7482 *J. Phys. Chem. A* **2003,** *107,* 7482−7489

Generalized Graph Matrix, Graph Geometry, Quantum Chemistry, and Optimal Description of Physicochemical Properties

Ernesto Estrada*

Safety & Environmental Assurance Centre (SEAC), Unilever, Colworth House,
Sharnbrook, Bedford MK44 1LQ, U.K.

Received: March 14, 2003; In Final Form: June 19, 2003

The generalized graph matrix $\mathbf{\Gamma}(x,v)$ (Estrada, E. *Chem. Phys. Lett.* **2001**, *336*, 247) is shown to encompass several of the applications of graph theory in physical chemistry in a more compact and effective way. It defines several n-Euclidean graph metrics, which simulate a graph defolding by changing the exponent v from 0 to 0.5 in a continuous way. This matrix is included in the formalism of the Hückel molecular orbital approach by considering that the resonance integrals between nonneighbor atoms are a function of the topological distance in terms of β. In doing so, the isospectrality between graphs disappears by changing the x parameter in this matrix as a consequence of considering the interactions between nonneighbor atoms. The $\mathbf{\Gamma}(x,v)$ matrix permits several of the "classical" topological indices to be (re)defined using only one graph invariant. These indices include the connectivity index, Balaban J index, Zagreb indices, Wiener index, and Harary indices, which are represented in an 8-dimensional space of parameters to show their similarities and differences. The indices can be optimized to describe physicochemical properties by changing in a systematic way the parameters of the generalized graph matrix and vectors. We show here how a dramatic improvement is obtained by optimizing the Wiener index for describing octane boiling points (from $R = 0.53$ to $R = 0.94$), also providing a structural interpretation of the model found.

Introduction

There are several sources of connection between chemical applications of graph theory and physical chemistry.[1] For instance, the Hückel molecular orbital (HMO) approach and the eigenvalue problem of the adjacency matrix is a well-documented chapter of these interconnections.[2] The graphical rules introduced by Sinanoğlu on the basis of a quantum chemical molecular orbital approach is another example of this interrelation.[3−5] On the other hand, there are several molecular descriptors based on a graph theoretic formulation of the molecular structure, the so-called topological indices, which show great utility in describing physicochemical properties of chemical compounds.[6] However, because of the lack of a general theory that accounts for the whole use of graph theory in physical chemistry, all these applications of graph theory to chemistry appear in disconnected ways. One of the undesired effects of this disconnection is the proliferation of descriptors and approaches using graphs to represent the molecular structure. Randić tried to introduce a sort of rationality in the development of topological indices (TIs) more than 10 years ago, giving some attributes that the indices should have.[7] However, up to now attempts to introduce "new" TIs, which in most cases do not fulfill these attributes, have not stopped.

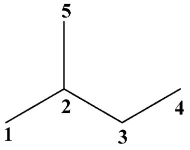
At the moment some researchers have adopted another direction for searching better quantitative structure−property and −activity relationships (QSPR and QSAR). Instead of producing new indices in an indiscriminate way, they prefer to optimize the existing ones for developing better models than with the original indices. The philosophy behind this strategy could be resumed in the fact that all TIs described today in the literature were introduced in an ad hoc way. That is to say, they are not necessarily optimal for describing a particular property. For instance, the connectivity index was introduced by Randić as an attempt to account for molecular branching, and it was shown that it correlated in a satisfactory way with the boiling point of alkanes.[8] However, several researchers have produced better results for this and other properties simply by optimizing the exponent in the Randić invariant.[9−11] By this means, Pogliani has introduced several series of semiempirical connectivity indices that describe in a much more effective way several physicochemical properties of chemical compounds.[12−18] Randić has also introduced variable molecular descriptors as a way to optimize their capacity of describing properties.[19−23] In this line Estrada introduced the concept of generalized TIs, in which a rational way to optimize these descriptors is provided.[24−26]

The main objective of the current work is to demonstrate that there are several applications of graph theory to physical chemistry which are interconnected and are not isolated islands in the middle of the ocean. Thus, we show that it is possible to build a general approach for the study of graph theoretic invariants and for their optimization to describe physicochemical properties. We also show that this approach gives an understanding of the failure of the HMO approach in describing certain classes of molecules, such as the so-called isospectral ones. Finally, but not least, we also demonstrate that under the umbrella of this general approach a graph geometry exists. All these new findings and generalizations can be useful ingredients for the future building of a theory that connects graph theory and physical chemistry with more basic and fundamental principles.

---

* E-mail: estrada66@yahoo.com.

10.1021/jp0346561 CCC: $25.00



$$\mathbf{\Gamma}(x,p)=\begin{bmatrix} 0 & 1 & (2x)^v & (3x^2)^v & (2x)^v \\ 1 & 0 & 1 & (2x)^v & 1 \\ (2x)^v & 1 & 0 & 1 & (2x)^v \\ (3x^2)^v & (2x)^v & 1 & 0 & (3x^2)^v \\ (2x)^v & 1 & (2x)^v & (3x^2)^v & 0 \end{bmatrix}$$

$$\mathbf{P}_1=\mathbf{A}=\begin{bmatrix} 0&1&0&0&0\\ 1&0&1&0&1\\ 0&1&0&1&0\\ 0&0&1&0&0\\ 0&1&0&0&0 \end{bmatrix}\quad \mathbf{P}_2=\begin{bmatrix} 0&0&1&0&0\\ 0&0&0&1&0\\ 1&0&0&0&1\\ 0&1&0&0&0\\ 1&0&1&0&0 \end{bmatrix}\quad \mathbf{P}_3=\begin{bmatrix} 0&0&0&1&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 1&0&0&0&1\\ 0&0&0&1&0 \end{bmatrix}$$

$$\mathbf{\Gamma}(x,v)=\mathbf{P}_1+(2x)^v\mathbf{P}_2+(3x^2)^v\mathbf{P}_3$$

Figure 1. Molecular graphs representing 2-methylbutane and its generalized graph and shortest path matrixes.

Theoretical Approach

In this section we introduce some of the concepts and mathematical formalisms that we use in the other sections of this work. We start by introducing a generalized graph matrix and some of its mathematical properties.

