Molecular Topology. 14. Molord Algorithm and Real Number Subgraph Invariants

Mircea V. Diudea, Dragos Horvath

Department of Chemistry, »Babes-Bolyai« University, Cluj, Romania

and

Danail Bonchev

Higher School of Chemical Technology, 8010 Burgas, Bulgaria

An algorithm, MOLORD, is proposed for defining real number invariants for subgraphs of various sizes in molecular graphs. The algorithm is based on iterative line derivatives and accounts for heteroatoms by means of their electronegativities. It can be used in topological equivalence perception as well as to provide local and global descriptors for QSPR or QSAR studies. The algorithm is implemented on a TURBO PASCAL, TOPIND program and examplified on a set of selected graphs.

INTRODUCTION

In a previous paper of this series,[1] we presented a powerful algorithm, called MOLCEN, which allows us to find the centricity and centrocomplexity of subgraphs of various size in a graph G. MOLCEN made use od iterative line derivatives,[2] $L_n(G)$, LOVIs (LOcal Vertex Invariants), and TIs (Topological Indices) derived on the ground of layer matrices, LM.

A layer matrix, LM,[3–5] collects the properties (topological and chemical) of vertices u located on concentric shells (layers) at a distance j around each vertex $i \in$ G. The j-th layer of vertex i, $G(u)_j$ is defined by

$$G(u)_j = \left\{ u: d_{iu} = j \right\} \tag{1}$$

whereas the layer matrix entries are denoted by

$$lm_{i,j} = \underset{u \in G(u)_j}{\Omega} m_u . \tag{2}$$

Hence, the layer matrix can be written as

$$\mathbf{LM}(G) = \left\{ lm_{i,j} \, ; i \in [l,n] \, ; j \in [0,d] \right\} \tag{3}$$

where:

m, M are labels for a given property and the corresponding matrix, respectively

Ω stands for the mathematical operator acting on the vertices $u \in G(u)j$, at each j-layer; usually Ω is Σ, and

d is the diameter of G, *i.e.* the largest distance in the graph.

We defined[4] two types of LOVIs on the $\mathbf{LM}$ matrices: one of centricity, $c(\mathbf{LM})_i$, and the other of centrocomplexity, $x(\mathbf{LM})_i$, according to Eqs. (4) and (5)

$$c(\mathbf{LM})_i = \left[\sum_{j=1}^{ecc_i} (lm_{ij})^{j/dsp} \right]^{-1} \tag{4}$$

$$x(\mathbf{LM})_i = \left[\sum_{j=0}^{ecc_i} lm_{ij} \cdot 10^{-zj} \right]^{-1} \cdot t_i \tag{5}$$

where:

dsp is a specified topological distance, usually larger than the graph diameter (within this work, $dsp = 10$)

z is the number of digits of the (integer part of) max lm_{ij} – value in the graph

t_i is a weighting factor, accounting for heteroatoms (*e.g.* a Sanderson type of electronegativity,[6] see below)

Summation of the LOVIs values over all i vertices in G provides the corresponding global indices (TIs), denoted by capital letters.

Within the MOLCEN algorithm, the values of the above mentioned LOVIs are normalized within the range (0 – 1) by being divided by their largest value in the graph. They are useful in the centric ordering of vertices in a graph and in the topological equivalence perception. However, these LOVIs are not useful for the characterization of similar local neighbourings within a set of molecular graphs.

The present paper offers a modified algorithm, called MOLORD, which provides a spectrum of non-normalized LOVIs and TIs. The new algorithm is examplified on sone representative molecular graphs. As the algorithm is based on iterative line derivatives of a graph, formulas are given for the calculation of the vertex degree and number of edges in line derivatives of regular graphs. A set of topological indices given by MOLORD was tested for correlating ability with the critical pressure of octane isomers.

LINE DERIVATIVES OF GRAPHS

The line derivative of a graph is obtained by representing its lines by points, and then by joining two such points with a line, if the lines they represent were adjacent in the original graph G (which is zero order derivative, L_0). By repeating this procedure n times, one obtains the iterative line derivative of the order n, L_n. The number of vertices p_{n+1} and edges q_{n+1} in L_{n+1} is given by the following relations:[2,7]

$$p_{n+1} = q_n \tag{6}$$

$$q_{n+1} = -q_n + 1/2 \sum_{i \in L_n} (k_i)^2 \tag{7}$$

$$q_{n+1} = \sum_{i \in L_n} \binom{k_i}{2} = \sum_{i \in L_n} k_i(k_i - 1)/2 = B_n \tag{8}$$

with k_i being the vertex degrees and B_n – Bertz's branching index,[2] which is just the number of edges in the L_n derivative.

In regular graphs (*i.e.* graphs in which all vertices have the same degree), the number of edges q_{n+1} can be calculated by a recursive relation derived from Eq. (7) or Eq. (8) by substituting the value for the vertex degree (see also Ref. 2):

$$k_n = 2q_n/p_n = 2q_n/q_{n-1} \tag{9}$$

$$q_{n+1} = -q_n + 2q_n^2/q_{n-1} . \tag{10}$$

The number of edges in L_{n+1} can also be calculated by

$$q_{n+1} = 1/2 \cdot k_{n+1} \cdot p_{n+1} . \tag{11}$$

Since in regular graphs

$$k_{n+1} = 2(k_n - 1) \tag{12}$$

and taking into account Eq. (6), Eq. (11) becomes

$$q_{n+1} = q_n(k_n - 1) . \tag{13}$$

From relations (12) and (13) we were able to derive relationships for computing k_n and q_n from the starting parameters, k_0 and q_0 only (*i.e.* from the degree and number of edges in the initial graph, L_0)

$$k_n = 1 + 2^n \cdot k_o - \sum_{e=0}^{n} 2^e \tag{14}$$

$$q_n = q_o \cdot \prod_{m=0}^{n-1} (k_m - 1) = q_0 \cdot \prod_{m=0}^{n-1} \left(2^m \cdot k_o - \sum_{e=0}^{m} 2^e \right) . \tag{15}$$

MOLORD ALGORITHM

The aim of the MOLORD algorithm is to characterize vertices or subgraphs (of various size) of the initial graph G by means of invariants derived from the topology of a series of line derivatives of this graph, $L_0(=G)$, $L_1....,L_m$. Before detailing the algorithm, some notations need to be defined.

First, recall that the vertices i_n of graph L_n (*i.e.* the current derivative graph) denote pairs of vertices of the lower-order derivative graph, L_{n-1}: ·

$$i_n = (j_{n-1}, k_{n-1}) \tag{16}$$

where the two points j and k in L_{n-1} are necessarilly connected by an edge of this graph and are themselves pairs of vertices of graph L_{n-2}. We can write that $j_{n-1} \in j_n$ and $k_{n-1} \in i_n$. The relatedness of vertices (subgraphs) in the process of derivatization can be expressed by

$$\delta(i_n, i_{n+1}) = \begin{vmatrix} 1 & \text{if } (i_n \in i_{n+1}) \\ 0 & \text{otherwise} \end{vmatrix} \tag{17}$$

The definition can be easily extended for any two arbitrary ranks n and $m \geq n$, stating that $\delta(i_n, i_m) = 1$ only if the vertex n_i appears in at least one of the subsets defining vertex i_m. In going back to L_0, one can see that i_n denotes a subgraph, consisting of n edges, in L_0.

