Chemical graphs with degenerate topological indices based on information on distances

Ovidiu Ivanciuc [a], Teodor-Silviu Balaban [b] and Alexandru T. Balaban [a]

[a] *Department of Organic Chemistry, Polytechnic Institute, Faculty of Chemistry,*
Splaiul Independentei 313, 77206 Bucharest, Romania
[b] *Centre of Organic Chemistry, Splaiul Independentei 202, Bucharest, Romania*

Recently, four new types of vertex invariants, namely u, v, x, and y, were defined on the basis of information on graph distances. They were combined to give four highly selective topological indices: U, V, X, and Y. The degeneracy, i.e. equal values for nonisomorphic graphs, of the four topological indices is investigated. A structural condition and a graphical method which gives pairs of molecular graphs with identical U, V, X, and Y topological indices are introduced. The smallest pair of 4-trees representing alkanes having degenerated U, V, X, and Y values consists of trees with eighteen vertices.

1. Introduction

In recent years many different graph invariants have been proposed for characterization of structural features of chemical species. These are usually referred to in the chemical literature as topological indices (TIs) [1–6]. Such indices reflect in different ways the size and shape of the molecules they characterize and also provide some measure of the degree of molecular branching.

By removing all hydrogen atoms from the chemical formula of a chemical compound containing covalent bonds we obtain the hydrogen-depleted graph (or molecular graph) of that compound, whose vertices correspond to non-hydrogen atoms. In the particular case of hydrocarbons the vertices of the molecular graph denote carbon atoms.

A number of useful graph definitions will be introduced. Let $G = (V, E)$ be a graph G with N vertices, q edges, the cyclomatic number $\mu = q - N + 1$, the vertex set $V(G)$ and the edge set $E(G)$. The cyclomatic number μ represents the number of cycles in the graph. A tree is a graph without cycles. DEG_i denotes the degree of the vertex i in G, i.e. the number of edges incident with the vertex i. A k-tree is a tree with the maximum degree k. In chemical graph theory alkanes are represented as 4-trees. The distance between the vertices i and j is denoted by d_{ij} and is equal to the number of bonds on the shortest path between the vertices i and j. Distances d_{ij} are elements of the distance matrix of G, $\mathbf{D} = \mathbf{D}(G)$.

Let d_{max} be the maximum topological distance of the vertex i, i.e. the largest element of the ith row of the distance matrix of a molecular graph and let h_{ij} be the number of vertices in G at distance j from i. The sequence $(h_{i1}, h_{i2}, \ldots, h_{ij}, \ldots)$ is called the distance degree sequence of the vertex i in G, and is denoted DDS(i). Note that $h_{i1} = \mathrm{DEG}_i$. The number of elements in the sequence DDS(i) is equal to d_{max}.

The distance sum of the vertex i is the sum of the topological distances between vertex i and every vertex in the molecular graph [7],

$$s_i = \sum_{j=1}^{N} d_{ij} \, . \tag{1}$$

2. Topological indices

Although in their origins TIs were developed for the purpose of obtaining correlations with a great variety of physicochemical properties of chemical substances, namely QSPR (quantitative structure–property relationships), their range of applications has extended to bibliographical classification of chemical compounds and QSAR (quantitative structure–activity relationships). A brief presentation of the most important TIs will follow.

The Wiener index W was introduced in 1947 by Wiener [8], for predicting the alkane boiling points. In subsequent studies [9,10], Wiener extended the application of the W index to other physical properties of alkanes such as heats of formation, heats of vaporization, molar volumes and molar refractions. The Wiener index is equal to the half-sum of the elements of the distance matrix $\mathbf{D}$ [11], or to the half-sum of the distance sums:

$$W = \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} = \tfrac{1}{2} \sum_{i=1}^{N} s_i \, . \tag{2}$$

The relatively high degeneracy of W is due to the loss of information on graph distances when they are summed together. The degeneracy of the W index may be reduced by using information theory, as shown by Bonchev and Trinajstić [12], who devised ways for increasing the correlation ability while decreasing the degeneracy of topological indices.

The Randić connectivity index was defined as [13]

$$\chi = \sum_{E(G)} (\mathrm{DEG}_i \mathrm{DEG}_j)^{-1/2} \, , \tag{3}$$

where DEG_i and DEG_j denote the degrees of the two endpoints of an edge e_{ij} in the molecular graph. The summation is extended over all edges in the molecular

graph. The χ TI is the one most used as a molecular descriptor in structure–property and structure–activity relationship studies [14].

