

Edge-Connectivity Indices in QSPR/QSAR Studies. 2. Accounting for Long-Range Bond Contributions

Ernesto Estrada

Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela,
15706 Santiago de Compostela, Spain

Received April 5, 1999

A novel graph theoretical polynomial, $P_e(G, x)$, counting the edge connectivity is introduced. The first derivative of this polynomial evaluated for $x = 0$ is equal to the edge-connectivity index of the molecular graph. A series of edge-connectivity indices modified to include long-range bond contributions, $\epsilon^c(x)$, are obtained by using values of x different from zero. These indices form three independent (orthogonal) groups containing different structural information as proved by the principal component analysis. The use of the $\epsilon^c(x)$ indices in QSPR studies produced significant improvements in the statistical results compared to the original edge-connectivity index, ϵ , for six of the seven properties studied. Some of the novel descriptors are able to describe physicochemical properties poorly correlated to the original ϵ index.

INTRODUCTION

The edge-connectivity index, ϵ , was defined in full analogy with the vertex-connectivity index, χ , by using edge degrees instead of the vertex degrees.¹ The main reason for the selection of the Randic-type formula² in the definition of the edge-connectivity index is the great success in quantitative structure–property/structure–activity relationship (QSPR/QSAR) studies that the vertex-connectivity index has maintained during almost 25 years.^{3,4} From its definition in 1995 the ϵ index has been used in several QSPR/QSAR studies, showing good correlating abilities.^{1,5–12} By this way this index has been claimed as a very promising topological index for QSPR/QSAR studies.¹⁰

In the precedent work of this series we demonstrate that the ϵ index is orthogonal to almost all the other 35 first-, second-, and third-generation topological indices studied for a data set of octane isomers.¹³ On the other hand, the ϵ index was the best single descriptor, from the pool of 36 topological indices, in correlating five of the seven physicochemical properties studied for octane isomers. These properties were the molar volume, molar refraction, critical temperature, critical pressure, and surface tension. The other two properties, boiling point and heat of vaporization, were poorly described by the edge-connectivity index, and the best models were obtained by using the Hosoya's Z number¹⁴ and the Randic's χ index.²

The edge-connectivity index, like the vertex-connectivity one, is a locally defined topological index. This index contains information only about the adjacent bonds in the molecule obtained from the multiplication of edge (bond) degrees of adjacent edges. Here we are interested in investigating if the structural information coming from more distant bonds, i.e., those not adjacent in the molecule, can be of significant importance for the description of physicochemical and biological properties of molecules. To account for these “long-range bond contributions” in the scheme of the edge-connectivity approach, we propose the introduction

of the edge-connectivity counting polynomial. The modification of the edge-connectivity index is carried out by using the first derivative of this polynomial. A similar procedure has been followed for the present author for the vertex-connectivity index.^{15,16}

Here, as in the precedent work of this series, we continue by using the octane data set of compounds for testing the modifications introduced to the edge connectivity index. This data set of compounds has been advocated by Randic and Trinajstić as a very suitable data set for testing topological indices.^{17,18} These kinds of molecular descriptors are designed to have gradual changes with the changes in the molecular weight of compounds. Consequently, it is not a surprise to obtain good correlations with physicochemical properties for data sets of chemicals with great variability in the molecular weight. These correlations are not necessarily observed when the same indices are used in isomeric data sets of compounds, such as the octane data set. On the other hand, most of the physicochemical properties commonly studied in QSPR analysis with topological indices are interrelated for data sets of compounds with different molecular weight, for instance for alkanes with two to nine carbon atoms.¹⁷ This interrelatedness is hardly observed when the data set of octane isomers is analyzed.

THEORETICAL APPROACHES

The edge-connectivity index $\epsilon = \epsilon(G)$ of a (molecular) graph G is defined by using the edge degrees $\delta(e_i)$ instead of the vertex degrees in the Randic-type formula or invariant as follows:¹

$$\epsilon = \sum_{\text{adjacent edges}} [\delta(e_i) \delta(e_j)]^{-1/2} \quad (1)$$

The sum is carried out for all adjacent edges in the graph. Two edges are adjacent if they have a common vertex. It is obvious that only the contributions coming from pairs of adjacent bonds in the molecule are taken into account for

