Journal of Mathematical Chemistry 2(1988)347 – 375

TOPOLOGICAL INDICES FOR MOLECULAR FRAGMENTS AND NEW GRAPH INVARIANTS

Ovanes MEKENYAN and Danail BONCHEV

Department of Physical Chemistry, Higher School of Chemical Technology, 8010 Burgas, Bulgaria

and

Alexandru BALABAN

Organic Chemistry Department, Polytechnic Institute, 76206 Bucharest, Roumania

Received 4 April 1988

Abstract

Whereas the internal fragment topological index (IFTI) is calculated in the normal manner as for any molecule, the external fragment topological index (EFTI) is calculated so as to reflect the interaction between the excised fragment F and the remainder of the molecule $(G - F)$. For selected topological indices (TIs), a survey of EFTI values, formulas and examples is presented. Some requirements as to the fragment indices are formulated and examined. In the discussion of the results, it is shown that for some TIs regularities exist in the dependence of EFTI values upon the branching of fragment F, or upon the marginal versus central position of the fragment F in the graph G. New vertex invariants can be computed as EFTI values for one-atom fragments over all graph vertices; by iteration, it is in principle possible to devise an infinite number of new vertex invariants.

Abbreviations and notation

a_{ij} – entries (Kronecker delta) in A
A – adjacency matrix
d_{ij} – topological distance between vertices i and j, entries in $\boldsymbol{D}$
$\boldsymbol{D}$ – distance matrix
e_{ij} – number of edges adjacent to edge $\{ij\}$
EA – edge adjacency, EA = N_2
EFTI – external fragment topological index
F – fragment
FTI – fragment topological index
G – graph
HOC – hierarchically ordered extended connectivities
IFTI – internal fragment topological index
I_D^M, I_D^E, I_{CHR}, I_{ORB}, I_Z, I_χ – information theoretic indices (see text)
$\{ij\}$ – edge between vertices i,j
J – average distance sum connectivity TI
M_1 – Zagreb group TI
ΣM_i – sum of S_i values
$\mathcal{M}$ – topological index based on HOC vertex labels, ΣM_i
N_2 – topological index proposed by Gordon and Scantlebury
$\mathcal{N}$ – topological index based on HOC vertex labels, ΣM_i^2
NEFTI – normalized external fragment TI
NIFTI – normalized internal fragment TI
p, p' – number of vertices in G and F, respectively
q – number of edges in the graph
Q – quadratic TI
s_i – distance sum for vertex i
S_i – sum of HOC vertex labels up to vertex i
TI – topological index
v_i – degree of vertex i
VA – vertex adjacency
W – Wiener's TI (half the sum of all entries in $\boldsymbol{D}$)
Z – Hosoya's TI
${}^1\chi$ – Randić's TI, molecular connectivity TI
${}^k\chi$ – generalized molecular connectivity TI
${}^k\chi v$ – valence connectivity TI

1. Introduction

Topological indices (TIs) are numbers associated with molecular structures which serve for quantitative relationships between chemical structure and properties. The first such index was published by Wiener [1], but the name topological index was invented by Hosoya [2]. A very successful TI, the molecular connectivity, was devised by Randić [3]. A drawback of TIs is their degeneracy, i.e. the fact that two or more TIs can have the same value. In the search for TIs with lower degeneracy, information theory was applied by Bonchev and Trinajstić [4] and indices with low degeneracy were proposed by Balaban [5] and Randić [6]. Several reviews of topological indices are available [3–14]. In some cases, especially for drug design, it is desirable to characterize topologically a fragment of a molecule. Until now, this was done by applying to fragments some of the general procedures used for whole molecules, without paying special attention to the interaction between the fragment(s) and the remainder of the molecule. It is known that such interactions can have important consequences, e.g. hydroxy groups are strongly influenced by the group to which they are bonded: NO_2, Hal, H, Alk, Ar, RCO, etc.

2. Mathematical formalism

In practice, topological indices are derived by certain procedures starting from the hydrogen-depleted graphs representing the skeleton of organic molecules. Let G be such a molecular graph having p vertices symbolizing non-hydrogen atoms, and let F be a subgraph (fragment) with p' vertices ($1 \leqslant p' < p$). In the following, we shall abbreviate Fragment Topological Index as FTI.

A few more graph-theoretical definitions and the corresponding notation are necessary. The number of graph edges (symbolizing covalent bonds) will be denoted by q. If v_i edges meet in vertex i of a graph, the vertex is said to have degree v_i. The adjacency matrix A of a graph is a $p \times p$ symmetrical square matrix with entries a_{ij} equal to 1 if two vertices i and j are adjacent, and zero otherwise. The topological distance between two vertices i and j is the number of edges between these vertices along the shortest path between them. The distance matrix $\boldsymbol{D}$ of a graph is a $p \times p$ symmetrical square matrix with entries d_{ij} equal to topological distances between vertices i and j. On adding all entries on row i of A one obtains v_i, the vertex degree; on adding all entries on row i of $\boldsymbol{D}$ one obtains another graph invariant, the distance sum S_i.

Several problems arise in connection with devising topological indices for molecular fragments. The first problem is to characterize the fragment in all its topological (constitutional) detail. Thus, if one treats a fragment like a molecule, one cannot differentiate between n-butyl and s-butyl groups, or between n-propyl and isopropyl groups. This problem may be solved by having recourse to rooted graphs:

the root (to be differentiated numerically) corresponds to the point of attachment of a univalent molecular fragment [15]. In some other cases, however, one may be interested in fragments having several points of attachment.

The second problem is to consider the interaction between fragments and the remainder of the molecule. In this respect, one has to consider two types of topological indices for molecular fragments:

(i) Indices which consider only the atoms and bonds belonging to the fragment, i.e. the internal fragment topology; we shall call such indices "internal" FTIs, IFTI, or intra-indices.

(ii) Indices which describe a fragment in connection with the remainder of the molecule; we shall call such indices "external FTIs", EFTI, or inter-indices.

A requirement as to the IFTI range of values for both the fragment F and the remainder of the molecule $G - F$ may be assumed:

$$\begin{aligned} &0 \leqslant \mathrm{IFTI}\,(F) < \mathrm{TI}\,(G) \\ &0 \leqslant \mathrm{IFTI}\,(G-F) < \mathrm{TI}\,(G), \end{aligned} \tag{1}$$

where TI may denote those out of the hundred odd topological indices described so far which increase with the increase in the number of graph vertices p. The lower bound is reached for one-vertex fragments.

In the next section, it will be shown that the majority of known indices meets requirments (1). This is not, however, the case for the information theoretic indices I_i^{E} based on the equivalence of different graph elements or characteristics [4], for which some IFTI (F) or IFTI $(G - F)$ are larger than the respective TI for the whole graph. The Balaban index J [5] disobeys condition (1) for another reason. It is not an increasing function of p, and hence the reverse inequality IFTI $(F) >$ TI (G) occurring for J does not prevent the applicability of this fragment index.

Related to point (ii), the external fragment topological indices EFTI (F) are specified as the difference in value between the topological index for the whole graph and the internal fragment indices for both the fragment and the remainder of the molecule:

$$\mathrm{EFTI}\,(F) = \mathrm{TI}\,(G) - [\mathrm{IFTI}\,(F) + \sum_k \mathrm{IFTI}\,(G-F)_k]. \tag{2}$$

Here, the sum incorporates as many terms as the number of $(G - F)$ disconnected components. Indeed, there will be only one term when $(G - F)$ is a connected graph.

The idea of EFTI-indices may best be illustrated by those topological indices which are based on the adjacency matrix (vertex and edge adjacency, Zagreb index, molecular connectivity, etc.) or the distance matrix of the molecular graph (Wiener

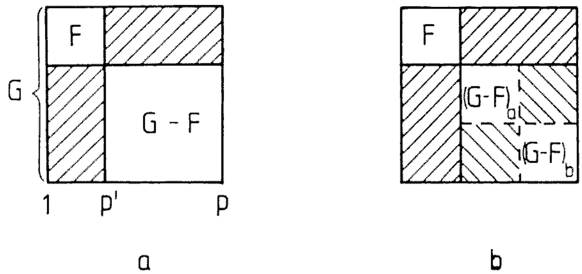


Fig. 1. A scheme for a topological matrix (adjacency or distance matrix) of a graph G with p vertices, from which a fragment F with p' vertices is selected. Two cases are considered, in which the remainder of the molecule is: (a) connected, (b) disconnected graph.

index, information index on the distance magnitude, etc.). As shown in fig. 1, these are $p \times p$ symmetrical matrices. If the fragment F has p' vertices, the IFTI (F) is defined by operations on the submatrix F, while the IFTI $(G - F)$ is similarly specified on the submatrix $(G - F)$ having $p - p'$ vertices. The EFTI-indices are defined by operations on the hatched portions of the matrix. Two cases occur in calculating EFTIs, depending on whether the remainder of molecule $G - F$ is a connected graph (fig. 1a) or whether it is a disconnected one (fig. 1b). In the first case, EFTI describes the interrelation between F and $G - F$ (adjacency and distances between vertices from the two parts of the graph). When $(G - F)$ comprises two or more disjoint subgraphs, the interaction between these subgraphs (the additional hatched portions in fig. 1b) is not taken into account in specifying IFTI $(G - F)$, since they are connected only by virtue of the fragment F. In dealing with adjacency matrix-based IFTIs, this interaction is zero (disjoint subgraphs), while the distance between the vertices of $(G - F)_a$ and $(G - F)_b$ is by definition infinity. Thus, according to eq. (2), the EFTIs based on the distance matrix of G accounts also for the distances between the vertices of $(G - F)_a$ and $(G - F)_b$ in the initial connected graph G. Indeed, eq. (2) covers also TIs which are not matrix-based, for example, the Hosoya index, i.e. all EFTIs are calculated by summing the IFTIs over all $(G - F)$-components, despite the possibilities for some TIs and the $(G - F)_k$-components to be treated in a common scheme.

