match no. 28 pp. 151-164 1992

DESIGN OF TOPOLOGICAL INDICES. PART 7. ANALYTICAL FORMULAE FOR LOCAL VERTEX INVARIANTS OF LINEAR AND MONOCYCLIC MOLECULAR GRAPHS[1]

Ovidiu Ivanciuc,[a] Teodor-Silviu Balaban,[b]
Petru Filip[b] and Alexandru T. Balaban[a]

[a] Polytechnic Institute, Faculty of Chemistry, Organic Chemistry Department, Splaiul Independentei 313, 77206 Bucharest, Roumania

[b] Institute of Chemistry of the Roumanian Academy Splaiul Independentei 202, Bucharest, Roumania

(Received: March 1992)

Analytical formulae are presented which enable the computation of a set of graph theoretical invariants, for example **AZDEG**-LOIS, for a generalized linear and cyclic molecular graph, representing linear and monocyclic chemical compounds.

1. INTRODUCTION

The central problem in theoretical chemistry is the deduction of molecular properties from the structural features of molecules. The first level of information about the molecular structure considers the atom-to-atom connectivity or topology of a molecule, disregarding geometrical features such as bond lengths, bond angles, and torsional angles.

Molecular topology determines a large number of molecular properties ranging from physico-chemical and thermodynamic properties to chemical reactivity and biological activity. In this respect, chemical applications of graph theory have undergone dramatic expansions in recent years.[2-7]

From a practical point of view, a convenient way of (at least, partly) characterizing (and thereby, in a way, expressing) the topology of a chemical structure is represented by a significant class of graph invariants, usually referred to as topological indices.[8-12] A topological index is a number which expresses some characteristic feature of the bonding topology of a molecule.

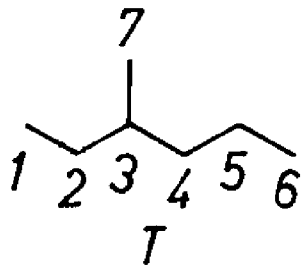
A graph (molecular graph) is a one-to-one mapping of a structure in which atoms are represented by points called vertices, and the chemical bonds are represented by lines (edges) joining the vertices. All N vertices of a graph G form the vertex set V = V(G), and the M edges are elements of the edges set E = E(G). A tree is a graph without cycles. A k-tree is a tree with the maximum degree k. In chemical graph theory alkanes are represented as 4-trees. DEG(i) denotes the degree of the vertex i in G, i.e. the number of edges

incident with the vertex i.

The topology of a chemical structure can be coded in matrix form by the use of the adjacency matrix.[13] The adjacency matrix of a graph G with N vertices, **A**(G) = **A**, is the square NxN symmetric matrix which contains information about the connectivity of vertices in G. Its entries are defined as:

$$(\mathbf{A})_{ij} = \begin{cases} 1, \text{ for vertices } i, j \text{ adjacent} \\ 0, \text{ otherwise} \end{cases}$$

The sum of entries over row i or column i in **A**(G) is DEG(i). As example, the molecular graph of 3-methylhexane (T) is given below. This graph is the smallest identity tree because its only symmetry operation is the identity (no vertices are equivalent), and it has the smallest number of vertices among all identity trees.



T

Another possibility of (partly) encoding the topology of a chemical graph uses graph theoretical polynomials,[14] which play a significant role in the area of topological chemistry, with applications to diverse areas such as the topological resonance energy (TRE)[15,16] and the topological effect on molecular orbitals (TEMO).[17] The characteristic polynomial of a graph is expressed as follows:

$$\mathrm{Ch}(x,G) = \det|x\cdot\mathbf{I} - \mathbf{A}| \qquad (1)$$

where **I** is the unit matrix. L(x,N) will denote the characteristic polynomial of a linear graph consisting of N vertices.

We will make use of the generalized characteristic polynomial, recently defined[18] and used in the context of spanning-tree enumeration[19,20] and Ulam's conjecture on reconstruction of the characteristic polynomial.[21] The generalized characteristic polynomial of a graph with N vertices is defined as:

$$Ch(\mathbf{X},N) = \det|\mathbf{X} - \mathbf{A}| \qquad (2)$$

where **X** is a diagonal NxN matrix of vertex parameters, with variable diagonal elements x_{ii} (where, in chemical applications, $x_{ii} \neq 0$) while $x_{ij} = 0$ if $i \neq j$.

