Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume

Ernesto Estrada

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

Received March 22, 1994[⊗]

Adjacency between edges in molecular graphs is represented by a square and symmetric edge-adjacency matrix, and one index is derived using the Randić-type graph-theoretical invariant. The novel index, which is very simple to calculate, is highly correlated with molar volume and related properties; it also has good discrimination between isomers and is orthogonal with other indices described in literature.

1. INTRODUCTION

The use of graph-theoretical invariants in quantitative structure−property relationships (QSPR) and quantitative structure−activity relationships (QSAR) studies[1] have become of major interest in the recent years, and several review papers on topological indices have been published.[2−7]

Most of the existing topological indices are related either to a vertex adjacency matrix, known simply as adjacency matrix, or to a distance matrix in graphs. There is a great number of topological indices; among these indices we can find Wiener's number W,[8] Hozoya's, Z index,[9] Randić's molecular connectivity,[10] Balaban's J index,[11] a molecular ID number,[12] Kier's κ indices,[13] electrotopological state,[14] Harary number,[15] and so on.

The proliferation of topological indices can be compared with the effect produced on quantum chemical parameters by a change of molecular orbital basis, and principal problems are related to its physical meaning to avoid a duplication of information contained in them.[16]

Randić states[17] that novel molecular descriptors need to be simpler, add some insights to the problem, or solve a problem that was not solvable with alternative schemes.

One of the possible solutions to this problem is the introduction of a new representation of molecular graphs, for instance, topographic matrices proposed by Randić,[18,19] Bogdanov et al.,[20] and Estrada[21,22] to incorporate stereochemical features of molecules.

We propose in this paper a new matrix representation of graphs based on the adjacency of edges, and one topological index derived therefrom is used to correlate molar volumes and molar fraction of alkanes.

2. THEORETICAL SECTION

Let $G = \{V,E\}$ be a connected graph, where V is the vertex set and E the edge set. The vertex-adjacency matrix, commonly called the adjacency matrix of G will be a square and symmetric matrix $A = [a_{ij}]_{nxn}$, where n is the number of vertices, and elements a_{ij} are defined as follows:

$$a_{ij} = \begin{cases} 1 & \text{if} \quad e_k \in E/e_k \sim v_i,v_j \\ 0 & \text{otherwise} \end{cases}$$

i.e., if there is an edge e_k incident with vertices v_i and v_j ($e_k \sim v_i,v_j$), then both vertices are adjacent and $a_{ij} = 1$.

The adjacency relationship is not an exclusive property of elements of the vertex set V, because two edges e_i and e_j, can be adjacent if there is one vertex v_k incident with both edges. Hence, we can build an edge-adjacency matrix $E = [g_{ij}]_{mxm}$, where m is the number of edges in the graph and elements g_{ij} are defined as follows:

$$g_{ij} = \begin{cases} 1 & \text{if} \quad v_k \in V/v_k \sim e_i,e_j \\ 0 & \text{otherwise} \end{cases}$$

An edge's degree $\delta(e_k)$, can be defined, in the same way that vertex degrees are, but using elements g_{ij} of E matrix:

$$\delta(e_k) = \sum_i g_{ik} = \sum_j g_{kj}$$

If e_k is incident with vertices v_i and v_j, then an edge degree of e_k can be expressed as a function of vertex degree v_i and v_j as

$$\delta(e_k) = \delta(v_i) + \delta(v_j) - 2$$

Now, we define a new topological (graph-theoretical) index ϵ, using the Randić-type[10] graph-theoretical invariant,

$$\epsilon = \sum_l [\delta(e_i)\ \delta(e_j)]_l^{-1/2}$$

where the sum is over all l adjacent edges in the graph.

The calculation technique of the ϵ index is illustrated with 2,3-dimethylbutane in Figure 1, where the hydrogen-suppressed graph is also shown, as well as its edge's adjacency matrices.

3. CALCULATIONS

The ϵ index was calculated for 69 C_5−C_9 alkanes by using the STAR software system,[23] and its values were correlated with molar volumes (MV) and molar refraction (MR). Molar volumes were calculated as MW/d, where MW is the molecular weight and d is the density (g/cm^3) at 20 °C. Molar refractions were calculated using the Lorentz−Lorenz expression:

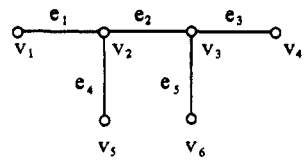
$$MR = \frac{\eta_{o-1}}{r_o^2 + 2} MV$$

where η_o is the index of refraction and MV the molar volume.