Now let the requirement as to the EFTI values be formulated similarly to (1):

$$\mathrm{EFTI}\,(F) < \mathrm{TI}\,(G), \tag{3}$$

where again TI (G) is an increasing function of the number of graph vertices. The lower bound of EFTI values is not specified for two reasons. EFTI $(F) \neq 0$ always holds because there always exists some interrelation between F and $G - F$ due to the graph connectedness. Hence, EFTI $(F) > 0$ should be expected for all TIs which display additivity or, more generally, for which inequality (4) holds:

$$\mathrm{IFTI}\,(F) + \sum_k \mathrm{IFTI}\,(G-F)_k < \mathrm{TI}\,(G). \tag{4}$$

This is the case with topological indices such as the Hosoya index, Zagreb index, HOC index, Wiener index, etc.

The inverse inequality (4) could, however, also occur for some topological indices, thus resulting in negative EFTI values. This is the case with the Randić index, the information indices for the magnitude of the respective graph characteristics, etc. Indeed, the negative EFTI values do not violate requirement (3) and can be used for practical purposes. In applying I_D^M, ${}^1\chi$, J, and other indices having similar mathematical formulation, one should take into account some more details. Thus, ${}^1\chi$ and J indices contain terms of the kind $(x_i x_j)^{-1/2}$, x_i being the vertex degree v_i and the vertex distance sum s_i, respectively. With fragment removal, some terms disappear (some bonds are cut), but some v_i and all s_i diminish, thus enlarging in value the remaining terms. These two opposing trends usually prevent the regular behaviour of these EFTIs. In the case of ${}^1\chi$, v_i are small in value and the increase in the remaining terms is larger, thus causing the appearance of positive, along with the negative, EFTI values. In dealing with J, one almost always obtains negative EFTIs due to larger s_i values (i.e. smaller $(s_i s_j)^{-1/2}$ terms). Similar difficulties may arise with I_D^M, where again two opposing trends may appear.

Another general criterion for the applicability of the fragment topological indices may be formulated proceeding from the idea that the FTI should reflect the structure topology in the same manner as the topological index of the whole structure does. More, specifically, if

$$\mathrm{TI}\,(G_1) > \mathrm{TI}\,(G_2),$$

then

$$\mathrm{EFTI}\,(F \subset G_1) > \mathrm{EFTI}\,(F \subset G_2) \tag{5}$$

should hold when keeping constant the remaining factors: the fragment centric location, the $(G-F)$ branching and cyclicity, etc. This requirement will be discussed in detail in sect. 4.

The fragment topological indices may be normalized by dividing them by TI (G):

$$\begin{aligned} \mathrm{NIFTI}\,(F) &= \mathrm{IFTI}\,(F)/\mathrm{TI}\,(G) \\ \mathrm{NEFTI}\,(F) &= \mathrm{EFTI}\,(F)/\mathrm{TI}\,(G) \end{aligned} \tag{6}$$

Here, N is added in the beginning of the abbreviations, denoting Normalized.

Since the fragment F can be as small as a one-atom (non-hydrogen) fragment, in which case most of the internal indices are equal to zero, we have:

$$0 \leqslant \text{NIFTI}\,(F) < 1, \tag{7}$$

as it will be illustrated in the next two sections.

A similar range

$$0 < \text{NEFTI}\,(F) < 1 \tag{8}$$

can be specified for those indices which obey inequality (4), while the lower bound is -1 for the indices obeying the reversed inequality (4):

$$-1 < \text{NEFTI}\,(F) < 1. \tag{9}$$

Topological indices which are not a continuous increasing function of the number of graph vertices could have NEFTI values out of this range, as is the case with the Balaban index J (see table 3).

One should note that IFTI (F) is a constant for a given fragment of any molecule, whereas NIFTI (F) depends on the whole molecule. On the other hand, both EFTI (F) and NEFTI (F) depend upon the molecule as a whole in a more subtle fashion, whose analysis is the main object of the present paper.

In the following section, we show how to calculate these indices, both with general formulas and with selected examples.

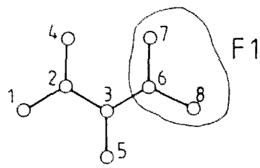
3. Selected fragment topological indices

Three examples are chosen to illustrate how the fragment topological indices are to be calculated: an acyclic graph **1** (2,3,4-trimethylpentane), a monocyclic graph **2** (2-sec-butylcyclohexane), and a tricyclic graph **3** (perhydroantracene), as shown in fig. 2. We shall include only certain topological indices, omitting others. One such omitted TI is the largest eigenvalue because possible disconnected fragments in $G - F$ have no clearly defined such TI. Other topological indices will be shown to violate some of the requirements given in sect. 2.

3.1. VERTEX ADJACENCY (VA) FRAGMENT TIs [11,16]

$$\begin{aligned} &\text{VA.IFTI}\,(F) = \tfrac{1}{2} \sum_{ij \in F} a_{ij}\,; \quad \text{VA.IFTI}\,(G-F) = \tfrac{1}{2} \sum_{ij \in \{G-F\}} a_{ij}\,; \\ &\text{VA.EFTI}\,(F) = \tfrac{1}{2} \left[\text{VA}(G) - \sum_{ij \in F} a_{ij} - \sum_{ij \in \{G-F\}} a_{ij} \right] = N_{\text{fce}}\,, \end{aligned} \tag{10}$$

Block 1

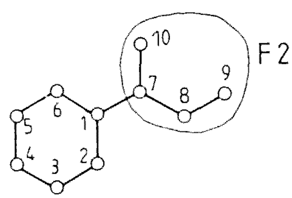


1

	1	2	3	4	5	6	7	8
1	0	1						
2	1	0	1	1				
3		1	0		1	1		
4		1		0				
5			1		0			
6			1			0	1	1
7						1	0	
8						1		0

	1	2	3	4	5	6	7	8
1	0	1	2	2	3	3	4	4
2	1	0	1	1	2	2	3	3
3	2	1	0	2	1	1	2	2
4	2	1	2	0	3	3	4	4
5	3	2	1	3	0	2	3	3
6	3	2	1	3	2	0	1	1
7	4	3	2	4	3	1	0	2
8	4	3	2	4	3	1	2	0

Block 2



2

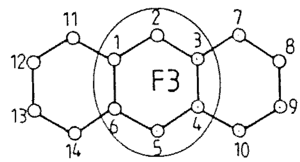
	1	2	3	4	5	6	7	8	9	10
1	0	1				1	1			
2	1	0	1							
3		1	0	1						
4			1	0	1					
5				1	0	1				
6	1				1	0				
7	1						0	1		1
8							1	0	1	
9								1	0	
10							1			0

	1	2	3	4	5	6	7	8	9	10
1	0	1	2	3	2	1	1	2	3	2
2	1	0	1	2	3	2	2	3	4	3
3	2	1	0	1	2	3	3	4	5	4
4	3	2	1	0	1	2	4	5	6	5
5	2	3	2	1	0	1	3	4	5	4
6	1	2	3	2	1	0	2	3	4	3
7	1	2	3	4	3	2	0	1	2	1
8	2	3	4	5	4	3	1	0	1	2
9	3	4	5	6	5	4	2	1	0	3
10	2	3	4	5	4	3	1	2	3	0

Fig. 2.

