communications in mathematical
match
and in computer chemistry

no. 35, March 1997

ISSN 0340-6253

MATCDY (35) 157-167 (1997)

MATRIX ALGEBRAIC MANIPULATION OF MOLECULAR GRAPHS. 2. HARARY- AND MTI–LIKE MOLECULAR DESCRIPTORS

Ernesto ESTRADA* and Lissette RODRÍGUEZ

Departamento de Diseño de Fármacos, Centro de Bioactivos Químicos,
Universidad Central de Las Villas,
Santa Clara 54830, Villa Clara, Cuba

(Received: July 1996)

Abstract: The Harary *H* number and *MTI* descriptor were expressed by means of a matrix-vector-matrix multiplication procedure. Two new series of topological indices analogous to them were defined. *MTI*-like descriptors were expressed as sums of other topological indices defined in the literature. Both series of molecular descriptors were used to describe boiling points of 21 alkanes. Very significant improvements relative to Harary number were obtained by using some of the novel Harary-like numbers. Almost all *MTI*-like descriptors show better correlations with boiling point than the original *MTI*. All the molecular descriptors defined here are expressed in terms of local vertex properties, such as degrees and distance numbers.

INTRODUCTION

The study of topological indices is one of the most important branches of research in chemical graph theory. These indices are simple numbers that characterize the chemical structure as a whole. Generally, they are obtained from structure descriptive matrices such as distance and vertex-adjacency ones [1,2].

The first index defined in a chemical context to describe physical properties of molecules was the number introduced by H. Wiener in 1947 [3]. This index was later expressed by Hosoya

[4] by means of the distance matrix. From that time up to now, distance matrix has been widely used for generation of topological indices that characterize to some extent the chemical structure [5]. The recent definition of some distance-based topological indices illustrates the continued interest in such descriptors. Among these new topological indices we can mention the Harary number introduced by Plavsic et al. [6], the Szeged index defined by Gutman et al. [7], and the Hyper-Wiener index suggested by Randic [8].

Other topological indices have been defined by using a combination of distance and vertex-adjacency matrices. The most studied of these indices is the *MTI* introduced by Schultz in 1989 [9]. Klein et al. [10] and later Gutman [11] proved that for 4-trees representing alkanes this index was directly related to the Wiener number. The difference between *MTI* and 4W depends only on the number of vertices of the molecular graph. An edge-based *MTI* index was defined by Estrada and Gutman [12]. It showed excellent correlations with the heat of formation of alkanes and no correlation with Wiener number.

In the previous paper of this series [13] we used a vector-matrix-vector multiplication procedure to define a series of topological indices. We showed that the Wiener number can be defined through this procedure by using unit vectors and distance matrix. Other topological indices, such as the Zagreb group indices [14], can also be defined by using this approach. Some of the novel descriptors represent an improvement in statistical quality of correlations describing boiling points of alkanes with respect to the Wiener and Zagreb group indices. The structural selectivity of these indices was also improved by some of the novel graph invariants [13].

In the present work we propose the use of the vector-matrix-vector multiplication procedure to study a series of graph invariants analogous to Harary number [6] and *MTI* descriptors [9]. We show that some of the novel descriptors represent significant improvements to Harary numbers. We also show that MTI-like descriptors are combinations of other topological indices that can be defined by vector-matrix-vector multiplication of distance and vertex-adjacency matrices, independently.

HARARY NUMBER-LIKE DESCRIPTORS

Let $G = (V, E)$ be a molecular graph, with $V = \{v_1, v_2, \ldots v_n\}$ and $E = \{e_1, e_2 \ldots e_m\}$ be the vertex- and edge sets, respectively. The distance *d(i, j)* between two vertices *i* and *j* is defined as the length of the shortest path between the corresponding vertices. The non diagonal entry (*i, j*) of the distance **D** matrix is the distance between vertices v_i and v_j in the graph. On the other hand, the elements $a_{i,j}$ of the vertex-adjacency matrix **A** are ones or zeros depending on whether the vertices *i* and *j* are adjacent or not. Two vertices are adjacent if they are connected by an edge.