In order to increase the discriminating power of the index χ, while preserving its correlation abilities, the TI J (average distance sum connectivity) was defined by the formula [15–17]

$$J(G) = \frac{q}{\mu+1} \sum_{E(G)} (s_i s_j)^{-1/2},\tag{4}$$

where s_i and s_j denote the distance sums of the endpoints of an edge in the molecular graph and the summation goes over all edges in the molecular graph.

3. Local graph invariants: information on distances

A general problem of TIs is that they are more or less degenerate, i.e. two or more nonisomorphic structures may lead to the same value for a given TI. The discriminating ability or structural selectivity of a TI is inversely related to its degeneracy. A study on the structural selectivity of six TIs: Wiener's index W, the Zagreb group index M_1, Randić's connectivity index χ, the information on distance index I_D^E and Balaban's index J, reveals that J has the greatest selectivity among the studied indices [18]. Of course, high selectivity, although highly desirable, is not sufficient to obtain good results in QSPR and QSAR if the values of TIs do not properly reflect structural information in conjunction with the property under investigation.

Using a theorem concerning pairs of graphs with the same distance sum sequence, six pairs of molecular graphs representing alkanes with twelve carbon atoms were found to have identical index J [17]. That there are no other minimal order J-equivalent alkanes was confirmed by calculation of the J index of all alkanes with up to twelve carbon atoms. This finding shows that a dumping together of distances by such a crude method as summation loses information contained in the sequence of distances from a given vertex to all other graph vertices.

Recently new vertex invariants based upon the DDS were proposed [19]. In order to replace the crude operation of global distance summation by a more refined approach, the DDS vector is converted into new local vertex graph invariants by means of information theory. On applying Shannon's formula to the information content in DDS vectors we obtain for each vertex i the mean local information on the magnitude of distances,

$$u_i = -\sum_{j=1}^{d_{\max}} \frac{j\text{DDS}(i,j)}{s_i} \log_2 \frac{j}{s_i}\tag{5}$$

and the local information on the magnitude of distances,

$$v_i = s_i \log_2 s_i - u_i.\tag{6}$$

$\text{DDS}(i,j)$ represents the element j in the distance degree sequence of the vertex i.

For obtaining more convenient numbers, albeit not derived from rigorous information theory formula, two new related vertex invariants were proposed, called the extended local information on distance magnitude,

$$x_i = s_i \log_2 s_i - y_i \qquad (7)$$

and the mean extended local information on distance magnitude,

$$y_i = \sum_{j=1}^{d_{\max}} j\text{DDS}(i,j) \log_2 j. \qquad (8)$$

4. Topological indices U, V, X and Y

From local graph invariants one may form TIs either by means of simple operations confined to one vertex (e.g., summation of invariant, of its powers, of its inverse, etc.) or by more sophisticated operations involving more than one vertex at a time. Such operations are involved in computing the Randić connectivity index χ [13], the Balaban index J [15–17], the identification numbers ID [20], WID [21] and SID [5].

Summation is likely to lead to high operational degeneracy; therefore by analogy with the Randić connectivity index with the J index four new TIs were defined [19] on the basis of the four local graph invariants u_i, v_i, x_i and y_i,

$$U(G) = \frac{q}{\mu+1} \sum_{E(G)} (u_i u_j)^{-1/2}, \qquad (9)$$

$$V(G) = \frac{q}{\mu+1} \sum_{E(G)} (v_i v_j)^{-1/2}, \qquad (10)$$

$$X(G) = \frac{q}{\mu+1} \sum_{E(G)} (x_i x_j)^{-1/2}, \qquad (11)$$

$$Y(G) = \frac{q}{\mu+1} \sum_{E(G)} (y_i y_j)^{-1/2}. \qquad (12)$$

In all these formulas summations are over all edges in the molecular graph. Table 1 presents the indices U, V, X, and Y for the five hexane isomers.

5. Graphs with degenerate topological indices U, V, X and Y

An interesting fact was observed [19] for the molecular graph of 4-ethyloctane:

Table 1
Topological indices U, V, X and Y for alkane isomers with six carbon atoms.

