

HIGHLY DISCRIMINATING DISTANCE-BASED TOPOLOGICAL INDEX

Alexandru T. BALABAN

The Polytechnic, Department of Organic Chemistry, 76206 Bucharest, Roumania

Received 13 April 1982

A new topological index J (based on distance sums s_i as graph invariants) is proposed. For unsaturated or aromatic compounds, fractional bond orders are used in calculating s_i . The degeneracy of J is lowest among all single topological indices described so far. The asymptotic behaviour of J is discussed, e.g. when $n \rightarrow \infty$ in C_nH_{2n+2} , $J \rightarrow \pi$ for linear alkanes, and $J \rightarrow \infty$ for highly branched ones.

1. Introduction

Topological indices (TIs) are a convenient means of translating chemical constitution into numerical values which can be used for correlations with physical properties or for biochemical quantitative structure-activity relationships. Revived interest in TIs is apparent from numerous recent reviews [1-5]. The reviews stress the need for new, better TIs, each with its optimal range of uses. A major drawback of most TIs is their degeneracy: two or more chemical constitutions (usually isomers) give rise to the same TI

value. In this communication, a new TI with very low degeneracy is introduced.

Table 1 presents a brief survey of the most representative TIs grouped into four categories whose significance will become apparent from the discussion below.

The constitution of the molecular skeleton (hydrogen-depleted graph) can be converted either into (A) the adjacency matrix (AM) whose entries are 1 for adjacent non-hydrogen atoms (points or vertices of the graph) and 0 otherwise, or (B) into the distance matrix (DM). The (topological) distance in a

Table 1
Main topological indices (TIs) proposed till now

Matrix	Type	
	book-keeping, I	structural, II
Adjacency (AM)	M_1 (Zagreb group TI) [6] F (First-neighbor sum) [7,8] Q (Quadratic TI) [9] ^{a)} C (Information content of F) [10]	Z (non-adjacent number sum) [11] χ (molecular connectivity) [12] χ_1 (first, largest, eigenvalue) [13]
Distance (DM)	w (path number) [14,15] I_B^w (information content of w) [16] ^{b)} B (centric TI) [9] I_B^D (information content of distances) [16]	^{c)}

^{a)} In addition, normalized and binormalized quadratic and/or centric TIs [9].

^{b)} In addition, mean information content \bar{I}_D for any I_D [16].

^{c)} Hosoya [17] proposed a distance polynomial (which is not a TI) for characterizing graphs.

graph is the number of bonds (lines or edges of the graph) in the shortest path between two vertices. Both the AM and the DM are square $n \times n$ matrices when there are n non-hydrogen atoms, i.e. n vertices in the graph. The entries in the DM are the distances between vertices, so that in a constitutional graph all entries on the main diagonal are 0 and the other ones are non-zero integers. Further graph-theoretical definitions may be found in monographs [18] or reviews [19].

The sums of entries in each row or column of the AM constitute graph invariants (independent of the initial arbitrary vertex numbering) called vertex degrees, i.e. the number of vertices adjacent to each vertex. Many topological indices are based on simple ("book-keeping") numerical transformations of vertex degrees, and fewer on conversions which take the constitution (connectivity) into account (either by using the Randić formula involving pairs of adjacent vertices, or by computing the coefficients of the characteristic polynomial or the eigenvalues corresponding to the AM). For molecules possessing the same inventory of graph invariants, topological indices belonging to the former type (I) are automatically degenerate, but the latter type (II) TIs may be different. The distribution of distances i in the graph is indicated by the numbers $2g_i$ of times each distance $i = 1, 2, \dots$ appears in the DM.

For example, in fig. 1, graphs 1, 2; 3, 4; 5, 6; 7, 8 have pairwise the same inventory (distribution) of vertex degrees; graphs 9 and 10 have the same distribution of distances, while graphs 3 and 4 have the same distribution of distance sums (defined below).