Table 1. Edge-Connectivity Counting Polynomials of Octane Isomers

molecule	edge-connectivity counting polynomial
8	$4.5000 + 3.4142x + 2.9142x^2 + 2.4142x^3 + 1.9142x^4 + 1.4192x^5 + x^6$
2M7	$3.8333 + 3.4317x + 2.6154x^2 + 2.1154x^3 + 1.5774x^4 + 1.4142x^5$
3M7	$4.1667 + 3.3423x + 3.3081x^2 + 2.1927x^3 + 1.9916x^4 + x^5$
4M7	$4.1667 + 3.3803x + 3.1783x^2 + 3.0689x^3 + 1.4142x^4 + x^5$
3E6	$4.5000 + 3.2698x + 3.7032x^2 + 3.5689x^3 + 2x^4$
22MM6	$3.2500 + 3.4267x + 2.2854x^2 + 1.7247x^3 + 1.7321x^4$
23MM6	$3.5833 + 3.3730x + 3.2474x^2 + 2.2071x^3 + 1.4142x^4$
24MM6	$3.6667 + 3.3767x + 2.9175x^2 + 2.5774x^3 + 1.4142x^4$
25MM6	$3.1667 + 3.4492x + 2.3333x^2 + 1.6330x^3 + 2x^4$
33MM6	$3.6667 + 3.2988x + 3.3247x^2 + 2.3618x^3 + x^4$
34MM6	$3.9167 + 3.2556x + 4.0640x^2 + 2.5689x^3 + x^4$
2M3E5	$3.9167 + 3.2725x + 3.7877x^2 + 3.8284x^3$
3M3E5	$4.0833 + 3.1160x + 4.7321x^2 + 3.0000x^3$
223MMM5	$3.2000 + 3.3344x + 3.6038x^2 + 1.7321x^3$
224MMM5	$2.5833 + 3.4711x + 1.7071x^2 + 2.4495x^3$
233MMM5	$3.1166 + 3.2116x + 3.9420x^2 + 1.4142x^3$
234MMM5	$3.0000 + 3.3716x + 3.4142x^2 + 2.0000x^3$
2233MMMM4	$2.1667 + 3.4142x + 3.0000x^2$

the edge-connectivity index. To account for long-range bond contributions, we introduce the following polynomial.

e_i and e_j denote a pair of distinct edges in the (molecular) graph G and v_p, v_q and v_r, v_s are the end vertices of e_i and e_j , respectively. Then the edge-connectivity counting polynomial $P_\epsilon(G, x)$ is defined as follows:

$$P_\epsilon(G, x) = \sum_{i=1}^m \sum_{j \geq i}^m [\delta(e_i) \delta(e_j)]^g x^{d(e_i, e_j)} \quad (2)$$

where $d(e_i, e_j)$ is the distance between edges e_i and e_j . This distance has been defined as follows:¹⁹

$$d(e_i, e_j) = \min\{d(p, r), d(p, s), d(q, r), d(q, s)\} + 1 \quad (3)$$

and in the particular case of acyclic graphs this distance is the average of the distances between the corresponding end-vertices of these edges:¹⁹

$$d(e_i, e_j) = 1/4[d(p, r) + d(p, s) + d(q, r) + d(q, s)] \quad (4)$$

The exponent g in expression 2 can take any value similar to that used in the generalization of the connectivity index done by Altenburg.²⁰ It has been proved that the use of values different from $g = -1/2$ in the vertex-connectivity index,^{21–23} as well as in the edge-connectivity one,²² increases the discrimination of isomers and improves the quality of the QSPR models obtained. However, we will consider here the traditionally used value of $g = -1/2$ in order to permit the comparison with the original ϵ index. The edge-connectivity polynomial has been introduced in this way in order to ensure that its first derivative evaluated for $x = 0$ gives the edge connectivity index and the variation in the values of $x \neq 0$ produce new indices representing modifications to the original one. The edge-connectivity counting polynomials of the 18 octane isomers by using $g = -1/2$ are given in Table 1.

The use of graph polynomials in theoretical chemistry is a well-known area of research in chemical graph theory.^{24,25} These polynomials have also been used in the generation or generalization of topological indices.^{26,27} For these reasons there are many possibilities of uses and applications in chemistry and drug research for the novel counting polynomial introduced here. However, in the present study we

will be limited to its use for the generalization of the edge-connectivity index in order to include long-range bond contributions in its definition.

It is straightforward to realize that the first derivative of the edge-connectivity counting polynomial, $P'_\epsilon(G, x) = d(P_\epsilon(G, x))/dx$, evaluated for $x = 0$, is identical to the edge-connectivity index of the (molecular) graph: $P'_\epsilon(G, 0) = \epsilon$. However, if we consider the first derivative of this polynomial evaluated for $x \neq 0$ and $P'_\epsilon(G, x \neq 0)$, we can obtain a series of edge-connectivity indices containing structural information not only on the adjacent bonds but also from the more distant ones. If we design the distance between any pair of edges simply as d , these novel indices, ϵ^c , will have the following general expression:

$$\epsilon^c = \epsilon + \sum_{d=2}^{d_{\max}} \sum_i \sum_{j>i} dx^{d-1} [\delta(e_i) \delta(e_j)]^g \quad (5)$$

where d_{\max} is the maximal distance between two edges and m is the total number of edges in the graph. Here, the second and third sum are carried out only for nonadjacent bonds, and the third sum is carried out only for $j > i$ instead of for $j \geq i$, because the particular case $j = i$ corresponds to the constant term in the polynomial which disappears when the derivative of the polynomial is obtained. We have expressed the long-range edge-connectivity index as the sum of two terms only to show that $\epsilon^c = \epsilon$ in the particular case when $x = 0$.