Alkane	U	V	X	Y
C_6	11.59807	0.73698	1.08361	1.95797
2-Me-C_5	11.43376	0.88214	1.20076	2.63600
3-Me-C_5	11.29896	0.95477	1.24406	3.13516
2,2-Me$_2$-C_4	11.12333	1.20253	1.40951	5.18377
2,3-Me$_2$-C_4	11.19038	1.08782	1.33988	3.97067

two nonequivalent vertices present identical distance degree sequence. Using the mathematical formulas outlined in eqs. (5)–(8), these nonequivalent vertices with the same distance degree sequence will be assigned identical (degenerate) local invariants. Nethertheless, on combining local vertex invariants into TIs U, V, X and Y, these indices are found to have no degeneracy up to, and including $C_{15}H_{32}$.

We will present a theorem concerning the degeneracy of the TIs U, V, X and Y and we will give some examples of graphs with degenerate local graph invariants, u, v, x and y, as well as pairs of graphs with degenerate TIs U, V, X and Y.

By r_i we will denote a local topological invariant of the vertex i defined on the basis of the distance degree sequence of the vertex i,

$$r_i = f[\text{DDS}(i)]. \tag{13}$$

Also, R will denote a TI defined on the basis of the r_i vertex invariant using a similar formula to the TI J,

$$R(G) = \frac{q}{\mu + 1} \sum_{E(G)} (r_i r_j)^{-1/2}. \tag{14}$$

THEOREM 1

If G and H are connected graphs having the same distance degree sequence and $e \rightarrow F(e)$ is a one-to-one function of $E(G)$ onto $E(H)$ such that

$$[\text{DDS}(a), \text{DDS}(b)] = [\text{DDS}(c), \text{DDS}(d)]$$

for each edge $e = \{a, b\}$ and $F(e) = \{c, d\}$, then G and H are R-equivalent.

Proof

First, if $\text{DDS}(G) = \text{DDS}(H)$, then G and H have the same number of vertices n, the same number of edges q, and the same cyclomatic number μ. Then, the term $q/(\mu + 1)$ is the same for G and H.

Second, if for each edge $e = \{a, b\}$ from G and $F(e) = \{c, d\}$ from H, $[\text{DDS}(a), \text{DDS}(b)] = [\text{DDS}(c), \text{DDS}(d)]$, then

$$(r_a, r_b) = (r_c, r_d) \, .$$

From the definition of the TI R we have

$$\sum_{E(G)} (r_a r_b)^{-1/2} = \sum_{E(H)} (r_c r_d)^{-1/2} \, .$$

Therefore, G and H are R-equivalent.

Theorem 1 is equivalent to the one presented in ref. [17] for the index J. In order to apply theorem 1, we need a method for obtaining pairs of graphs exhibiting the same distance degree sequence. The method was also applied for obtaining J-equivalent graphs [17], and was introduced by Slater [22].

THEOREM 2

Let A be a graph such that two topologically nonequivalent vertices a_1 and a_2 in A have the same distance degree sequence. Let b_1 be a vertex in a graph B_1 and b_2 be a vertex in a graph B_2 such that b_1 and b_2 have the same distance degree sequence in B_1 and B_2 respectively.