Definition 1: Let $\mathbf{\Gamma}(x,v) = [g_{ij}(x,v)]_{\mathrm{nxn}}$ be the generalized graph matrix, which is defined as a square symmetric matrix with elements g_{ij}:[24,25]

$$g_{ij}=\begin{cases}1 & \text{if } d_{ij}=1\\ (d_{ij}x^{d_{ij}}-1)^v & \text{if } i\neq j \text{ and } d_{ij}\neq 1\\ 0 & \text{otherwise}\end{cases}$$

Definition 2: Let $\mathbf{P}_k = [p_{ij}(k)]_{\mathrm{nxn}}$ be the shortest path matrix of order k whose elements p_{ij} are 1 if and only if there exists a shortest path of order k between vertexes i and j in the graph. It is straightforward to realize that the adjacency matrix **A** of the graph is equal to $\mathbf{P}_1$.

Definition 3: The generalized graph matrix $\mathbf{\Gamma}(x,v)$ can be expressed as a sum of shortest path matrixes of different orders in the following way:

$$\mathbf{\Gamma}(x,v)=\sum_{i=1}(ix^{i-1})^v\mathbf{P}_i$$

An example of the generalized graph matrix as well as the shortest path matrixes for 2-methylbutane are given in Figure 1.

According to the definitions given it is straightforward to realize that several of the main matrixes used in chemical graph theory today are particular cases of the generalized matrix. For instance, the adjacency matrix **A**, the distance matrix **D**, and the Harary matrixes **H1** and **H2** are obtained as particular cases of $\mathbf{\Gamma}(x,v)$. We recall that the Harary matrixes are the matrixes with inverse and squared inverse distances between vertexes in the graph.

$$\mathbf{\Gamma}(0,1)=\mathbf{A},\qquad \mathbf{\Gamma}(1,1)=\mathbf{D}$$

$$\mathbf{\Gamma}(1,-1)=\mathbf{H1},\qquad \mathbf{\Gamma}(1,-2)=\mathbf{H2}$$

One of the consequences of this generalization is the fact that there is an infinite number of graph theoretic matrixes from which the before-mentioned ones are only particular cases.

Generalized Matrix Metrics

A metric on a graph is defined as a function from the Cartesian product $V \times V$ to the nonnegative real numbers such that for any $i, j, k \in V$[27]

$$\rho(i,j)=0 \Leftrightarrow i=j$$

$$\rho(i,j)=\rho(j,i)\geq 0$$

$$\rho(i,j)+\rho(j,k)\geq\rho(i,k)$$

The last one is known as the triangle inequality. From the graph metrics analyzed by Klein the shortest path metric and the square-rooted shortest path metric are straightforwardly defined under the current approach.[28−31] The shortest path metric is that defined by the "classical" topological distance matrix $\mathbf{\Gamma}(1,1) = \mathbf{D}$.[32] This is a particular class of a more general metric existing for graphs. If we consider the general case $\mathbf{\Gamma}(1,v)$, we can investigate which of these matrixes represent metrics for graphs.[32] Using the Blumenthal theorem for the construction of metrics, we can obtain the following generalization.

Theorem (Blumenthal):[27] Given a metric ρ, the functions ρ_v with $\rho_v(i,j) \equiv [\rho(i,j)]^v$ are n-Euclidean metrics for all $n \geq 0$ if $0 \leq v \leq 0.5$.

We understand n-Euclidean to mean a function ρ defined on pairs of distinct vertexes for which every subset of $n + 1$ elements of the vertex set V is isometrically embeddable into n-dimensional Euclidean space ϵ_n. The functions ρ are considered from the Cartesian product $V \times V$ to the real numbers. Then ρ establishes an isometry on a subset of $n + 1$ vertexes if and only if these $n + 1$ vertexes can be embedded in n-dimensional Euclidean space ϵ_n, such that the ρ values for pairs of vertexes of this subset are identical to the corresponding Euclidean distances in ϵ_n.[28]

The square-rooted metric is a particular case of the following metrics: $\mathbf{\Gamma}(1,0.5) = \mathbf{D}_{1/2}$, as analyzed by Klein in several works.[28−31] In this matrix the elements are the square roots of the topological distances between the vertexes in the graph. However, all the generalized matrixes $\mathbf{\Gamma}(1,v)$ with $0 \leq v \leq 0.5$ involve n-Euclidean metrics for all $n \geq 0$.

Several graph geometric invariants were proposed by Zhu and Klein for characterizing the metric properties of graphs.[29] We use here two of these invariants to illustrate the metric properties of certain graphs as deduced from the generalized matrix of graphs $\mathbf{\Gamma}(1,v)$ with $0 \leq v \leq 0.5$. These two graph geometric invariants are the *linear curvature* and the *curvilinear torsion*.

The net linear curvature k_i is associated with each vertex of the graph. It is the sum of the angles θ formed by all combinations of three successive vertexes i, j, and k. The angle θ is obtained as[29]

$$\cos(\pi-\theta)=({\rho_{ik}}^2-{\rho_{ij}}^2-{\rho_{jk}}^2)/2\rho_{ij}\rho_{jk}$$

in which i, j, and k are the vertexes of a 3-path in the graph. The curvature sum is simply the sum of all net curvatures in the graph.

The curvilinear torsion is defined for the i, j, k, and l vertexes in a 3-path in the following way:[29]

$$\cos(\phi)=\left|\frac{\Delta(ikj)\,\Delta(ljk)+2{\rho_{jk}}^2[\Delta(ikj)-\Delta(ilj)]}{16A(ijk)\,A(jkl)}\right|$$

TABLE 1: Linear Curvature Sums Divided by 2π for Path Graphs in the Generalized Matrix Approach at Different v Values ($x = 1$)

v	L4	L5	L6	L7	L8
0	0.3333	0.5000	0.6667	0.8333	1.0000
0.1	0.3601	0.5401	0.7201	0.9001	1.0802
0.2	0.3895	0.5843	0.7790	0.9738	1.1685
0.3	0.4222	0.6333	0.8443	1.0554	1.2665
0.4	0.4587	0.6880	0.9173	1.1467	1.3760
0.5	0.5000	0.7500	1.0000	1.2500	1.5000