The *Harary number H* of a molecular graph was introduced by Plavsic et al. [6] in honor of the 70th birthday of Professor Frank Harary. It is defined as follows:

$$H = \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} \left(d_{ij}\right)^{-2} \tag{1}$$

where $\left(d_{ij}\right)^{-2}$ are the squares of the reciprocal distances in G. They are the elements of the $\mathbf{D}^{-2}$ matrix. Mihalic and Trinajstic [15] have stated that the $\mathbf{D}^{-2}$ matrix can be considered as the distance matrix of a class of specially weighted graphs in which weights between vertices in G mimic the Coulomb law between the sites in the corresponding structure.

Harary number can be defined following our vector-matrix-vector multiplication procedure as in expression 2, in which **u** is a unit vector and $\mathbf{u}^T$ is the transpose of **u**.

$$H = \frac{1}{2} \left(\mathbf{u}\mathbf{D}^{-2}\mathbf{u}^T\right) \tag{2}$$

Similarly to in our previous paper [13] we can introduce a new series of topological indices by using different vectors **p** and **q**, and expressions of the form:

$$H(no.) = \frac{1}{2} \left(\mathbf{p}\mathbf{D}^{-2}\mathbf{q}^T\right) \tag{3}$$

Vectors **p** and **q** can be selected in different ways. We will use the vertex degree **v** and the distance numbers **d** vectors in the present study. The element δ_i of the **v** vector is the degree of the vertex *i*, in other words it is the sum of all entries of the *i*th row or column of **A** matrix. Elements of the distance **d** vector are the corresponding distance numbers D_i for the different vertices in G, where the *distance number* D_i is the sum of all entries in the *i*th row or column of

the distance matrix. Five new graph theoretical invariants will be introduced by using expression 3. This number is a consequence of the equality $\mathbf{pDq}^{T} = \mathbf{qDp}^{T}$, which reduces the number of topological indices to those defined in Table 1. We will denote the Harary number as $H(0)$ in order to indicate that it is also defined by using expression 3.

Table 1. Harary-like Graph Theoretical Invariants.

H(no	invariant	H(no	invariant
1	$\frac{1}{2}\left(\mathbf{vD}^{-2}\mathbf{u}^{T}\right.$	4	$\frac{1}{2}\left(\mathbf{vD}^{-2}\mathbf{d}^{T}\right)$
2	$\frac{1}{2}\left(\mathbf{dD}^{-2}\mathbf{u}^{T}\right.$	5	$\frac{1}{2}\left(\mathbf{dD}^{-2}\mathbf{d}^{T}\right)$
3	$\frac{1}{2}\left(\mathbf{vD}^{-2}\mathbf{v}^{T}\right.$		

Graph theoretical invariants given in Table 1 can be expressed in terms of vertex degrees and distance numbers of the vertices of the corresponding molecular graphs. We have shown [13] that this kind of expression can be very useful for the definition of novel graph theoretical descriptors. These expressions are given below:

$$\frac{1}{2}\left(\mathbf{vD}^{-2}\mathbf{u}^{T}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\delta_i}{d_{ij}^2} \tag{4}$$

$$\frac{1}{2}\left(\mathbf{dD}^{-2}\mathbf{u}^{T}\right) = \frac{1}{2}\sum_{i=1}^{n} D_i D_i^{-2} \tag{5}$$

$$\frac{1}{2}\left(\mathbf{vD}^{-2}\mathbf{v}^{T}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\delta_i\delta_j}{d_{ij}^2} \tag{6}$$

$$\frac{1}{2}\left(\mathbf{vD}^{-2}\mathbf{d}^{T}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\left(D_i\delta_j + D_j\delta_i\right)}{d_{ij}^2} \tag{7}$$

$$\frac{1}{2}\left(\mathbf{dD}^{-2}\mathbf{d}^{T}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{D_i D_j}{d_{ij}^2} \tag{8}$$

Expressions (6) and (8) have some resemblance with the electrostatic energy of a molecule, in which degrees and distance number of vertices can be considered as measures of atomic charges. The further manipulation of expression 4-8 can produce some novel structure-descriptive topological indices to be used in QSPR studies.