Here we will employ an empirical approach to the use of the long-range edge-connectivity indices in QSPR studies. Consequently, we will study different values for the term x in expression 5 in order to find what the values are that produce the best QSPR models for describing the physico-chemical properties studied. By this way three cases can be observed. The simplest case is when $x = 0$, which is equivalent to the nonconsideration of long-range contributions. When $x < 0$, contributions coming from fragments containing an even number of bonds, i.e., d is an odd number, are summed to the connectivity index, while contributions coming from fragments having an odd number of atoms, i.e., d is an even number, are subtracted to the original ϵ index. The final case is when $x > 0$, in which the contributions coming from all distant pairs of bonds are summed to the edge connectivity index.

Table 2. Values the Edge-Connectivity Indices of Octane Isomers, Corrected for Long-Range Bond Contributions

molecule	$\epsilon^c(x)$ indices for given x values										
	-1.0	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
8	-1.7574	-0.4503	0.3967	0.9462	1.3205	1.6080	1.8712	2.1535	2.4864	2.8968	3.4142
2M7	5.3085	3.9040	2.9744	2.4133	2.1314	2.0561	2.1320	2.3205	2.6002	2.9665	3.4317
3M7	0.3378	0.1890	0.2285	0.4023	0.6680	0.9954	1.3665	1.7749	2.2265	2.7390	3.3423
4M7	5.5736	4.2735	3.3390	2.7022	2.3069	2.1091	2.0767	2.1897	2.4400	2.8316	3.3803
3E6	-1.4299	-0.5555	0.1010	0.5876	0.9524	1.2433	1.5083	1.7955	2.1528	2.6282	3.2698
22MM6	-2.8980	-1.5464	-0.4655	0.3863	1.0506	1.5690	1.9830	2.3342	2.6642	3.0145	3.4267
23MM6	-2.1573	-1.2329	-0.4815	0.1308	0.6379	1.0738	1.4725	1.8677	2.2936	2.7841	3.3730
24MM6	-0.3829	0.2645	0.7610	1.1407	1.4374	1.6852	1.9178	2.1694	2.4737	2.8649	3.3767
25MM6	-4.3184	-2.6146	-1.2447	-0.1609	0.6849	1.3407	1.8544	2.2741	2.6478	3.0235	3.4492
33MM6	-0.2652	0.1375	0.4659	0.7441	0.9959	1.2455	1.5167	1.8337	2.2203	2.7007	3.2988
34MM6	-1.1657	-0.7332	-0.3625	-0.0297	0.2892	0.6183	0.9815	1.4028	1.9063	2.5159	3.2556
2M3E5	7.1823	5.7577	4.5627	3.5975	2.8619	2.3561	2.0800	2.0335	2.2168	2.6298	3.2725
3M3E5	2.6518	1.8882	1.3046	0.9011	0.6775	0.6339	0.7703	1.0867	1.5832	2.2596	3.1160
223MMM5	1.3231	1.0566	0.8940	0.8353	0.8805	1.0297	1.2828	1.6398	2.1007	2.6656	3.3344
224MMM5	7.4054	6.3506	5.4428	4.6819	4.0680	3.6011	3.2812	3.1082	3.0822	3.2032	3.4711
233MMM5	-0.4298	-0.4475	-0.3803	-0.2283	0.0085	0.3303	0.7368	1.2282	1.8045	2.4656	3.2116
234MMM5	2.5432	2.0860	1.7489	1.5317	1.4346	1.4574	1.6002	1.8631	2.2459	2.7488	3.3716
2233MMMM4	-2.5858	-1.9858	-1.3858	-0.7858	-0.1858	0.4142	1.0142	1.6142	2.2142	2.8142	3.4142

molecule	$\epsilon^c(x)$ indices for given x values									
	+0.1	+0.2	+0.3	+0.4	+0.5	+0.6	+0.7	+0.8	+0.9	+1.0
8	4.0779	4.9441	6.0931	6.0931	7.6369	9.7256	12.5554	16.3754	21.4948	28.2903
2M7	4.0253	4.7935	5.7997	5.7997	7.1242	8.8643	11.1341	14.0648	17.8047	22.5188
3M7	4.0782	5.0004	6.1748	6.1748	7.6791	9.6032	12.0489	15.1299	18.9720	23.7131
4M7	4.1142	5.0731	6.3091	6.3091	7.8860	9.8799	12.3785	15.4820	19.3021	23.9630
3E6	4.1255	5.2433	6.6713	6.6713	8.4574	10.6497	13.2961	16.4446	20.1432	24.4400
22MM6	3.9424	4.6032	5.4506	5.4506	6.5261	7.8715	9.5282	11.5377	13.9418	16.7819
23MM6	4.0943	4.9821	6.0701	6.0701	7.3924	8.9828	10.8754	13.1041	15.7028	18.7054
24MM6	4.0432	4.8982	5.9758	5.9758	7.3099	8.9344	10.8832	13.1903	15.8896	19.0151
25MM6	3.9729	4.6425	5.5061	5.5061	6.6117	8.0073	9.7408	11.8603	14.4138	17.4493
33MM6	4.0386	4.9441	6.0393	6.0393	7.3482	8.8949	10.7032	12.7972	15.2010	17.9384
34MM6	4.1495	5.2215	6.4956	6.4956	7.9959	9.7463	11.7708	14.0935	16.7383	19.7292
2M3E5	4.1449	5.2470	6.5788	6.5788	8.1403	9.9315	11.9524	14.2030	16.6833	19.3934
3M3E5	4.1524	5.3688	6.7653	6.7653	8.3417	10.0981	12.0345	14.1509	16.4474	18.9238
223MMM5	4.1071	4.9838	5.9643	5.9643	7.0488	8.2373	9.5296	10.9259	12.4261	14.0302
224MMM5	3.8860	4.4479	5.1567	5.1567	6.0125	7.0153	8.1651	9.4618	10.9055	12.4962
233MMM5	4.0424	4.9581	5.9586	5.9586	7.0440	8.2143	9.4693	10.8093	12.2341	13.7437
234MMM5	4.1144	4.9773	5.9601	5.9601	7.0630	8.2858	9.6286	11.0915	12.6743	14.3772
2233MMMM4	4.0142	4.6142	5.2142	5.2142	5.8142	6.4142	7.0142	7.6142	8.2142	8.8142