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1994.

0095-2338/95/1635-0031$09.00/0

a) Graph G with labeled vertices and edges:



b) The edge adjacency matrix $\mathbb{E}$ of G:

$$\mathbf{E} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

c) Vertex and edge degrees:

$$\delta(v_1) = \delta(v_4) = \delta(v_5) = \delta(v_6) = 1$$
$$\delta(v_2) = \delta(v_3) = 3$$

$$\delta(e_1) = \delta(e_3) = \delta(e_4) = \delta(e_5) = 2$$
$$\delta(e_2) = 4$$

d) The ϵ index of G:

$$\epsilon = 4 \cdot (2 \cdot 4)^{-1/2} + 2 \cdot (2 \cdot 2)^{-1/2} = 2.4142$$

Figure 1. The computation of the ϵ index for a graph G representing the carbon skeleton of 2,3-dimethylbutane.

These properties are important with regard to the relationships of the ϵ index with molecular volume and molecular polarizability.[24]

Regression equations and statistical parameters for correlation of both properties and the ϵ index are shown below:

$$MV = 57.8501 + 30.8559\epsilon$$

$$r = 0.9931 \qquad s = 2.032 \qquad f = 4831 \qquad (1)$$

$$MR = 7.6512 + 9.3742\epsilon$$

$$r = 0.9913 \qquad s = 0.698 \qquad F = 3782 \qquad (2)$$

where r, s, and F represent the correlation coefficient, standard deviation of regression, and Fisher ratio, respectively. All values of the ϵ index as well as experimental and calculated properties are depicted in Table 1. In general, statistics indicate that eqs 1 and 2 represent very good models for calculating *MV* and *MR*, because more than 98% of the variance in the properties is accounted for by both equations; standard errors are 1.2 and 1.8% of the mean values of molar volumes and molar refraction, respectively. The use of the Kolmogorov–Smirnov test also proved that residuals are normally distributed.

4. DISCUSSION

In a recent paper, Randić[26] proposed a list of desirable attributes for topological indices that can be considered as a methodology for designing new topological indices. The first criterion examined was its direct structural interpretation to help us in the comprehension of convoluted and complex properties in terms of the structure. The good correlation between the ϵ index and molar volume or related properties, i.e., molar refractivity, considered as molecular bulkiness or steric parameters,[27,28] permits us to interpret this index as a good measurement of molecular volume.

Table 1. Experimental Molar Volume (MV) and Molar Refraction (MR) for C_5–C_9 Alkanes and Predicted Values Based on Simple Linear Regression for the ϵ Index