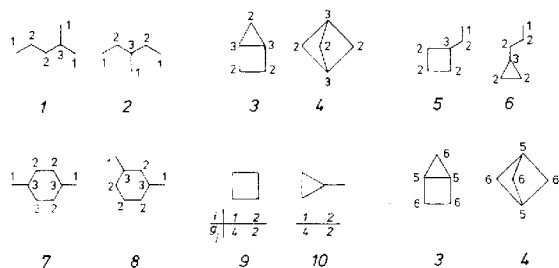


Fig. 1. Pairs of graphs having the same distribution of vertex degrees: 1 and 2; 3 and 4; 5 and 6; 7 and 8; the same distribution of distances: 9 and 10; the same distance sums: 3 and 4. Calculated J values are: 1, 2.6270; 2, 2.7540; 3, 2.1939; 4, 2.1909; 5, 2.0143; 6, 1.8763; 7, 2.1924; 8, 2.2307; 9, 2.0000; and 10, 2.1711.

Therefore TIs of type IA (table 1) will be degenerate for the four pairs of graphs 1–8. Similarly, TIs of type IB (excepting centric indices) will be degenerate for 9 and 10. However, the Randić index χ (a type IIA TI) is different for 3 and 4, but is accidentally degenerate for 5, 6 or for 7, 8.

2. The new topological index J

The distance matrix DM is richer in information than the adjacency matrix AM, yet all TIs so far described based on the DM are of the simple, book-keeping type I, therefore they have high degeneracy, i.e. little discriminating ability between constitutional isomers (stereoisomerism is till now beyond the scope of TIs). The new TI proposed here is based on the DM and is of type IIB corresponding to the empty case of table 1.

The distance sum (DS) for each vertex results by adding entries in the row or column of the DM corresponding to that vertex. Distance sums s_j ($j = 1, 2, \dots, n$) are thus akin to vertex degrees resulting by similar summation in the AM, so that distance sums s_j also constitute graph invariants; they are sometimes called distance degrees and they have the advantage of unlimited variation (unlike vertex degrees which in organic formulas can have only the four integer values 1–4). Averaged distance sums \bar{s}_j result when dividing distance sums by the number q of vertex adjacencies (i.e. by the number of edges in simple graphs of saturated molecules): $\bar{s}_j = s_j/q$.

We have found that a new TI which we call the *averaged distance sum connectivity* and denote by J , is the most discriminating single topological index so far proposed. Only the "superindex" (a linear combination of several TIs) [20] has comparative discriminating ability among chemical structures. The new index J is based on a Randić type formula, replacing vertex degrees by averaged distance sums:

$$J_{\text{acyclic}} = \sum_{\text{adjac. } i, j} (\bar{s}_i \bar{s}_j)^{-1/2} = q \sum_{\text{adjac. } i, j} (s_i s_j)^{-1/2} \quad (1)$$

For simple graphs of saturated acyclic molecules the application of formula (1) is straightforward, as seen in fig. 2. There are practically no degeneracies. Index J arranges the alkane isomers in an order which differs from that induced by all other TIs but bears

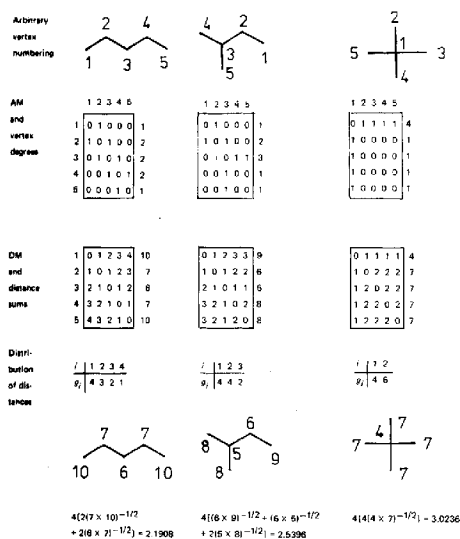


Fig. 2. Exemplification of adjacency matrix AM, distance matrix (DM), distance sums (DSs), and the new distance-based TIs for the three isomers of pentane.

a closer resemblance to that induced by I_D^W . As seen from fig. 2 and table 2, J increases with increasing branching.