Fig. 2 (continued)

Block 3



3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	1				1					1			
2	1	0	1											
3		1	0	1			1							
4			1	0	1					1				
5				1	0	1								
6	1				1	0								1
7			1				0	1						
8							1	0	1				0	
9								1	0	1				
10				1					1	0				
11	1										0	1		
12								0			1	0	1	
13												1	0	1
14						1							1	0

	1	2	3	4	5	6	7	8	9	10	11	12	12	14
1	0	1	2	3	2	1	3	4	5	4	1	2	3	2
2	1	0	1	2	3	2	2	3	4	3	2	3	4	3
3	2	1	0	1	2	3	1	2	3	2	3	4	5	4
4	3	2	1	0	1	2	2	3	2	1	4	5	4	3
5	2	3	2	1	0	1	3	4	3	2	3	4	3	2
6	1	2	3	2	1	0	4	5	4	3	2	3	2	1
7	3	2	1	2	3	4	0	1	2	3	4	5	6	5
8	4	3	2	3	4	5	1	0	1	2	5	6	7	6
9	5	4	3	2	3	4	2	1	0	1	6	7	6	5
10	4	3	2	1	2	3	3	2	1	0	5	6	5	4
11	1	2	3	4	3	2	4	5	6	5	0	1	2	3
12	2	3	4	5	4	3	5	6	7	6	1	0	1	2
13	3	4	5	4	3	2	6	7	6	5	2	1	0	1
14	2	3	4	3	2	1	5	6	5	4	3	2	1	0

Fig. 2. The three graphs to exemplify the selected TIs and their adjacency and distance matrices.

where VA (G) stands for the sum of entries in the adjacency matrix of graph G, and where N_{fce} denotes the number of the fragment cut edges.

In the three examples:

$$\mathrm{VA}\,(G1) = 7, \mathrm{VA.IFTI}\,(F1) = 2;\ \mathrm{VA.IFTI}\,(G1 - F1) = 4;\ \mathrm{VA.EFTI}\,(F1) = 1;$$

$$\mathrm{VA}\,(G2) = 10, \mathrm{VA.IFTI}\,(F2) = 3;\ \mathrm{VA.IFTI}\,(G2 - F2) = 6;\ \mathrm{VA.EFTI}\,(F2) = 1;$$

$$\mathrm{VA}\,(G3) = 16, \mathrm{VA.IFTI}\,(F3) = 6;\ \mathrm{VA.IFTI}\,(G3 - F3) = 6;\ \mathrm{VA.EFTI}\,(F3) = 4.$$

3.2. ZAGREB INDEX (M_1), FRAGMENT TIs [17]

$$M_1.\mathrm{IFTI}\,(F) = \sum_{i \in F} v_i^2\ ;\quad M_1.\mathrm{IFTI}\,(G - F) = \sum_{i \in \{G-F\}} v_i^2\ . \tag{11}$$

Here, the vertex degrees refer to subgraphs after cutting the cut edges.

$$M_1.\mathrm{EFTI}\,(F) = M_1(G) - \sum_{i \in F} v_i^2 - \sum_{i \in \{G-F\}} v_i^2$$

$$= 2 \sum_{\{ij\} \in \{\mathrm{fce}\} \in G} (v_i + v_j - 1), \tag{12}$$

where $\{ij\} \in \{\mathrm{fce}\}$ denotes the endpoints of fragment cut edges and the final expression refers to the degrees of these vertices in the whole graph (before cutting).

In the three examples,

$M_1.\mathrm{IFTI}\,(F1) = 1^2 + 1^2 + 2^2 = 6$; $M_1.\mathrm{IFTI}\,(G1 - F1) = 3.1^2 + 3^2 + 2^2 = 16$;

$M_1(G1) = 5.1^2 + 3.3^2 = 34$; $M_1.\mathrm{EFTI}\,(F1) = 10$;

$M_1.\mathrm{IFTI}\,(F2) = 2.2^2 + 2.1^2 = 10$; $M_1.\mathrm{IFTI}\,(G2 - F2) = 6.2^2 = 24$;

$M_1(G2) = 6.2^2 + 2.3^2 + 2.1^2 = 44$; $M_1.\mathrm{EFTI}\,(F2) = 10$;

$M_1.\mathrm{IFTI}\,(F3) = 6.2^2 = 24$; $M_1.\mathrm{IFTI}\,(G3 - F3) = 2(2.1^2 + 2.2^2) = 20$;

$M_1(G3) = 10.2^2 + 4.3^2 = 76$; $M_1.\mathrm{EFTI}\,(F3) = 32$.

Closely related to the Zagreb index are two other TIs, namely the Gordon–Scantlebury index N_2 [18] and the quadratic index Q [19]. The following relationships exist between these TIs for acyclic graphs:

$$Q = N_2 - p + 2; \quad M_1 = 2(N_2 + p - 1). \tag{13}$$

Since Q and N_2 depend linearly on N_2 and on the number p of vertices in the graph, we do not present these two TIs in detail.

3.3. MOLECULAR CONNECTIVITY ($^1\chi$, RANDIĆ INDEX [13] FRAGMENT TIs

$$\begin{aligned} &{}^1\chi.\mathrm{IFTI}\,(F) = \sum_{\{ij\} \in F} (v_i v_j)^{-1/2}; \quad {}^1\chi.\mathrm{IFTI}\,(G-F) \sum_{\{ij\} \in G-F} (v_i v_j)^{-1/2}; \\ &{}^1\chi.\mathrm{EFTI}\,(F) = {}^1\chi(G) - \sum_{\{ij\} \in F} (v_i v_j)^{-1/2} - \sum_{\{ij\} \in (G-F)} (v_i v_j)^{-1/2}. \end{aligned} \tag{14}$$

In the final expression, the vertex degrees in $^1\chi\,(G)$ refer to the whole graph, before cutting off the fragment(s), whereas vertex degrees in F and $G - F$ refer to subgraphs after cutting. Therfore, a "simpler" expression of $^1\chi.\mathrm{EFTI}\,(F)$ is less easy is manipulate: it contains the sum of $(v_i v_j)^{-1/2}$ terms for cut edge(s) and of terms

$$\sum_k \left[(v_i v_k)^{-1/2} - (v_i v_k - v_k)^{-1/2} + (v_j v_k)^{-1/2} - (v_j v_k - v_k)^{-1/2}\right] ,$$

where i, j are the endpoints of the cut edge and k is an adjacent point in F or in $G - F$, with all vertex degrees referring to the whole graph.

For the three examples,

$$^1\chi.\text{IFTI}\,(F1) = 2(1 \times 2)^{-1/2} = 1.4142; \quad ^1\chi.\text{IFTI}\,(G1 - F1) = 2(1 \times 3)^{-1/2}$$

$$+ \ (1 \times 2)^{-1/2} + (2 \times 3)^{-1/2} = 2.2700$$

$$^1\chi(G1) = 3.5536; \qquad ^1\chi.\text{EFTI}\,(F1) = -0.1306$$

$$^1\chi.\text{IFTI}\,(F2) = 2(1 \times 2)^{-1/2} + (2 \times 2)^{-1/2} = 1.9142; \quad ^1\chi.\text{IFTI}\,(G2 - F2)$$

$$= 6(2 \times 2)^{-1/2} = 3.0000$$

$$^1\chi(G2) = 4.8427; \qquad ^1\chi.\text{EFTI}\,(F2) = -0.0715$$

$$^1\chi.\text{IFTI}\,(F3) = 6(2 \times 2)^{-1/2} = 3.00; \quad ^1\chi.\text{IFTI}\,(G3 - F3) = 2[2(1 \times 2)^{-1/2}$$

$$+ \ (2 \times 2)^{-1/2}] = 3.8284$$

$$^1\chi(G3) = 6.9327; \qquad ^1\chi.\text{EFTI}\,(F3) = 0.1043.$$

3.4. GENERALIZED CONNECTIVITY [9,20] ($^h\chi$) FRAGMENT TIs

If instead of edges (paths of length one) in the Randić index one takes into account larger paths (length $h = 2, 3$, etc.), one obtains by a formula similar to that of the Randić index the generalized connectivity $^h\chi$:

$$^h\chi.\text{IFTI}\,(F) = \sum_{(h)\,\text{paths} \in F} (v_i v_j \ldots v_{h+1})^{-1/2}$$

$$^h\chi.\text{IFTI}\,(G - F) = \sum_{(h)\,\text{paths} \in (G - F)} (v_i v_j \ldots v_{h+1})^{-1/2} \tag{15}$$

$$^h\chi.\text{EFTI}\,(F) = {}^h\chi(G) - \sum_{(h)\,\text{paths} \in F} (v_i v_j \ldots v_{h+1})^{-1/2}$$

$$- \sum_{(h)\,\text{paths} \in (G - F)} (v_i v_j \ldots v_{h+1})^{-1/2} .$$

3.5. VALENCE CONNECTIVITY [9,21] (${}^h\chi^v$) FRAGMENT TIs

If, according to Kier and Hall, one employs in the previous cases the "atom connectivity" Δ_i^v, defined according to the chemical nature of the atom including its unshared electrons and its multiple bonding for edges ($h = 1$) or larger paths ($h > 1$), instead of the vertex degree, one obtains the valence connectivity.