If G is the graph constructed from A, B_1 and B_2 by identifying vertices a_1 with b_1 and identifying a_2 with b_2 and H is the graph constructed from A, B_1 and B_2 by identifying a_1 with b_2 and identifying a_2 with b_1, then G and H have the same distance degree sequence.
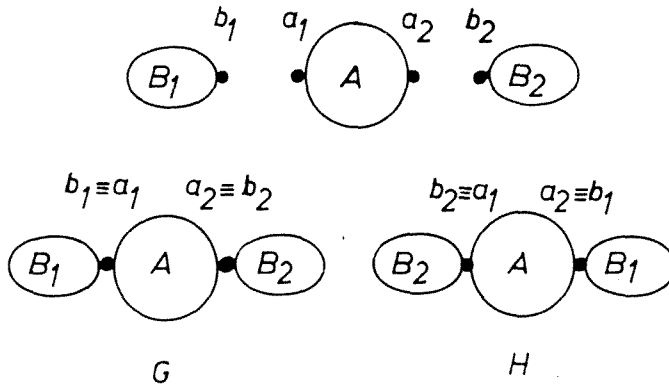
The whole set of 4-trees with up to fourteen vertices was searched in order to find graphs with topologically nonequivalent vertices exhibiting the same distance degree sequence. A brute force approach was used, which consisted in generating all 4-trees with up to fourteen vertices, and then calculating the paths