TABLE 2: Net Torsion Sums Divided by 2π for Path Graphs in the Generalized Matrix Approach at Different v values ($x = 1$)

v	L4	L5	L6	L7	L8
0	0.1959	0.3918	0.5878	0.7837	0.9796
0.1	0.2093	0.4186	0.6279	0.8372	1.0465
0.2	0.2212	0.4423	0.6635	0.8847	1.1058
0.3	0.2318	0.4635	0.6953	0.9270	1.1588
0.4	0.2413	0.4826	0.7238	0.9651	1.2064
0.5	0.2500	0.5000	0.7500	1.0000	1.2500

where the "Pythagorean defect" $\Delta(abc)$ and the area of a triangle $A(abc)$ are defined as

$$\Delta(abc) = {\rho_{ab}}^2 - {\rho_{ac}}^2 - {\rho_{bc}}^2$$

$$A(abc) = [s(s - \rho_{ab})(s - \rho_{bc})(s - \rho_{ca.})]^{1/2}$$

where $s = (\rho_{ab} + \rho_{bc} + \rho_{ac})/2$. The net torsion sum is the sum of net torsions at each edge of the graph.

In Tables 1 and 2 we give the values of the net linear curvature and the curvilinear torsion sums, respectively. As can be seen in these tables the net linear curvature and the curvilinear torsion sums increase their values as the value of v increases.

In the particular case of the 3-path (L3) these metrics are represented by embedding this graph in a 3-Euclidean space, as in Figure 2. This path was defined by Balaban and Rücker as a protochiron for the 3-dimensional coding of certain structures.[33] In this way, the linear curvature represents the net angles between the vertexes in the graph and the curvilinear torsion represents the total number of twisting as given by the dihedral angles. As can be observed the net effect of increasing the value of v is the "defolding" of the chain from a more compact structure to one in which vertexes are more separated in space. This "graph expansion" follows a linear dependence for the value of v as illustrated in Figure 3.

Generalized Matrix and Quantum Chemistry

The relationship between the adjacency matrix of graphs representing conjugated molecules and the HMO method is well-known and documented in the literature.[2] In the HMO approximation all overlap integrals are set equal to 0, all resonance integrals between nonneighbors are set equal to 0, and the remaining resonance integrals are set equal to β.[34]

In the current approach, we consider again that all overlap integrals are set equal to 0 but that the resonance integrals are a function of the topological distance separating atoms i and j:

$$\beta_{ij} = (d_{ij} x^{dij})^{v} \beta$$

The idea of not neglecting the off-diagonal overlaps and Hamiltonian matrix elements in the HMO secular determinant is very old and can be resumed in Streitweiser's book.[35] The current approach based on the generalized graph matrix also has some resemblance to the extended Hückel theory (EHT) introduced by Hoffmann,[36] where the nondiagonal elements of the Hamiltonian are refined to be an explicit function of the molecular geometry through the dependence of the overlap matrix on the internuclear distances.[37] However, the main value of the current approach is that it is defined under the umbrella of a graph theoretical approach that encompasses aspects of molecular geometry, quantum chemistry, and a description of physicochemical properties through structure−property relationships.

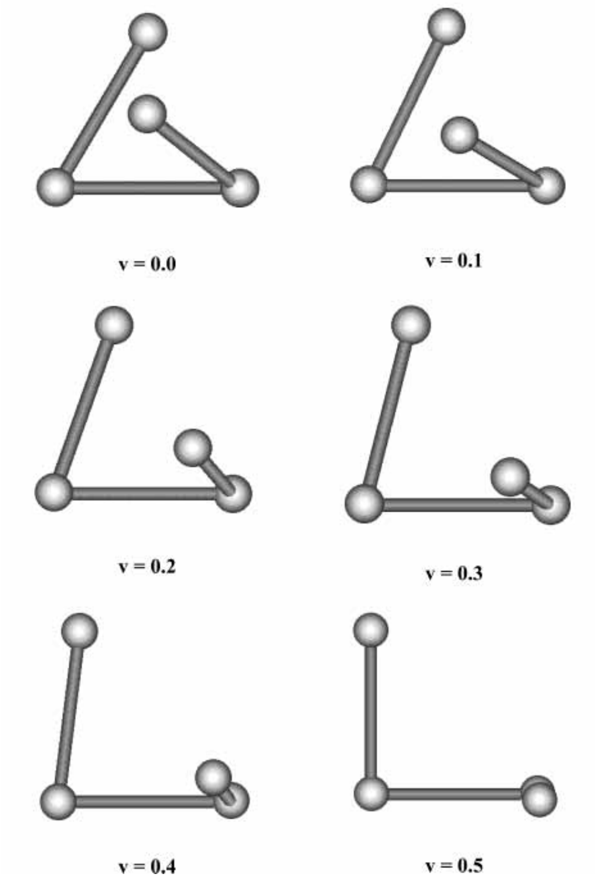


Figure 2. Embedding of the 3-path graph into a 3-Euclidean space and simulation of the graph expansion produced by the continuous change of the v parameter.

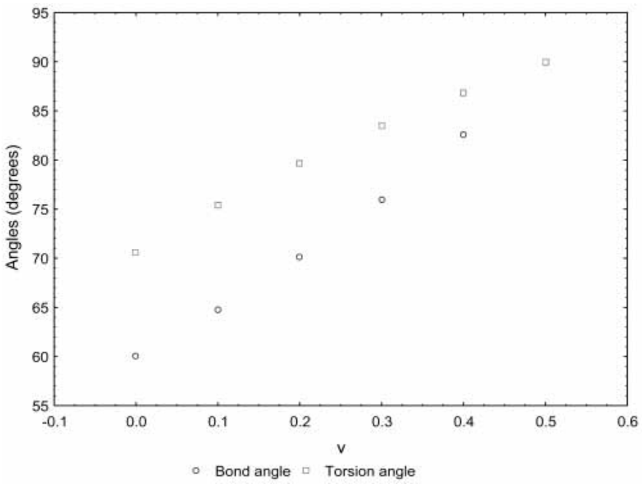


Figure 3. Linear relation between the v exponent and the graph expansion as measured by the "bond angle" and "torsion angle" in an embedding of the 3-path in a 3-Euclidean space.