Since for monocyclic and especially for polycyclic

Table 2
Values of the topological index J for alkanes C_nH_{2n+2}

C_n		J
C_2	ethane	1.0000
C_3	propane	1.6330
C_4	<i>n</i> -butane	1.9746
	isobutane	2.3238
C_5	<i>n</i> -pentane	2.1908
	2-methylbutane	2.5396
	2,2-dimethylpropane	3.0236
C_6	<i>n</i> -hexane	2.3840
	2-methylpentane	2.6270
	3-methylpentane	2.7540
	2,2-dimethylbutane	3.1685
	2,3-dimethylbutane	2.9935
C_7	<i>n</i> -heptane	2.4474
	2-methylhexane	2.6784
	3-methylhexane	2.8320
	2,4-dimethylpentane	2.9532
	3-ethylpentane	2.9922
	2,3-dimethylpentane	3.1440
	2,2-dimethylpentane	3.1542
	3,3-dimethylpentane	3.3606
	2,2,3-trimethylbutane	3.5412

simple graphs the distance sums are smaller than for acyclic graphs with the same n , J increases with the number of rings. Like most TIs $\sum (s_i s_j)^{-1/2}$ increases with branching, with the number of atoms and of rings [21]. Since, however, the two last numbers tend to obscure the variation due to branching and since they can be expressed by simple numerical parameters, we wish to construct a TI where their weight would be compensated. By taking averaged distance sums \bar{s} instead of s we compensate through q the increase with n . In order to have better comparability between acyclic and cyclic graphs with the same number n of vertices, the cyclomatic number μ is calculated using the formula

$$\mu = q - n + 1. \quad (2)$$

The expression

$$J = q(\mu + 1)^{-1} \sum_{\text{adjac. } i, j} (s_i s_j)^{-1/2} \quad (3)$$

is the final ring-corrected formula for J ; for acyclic graphs, (3) reduces to formula (1).

For unsaturated or aromatic molecules, instead of simple graphs having only single bonds, we have to use graphs with multiple bonds (multigraphs). We propose to introduce into the DM fractional numbers in this case: the entry in the DM for two adjacent vertices with bond order b is $1/b$, i.e. for triply bonded vertices, $1/3$; for doubly-bonded vertices, $1/2$; and for adjacent vertices in aromatic rings, $2/3$. With these entries the distance sums are calculated (paying special attention to the shortest path, i.e. smallest DS, in cyclic graphs), and then formula (3) is applied. For multigraphs, q is, as before, the number of pairs of adjacent vertices, but it no longer equals the number of edges as in simple graphs (i.e. in calculating q , a multiple bond is counted as one bond irrespective of its multiplicity). Fig. 3 and table 3 present examples of calculating J for mono- and poly-cyclic graphs and multigraphs.

Hundreds of J values were calculated for various constitutional graphs. Skeletal and positional isomerism of alkenes is reflected in J values: 1-pentene, 3.0249; 2-pentene, 3.2780; 3-methyl-1-butene, 3.5322; 2-methyl-1-butene, 3.5592; and 2-methyl-2-butene, 3.9290.

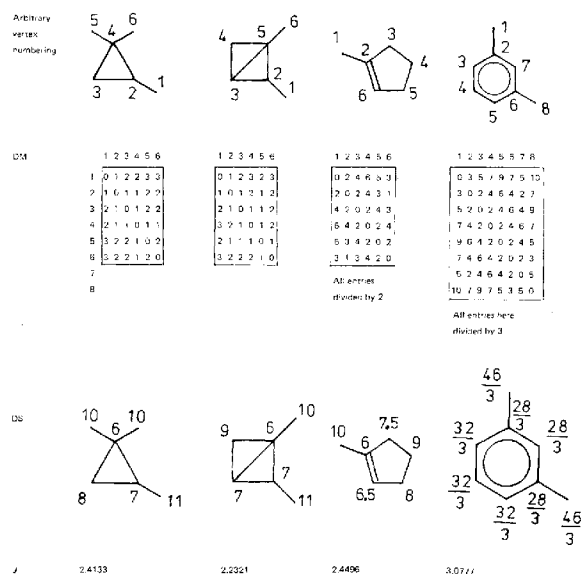


Fig. 3. Calculation of topological index J for mono- and polycyclic simple graphs and multigraphs (for comparison with the last graph, J values for ortho- and para-xylene are 3.1349 and 3.0325, respectively, while ethylbenzene has $J = 2.8321$).