The implications of the consideration of this kind of long-range contribution between bonds in the edge-connectivity theory to the study of physicochemical properties of organic compounds will be considered in the following.

APPLICATIONS TO QSPR STUDIES

A series of topological indices derived from the edge-connectivity counting polynomial were obtained by varying the values of x from -1 to $+1$. Consequently, 20 ϵ^c indices (excluding the original ϵ index) were generated, and they are shown in Table 2. These indices containing contributions coming from bonds separated at distances greater than 1 were analyzed for linear dependence by factor analysis with the principal component method.²⁸ In the preceding paper in this issue we study the linear dependence of the edge-connectivity index to a series of 37 first-, second-, and third-generation topological indices.¹³ These indices will be used in the present work to analyze the interrelation of the ϵ^c index to other molecular descriptors. For definitions and references of the topological indices used see the precedent work of this series.¹³

When the 56 topological descriptors (36 from the literature and 20 ϵ^c indices) were analyzed by the factor analysis, the results shown in Table 3 were obtained. The indices taken

Table 3. Results of the Factor Analysis by Using the Principal Component Method for 36 First-, Second-, and Third-Generation Topological Indices as Well as the 20 Edge-Connectivity Indices Corrected for Long-Range Bond Contributions for Octane Isomers

	eigenvalue	% total variance	cumul %
1	30.256	52.16	52.16
2	12.121	20.89	73.06
3	5.627	9.70	82.76
4	2.460	4.24	87.00

from the literature include first-, second-, and third-generation topological indices. The first-generation TI's are those integer numbers obtained from graph-theoretical invariants involving integer numbers. The second-generation TI's are real numbers obtained from manipulation of local graph properties that are integer numbers. Finally, those real numbers obtained from graph invariants involving real numbers compose the third generation of topological indices. Four factors were extracted, explaining 87% of the variance.

The factor loadings from the principal component analysis, after a Quartimax rotation of the factors, are shown in Table 4. The objective of this rotational procedure is to obtain a clear pattern of loadings, that is, factors that are somehow clearly marked by high loadings for some variables and low loadings for others. The Quartimax rotation is directed to maximize the variances of (raw) factor loadings across factors

Table 4. Factor Loadings for 36 First-, Second-, and Third-Generation Topological Indices as Well as 20 Edge-Connectivity Indices Corrected for Long-Range Bond Contributions for Octane Isomers

