match no. 30 pp. 141-152 1994

DESIGN OF TOPOLOGICAL INDICES. PART 8.[1] PATH MATRICES AND DERIVED MOLECULAR GRAPH INVARIANTS

Ovidiu Ivanciuc and Alexandru T. Balaban

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

(received: May 1992)
first version

ABSTRACT

The maximum path matrix of a molecular graph and derived vertex and molecular graph theoretical invariants are defined, and their relation with previously defined graph invariants is discussed.

1. INTRODUCTION

Chemical graph theory[2-5] is usually applied to the graphical representation of chemical structures (molecules) whose atoms are depicted by points (vertices) while the chemical bonds are represented by lines (edges) connecting the points. Commonly, hydrogen atoms are ignored, and the resulting hydrogen-suppressed graph is used.

The bond lengths, bond angles, and stereoisomerism are not taken into account in the graph representation of chemical structures, and the adjacency of atoms, i.e. the constitution of the molecule, is regarded as being their most important topological feature.

From a practical point of view, a convenient way of expressing the topology of a chemical structure is represented by a significant class of graph invariants, usually referred to as topological indices.[6-8] A *topological index* is a number that is computed by a mathematical algorithm from the constitutional graph of a molecule. Since isomorphic graphs possess identical values for any given topological index, these indices are referred to as *graph invariants*. They do not characterize graphs up to isomorphism. Topological indices usually reflect both molecular size and shape.

A few useful graph theoretical definitions will be introduced. A *graph* G is an ordered pair consisting of two sets V = V(G) and E = E(G). Elements of the set V(G) are called *vertices* and elements of the set E(G), involving a binary relation (adjacency) between the vertices, are called *edges*. The number of elements of the set V(G) is denoted by N, the number of vertices in the molecular graph. Two vertices v_i and v_j of a graph G are said to be *adjacent* if there is an edge joining them; the vertices v_i and v_j (edge endpoints) are then said to be *incident* to such an edge. Similarly, two distinct edges of G are adjacent if they have at least one vertex in common.

A *walk* in a graph is a sequence of edges, starting from any vertex and ending on any vertex, with the property that the ending vertex of the jth edge in the sequence is the

beginning of the (j+1)th edge. A walk is called a *path* if all the vertices (and thus necessarily all the edges) are distinct.

The length of the shortest path which connects two vertices v_i and v_j in a graph G is denoted by d_{ij} and is called the *topological distance* (or the *distance*) between these two vertices; of course, $d_{ii} = 0$. The distances $d_{ij} = (D)_{ij}$ represent the elements of the distance matrix $D = D(G)$ of the graph G.

We consider the molecular graph to be arbitrarily numbered from 1 to N, where N is the number of vertices. Although the distance matrix is not invariant to different numberings of the molecular graph, the various graph theoretic invariants which will be used or defined in this paper will exhibit constant values for all N! possible numberings of the vertices in the molecular graph.

The *distance sum*[9,10] of the vertex v_i of the graph G, denoted by DS(G,i), is defined as the sum of the topological distances between vertex v_i and every vertex in the molecular graph, i.e. the sum over row i or column i in D(G).

$$DS(G,i) = \sum_{j=1}^{N} (D)_{ij}$$

Wiener[11-13] has introduced in his studies on additive physical parameters of molecules the topological index W that is equal to the half-sum of the elements of the distance matrix D. The Wiener index may be expressed also as the half-sum of the distance sums of the molecular graph:

$$W = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (D)_{ij} = \frac{1}{2} \sum_{i=1}^{N} DS(G,i)$$

The Wiener index is widely used in devising correlations between the chemical structure and various physico-chemical properties of classes of substances. A number of variations of the Wiener index were defined, e.g. on the basis of information theory[14,15] or as expanded Wiener number.[16] The applications of the distance matrix and of the Wiener index were recently reviewed.[17]

2. DEFINITION OF PATH MATRICES

The length of the longest path connecting two vertices v_i and v_j is called the *elongation* between the two vertices.[18,19].