In the case of adjacent atoms the resonance integral is equal to β as in the original HMO approach, and in the particular

TABLE 3: Experimental Ionization Energies (I_j^m, eV) and Eigenvalues of the Generalized Graph Matrix with Different Values of the x Parameter ($v = 1$)

I_j^m	$x = 0$	$x = 0.05$	$x = 0.1$	$x = 0.15$	$x = 0.2$
		1,4-Divinylbenzene			
8.11	0.539	0.384	0.218	0.042	-0.140
9.18	1.000	0.892	0.770	0.632	0.480
9.80	1.000	0.930	0.840	0.727	0.589
11.00	1.675	1.724	1.765	1.794	1.807
	2.214	2.486	2.790	3.131	3.514
		2-Phenylbutadiene			
8.60	0.539	0.409	0.269	0.118	−0.041
9.28	1.000	0.891	0.766	0.622	0.462
9.50	1.000	0.893	0.770	0.632	0.480
11.54	1.675	1.743	1.799	1.838	1.853
	2.214	2.481	2.783	3.123	3.507
R^{2a}	0.945	0.958	0.963	0.964	0.962

[a] Correlation coefficient between the experimental ionization energies and the eigenvalues of the generalized matrix. The case $x = 0$ corresponds to the simple HMO approach.

case of $x = 0$ the secular determinant is exactly the same as for HMO. To keep the physicality of the approach, we will think about only those values of x (v will be kept equal to 1 for the sake of simplicity) which make the interactions between nonbonded atoms weaker than those between bonded atoms. This means that we select the values of x in the range $0 \leq x < 0.5$. The upper limit is based on the fact that when $x = 0.5$ the resonance integrals β_{ij} for the atoms separated at distance 2 take exactly the same value as for the atoms at distance 1.

One interesting application of the current approach is concerned with the so-called isospectral graphs.[38−43] These graphs are nonisomorphic graphs having exactly the same spectra for their adjacency matrixes. A well-known example of a pair of isospectral graphs is that for 1,4-divinylbenzene and 2-phenylbutadiene. These two molecules are conjugated hydrocarbons, and it is possible to use the HMO approach for their study. In 1978 Heilbronner and Jones[44] determined the ionization energies of this pair of molecules using photoelectron spectra, finding that they “differ at least as much as those of any other ‘nonisospectral’ pair having π systems of comparable size”. Heilbronner and Jones also showed that MO models at the STO-3G level were able to differentiate correctly the ionization energies of these two molecules.[44] On this basis they concluded “that the graph theoretical statement that two molecules are ‘isospectral’ is not of relevance for their physical and chemical behavior”.[44] However, the point here is the following: If the HMO reproduces reasonably well the electronic features of conjugated systems, why does it fail in explaining the differences between “isospectral” graphs?

To investigate this question, we carried out calculations on the framework of the generalized graph matrix for these two molecules by changing the value of x as shown in Table 3. As can be seen in this table not only does the change of the x parameter from 0 to 0.05 produce a differentiation of the spectra of both molecules but this change is in the correct direction. According to the experimental values shown as well in this table, 2-phenylbutadiene has a higher HOMO energy than 1,4-divinylbenzene, which is well reproduced by the generalized matrix spectra with $x > 0$. We also provide in this table the correlation coefficient between the experimental values of the ionization energies and the eigenvalues of the generalized matrix. As can be seen when $x = 0.15$ is used as a parameter a correlation coefficient of $R^2 = 0.964$ is obtained. It is still not as good as that obtained by MO STO-3G calculations reported by Heilbronner and Jones ($R^2 = 0.982$)[44] but represents a significant improvement compared to the HMO results in both a qualitative and a quantitative way ($R^2 = 0.945$).

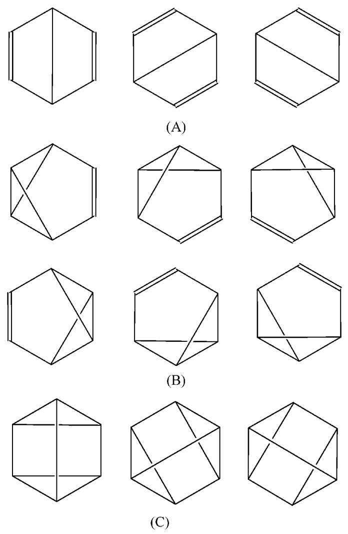
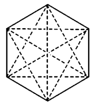


Figure 4. Valence isomers of benzene showing the different topologically equivalent structures.

The interpretation of the results obtained by using the generalized matrix is based on the fact that we are considering not only the resonance integrals between the bonded atoms but also those between nonneighbors. Consequently, if we consider the simple case of the C_6 graph representing the molecule of benzene, what we are analyzing in reality is a graph with “strong” connections between bonded atoms and some “weak” connections between the nonbonded ones as illustrated below:



If $x = 0.1$, for instance, the strong connections take values of 1 and the weak connections take values of 0.2 and 0.03 for atoms at distances 2 and 3, respectively. In this particular case of benzene we can think that the contributions for nonneighbor atoms come from the contributions made by the valence isomers of benzene,[45−47] such as Dewar structures, benzvalene structures, and prismane structures. For instance, there are three topologically equivalent Dewar benzene structures in which connections between atoms at distance 3 are considered. In benzvalene each structure has a couple of connections between atoms separated at distance 2, which makes six topologically equivalent benzvalene structures. Finally, there are three topologically equivalent prismane structures in which connections between atoms separated at distances 2 and 3 are considered. These structures are illustrated in Figure 4. Using these structures as well as the two Kekule structures of benzene, we can calculate the Pauling bond orders not only for the bonded atoms but for the nonneighbor atoms as well, giving the values of 1.214, 0.214, and 0.143 for atoms at distances 1, 2, and 3, respectively. This is a result similar to that obtained here when $x \neq 0$, which means that the interactions between nonneighbor atoms make a contribution to the ionization energies of these two molecules.

Generalized Topological Indices

As commented on in the Introduction, one of the current main problems in the research of topological indices is an indiscriminate proliferation of such molecular descriptors. To avoid the continuation of this practice, we introduce a strategy that permits optimization of several of the well-known topological indices. This optimization is carried out to predict in a more effective way the physicochemical or biological properties of chemicals under study as well as to facilitate the interpretation of these descriptors in a physicochemical context.