MTI-LIKE DESCRIPTORS

The *molecular topological index MTI* was introduced by Schultz [9] for the characterization of alkanes. The *MTI* appears to be an attractive graph-theoretical descriptor, that is integer-valued, easy to compute and with structural significance. It can be defined as follows:

$$MTI = \sum_{j=1}^{n} s_j \tag{9}$$

where

$$s_j = \sum_{i=1}^{n} \delta_i \left(a_{ij} + d_{ij} \right) \tag{10}$$

Using our vector-matrix-vector multiplication scheme the *MTI* can be defined in the following way:

$$MTI = \mathbf{v}(\mathbf{A} + \mathbf{D})\mathbf{u}^{\mathrm{T}} \tag{11}$$

A series of *MTI*-like descriptors can be introduced now in a similar way as for the Wiener index and Harary number. This series of graph theoretical invariants is given in Table 2. The Schultz's *MTI* will be denoted here as MTI(0).

Table 2. MTI-like Graph Theoretical Invariants Based on Vector-Matrix-Vector Multiplication.

MTI(no	invariant	*MTI*(no	invariant
1	$\mathbf{u}(\mathbf{A} + \mathbf{D})\mathbf{u}$	4	$\mathbf{v}(\mathbf{A} + \mathbf{D})\mathbf{d}$
2	$\mathbf{d}(\mathbf{A} + \mathbf{D})\mathbf{u}$	5	$\mathbf{d}(\mathbf{A} + \mathbf{D})\mathbf{d}$
3	$\mathbf{v}(\mathbf{A} + \mathbf{D})\mathbf{v}$		

The MTI-like descriptors are expressed as simple sums of some topological indices defined in the literature or in our previous paper about matrix algebraic manipulation of molecular graphs. These expressions are given below for *MTI*(0)-*MTI*(5):

$$MTI(0) = \mathbf{vAu}^T + \mathbf{vDu}^T = M_1 + TI(1) \quad (12)$$

$$MTI(1) = \mathbf{uAu}^T + \mathbf{uDu}^T = 2(m + W) \quad (13)$$

$$MTI(2) = \mathbf{dAu}^T + \mathbf{dDu}^T = \mathbf{vDu}^T + \mathbf{dDu}^T = TI(1) + TI(2) \quad (14)$$

$$MTI(3) = \mathbf{vAv}^T + \mathbf{vDv}^T = 2M_2 + TI(3) \quad (15)$$

$$MTI(4) = \mathbf{vAd}^T + \mathbf{vDd}^T = TI(8) + TI(4) \quad (16)$$

$$MTI(5) = \mathbf{dAd}^T + \mathbf{dDd}^T = TI(9) + TI(5) \quad (17)$$

In expressions (12)-(17), M_1 and M_2 are the Zagreb group indices introduced by Gutman et al. [14], m is the number of edges in the graph, W is the Wiener number [3], and *TI*(no.) are topological indices defined previously [13]. As can be seen from Eqs. 12-17, all MTI-like descriptors can be expressed as combinations of other topological indices based on distance and vertex-adjacency matrices independently. Expression 12 relating Schultz's *MTI* to Zagreb group M_1 index and the relation found by Klein et al. [10] and later by Gutman [11] that relates this index to Wiener number, prove that *MTI* cannot be considered as an independent topological index but as an index based on combination of other descriptors. However, several authors [16] have shown that there are some properties that are better described by using a combination of topological indices than by using the simple descriptors.