for every vertex in a 4-tree. Finally, a search was done for the detection of topologically nonequivalent vertices exhibiting identical distance degree sequences. The result of this search is a set of sixty-two 4-trees with one or more pairs of vertices with the same
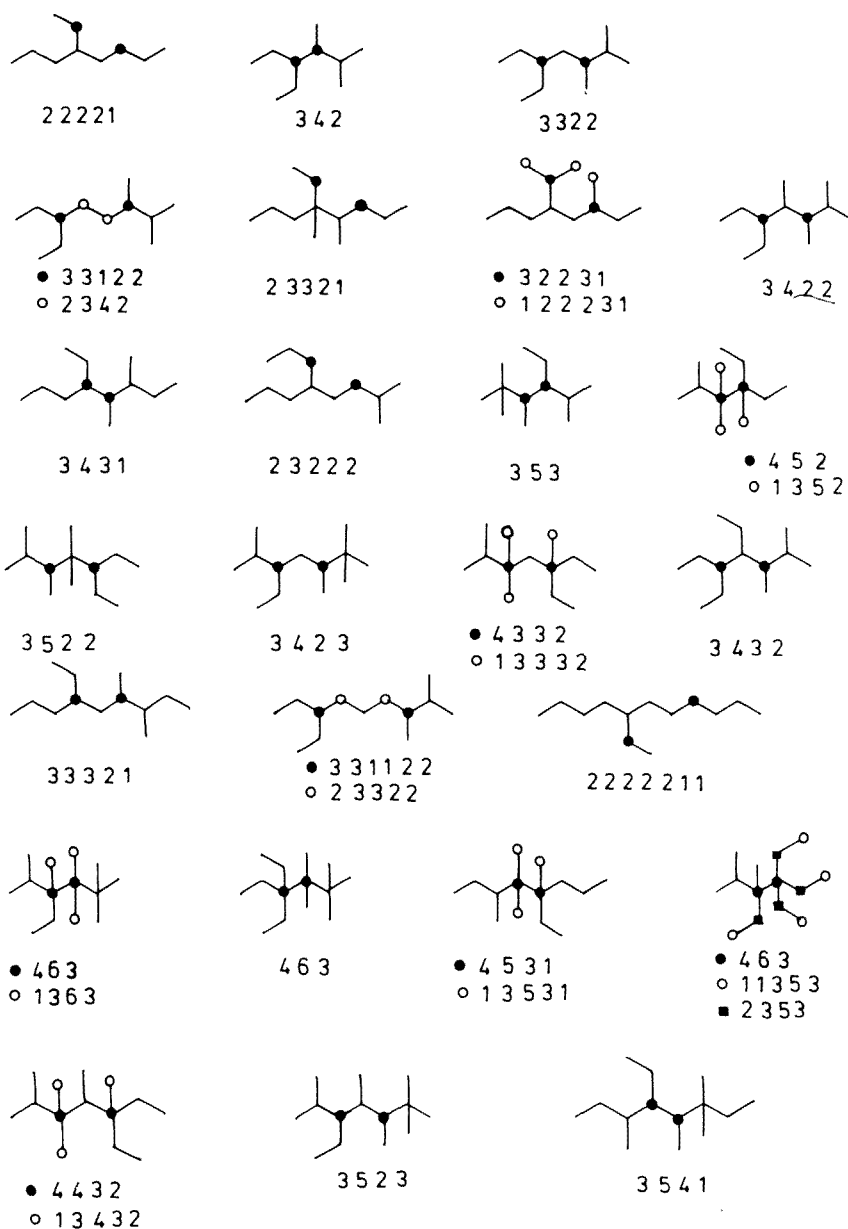
Fig. 1. 4-trees up to fourteen vertices with nonequivalent vertices presenting the same distance degree sequence. Indicated nonequivalent vertices have the same distance degree sequence, depicted under the corresponding graph. For brevity, the zero distance count is omitted.
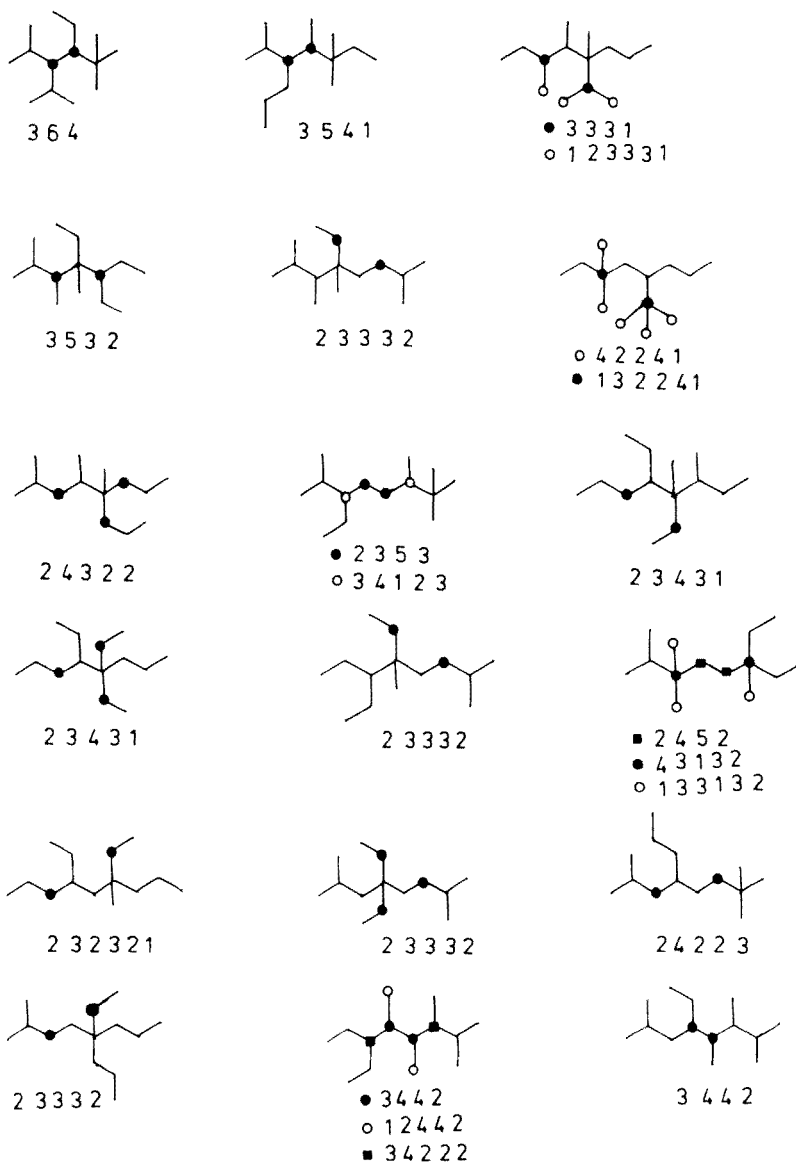
3 6 4