The algorithm consists of the following four steps:

Step 1, computes local, $I(i_n)$, and global, $GI(L_n)$ classical invariants on each L_n within the set of derivative graphs L_0 to L_m:

$$GI(L_n) = \sum_{i_n} I(i_n) . \tag{18}$$

Step 2, evaluates a partial local invariant $PI_m(i_n)$ of a vertex i_n with respect to the mth order derivative graph, L_m

$$PI_m(i_n) = \frac{GI(L_n)}{GI(L_m)} \cdot \sum_{i_m} I(i_m) \cdot \delta(i_n, i_m) . \tag{19}$$

Here, $I(i_m)$ denotes the function used to calculate a given local invariant of vertex i_m with respect to the topology of graph L_m containing it. Further, the partial invariant of i_n with respect to L_m is calculated by summing up all the local invariants $I(i_m)$ of those vertices in L_m which are »related« to i_n according to the m–n successive derivatives, $L_n,...,L_m$. The ratio $GI(L_n)/GI)(L_m)$ is used as a scaling factor meant to ensure that the resulting PI values can be compared with each other irrespective current L_m for which they are evaluated.

Step 3, computes a synthetic local invariant of vertex i_n, in a series of successive derivative graphs, $L_n,...,L_m$:

$$SI_m(i_n) = \sum_{k=n}^{m} PI_k(i_n) \cdot f^{n-k} . \tag{20}$$

Subscript m in $SI_m(i_n)$ means that the last L_m derivative graph has been taken into account. The empirical factor f can be used to give a different weighting to the contributions arising from derivatives of various ranks. Note that in the case of $n = m$, the synthetic invariant $SI_m(i_n)$ reduces to the classical invariant $I(i_n)$.

Step 4, evaluates the final expression for the global synthetic index of a graph L_n

$$GSI_m(L_n) = \sum_{i_n} SI_m(i_n) . \tag{21}$$

Within this work, the algorithm is examplified by using LOVIs and TIs derived on the ground of the **LDS** (layer matrix of distance sum)[4,8,9] whose entries are given by

$$lm_{i,j} = \sum_{u \in G(u)_j} DS_u \tag{22}$$

with DS_u being the sum of distances from vertex u to all other vertices in the graph. In calculating $x(LDS_i)$ (see Eq. (5)), the values of lm_{ij} are divided by the values of the vertex degree, k_i.

MOLORD algorithm provides a spectrum of local, $SI_m(i_n)$ (per subgraph of various size) and global values $GSI_m(L_n)$, (with m varying from n to a selected m) for a given topological index, TI. It is examplified on 2-methylbutane (**1**), for which Figure 1 shows the derivative graphs L_0 to L_3 along with the corrresponding **LDS** matrices.

Note that the subgraph 1235 of L_0 is characterized by three points in $L_3(1)$ so that the value 2.11958 represents the sum of their corresponding values.

CENTRIC ORDERING OF VERTICES (SUBGRAPHS)

As pointed out above, the MOLORD algorithm is capable of ordering the vertices in molecular graphs either in terms of the centricity »c« or centrocomplexity »x«. Table I lists the centric ordering of the subgraphs (consisting of 0 to 3 edges) of 2,2-dimethylnonane (**2**) induced by $c(LDS)_i$ (values $SI_m(i_n)$, considered in decreasing order).

In order to obtain a comparison term for the above centric ordering of subgraphs in graph **2**, consider the *1D – 3D centric criteria* of Bonchev et al.[10] They are as follows

$L_0\,(\mathbf{1})$ $L_1\,(\mathbf{1})$ $L_2\,(\mathbf{1})$ $L_3\,(\mathbf{1})$

$L_0(\mathbf{1})$	$L_1(\mathbf{1})$	$L_2(\mathbf{1})$	$L_3(\mathbf{1})$
$\mathbf{\textit{LDS}}(L_0(1))$	$\mathbf{\textit{LDS}}(L_1(1))$	$\mathbf{\textit{LDS}}(L_2(1))$	$\mathbf{\textit{LDS}}(L_3(1))$

1 8 5 14 9				
2 5 22 9 0	(1,2) 4 7 5	(1,2,3) 3 11 0	(1,2,3,4) 5 14 5	
3 6 14 16 0	(2,3) 3 13 0	(1,2,5) 4 6 4	(1,2,3,5) 4 20 0	
4 9 6 5 16	(2,5) 4 7 5	(2,3,4) 4 6 4	(1,2,3,5) 5 14 5	
5 8 5 14 9	(3,4) 5 3 8	(2,3,5) 3 11 0	(1,2,3,5) 5 14 5	
			(2,3,4,5) 5 14 5	

Step 1; values $I(i_n)$

1 0.12420	(1,2) 0.49134	(1,2,3) 0.96463	(1,2,3,4) 0.58360
2 0.57461	(2,3) 0.95847	(1,2,5) 0.49256	(1,2,3,5) 2.11958
3 0.32565	(2,5) 0.49134	(2,3,4) 0.49256	(2,3,4,5) 0.58360
4 0.11037	(3,4) 0.19878	(2,3,5) 0.96463	
5 0.12420			

$GI(L_n)$ 1.25903   2.13992   2.91439   3.28678

Step 2; values $PI_2(i_1)$

(1,2) $(0.96463 + 0.49256) \times (2.13992/2.91439) = 1.06996$
(2,3) $(0.96463 + 0.49256 + 0.96463) \times (2.13992/2.91439) = 1.77825$
(2,5) $(0.96463 + 0.49256) \times (2.13992/2.91439) = 1.06996$
(3,4) $0.49256 \times (2.13992/2.91439) = 0.36167$

values $PI_3(i_1)$

(1,2) $(0.58360 + 2.11958) \times (2.13992/3.28678) = 1.75996$
(2,3) $(0.58360 + 2.11958 + 0.58360) \times (2.13992/3.28678) = 2.13992$
(2,5) $(0.58360 + 2.11958) \times (2.13992/3.28678) = 1.75996$
(3,4) $(0.58360 + 0.58360) \times (2.13992/3.28678) = 0.75993$

Steps 3 and 4; values $SI_m(i_1)$ and $GSI_m(L_1)$; $f = 10$:

$SI_1(i_1)$ $SI_2(i_1)$

(1,2) $0.49134 \times 10^{(1-1)}$ $0.49134 \times 10^{(1-1)} + 1.06996 \times 10^{(1-2)} = 0.59834$
(2,3) $0.95847 \times 10^{(1-1)}$ $0.95847 \times 10^{(1-1)} + 1.77825 \times 10^{(1-2)} = 1.13629$
(2,5) $0.49134 \times 10^{(1-1)}$ $0.49134 \times 10^{(1-1)} + 1.06996 \times 10^{(1-2)} = 0.59834$
(3,4) $0.19878 \times 10^{(1-1)}$ $0.19878 \times 10^{(1-1)} + 0.36167 \times 10^{(1-2)} = 0.23494$