		MV[b]		MR	
compd[a]	ϵ index	obsd	calcd	obsd	calcd
5	1.9142	115.205	116.914	25.2656	25.5953
2M4	1.8937	116.426	116.282	25.2923	25.4031
22MM3	1.9998	122.074	119.556	25.7243	26.3977
6	2.4142	130.688	132.342	29.9066	30.2824
2M5	2.4317	131.933	132.882	29.9459	30.4464
3M5	2.3043	129.717	128.951	29.8016	29.2522
22MM4	2.3660	132.744	130.855	29.9347	29.8306
23MM4	2.4142	130.240	132.342	29.8104	30.2824
7	2.9142	146.540	147.770	34.5504	34.9695
2M6	2.9317	147.656	148.310	34.5908	35.1335
3M6	2.8423	145.821	145.552	34.4597	34.2955
3E5	2.7318	143.517	142.142	34.2827	33.2596
22MM5	2.9267	148.695	148.156	34.6166	35.0867
23MM5	2.8350	144.153	145.327	34.3237	34.2270
24MM5	2.9661	148.949	149.372	34.6192	35.4560
33MM5	2.7381	144.530	142.337	34.3323	33.3187
223MMM4	2.9069	145.191	147.545	34.3736	34.9011
8	3.4142	162.592	163.198	39.1922	39.6566
2M7	3.4317	163.663	163.738	39.2316	39.8206
3M7	3.3423	161.832	160.980	39.1001	38.9826
4M7	3.3803	162.105	162.152	39.1174	39.3388
3E6	3.2698	160.072	158.743	38.9441	38.3030
22MM6	3.4267	164.285	163.584	39.2525	39.7738
23MM6	3.3730	160.395	161.927	38.9808	39.2704
24MM6	3.3767	163.093	162.041	39.1300	39.3051
25MM6	3.4492	164.697	164.278	39.2596	39.9847
33MM6	3.2988	160.879	159.638	39.0087	38.5748
34MM6	3.2556	158.814	158.304	43.6870	43.1698
23ME5	3.2725	158.794	158.826	38.8362	38.3283
33ME5	3.3154	157.026	160.150	38.7171	38.7304
223MMM5	3.3344	159.526	160.736	38.9449	38.9085
224MMM5	3.4711	165.083	164.954	39.2617	40.1900
233MMM5	3.2116	157.292	156.947	38.7617	37.7574
234MMM5	3.3716	158.852	161.884	38.8681	39.2572
9	3.9142	178.713	178.626	43.8423	44.3437
2M8	3.9317	179.773	179.166	43.8795	44.5077
3M8	3.8423	177.952	176.408	43.7296	43.6697
4M8	3.8803	178.150	177.580	43.7687	44.0259
3E7	3.7698	176.410	174.171	43.6420	42.9900
4E7	3.8078	175.685	175.343	43.4907	43.3463
22MM7	3.9267	180.507	179.012	43.9138	44.4609
23MM7	3.8730	176.653	177.355	43.6269	43.9575
24MM7	3.9147	179.120	178.624	43.7393	44.3484
25MM7	3.8598	179.371	176.948	43.8484	43.8337
26MM7	3.9492	180.914	179.706	43.9258	44.6718
33MM7	3.7988	176.897	175.065	43.6870	43.2619
34MM7	3.7936	175.349	174.905	43.5473	43.2132
35MM7	3.7873	177.386	174.711	43.6378	43.1541
44MM7	3.8595	176.897	176.938	43.6022	43.8309
23ME6	3.8105	175.445	175.426	43.6550	43.3716
24ME6	3.8042	177.386	175.232	43.6472	43.3125
33ME6	3.8761	173.077	177.451	43.2680	43.9865
34ME6	3.6931	172.844	171.804	43.3746	42.2710
223MMM6	3.8724	175.878	177.336	43.6226	43.9518
224MMM6	3.8817	179.220	177.623	43.7638	44.0390
225MMM6	3.9442	181.346	179.552	43.9356	44.6249
233MMM6	3.8438	173.780	176.454	43.4347	43.6838
234MMM6	3.6231	173.498	169.644	43.3917	41.6149
235MMM6	3.9074	177.656	178.416	43.6474	44.2799
244MMM6	3.7695	177.187	174.161	43.6598	42.9872
334MMM6	3.6231	172.055	169.644	43.3407	41.6149
33EE5	3.5000	170.185	165.846	43.1134	40.4609
223MME5	3.7788	174.537	174.448	43.4571	43.0744
233MME5	3.6652	170.093	170.943	42.9542	42.0095
234MEM5	3.8191	173.804	175.692	43.4037	43.4522
2233(M)5	3.7932	169.495	174.893	43.2147	43.2094
2234(M)5	3.8751	173.557	177.420	43.4359	43.9772
2244(M)5	3.9820	178.256	180.718	43.8747	44.9793
2334(M)5	3.8309	169.928	176.056	43.2016	43.5628

[a] Abbreviations as in ref 1. [b] Experimental values of density and index of refraction used to calculate MV and MR from ref 25.

This interpretation is possible by considering that the alkanes are nonpolar compounds and a number of complexities due to polarity, polarizability, and/or hydrogen bonding that are present in more polar compounds are absent.

On the other hand, the ϵ index has a good discrimination of isomers, which is another important feature required for topological indices. Considering all C_3-C_9 alkanes, there are only two pairs of isomers with the same value of this index: (i) *n*-hexane and 2,3-dimethylbutane, $\epsilon = 2.4142$; (ii) 2,3,4-trimethylhexane and 3,3,4-trimethylhexane, $\epsilon = 3.6231$.

The problem of degeneracy is found in several topological indexes too; for instance, for the studied series of alkanes there are seven pairs of isomers with the same values of Randić index, and even highly discriminative indices such as ID numbers have two couples of isomers with degenerated values of the index.

Linear independence, or orthogonality, of the ϵ index can be expressed by the fact that molar volume is a property not successfully correlated with other topological descriptors. The above may be illustrated by the best two equations of a single variable found by Seybold et al. in terms of topological indices for the same series of alkanes.[29] The first model, using the modified Wiener index W_{mod}, produces a correlation coefficient of 0.976 and standard deviation of 3.8, and the second model, correlating *MV* with the zero-order connectivity index ${}^0\chi$, gave $r = 0.962$ and $s = 4.8$. Similar results were obtained in equations with two topological indices; for instance the model found by Seybold using ${}^0\chi$ and ${}^1\chi$ gave $r = 0.9823$ for the correlation coefficient and $s = 3.3$ for the standard deviation.