For a linear graph with N vertices, $L(\mathbf{X},i:j)$ will denote the generalized characteristic polynomial of the subgraph of the linear graph which consists of the vertices between vertices i and j, including both these vertices. The matrix of vertex parameters **X** is labelled from i through j, and the element x_{kk} ($i \leq k \leq j$) is the parameter associated with the vertex k. If $i > j$ then by definition $L(\mathbf{X},i:j) = 1$, and if $i = j$, then $L(\mathbf{X},i:i) = x_{ii}$. Also, if $j = i + 1$, then $L(\mathbf{X},i:j) = x_{ii} \cdot x_{jj} - 1$.

For a linear graph consisting of four vertices, with the matrix of vertex parameters **X**:

$$\mathbf{X} = \begin{bmatrix} x_{11} & 0 & 0 & 0 \\ 0 & x_{22} & 0 & 0 \\ 0 & 0 & x_{33} & 0 \\ 0 & 0 & 0 & x_{44} \end{bmatrix}$$

the generalized characteristic polynomial is given by the following equation:

$$L(\mathbf{X},1:4) = x_{11}x_{22}x_{33}x_{44} - x_{11}x_{22} - x_{11}x_{44} - x_{33}x_{44} + 1$$

2. TOPOLOGICAL INDICES: GRAPH POTENTIALS

A new class of vertex topological indices, called graph potentials U(G), were defined as the solutions of the system of linear equations[22]

$$F \cdot U = C \qquad (3)$$

where F is a square NxN matrix and U and C are column vectors. The graph topological matrix $F(G)=(f_{ij})$ of the graph G is calculated according to the expression

$$f_{ij} = \begin{cases} 0 & \text{if } i \neq j \text{ and } e_{ij} \notin E(G) \\ -w_{ij} & \text{if } i \neq j \text{ and } e_{ij} \in E(G) \\ w_{ii} + \sum w_{ik} & \text{if } i = j \end{cases} \qquad (4)$$

where $w_{ii} > 0$ is the weight of the vertex i, and $w_{ij} > 0$ is the weight of the edge e_{ij}. In the simplest case all weights are set equal to 1, then

$$f_{ij} = \begin{cases} 0 & \text{if } i \neq j \text{ and } e_{ij} \notin E(G) \\ -1 & \text{if } i \neq j \text{ and } e_{ij} \in E(G) \\ 1 + DEG(i) & \text{if } i = j \end{cases} \qquad (5)$$

The matrix F defines the graph G up to isomorphism. Let A(G) be the adjacency matrix, I be the NxN identity matrix, and $B(G) = (b_{ij})$ the degree matrix, i.e., the diagonal matrix with $b_{ii} = DEG(i)$. Then the F matrix can be rewritten in the form

$$F = B - A + I \qquad (6)$$

Different vertex invariants can be selected as components of the column vector **C**. Due to the fact that the **F** matrix determinant is always greater than zero, equation (3) always has a unique solution and there always exists the inverse matrix $H = F^{-1}$.

Extensive computations showed that the diagonal of the **H** matrix

$$D = (h_{11}, h_{22}, \ldots, h_{NN}) \tag{7}$$

is a very sensitive vertex invariant that discriminates topologically nonequivalent vertices of many regular graphs. Therefore sensitive vertex potentials are obtained by the solution $U = H \cdot D^T$ of the system of linear equations

$$F \cdot U = D^T \tag{8}$$

where D^T (a column vector) is the transpose of **D**.

The system of linear equations (3), with weights all equal to 1, is exemplified for the tree T. The vector **C** was considered the vector D^T.

$$\begin{cases} 2u_1 - u_2 = 0.61473 ; & u_1 = 0.55340 \\ -u_1 + 3u_2 - u_3 = 0.45892 ; & u_2 = 0.49206 \\ -u_2 + 4u_3 - u_4 - u_7 = 0.36827 ; & u_3 = 0.46386 \\ -u_3 + 3u_4 - u_5 = 0.43909 ; & u_4 = 0.46714 \\ -u_4 + 3u_5 - u_6 = 0.47025 ; & u_5 = 0.49847 \\ -u_5 + 2u_6 = 0.61756 ; & u_6 = 0.55802 \\ -u_3 + 2u_7 = 0.59207 ; & u_7 = 0.52796 \end{cases}$$