For the current purpose of generalizing TIs we introduce a change of variable in $\mathbf{\Gamma}(x,\upsilon)$ so that x is substituted by y or z and $\upsilon = 1$. Thus, we obtain the following matrixes: $\mathbf{\Gamma}(y,1)$ and $\mathbf{\Gamma}(z,1)$. Using these matrixes, two new graph theoretical vectors used in the generalization of TIs are defined as follows.[24-26]

Definition 4: Let $\mathbf{y}(w,y,q)$ and $\mathbf{z}(s,z,r)$ be two vectors of order n whose elements y_i and z_i are defined as follows:

$$y_i = (w_i + \sum_j g_{ij}(y,1))^q \qquad z_i = (s_i + \sum_j g_{ij}(z,1))^r$$

where w_i and s_i are weights to be assigned to the corresponding vertex to differentiate heteroatoms or chiral centers.

The generalized vectors of the 2-methylbutane molecule, which was previously illustrated in Figure 1, are

$$\mathbf{y}(w,y,q) = [(w_1 + 4y + 3y^2)^q \;\; (w_2 + 3 + 2y)^q \;\; (w_3 + 2 + 4y)^q \;\; (w_4 + 1 + 2y + 6y^2)^q \;\; (w_5 + 1 + 4y + 3y^2)^q]$$

$$\mathbf{z}(s,z,r) = [(s_1 + 4z + 3z^2)^r \;\; (s_2 + 3 + 2z)^r \;\; (s_3 + 2 + 4z)^r \;\; (s_4 + 1 + 2z + 6z^2)^r \;\; (s_5 + 1 + 4z + 3z^2)^r]$$

The following definition based on a vector−matrix−vector (VMV) multiplication procedure introduced by Estrada[48,49] is used for the generalization of the TIs.

Definition 5: Let $\mathcal{T}$ be a generalized vector−matrix−vector invariant defined as[24]

$$\mathcal{T} = C[\mathbf{y}(y,w,q)\;\mathbf{\Gamma}(x,\upsilon)\;\mathbf{z}(z,s,r)]$$

By using this expression, we can obtain several adjacency-based and distance-based TIs using only one graph theoretic invariant. The mathematical expressions for these classical indices are

$$W = {}^1/_2[\mathbf{y}(1,0,0)\;\mathbf{\Gamma}(1,1)\;\mathbf{z}(1,0,0)]$$

$$J = {}^1/_2[\mathbf{y}(1,0,-0.5)\;\mathbf{\Gamma}(0,1)\;\mathbf{z}(1,0,-0.5)]$$

$$H_1 = {}^1/_2[\mathbf{y}(1,0,0)\;\mathbf{\Gamma}(1,-1)\;\mathbf{z}(1,0,0)]$$

$$H_2 = {}^1/_2[\mathbf{y}(1,0,0)\;\mathbf{\Gamma}(1,-2)\;\mathbf{z}(1,0,0)]$$

$$M_1 = [\mathbf{y}(0,0,1)\;\mathbf{\Gamma}(0,1)\;\mathbf{z}(0,0,0)]$$

$$M_2 = {}^1/_2[\mathbf{y}(0,0,1)\;\mathbf{\Gamma}(0,1)\;\mathbf{z}(0,0,1)]$$

$$\chi = {}^1/_2[\mathbf{y}(0,0,-0.5)\;\mathbf{\Gamma}(0,1)\;\mathbf{z}(0,0,-0.5)]$$

$$\chi^{\mathrm{V}} = {}^1/_2[\mathbf{y}(0,N,-0.5)\;\mathbf{\Gamma}(0,1)\;\mathbf{z}(N,0,-0.5)]$$

where N in the last expression stands for the count of electrons in π bonds and in lone-pair orbitals. Here W is the Wiener index,[50] J is the Balaban index,[51] H_1 and H_2 are the so-called Harary indices,[52,53] M_1 and M_2 are the Zagreb indices,[54] χ is the connectivity index,[8] and χ^{V} is the valence connectivity index.[55]

This result has three important consequences. The first is that there are not so many graph theoretic invariants as we think, many being the same invariant with different weights. The second consequence is that from a practical point of view we can optimize these topological indices, previously introduced in an ad hoc way, to describe the properties we are interested in. Both consequences can be a guide for stopping the indiscriminate generation of new topological indices. Instead of trying to generate a new index, one can try to optimize the existing ones. If you do not succeed with the optimization, then you can try introducing a new one. But take into account that by changing these eight parameters at the same time in a continuous way you will generate an infinite number of choices for describing the property. Finally, we can seek a physicochemical interpretation of the TIs in a global way. The identification of some physicochemical basis for the connectivity index[56] can give us some clues about the way in which this interpretation can be carried out.

Representation of Topological Indices

As we can see, the first consequence of the generalized vector−matrix−vector multiplication procedure is that all TIs studied here are points in an 8-dimensional space. This is a completely new picture in the study of TIs, because most of these descriptors have appeared as disconnected from each other, eluding generalizations and interpretations in a united way and complicating the computer programs to calculate them by using several different algorithms. Now, we have only one algorithm depending on eight different parameters and can generate a series of the most important TIs described in the literature. This fact also will permit a better interpretation of these descriptors than if we attempt to make this interpretation of the indices one by one.

A graphical representation of the TIs studied here with radial graphics is given in Figure 5. These graphics are built by using eight axes, which correspond to the eight parameters in the current approach, y, w, q, x, υ, z, s, and r, and an octagonal figure, formed by joining these points representing the topological indices.

The main utility of these graphics is to orient the optimization process of the TIs in a rational way. Suppose one knows a priori that the Balaban index J describes property P in a better way than the Wiener W index. Then, by analyzing the differences between the diagrams for these two indices, one can select as a first choice the optimization of q, x, and r, which make the differences between both indices, instead of optimizing the eight parameters.

Optimization of Topological Indices

The optimization of an index according to our approach of the generalized vector−matrix−vector multiplication procedure consists simply in changing systematically the values of the eight parameters on which any of the studied indices depend. This procedure permits the definition of an infinite number of TIs, the known ones of which are particular cases. However, if we consider changing systematically all eight parameters of the VMV multiplication procedure, we can generate a huge number of descriptors, creating a big problem for their computation, storage, and manipulation. For instance, if we consider simultaneous variations of the eight parameters from 0 to 0.9 with a step of 0.1, we generate 100000000 indices.