We define the *maximum path matrix* of a graph G with N vertices, MP(G) = MP, the square NxN symmetrical matrix, whose entries, $(MP)_{ij}$, are equal to the elongation between the two vertices v_i and v_j. All the elements of type $MP_{ii}(G)$ are, by definition, zero. Thus, all the entries of the maximum path matrix are integers.

$$(MP)_{ij} = \begin{cases} \text{if } i \neq j, & \text{length of the elongation} \\ & \text{between vertices } v_i \text{ and } v_j \\ \text{if } i = j, & 0 \end{cases}$$

As an example, the maximum path matrix of the molecular graph of 1-ethyl-2-methylcyclopropane (graph G_4 in Figure 1) is given below. This molecular graph is the smallest identity monocyclic graph, because it is the smallest monocyclic graph with no two topologically equivalent vertices, hence it has the identity operation as the only symmetry element.

$$MP(G_4) = \begin{bmatrix} 0 & 2 & 2 & 1 & 2 & 3 \\ 2 & 0 & 2 & 3 & 4 & 1 \\ 2 & 2 & 0 & 3 & 4 & 3 \\ 1 & 3 & 3 & 0 & 1 & 4 \\ 2 & 4 & 4 & 1 & 0 & 5 \\ 3 & 1 & 3 & 4 & 5 & 0 \end{bmatrix}$$

The *maximum/minimum path matrix* of a graph G with N vertices, **MmP**(G) = **MmP**, is a square NxN nonsymmetrical matrix, whose entries, $(\mathbf{MmP})_{ij}$, are defined as:

$$(\mathbf{MmP})_{ij} = \begin{cases} \text{if } i > j, & \text{length of the longest path (elongation)} \\ & \text{between vertices } v_i \text{ and } v_j \\ \text{if } i = j, & 0 \\ \text{if } i < j, & \text{length of the shortest path (distance)} \\ & \text{between vertices } v_i \text{ and } v_j \end{cases}$$

From the definition of the **MmP** matrix we easily observe that the upper triangle of the matrix is identical with the one of the maximum path matrix, while the lower triangle of the matrix coincides with the lower triangle of the distance matrix.

As example, the maximum/minimum path matrix of the graph G_4 is given below.

$$\mathbf{MmP}(G_4) = \begin{bmatrix} 0 & 2 & 2 & 1 & 2 & 3 \\ 1 & 0 & 2 & 3 & 4 & 1 \\ 1 & 1 & 0 & 3 & 4 & 3 \\ 1 & 2 & 2 & 0 & 1 & 4 \\ 2 & 3 & 3 & 1 & 0 & 5 \\ 2 & 1 & 2 & 3 & 4 & 0 \end{bmatrix}$$

We have to note that both the maximum path matrix andthe maximum/minimum path matrix are identical to the distance

matrix for all acyclic molecular graphs.

3. VERTEX THEORETICAL INVARIANTS BASED ON PATH MATRICES

The *maximum path sum* of the vertex v_i of the graph G, denoted by MPVS(G,i), is defined as the sum of the length of the longest path between vertex v_i and every vertex in the molecular graph, i.e. the sum over row i or column i in MP(G).

$$MPVS(G,i) = \sum_{j=1}^{N} (MP)_{ij}$$

The *maximum/minimum path sum* of the vertex v_i of the graph G, denoted by MmPVS(G,i), is defined as the sum of the length of the longest and shortest paths between vertex i and every vertex in the molecular graph, i.e. the sum over row i and column i in MmP(G), or, alternatively, the sum of the distance sum and maximum path sum of the vertex v_i:

$$MmPVS(G,i) = \sum_{j=1}^{N} (MmP)_{ij} + \sum_{j=1}^{N} (MmP)_{ji} = DS(G,i) + MPVS(G,i)$$

Table 1 presents the MPVS, DS and MmPVS for the six cyclic graphs in Figure 1. There exist molecular graphs with topologically inequivalent vertices exhibiting identical local vertex invariants. Such a degeneracy appears for G_1, where $MPVS(G_1,3) = MPVS(G_1,4)$, but $DS(G_1,3) \neq DS(G_1,4)$, and $MmPVS(G_1,3) \neq MmPVS(G_1,4)$. DS presents also degenerate values for G_4, where $MPVS(G_4,3) \neq MPVS(G_4,4)$, but $DS(G_4,3) = DS(G_4,4)$ and $MmPVS(G_4,3) \neq MmPVS(G_4,4)$.