QUANTITATIVE STRUCTURE-PROPERTY STUDIES

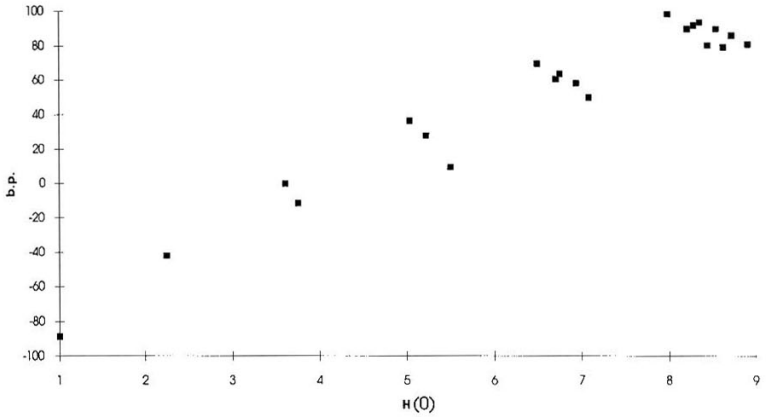
Graph theoretical invariants defined in the present work were calculated for a series of 21 alkanes with 2-7 carbon atoms. All descriptors were used to describe the experimental boiling points of the studied alkanes taken from ref. [17]. We explore different functional forms for expressions that relate boiling points to topological indices. These functional forms include the square, cubic, fourth, fifth and sixth roots of the indices. The square root of the Harary number

shows the best correlation of this index with the boiling points of studied alkanes. The model found has the following form:

$$b.\,p.\,(^{\circ}C) = 92.263[H(0)]^{1/2} - 181.651 \qquad (18)$$
$$r = 0.9774 \quad s = 10.88 \quad F = 405$$

This equation is similar to that obtained by Mihalic and Trinajstic [15] for the description of boiling points of 40 alkanes by using the same descriptor. These authors obtained a regression coefficient of 0.9864 and a standard deviation of 10.46 for boiling points of alkanes up to 8 carbon atoms. In Fig. 1 the Harary number of the 21 alkanes studied here is plotted versus the boiling points.

Figure 1. Correlation Between Experimental Boiling Points of Alkanes and the Harary Number.



The use of some of the Harary-like topological indices produce significant improvements in the statistical parameters of the regressions describing boiling points of alkanes. These improvements are observed only for those indices in which the $\mathbf{D}^{-2}$ matrix is multiplied by the **d** vector. No one improved model could be found by using H(1) and H(3) indices. The models found with the use of H(2), H(4) and H(5) are given below:

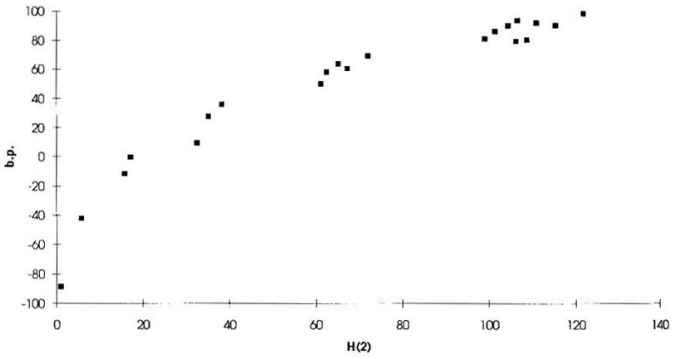
$$b.\,p.\,(^{\circ}C) = 78.948[H(2)]^{1/4} - 166.191 \quad (19)$$
$$r = 0.9948 \quad s = 5.25 \quad F = 1803$$

$$b.\,p.\,(^{\circ}C) = 61.628[H(4)]^{1/4} - 149.955 \quad (20)$$
$$r = 0.9855 \quad s = 8.72 \quad F = 642$$

$$b.\,p.\,(^{\circ}C) = 97.607[H(5)]^{1/7} - 186.027 \quad (21)$$
$$r = 0.9960 \quad s = 4.59 \quad F = 2362$$