As a working example we investigate the optimization of the Wiener W index for describing the boiling points of octanes. For the sake of simplicity we optimize only the parameters x

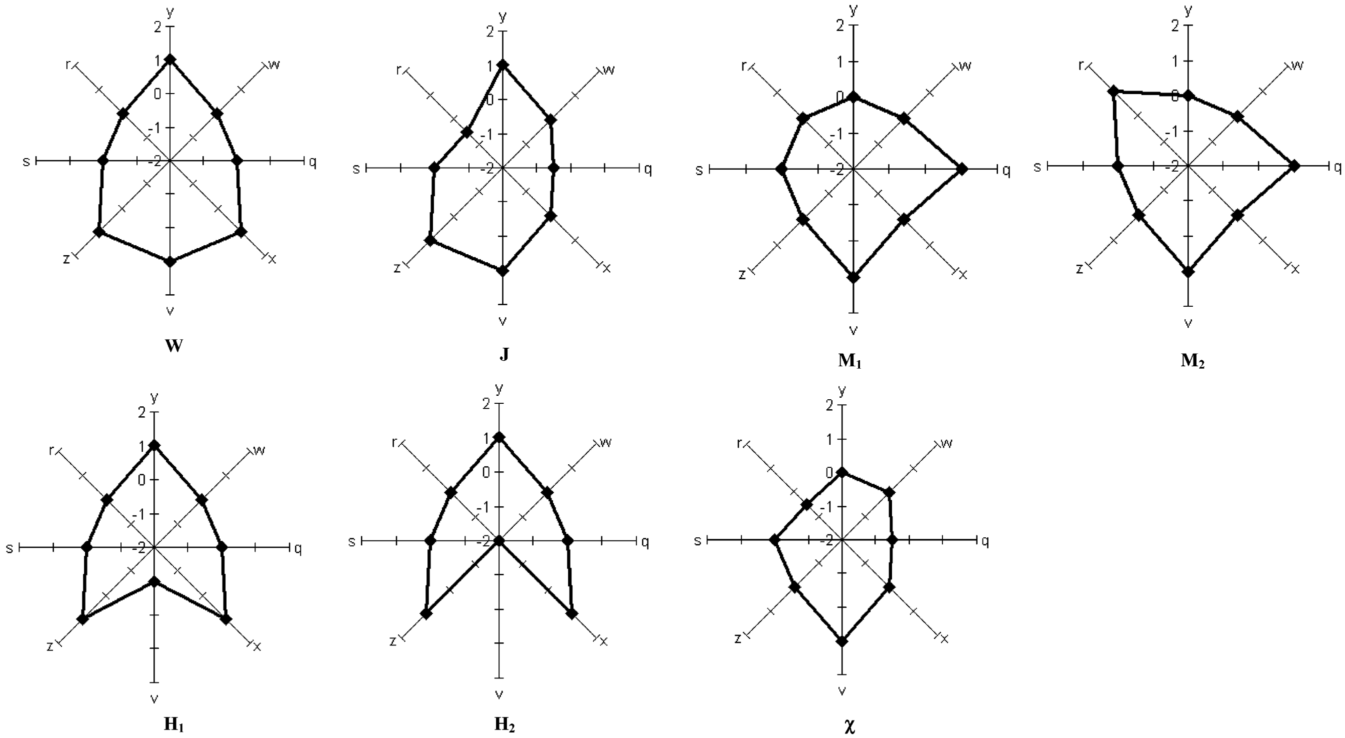


Figure 5. Representation of the topological indices in 8-dimensional spaces of parameters arising from the generalized graph matrix and vectors. The axes start at the value of −2.

and v, reducing the problem to the generation of the optimal index of the form

$$W = {}^{1}/_{2}[\mathbf{y}(1,0,0)\ \boldsymbol{\Gamma}(x,v)\ \mathbf{z}(1,0,0)]$$

The optimization procedure consists of the following steps:

(a) select a range for the x and v parameters ($x_{\min} \le x \le x_{\max}$; $v_{\min} \le v \le v_{\max}$),

(b) select steps for the change of the x and v parameters, e.g., step 1 and step 2, respectively,

(c) calculate the indices $W(x,v)$ for the data set of compounds (e.g., octanes) according to the VMV multiplication procedure,

(d) find linear regression models for each of the $W(x,v)$ indices and the property to be studied (e.g., boiling point),

(e) select the values of the x and v parameters of the best model,

(f) select a new range for x and v closer to the values obtained in (e) as well as new steps,

(g) turn to (c) until the optimal values of x and v are obtained.

We first explored the region around the original Wiener index using the following x and v values: $1 \le x \le 5$ and $1 \le v \le 5$ with step 1 = step 2 = 1. The correlation coefficients in this region were not significantly better than that obtained for the Wiener index as illustrated in Figure 6A. Then, we moved to the following parameter regions: $-5 \le x \le -1$ and $-5 \le v \le -1$ with step 1 = step 2 = 1. The results of this second optimization are given in a graphical form in Figure 6B.

The original W index has a correlation coefficient of only $R = 0.539$ with the boiling point of octanes. It is shown at the extreme right corner of Figure 6A, i.e., $x = 1$ and $v = 1$. The best result obtained with the current gross optimization is for the case $x = -2$ and $v = -1$ for which $R = 0.901$. This index, $W(-2,-1)$ explains more than 81% of the variance in the boiling points of octanes, which represents an improvement of 52% over the original Wiener W index.

A “finer” optimization of this index was then conducted by changing systematically the value of x from −3 to −1 and keeping constant the value of $v = -1$. The results obtained in this optimization step are illustrated in Figure 7, where we plot the correlation coefficient versus the value of x.