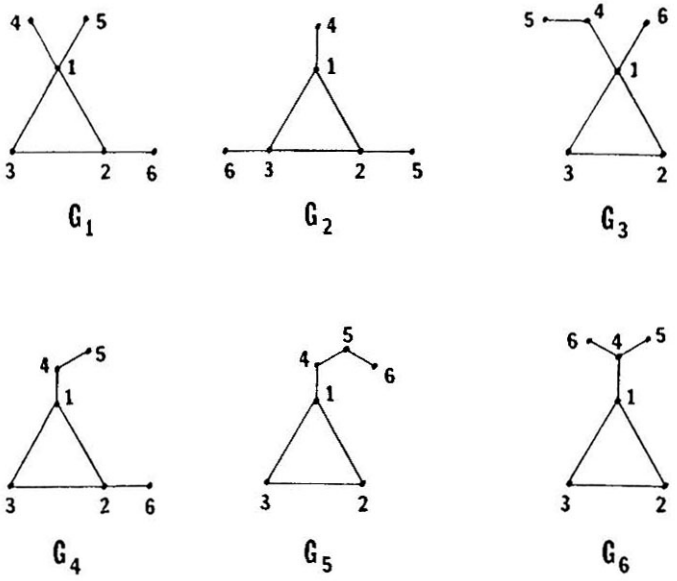

Fig. 1. Molecular graphs of the six 3-membered cycloalkanes used in computing the graph-theoretical invariants presented in Table 1.

Let n_{ij} be the number of vertices in the molecular graph G situated at the maximum path of length j from the vertex v_i; then the sequence $(n_{i0}, n_{i1}, n_{i2}, \ldots, n_{ik})$ is called the *maximum path degree sequence* of the vertex v_i in G, and k represents the length of the longest path in G; note that $n_{i0} = 1$. The maximum path degree sequence of the vertex v_i in the graph G is denoted by MPDS(G,i).

The maximum path degree sequence of vertex 1 in graph G_4 is given below:

$$MPDS(G_4,1) = (1, 1, 3, 1, 0, 0)$$

Let m_i be the number of maximum paths of length i in the molecular graph G; then the sequence $(m_0, m_1, m_2, \ldots, m_k)$ is called the *maximum path frequency sequence*, and k represents the length of the longest path (maximum elongation) in G; note that $m_0 = N$, the number of vertices in G. The maximum path frequency sequence of the graph G is denoted by MPFS(G).

The maximum path frequency sequence of the graph G_4 is:

$$MPFS(G_4) = (6, 3, 4, 4, 3, 1)$$

4. GRAPH THEORETICAL INVARIANTS BASED ON PATH MATRICES

By analogy with the Wiener index, we define the MPS topological index as the sum of the number of bonds on the longest path between any two vertices in the molecular graph, i.e. the half-sum of the elements of the maximum path matrix **MP**, or the half-sum of the maximum path sums MPVS of the molecular graph:

$$MPS(G) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{MP})_{ij} = \frac{1}{2}\sum_{i=1}^{N} MPVS(G,i)$$

The MmPS topological index is defined as the sum of the longest and shortest path between any two vertices in the molecular graph, i.e. the sum of the elements of the maximum/minumum path matrix **MmP**, or the sum of the maximum/minimum path sums MmPVS of the molecular graph, or, equally well, the sum of the Wiener index and MPS topological index of the molecular graph:

$$MmPS(G) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{MmP})_{ij} = \frac{1}{2}\sum_{i=1}^{N} MmPVS(G,i) = W(G) + MPS(G)$$

We have to note that if G is an acyclic graph, then

$$\mathbf{MPS}(G) = \mathbf{MmPS}(G)/2 = \mathbf{W}(G)$$

The topological indices W, MPS, and MmPS of the six cyclic graphs in Figure 1 are presented in Table 1. As is apparent from the Table 1, some degenerate values for indices W and MPS appear, but when they are considered together in MmPS, they give different numerical values. For example, $W(G_2) = W(G_3)$, $MPS(G_2) \neq MPS(G_3)$, and $MmPS(G_2) \neq MmPS(G_3)$. In another example, $W(G_4) \neq W(G_5)$, $MPS(G_4) = MPS(G_5)$, and $MmPS(G_4) \neq MmPS(G_5)$. However, the reverse situation may appear, when both W and MPS indices are different for two graphs, but MmPS has identical values: $W(G_1) \neq W(G_3)$, $MPS(G_1) \neq MPS(G_3)$, and $MmPS(G_1) = MmPS(G_3)$.