$GSI_1(L_1)$ 2.13992 $GSI_2(L_1)$ 2.56791

$SI_3(i_1)$

(1,2) $0.49134 \times 10^{(1-1)} + 1.06996 \times 10^{(1-2)} + 1.75996 \times 10^{(1-3)} = 0.61594$
(2,3) $0.95847 \times 10^{(1-1)} + 1.77825 \times 10^{(1-2)} + 2.13992 \times 10^{(1-3)} = 1.15769$
(2,5) $0.49134 \times 10^{(1-1)} + 1.06996 \times 10^{(1-2)} + 1.75996 \times 10^{(1-3)} = 0.61594$
(3,4) $0.19878 \times 10^{(1-1)} + 0.36167 \times 10^{(1-2)} + 0.75993 \times 10^{(1-3)} = 0.24254$

$GSI_3(L_1)$ 2.63210

Figure 1. MOLORD algorithm applied to 2-methylbutane (**1**). **LDS** matrices; values derived for $I=X(\mathbf{LDS})$; $t_i = 1$.

TABLE I

Centric ordering given by $c(\boldsymbol{LDS})_i$ (values $SI_m(i_m)$) in 2,2-Dimethylnonane (**2**); $f=10$

$SI_0(i_0)$	$SI_1(i_0)$	$SI_2(i_0)$	$SI_3(i_0)$
5; 4; 6; 3; 7; 2; 8; {1; 10; 11}; 9	5; 4; 6; 3; 7; 2; 8; {1; 10; 11}; 9	5; 4; 6; 3; 7; 2; 8; {1; 10; 11}; 9	5; 4; 6; 3; 7; 2; 8; {1; 10; 11}; 9

$SI_1(i_1)$	$SI_2(i_1)$	$SI_3(i_1)$
(4,5); (5,6); (3,4); (6,7); (2,3); (7,8); {(1,2); (2,10); (2,11)}; (8, 9)	(4,5); (5,6); (3,4); (6;7); (2,3); (7,8); {(1,2); (2,10); (2,11)}; (8, 9)	(4,5); (5,6); (3,4); ˙(6,7); (2,3); (7,8); {(1,2); (2,10); (2,11)} (8, 9)

$SI_2(i_2)$	$SI_3(i_2)$
(3, 4, 5); (4, 5, 6); (2, 3, 4); (5, 6, 7); {(1, 2, 3); (2, 3, 10); (2, 3, 11)}; (6, 7, 8); {(1, 2, 10); (1, 2, 11); (2, 10, 11)}; (7, 8, 9);	(3, 4, 5); (4, 5, 6); (2, 3, 4); {(1, 2, 3); (2, 3, 10); (2, 3, 11)}; (5, 6, 7); {(1, 2, 10); (1, 2, 11); (2, 10, 11)}; (6, 7, 8); (7, 8, 9)

$SI_3(i_3)$

{(2, 3, 10, 11); (1, 2, 3, 10); (1, 2, 3, 11)}; (1, 2, 10, 11); (2, 3, 4, 5); (3, 4, 5, 6); {(1, 2, 3, 4); (2, 3, 4, 10); (2, 3, 4, 11)}; (4, 5, 6, 7); (5, 6, 7, 8); (6, 7, 8, 9)

1D :   minimal vertex eccentricity, ecc_i = min.

2D :   minimal vertex distance sum, DS_i = min.

3D :   minimal number of occurrences of the largest distance (or, when this is identical for two or more vertices, the next largest distance, *etc.*)

Criteria 1D – 3D are applied heirarchically. The last criterion is sometimes non-decisive since there are graphs with pair degenerate distance degree sequences for nonequivalent vertices.[9,11] However, in our test, application of 1D – 3D criteria to matrix $\boldsymbol{LDS}(L_2(\mathbf{2}))$ (see below) results in the same ordering as given by values $SI_2(i_2)$

$\boldsymbol{LDS}\ (L_2(\mathbf{2}))$:

(1,2,3)	27	144	59	30	36	44	54	0
(1,2,10)	33	120	51	26	30	36	44	54
(1,2,11)	33	120	51	26	30	36	44	54
(2,3,10)	27	144	59	30	36	44	54	0
(2,3,11)	27	144	59	30	36	44	54	0
(2,3,4)	24	107	129	36	44	54	0	0
(2,10,11)	33	120	51	26	30	36	44	54
(3,4,5)	26	54	117	143	54	0	0	0
(4,5,6)	30	62	68	135	99	0	0	0
(5,6,7)	36	74	80	24	81	99	0	0
(6,7,8)	44	90	30	26	24	81	99	0
(7,8,9)	54	44	36	30	26	24	81	99

In general, the ordering induced by $SI_m(i_n)$ values obey the 1D – 3D criteria when $n = m$. When $n < m$, the ordering could be different (compare the ordering given by the values $SI_2(i_2)$ and $SI_3(i_2)$, respectively – Table I).

An additional example is given for cuneane (**3**). Table II shows the centric ordering induced by $c(\mathbf{LDS})_i$, (values $SI_m(i_n)$) in this graph ($f = 5.000$).

TABLE II

Centric ordering given by $c(\mathbf{LDS})_i$ (values $SI_m(i_n)$) in cuneane (**3**); $f = 5.000$
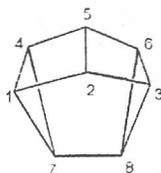
3

$SI_0(i_0)$	$SI_1(i_0)$; $SI_2(i_0)$; $SI_3(i_0)$; $SI_4(i_0)$;
(2, 5, 7, 8); {1, 3, 4, 6}	{2, 5}; {7, 8}; {1, 3, 4, 6}

$SI_1(i_1)$; $SI_2(i_1)$; $SI_3(i_1)$; $SI_4(i_1)$:

(7, 8); (2, 5); {(1, 2); (2, 3); (4, 5); (5, 6)}
{(1, 7); (3, 8); (4, 7); (6, 8)}; {(1, 4); (3, 6)}

$SI_2(i_2)$; $SI_3(i_2)$; $SI_4(i_2)$:

{(1, 2, 3); (4, 5, 6)}; {(1, 7, 8); (3, 7, 8); (4, 7, 8); (6, 7, 8)}
{(1, 2, 7); (2, 3, 8); (4, 5, 7); (5, 6, 8)}; {(1, 2, 5);(2, 3, 5); (2, 4, 5); (2, 5, 6)};
{(1, 2, 4); (1, 4, 5); (2, 3, 6); (3, 5, 6)}; {(1, 4, 7); (4, 1, 7); (3, 6, 8); (6, 3, 8)};
{(1, 7, 4); (3, 8, 6)}

$SI_3(i_3)$; $SI_4(i_3)$*:

(a)   {(1, 2, 3, 5); (2, 4, 5, 6)}; (b) {(1, 4, 7, 8); (3, 6, 7, 8)};
(c)   {(1, 2, 4, 7); (1, 4, 5, 7); (2, 3, 6, 8); (3, 5, 6, 8)}; (d) {(1, 4, 7); (3, 6, 8)};
(e)   {(1, 7, 8, 6); (3, 8, 7, 4)}; (f) {(3, 2, 1, 4); (1, 2, 3, 6); (1, 4, 5, 6); (3, 6, 5, 4)};
(g)   {(3, 2, 1, 7); (1, 2, 3, 8); (6, 5, 4, 7); (4, 5, 6, 8)};
(h)   {(2, 1, 7, 8); (2, 3, 8, 7); (5, 4, 7, 8); (5, 6, 8, 7)} (i) {(1, 2, 5, 6); (3, 2, 5, 4)};
(j)   {(1, 7, 8, 3); (4, 7, 8, 6)}; (k) {(5, 2, 1, 7); (5, 2, 3, 8); (2, 5, 4, 7); (2, 5, 6, 8)};
(l)   {(4, 1, 7, 8); (1, 4, 7, 8); (6, 3, 8, 7); (3, 6, 8, 7)};
(m) {(4, 1, 2, 5); (1, 4, 5, 2); (5, 2, 3, 6); (2, 5, 6, 3)}; (n) {(2, 1, 4, 5); (2, 3, 6, 5)};
(o)   {(1, 2, 5, 4); (3, 2, 5, 6)}; (p) {(2, 1, 4, 7); (5, 4, 1, 7); (2, 3, 6, 8); (5, 6, 3, 8)};
(q)   {(2, 3, 8, 6); (3, 8, 6, 5); (1, 7, 4, 5); (2, 1, 7, 4)}

* see Figure 2

Figure 2. presents the cuneane subgraphs i_3 (of three edges – labelled a to q) ordered according to their values $SI_3(i_3)$.
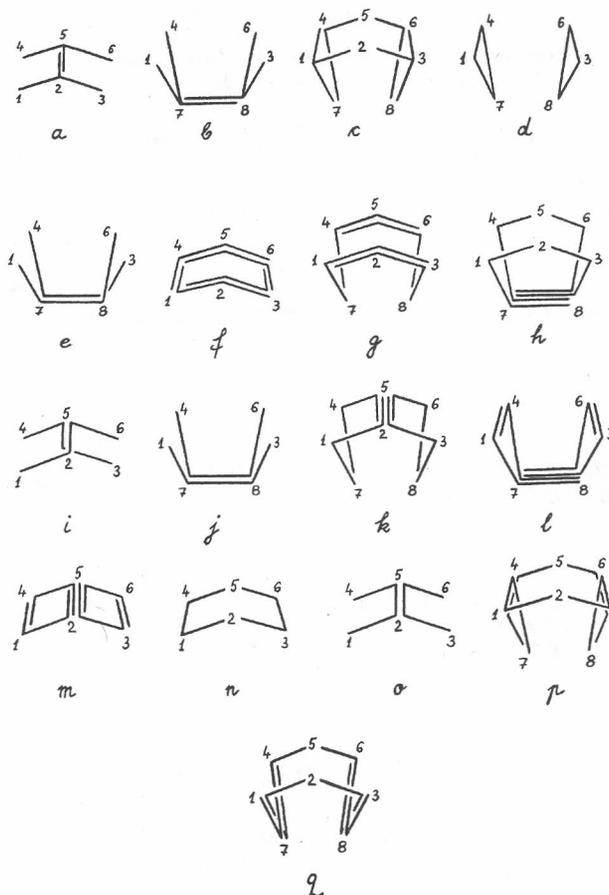
Figure 2. Pictorial representation of cuneane subgraphs i_3 (e to q) ordered according to values $SI_3(i_3)$

CENTROCOMPLEXITY ORDERING OF VERTICES (SUBGRAPHS)

The »x«-type operators provide an interesting ordering, which states a »centre of complexity« (or a centre of importance, *i.e.* a heteroatom). Table III lists the ordering induced by the $x(\boldsymbol{LDS})_i$ values in graph 2.

HETEROATOM PERCEPTION

The »x«-type operators are sensitive to heteroatoms by means of the t_i factor which, within this work, represents EVG_i (valence group electronegativities) values, defined[6] as follows

$$ESG_i = [ESA_i \cdot (ESH)^{h_i}]^{1/(1+h_i)} \tag{23}$$

TABLE III

Centrocomplexity ordering given by $x(\boldsymbol{LDS})_i$ (values $SI_m(i_n)$)
in 2,2-Dimethylnonane (**2**); $f = 10$.

$SI_0(i_0)$:	$SI_1(i_0)$; $SI_2(i_0)$; $SI_3(i_0)$:
2; 4; 5; 3; 6; 7; 8; {1; 10; 11}; 9	2; 3; 4; 5; 6; 7; 8; {1; 10; 11}; 9

$SI_1(i_1)$:	$SI_2(i_1)$; $SI_3(i_1)$:
(2,3); {(1,2);(2,10);(2,11)} (4,5); (3,4); (5,6); (6,7); (7,8); (8,9)	(2,3);{(1,2);(2,10);(2,11)}; (3,4); (4,5); (5,6); (6,7); (7,8); (8,9)

$SI_2(i_2)$; $SI_3(i_2)$:

{(1,2,3);(2,3,10);(2,3,11)}; (2,3,4); {(1,2,10);(1,2,11);(2,10,11)};
(3,4,5); (4,5,6); (5,6,7); (6,7,8); (7,8,9)

$SI_3(i_3)$:

{(2,3,10,11);(1,2,3,10);(1,2,3,11)}; (1,2,10,11);
{(1,2,3,4);(2,3,4,10);(2,3,4,11)}; (2,3,4,5);
(3,4,5,6); (4,5,6,7); (5,6,7,8); (6,7,8,9)

$$h_i = (8 - GA_i) - k_i \tag{24}$$

$$EVG_i = (ESG_i)^{1/(1 + k_i)} \tag{25}$$

where:

ESA, ESH the Sanderson electronegativities[12] for atoms A_i and hydrogen,
respectively

h_i number of hydrogen atoms belonging to group G_i

GA_i group number in the Periodic System for atom A belonging to group G_i

k_i degrees of vertex i (*i.e.* group G_i; when $k_i > (8 - GA_i)$, then $h_i = 0$

ESG_i Sanderson electronegativities for group G_i (*i.e.* the geometric mean of
electronegativities of the atoms belonging to group G_i)

Factor t_i (Eq. 5) is computed for the vertices i_n in L_n as the geometric mean of
EVG_i values of the vertices (of L_0) which i_n represents (see Refs. 1,12,13). Heteroa-
tom perception is examplified on cuneane and some of its N-rooted congeners
(graphs **3** and **7** – Figure 3 and Tables IV and V).