3. Degeneracy of J

Since J was meant to reflect primarily the branching of molecules, the compensating term $q(\mu + 1)^{-1}$ was introduced into formula (3). This term induces degeneracies among non-isomeric systems: all $2m$ -membered cycloalkanes have $J = 2$; all localized $[4m]$ -

annulenes have $J = 8/3$; all delocalized $[4m]$ annulenes have $J = 3$, irrespective of the ring size. On the other hand, $(2m + 1)$ -membered cycloalkanes have $J = (2m + 1)^2 / (2m^2 + 2m)$, localized $[4m + 2]$ annulenes have $J = 4(2m + 1)^2 / (6m^2 + 6m + 1)$, and delocalized $[4m + 2]$ annulenes have $J = 3(2m + 1)^2 / (4m^2 + 4m + 1)$.

There also exist true degeneracies for J in the case of isomeric systems. The smallest pair of isomeric alkanes with the same J value appears for $C_{12}H_{26}$ (i.e. one pair out of the 355 constitutionally isomeric dodecanes!). For cyclic systems the degeneracy appears at lower numbers of vertices: eight for monocyclic and bicyclic graphs, namely the pairs 11 and 12 (monocyclic) and 13 and 14 (bicyclic). These constitutions can be generalized to 15 and 16 (monocyclic, where R and R' can be different or identical alkyls) and to 17 and 18 (bicyclic, $n \geq 1$) or to 19 and 20 (bicyclic, with $m, n \geq 1$) which have pairwise degenerate J values. For tri- and tetra-cyclic systems, the smallest degenerate systems have six vertices; these are 21 and 22 (tricyclic) and two tetracyclic pairs of planar graphs: 23 and 24; 25 and 26 (all formulas are presented in fig. 4).

Despite the above degeneracies of J , this new topological index is by far the least degenerate TI proposed till now. This is illustrated by table 4 which compares J with the next best TI, namely the Randić index χ : a comparison [20] between ten TIs for 427 acyclic, mono- and bi-cyclic graphs with 4–8 vertices had re-

Table 3
Values of J for alkylcyclohexanes and alkylbenzenes

C_n	Alkyl groups	Cyclohexane	Benzene
C_6	none	2.0000	3.0000
C_7	methyl	2.1229	3.0215
C_8	ethyl	2.1250	2.8321
	1,2-dimethyl	2.2794	3.1349
	1,3-dimethyl	2.2307	3.0777
	1,4-dimethyl	2.1924	3.0325
C_9	<i>n</i> -propyl	2.0779	2.6149
	isopropyl	2.2284	2.8483
	1-ethyl-2-methyl	2.2973	3.0065
	1-ethyl-3-methyl	2.2317	2.9369
	1-ethyl-4-methyl	2.1804	2.8816
	1,2,3-trimethyl	2.4017	3.4729
	1,2,4-trimethyl	2.4072	3.1717
	1,3,5-trimethyl	2.3409	3.1657

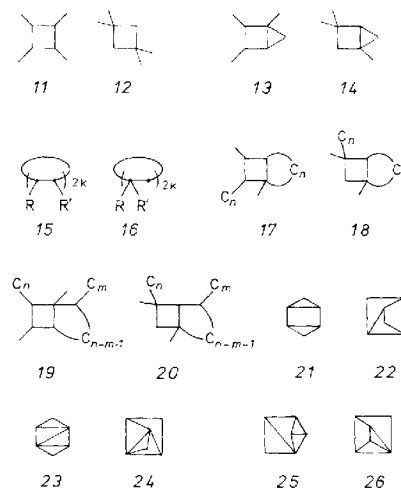


Fig. 4. Pairs of isomeric graphs possessing degenerate J values: 2.4220 for 11 and 12; 2.1610 for 13 and 14; 2.9430 for 23 and 24; 2.9527 for 25 and 26; and 2.1508 for 21 and 22.