Since both the local vertex invariants and the resulting topological indices are integers, these belong to the so-called "first generation" topological indices.[20]

5. CONCLUSIONS

Two molecular graph matrices were defined, namely the maximum path matrix and the maximum/minimum path matrix. Based on the path matrices, several vertex invariants and topological indices were defined, showing interesting properties when compared with graph invariants based on the distance matrix.

REFERENCES

1. Part 7: O. Ivanciuc, T.-S. Balaban, P. Filip and A.T. Balaban, *MATCH*, in press.
2. A.T. Balaban, Editor, *Chemical Applications of Graph Theory*, Academic Press, London, 1976.
3. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, Florida, 1983.
4. 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).
5. N. Trinajstić, D.J. Klein and M. Randić, *Int. J. Quantum Chem.: Quantum Chem. Symp*, **20**, 699 (1986).
6. A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, *Top. Curr. Chem.*, **114**, 21 (1983).
7. P.J. Hansen and P.C. Jurs, *J. Chem. Educ.*, **65**, 574 (1988).
8. 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).

9. A.T. Balaban, *Chem. Phys. Lett.*, **89**, 399 (1982).
10. P.G. Seybold, *Int. J. Quantum Chem., Quantum Biol. Symp.* **10**, 95, 103 (1983).
11. H. Wiener, *J. Am. Chem. Soc.*, **69**, 17, 2636 (1947).
12. H. Wiener, *J. Chem. Phys.*, **15**, 766 (1947).
13. H. Wiener, *J. Phys. Chem.*, **52**, 425, 1082 (1948).
14. D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, **67**, 4517 (1977).
15. D. Bonchev, *Information-Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press, Chichester, 1983.
16. S.S. Tratch, M.I. Stankevitch and N.S. Zefirov, *J. Comput. Chem.*, **11**, 899 (1990).
17. D.H. Rouvray, in *Mathematical and Computational Concepts in Chemistry*, Ed. N. Trinajstić, Ellis Horwood, Chichester, 1986, Chapter 25, p 295.
18. O. Ore, *Theory of Graphs*, Am. Math. Soc. Colloquium Publ., Vol. 38, Am. Math. Soc., Providence, 1962.
19. V.A. Skorobogatov and A.A. Dobrynin, *MATCH*, **23**, 105 (1988).
20. A.T. Balaban, *J. Chem. Inf. Comput. Sci.*, **32**, 23 (1992).

Table 1. Vertex theoretical indices and topological indices of alkyl-substituted cyclopropanes depicted in Figure 1.

	Vertex label						W	MPS	MmPS
	1	2	3	4	5	6			
G_1 1,1,3-Trimethyl-cyclopropane							26	37	63
MPVS	9	11	13	13	13	15			
DS	6	7	8	10	10	11			
MmPVS	15	18	21	23	23	26			
G_2 1,2,3-Trimethyl-cyclopropane							27	39	66
MPVS	11	11	11	15	15	15			
DS	7	7	7	11	11	11			
MmPVS	18	18	18	26	26	26			
G_3 1-Ethyl-1-methyl-cyclopropane							27	36	63
MPVS	8	14	14	10	14	12			
DS	6	9	9	8	12	10			
MmPVS	14	23	23	18	26	22			
G_4 1-Ethyl-2-methyl-cyclopropane							29	40	69
MPVS	10	12	14	12	16	16			
DS	7	8	9	9	13	12			
MmPVS	17	20	23	21	29	28			
G_5 Propyl-cyclopropane							31	40	71
MPVS	10	16	16	10	12	16			
DS	8	11	11	8	10	14			
MmPVS	18	27	27	18	22	30			
G_6 Isopropyl-cyclopropane							28	37	65
MPVS	9	15	15	9	13	13			
DS	7	10	10	7	11	11			
MmPVS	16	25	25	16	24	24			