The intramolecular ordering induced by $x(\boldsymbol{LDS})_i$ index (*i.e.* values $SI_1(i_0)$) in
graph **2** follows the same trend as the coefficients of the first eigenvector. A com-
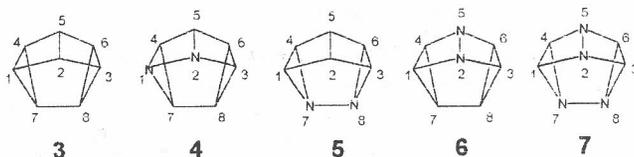
Figure 3. Cuneane and N-rooted cuneane (**3** to **7**).

parison of the two sets of values is given in Table VI . The only difference is that the $x(LDS)_i$ index puts vertices 1, 10, 11, and 9 in the set of »monovalent« vertices, after the divalent ones. For other considerations about the use of coefficients of the eigenvectors as topological invariants, consult Refs. 9,14–17.

TABLE IV

Centrocomplexity index $x(LDS)_i$ and $X(LDS)$ in cuneanes (**3** to **7**);
values $SI_m(i_0)$ and $GSI_m(L_0)$; $f = 10$.

Graph	**3**		**4**		**5**		**6**		**7**
values $SI_0(i_0)$:									
2	0.337702	2	0.353210	7	0.353210	2	0.353210	2	0.353210
5	0.337702	5	0.337702	8	0.353210	5	0.353210	5	0.353210
7	0.337702	7	0.337702	2	0.337702	7	0.337702	7	0.353210
8	0.337702	8	0.337702	5	0.337702	8	0.227702	8	0.353210
1	0.310649	1	0.324915	1	0.310649	1	0.310649	1	0.310649
3	0.310649	3	0.310649	3	0.310649	3	0.310649	3	0.310649
4	0.310649	4	0.310649	4	0.310649	4	0.310649	4	0.310649
6	0.310649	6	0.310649	6	0.310649	6	0.310649	6	0.310649
values $GSI_0(L_0)$:									
	2.593402		2.623177		2.624419		2.624419		2.655436
values $SI_1(i_0)$:									
2	0.404711	2	0.422259	7	0.420017	2	0.422272	2	0.422311
5	0.404711	5	0.405217	8	0.420017	5	0.422272	5	0.422311
7	0.402482	7	0.402963	2	0.404749	7	0.402493	7	0.420029
8	0.402482	8	0.402481	5	0.404749	8	0.402493	8	0.420029
1	0.374424	1	0.390656	1	0.374943	1	0.374943	1	0.375461
3	0.374424	3	0.374931	3	0.374943	3	0.374943	3	0.375461
4	0.374424	4	0.374883	4	0.374943	4	0.374943	4	0.375461
6	0.374424	6	0.374426	6	0.374943	6	0.374943	6	0.375461
values $GSI_1(L_0)$:									
	3.112083		3.147812		3.149303		3.149303		3.186523

TABLE V

Centrocomplexity index $x(\boldsymbol{LDS})_i$ and $X(\boldsymbol{LDS})_i$ in cunaenes (**3**) to (**7**); values $SI_m(i_1)$ and $GSI_m(L_1)$; $f = 10$.

Graph	**3**		**4**		**5**		**6**		**7**
values $SI_1(i_1)$:									
(2, 5)	0.268012	(1, 2)	0.280276	(7, 8)	0.280261	(2, 5)	0.280320	(2, 5)	0.280320
(1, 2)	0.267970	(2, 5)	0.274097	(2, 5)	0.268012	(1, 2)	0.274054	(7, 8)	0.280261
(2, 3)	0.267970	(2, 3)	0.274054	(1, 2)	0.267970	(2, 3)	0.274054	(1, 2)	0.274054
(4, 5)	0.267970	(4, 5)	0.267970	(2, 3)	0.267970	(4, 5)	0.274054	(2, 3)	0.274054
(5, 6)	0.267970	(5, 6)	0.267970	(4, 5)	0.267970	(5, 6)	0.274054	(4, 5)	0.274054
(7, 8)	0.267956	(7, 8)	0.267956	(5, 6)	0.267970	(7, 8)	0.267956	(5, 6)	0.274054
(1, 7)	0.254624	(1, 7)	0.260405	(1, 7)	0.260405	(1, 7)	0.254624	(1, 7)	0.260405
(3, 8)	0.254624	(3, 8)	0.254624	(3, 8)	0.260405	(3, 8)	0.254624	(3, 8)	0.260405
(4, 7)	0.254624	(4, 7)	0.254624	(4, 7)	0.260405	(4, 7)	0.254624	(4, 7)	0.260405
(6, 8)	0.254624	(6, 8)	0.254624	(6, 8)	0.260405	(6, 8)	0.254624	(6, 8)	0.260405
(1, 4)	0.242557	(1, 4)	0.248064	(1, 4)	0.242557	(1, 4)	0.242557	(1, 4)	0.242557
(3, 6)	0.242557	(3, 6)	0.242557	(3, 6)	0.242557	(3, 6)	0.242557	(3, 6)	0.242557
values $GSI_1(L_1)$:									
	3.111455		3.147218		3.146885		3.146099		3.183529
values $SI_2(i_1)$:									
(2, 5)	0.320767	(1, 2)	0.334647	(7, 8)	0.334616	(2, 5)	0.334692	(2, 5)	0.334693
(1, 2)	0.320733	(2, 5)	0.327854	(1, 2)	0.320936	(1, 2)	0.327828	(7, 8)	0.334628
(2, 3)	0.320733	(2, 3)	0.327824	(2, 3)	0.320936	(2, 3)	0.327828	(1, 2)	0.328032
(4, 5)	0.320733	(4, 5)	0.321131	(4, 5)	0.320936	(4, 5)	0.327828	(2, 3)	0.328032
(5, 6)	0.320733	(5, 6)	0.320937	(5, 6)	0.320936	(5, 6)	0.327828	(4, 5)	0.328032
(7, 8)	0.320704	(7, 8)	0.320907	(2, 5)	0.320771	(7, 8)	0.320717	(5, 6)	0.328032
(1, 7)	0.305766	(1, 7)	0.312525	(1, 7)	0.312524	(1, 7)	0.305978	(1, 7)	0.312737
(3, 8)	0.305766	(4, 7)	0.306145	(3, 8)	0.312524	(3, 8)	0.305978	(3, 8)	0.312737
(4, 7)	0.305766	(3, 8)	0.305970	(4, 7)	0.312524	(4, 7)	0.305978	(4, 7)	0.312737
(6, 8)	0.305766	(6, 8)	0.305771	(6, 8)	0.312524	(6, 8)	0.305978	(6, 8)	0.312737
(1, 4)	0.293140	(1, 4)	0.299612	(1, 4)	0.293518	(1, 4)	0.293543	(1, 4)	0.293918
(3, 6)	0.293140	(3, 6)	0.293340	(3, 6)	0.293518	(3, 6)	0.293543	(3, 6)	0.293918
Values $GSI_2(L_1)$:									
	3.733746		3.776662		3.776261		3.777719		3.820234

TABLE VI

Comparison between the eigenvector and $x(\boldsymbol{LDS})_i$ in graph **2**; values $SI_1(i_0)$; $f = 1$.

vertex	2	3	4	5	6	7	8
eigenvector	0.6177	0.4357	0.3061	0.2133	0.1461	0.0965	0.0585
$x(\boldsymbol{LDS})i$	0.1603	0.0889	0.0864	0.0832	0.0750	0.0643	0.0521

vertex	1	10	11	9
eigenvector	0.2913	0.2913	0.2913	0.0276
$x(\boldsymbol{LDS})_i$	0.0318	0.0318	0.0318	0.0208