$$
\begin{aligned}
{}^h\chi^v.\mathrm{IFTI}\,(F) &= \sum_{(h)\,\mathrm{paths}\,\in F} (\Delta_i^v\,\Delta_j^v \ldots \Delta_{h+1}^v)^{-1/2} \\
{}^h\chi^v.\mathrm{IFTI}\,(G-F) &= \sum_{(h)\,\mathrm{paths}\,\in (G-F)} (\Delta_i^v\,\Delta_j^v \ldots \Delta_{h+1}^v)^{-1/2} \qquad (16) \\
{}^h\chi^v.\mathrm{EFTI}\,(F) &= {}^h\chi^v(G) - \sum_{(h)\,\mathrm{paths}\,\in F} (\Delta_i^v\,\Delta_j^v \ldots \Delta_{h+1}^v)^{-1/2} \\
&\quad - \sum_{(h)\,\mathrm{paths}\,\in (G-F)}{}' (\Delta_i^v\,\Delta_j^v \ldots \Delta_{h+1}^v)^{-1/2}.
\end{aligned}
$$

3.6. EDGE ADJACENCIES [18] (EA, GORDON – SCANTLEBURY INDEX) FRAGMENT TIs

$$
\mathrm{EA.IFTI}\,(F) = \tfrac{1}{2}\sum_{\{ij\}\in F} e_{ij}; \quad \mathrm{EA.IFTI}\,(G-F) = \tfrac{1}{2}\sum_{\{ij\}\in (G-F)} e_{ij}. \qquad (17)
$$

In the above expression, e_{ij} indicates the number of edges adjacent to edge $\{ij\}$ in the subgraphs. The index EA is half the first neighbour sum and half the Platt index [22]

$$
\begin{aligned}
\mathrm{EA.EFTI}\,(F) &= \tfrac{1}{2}\left[\sum_{\{ij\}\in G} e_{ij} - \sum_{\{ij\}\in F} e_{ij} - \sum_{\{ij\}\in (G-F)} e_{ij}\right] \\
&= 2\sum_{\{ij\}\in \mathrm{fce}} (v_i + v_j - 2). \qquad (18)
\end{aligned}
$$

For the three examples,

EA.IFTI $(F1) = 2$; EA.IFTI $(G1 - F1) = 8$; $\mathrm{EA}(G1) = \sum_{\{ij\}\in G} e_{ij} = 18$,

EA.EFTI $(F1) = 8$;

EA.IFTI $(F2) = 4$; EA.IFTI $(G2 - F2) = 12$; $\mathrm{EA}(G2) = 24$,

EA.EFTI $(F2) = 8$;

$$\text{EA.IFTI}\,(F3) = 12; \quad \text{EA.IFTI}\,(G3 - F3) = 8; \quad \text{EA}(G3) = 44,$$

$$\text{EA.EFTI}\,(F3) = 24.$$

3.7. WIENER INDEX [1] (W) FRAGMENT TIs

$$W.\text{IFTI}\,(F) = \tfrac{1}{2} \sum_{ij \in F} d_{ij}; \quad W.\text{IFTI}\,(G - F) = \tfrac{1}{2} \sum_{k} \sum_{ij \in (G-F)_k} d_{ij}$$

$$W.\text{EFTI}\,(F) = W(G) - \tfrac{1}{2} \sum_{ij \in F} d_{ij} - \tfrac{1}{2} \sum_{k} \sum_{ij \in (G-F)_k} d_{ij}\,. \tag{19}$$

In the above expressions, $W(G)$ is half the sum of all entries in the distance matrix $\boldsymbol{D}(G)$. Taking fig. 1 into consideration, it is evident that W.EFTI (F) is the sum of entries in one hatched area of fig. 1a or in two such areas of fig. 1b. In fact, it is the sum of topological distances (before fragmentation) between vertices of subgraphs which may become disconnected on fragmentation. Another important item to be noted is that when $G - F$ is a disconnected graph, each IFTI $(G - F)_k$ contains the distances between its own vertices, but the distances between vertices belonging to disconnected $(G - F)_k$ are not taken into consideration, since they contribute to EFTI (F).

$$W.\text{IFTI}\,(F1) = 4; \quad W.\text{IFTI}\,(G1 - F1) = 18; \quad W(G1) = 65;$$

$$W.\text{EFTI}\,(F1) = 43,$$

$$W.\text{IFTI}\,(F2) = 10; \quad W.\text{IFTI}\,(G2 - F2) = 27; \quad W(G2) = 121;$$

$$W.\text{EFTI}\,(F2) = 84,$$

$$W.\text{IFTI}\,(F3) = 27; \quad W.\text{IFTI}\,(G3 - F3) = 20; \quad W(G3) = 279;$$

$$W.\text{EFTI}\,(F3) = 232.$$

3.8. INFORMATION CONTENT FOR THE MAGNITUDE OF DISTANCES [4,10] (I_{D}^{M}) FRAGMENT TIs

In order to reduce the degeneracy of TIs obtained from graph invariants, one may apply information theory to the set of numbers from which the TI is obtained, taking into account the distribution of their magnitudes or their (in)equality; as is known, the larger the inequality, the larger the information content according to the Shannon formula. Bonchev and Trinajstić [4] first used information theory for the purpose of improving TIs based on the distance matrices. The information indices for the magnitude of distances i lead to the fragment TIs:

$$I_D^M.\mathrm{IFTI}\,(F) = -\sum_{i\in F} g_i \frac{i}{W} \log_2 \frac{i}{W}\ ;$$

$$I_D^M.\mathrm{IFTI}\,(G-F) = -\sum_k \sum_{j\in(G-F)_k} g_j \frac{j}{W} \log_2 \frac{j}{W}\ . \tag{20}$$

where each distance i and j within F or $(G - F)$, respectively, occurs g_i times,

$$I_D^M.\mathrm{EFTI}\,(F) = I_D^M(G) + \sum_{i\in F} g_i \frac{i}{W} \log_2 \frac{i}{W} + \sum_k \sum_{j\in(G-F)_k} g_j \frac{j}{W} \log_2 \frac{j}{W}\ . \tag{21}$$

The amount of information is calculated in eqs. (20) and (21) in bits per unit distance.

$I_D^M.\mathrm{IFTI}\,(F1) = 1.5$; $\quad I_D^M.\mathrm{IFTI}\,(G1-F1) = 3.1974$

$4\,\{2\times 1,\ 1\times 2\}$ $\quad 18\,\{4\times 1,\ 4\times 2,\ 2\times 3\}$

$I_D^M(G1) = 4.6679$; $\quad I_D^M.\mathrm{EFTI}\,(F2) = 4.6679 - 1.5 - 3.1974 = -0.0292$

$65\,\{7\times 1,\ 9\times 2,\ 8\times 3,\ 4\times 4\}$

$I_D^M.\mathrm{IFTI}\,(F2) = 2.4464$; $\quad I_D^M.\mathrm{IFTI}\,(G2-F2) = 3.7821$

$10\,\{3\times 1,\ 2\times 2,\ 1\times 3\}$ $\quad 27\,\{6\times 1,\ 6\times 2,\ 3\times 3\}$

$I_D^M(G2) = 5.3135$; $\quad I_D^M.\mathrm{EFTI}\,(F2) = -0.9150$

$121\,\{10\times 1,\ 12\times 2,\ 11\times 3,\ 7\times 4,\ 4\times 5,\ 1\times 6\}$

$I_D^M.\mathrm{IFTI}\,(F3) = 3.7821$; $\quad I_D^M.\mathrm{IFTI}\,(G3-F3) = 4.8929$

$27\,\{6\times 1,\ 6\times 2,\ 3\times 3\}$ $\quad 108\,\{6\times 1,\ 4\times 2,\ 2\times 3,\ 2\times 4,\ 6\times 5,\ 6\times 6,\ 2\times 7\}$

$I_D^M(G3) = 6.3178$; $\quad I_D^M.\mathrm{EFTI}\,(F3) = -2.3585$

$279\,\{16\times 1,\ 22\times 2,\ 21\times 3,\ 14\times 4,\ 10\times 5,\ 6\times 6,\ 2\times 7\}$

3.9 HOSOYA INDEX [2] (Z) FRAGMENT TIs

Hosoya invented an interesting and useful TI by summing the non-adjacency number $p(G, l)$ for all l values, where $p(G2)$ is the number of ways in which l edges may be chosen from the graph G so that no two of them are adjacent. By definition, $p(G, 0) = 1$ and $p(G, 1) = q$, the number of graph edges.

$$Z.\mathrm{IFTI}\,(F) = \sum_l p(F, l); \quad Z.\mathrm{IFTI}\,(G - F) = \sum_k \sum_l p[(G - F), l]$$

$$Z.\mathrm{EFTI}\,(F) = Z(G) - \sum_l p(F, l) - \sum_k \sum_l p[(G - F)_k, l] \,. \tag{22}$$

