communications in mathematical
match
and in computer chemistry

no. 35, March 1997

ISSN 0340-6253

MATCDY (35) 145-156 (1997)

MATRIX ALGEBRAIC MANIPULATION OF MOLECULAR GRAPHS. 1. DISTANCE AND VERTEX–ADJACENCY MATRICES

Ernesto ESTRADA*, Lissette RODRÍGUEZ and Amauri GUTIÉRREZ

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 Wiener number is expressed in terms of vector-matrix multiplication procedure by using distance matrix and unit vectors. This approach is extended to the definition of a new series of graph theoretical invariants based on distance and vertex-adjacency matrices. The performance of these topological indices, which include the Zagreb group indices, for the description of boiling points of alkanes is analyzed. It is shown that the Wiener index is one of the best descriptors that can be defined by this procedure. However, some improvements to it are obtained by using some of the novel graph invariants. One of the invariants based on vertex-adjacency matrix represents a very significant improvement relative to Zagreb group indices. The structural selectivity of all descriptors is also analyzed.

INTRODUCTION

One of the main objectives of chemical graph theory is to find different ways to numerically characterize the chemical structure. The chemical structure can be reduced to mathematical objects by transforming the chemical graphs into graph theoretical invariants. These invariants are numbers or set of numbers that do not depend on the enumeration of vertices or edges in the graph. The structure-descriptive matrices, such as distance and vertex-adjacency matrices, are commonly used as sources of these invariants [1,2].

Graph theoretical invariants of molecular graphs contain important structural information about molecules. These invariants can be used as molecular descriptors in structure-property or structure-activity studies. Among the best known and most often used descriptors are the so called topological indices. These indices are simple numbers that characterize the chemical structure as a whole [3,4].

The first reported use of a topological index in a chemical context was by Wiener [5] in 1947. This quantity is defined as the half-sum of the elements of the distance matrix. It is known as *Wiener number* or *Wiener index* and denoted by W. There is a lot of published papers [6-8] treating the Wiener number, some of them related to the mathematical aspects of its calculation in certain classes of chemical graphs and others to its applications in chemical problems.

From this successful beginning of topological indices, various authors have devised several ways to obtain such kind of molecular descriptors. At present, there are more than 120 topological indices described in the literature [9].

The proliferation of graph theoretical indices can be explained by the efforts of researchers in the search for better descriptors that contain more structural information. However, there is an appreciable quantity of topological indices with duplicate structural information and some of them are trivially related to each other [10]. In order to avoid these problems, Randic [11] has published a list of desirable attributes for topological indices.

An interesting route to find novel topological indices is to explore the algebraic manipulation of the known graph theoretical matrices. The molecular topological index proposed by Schultz [12] is one of the most recent examples of this approach. Some years ago, Balaban et al. [13] studied the vector-matrix multiplication as a new way to obtain local vertices' invariants and generate novel topological indices, which resulted in very discriminative descriptors with good ability to describe boiling points of alkanes.

In the present paper we begin the study of the generation of graph invariants based on algebraic work on distance and vertex-adjacency matrices of molecular graphs. Our main objective is to study invariants analogous to the Wiener index by using a vector-matrix multiplication procedure. We will show that some improvements in the ability of Wiener-like topological indices to correlate boiling points of alkanes and to discriminate among different structures can be introduced by using this approach.

THEORETICAL APPROACH

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

The Wiener number is defined as the half-sum of the elements of D matrix. It can be mathematically defined as follows:

$$W = \frac{1}{2}\sum_i \sum_j d(i, j) = \frac{1}{2}\sum_{v \in V} D(v) \qquad (1)$$

where $D(v)$ is the sum of all entries in the vth row or column of the distance matrix. This number is called the *distance number* of the vertex v.

