Revue Roumaine de Chimie, 34, 6, 1361–1368 (1989)

DESIGN ON TOPOLOGICAL INDICES. 1

DEFINITION OF A VERTEX TOPOLOGICAL INDEX IN THE CASE OF 4-TREES

OVIDIU IVANCIUC
Polytechnic Institute "Traian Vuia", Faculty of Chemical Technology, Timişoara *

Received July 25, 1988

Distance matrix of a 4-tree was used in defining eighteen vertex topological indices. The best definition was chosen, which ensures a low degeneracy and a large interval of values.

1. INTRODUCTION

Molecular topology determines a large number of molecular properties ranging from physicochemical and thermodynamic properties to chemical reactivity and biological activity. Organic molecules are represented by hydrogen-depleted graphs depicting the covalent bonds between non-hydrogen atoms. In graph theory, alkanes are represented as 4-trees. A 4-tree is a connected graph without cycles and with no vertex with the degree greater than 4. The topology of a chemical structure can be coded in matrix form by the use of the adjacency matrix and the distance matrix.[1]

The adjacency matrix [2] of a graph G with N vertices, $A(G) = A$, is the square $N \times N$ symmetric matrix which contains information about the connectivity of vertices in G. Its entries are defined as :

$$a_{ij} = \begin{cases} 1, & \text{for vertices i, j adjacent} \\ 0, & \text{otherwise} \end{cases}$$

The distance matrix of a graph G with N vertices, $D(G) = D$, is a square $N \times N$ symmetric matrix, whose entries, d_{ij}, are equal to the number of edges connecting vertices i and j on the shortest path between them. The two types of matrices are interrelated and the distance matrix can be computed from the adjacency matrix using a simple algorithm based on the adjacency matrices of higher orders.[3]

From the practical point of view, an efficient way of coding the topology of a chemical structure is represented by the topological indices.[4,5] A topological index (TI) is a numerical quantity which characterizes the bonding topology of a molecule.

The problem of chemical species classification is a difficult task.

* Correspondence address : strada N. Grigorescu, 18/2 2750 Hunedoara, Romania

For example, the number of alkane constitutional isomers increases very rapidly with the number of carbon atoms, being equal with [6] 75 for $C_{10}H_{22}$, 366,319 for $C_{20}H_{42}$, 4, 111, 846, 763 for $C_{30}H_{62}$, 62, 481, 801, 147, 341 for $C_{40}H_{82}$ and 1, 117, 743, 651, 746, 953, 270 for $C_{50}H_{102}$. In order to classify such a great number of structures, topological indices with high discriminating power are strongly needed.

An almost general deficiency of topological indices is that they do not characterize uniquely the topology of a molecular graph, but are more or less degenerate, i.e. two or more nonisomorphic structures may lead to the same numerical value for a certain topological index. There are two possible exceptions: the smallest binary notation (SBN) of a graph, introduced by Randić[7–10] and according to Hosoya's conjecture,[11] the distance polynomial and the distance polynomial index Z'.

By definition, the vertex topological index (VII) is a numerical quantity associated with a certain vertex of the chemical graph, which expresses the local topology of the chemical structure.

In the present paper, a design of a VII for 4-trees is presented. The definition obtained ensures a minimum degeneracy and a good distribution of the values. A number of informational TI's are computed, based on the defined VTI's. The rules of branching[12] at constant number of vertices are tested for the computed TI's.

2. THE DEFINITION OF THE VERTEX TOPOLOGICAL INDEX

The molecular topology will be defined by the aid of the distance matrix, because it contains a larger quantity of information concerning the molecular structure. In equations (1)—(18), eighteen types of VTI's are defined. The valence (degree) v_i of the vertex i is equal to the number of neighbours of the vertex i.

$$\text{VTI-1}_i = \sum_{j=1}^{N} d_{ij} \tag{1}$$

$$\text{VTI-2}_i = \sum_{j=1}^{N} d_{ij} v_i \tag{2}$$

$$\text{VTI-3}_i = \sum_{j=1}^{N} d_{ij} v_i^{-1} \tag{3}$$

$$\text{VTI-4}_i = \sum_{j=1}^{N} d_{ij} v_j \tag{4}$$

$$\text{VTI-5}_i = \sum_{j=1}^{N} d_{ij} v_j^{-1} \tag{5}$$

$$\text{VTI-6}_i = \sum_{j=1}^{N} d_{ij} v_i v_j \tag{6}$$

$$\mathrm{VTI\text{-}7}_i = \sum_{j=1}^{N} d_{ij} v_i v_j^{-1} \tag{7}$$

$$\mathrm{VTI\text{-}8}_i = \sum_{j=1}^{N} d_{ij} v_i^{-1} v_j \tag{8}$$

$$\mathrm{VTI\text{-}9}_i = \sum_{j=1}^{N} d_{ij} v_i^{-1} v_j^{-1} \tag{9}$$

$$\mathrm{VTI\text{-}10}_i = \sum_{j=1}^{N} d_{ij}^{-1} \tag{10}$$

$$\mathrm{VTI\text{-}11}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_i \tag{11}$$