3 5 4 1

● 3 3 3 1
○ 1 2 3 3 3 1

3 5 3 2

2 3 3 3 2

○ 4 2 2 4 1
● 1 3 2 2 4 1

2 4 3 2 2

● 2 3 5 3
○ 3 4 1 2 3

2 3 4 3 1

2 3 4 3 1

2 3 3 3 2

■ 2 4 5 2
● 4 3 1 3 2
○ 1 3 3 1 3 2

2 3 2 3 2 1

2 3 3 3 2

2 4 2 2 3

2 3 3 3 2

● 3 4 4 2
○ 1 2 4 4 2
■ 3 4 2 2 2

3 4 4 2

Fig. 1. Continued.

3 4 4 2

3 4 3 2 1

3 4 4 2

3 4 3 3

3 5 4 1

3 4 4 2

2 2 3 3 2 1

3 4 3 2 1

3 4 4 2

● 3 3 2 3 2
○ 1 2 3 2 3 2

● 2 4 3 2 2
○ 3 3 2 1 2 2

○ 3 3 2 3 2
■ 2 2 2 2 3 2
○ 1 1 2 2 2 3 2

3 4 4 2

2 3 4 3 1
3 3 2 2 2 1

● 2 2 3 4 2
○ 2 3 3 1 2 2
■ 3 3 1 1 1 2 2
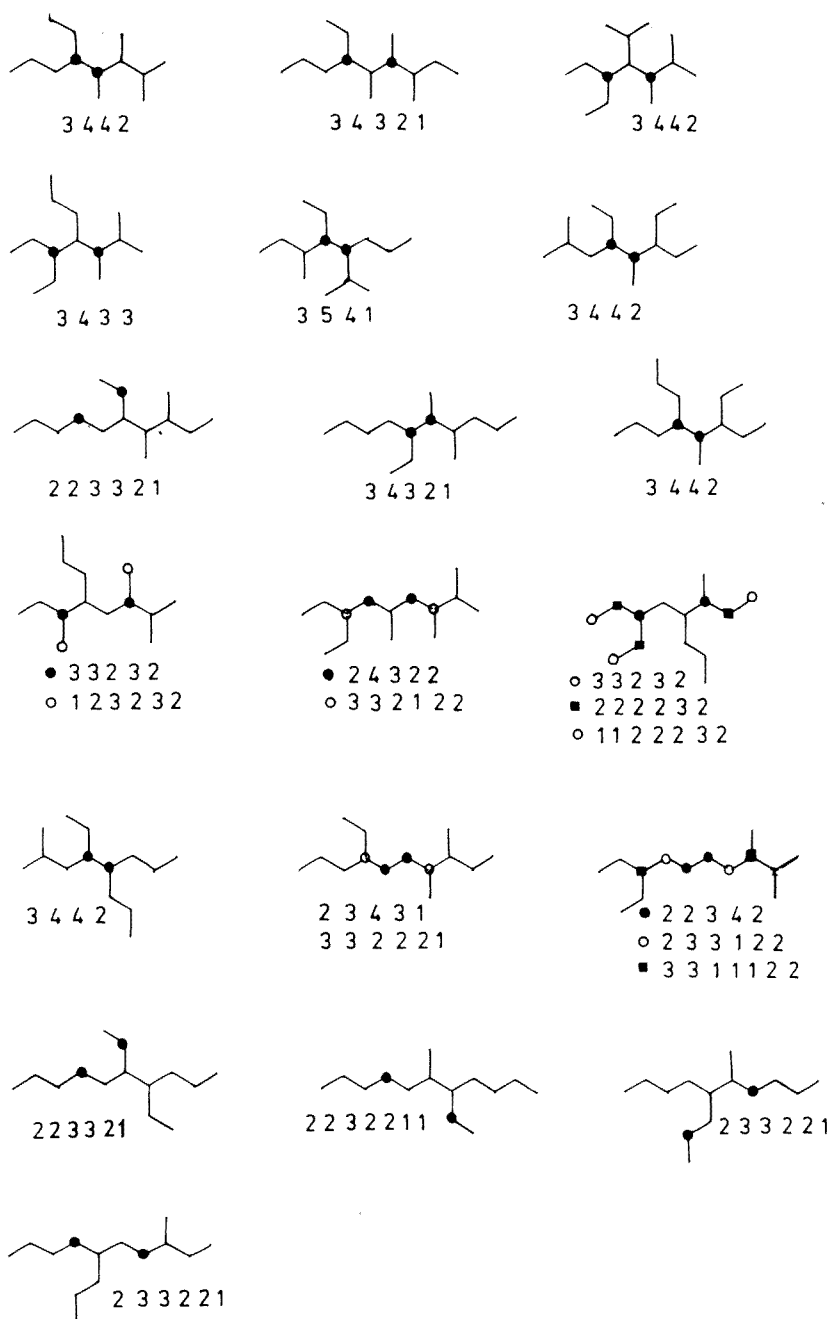
2 2 3 3 2 1