It is very difficult to find a property (either a physical or a theoretical one) that can be correlated with the values derived from higher L_n (*i.e.* $SI_1(i_0)$), particularly in regular graphs (in which our LOVIs demonstrated their discriminating ability). The difficulty arises from the fact that such graphs show degenerate vertex properties (*i.e.* walk degree and distance degree sequences[18]). However, the cyclicity, expressed by self-returning walks of elongation (e), $srw_i^{(e)}$ (*i.e.* the diagonal elements of the powers of adjacency matrix, A^e) could be such a (nondegenerate) property. Table VII shows the correlation (about 0.973) between the values $SI_1(i_0)$ of the LOVIs built up on matrix **LDS** and $srw_i^{(6)}$, the minimal elongation at which the vertices are differentiated, within graph **8**. As we have shown elsewhere,[9] our layer matrices differentiate quasi equivalent vertices at low values of elongation, so that, when combined with derivative graphs L_n, n values no larger than 2 suffice for this purpose.

TABLE VII

Self-Returning Walks $srw_i^{(6)}$ $c(\textbf{LDS})_i$, $x(\text{LDS})_i$, (values $SI_1(i_0)$) and their correlation within graph **8**.
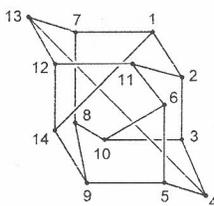
8

Vertex	$srw_i^{(6)}$	$c(\textbf{LDS})_i$	$x(\textbf{LDS})_i$
1	93	0.1407856	0.1327922
2	97	0.1433617	0.1329572
3	99	0.1433550	0.1329574
4	103	0.1459216	0.1331226
5	99	0.1433550	0.1329574
6	93	0.1407856	0.1327922
7	99	0.1433550	0.1329574
8	97	0.1433617	0.1329572
9	97	0.1433617	0.1329572
10	93	0.1407856	0.1327922
11	97	0.1433617	0.1329572
12	99	0.1433550	0.1329574
13	103	0.1459216	0.1331226
14	93	0.1407856	0.1327922

Correlations:

$srw^{(6)}_i$ *vs.* $c(\textbf{LDS})_i$: $R = 0.97300$; $S = 0.81834$

$srw^{(6)}_i$ *vs.* $x(\textbf{LDS})_i$: $R = 0.97329$; $S = 0.81403$

Another example illustrates the capability of indices calculated on **LDS** of derivative graphs L_0 and L_1 of octane isomers ($C(LDS(L_0))$, $X(LDS(L_0))$, $C(LDS(L_1))$, and $X(LDS(L_1))$) – denoted in Table VIII to X as $C_0 . X_0 . C_1$ and X_1) to estimate the critical pressure CP of these hydrocarbures.[19]

TABLE VIII

Values of the basic topological indices within the set of octane isomers and their critical pressures (CP)

	C_0	C_1	X_0	X_1	$\chi W^{(1)}$	CP
C8	0.4903	0.6411	0.7113	0.8035	7.8284	24.54
2MC7	0.6144	0.7869	0.7629	1.0081	7.5401	24.52
3MC7	0.6510	0.8480	0.8046	1.1110	7.6161	25.13
4MC7	0.6816	0.9002	0.8211	1.1549	7.6161	25.09
3EC6	0.7983	1.1329	0.8656	1.2558	7.6921	25.74
25M2C6	0.7565	1.0666	0.8219	1.2461	7.2518	24.54
24M2C6	0.8085	1.1487	0.8709	1.3813	7.3278	25.23
23M2C6	0.8154	1.1606	0.8917	1.4474	7.3615	25.94
34M2C6	0.8640	1.2444	0.9267	1.5451	7.4375	26.57
3E2MC5	1.0849	1.5449	0.9448	1.5889	7.4375	26.65
22M2C6	0.7856	1.1013	0.8725	1.5540	7.1213	24.96
33M2C6	0.8760	1.2566	0.9476	1.7956	7.2426	26.19
234M3C5	1.1052	1.5837	0.9745	1.8022	7.1068	26.94
3E3MC5	1.1592	1.7009	1.0084	1.9854	7.3640	27.71
224M3C5	1.0667	1.4709	0.9503	1.8949	6.8330	25.34
223M3C5	1.1348	1.6292	1.0180	2.1335	6.9628	26.94
233M3C5	1.1890	1.7543	1.0426	2.2288	7.0081	27.83
2233M4C4	1.4589	2.4852	1.1284	2.8649	6.5000	28.30

M = methyl; E = ethyl.

TABLE IX

Orthogonal indices $\Omega(TI, \chi W^{(1)})$ and $\Omega(TI, \chi W^{(2)})$ built up on C_0, C_1, X_0 and X_1 as origin TI and $\chi W^{(e)}$; (values multiplied by 10^{-2}).

	Ω $(C_0, \chi W^{(1)})$	Ω $(C_0, \chi W^{(2)})$	Ω $(C_1, \chi W^{(1)})$	Ω $(C_1, \chi W^{(2)})$	Ω $(X_0, \chi W^{(1)})$	Ω $(X_0, \chi W^{(2)})$	Ω $(X_1, \chi W^{(1)})$	Ω $(X_1, \chi W^{(2)})$
C8	7.9336	6.1156	12.4379	10.5179	3.2395	0.7935	7.1942	10.3981
2MC7	−7.2926	1.2642	−7.3042	1.8524	−12.4321	−4.3901	−9.6930	−0.1835
3MC7	4.3100	−3.3301	4.0980	−3.0396	5.8163	−2.7241	3.9126	−0.7571
4MC7	7.6704	6.5145	7.3540	6.6430	10.0310	8.0496	6.4736	8.7571
3EC6	28.0533	7.0281	29.4500	8.2103	28.9705	8.1079	19.9625	5.8508
25M2C6	−20.5525	−1.3960	−18.7119	5.0129	−26.1999	−7.3371	−24.6358	−8.3367
24M2C6	−7.2516	−1.9530	−5.9965	−0.7049	−6.1114	−1.4210	−9.1423	−4.4662
23M2C6	−3.1290	−6.5999	−1.8827	−5.3954	2.5511	−1.9331	−1.9148	−5.3231
34M2C6	9.8013	−13.0161	10.9360	−12.0678	19.0825	−4.7182	11.3908	−9.1165
3E2MC5	34.0026	10.5922	29.6592	6.5384	23.6918	2.7751	13.9421	−3.2416
22M2C6	−30.4132	0.0435	−29.5970	1.1101	−26.3461	2.5844	−19.7134	5.1272
33M2C6	−8.3731	−6.7304	−7.7915	−6.1507	4.9381	4.7755	6.5192	6.1957
234M3C5	3.1573	−2.8799	−0.9898	−6.6379	−1.8025	−6.5714	−6.6729	−13.3221
3E3MC5	34.7961	−3.1048	32.0320	−6.0222	32.5589	−3.5138	29.7365	−2.8828
224M3C5	−28.4369	14.3693	−35.4004	8.5776	−35.3375	8.1388	−28.6423	7.0542
223M3C5	−7.9920	−3.3739	−12.5585	−7.4653	−5.1089	−2.6801	−1.7407	−0.8062
233M3C5	2.4708	−8.9922	−0.2325	−11.7335	5.6940	−5.1810	8.3522	−5.0724
22333M4C	−18.7545	5.4492	−5.5022	15.2664	−23.2354	2.8331	−5.3286	10.1249