$$\mathrm{VTI\text{-}12}_i = \sum_{j=1}^{N} d_{ij}^{-1} d v_i^{-1} \tag{12}$$

$$\mathrm{VTI\text{-}13}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_j \tag{13}$$

$$\mathrm{VTI\text{-}14}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_j^{-1} \tag{14}$$

$$\mathrm{VTI\text{-}15}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_i v_j \tag{15}$$

$$\mathrm{VTI\text{-}16}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_i v_j^{-1} \tag{16}$$

$$\mathrm{VTI\text{-}17}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_i^{-1} v_j \tag{17}$$

$$\mathrm{VTI\text{-}18}_i = \sum_{j=1}^{N} d_{ij}^{-1} v_i^{-1} v_j^{-1} \tag{18}$$

We have to note that VTI1 has the same definition as the distance sum index.[13]

3. THE SELECTION OF THE BEST VERTEX TOPOLOGICAL INDEX

Computations of the VTI's were done for all 147 4-trees with $N = 4$ to 10 vertices. For the case of 4-trees with ten vertices, table 1 presents the minimum value, the maximum value, the average value, the standard deviation and the dispersion of the respective VTI. The dispersion of the values of a VTI is defined as :

$$D = \frac{\mathrm{VTImax} - \mathrm{VTImin}}{\mathrm{VTIav}} \tag{19}$$

Table 1

The minimum value, the maximum value, the average value, the standard deviation and the dispersion of the VTI's in the case of 4-trees with ten vertices

No.	VTI	Minimum value	Maximum value	Average value	Standard deviation	Dispersion
1.	VTI-1	14.	45.	25.90	35.24	1.197
2.	VTI-2	22.	96.	42.81	228.77	1.729
3.	VTI-3	3.5	45.	19.26	119.56	2.155
4.	VTI-4	19.0	81.0	42.81	140.95	1.448
5.	VTI-5	11.83	30.0	19.26	12.55	0.943
6.	VTI-6	35.0	156.0	69.42	550.64	1.743
7.	VTI-7	17.25	69.33	32.26	139.28	1.615
8.	VTI-8	4.75	81.0	32.26	374.53	2.364
9.	VTI-9	2.96	30.0	14.19	58.89	1.906
10.	VTI-10	2.83	6.5	4.26	0.729	0.863
11.	VTI-11	2.83	26.0	8.37	35.21	2.767
12.	VTI-12	1.36	4.17	2.74	0.667	1.025
13.	VTI-13	5.55	12.0	8.37	1.76	0.771
14.	VTI-14	1.47	5.08	2.74	0.629	1.321
15.	VTI-15	5.55	46.0	15.74	92.64	2.571
16.	VTI-16	1.47	20.33	5.61	20.59	3.365
17.	VTI-17	1.93	9.67	5.61	4.60	1.380
18.	VTI-18	1.02	2.71	1.70	0.180	0.995

It is not practical to choose a VTI with a small interval of values, or with a small standard deviation, or with a small dispersion, like VTI-10, VTI-12, VTI-13, VTI-14 or VTI-18.

Another comparison considers the distribution of numerical values of the VTI's, exemplified for the 75 decane isomers in table 2. Some of the definitions used give a very uneven distribution of the VTI's: VTI-2, VTI-3, VTI-6, VTI-7, VTI-11, VTI-15, VTI-16 in the lower range; VTI-9, VTI-12 a distribution with gaps T; and VT-14, VTI-5 with a central distribution. More evenly distributed are VTI-1, VTI-8, VTI-10 and VTI-18.

Table 3 presents the number of degenerations of the VTI's for all the 4-trees with N = 4 to 10 vertices. It may be seen that VTI-1, VTI-4, VTI-6, VTI-2, VTI-3 and VTI-8 have a high degeneracy and low selectivity. On the other hand, VTI-15, VTI-17, VTI-19, VTI-7 and VTI-13 have a fairly low degeneracy and high selectivity. VTI-15 and VTI-17 present degeneracy for the same two alkanes: 4-Et-octane and 4-Et-2,3-di-Me-hexane. VTI-9 presents degeneracy for the 4-trees representing 3-Me-nonane, 2,6-di-Me-octane and 3-Et-4-Me-heptane.