We point out that the Wiener number can be obtained by vector-matrix multiplication as in expression 2, where **u** is a unit vector of order n and $\mathbf{u}^T$ is its transpose.

$$W = \frac{1}{2}(\mathbf{uDu}^T) \qquad (2)$$

Taking into account this expression we propose the generation of a new series of topological indices of the form:

$$TI = \mathbf{pDq}^T \qquad (3)$$

In expression 3, **p** and **q** are two selected vectors containing structural information on the graph or, in the most simple cases, the unit vector of order n. From elementary matrix algebra it is straightforward to realize that $\mathbf{pDq}^T = \mathbf{qDp}^T$.

A new series of topological indices (TI) can be obtained from the expression 3 by using the vertex degree **v** vector and the distance **d** vector. 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 ith row or column of **A** matrix. Elements of the distance **d** vector are the corresponding distance numbers $D(v)$ for the different vertices in G.

Using expression 3 and vectors **v** and **d** we can build a total of 5 topological indices. Other 4 descriptors can be obtained by using the **A** matrix instead of **D** in expression 3. All these graph theoretical invariants are illustrated in Table 1.

Table 1. Graph Theoretical Invariants Based on Algebraic Manipulation of Distance and Vertex-Adjacency Matrices.[a]

no.	invariant	no.	invariant
1	$\mathbf{vDu}^T$	6	$\mathbf{vAu}^T$
2	$\mathbf{dDu}^T$	7	$\mathbf{vAv}^T$
3	$\mathbf{vDv}^T$	8	$\mathbf{vAd}^T$
4	$\mathbf{vDd}^T$	9	$\mathbf{dAd}^T$
5	$\mathbf{dDd}^T$		

[a] Topological indices are defined as TI (no.) = invariant.

The number of topological indices is reduced for **A** matrix as a consequence of the equality of two vector-matrix multiplications $\mathbf{Du}^T = \mathbf{d}^T$ and $\mathbf{Au}^T = \mathbf{v}^T$, which produces the following identity: $\mathbf{vDu}^T = \mathbf{dAu}^T$.

Graph theoretical invariants given in Table 1 can be expressed in terms of vertex degrees and distance numbers of the corresponding vertices of the molecular graphs. These expressions will permit the future generation of novel graph theoretical invariants similar to those reported nowadays in the literature. These expressions are as follows:

$$\mathbf{vDu}^T = \mathbf{vd}^T = \sum_{i=1}^{n} \delta_i D_i = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i d_{ij} \tag{10}$$

$$\mathbf{dDu}^T = \mathbf{dd}^T = \sum_{i=1}^{n} (D_i)^2 \tag{11}$$

$$\mathbf{vDv}^T = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j d_{ij} \tag{12}$$

$$\mathbf{dDd}^T = \sum_{i=1}^{n} \sum_{j=1}^{n} D_i D_j d_{ij} \tag{13}$$

$$\mathbf{vDd}^T = \sum_{i=1}^{n} \sum_{j>i}^{n} d_{ij} \left(D_i \delta_j + D_j \delta_i \right) \tag{14}$$

$$\mathbf{vAu}^T = \mathbf{vv}^T = \sum_{i=1}^{n} (\delta_i)^2 = M_1 \tag{15}$$

$$\mathbf{vAv}^T = 2 \sum_{k=1}^{m_1} \left(\delta_i \delta_j \right)_k = 2 M_2 \tag{16}$$

$$\mathbf{vAd}^{\mathrm{T}} = \sum_{k=1}^{m} \left(D_i \delta_j + D_j \delta_i \right)_k \tag{17}$$

$$\mathbf{dAd}^{\mathrm{T}} = 2 \sum_{k=1}^{m} \left(D_i D_j \right)_k \tag{18}$$

In expressions 16-18 the index k runs over the number of adjacent vertices in the graph. As can be appreciated graph theoretical invariants 6 and 7 (see Table 1) are identical to the Zagreb Group indices M_1 and M_2 introduced by Gutman et al. [14]. On the other hand, the graph theoretical invariant 9 (see Table 1 and expression 18) is intimately related to the J index defined by Balaban [15] as:

$$J = \frac{m}{\mu + 1} \sum_{k=1}^{m} \left(D_i D_j \right)_k^{-0.5} \tag{19}$$

All these facts point out that expressions 10-18 can be useful in the definition of other sets of novel graph theoretical invariants.