3. LOCAL VERTEX INVARIANTS

Independently, a more general class of vertex topological indices were investigated.[23] These invariants were obtained as the solution of the system of linear equations

$$\mathbf{Q} \cdot \mathbf{S} = \mathbf{R} \qquad (9)$$

where **Q** is a matrix derived from a graph topological matrix, **R** is a column vector, and **S** is the column vector of LOcal Vertex Invariants (LOVI). Matrix **Q** is derived from a graph topological matrix **M**, by replacing its diagonal elements m_{ii} (associated with vertex i) with the components p_i of a nonzero column vector **P** describing a certain property of vertices in the graph G. The vertex property encoded in the column vectors **P** and **R** can be either topological, e.g. the vertex degrees, the distance sums, the total number of vertices in the graph, or chemical, e.g. the atomic number, electronegativity, ionization potential, etc. A simple way to denote the system (9) and the type of LOVI obtained is by a three-element notation **MPR**, where **M** represents the type of graph topological matrix, **P** the vertex property to be introduced on the main diagonal of the matrix **M**, and **R** the type of parameters used as free term. For example, the LOVI obtained using the adjacency matrix **A**, the atomic number Z as the property **P**, and the vertex degree DEG as property **R**, is denoted by **AZDEG**.

The system of linear equations (9) is exemplified for the tree T:

$$
\begin{cases}
6s_1 + s_2 = 1 ; s_1 = 0.12589 \\
s_1 + 6s_2 + s_3 = 2 ; s_2 = 0.24466 \\
s_2 + 6s_3 + s_4 + s_7 = 3 ; s_3 = 0.40614 \\
s_3 + 6s_4 + s_5 = 2 ; s_4 = 0.21954 \\
s_4 + 6s_5 + s_6 = 2 ; s_5 = 0.27665 \\
s_5 + 6s_6 = 1 ; s_6 = 0.12056 \\
s_3 + 6s_7 = 1 ; s_7 = 0.09898
\end{cases}
$$

By using various mathematical operations with the set of LOVI, a large number of topological indices is generated. Excellent correlations were obtained for the boiling point and vaporization enthalpies of alkanes using a topological index resulting from a simple summation of the local vertex invariants of the type **AZDEG**. This and other correlations, with NMR chemical shifts, liquid phase density, partial molal volumes, motor octane numbers of alkanes or cavity surface areas of alcohols, emphasize the potential of this approach.[23]

On the basis of LOVI, one may devise a wide variety of topological indices, characterizing the whole molecular graph by a single number. The devised topological indices are expected to correlate with experimental molecular data, in a similar manner as the indices obtained on the basis of **AZDEG**. This endeavour is being done in our group.

4. ANALYTICAL FORMULAE FOR LOVI

We will present, for vertex-weighted linear and monocyclic graphs, some analytical formulae for LOVI of the type **APR**, where **A** represents the adjacency matrix of the molecular graph, while the column vectors **P** and **R** represent certain properties of the vertices in the molecular graph.

We will solve analytically the linear system **APR** according to Cramer's rule:

$$LOVI_i = \frac{\Delta_i}{\Delta} \qquad (10)$$

where Δ is the determinant of the matrix **Q** of the linear system matrix of equations, and Δ_i is the determinant obtained from the matrix **Q** by replacing its i-th column with the column of the constant terms of the system, the vector **R**.

For a linear graph with N vertices of equal weight, all diagonal elements of the **X** matrix are equal, i.e. $x_{11} = x_{22} = \ldots = x_{NN} = x$, and the value of the determinant Δ is equal to:[25-28]

$$\Delta = L(x,N) \qquad (11)$$

while Δ_i is expressed by the equation:

$$\Delta_i = \sum_{j=1}^{i} (-1)^{i+j} \cdot R_j \cdot L(x,N-i) \cdot L(x,j-1) + \sum_{j=i+1}^{N} (-1)^{i+j} \cdot R_j \cdot L(x,i-1) \cdot L(x,N-j) \qquad (12)$$

It follows that:

$$LOVI_i = \left[\sum_{j=1}^{i} (-1)^{i+j} \cdot R_j \cdot L(x,N-i) \cdot L(x,j-1) + \right. \qquad (13)$$
$$\left. + \sum_{j=i+1}^{N} (-1)^{i+j} \cdot R_j \cdot L(x,i-1) \cdot L(x,N-j) \right] \cdot L(x,N)^{-1}$$