The optimal Wiener index obtained by this process is $W(-1.4,-1)$, which gives a correlation coefficient of 0.946. The linear regression model obtained with this index is given below together with the statistical parameters of the model:

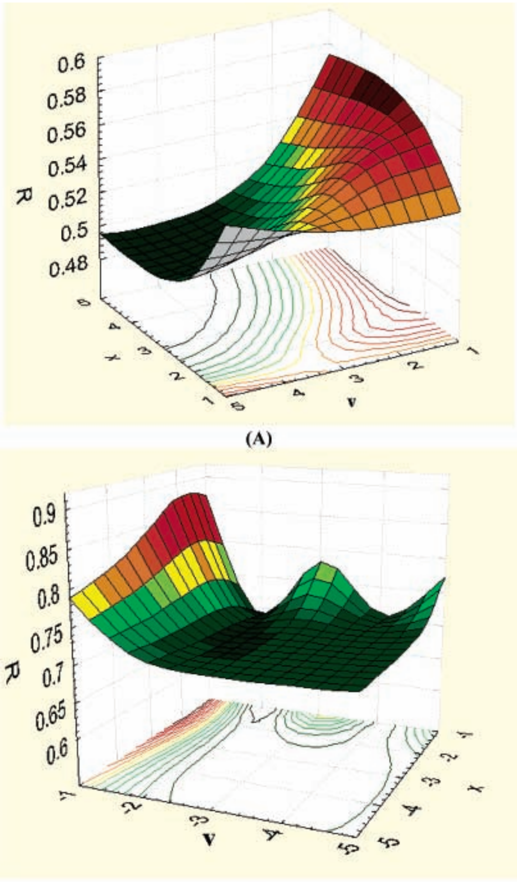
$$\text{bp }(^{\circ}\text{C}) = 48.0023 + 13.5271W(-1.4,-1)$$

$$n = 18,\quad R = 0.9464,\quad s = 2.04,\quad F = 137.5$$

This model represents an improvement of 14.3% in the standard deviation over the previous best model with a Wiener-like index obtained by Ivanciuc et al.[57] with odd/even Wiener indices. It also represents an improvement of 7.7% with respect to the long-range connectivity index developed by Estrada[58] that was the best previous model to describe boiling points of octanes with one topological index. In fact, the correlation obtained here is the best QSPR model for describing boiling points of octanes with one TI. However, we have to say that this index is not necessarily the best of all Wiener indices $W(x,v)$ that can be obtained for describing the boiling points of octanes because we have explored only a small portion of the parameter space for optimizing the W index.

In terms of the interpretation of the model developed using the generalized Wiener index for the boiling point of octane, we have to consider first the expression of $W(x,v)$ in terms of the number of vertexes separated at the same distance. This expression is given below for the general values of x and v in terms of the number η_i of vertexes separated at distance i:

$$W(x,v) = \sum_{i=1} (ix^{i-1})^{v} \eta_i$$



(A)

(B)

Figure 6. Optimization of the Wiener index by changing the parameters x and υ in the ranges $1 \leq x \leq 5$ and $1 \leq \upsilon \leq 5$ (A) as well as $-5 \leq x \leq -1$ and $-5 \leq \upsilon \leq -1$ (B) for improving the correlation coefficient for the boiling points of octanes.

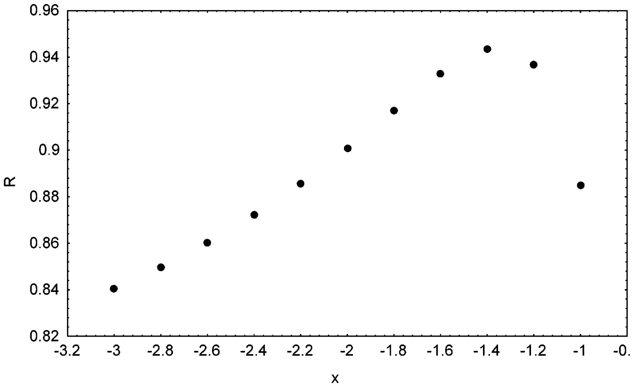


Figure 7. "Finest" optimization of the Wiener index by changing the values of the parameter x.

This means that the particular case of $W(-1.4,-1)$ is given by the following expression:

$$W(-1.4,-1) = \eta_1 - 0.357\eta_2 + 0.170\eta_3 - 0.091\eta_4 + ... + [i(-1.4)^{i-1}]^{-1}$$

By substituting this expression into the QSPR model for the boiling point of octanes, we obtain an additive contribution of the paths of different lengths in the octanes to their boiling points:

$$\text{bp } (^\circ\text{C}) = 48.00 + 13.527\eta_1 - 4.829\eta_2 + 2.34\eta_3 - 1.23\eta_4 + ...$$

This means that the main contribution is carried out by the number of bonds, η_1 (13.527 °C). Then, a negative contribution comes from fragments formed by two bonds. These fragments increase dramatically with branching in the molecule. For instance, in a path of order 4 (*n*-butyl group) we have three fragments of this type, but in the star of five vertexes (a *tert*-butyl group) it is increased to 6. It is well-known that branching decreases the boiling point due to a smaller number of intermolecular contacts in the liquid as predicted by our model that assigns a negative coefficient to these fragments. With the increase of the size of the fragments their contributions diminish, representing only minor corrections to the boiling point of octanes in the current approach. The idea of assigning decreasing contributions according to increasing distance in TIs appears to have been first introduced by Diudea and Balaban 10 years ago.[59,60] Other similar long-range contributions to connectivity indices were also reviewed recently by Estrada. [61]

Conclusions

The conclusions of this work are the following:

(1) The generalized graph matrix encompasses in a more compact way the structural information contained in several traditional graph theoretic matrixes.

(2) The generalized graph matrix for $x = 1$ and $0 \leq \upsilon \leq 0.5$ defines several graph metrics, which are n-Euclidean for all $n \geq 0$. The net effect of changing the exponent υ from 0 to 0.5 in a continuous way represents a graph expansion or defolding from a more compact "conformation" to a more expanded one.

(3) The generalized graph matrix can be included in the formalism of the HMO approach by considering that the resonance integrals between nonneighbor atoms are a function of the topological distance in terms of β.

(4) The isospectrality between graphs is produced by the neglect of interactions between nonneighbor atoms in the HMO approach, and it disappears as soon as we change the parameter x in the generalized graph matrix.

(5) There are several TIs that can be obtained using the same graph invariant with the generalized graph matrix. They include the connectivity index, Balaban *J* index, Zagreb indices, Wiener index, and Harary indices.

(6) The TIs can be represented in an 8-dimensional space based on the parameters of the generalized graph matrix and vectors. This representation can help in orienting a rational optimization of these indices.