CHEMICAL APPLICATIONS

The graph theoretical invariants proposed in the present work were calculated for a series of 21 alkanes with 2-7 carbon atoms. Values of topological indices for the studied series of compounds are illustrated in Table 2.

Table 2. Calculated Values of Graph Theoretical Indices.[a]

alkane[b]	W(G)	TI(1)	TI(2)	TI(3)	TI(4)	TI(5)	TI(6)	TI(7)	TI(8)	TI(9)
2	1	2	2	2	2	2	2	2	2	2
3	4	10	22	12	28	60	6	8	16	24
4	10	28	104	38	148	536	10	16	48	128
2M3	9	24	84	30	114	390	12	18	54	90
5	20	60	334	88	508	2764	14	24	106	448
2M4	18	52	270	72	386	2012	16	28	110	328
22MM3	16	44	212	56	296	1400	20	32	128	224
6	35	110	854	170	1358	10330	18	32	198	1218
2M5	32	98	712	146	1104	7864	20	36	202	952
3M5	31	94	670	138	1030	7190	20	38	192	874
22MM4	28	82	544	114	808	5256	24	44	214	648
23MM4	29	86	582	122	874	5818	22	42	202	714
7	56	182	1876	292	3080	31140	22	40	332	2800
2M6	52	166	1612	260	2600	24800	24	44	338	2304
3M6	50	158	1476	244	2392	22210	24	46	318	2088
3E5	48	150	1380	228	2184	19720	24	48	300	1872
22MM5	46	142	1257	214	1956	17040	28	52	349	1692
23MM5	46	142	1260	212	1968	17170	26	52	318	1656
24MM5	48	150	1366	228	2156	19360	26	48	340	1848
223MMM4	42	126	1044	180	1584	12940	30	60	342	1272
33MM5	44	134	1156	196	1784	15100	28	56	326	1472

[a] W (G) is the Wiener index, for the definition of TI (no.) see Table 1.

[b] Numbers before M or E indicate the position of methyl or ethyl substituent in the alkyl chain of length expressed by the last number, for instance 23MM5 means 2,3-dimethylpentane.

Experimental boiling points of these 21 alkanes taken from ref. [16] were used to test the usefulness of the studied graph theoretical invariants in quantitative structure-property relationships (QSPR) studies. An important aspect to be considered in a QSPR study is the functional form of expressions relating topological indices and the physical property. The number of functional forms is unlimited and the exploration of a great number of them is outside the scope of this work. As a consequence we will take the work of Randic et al. [17] about the search of graph theoretical invariants as a starting point in our research. These authors have explored the square, cube, fourth, and fifth roots of the Wiener index. They showed that the fourth root of W number yields a standard deviation of 4.42 in the description of boiling points of the C_2-C_7 alkanes. This improvement represents a drop of near 75 % from the original value of 17.09 obtained from a linear model.

Standard deviations and regression coefficients of models found by using cube, fourth, fifth, and sixth roots of graph theoretical invariants will be used here as statistical criteria of quality. Table 3 illustrates these statistical parameters for models obtained from graph invariants that involve the distance matrix of molecular graphs.

As measured by the standard deviation of 4.31, the $[T.I(1)]^{1/4}$ model represents an improvement of 2.5 % with respect to the Wiener number, and it is the best model developed in this study. On the other hand, the $[T.I(2)]^{1/6}$ gave nearly identical results to W number. The best models obtained with T.I.(3) and T.I.(4) are approximately 5 % worse than the best model obtained here. Finally, there was no improved model with the use of T.I.(6), the best model developed by using this index was nearly 36 % worse than that reported for the Wiener index.