For a vertex-weighted linear graph, the analytical formula for LOVI is obtained in a similar manner:

$$LOVI_i = \left[\sum_{j=1}^{i} (-1)^{i+j} \cdot R_j \cdot L(X,i+1:N) \cdot L(X,1:j-1) + \right. \qquad (14)$$
$$\left. + \sum_{j=i+1}^{N} (-1)^{i+j} \cdot R_j \cdot L(X,1:i-1) \cdot L(X,j+1:N) \right] \cdot L(X,1:N)^{-1}$$

TABLE 1 presents the **AZDEG**-LOVI values for a few linear molecular graphs of N-alkanes (for carbon, Z = 6) and of their heteroatom derivatives: amines, alcohols and ethers (for nitrogen and oxigen, Z values are 7 and 8, respectively).

TABLE 1. AZDEG LOVI for linear molecular graphs containing C, N, and O.

Graph	AZDEG LOVI					
C-C-C	0.1176	0.2941	0.1176			
C-C-C-C	0.1220	0.2683	0.2683	0.1220		
C-C-C-C-C	0.1212	0.2727	0.2424	0.2727	0.1212	
C-C-C-C-C-C	0.1213	0.2720	0.2469	0.2469	0.2720	0.1213
N-C-C-C-C	0.1035	0.2758	0.2419	0.2728	0.1212	
C-N-C-C-C	0.1280	0.2318	0.2494	0.2715	0.1214	
C-C-N-C-C	0.1202	0.2790	0.2060	0.2790	0.1202	
O-C-C-C-C	0.0902	0.2780	0.2415	0.2729	0.1212	
C-O-C-C-C	0.1331	0.2015	0.2546	0.2706	0.1216	
C-C-O-C-C	0.1194	0.2836	0.1791	0.2836	0.1194	

For a cyclic graph, denoted by C_N, consisting of N vertices of the same type, we can express Δ with a recurrence formula:

$$\Delta(C_N) = L(x,N) - L(x,N-2) - 2\cdot(-1)^N \qquad (15)$$

Hence we obtain the following expression for LOVI:

$$LOVI_1 = \left\{ R_1 \cdot L(x,N-1) + \sum_{j=2}^{N} (-1)^{j+1} \cdot R_j \left[L(x,N-j) + (-1)^N \cdot L(x,j-2) \right] \right\} \cdot \left[L(x,N) - L(x,N-2) - 2\cdot(-1)^N \right]^{-1} \qquad (16)$$

For the vertex-weighted monocyclic graph we obtain the following analytical formula for LOVI:

$$LOVI_1 = \left\{ R_1 \cdot L(X,2{:}N) + \sum_{j=2}^{N} (-1)^{j+1} \cdot R_j \left[L(X,j+1{:}N) + (-1)^N \cdot L(X,2{:}j-1) \right] \right\} \cdot \left[L(X,1{:}N) - L(X,2{:}N-1) - 2 \cdot (-1)^N \right]^{-1} \quad (17)$$

By a cyclic permutation of the numbering of vertices in the cycle, all LOVI values may be computed.

TABLE 2 presents the **AZDEG**-LOVI values for five- and six-membered monocyclic heterocompounds represented by graphs G_1 and G_2, with one heteroatom Y, where Y represents B, N, and O, respectively. All the **AZDEG**-LOVI values for cycloalkanes are equal to 0.25, irrespective of the number of carbon atoms in the cycle.

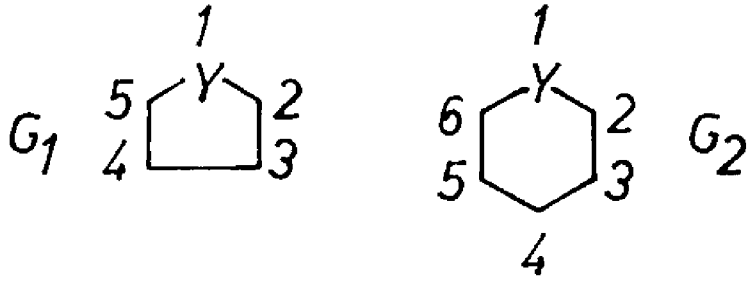


TABLE 2. AZDEG LOVI for five- and six-membered monocyclic molecular graphs containing B, C, N and O.