As can be observed, the $\frac{1}{2}(\mathbf{dD}^{-2}\mathbf{d}^{T})$ and $\frac{1}{2}(\mathbf{dD}^{-2}\mathbf{u}^{T})$ indices diminished the standard deviation by more than 50 % relative to the original Harary number, while the improvement produced by $\frac{1}{2}(\mathbf{vD}^{-2}\mathbf{d}^{T})$ index is of nearly 20 %. In Fig. 2 we can appreciate the correlation of H(2) index versus boiling points of studied alkanes.

Figure 2. Correlation Between Experimental Boiling Points of Alkanes and the H(2) Number.



If we compare this figure to Fig. 1, we can appreciate the lower dispersion of data points that is obtained with the H(2) index. This is a consequence of the lower clustering of H(2) index compared to the original Harary number. Clustering effect is also observed in other topological indices, such as the ID numbers [18] of Randic. This undesirable effect was studied by Balaban [19] for the ID numbers. This author used the distance numbers instead of vertex degrees in the calculation of the ID numbers, resulting in ID' values. In this case the clustering disappears. A similar effect has been obtained here by introducing the distance sum vector for the calculation of H(2), H(4) and H(5). Considering the statistical quality of the regression models and the

reduction of clustering, we can conclude that these three indices represent significant improvements relative to the original Harary number.

The MTI-like descriptors were also used to describe boiling points of the same series of alkanes. The respective statistical parameters are given in Table 3. We recall that MTI(0) is the original Schultz's MTI.

Table 3. Statistical Parameters[a] of Models Describing Boiling Points with MTI-like Descriptors.[b]

M.T.I.	exponent			
	1/3	1/4	1/5	1/6
0	0.9943	0.9950	0.9944	0.993
	5.49	5.14	5.44	5.84
1	0.9947	0.9960	0.9961	0.995
	5.27	4.57	4.56	4.74
2	0.9839	0.992	0.9950	0.996
	9.19	6.51	5.14	4.53
3	0.9951	0.9962	0.9956	0.994
	5.08	4.49	4.83	5.34
4	0.9837	0.9919	0.9948	0.995
	9.25	6.53	5.25	4.81
5	0.9639	0.9809	0.9889	0.992
	13.70	10.00	7.65	6.15

[a] Top value is the regression coefficient and botton value the standard deviation.
[b] Models of the form: $y = a + b\ (T.I.)^{exp}$

As can be appreciated in Table 3, all models obtained with MTI-like indices describing boiling points of studied alkanes represent improvement with respect to the original index. The only exception is the MTI(5) index, for which no improved model could be found by using exponents given in this table. However, $[MTI(5)]^{1/8}$ model produces a drop of 8.4 % in the standard deviation with respect to the best model found with Schultz's MTI. This model has r = 0.9958 and s = 4.71.

In a previous section we proved that all *MTI* descriptors are combinations of other topological indices defined from distance and vertex-adjacency matrices independently (see expressions 12-17). Taking this fact into account, we can observe that the simple sum of Wiener index and the number of edges in the graph, which is designed as MTI(1), produces better correlations with boiling points of alkanes than the sum $4W + 2p_2 - (n-1)(n-2)$ [20], which is the original *MTI* index [10].

On the other hand, there are other combinations of simple topological indices that produce better correlations with boiling points than the *MTI* index. An example is given by MTI(2), which is a combination of TI(1) and TI(2) defined in the previous paper of this series [13]. The

expression of Schultz's *MTI* as a sum of the Zagreb group M_1 and the TI(1) defined by us (see expression 11) permits to gain insight into the characteristic features of this topological index. While TI(1) index produces good models with boiling points of alkanes [13], the M_1 index is poorly correlated to this property. This fact explains why the use of TI(2) instead of M_1 in the definition of MTI(2) produces a significant improvement when it is used to describe boiling points of alkanes.

