \tag{5}$$

In eq. (5) $\mathbf{v}_i^e$ is the degree of the i-th edge of the molecular graph G, whereas $(\mathbf{A}^e)_{ij}$ and $(\mathbf{D}^e)_{ij}$ are the elements of the edge-adjacency and edge-distance matrices, respectively.

In Figure 2 the edge-based molecular topological index is plotted versus the heats of formation of the same set of 31 alkanes as in Figure 1. The fact that the correlation of heats of formation with $MTI(E)$ is significantly better than the correlation with $MTI(V)$ is evident without any further statistical analysis. The correlation between $MTI(E)$ and $-\Delta H_f^\circ$ is clearly curvilinear; as seen from the data shown below, $-\Delta H_f^\circ$ is, as a good approximation, a linear function of $\sqrt{MTI(E)}$. No attempt to optimize the exponent 1/2 in formula 6 has been made.

The data shown in Figure 2 conform to the equation

$$-\Delta H_f^\circ (\text{kJ/mol}) = 96.568 + 8.080 MTI(E)^{0.5} \tag{6}$$

$$r = 0.9948 \quad s = 3.18 \quad F = 2747$$

As can be seen from the statistical parameters of eq 6, $MTI(E)$ possesses the basic quality required for topological indices: good correlation with at least one molecular property. Another desirable feature of topological indices is their (mutual) linear independence.[23] This requirement secures that novel topological indices are not trivially related to other previously defined topological indices. Klein et al.[21] and later one of the present authors[24] proved that the $MTI(V)$ and the vertex-distance-based Wiener index, $W(V)$, of acyclic molecules are closely related graph invariants: the difference between $MTI(V)$ and $4W(V)$ depends only on the number of vertices of the molecular graph. It is thus no surprise that very good linear correlations ($r \geq 0.999$) were observed between $MTI(V)$ and $W(V)$.[25,26]

In order to test if the novel $MTI(E)$ is related to the Wiener index $W(E)$, based on edge distances, we examined their correlation for the series of alkanes employed in Figures 1
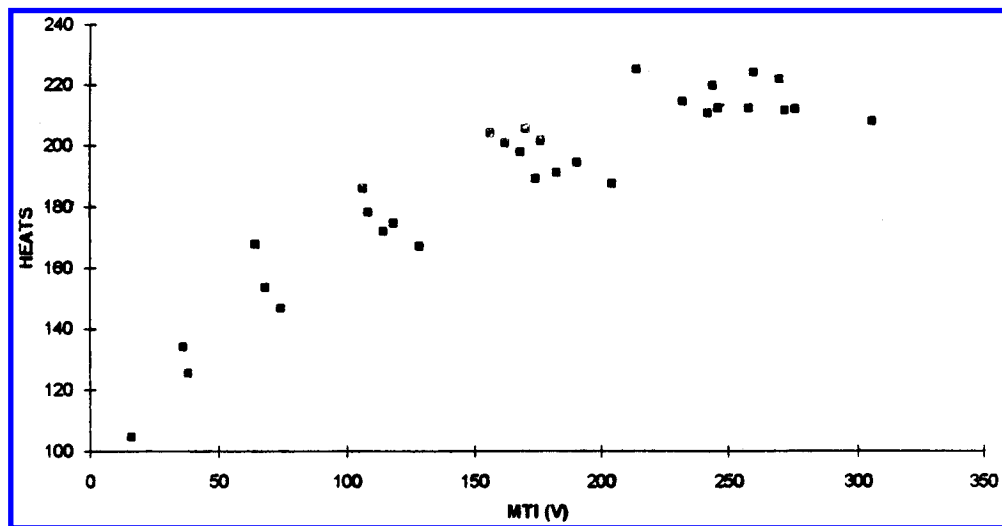
Figure 1. Correlation between experimental heats of formation of alkanes and the vertex-distance-based molecular topological index *MTI*(*V*); the heats of formation are taken from ref 22 and are given in kJ/mol.
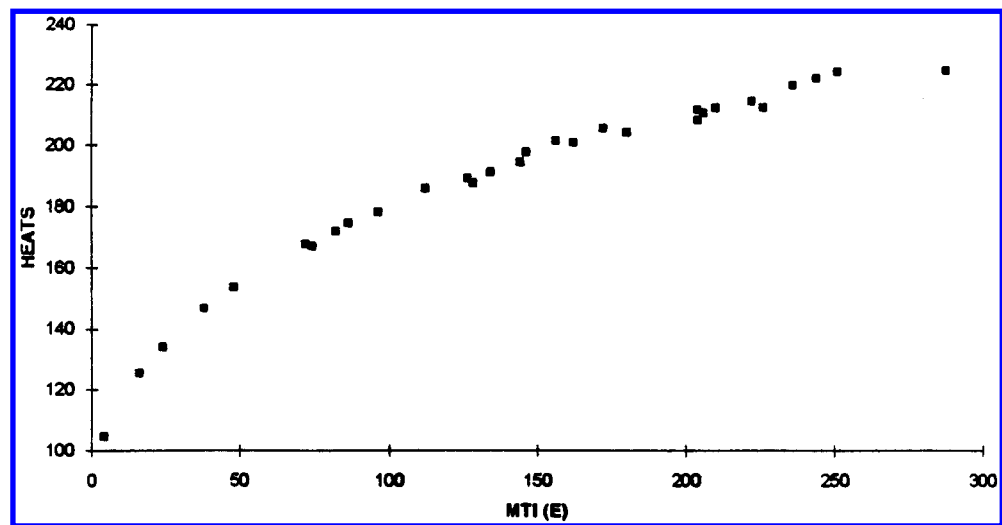
Figure 2. Correlation between experimental heats of formation of alkanes and the edge-distance-based molecular topological index *MTI*(*E*); the heats of formation are taken from ref 22 and are given in kJ/mol.