Table 4
Mean degeneracy for mono- and bi-cyclic graphs with n vertices

n	Number N of graphs	Mean degeneracy	
		Randić index χ	Index J
5	10	1	1
6	29	1.2	1
7	85	1.6	1
8	250	1.6	1.008 ^{a)}

a) Graphs 11–14 are the two pairs out of 250 graphs leading to two degenerate J values.

vealed that χ had the lowest degeneracy, since for cyclic graphs its degeneracy started with 6 vertices, while all other TIs were degenerate for lower numbers of vertices. The criterion used in table 4 is the “mean degeneracy” [20], i.e. the ratio between the number N of graphs and the number of distinct values of the respective TI. A mean degeneracy equal to 1 indicates that no two graphs have degenerate values. Table 4 shows that J surpasses by far index χ . This is easy to understand: any edge in an organic hydrogen-depleted graph with vertices of degrees 1–4 can give rise in the Randić index χ to 10 possibilities for products of the vertex degrees of edge endpoints. By contrast, formulas (1) and (3) have an infinite number of possibilities for products $s_i s_j$ characterizing edges. Sometimes, degeneracies of J and χ coincide (e.g. 21–26), and sometimes they do not (e.g. 11–14, or regular graphs with non-degenerate J and degenerate χ values).

4. Asymptotic values of J

From the general formulas presented in section 3 for J in the case of cycloalkanes, localized and delocalized annulenes, it is easy to see the asymptotic convergence of J for very large m values in each of these three cases. It can be demonstrated that index J tends asymptotically towards π for linear alkanes $C_n H_{2n+2}$ as n increases; linear conjugated polyenes have J values tending asymptotically towards $4\pi/3$, while conjugated polyenes have J tending towards $3\pi/2$ when the chain increases indefinitely. On the other hand, for highly branched alkanes, J increases without limit.

For an alkane with one vertex of degree three and

with three infinitely long linear branches, the asymptotic J value is $4.1033 = (18/5^{1/2}) \arctan(5^{1/2}/4)$, while for the related case of one point of degree four with four infinitely long linear branches, the asymptotic J value is $5.1480 = 16 \arctan(1/3)$.

5. Conclusions

A comparison between J and the best or most used TIs so far proposed, namely χ , Z , I_D^W and I_D^E , reveals that (i) whereas all these TIs present many degeneracies for isomers, J is practically non-degenerate, especially for acyclic systems; (ii) unlike Z which normally requires in the general case a computer program, J can be calculated easily with a pocket calculator; (iii) unlike most TIs (except Bertz's indices [10,21]), J discriminates between graphs and multigraphs, extending the area of possible correlations; (iv) like any new TI, J induces a new definition of branching [22] and a new means of ordering isomeric molecules [23].

A computer program was devised for calculating rapidly J for large numbers of graphs. Input data are triplets of numbers indicating adjacent vertex numberings and bond order; as output one obtains the distance matrix and J (and optionally the adjacency matrix and χ) [24].

Further extensions and applications of J will involve (i) consideration of longer paths than edges (which are paths of length 1), analogously to $^h\chi$ [5]; (ii) the information content of J ; (iii) application to molecules containing heteroatoms, i.e. to vertex-labelled or chromatic graphs, e.g. by means of including unshared electron pairs [5] or bond orders which depend on the nature of the heteroatom [25]; and (iv) exploration of its correlational ability with properties [25]. Some properties such as the octane numbers of alkanes depend mainly on branching, and correlate only with some TIs [26]. It is to be hoped that by expressing through J the “topological shape” of molecules or of molecular fragments, a useful tool for quantitative structure–activity relationships and drug design has become available. The advantage of using J to account for the shape (branching) together with other parameters such as the number of vertices is that one can estimate the weight of each of these factors in the correlation.