At this level of information it is hard to choose between VTI-15, VTI-17 and VTI-9. It is interesting to see if they carry the same type of information on molecular topology. Table 4 presents the intercorrelation matrix of the VTI's. The intercorrelation between VTI's varies from insignificant (for example, between VTI-1 and VTI-7, VTI-2 and VTI-4,

Table 2

Histogram presenting the distribution of the VTI's in the case of 4-trees with ten vertices

	INTERVAL									
	1	2	3	4	5	6	7	8	9	10
VTI-1	52	98	129	139	138	88	56	33	11	6
VTI-2	186	138	122	92	90	66	27	19	7	3
VTI-3	135	148	100	13	33	117	108	60	26	10
VTI-4	52	98	129	139	138	88	56	33	11	6
VTI-5	36	84	123	130	136	109	78	35	14	5
VTI-6	150	163	134	113	85	51	26	16	7	5
VTI-7	222	124	75	101	86	61	38	24	13	6
VTI-8	141	161	84	14	58	124	91	53	18	6
VTI-9	129	104	148	15	0	51	145	106	43	9
VTI-10	58	116	156	95	83	98	47	52	26	19
VTI-11	354	2	206	29	10	76	28	0	17	28
VTI-12	61	99	86	130	20	26	79	86	109	54
VTI-13	34	70	98	125	129	137	81	48	20	8
VTI-14	65	162	134	127	91	55	50	27	26	13
VTI-15	350	18	113	106	32	45	40	18	17	11
VTI-16	354	34	202	1	43	69	2	1	25	19
VTI-17	51	104	100	112	39	51	84	89	68	52
VTI-18	68	145	128	70	68	75	75	71	28	22

Table 3

Number of degenerations of the VTI's and of the informational TI's E, and $\bar{I}$, defined on the respective VTI's, in the case of 4-trees with $N = 4$ to 10 vertices. There are no degeneracies for $\bar{I}$

No	VTI	Number of degenerations of VTI	Number of degenerations of E	Number of degenerations of I
1	VTI-1	115	5	1
2	VTI-2	39	1	0
3	VTI-3	32	2	1
4	VTI-4	115	5	0
5	VTI-5	16	3	0
6	VTI-6	48	1	0
7	VTI-7	5	0	0
8	VTI-8	32	4	0
9	VTI-9	3	1	0
10	VTI-10	13	0	0
11	VTI-11	7	0	0
12	VTI-12	7	2	0
13	VTI-13	6	2	1
14	VTI-14	18	0	0
15	VTI-15	2	0	0
16	VTI-16	18	0	0
17	VTI-17	2	1	0
18	VTI-18	18	2	0

Table

The matrix displaying the intercorrelation coefficients

	VTI-1	VTI-2	VTI-3	VTI-4	VTI-5	VTI-6	VTI-7	VTI-8	VTI-9
VTI-1	1.000	0.112	0.834	0.998	0.977	0.245	−0.008	0.846	0.803
VTI-2		1.000	−0.418	0.073	0.117	0.987	0.982	−0.395	−0.455
VTI-3			1.000	0.852	0.814	−0.302	−0.507	0.999	0.992
VTI-4				1.000	0.968	0.210	−0.051	0.865	0.819
VTI-5					1.000	0.241	0.025	0.822	0.806
VTI-6						1.000	0.945	−0.278	−0.343
VTI-7							1.000	−0.486	−0.530
VTI-8								1.000	0.988
VTI-9									1.000
VTI-10									
VTI-11									
VTI-12									
VTI-13									
VTI-14									
VTI-15									
VTI-16									
VTI-17									
VTI-18									

VTI-13 and VTI-17, VTI-13 and VTI-18), through weak (between VTI-2 and VTI-15, VTI-13 and VTI-17), to high (between VTI-1 and VTI-4, VTI-11 and VTI-16, VTI-12 and VTI-17). The intercorrelation coefficients indicate that VTI-9 and VTI-7 express approximately the same type of information, different from the information contained in VTI-15. Thus, the use of VTI-9 and VTI-17 in developing a TI is almost equivalent and we will prefer VTI-17 for its lower degeneracy.

It is well known that informational TI's are suited for characterizing molecular branching.[12] In order to test the branching rules at constant number of vertices, three types of informational TI's, based on the eighteen types of VTI'-s defined, were computed: the informational energy content [14]

$$E = \sum p_i^2 \tag{20}$$

the total information content [15]

$$I = \sum VTI_i \log_2 \sum VTI_i - \sum VTI_i \log_2 VTI_i \tag{21}$$

the mean information content[16]

$$\bar{I} = - \sum_i p_i \log_2 p_i \tag{22}$$

where $p_i = \dfrac{VTI_i}{\sum VTI_i}$

Table 3 shows the number of degenerations of the informational TI's computed. The informational energy content has a rather high number of degenerations, in contrast with the total information content index which