2 2 3 2 2 1 1

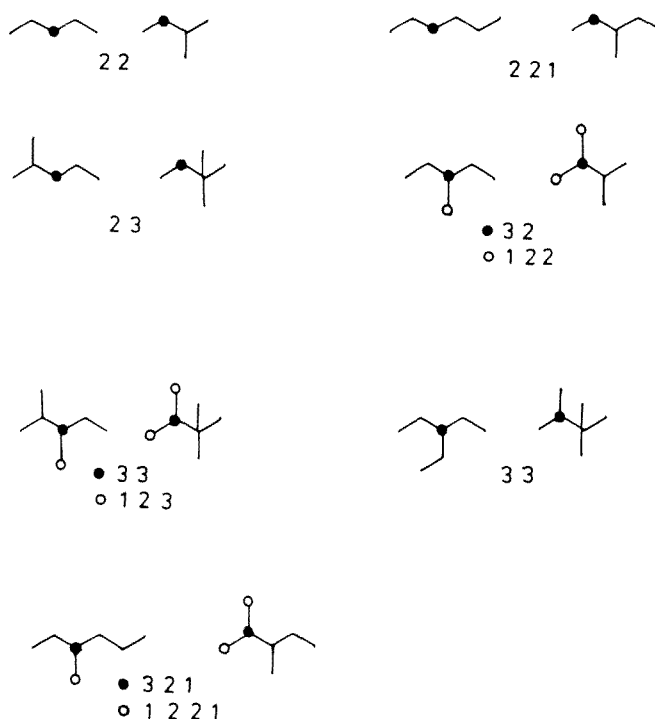2 3 3 2 2 1

2 3 3 2 2 1

Fig. 1. Continued.

Fig. 2. Pairs of 4-trees with vertices having identical distance degree sequence. Indicated vertices for each pair have the same distance degree sequence, depicted under the corresponding pair of 4-trees. For brevity, the zero distance count is omitted.

distance degree sequence: two with ten vertices, one with eleven vertices, eight with twelve vertices, seven with thirteen vertices, and fortyfour with fourteen vertices. All these graphs are depicted in fig. 1. Nonequivalent vertices indicated with the same symbol have identical distance degree sequence. Below each graph, the distance degree sequence of the indicated vertices is presented.

In order to apply theorems 1 and 2, one needs also to find pairs of vertices in different graphs, exhibiting identical distance degree sequence. Figure 2 presents such pairs of 4-trees. Indicated vertices in a pair of 4-trees have the same distance degree sequence. Below each graph, the distance degree sequence of the indicated vertices is presented.

Applying theorems 1 and 2 to the set of graphs in figs. 1 and 2, an enormous number of *R*-equivalent pairs of trees in obtained. We restrict our presentation to the six pairs of graphs from fig. 3.
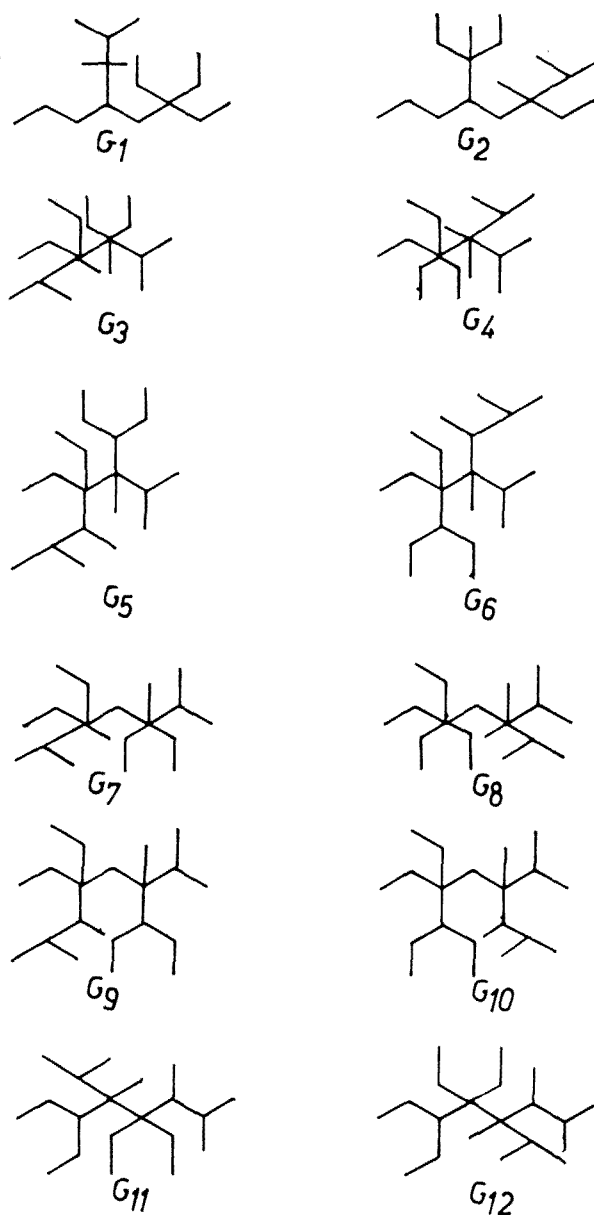
Fig. 3. Pairs of trees with degenerate U, V, X and Y topological indices.