TABLE X

Statistics of the correlation of critical pressure (CP) of octanes with $C_0, X_0, \chi W^{(e)}, C_1, X_1$, and their orthogonal indices (denoted $\Omega(TI_1, TI_2)$; see text)

Variable	R	S
1 C_0	0.89810	0.54235
2 X_0	0.92414	0.47115
3 $\chi W^{(1)}$	0.57221	1.01135
4 C_1	0.89758	0.54366
5 X_1	0.86791	0.61255
6 C_0 ; X_0	0.92422	0.48635
7 C_0 ; $\chi W^{(1)}$	0.94179	0.42819
8 X_0 ; $\chi W^{(1)}$	0.97180	0.30031
9 C_0 ; $\Omega(C_0, \chi W^{(1)})$	0.94179	0.42819
10 C_1 ; $\Omega(C_1, \chi W^{(1)})$	0.94628	0.41185
11 X_0 ; $\Omega(X_0, \chi W^{(1)})$	0.97180	0.30031
12 X_1 ; $\Omega(X_1, \chi W^{(1)})$	0.97173	0.30068
13 C_0 ; $\Omega(C_0, \chi W^{(2)})$	0.93587	0.44876
14 C_1 ; $\Omega(C_1, \chi W^{(2)})$	0.93804	0.44135
15 X_0 ; $\Omega(X_0, \chi W^{(2)})$	0.93309	0.45806
16 X_1 ; $\Omega(X_1, \chi W^{(2)})$	0.88941	0.58215
17 C_0 ; X_0 ; $\chi W^{(1)}$	0.97339	0.30213
18 C_0 ; X_1 ; $\chi W^{(1)}$	0.97921	0.26746
19 C_1 ; X_0 ; $\chi W^{(1)}$	0.97835	0.27287
20 C_1 ; X_1 ; $\chi W^{(1)}$	0.97760	0.27747
21 C_0 ; $\Omega(C_0, \chi W^{(1)})$; $\Omega(C_0, \chi W^{(2)})$	0.97788	0.27577
22 X_0 ; $\Omega(X_0, \chi W^{(1)})$; $\Omega(X_0, \chi W^{(2)})$	0.98032	0.26027
23 C_1 ; $\Omega(X_1, \chi W^{(1)})$; $\Omega(X_1, \chi W^{(2)})$	0.98129	0.25381
24 X_0 ; $\Omega(X_1, \chi W^{(1)})$; $\Omega(X_1, \chi W^{(2)})$	0.98225	0.24727
25 C_1 ; $\Omega(C_1, \chi W^{(1)})$; $\Omega(C_1, \chi W^{(2)})$	0.98474	0.22944
26 C_1 ; $\Omega(C_0, \chi W^{(1)})$; $\Omega(C_0, \chi W^{(2)})$	0.98521	0.22590
27 X_1 ; $\Omega(X_1, \chi W^{(1)})$; $\Omega(X_1, \chi W^{(2)})$	0.99100	0.17649
28 X_1 ; $\Omega(X_0, \chi W^{(1)})$; $\Omega(X_0, \chi W^{(2)})$	0.99113	0.17517
29 X_1 ; $\Omega(C_0, \chi W^{(1)})$; $\Omega(C_0, \chi W^{(2)})$	0.99149	0.17161
30 X_1 ; $\Omega(C_1, \chi W^{(1)})$; $\Omega(C_1, \chi W^{(2)})$	0.99214	0.16501

From Table X one can see that our basic indices show rather poor correlations with CP, (lower than 0.93 – entries 1; 2 and 4; 5) but they are still better than that given by Randić's index[20] $\chi(0.57221$ – entry 3), denoted here as $\chi W^{(1)}$. Symbol $\chi W^{(e)}$ represents an extended connectivity index, calculated by Razinger[21] with Randić's formula by using walk degrees, $w_i^{(e)}$, of various elongation (e) (see also Ref. 18). In this work, indices $\chi W^{(e)}$ are calculated *per vertex* by

$$\chi W_i^{(e)} = \sum_{j:(i,j) \in E(G)} (W_i^{(e)} \cdot W_j^{(e)})^{-1/2} \qquad (26)$$

$$\chi W^{(e)} = \sum_i \chi W_i^{(e)} \qquad (27)$$

The summation in Eq. (26) runs over all j vertices adjacent to vertex i. It is obvious that $\chi W^{(e)}$ (G) is two times larger than the corresponding *per edge* connectivity index.

To improve the correlation , we used orthogonal indices, $\Omega(TI_1, TI_j)$ (Table IX) built up according to the Randić's procedure for orthogonalization of a set of ordered TIs: TI_1 is the origin index and TI_j are made orthogonal in a sequential process, one descriptor at a time (for details the reader can consult Refs. 22–24). The procedure enables separation of information brought by each descriptor within the set of orthogonal indices. Here, TI_1 is one of the basic indices C_0, X_0, C_1 and X_1 and TI_j belongs to the set $\chi W^{(e)}$, $e = 1, 2, ..., 7$. Among the resulting orthogonal descriptors, only $\Omega(TI_1, \chi W^{(1)})$ and $\Omega(TI_1, \chi W^{(2)})$ improved the correlation and are therefore, shown in Tables IX and X.

Thus, in two variable regressions, the orthogonal descriptors $\Omega(TI_1, \chi W^{(1)})$ give exactly the same correlation as that given by a TI_1 and the nonorthogonal $\chi W^{(1)}$ (compare entries 7 and 8 with 9 and 11, respectively – Table X). This is not surprising since $\Omega(TI_1, \chi W^{(1)})$ are produced in the first step of the orthogonalization process and represent just the part of $\chi W^{(1)}$ nonexplained by TI_1 according to the regression. The maximum correlation value obtained was 0.97180., (with variables X_0 and $\chi W^{(1)}$) (entry 8) and X_0 and $\Omega(TI_1, \chi W^{(1)})$ (entry 11), respectively). The second set of orthogonal indices $\Omega(TI_1, \chi W^{(2)})$ shows slightly lower correlations (entries 13 and 16) in comparison with those given by $\Omega(TI_1, \chi W^{(1)})$.