$Z.\mathrm{IFTI}\,(F1) = 1 + 2 = 3;$ $\quad Z.\mathrm{IFTI}\,(G1 - F1) = 1 + 4 + 2 = 7$

$Z(G1) = 1 + 7 + 12 + 4 = 24;$ $\quad Z.\mathrm{EFTI}\,(F1) = 14$

$Z.\mathrm{IFTI}\,(F2) = 1 + 3 + 1 = 5;$ $\quad Z.\mathrm{IFTI}\,(G2 - F2) = 1 + 6 + 9 + 2 = 18$

$Z(G2) = 1 + 10 + 33 + 42 + 18 + 2 = 106;$ $\quad Z.\mathrm{EFTI}\,(F2) = 83$

$Z.\mathrm{IFTI}\,(F3) = 18;$ $\quad Z.\mathrm{IFTI}\,(G3 - F3) = 2.5 = 10$

$Z(G3) = 1 + 16 + 95 + 290 + 429 + 294 + 76 + 4 = 1205;$ $\quad Z.\mathrm{EFTI}\,(F3) = 1177.$

3.10 HIERARCHICALLY ORDERED EXTENDED CONNECTIVITIES (HOC) INDICES [23] ($\mathcal{M}$ AND $\mathcal{N}$) FRAGMENT TIs

It was shown [23] that the HOC vertex ordering can be used for devising two TIs, denoted by $\mathcal{M}$ and $\mathcal{N}$, respectively. Both indices start from a unique vertex numbering and result in a sequence of sums (S_i) of vertex numbering (each vertex k is labelled with the sum of vertex labels for its adjacent vertices except for vertices with labels lower than k. These initial sequences of sums (the first line in the calculation of each example) are then transformed following Muirhead [24] in their final form (the second line in the three examples).

$$M_i = \sum_j S_j, \quad \mathcal{M} = \sum_i M_i, \quad \mathcal{N} = \sum_i M_i^2 \,. \tag{23}$$

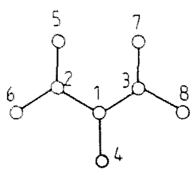
For the fragment topological indices $\mathcal{M}$ we have:

$$\mathcal{M}.\mathrm{IFTI}\,(F) = \sum_{i\in F} M_i; \quad \mathcal{M}.\mathrm{IFTI}\,(G-F) = \sum_{k} \sum_{j\in (G-F)_k} M_j$$

$$\mathcal{M}.\mathrm{EFTI}\,(F) = \mathcal{M}(G) - \sum_{i\in F} M_i - \sum_{k} \sum_{j\in (G-F)_k} M_j\,. \tag{24}$$

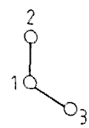
On replacing M_1 in the above three formulas by M_1^2, we obtain analogously the $\mathcal{M}$.FTIs.

Block 4



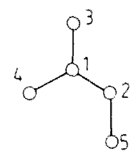
9, 11, 15, 0, 0, 0, 0, 0
9, 20, 35, 35, 35, 35, 35, 35

Block 5



5, 0, 0
5, 5, 5

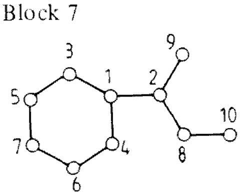
Block 6



9, 5, 0, 0, 0
9, 14, 14, 14, 14

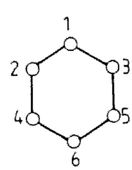
$\mathcal{M}$.IFTI ($F1$) = 15; $\mathcal{M}$.IFTI ($G1 - F1$) = 65
$\mathcal{M}(G1)$ = 239; $\mathcal{M}$.EFTI ($F1$) = 159
$\mathcal{N}$.IFTI ($F1$) = 75; $\mathcal{N}$.IFTI ($G1 - F1$) = 865
$\mathcal{N}(G1)$ = 7831; $\mathcal{N}$.EFTI ($F1$) = 6891

Block 7



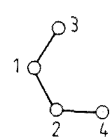
9, 17, 5, 6, 7, 7, 0, 10, 0, 0
9, 26, 31, 37, 44, 51, 51, 61, 61, 61

Block 8



5, 4, 5, 6, 6, 0
5, 9, 14, 20, 26, 26

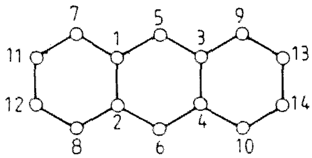
Block 9



5, 4, 0, 0
5, 9, 9, 9

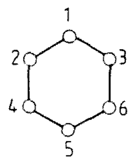
$\mathcal{M}$.IFTI ($F2$) = 32; $\mathcal{M}$.IFTI($G2 - F2$) = 100
$\mathcal{M}(G2)$ = 432; $\mathcal{M}$.EFTI ($F2$) = 300
$\mathcal{N}$.IFTI ($F2$) = 268; $\mathcal{N}$.IFTI ($G2 - F2$) = 2054
$\mathcal{N}(G2)$ = 21388; $\mathcal{N}$.EFTI ($F2$) = 19066

Block 10



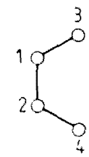
14, 14, 18, 16, 0, 0, 11, 12, 13, 14, 12, 0, 14, 0
14, 28, 46, 62, 62, 62, 73, 85, 98, 112, 124, 124, 138, 138

Block 11



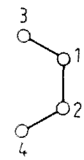
5, 4, 5, 6, 6, 0
5, 9, 14, 20, 26, 26

Block 12



5, 4, 0, 0
5, 9, 9, 9

Block 13



5, 4, 0, 0
5, 9, 9, 9

M.IFTI ($F3$) = 100; M.IFTI ($G3 - F3$) = 2.32 = 64
$M(G3)$ = 1166; M.EFTI ($F3$) = 1037
N.IFTI ($F3$) = 2054; N.IFTI ($G3 - F3$) = 2.268 = 536
$N(G3)$ = 118170; N.EFTI ($F3$) = 115580.

3.11. AVERAGE DISTANCE SUM CONNECTIVITY [5,14,25,26] (BALABAN INDEX J) FRAGMENT TIs

The index J results by applying a Randić-type formula to average distance sums s_i/q (instead of vertex degrees as for χ), and by normalizing with respect to the cyclomatic number μ. For the present application, no normalization factor is used because the number of graph edges q and cycles μ could undergo drastic changes upon fragmentation.

$$J.\mathrm{IFTI}(F) = \sum_{\{ij\}\in F} (s_i s_j)^{-1/2}; \quad J.\mathrm{IFTI}(G-F) = \sum_k \sum_{\{ij\}\in (G-F)_k} (s_i s_j)^{-1/2}$$

$$J.\mathrm{EFTI} = \sum_{\{ij\}\in G} (s_i s_j)^{-1/2} - \sum_{\{ij\}\in F} (s_i s_j)^{-1/2} - \sum_k \sum_{\{ij\}\in (G-F)_k} (s_i s_j)^{-1/2} \tag{25}$$

$J.\mathrm{IFTI}(F1) = 2(2\times 3)^{-1/2} = 0.8165;$

$J.\mathrm{IFTI}(G1 - F1) = 2(5\times 8)^{-1/2} + (5\times 6)^{-1/2} + (6\times 9)^{-1/2} = 0.6349$

$J(G1) = 4(13\times 19)^{-1/2} + 2(13\times 11)^{-1/2} + (11\times 17)^{-1/2} = 0.4949;$

$J.\mathrm{EFTI}(F1) = -0.9565$

$J.\mathrm{IFTI}(F2) = 2(4\times 6)^{-1/2} + (4\times 4)^{-1/2} = 0.6582;$

$J.\mathrm{IFTI}(G2 - F2) = 6(9\times 9)^{-1/2} = 0.6661$

$J.\mathrm{EFTI}(F2) = -0.8770$

$$J(G2) = 2(17.21)^{-1/2} + 2(21.25)^{-1/2} + 2(25.29)^{-1/2} + (17.19)^{-1/2} + (19.25)^{-1/2}$$

$$+ \ (19.27)^{-1/2} + (25.33)^{-1/2} = 0.4479$$

$$J.\mathrm{IFTI}\,(F3) = 6(9 \times 9)^{-1/2} = 0.6667;$$

$$J.\mathrm{IFTI}\,(G3 - F3) = 2\,[2(6 \times 4)^{-1/2} + (4 \times 4)^{-1/2}] = 1.3165$$

$$J(G3) = 6(33.33)^{-1/2} + 4(33 \times 41)^{-1/2} + 4(41.49)^{-1/2} + 2(49.49)^{-1/2} = 0.4206$$

$$J.\mathrm{EFTI}\,(F3) = -1.5626.$$

3.12. OTHER INFORMATION THEORETIC [4,11,27] (I_i^{E}) FRAGMENT TIs

Several information theoretic TIs have been defined on the basis of the equivalency (equality) of graph distances ($I_{\mathrm{D}}^{\mathrm{E}}$), graph vertices according to their chromatic ($I_{\mathrm{CHR}}^{\mathrm{E}}$), orbital ($I_{\mathrm{ORB}}^{\mathrm{E}}$), or centric ($I_{\mathrm{C}}^{\mathrm{E}}$) partitioning, Hosoya's non-adjacency numbers ($I_{\mathrm{Z}}^{\mathrm{E}}$), partial Randić connectivities (I^{E}), etc.