Graph	AZDEG LOVI					
$c(B\text{-}C_4)$	0.3037	0.2408	0.2513	0.2513	0.2408	
$c(B\text{-}C_5)$	0.3037	0.2408	0.2516	0.2495	0.2516	0.2408
$c(N\text{-}C_4)$	0.2125	0.2564	0.2491	0.2491	0.2564	
$c(N\text{-}C_5)$	0.2124	0.2564	0.2489	0.2504	0.2489	0.2564
$c(O\text{-}C_4)$	0.1847	0.2611	0.2484	0.2484	0.2611	
$c(O\text{-}C_5)$	0.1847	0.2612	0.2480	0.2507	0.2480	0.2612

REFERENCES

1. Part 6: O. Ivanciuc and A.T. Balaban, J. Math. Chem. in press.
2. N. Trinajstić, Chemical Graph Theory, CRC Press, Boca Raton, Florida, 1983.
3. A.T. Balaban, Editor, Chemical Applications of Graph Theory, Academic Press, London, 1976.
4. J.W. Kennedy and L.V. Quintas, Editors, Applications of Graphs in Chemistry and Physics, North Holland, Amsterdam, 1988.
5. A.T. Balaban, J. Chem. Inf. Comput. Sci. **25**, 334 (1985); A.T. Balaban, J. Mol. Struct. (Theochem) **120**, 117 (1985); A.T. Balaban, J. Mol. Struct. (Theochem) **165**, 243 (1988).
6. N. Trinajstić, D.J. Klein and M. Randić, Int. J. Quantum Chem.: Quantum Chem. Symp **20**, 699 (1986).
7. K. Balasubramanian, Chem. Rev. **85**, 599 (1985).
8. A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Top. Curr. Chem. **114**, 21 (1983).
9. P.J. Hansen and P.C. Jurs, J. Chem. Educ. **65**, 574 (1988).
10. D.H. Rouvray, Acta Pharm. Jugosl. **36**, 239 (1986); D.H. Rouvray, J. Comput. Chem. **8**, 470 (1987); D.H. Rouvray, J. Mol. Struct. (Theochem) **185**, 187 (1989).
11. L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic, New York, 1976.
12. L.B. Kier and L.H. Hall, Molecular Connectivity in Structure-Activity Analysis, Wiley, New York, 1986.
13. J.J. Sylvester, Am. J. Math., **1**, 64 (1874).
14. J.V. Knop and N. Trinajstić, Int. J. Quantum Chem.: Quantum Chem. Symp. **14**, 503 (1980).

15. I. Gutman, M. Milun and N. Trinajstić, J. Am. Chem. Soc. **99**, 1692 (1977).

16. J. Aihara, J. Am. Chem. Soc. **98**, 2750 (1976).

17. O.E. Polansky and M. Zander, J. Mol. Struct. **84**, 361 (1982); O.E. Polansky, G. Mark and M. Zander, Der topologische Effekt an Molekülorbitalen (TEMO), Schriftenreihe des Max-Planck-Instituts für Strahlenchemie, Mülheim a.d. Ruhr, 1987.

18. D. Cvetković, M. Doob and H. Sachs, Spectra of Graphs: Theory and Application, Academic Press, London, 1980, p. 30.

19. I. Gutman, R.B. Mallion and J.W. Essam, Mol. Phys. **50**, 859 (1983).

20. B. O'Leary and R.B. Mallion, in Graph Theory and Topology in Chemistry, R.B. King and D.H. Rouvray (Eds.), Elsevier, Amsterdam, 1987, pp. 544-555.

21. P. Krivka, R.B. Mallion and N. Trinajstić, J. Mol. Struct. (Theochem) **164**, 363 (1988).

22. V.E. Golender, V.V. Drboglav and A.B. Rozenblit, J. Chem. Inf. Comput. Sci. **21**, 196 (1981).

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

24. A.T. Balaban and V. Feroiu, Reports Mol. Theory **1**, 133 (1990).

25. C.L. Coates, IRE Trans. Circuit Theory CT-6, 170 (1959).

26. F. Harary, SIAM Rev. **4**, 202 (1962).

27. D. Cvetković, Matemat. Vesnik **12**, 333 (1975).

28. A. Graovac and I. Gutman, MATCH **6**, 49 (1979).