Acknowledgement

Discussions with Drs. D. Bonchev, P. Filip, N. Ionescu-Pallas, I. Motoc, L.V. Quintas, S. Teleman and V.G. Velculescu are gratefully acknowledged.

References

- [1] A.T. Balaban, A. Chiriac, I. Motoc and Z. Simon, Steric fit in quantitative structure-activity relations, Lecture Notes in Chemistry No. 15 (Springer, Berlin, 1980) p. 22.
- [2] N. Trinajstić, Chemical graph theory (CRC Press, Cleveland), to be published;
A. Sablić and N. Trinajstić, Acta Pharm. Jugoslav. 31 (1981) 189.
- [3] D.H. Rouvray and A.T. Balaban, in: Applications of graph theory, eds. R.J. Wilson and L.W. Beineke (Academic Press, New York, 1979) p. 177.
- [4] D.H. Rouvray, Am. Sci. 61 (1973) 729; Math. Chem. 1 (1975) 125.
- [5] L.B. Kier and L.H. Hall, Molecular connectivity in chemistry and drug research (Academic Press, New York, 1976).
- [6] I. Gutman, B. Ruscic, N. Trinajstić and C.F. Wilcox Jr., J. Chem. Phys. 62 (1975) 3399.
- [7] J.R. Platt, J. Chem. Phys. 15 (1947) 419.
- [8] M. Gordon and G.R. Scantlebury, Trans. Faraday Soc. 60 (1964) 604.
- [9] A.T. Balaban, Theoret. Chim. Acta 53 (1979) 355.
- [10] S.H. Bertz, J. Am. Chem. Soc. 103 (1981) 3599.
- [11] H. Hosoya, Bull. Chem. Soc. Japan 44 (1971) 2332.
- [12] M. Randić, J. Am. Chem. Soc. 97 (1975) 6609.
- [13] L. Lovasz and J. Pelikan, Period. Math. Hung. 3 (1973) 175.
- [14] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17, 2636.
- [15] D.H. Rouvray and B.C. Crafford, S. African J. Sci. 72 (1976) 47.
- [16] D. Bonchev and N. Trinajstić, J. Chem. Phys. 67 (1977) 4517.
- [17] H. Hosoya, M. Murakami and M. Gotoh, Nat. Sci. Rep. Ochanomizu Univ. 24 (1973) 27.
- [18] F. Harary, in: Chemical applications of graph theory, ed. A.T. Balaban (Academic Press, New York, 1976) p. 5; Graph theory (Addison-Wesley, Reading, 1969).
- [19] J.W. Essam and M.E. Fisher, Rev. Mod. Phys. 42 (1970) 272.
- [20] D. Bonchev, O. Mekenyan and N. Trinajstić, J. Comput. Chem. 2 (1981) 127.
- [21] S.H. Bertz, Chem. Commun. (1981) 818.
- [22] D. Bonchev, J.V. Knop and N. Trinajstić, Math. Chem. 6 (1979) 21.
- [23] M. Randić, Math. Chem. 7 (1979) 5;
M. Randić, G.M. Bissey, R.B. Spencer and C.L. Wilkins, Comput. Chem. 3 (1979) 5;
R.E. Merrifield and H.E. Simmons, Theoret. Chim. Acta 55 (1980) 55;
D. Bonchev, A.T. Balaban and M. Randić, Intern. J. Quantum Chem. 19 (1981) 61;
D. Bonchev, A.T. Balaban and O. Mekenyan, J. Chem. Inf. Comput. Sci. 20 (1980) 106.
- [24] A.T. Balaban and P. Filip, to be published.
- [25] A.T. Balaban and I. Motoc, to be published.
- [26] A.T. Balaban and I. Motoc, Math. Chem. 5 (1979) 197.