The quality of the QSPR models can be conveniently measured by the correlation coefficient r and the standard deviation s, and a good QSPR model must have, as suggested by Mihalic and Trinajstić[7] in a recent strategy to design QSPR models, $r > 0.99$, while s depends on the property under study. Equations 1 and 2 satisfy this condition, and they are the best equations to describe molar volume and molar refraction using one topological index.

Other attributes claimed by Randić, such as simplicity, not based on physicochemical properties, not trivially related to other indices, to show a correct size dependence, etc., are well satisfied by the ϵ index.

Other features that can be included in the ϵ index are the generalization to "higher" analogues[30] and the inclusion of heteroatoms.[31]

5. CONCLUDING REMARKS

The adjacency of edges in a graph can be represented in a square and symmetric matrix E, and the edge degrees, defined as the sum of elements of each row or column, has been used to generate a new topological index ϵ. The proposed ϵ index promises to be a useful parameter in quantitative structure−property (QSPR) and quantitative structure−activity relationships (QSAR). Our hope is based on the features of this index since it is very related with molecular volume, it has good discrimination of isomers, and it is very simple to calculate.

It is important that this simple topological index has a direct structural interpretation, contrasting with the lack of physical meaning of several molecular descriptors (not only graph-theoretical indices) and the unclear significance of correlations where they appear.

The applications of edge-adjacency relationships in molecular graphs open new possibilities for generating other topological (graph-theoretical) indices to be used in QSPR and QSAR and to give continuity to the development of chemical graph theory.

ACKNOWLEDGMENT

I would like to thank Professor Milan Randić for sending me several reprints of his papers about chemical graph theory.

REFERENCES AND NOTES

(1) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(2) Sabljic, A.; Trinajstić, N. *Acta Pharm. Jugosl.* **1981**, *31*, 189.
(3) Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334.
(4) Balaban, A. T. In *Graph Theory and Topology in Chemistry*; King, R. B., Rouvray, D. H., Eds.; Elsevier: Amsterdam, 1987; p 159.
(5) Stakevich, M. I.; Stakevich, I. V.; Zefirov, I. V. *Usp. Khim.* **1988**, *5*, 337.
(6) Hansen, P. J.; Jurs, P. C. *J. Chem. Educ.* **1988**, *65*, 574.
(7) Mihalic, Z.; Trinajstić, N. *J. Chem. Educ.* **1992**, *69*, 701.
(8) Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
(9) Hosoya, H. *Bull Chem. Soc. Jpn.* **1971**, *44*, 2332.
(10) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
(11) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
(12) Randić, M. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164.
(13) Kier, L. B. *Acta Pharm. Jugosl.* **1986**, *36*, 172.
(14) Hall, L. H.; Mohney, B.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76.
(15) Plavsic, D.; Nikolic, S.; Trinajstić, N. *J. Math. Chem.*, submitted for publication (from ref 7).
(16) Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R.; Veith, G. D. *Math. Model* **1987**, *8*, 300.
(17) Randić, M.; Hansen, P. J.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60.
(18) Randić, M. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1988**, *28*, 60.
(19) Randić, M. In *MATH/CHEM/COMP 1987*, ed. Lacher, R. C. Studies in Physical Theoretical Chemistry; Elsevier: Amsterdam, 1988; Vol. 54, p 101.
(20) Bogdanov, B.; Nikolic, S.; Trinajstić, N. *J. Math. Chem.* **1989**, *3*, 299.
(21) Estrada, E.; Montero, L. A. *Mol. Eng.*, in press.
(22) Estrada, E. Work in progress.
(23) Rodriguez, L.; Estrada, E. Program STAR (Structural Topology and Applied Regression) Version 1.0. Universidad Central de Las Villas, 1993.
(24) Miller, K. J.; Savchic, J. A. *J. Am. Chem. Soc.* **1979**, *101*, 7206.
(25) *Handbook of Chemistry and Physics*, 61st ed.; CRC Press: Boca Raton, FL, 1980.
(26) Randić, M. *J. Math. Chem.* **1991**, *7*, 155.
(27) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam 1984.
(28) Kalizan, R. *Anal. Chem.* **1992**, *64*, 6194.
(29) Needham, N. D.; Wei, I.-C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186.
(30) Estrada, E. Work in progress.
(31) Estrada, E. Submitted for publication.