Table 3. Statistical Parameters of Models to Describe Boiling Points with Topological Indices Based on Distance Matrix.[a]

T.I.	exponent			
	1/3	1/4	1/5	1/6
W	0.9951	0.9963	0.9960	0.9953
	5.08	4.42	4.60	4.98
1	0.9947	0.9965	0.9963	0.9955
	5.27	4.31	4.43	4.85
2	0.9834	0.9921	0.9952	0.9963
	9.33	6.47	5.03	4.45
3	0.9926	0.9957	0.9961	0.9957
	6.24	4.75	4.53	4.78
4	0.9813	0.9912	0.9948	0.9961
	9.90	6.83	5.21	4.53
5	0.9631	0.9809	0.9891	0.9931
	13.85	10.01	7.56	6.01

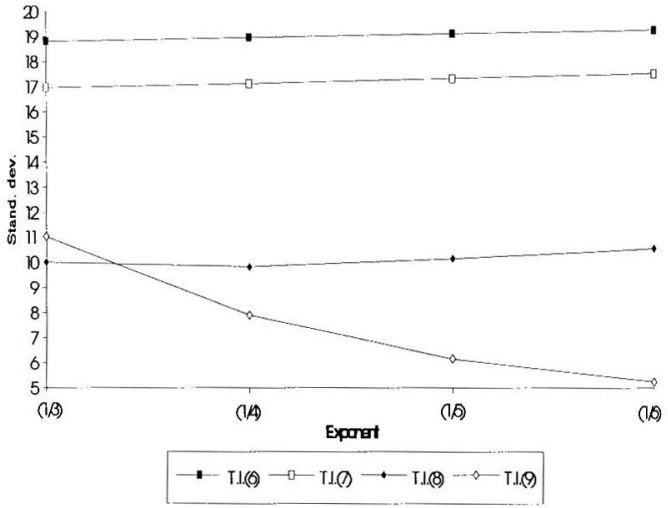
[a] Models of the form: $y = a + b\,(T.I.)^{exp}$

It is interesting that, even when the Wiener index was not obtained by using any algebraic manipulation of distance matrix, *it is one of the best graph theoretical invariants that can be obtained by vector-matrix multiplication procedures of the form studied here.*

The following step in our present analysis is to compare the quality of models involving the vertex-adjacency matrix. We recall here that the topological indices T.I.(6) and T.I.(7) are identical to M_1 and $2M_2$, which are the Zagreb group indices [14]. For

topological indices based on vertex-adjacency matrix the improvements in the statistical quality of models found are dramatic from one index to other. These improvements are better appreciated in a graphical form as in Figure 1.

Figure 1. Standard Deviations of Models to Describe Boiling Points with Topological Indices Based on Vertex-Adjacency Matrix.



As can be appreciated in Fig. 1, the standard deviations of 18.81 and 16.96 obtained with M_1 and M_2 are dropped to 5.24 by using the sixth root of T.I.(9). This represents an improvement of more than 70 %. The $[T.I(9)]^{1/6}$ model is only 18 % worse than the best model obtained with W number.

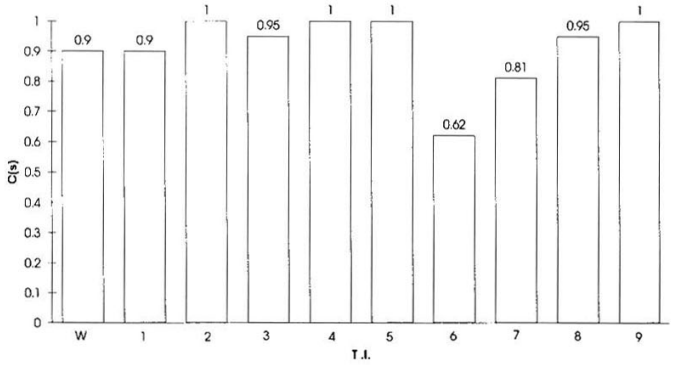
In the previous section we remark the resemblance between expression 18, that relates the T.I.(9) index and the distance numbers, with the Balaban J index. This is an important fact because the J index is a very discriminative descriptor that has not a good

ability to describe boiling points of alkanes [18]. The T.I.(9) index has no duplicated value for the studied series of alkanes, and in addition it has a good correlation with this physical property. Mihalic and Trinajstic [18] heve suggested a quadratic model as the most successful to describe boiling points by using the *J* index. Following this result we have developed a similar model for the studied series of alkanes that has r = 0.8407 and s = 28.61. In this sense, the $[T.I(9)]^{1/6}$ model represents an improvement of more than 80 %, and it is straightforward to define a topological index like *J* but having exponent 1 instead of -1/2 in expression 19.