index	f1	f2	f3	f4	index	f1	f2	f3	f4
W	0.947	0.090	0.005	0.037	X	-0.948	-0.067	0.040	-0.008
Z	0.779	-0.539	0.036	0.040	Y	-0.919	-0.105	0.098	0.073
\bar{W}	0.925	0.135	0.044	0.072	VEA1	0.461	0.342	0.022	0.759
MTI	0.940	0.135	0.014	0.036	VEA3	0.459	0.340	0.019	0.762
WW	0.937	0.119	0.027	0.055	VED1	-0.162	0.713	0.050	0.394
$^2W_{De}$	0.942	0.084	0.019	0.051	VED3	0.028	0.408	0.178	-0.712
$^2W_{We}$	0.904	0.277	0.028	0.081	VRA1	0.625	-0.077	0.016	-0.715
$^2W_{Cju}$	0.939	0.102	0.024	0.054	VRD1	-0.928	-0.195	0.035	-0.016
$^2WW_{Dp}$	0.915	0.114	0.054	0.090	$\epsilon^c(-1.0)$	-0.037	-0.139	-0.899	0.055
$^2WW_{Wp}$	0.914	0.176	0.050	0.095	$\epsilon^c(-0.9)$	0.007	-0.090	-0.929	0.047
D^2-CJ_u	0.908	0.146	0.061	0.094	$\epsilon^c(-0.8)$	0.062	-0.016	-0.947	0.031
χ	0.828	-0.467	-0.037	-0.013	$\epsilon^c(-0.7)$	0.129	0.088	-0.944	0.009
ϵ	0.290	0.856	-0.100	0.099	$\epsilon^c(-0.6)$	0.204	0.225	-0.904	-0.018
ID	0.916	-0.255	-0.037	-0.019	$\epsilon^c(-0.5)$	0.278	0.386	-0.815	-0.043
R_1	-0.904	0.258	0.062	0.098	$\epsilon^c(-0.4)$	0.333	0.548	-0.680	-0.059
IBC	0.898	-0.217	-0.076	0.008	$\epsilon^c(-0.3)$	0.357	0.687	-0.521	-0.061
AIBC	0.868	-0.211	-0.010	-0.067	$\epsilon^c(-0.2)$	0.348	0.792	-0.363	-0.044
H	-0.947	0.064	0.055	-0.015	$\epsilon^c(-0.1)$	0.313	0.860	-0.224	-0.006
J	-0.945	-0.095	0.039	0.004	$\epsilon^c(+0.1)$	0.063	-0.862	0.057	0.171
H^V	0.773	0.071	-0.235	-0.090	$\epsilon^c(+0.2)$	0.092	-0.942	-0.012	0.027
H^D	-0.785	-0.130	-0.087	0.102	$\epsilon^c(+0.3)$	0.227	-0.919	-0.053	-0.022
I_D^E	0.885	0.023	-0.046	-0.082	$\epsilon^c(+0.4)$	0.227	-0.919	-0.053	-0.022
X(D)	-0.919	0.208	0.074	-0.026	$\epsilon^c(+0.5)$	0.381	-0.862	-0.072	-0.051
X(LK)	-0.939	0.115	0.065	-0.017	$\epsilon^c(+0.6)$	0.531	-0.777	-0.075	-0.065
DM ¹	-0.913	0.156	0.108	0.038	$\epsilon^c(+0.7)$	0.658	-0.675	-0.066	-0.066
C(LM3D)	-0.927	0.114	0.003	0.001	$\epsilon^c(+0.8)$	0.755	-0.567	-0.049	-0.057
U	0.882	0.341	0.015	-0.042	$\epsilon^c(+0.9)$	0.823	-0.466	-0.030	-0.042
V	-0.941	-0.096	0.054	0.019	$\epsilon^c(+1.0)$	0.866	-0.377	-0.011	-0.024

for each variable. This strategy makes the structure of factors pattern as simple as possible, permitting a clearer interpretation of the factors without loss of orthogonality between them.

The structure of the factor space obtained for this series of topological indices is very similar to that obtained in our preceding work with only one difference being that a new factor has appeared. As in the preceding work of this series,¹³ most of the first- and second-generation topological indices are loaded in the factor 1. The second factor is intimately related to the edge-connectivity¹ and the Balaban's VED1 indices.²⁹ The fourth factor corresponds to factor 3 in the preceding work, where some of the third-generation Balaban's indices are loaded.²⁹ However, factor 3 here corresponds to a new axis related to some of the $\epsilon^c(x)$ indices introduced here. It is interesting to note that the variation of the value of x in the edge-connectivity indices containing long-range information produces three series of independent topological indices. The first group corresponds to the values of ϵ^c that are highly loaded in factor 3. These indices are not only orthogonal to the other two groups of $\epsilon^c(x)$ indices but also to the rest of the first-, second-, and third-generation topological indices. The second group of $\epsilon^c(x)$ indices is that obtained for $-0.2 \leq x \leq +0.6$, which are loaded in factor 2, i.e., the factor related to the edge-connectivity index. The third group of $\epsilon^c(x)$ indices corresponds to the values x between 0.8 and 1.0. These indices are loaded in the first factor, that mainly related to most of the first and second generation of topological indices. There are some borderline cases such as $\epsilon^c(-0.4)$ and $\epsilon^c(-0.3)$ which are between groups 1 and 2, and the index $\epsilon^c(0.7)$ that is between groups 2 and 3.

We recall that the topological indices loaded in factors 1 and 2 in the precedent work were able to describe all seven physicochemical properties of octanes studied there.¹³ Con-

sequently, we can expect that the $\epsilon^c(x)$ indices, which are loaded in these two factors, will be of significant importance in describing these properties through QSPR models. To prove this hypothesis, we will study the description of boiling points (BP), molar volumes (MV), molar refraction (MR), heat of vaporization (HV), critical temperature (CT), critical pressure (CP), and surface tension (ST) of octane isomers³⁰ with the $\epsilon^c(x)$ indices.

To select the best QSPR models for describing the physicochemical properties studied, we obtain all the single-variable regression equations with the $\epsilon^c(x)$ indices varying x from -1.0 to $+1.0$. By this way, we found three different cases for the optimal QSPR models corresponding to values of $x < 0$, $x > 0$, and $x = 0$. The first case, that is when the best QSPR model to describe a physicochemical property is obtained for a negative value of x , was observed for molar volume and molar refraction. In these two cases the optimal x value was -0.1 . In Figure 1A, we show the plot of the standard deviation of the regression versus the x values for the molar volume. The second case corresponded to those properties for which the optimal QSPR models were obtained by using positive x values. The properties included in this group are, together with the optimal value of x in parentheses: BP (0.7); HV (1.0); TC (0.2); ST (0.2). In Figure 1B,C we illustrate the plots of the standard deviation of the regression versus x for HV and ST. The last case corresponds to those properties for which the best QSPR models are obtained with $x = 0$; i.e., there is no improvement with the use of $\epsilon^c(x \neq 0)$ with respect to the original ϵ index. This case was only observed for the critical pressure, which is plotted in Figure 1D.