In all cases, one applies Shannon-type formulas for the finite probability schemes based on the respective distribution, as shown above for $I_{\mathrm{D}}^{\mathrm{M}}$.

In the following, we give a detailed example for the calculation of $I_{\mathrm{Z}}^{\mathrm{E}}$ fragment indices for the three examples, as well as the numerical results for five other I^{E}-type information indices. In all cases, the average values of these indices are presented (in bits per vertex, edge, distance, etc.).

$$I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{IFTI}\,(F1) = \underset{3\{1,2\}}{0.9183}; \qquad I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{IFTI}\,(G1 - F1) = \underset{7\{1,4,2\}}{1.3788}$$

$$I_{\mathrm{Z}}^{\mathrm{E}}(G1) = \underset{24\{1,7,12,4\}}{1.6403}; \qquad I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{EFTI}\,(F1) = -0.6568$$

$$I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{IFTI}\,(F2) = \underset{5\{1,3,1\}}{1.3709}; \qquad I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{IFTI}\,(G2 - F2) = \underset{18\{1,6,9,2\}}{1.6122}$$

$$I_{\mathrm{Z}}^{\mathrm{E}}(G2) = \underset{105\{1,10,33,42,18,2\}}{1.9806}; \qquad I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{EFTI}\,(F2) = -1.0025$$

$$I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{IFTI}\,(F3) = \underset{18\{1,6,9,2\}}{1.6122}; \qquad I_{\mathrm{Z}}^{\mathrm{E}}.\mathrm{IFTI}\,(G3 - F3) = \underset{5\{1,3,1\} + 5\{1,3,1\}}{2.1.3709} = 2.7418$$

$$I_Z^E(G3) = 2.1806; \qquad I_Z^E.\mathrm{EFTI}\,(F3) = -2.1732$$

$$1205\{1, 16, 95, 290, 429, 294, 76, 4\}.$$

The results obtained for some other information theoretic indices are given in table 1.

Table 1

Five fragment information indices based on the equivalence (equality) of specified graph characteristics

Index	Graph	IFTI (F)	IFTI ($G-F$)	TI (G)	EFTI (F)
I_D^E	$G1$	0.9183	1.5219	1.9438	−0.4964
	$G2$	1.4592	1.5219	2.3375	−0.6435
	$G3$	1.5219	2.9179	0.9663	−3.4734
I_{CHR}^E	$G1$	0.9183	0.9709	0.9544	−0.9349
	$G2$	1.	1.	1.	−1.
	$G3$	1.	2.	1.	−2.
I_{ORB}^E	$G1$	0.9183	1.9219	1.7500	−1.0902
	$G2$	1.	0.	2.9219	1.9219
	$G3$	0.	2.	1.9502	−0.0498
I_C^E	$G1$	0.9183	1.9219	0.9544	−1.0902
	$G2$	1.	0.	2.9219	1.9219
	$G3$	0.	2.	1.9502	−0.0498
I_X^E	$G1$	0.	1.5	0.8631	−0.6369
	$G2$	0.9183	0.	2.0464	1.1281
	$G3$	0.	1.8355	1.4059	0.4873

It can clearly be seen from all examples and tables 1–3 that fragment topological indices obtained as indicated in the present paper evidence interesting regularities.

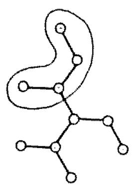
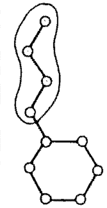
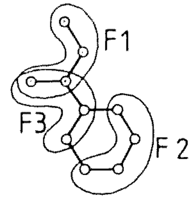
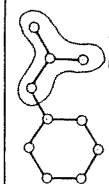
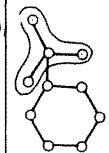
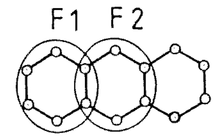
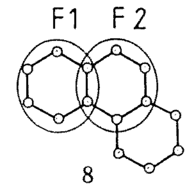
4. Results and discussion

4.1. FURTHER NUMERICAL RESULTS

In addition to the illustrative examples given in the preceding section, we selected a few FTIs and a few graphs to investigate the general behaviour of IFTI(F) and EFTI (F) values, as well as their changes depending on two structural features: branching and the more or less central position of the fragment in the molecule.

Table 2

Topological indices for graphs **2**–**8**, showing one below the other: TI (G), IFTI (F), and IFTI ($G-F$)

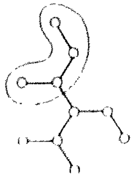
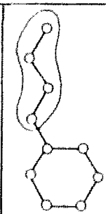
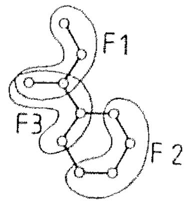
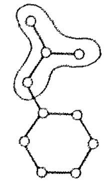
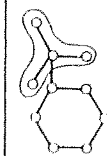
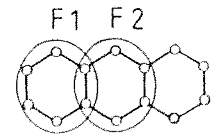
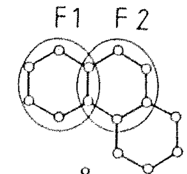
Graphs and their fragments / Topological indices		 4	 5	 2			 6	 7	 3		 8	
		F	F	F_1	F_2	F_3	F	F	F_1	F_2	F_1	F_2
Z	G	67	114	106			96	80	1205		1230	
	F	5	5	5			4	4	18		18	
	$G-F$	11	18	18	12	7	18	18	38	10	44	10
$\mathcal{N}$	G	21138	16257	21388			25166	27548	118170		116962	
	F	268	268	268			324	324	2054		2054	
	$G-F$	1917	2054	2054	1877	276	2054	2054	9510	536	7110	536
M_1	G	40	42	44			44	48	76		76	
	F	10	10	10			12	12	24		24	
	$G-F$	20	24	24	20	12	24	24	36	20	34	20
W	G	122	133	121			126	144	279		271	
	F	10	10	10			9	9	27		27	
	$G-F$	32	27	27	31	11	27	27	60	20	64	20

I_D^M	G	5.3415	5.2732	5.3135			5.2958	5.3317	6.3165		6.3316	
	F	2.4464	2.4464	2.4464			2.5033	2.5033	3.7821		3.7821	
	$G-F$	3.7417	3.7821	3.7821	3.7600	2.4464	3.7821	3.7821	4.6729	4.8929	4.6423	4.8929
$^1\chi$	G	4.6292	4.9318	4.8427			4.7877	4.6052	6.9327		6.9495	
	F	1.9142	1.9142	1.9142			1.7321	1.7321	3.00		3.00	
	$G-F$	2.7700	3.0000	3.0000	2.8080	2.9142	3.0000	3.0000	3.8045	3.8284	3.9318	3.8284
J	G	0.4220	0.4034	0.4479			0.4263	0.4778	0.4206		0.4351	
	F	0.6582	0.6582	0.6582			0.7746	0.7746	0.6667		0.6667	
	$G-F$	0.5254	0.6667	0.6667	0.5508	1.6582	0.6667	0.6667	0.5698	1.3165	0.5312	1.3165

Table 2 (continued)

Table 3

Fragment topological indices for graphs **2** – 8: EFTI (F), NIFTI (F), values

Graphs and their fragments / Fragment topological indices		 4	 5	 2			 6	 7	 3		 8	
		F	F	F_1	F_2	F_3	F	F	F_1	F_2	F_1	F_2
Z	NIFTI	0.0746	0.0439	0.0472			0.0417	0.0500	0.0149		0.0146	
	EFTI	51	91	83	89	94	74	58	1149	1177	1168	1202
	NEFTI	0.7612	0.7982	0.7830	0.8396	0.8868	0.7708	0.7250	0.9535	0.9768	0.9496	0.9772
$\mathcal{N}$	NIFTI	0.0127	0.0165	0.0125			0.0129	0.0118	0.0168		0.0176	
	EFTI	18953	13935	19066	19243	20844	22788	25170	106606	115580	107798	114372
	NEFTI	0.8966	0.8572	0.8914	0.8997	0.9746	0.9055	0.9137	0.9021	0.9781	0.9216	0.9779
M_1	NIFTI	0.2500	0.2381	0.2273			0.2727	0.25	0.3158		0.3158	
	EFTI	10	8	10	14	22	8	12	16	32	18	32
	NEFTI	0.2500	0.1905	0.2273	0.3182	0.5000	0.1818	0.2500	0.2105	0.4211	0.2368	0.4211
W	NIFTI	0.0820	0.0752	0.0826			0.0714	0.0789	0.0968		0.0996	
	EFTI	80	96	84	80	100	90	78	192	232	180	224
	NEFTI	0.6557	0.7218	0.6942	0.6612	0.8264	0.7143	0.6842	0.6882	0.8315	0.6642	0.8266