and 2. We found

$$MTI(E) = 4.45W(E) + 27.12$$

$$r = 0.867 \quad s = 38.9$$

which is very different from what was obtained for the original Schultz and Wiener indices

$$MTI(V) = 3.59W(V) + 3.37$$

$$r = 0.9999 \quad s = 1.33$$

These findings provide a clear evidence that *MTI*(*E*) is statistically independent of the Wiener index. As a consequence, *MTI*(*E*) and Schultz's *MTI*(*V*) are only weakly correlated. Indeed, we found

$$MTI(V) = 0.873MTI(E) - 3.272$$

$$r = 0.916 \quad s = 31.5$$

The edge-based molecular topological index *MTI*(*E*) has, like the Schultz index *MTI*(*V*), a number of interesting features. It is an integer number with physical significance and easy to calculate. It appears to be not highly discriminative. For instance, in the studied series of alkanes there are

two pairs of isomers with degenerate values of this index. These are octane and 4-methylheptane, *MTI*(*E*) = 204, as well as 3-methylheptane and 3,4-dimethylhexane, *MTI*(*E*) = 210. However, these compounds have relatively close values of heats of formation, in agreement with the observation[27] that the discriminative power of topological indices is not necessarily related to their performance in QSPR studies.

The topological index *MTI*(*E*) can be straightforwardly modified into a topographic index, in which instead of vertex and edge distances we employ distances between atoms and chemical bonds in real molecules, and in which the distances between chemical bonds are determined using the present Theorem 1. This is also valid for the Wiener index *W*(*E*) and for other topological indices based on edge-distances in molecular graphs.

REFERENCES AND NOTES

(1) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31−33.
(2) Gutman, I.; Estrada, E. Topological Indices Based on the Line Graph of the Molecular Graph. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 541−543.
(3) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

A Topological Index

J. Chem. Inf. Comput. Sci., Vol. 36, No. 4, 1996 **853**

(4) Estrada, E. Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701−707.

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

(6) 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.

(7) Nikolić, S.; Trinajstić, N.; Mihalić, Z. The Wiener Index: Development and Applications. *Croat. Chem. Acta* **1995**, *68*, 105−130.

(8) Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227−228.

(9) Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. Molecular Topological Index. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 160−163.

(10) *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976.

(11) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; 2nd ed. 1992.

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

(13) *Chemical Graph Theory−Introduction and Fundamentals*; Bonchev, D., Rouvray, D. H., Eds.; Gordon & Breach: New York, 1991.

(14) Sana, M.; Leroy, G. Graph Theory, Electronic Structures and Reaction Mechanisms. *J. Mol. Struct. (Theochem)* **1984**, *109*, 251−269.

(15) Narumi, H.; Kita, H. Equivalent Bond Index. *Commun. Math. Chem. (MATCH)* **1994**, *30*, 125−142.

(16) Gutman, I. Topological Properties of Benzenoid Systems. IX. On the Sextet Polynomial. *Z. Naturforsch.* **1982**, *37a*, 69−73.

(17) Gutman, I.; El-Basil, S. Topological Properties of Benzenoid Systems. XXIV. Computing the Sextet Polynomial. *Z. Naturforsch.* **1984**, *39a*, 276−281.

(18) Mihalić, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701−712.

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

(20) Buckley, F.; Harary, F. *Distance in Graphs*; Addison-Wesley: Redwood, 1990.

(21) Klein, D. J.; Mihalić, Z.; Plavšić, D.; Trinajstić, N. Molecular Topological Index: A Relation with the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 304−305.

(22) Schultz, H. P.; Schultz, T. P. Topological Organic Chemistry. 3. Graph Theory, Binary and Decimal Adjacency Matrices, and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 144−147.

(23) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155−168.

(24) Gutman, I. Selected Properties of the Schultz Molecular Topological Index. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1087−1089.

(25) Mihalić, Z.; Nikolić, S.; Trinajstić, N. A Comparative Study of Molecular Descriptors Derived from the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37.

(26) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Klein, D. J. Relation between the Wiener Index and the Schultz Index for Several Classes of Chemical Graphs. *Croat. Chem. Acta* **1993**, *66*, 345−353.

(27) Razinger, M.; Chretien, J. R.; Dubois, J. E. Structural Selectivity of Topological Indexes in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 24−27.