Now we compare the structural selectivity of each topological index studied in the present work. As a quantitative measure of structural selectivity we use the selectivity coefficient defined by Razinger et al. [19] for a given set of structures. This coefficient C(s), is the number of distinct values, assigned by an index to a family of structures, divided by the number of structures in this family.

Figure 2 illustrates the values of this coefficient calculated for the 21 alkanes of the studied series. As can be appreciated in this figure, the values of *C*(s) are very similar for topological indices based on distance matrix. However, it is well known that the Wiener number is not a very discriminative index, specially for trees having odd number of atoms. In this sense, it is interesting to show that T.I.(2) has a higher structural selectivity than W index. We recall that the statistical quality of the best model found with this index is very similar to that obtained with the Wiener number. We can consider the statistical parameters together with the structural selectivity of the indices as a general indicator of the quality of a descriptor. On this basis we can observe that T.I.(1) and T.I.(2) represent some improvements respect to the original Wiener index.

Figure 2. Structural Selectivity of Topological Indices.



Topological indices based on vertex-adjacency matrix have very different coefficients of selectivity. Zagreb group indices have the lowest structural selectivity among all the studied descriptors. This selectivity is increased up to 1.0 for T.I.(9) which also has the best statistical parameters in the description of boiling points of the studied alkanes. By using the general quality criterion above mentioned, the T.I.(9) index is undoubtedly the best topological index based on vertex-adjacency matrix that has been developed in this work. However, it is necessary to state that the usefulness in QSPR studies of one topological index cannot be evaluated only by its ability to describe one physical property such as boiling point. For instance, M_1 index that has bad correlation with boiling points is a successful descriptor when it is used to correlate with heat of formation of alkanes. All these facts point out that the definition and selection of graph theoretical invariants as molecular descriptors is a complex process that requires the observation of many factors.

In closing, we have defined topological indices that maintain some of the original features of the Wiener number. They are integer numbers that vary regularly with the regular variations in chemical structures. On the other hand, expressions found by us

that relate these graph theoretical invariants to local properties of graphs, such as distance numbers and vertex degrees, permit the generation of novel theoretical descriptors to be used in a chemical context.

Finally, we want to express our acknowledgment to the pioneering work of H. Wiener that opened the doors of a novel branch of research in mathematical chemistry: the definition and study of topological indices, in which this modest contribution pretends to be included.

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] P. J. Hansen and P. C. Jurs, J. Chem. Educ. **65**, 574 (1988).

[4] A. Sabljic and N. Trinajstic, Acta Pharm. Jugosl. **31**, 189 (1981).

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

[6] I. Gutman and O. E. Polansky, Mathematical Concepts in Organic Chemistry, Springer-Verlag, Berlin 1986.

[7] D. H. Rouvray, in: N. Trinajstic (Ed.), Applications of Mathematical Concepts to Chemistry, Ellis Horwood, Chichester 1985, p. 295.

[8] I. Gutman, S.-L. Lee, Y.-N. Yeh, and Y.-L. Luo, Indian J. Chem. **32A**, 651 (1993).

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

[10] S. C. Basak, V. R. Magnuson, G. J. Niemi, and G. D. Veith, Math. Modell. **8**, 300 (1987).

[11] M. Randic, J. Math. Chem. **7**, 155 (1991).

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

[13] P. A. Filip, T.-S. Balaban, and A. T. Balaban, J. Math. Chem. **1**, 61 (1987).

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

[15] A. T. Balaban, Chem. Phys. Lett. **89**, 399 (1982).

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

[17] M. Randic, P. J. Hansen, P. C. Jurs, J. Chem. Inf. Comput. Sci. **28**, 60 (1988).

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

[19] M. Razinger, J. R. Chretien, and J. E. Dubois, J. Chem. Inf. Comput. Sci. **25**, 23 (1985).