I_D^M	NIFTI	0.4580	0.4639	0.4604			0.4727	0.4695	0.5988		0.5973	
	EFTI	− 0.8466	− 0.9553	− 0.9150	− 0.8929	0.4207	− 0.9896	− 0.9537	− 2.1385	− 2.3585	− 2.0928	− 2.3434
	NEFTI	− 0.1585	− 0.1812	− 0.1722	− 0.1680	0.0792	− 0.1869	− 0.1789	− 0.3386	− 0.3734	− 0.3305	− 0.3701
$^1\chi$	NIFTI	0.4135	0.3881	0.3953			0.3618	0.3761	0.4327		0.4317	
	EFTI	− 0.0550	0.0176	− 0.0715	0.1205	0.0143	0.0556	− 0.1269	0.1282	0.1043	0.0177	0.1211
	NEFTI	− 0.0119	0.0036	− 0.0148	0.0249	0.0030	0.0116	− 0.0276	0.0185	0.0150	0.0025	0.0174
J	NIFTI	0.1733	0.3263	0.2939			0.3635	0.3242	0.3963		0.3831	
	EFTI	− 0.7616	− 0.9215	− 0.8770	− 0.7611	− 1.8685	− 1.0150	− 0.9635	− 0.8159	− 1.5626	− 0.7628	− 1.5481
	NEFTI	− 1.8047	− 2.2843	− 1.9580	− 1.6993	− 4.1717	− 2.3810	− 2.0165	− 1.9398	− 3.7152	− 1.7532	− 3.5580

Table 3 (continued)

Graph **4** is acyclic, graphs **5**–**2** – **6**–**7** are monocyclic and isomeric, differing in the branching of the side chain, while the last two graphs (**3** and **8**) are again isomeric with one another and differ in the mode of ring condensation (angular versus linear). The position of the fragments in the molecule can differ, as indicated for **2**, **3**, and **8**.

Table 2 presents the topological indices for the whole initial graph G, for the fragment F, and for the remainder of the molecule $(G - F)$ for seven TIs. Table 3 indicates the fragment topological indices EFTI (F) and the corresponding normalized NIFTI and NEFTI indices.

4.2. INTERNAL FRAGMENT TOPOLOGICAL INDICES (IFTI) AND NIFTI

It can be seen that one and the same fragment has the same IFTI, irrespective of the molecule from which it originates (e.g. IFTI values are the same for graphs **4**, **5**, and **2**, or for graphs **6** and **7**, or for graphs **3** and **8**). On the other hand, different isomeric fragments have different IFTI values, as shown by comparing fragments in graphs **4**, **5**, and **2** with isomeric fragments in graphs **6** and **7**.

The normalized NIFTI indices have different values for one and the same fragment originating from different graphs; both the size and the shape of the whole molecule influence the NIFTI value. Only when one and the same fragment is cut out from the same molecule in different ways are the corresponding NIFTI values equal, as shown by the three fragmentation modes of graphs **2**, **3**, and **8** (actually, in these cases in table 3, the equal values are not repeated).

The examples given in subsect. 3.12 and table 1 provide another conclusion which is important for the applicability of the fragment topological indices. The fragment information indices based on equivalency (equality) of the graph elements (characteristics) I_i^{E} do not obey requirement (1) formulated in the foregoing text. As can be seen, for example, for graph **3**, all IFTI $(G - F)$ values are larger than those of the whole graph G. Also, $I_{\mathrm{CHR}}^{\mathrm{E}}$.IFTI $(F3) = I_{\mathrm{CHR}}^{\mathrm{E}}(G)$, etc. Thus, any of the information theoretic indices ($I_{\mathrm{D}}^{\mathrm{E}}$, $I_{\mathrm{Z}}^{\mathrm{E}}$, $I_{\mathrm{ORB}}^{\mathrm{E}}$, $I_{\mathrm{CHR}}^{\mathrm{E}}$, $I_{\mathrm{C}}^{\mathrm{E}}$, I_{χ}^{E}, etc.) based on equivalence relations of the distribution elements can be used as fragment topological indices. Instead, the graph characteristics partitioning used in specifying the I_i^{E} indices can be treated in a different manner, e.g. by means of a quadratic function, as is done, for instance, for the graph centric indices [27]. This, however, seems unreasonable, at least for some of these cases, since the centric, orbital, chromatic, etc. properties of graphs are radically changed upon the fragment excision.

4.3. EXTERNAL FRAGMENT TOPOLOGICAL INDICES EFTI AND NEFTI. GENERAL REGULARITY ANALYSIS

Here we trace how the examined external fragment indices follow the seniority relations occurring for the respective topological indices for the whole graph, as formulated by requirement (5). This requirement appears important because the

FTIs should reflect the fragment topology in the same manner as TIs do with the topology of the entire graph. An inspection of table 3 shows that with very few exceptions this is actually the case. The same trend in reflecting the graph and fragment structural patterns is found for the Hosoya index Z, the Wiener index W, the information index on the distance magnitude $I_{\mathrm{D}}^{\mathrm{M}}$, and the Balaban index J. For example,

	Graphs					
	5	2	6	7	3	8
$Z(G)$	114 >	106	96 >	80	1205 <	1230
Z.EFTI (F)	91 >	83	74 >	58	1177 <	1202
					1149 <	1168

Indeed, the comparison is made at a constancy of the structural factors: the fragment centric location, the $(G - F)$ branching and cyclicity, etc. For this reason, in the above examples we do not compare graphs **4** and **2** (or **5**), since the first one is acyclic when the second one is a cyclic graph. One exception to this rule is detected for fragment 1 when comparing the HOC index or the Randić index of graphs **3** and **8**. These deviations are small and should not be regarded as evidence against the applicability of the two fragment topological indices.

4.4. EFTI AND NEFTI INDICES AND MOLECULAR BRANCHING

By comparing EFTI (or NEFTI) indices for the graph series **5**, **2**, **6**, **7**, one could obtain some information on the degree to which these fragment topological indices reflect molecular branching as one of the major topological features of molecules. Molecular branching has been a subject of intensive graph-theoretical studies [2–14,17,18,28–33]. An attempt was made to express the essence of branching by a series of structural rules based on the graph distances (the Wiener number W) [4]. Most of these rules are reflected also by the Randić molecular connectivity and the Hosoya non-adjacency index; the ordering of isomeric compounds they provide was shown to be followed by many molecular properties [4,33]. In dealing with isomeric structures **5**, **2**, **6** and **7**, the different topological indices disagree as to which of graphs **2** and **6** is more branched. They do, however, indicate in full accord structure **5** as the least branched and structure **7** as the most branched. Thus, the ordering **5** > **6**, **2** > **7** is produced by W, ${}^1\chi$, and Z, and the reverse ordering **5** < **6**, **2** < **7** results for $I_{\mathrm{D}}^{\mathrm{M}}$, J, $\mathcal{N}$, and M_1. Only some of the fragment indices follow this order. These are the EFTI and NEFTI for the Wiener number, the Hosoya index, and the HOC index. The Zagreb index M_1, which is strongly degenerate, deviates from the expected ordering, showing the EFTI values of **5** and **6** to be the same, while the NEFTI values are even in a reversed order. The other three examined

EFTIs completely disagree with the ordering observed for the respective TIs, qualifying graph **6** as the least branched and graph **2** as the most branched one. The J.NEFTI and I_D^M.NEFTI follow the same trend, some improvement being found only for ${}^1\chi$.NEFTI, where structure **7** is restored as the most branched one.

The reason for the failure of J-, I_D^M-, and ${}^1\chi$-type fragment indices to reflect molecular branching correctly can be traced back to the mathematical functions used. ${}^1\chi$ and J are sums of terms. The number of these terms diminishes by one for each edge which is cut during fragment excision. The opposite influence on the magnitude of these indices, however, results from the increasing values of all the remaining terms, due to the decrease of the vertex degrees or distance sums which constitute the terms denominator (see eqs. (14) and (25)). The regularity in varying molecular branching could thus be lost in the counterbalance of these two opposing trends. On the other hand, $I_D^M(G)$ enhances with branching and so does I_D^M.IFTI (F), which is subtracted from $I_D^M(G)$. Two opposing factors thus again emerge (the IFTI $(G - F)$ term is constant) which may cause violations to the regular trend dictated by molecular branching.