(7) TIs can be optimized to describe physicochemical properties by systematically changing the parameters of the generalized graph matrix and vectors. The optimized indices can be interpreted structurally, giving physical interpretability to the models developed.

(8) An interconnection among several apparently isolated applications of graph theory in physical chemistry has been found by using the generalized graph matrix approach.

References and Notes

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

(2) Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer-Verlag: Berlin, 1987.

(3) Sinanoğlu, O. *Chem Phys. Lett.* **1984**, *103*, 315.

(4) Sinanoğlu, O.; Alia, J.; Hasting, M. *J. Phys. Chem.* **1994**, *98*, 5867.

(5) Shen, M. *Int. J. Quantum Chem.* **1990**, XXXVII, 125.

(6) *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 2001.

(7) Randić, M. *J. Math. Chem*. **1991**, *7*, 155.

(8) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 7, 6609.

(9) Randić, M.; Hansen, P. J.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1988**, *18*, 60.

(10) Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31.

(11) Amić, D.; Bešlo, D.; Lucic, B.; Nikolić, N.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 819.

(12) Pogliani, L. *J. Phys. Chem*. **1993**, *97*, 9029.

(13) Pogliani, L. *J. Phys. Chem*. **1994**, *98*, 1494.

(14) Pogliani, L. *J. Phys. Chem*. **1995**, *99*, 925.

(15) Pogliani, L. *J. Phys. Chem*. **1996**, *100*, 18065.

(16) Pogliani, L. *J. Phys. Chem. A* **1999**, *103*, 1598.

(17) Pogliani, L. *Chem. Rev*. **2000**, *100*, 3827.

(18) Pogliani, L. *J. Phys. Chem. A* **2000**, *104*, 9029.

(19) Randić, M.; Basak, S. C.; Pompe, M.; Novic, M. *Acta Chim. Slov*. **2001**, *48*, 169.

(20) Randić, M.; Pompe, M. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 631.

(21) Randić, M.; Plavsic, D.; Lers, N. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 657.

(22) Randić, M.; Basak, S. C. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614.

(23) Randić, M.; Dobrowolski, J. Cz. *Int. J. Quantum Chem*. **1998**, *70*, 1209.

(24) Estrada, E. *Chem. Phys. Lett*. **2001**, *336*, 247.

(25) Estrada, E. In *Topology in Chemistry*; Rouvray, D. H., King, R. B., Eds.; Horwood: Chichester, U.K., 2002; p 185.

(26) Estrada, E. *MATCH* **2001**, *44*, 155.

(27) Blumenthal, L. M. *Distance Geometry*; Chelsea: New York, 1970.

(28) Klein, D. J. In *Topology in Chemistry*; Rouvray, D. H., King, R. B., Eds.; Horwood: Chichester, U.K., 2002; p 292.

(29) Zhu, H.-Y.; Klein, D. J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1067.

(30) Klein, D. J. *MATCH* **1997**, *35*, 7.

(31) Klein, D. J.; Zhu, H.-Y. *J. Math. Chem*. **1998**, *23*, 179.

(32) Buckley, F.; Harary, F. *Distance in Graphs*; Addison-Wesley: Reading, MA, 1989.

(33) Balaban, A. T.; Rücker, C. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1145.

(34) Rauk, A. *Orbital Interaction Theory of Organic Chemistry*; Wiley: New York, 2001.

(35) Streitweiser, A. *Molecular Orbital Theory*; John-Wiley & Sons: New York, 1961.

(36) Hoffmann, R. *J. Chem. Phys*. **1963**, *39*, 1397.

(37) Calzaferri, G.; Forss, L.; Kamber, I. *J. Phys. Chem*. **1989**, *93*, 5366.

(38) Herndon, W. C.; Ellzey, M. L., Jr. *Tetrahedron* **1975**, *31*, 99.

(39) Zivković, T.; Trinajstić, N.; Randić, M. *Mol. Phys*. **1975**, *30*, 517.

(40) Randić, M.; Trinajstić, N.; Zivković, T. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 244.

(41) Randić, M.; Barysz, M.; Nowakowski, J.; Nikolićc, S.; Trinajstić, N. *J. Mol. Struct*.: *THEOCHEM* **1989**, *54*, 95.

(42) Babić, D. *J. Math. Chem*. **1993**, *12*, 137.

(43) Gutman, I.; Marković, S.; Grbović, V. *J. Serb. Chem. Soc*. **1991**, *56*, 553.

(44) Heilbronner, E.; Jones, J. B. *J. Am. Chem. Soc*. **1978**, *100*, 6506.

(45) Gutman, I.; Potgieter, J. H. *J. Chem. Educ*. **1994**, *71*, 222.

(46) Schulman, J. M.; Disch, R. L. *J. Am. Chem. Soc*. **1985**, *107*, 5059.

(47) Bettinger, H. F.; Schreiner, P. R.; Schaefer, H. F., III; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1998**, *120*, 5741.

(48) Estrada, E.; Rodriguez, L.; Gutierrez, A. *MATCH* **1997**, *35*, 145.

(49) Estrada, E.; Rodriguez, L. *MATCH* **1997**, *35*, 157.

(50) Wiener, H. *J. Am. Chem. Soc*. **1947**, *69*, 17.

(51) Balaban, A. T. *Chem. Phys. Lett*. **1982**, *89*, 399.

(52) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. *J. Math. Chem*. **1993**, *12*, 235.

(53) Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. *J. Math. Chem*. **1993**, *12*, 309.

(54) Gutman, I.; Ruš•ić, B.; Trinajstić, N.; Wilcox, C. F. *J. Chem. Phys*. **1975**, *62*, 3399.

(55) Kier; L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(56) Estrada, E. *J. Phys. Chem. A* **2002**, *106*, 9085.

(57) Ivanciuc, O.; Ivanciuc, T.; Klein, D. J.; Seitz, W. A.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 536.

(58) Estrada, E. *Chem. Phys. Lett*. **1999**, *312*, 556.

(59) Diudea, M. V.; Minailiuc, O.; Balaban, A. T. *J. Comput. Chem*. **1991**, *12*, 527.

(60) Balaban, A. T.; Diudea, M. V. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 421.

(61) Estrada, E. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: Amsterdam, 1999; p 403.