In three variable regressions, the use of orthogonal indices clearly improved the correlations in comparison with the regressions performed with nonorthogonal indices (entries 21 to 30 *vs.* 17 to 20). The presence of the basic indices C_1 or X_1 (*i.e.* indices calculated on derivative graphs L_1) in regression, either as nonorthogonal or as the origin of an orthogonal index, resulted in a supplementary rise of the coefficient of correlation (over 0.98 – entries 23 to 30) and a corresponding drop of the standard error of estimate (less than 0.26). The maximum correlation value obtained was 0.99214 (standard error 0.16501 – variables: X_1; $\Omega(C_1, \chi W^{(1)})$ and $\Omega(C_1, \chi W^{(2)})$. This fact demonstrates that the critical pressure is controlled by the topology of edges, as given by the indices constructed on derivative graphs L_1 and the connectivity indices $\chi W^{(1)}$ and $\chi W^{(2)}$.

Correlations of 0.903 and 0.971 (standard errors of 0.530 and 0.294, respectively) were reported by Balaban and Catana[19] by using two very elaborated indices XC and XC' (called distance-enhanced exponential connectivity indices), in single variable regressions.

DISCUSSION

The idea to »see« the total graph environment of each subgraph was developed by Diudea *et al.*[1] relative to the layer matrices of the line derivatives L_n of molecular graphs. At the vertex/atom level, the question was also considered by Hall and Kier[13] on the ground of the »topological state« matrix. Their algorithm offers a set of τ-indices with highly discriminating power, which are useful in topological equivalence perception and in the QSPR/QSAR studies.

The MOLORD algorithm provides a »spectrum« of invariants (*i.e.* $SI_m(i_n)$ and $GSI_m(L_n)$; $m = n$, $n+1$, $n+2$,...), derived from a set of successive derivative graphs, L_n. Among the proposed local invariants, the »c«-type ones enhance the contribution of more remote vertices whereas the »x«-type invariants that of the nearer neighbours.

As emphasized in our previous works,[1,4] the »c«-type indices, and particularly $c(\textbf{LDS})_i$, provide a centric ordering of the subgraphs in agreement with Bonchev's 1D – 3D criteria. It is well examplified in graphs 2 and 3. However, the ordering could change at lower values of factor f (see Eq. (20)). For example, with $f = 10$, the ordering given by $SI_2(i_2)$ values in graph **2** changes at higher m values, but it is retained with $f = 1.000$. This procedure has a meaning close to the hierarchical application of 1D – 3D criteria.

The partitioning of subgraphs of various sizes into classes of topological equivalence is reached, in general, at the level of $SI_m(i_n)$; $m = n+1$ and no additional derivative graphs (higher m values) are needed. In this respect, it is not essential that the ordering changes. The topological symmetry is well illustrated for cuneane in Figure 2.

By means of the t_i weighting factor, the $x(\textbf{LDS})_i$ index is able to discriminate various locations of heteroatoms in a molecular graph. This fact is illustrated in Figure 3 and Tables IV and V, for cuneane and its N-rooted congeners (graphs **3** to **7**; see also Ref. 25). Note that similarly located subgraphs (particularly atoms) have close values of $x(\textbf{LDS})_i$ within the considered set of graphs. Such LOVIs and the corresponding global indices were found to have a good correlating ability in QSPR/QSAR studies (see the results of this work and also Ref. 8).

Since the derivative graphs are based on connectivities, a large number of iterations results in an exponentail rise of the vertex degree (see Eq. (14)), excepting the graphs with max $k_i = 2$. As a consequence, the higher terms of L_n will stress the complexity of a given graph, L_0. Thus, it is easily conceivable that the »c«-type ordering will converge towards the one of »x«-type as n increases. The spectrum of values given by the MOLORD algorithm clearly follows the complexity trend of L_n. A way to limit the contribution of higher rank derivative graphs is to enlarge that f parameter sufficiently.

CONCLUSIONS

The MOLORD algorithm provides a »spectrum« of invariants, which are computed for subgraphs of various sizes extracting the topological information from the derivative graphs L_n. New relations are derived for the vertex degree and the number of edges in the L_n of regular graphs using parameters of the initial graph L_0. The classical invariants the algorithm uses are calculated on the layer matrices, $\textbf{LM}$ (particularly the layer matrix of distance sums, $\textbf{LDS}$). These invariant are real numbers, found to give good correlations in the QSPR/QSAR studies.

The MOLORD algorithm appears to be a powerful tool in topological equivalence perception and could be promising for the graph isomorphism testing (see Ref. 2).

REFERENCES

1. M. V. Diudea, D. Horvath, I. E. Kacso', O. M. Minailiuc, and B. Parv, *J. Math. Chem.* **11** (1992) 259.
2. S. H. Bertz, *Discr. Appl. Math.* **19** (1988) 65.
3. V. A. Skorobogatov and A. A. Dobrynin, *MATCH* **23** (1988) 105.
4. A. T. Balaban and M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **33** (1993) 421.

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

6. M. V. Diudea and I. Silaghi-Dumitrescu, *Rev. Roum. Chim.* **34** (1989) 1175.

7. N. Trinajstić, *Chemical Graph Theory*, CRC Press Inc., Boca Raton, Florida, 1983, p. 17.

8. M. V. Diudea, D. Horvath, and A. Graovac, *J. Chem. Inf. Comput. Sci.* **35** (1995) 129.

9. M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1064.

10. D. Bonchev, A. T. Balaban, and M. Randić, *Int. J. Quantum Chem.* **19** (1981) 61.

11. O. Ivanciuc, T. S. Balaban, and A. T. Balaban, *J. Math. Chem.* **14** (1993) 21.

12. R. T. Sanderson, *Polar Covalence*, Acad. Press, New York, 1983, p. 41.

13. L. H. Hall and L. B. Kier, *Quant. Struct.-Act. Relat.* **9** (1990) 115.

14. A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **32** (1992) 23.

15. A. T. Balaban, *ibid.* **34** (1994) 398.

16. A. T. Balaban, D. Ciubotariu, and M. Medeleanu, *ibid.* **31** (1991) 517.

17. C. Rucker and G. Rucker, *J. Math. Chem.* (in press).

18. M. V. Diudea, M. I. Topan, and A. Graovac, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1072.

19. A. T. Balaban and C. Catana, *J. Comput. Chem.* **14** (1993) 155.

20. M. Randić, *J. Amer. Chem. Soc.* **97** (1975) 6609.

21. M. Razinger, *Theor. Chim. Acta* **70** (1986) 365.

22. M. Randić, *New J. Chem.* **15** (1991) 517.

23. M. Randić, *J. Chem. Inf. Comput. Sci.* **31** (1991) 311.

24. M. Randić, *J. Molec. Struct. (Theochem)* **233** (1991) 45.

25. G. Rucker and C. Rucker, *J. Chem. Inf. Comput. Sci.* **30** (1990) 187.

SAŽETAK

Molekulska topologija. 14. Algoritam MOLORD
i realne invarijante podgrafova

Mircea V. Diudea, Dragos Horvath i Danail Bonchev

Predložan je algoritam, nazvan MOLORD za određivanje (realnih) invarijanti podgrafova u molekulskim grafovima. Algoritam uzima u obzir različitost atoma, na osnovi njihove elektronegativnosti i može se koristiti za uočavanje njihove ekvivalencije ili za definiranje lokalnih i globalnih deskriptora u QSPR i QSAR. Algoritam je implementiran u Turbo Pascalu, a njegova je primjena prikazana na nizu izabranih grafova.