4.5. EFTI AND NEFTI INDICES AND THE FRAGMENT CENTRIC LOCATION

Another test for the qualities of these fragment topological indices could be the comparison of their values in the case of different fragment locations in the molecule. One may expect a regular change in EFTI and NEFTI values upon a consecutive fragment removal from a more central position. With this purpose in mind, we compared fragmentations $F1-F3$ of graph **2** and fragmentations $F1$ and $F2$ of graphs **3** and **8**.

The anticipated regular increase in both EFTI and NEFTI indices was found for the Hosoya, HOC, and Zagreb indices on removing the fragment from a marginal to central position, i.e. in the series $F1-F2-F3$ for **2** or $F1-F2$ for **3** and **8**. This increase is due mainly to the fact that a more central fragment is formed by breaking more bonds, creating more endpoints in the fragment(s). The other topological indices deviate more or less from the regular trend. Thus, one such deviation (underlined) is found for the Wiener index $(\underline{F2-F1}-F3)$ and its information-theoretic analog $I_D^M(F2-F1)$, as well as for the Balaban index J assuming a reverse fragment ordering i.e. a decrease with the fragment in a more central position $(\underline{F2-F1}-F3, F1-F2)$. Once again, the Randić molecular connectivity index completely fails to reproduce the expected regular trend $(F1-\underline{F3-F2}, \underline{F2-F1}(G3), F1-F2(G8))$.

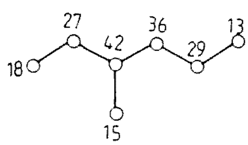
5. New local (vertex) graph invariants: an infinity of new vertex invariants based on EFTI

When considering a fragment of one non-hydrogen atom, the EFTI(1) value reduces to

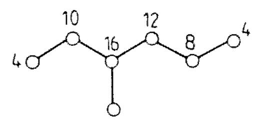
$$\mathrm{EFTI}(1) = \mathrm{TI}\,(G) - \mathrm{IFTI}\,(G') - a, \tag{26}$$

where G' is the vertex-excised graph, i.e. the initial graph from which the given vertex and its adjacent edges have been removed. a stands for IFTI (F) and is zero in the majority of cases or is a constant (e.g. a = 1 for the Hosoya index Z).

If one now moves the one-atom fragment along the graph, one obtains for each vertex an EFTI(1) value which is a vertex invariant based on the given TI. We illustrate this by the W.EFTI(1) and M_1.EFTI(1) values for the smallest identity tree on seven vertices. It is evident that these values vary consistently towards the graph center.



W.EFTI(1)



M_1.EFTI(1)

On applying to the newly obtained vertex invariants iteratively the formula for the same TI or for a different TI, and then on recalculating EFTI(1) for each vertex, it is possible to devise an infinite number of vertex invariants. More details on this subject will be given elsewhere [34].

6. Conclusions

We have presented a new method for calculating topological indices of molecular fragments which takes into account the topological interaction between the fragment and the remainder of the molecule by the EFTI values. Whenever these interactions are unimportant, only the internal fragment topological index (IFTI) should be considered; such an index is calculated according to the usual methods employed for obtaining TIs of whole molecules. Of course, the corresponding NIFTI reflects the relative weight of the fragment in the molecule, a useful fact in QSAR. When, however, one wishes to emphasize the mode of attachment of the fragment, one may also have recourse to graph-theoretical methods based on rooted graphs.

Some of the selected TIs are well suited for applications as fragment TIs, having a regular variation with respect to structural changes such as branching, central versus marginal location of the fragment, etc. These are the simpler TIs, such as Z, $\mathcal{N}$, W, or M_1. Other TIs, such as I_D^M or J, respond only partly in a regular manner to structural changes, while the Randić connectivity index $^1\chi$ perhaps needs to be renormalized so as to become widely applicable as a fragment topological index. However, depending on the experimental data to be correlated with the structure, even "irregularly" varying indices may be tested.

In practice, most correlations do not involve hydrocarbons, but molecules containing heteroatoms. In this case, one must parametrize graph constituents such as vertices or edges, and one must employ weighted (labelled) graphs. Several papers have described approaches to this end [9,26,35].

For fragments consisting of one non-hydrogen atom (Hal, NH_2, OH, etc.) most internal FTIs are zero, leading to non-trivial difficulties. However, EFTI (F) values differ from zero even for such fragments. Consequently, such EFTI values can some times be used as local characteristics of molecules (local vertex invariants).

Acknowledgement

This study was supported by the Bulgarian State Committee for Science (Grant No. 233).

References

[1] H. Wiener, J. Amer. Chem. Soc. 69(1947)17, 2636; J. Chem. Phys. 15(1947)766; J. Phys. Chem. 52(1948)425, 1082.

[2] H. Hosoya, Bull. Chem. Soc. Japan 44(1971)2332.

[3] M. Randić, J. Amer. Chem. Soc. 97(1975)6609.

[4] D. Bonchev and N. Trinajstić, J. Chem. Phys. 67(1977)4517; Int. J. Quant. Chem. Quant. Chem. Symp. 12(1978)293; ibid. 16(1982)463;
D. Bonchev, O. Mekenyan and N. Trinajstić, J. Comput. Chem. 2(1981)127.

[5] A.T. Balaban, Chem. Phys. Lett. 89(1982)399.

[6] M. Randić, J. Chem. Inf. Comp. Sci. 24(1984)164.

[7] A.T. Balaban, A. Chiriac, I. Motoc and Z. Simon, *Steric Fit in QSAR*, Lecture Notes in Chemistry 15 (Springer-Verlag, Berlin, 1980) ch. 2, p. 22.

[8] D.H. Rouvray, Sci. Amer. 254(1986)40; Amer. Sci. 61(1973)729; MATCH 1(1975)125; in: *Mathematical and Computational Concepts in Chemistry*, ed. N. Trinajstić (Horwood, Chichester, 1986) p. 295; J. Comput. Chem. 8(1987)470.

[9] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).

[10] N. Trinajstić, *Chemical Graph Theory*, Vol. 2 (CRC Press, Boca Raton, 1983) ch. 4, p. 105.

[11] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press, Chichester, 1983).

[12] A. Sabljić and N. Trinajstić, Acta Pharm. Jugosl. 31(1982)189.

[13] A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Top. Curr. Chem. 114(1983)21.

[14] A.T. Balaban, Pure Appl. Chem. 55(1983)199; Theor. Chim. Acta 53(1979)355.

[15] A.T. Balaban and P. Filip, MATCH 16(1984)163.

[16] D. Bonchev, O. Mekenyan and H. Fritsche, Cryst. Growth 49(1980)90.

[17] I. Gutman, M. Ruščić, N. Trinajstić and C.F. Wilcox, Jr., J. Chem. Phys. 62(1975)3399.

[18] M. Gordon and G.R. Scantlebury, Trans. Faraday Soc. 60(1964)605.

[19] A.T. Balaban, Theor. Chim. Acta 53(1979)355.

[20] L.B. Kier, L.H. Hall, W.J. Murray and M. Randić, J. Pharm. Sci. 64(1975)1971.

[21] L.B. Kier, W.J. Murray, M. Randić and L.H. Hall, J. Pharm. Sci. 65(1976)1226.

[22] J.R. Platt, J. Chem. Phys. 15(1947)419; J. Phys. Chem. 56(1952)328.
[23] O. Mekenyan, D. Bonchev and A.T. Balaban, J. Comput. Chem. 5(1984)629.
[24] R.M. Muirhead, Proc. Edinburgh Math. Soc. 19(1901)36; 21(1903)144; 24(1906)45.
[25] A.T. Balaban and L.V. Quintas, MATCH 14(1983)213;
A.T. Balaban, N. Jonescu Pallas and T.-S. Balaban, ibid. 17(1985)121.
[26] A.T. Balaban, MATCH 21(1986)115.
[27] D. Bonchev, A.T. Balaban and O. Mekenyan, J. Chem. Inf. Comput. Sci. 20(1980)106.
[28] L. Lovász and J. Pelikan, Period. Math. Hung. 3(1973)175.
[29] I. Gutman and M. Randić, Chem. Phys. Lett. 47(1977)15.
[30] D. Bonchev, J.V. Knop and N. Trinajstić, MATCH 6(1979)21.
[31] M. Barysz, J.V. Knop, S. Pejaković and N. Trinajstić, Polish J. Chem. 59(1985)405.
[32] R.E. Merrifield and H.E. Simmons, Proc. Natl. Acad. USA 78(1981)1329.
[33] D. Bonchev and O. Mekenyan, J. Chem. Soc. Trans. Faraday II 80(1984)695.
[34] A.T. Balaban, D. Bonchev and O. Mekenyan (to be published).
[35] M. Barysz, G. Jashari, R.S. Lall, V.K. Srivastava and N. Trinajstić, in: *Chemical Applications of Graph Theory and Topology*, ed. R.B. King (Elsevier, Amsterdam, 1983) p. 222.