It is clear from this study that we included only such topological indices that fulfill most of the desired attributes claimed by Randic for these kind of molecular descriptors. One of the most important of these requirements is the

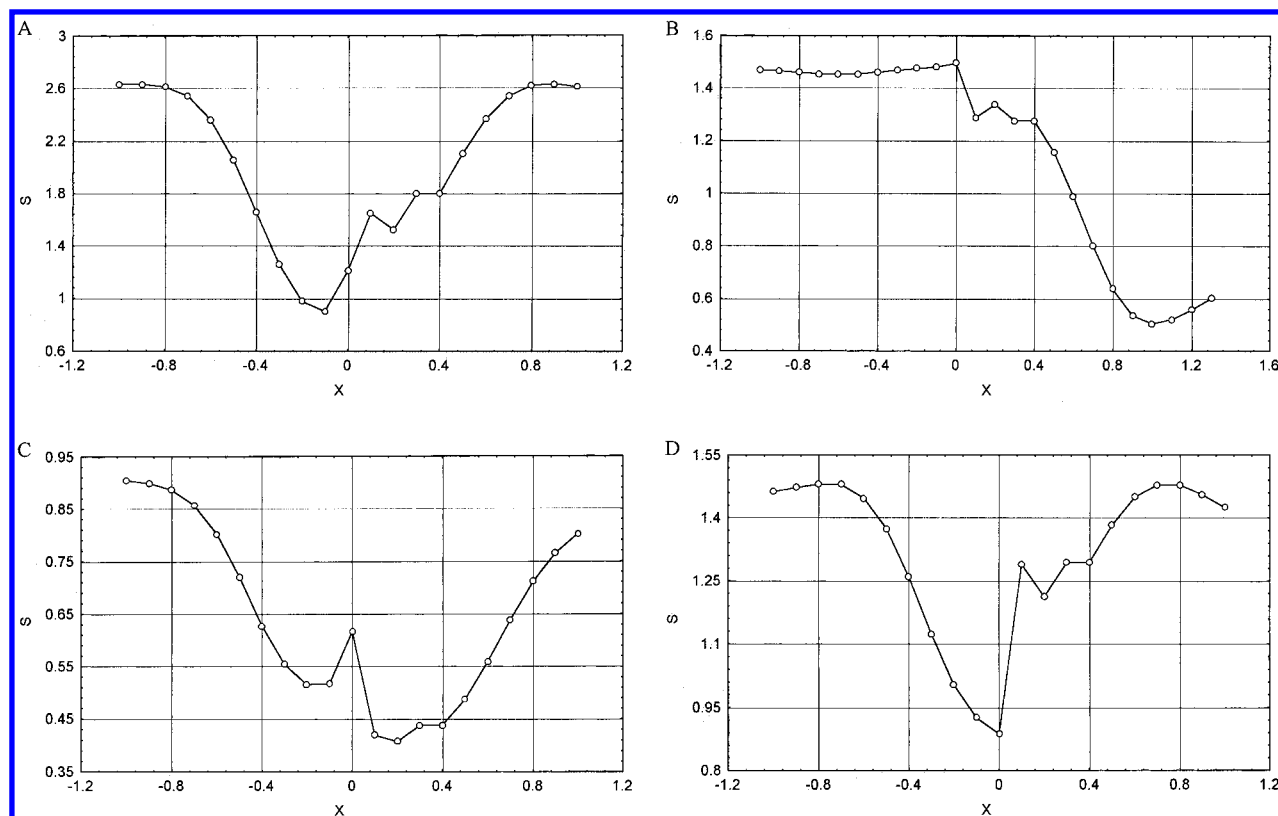


Figure 1. Plot of the standard deviation of the regression model against the x parameter in the edge-connectivity index corrected for long-range bond contributions for molar volume (A), heat of vaporization (B), surface tension (C), and critical pressure (D) of octane isomers.

gradual change that these descriptors need to have when the size of the molecule increases. All the topological indices studied here fulfill this requisite, and this way we have concentrated on the study of isomer variation of such indices in the data set of octane isomers. However, some confusion can be produced if some simple molecular descriptors are included in this study, such as the path number p_3 . This index produces excellent correlations with MV ($R = 0.973$, $s = 0.609$) and MR ($R = 0.986$, $s = 0.032$) for octane isomers. But, when data sets containing molecules with different sizes are studied, for instance alkanes with two to nine carbon atoms, these good correlations are hardly observed: $R = 0.693$ and $s = 1.874$ for MV and $R = 0.774$ and $s = 1.645$ for MR. In this case, we have proved before that the edge connectivity index produced excellent correlations with $R > 0.99$. At this point there is not doubt that the p_3 descriptor can be considered only as an auxiliary index to be used together with other "real" topological indices, such as the W number as originally proposed by Wiener.