The present approach of vector-matrix-vector multiplication has proved to be useful in the design and analysis of graph theoretical invariants. This methodology can also be applied to other graph theoretical invariants based on edge-distances [12, 21] and edge-adjacency matrices [22], and novel descriptors with improved ability to describe physical properties can, perhaps, be defined.

When to stop searching for novel graph theoretical inavariants? We cannot answer precisely to this question, but we can say that the search of novel graph invariants describing the molecular structure is necessary for the development of structure-descriptive models in chemistry. These models are necessary for the design and rational synthesis of new drugs, plastics, materials, etc. The basis of this search can be found in the XIX century [23], however the pioneering work of H. Wiener in 1947 with the introduction of the first topological index and its applications to structure-property studies can be considered as a very important step in the development of a predictive chemistry.

Acknowledgement. We thank Prof. Ivan Gutman for his valuable help and patience in the edition of this manuscript.

REFERENCES

[1] N. Trinajstic, Chemical Graph Theory, 2nd revised ed.; CRC: Boca Raton, FL, 1992; chapter 3.

[2] D. H. Rouvray, in: R. B. King (Ed.), Chemical Applications of Topology and Graph Theory, Elsevier, Amsterdam 1983, p. 159.

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

[4] H. Hosoya, Bull. Chem. Soc. Japan **44**, 2332 (1971).

[5] Z. Mihalic, S. Nikolic and N. Trinajstic, J. Chem. Inf. Comput. Sci. **32**, 28 (1992).

[6] D. Plavsic, S. Nikolic, N. Trinajstic and Z. Mihalic, J. Math. Chem. **12**, 235 (1993).

[7] I. Gutman, Graph Theory Notes New York, **27**, 9 (1994); P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. Dobrynin, I. Gutman and G. Domotor, J. Chem. Inf. Comput. Sci. **35**, 547 (1995).

[8] M. Randic, Chem. Phys. Lett. **211**, 478 (1993).

[9] H. P. Schultz, J. Chem. Inf. Comput. Sci. **29**, 227 (1989).

[10] D. J. Klein, Z. Mihalic, D. Plavsic and N. Trinajstic, J. Chem. Inf. Comput. Sci. **32**, 304 (1992).

[11] I. Gutman, J. Chem. Inf. Comput. Sci. **34**, 1087 (1994).

[12] E. Estrada and I. Gutman, J. Chem. Inf. Comput. Sci. In press.

[13] E. Estrada, L. Rodriguez and A. Gutierrez, Commun. Math. Chem. (MATCH), preceding paper.

[14] I. Gutman and Trinajstic, Chem. Phys. Lett. **17**, 535 (1972); I. Gutman, B. Ruscic, N. Trinajstic, and C. F. Wilcox, J. Chem. Phys. **62**, 3399 (1975).

[15] Z. Mihalic and N. Trinajstic, J. Chem. Educ. **69**, 701 (1992).

[16] See for instance: M. Randic, Croat. Chem. Acta, **64**, 43 (1991), and the references cited therein.

[17] R. C. Weast, CRC Handbook of Chemistry and Physics, 67th ed., 3rd printing, CRC, Boca Raton 1987.

[18] M. Randic, J. Chem. Inf. Comput. Sci. **24**, 164 (1984).

[19] A. T. Balaban, in: R. B. King and D. H. Rouvray (Eds.), Graph Theory and Topology in Chemistry, Elsevier, Amsterdam 1987, p. 159.

[20] The term p_2 is the number of path of length two in a tree. See ref. [10].

[21] I. Gutman and E. Estrada, J. Chem. Inf. Comput. Sci. **36**, 541 (1996).

[22] E. Estrada, J. Chem. Inf. Comput. Sci. **35**, 31 (1995); E. Estrada, J. Chem. Inf. Comput. Sci. **35**, 701 (1995).

[23] D. H. Rouvray, J. Mol. Struct. (Theochem) **185**, 187 (1989).