4

between the eighteen types of VTI's defined

VTI-10	VTI-11	VTI-12	VTI-13	VTI-14	VTI-15	VTI-16	VTI-17	VTI-18
−0.372	−0.469	0.515	−0.216	−0.460	−0.471	−0.475	0.458	0.417
0.715	0.746	−0.651	0.377	0.714	0.726	0.724	−0.667	−0.614
−0.685	−0.756	0.827	−0.393	−0.748	−0.763	−0.740	0.783	0.724
−0.427	−0.507	0.523	−0.277	−0.505	−0.515	−0.510	0.465	0.420
−0.304	−0.437	0.548	−0.115	−0.388	−0.429	−0.440	0.495	0.489
0.606	0.630	−0.575	0.290	0.600	0.607	0.608	−0.599	−0.554
0.801	0.829	−0.687	0.462	0.814	0.815	0.813	−0.695	−0.621
−0.683	−0.743	0.802	−0.408	−0.742	−0.753	−0.727	0.756	0.696
−0.679	−0.769	0.866	−0.359	−0.739	−0.771	−0.750	0.826	0.785
1.000	0.916	−0.585	0.830	0.952	0.956	0.882	−0.554	−0.484
	1.000	−0.780	0.558	0.954	0.985	0.991	−0.762	−0.693
		1 000	−0.105	−0.683	−0.743	−0.772	0.996	0.963
			1.000	0.675	0.669	0.500	−0.057	−0.015
				1.000	0.957	0.955	−0.655	−0.538
					1.000	0.961	−0.718	−0.657
						1.000	−0.755	−0.669
							1.000	0.968
								1.000

has no degenerate value for all VTI's. We have to note that VTI-15 has no degenerated value, unlike VTI-17 which has a degeneracy in the case of the informational energy index. This fact is an argument in the favour of VTI-15. The analysis based on the order induced by the total information content index, shows that both VTI-15 and VTI-17 closely follow the rules for branching at constant number of vertices. Their value increases when branching increases.

Based on the VTI-1 (the distance sum index) it was defined a very discriminating TI, J.[7,18] By calculations it was determined that there are no J-equivalent nonisomorphic pairs of 4-trees with 11 or less vertices. In the case of 4-trees with 12 vertices, there are six pairs of J-equivalent nonisomorphic graphs.[19] Taking into account that the first degeneracy for VTI-1 appears for a 4-tree with six vertices, representing 2-Me-pentane, we can conclude that a TI defined using VTI-15, whose first degeneracy appears for 4-trees with ten vertices, will be a more discriminating index. Work is in progress in this direction.

REFERENCES

1 D. H. Rouvray, in "Chemical Applications of Graph Theory", Ed. A. T. Balaban, Academic, New York, 1976, p. 175.

2 J. J. Sylvester, *Am. J. Math.*, **1**, 64 (1874).

3 M. Barysz, D. Plavšić and N. Trinajstić, *Match*, **19**, 89 (1986).

4 A. T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, *Top. Curr. Chem.*, **114**, 21 (1983), Eds. M. Charton and I. Motoc, Springer, Berlin.

5 D. H. Rouvray, *Stud. Phys. Theor. Chem.*, **28**, 159 (1983), Ed. R. B. King, Elsevier, Amsterdam.

[6] J. V. Knop, W. R. Muller, K. Szymanski and N. Trinajstić, "Computer Generation of Certain Classes of Molecules", SKTH, Zagreb, 1985, p. 38.
[7] M. Randić, *J. Chem. Phys.*, **60**, 3920 (1974).
[8] M. Randić, *J. Chem. Phys.*, **62**, 309 (1975).
[9] M. Randić, *J. Chem. Inf. Comp. Sci.*, **15**, 105 (1975).
[10] M. Randić, *Chem. Phys. Lett.*, **42**, 27 (1976).
[11] H. Hosoya, M. Murakami and M. Gotoh, *Nat. Sci. Rept. Ochanomizu Univ.*, **24**, 27 (1973).
[12] D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, **67**, 4517 (1977).
[13] D. Bonchev, A. T. Balaban and O. Mekenyan, *J. Chem. Inf. Comp. Sci.*, **20**, 106 (1980).
[14] O. Onicescu, *C.R. Acad. Sci. Paris*, Serie A, **263**, 841 (1966).
[15] L. Brillouin, "Science and Information Theory", Academic, New York, 1956.
[16] C. Shannon and W. Weaver, "Mathematical Theory of Communication", University of Illinois, Urbana, 1949.
[17] A. T. Balaban, *Chem. Phys. Lett.*, **89**, 399 (1982).
[18] A. T. Balaban, *Pure Appl. Chem.*, **55**, 199 (1983).
[19] A. T. Balaban and L. V. Quintas, *Match*, **14**, 213 (1983).