The improvements in the statistical parameters obtained by the $\epsilon^c(x)$ indices compared to the original edge-connectivity index¹ are given in Table 5. In the preceding paper of this series,¹³ we proved that the edge-connectivity index was the best single molecular descriptor from a pool of 38 topological indices to describe MV, MR, TC, PC, and ST of octanes. As can be seen in Table 5, there are significant improvements in the correlation coefficient and the standard deviation of the regression for these properties, except for PC, with the use of the $\epsilon^c(x)$ indices. For instance, improvements in the standard deviations of the molar volume and refraction are of 25.2 and 15.9%, respectively. These improvements are 14.9 and 34.0% for TC and ST, respec-

Table 5. Comparison of the Statistical Results for the QSPR Models Describing Physicochemical Properties of Alkanes with the Edge-Connectivity Index Corrected for Long-Range Bond Contributions and the Original Edge-Connectivity Index

index		BP	MV	MR	HV	TC	ST
$\epsilon^c(x)$	R	0.8296	0.9392	0.9306	0.9413	0.8476	0.8934
	s	3.523	0.904	0.069	0.505	5.333	0.407
	x	0.7	-0.1	-0.1	1.0	0.2	0.2
ϵ	R	0.4240	0.8885	0.9001	0.0359	0.7818	0.7318
	s	5.714	1.208	0.082	1.495	6.267	0.617

tively. By this way the $\epsilon^c(x)$ indices (x given in Table 5) produce the best single QSPR variables for these physicochemical properties of octanes from all the 36 first-, second-, and third-generation topological indices analyzed in the previous paper.

It was also shown in our preceding work that the edge connectivity was not able to describe the boiling point and heat of vaporization of octanes. In Table 5 we can see the dramatic improvements obtained with the $\epsilon^c(x)$ indices for these two properties compared to the ϵ index. The improvements in the standard deviation of the regression for these properties are of 38.3 and 66.2%, respectively. We recall that the best statistical parameters of QSPR models for BP reported in our preceding paper¹³ were $R = 0.8883$ and $s = 2.897$ with the Hosoya's Z index.¹⁴ The results obtained with the Randic's χ index² were $R = 0.8212$ and $s = 3.600$.¹³ In the case of HV, the best regression model was that using the Z index: $R = 0.9581$ and $s = 0.429$. As can be seen, the models obtained here by using the $\epsilon^c(x)$ indices are not better but at least comparable to the best QSPR models describing BP and HV of octanes. Consequently, the quality of the edge-connectivity index to be used in QSPR/QSAR studies has

been significantly improved with the inclusion of the long-range bond contributions accounted for by using the first derivative of the edge-connectivity counting polynomial. These improvements are even more significant than those reported by us with the use of modified vertex-connectivity indices, including long-range contributions between atoms in the molecule. For instance, the use of the $\chi^c(x)$ indices¹⁵ produced significant improvements for three physicochemical properties of octanes compared to the original χ index. However, only one of these QSPR models is ranked as the best QSPR model for one of these properties of octanes. This is the case of the boiling points for which $\chi^c(x = -0.3)$ gives $R = 0.9368$ and $s = 2.206$.¹⁵ The QSPR models obtained for each physicochemical property with the $\epsilon^c(x)$ indices are given below:

$$\text{BP} = 80.632 + 3.090 \epsilon^c(x = +0.7)$$

$$\text{MV} = 132.472 + 10.375 \epsilon^c(x = -0.1)$$

$$\text{MR} = 36.992 + 0.736 \epsilon^c(x = -0.1)$$

$$\text{HV} = 32.331 + 0.315 \epsilon^c(x = +1.0)$$

$$\text{TC} = 125.25 + 33.26 \epsilon^c(x = +0.2)$$

$$\text{ST} = 125.25 + 33.26 \epsilon^c(x = +0.2)$$

From the structural point of view, it is clear that the long-range bond contributions are different for each of the physicochemical properties of octane isomers studied. These contributions depend on the value of x in the definition of the $\epsilon^c(x)$ index and on the sign of the coefficient multiplying this index in the QSPR model. These coefficients for the physicochemical properties studied here were always positive, and the interpretation of the results is straightforward. By this way, the bonds separated at odd distance contribute positively to all the physicochemical properties studied, with the exception of the critical pressure for which long-range contributions are not significant. However, the contributions coming from bonds separated at even distance contribute positively to BP, TC, HV, and ST and negatively to MV and MR.