As is apparent from table 2, all pairs of trees from fig. 3 exhibit degenerate values for the topological indices U, V, X and Y.

The G_1 and G_2 pair of 4-trees with eighteen vertices represents the smallest pair

Table 2
Topological indices U, V, X and Y for the six R-index degenerate pairs of trees from fig. 3.

4-trees		U	V	X	Y
G_1	G_2	73.41206	0.87330	1.27427	2.70689
G_3	G_4	72.38707	1.22067	1.63273	4.64499
G_5	G_6	87.41797	1.04210	1.46177	3.54260
G_7	G_8	80.25789	1.05101	1.47389	3.54979
G_9	G_{10}	95.79533	0.92956	1.34845	2.93417
G_{11}	G_{12}	87.41797	1.04210	1.46177	3.54260

of alkanes with degenerate U, V, X and Y indices. The second pair, 5-trees G_3 and G_4, was originally constructed by Slater [22].

6. Conclusions

Knowing the structural causes of the degeneracy of topological indices, one can design new indices with better structural selectivity. Due to the global summation of distances, the Wiener index W is highly degenerate. Using a vertex partitioning of distances, the index J presents the smallest degenerate pairs of 4-trees for graphs with twelve vertices. The summation of distances to give distance sums is the cause of the degeneracy of the index J. Using information theory applied to distance degree sequences, the highly discriminating topological indices U, V, X and Y were obtained. The structural conditions for their degeneracy were demonstrated and a pair of 4-trees with eighteen vertices, exhibiting degenerate indices was obtained. Knowing the structural causes of the degeneracy of the indices U, V, X and Y, is it possible to design new, distance-based indices, with a higher structural selectivity? This is an open question, deserving a close consideration and further study.

References

[1] A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Topics Curr. Chem. 114 (1983) 21.
[2] A.T. Balaban, J. Chem. Inf. Comput. Sci. 25 (1985) 334; J. Mol. Struct. THEOCHEM 120 (1985) 117; 165 (1988) 243.
[3] P.J. Hansen and P.C. Jurs, J. Chem. Educ. 65 (1988) 574.
[4] M. Randić, J. Chem. Inf. Comput. Sci. 24 (1984) 164.
[5] A.T. Balaban, in: *Graph Theory and Topology in Chemistry*, Stud. Phys. Theor. Chem., Vol. 51, eds. R.B. King and D.H. Rouvray (Elsevier, Amsterdam, 1987) p. 159.
[6] M. Randić, B. Jerman-Blažić and N. Trinajstić, Comput. Chem. 14 (1990) 237.
[7] D. Bonchev, A.T. Balaban and O. Mekenyan, J. Chem. Inf. Comput. Sci. 20 (1980) 106.
[8] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17, 2636.
[9] H. Wiener, J. Chem. Phys. 15 (1947) 766.
[10] H. Wiener, J. Phys. Chem. 52 (1948) 425, 1082.

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

[12] D. Bonchev and N. Trinajstić, J. Chem. Phys. 67 (1977) 4517.

[13] M. Randić, J. Am. Chem. Soc. 97 (1975) 6609.

[14] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976); *Molecular Connectivity in Structure–Activity Analysis* (Wiley, New York, 1986).

[15] A.T. Balaban, Chem. Phys. Lett. 80 (1982) 399.

[16] A.T. Balaban, Pure Appl. Chem. 55 (1983) 189.

[17] A.T. Balaban and L.V. Quintas, MATCH 14 (1983) 213.

[18] M. Razinger, J.R. Chrétien and J.E. Dubois, J. Chem. Inf. Comput. Sci. 25 (1985) 23.

[19] A.T. Balaban and T.-S. Balaban, J. Math. Chem. 8 (1991) 383.

[20] M. Randić, J. Chem. Inf. Comput. Sci. 24 (1984) 164.

[21] K. Szymanski, W.R. Müller, J.V. Knop and N. Trinajstić, Int. J. Quant. Chem. Quant. Chem. Symp. 20 (1986) 173.

[22] P.J. Slater, Congressus Numerantium 33 (1981) 309.