CONCLUDING REMARKS

The introduction of the edge-connectivity counting polynomial has permitted to account for long-range bond contributions in the edge-connectivity index. We have shown that these long-range contributions are of great importance for the description of physicochemical properties of octane isomers. However, as normally occurs in science, the introduction of new concepts open different questions that should be answered for further development of the theory and applications of the concept introduced. The theoretical quantity obtained here is, for whatever reason, useful in describing the variance of most of the physicochemical properties studied here for octane isomers. In our preceding paper we have shown that the edge-connectivity index is more appropriated for describing "molecular volume" dependent properties, such as MV and MR, instead of those "molecular surface" properties, such as BP and HV. However, consideration of the long-range contributions to the edge connectivity index has permitted transformation the

original molecular volume descriptor into a molecular surface one. If we observe the QSPR models using the long-range edge-connectivity index to describe molecular volume properties, such as MV, MR, ST, and TC, we can see that the optimal values of x are very close to 0 ($x = -0.1$ for MV and MR and $x = +0.2$ for ST and TC). On the other hand, molecular surface properties are better described by greater values of x , such as 0.7 for BP and 1.0 for HV. This indicates that when the value of x is increased, the contributions coming from more distant bonds in the molecule are each time greater and better descriptions of the molecular surface features are obtained.

ACKNOWLEDGMENT

I want to thank both reviewers for critical comments that contributed significant improvement to this work.

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) Randic, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (4) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, 1986.
- (5) 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.
- (6) Estrada, E.; Ramirez, A. Edge adjacency relationships and molecular topographic descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837–843.
- (7) Estrada, E.; Guevara, N.; Gutman, I. Extension of edge connectivity index. Relationships to line graph indices and QSPR applications. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 428–431.
- (8) Casch, G. G. Correlation of physicochemical properties of alkylphenols with their graph-theoretical ϵ parameter. *Chemosphere* **1995**, *31*, 4307–4315.
- (9) Lekishvili, T.; Lekishvili, G.; Aleksidze, N. Mathematical investigation of the cancerogenic activity of some hormones. *Bull. Georgian Acad. Sci.* **1997**, *155*, 441–443.
- (10) Nikolic, S.; Trinajstić, N. Comparison between the vertex- and edge-connectivity indices for benzenoid hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 42–46.
- (11) Plavsic, D.; Trinajstić, N.; Amic, D.; Solisic, M. Comparison between the structure-boiling points relationships with different descriptors for condensed benzenoids. *New J. Chem.* **1998**, *22*, 1075–1078.
- (12) Nikolic, S.; Trinajstić, N. Modeling the aqueous solubility of aliphatic alcohols. *SAR QSAR Environ. Res.* **1998**, *9*, 117–126.
- (13) Estrada, E.; Rodríguez, L. Edge-Connectivity Indices in QSPR/QSAR Studies. 1. Comparison to Other Topological Indices in QSPR Studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037–1041.
- (14) Hosoya, H. Topological index. A newly proposed quantity characterizing topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (15) Estrada, E. Connectivity Counting Polynomial and Long-Range Contributions in the Molecular Connectivity Model. *Chem. Phys. Lett.*, in press.
- (16) Estrada, E. Novel Strategies in the Search of Topological Indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach: Reading, 1999, pp 503–553.
- (17) Randic, M.; Trinajstić, N. Viewpoint 4-Comparative structure-property studies: the connectivity basis. *J. Mol. Struct. (THEOCHEM)* **1993**, *284*, 209–221.
- (18) Randic, M.; Trinajstić, N. In search for graph invariants of chemical interest. *J. Mol. Struct. (THEOCHEM)* **1993**, *300*, 551–572.
- (19) Estrada, E.; Gutman, I. A topological index based on distances of edges of molecular graphs. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 850–853.
- (20) Altenburg, K. Eine Bemerkung zu dem Randicschen "Molekularen Bindungs-Index (Molecular Connectivity Index)". *Z. Phys. Chem. (Leipzig)* **1980**, *261*, 389–393.

- (21) Randic, M.; Hansen, P. J.; Jurs, P. C. Search for useful graph theoretical invariants of molecular structure. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 60–68.
- (22) Estrada, E Graph theoretical invariant of Randic revisited. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 1022–1025.
- (23) Amic, D.; Beslo, D.; Lucic, B.; Nikolic, S.; Trinajstic, N. The vertex-connectivity index revisited. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 819–822.
- (24) Gutman, I. Polynomials in Graph Theory. In *Chemical Graph Theory. Introduction and Fundamentals*; Bonchev, D., Rouvray, D. H., Eds.; Gordon & Breach: New York, 1991.
- (25) Klein, D. J.; Randic, M.; Babic, D.; Trinajstic, N. On conjugated-circuit polynomials. *Int. J. Quantum Chem.* **1994**, 50, 369–384.
- (26) Hosoya, H. On some counting polynomials in chemistry. *Discr. Appl. Math.* **1988**, 19, 239–257.
- (27) Estrada, E.; Ivanciuc, O.; Gutman, I.; Gutierrez, A.; Rodriguez, L. Extended Wiener indices. A new set of descriptors for quantitative structure-property studies. *New J. Chem* **1998**, 22, 819–822.
- (28) Malinowski, E. R.; Howery, D. G. *Factor Analysis in Chemistry*; Wiley: New York, 1980.
- (29) Balaban, A. T. Using real numbers as vertex invariants for third-generation topological indexes. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 23–28.
- (30) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular modeling of the physical properties of alkanes. *J. Am. Chem. Soc.* **1988**, 110, 4186–4194.

CI990031H