# NEW CENTRIC TOPOLOGICAL INDEXES FOR
# ACYCLIC MOLECULES (TREES) AND
# SUBSTITUENTS (ROOTED TREES),
# AND CODING OF ROOTED TREES

Alexandru T. Balaban[a], Sharon Bertelsen[b], and Subhash C. Basak[b]

[a]Polytechnic University
Department of Organic Chemistry
Splaiul Independenter 313
77206 Bucharest, Roumania

[b]Natural Resources Research Institute
University of Minnesota, Duluth
5013 Miller Trunk Highway
Duluth, MN 55811, USA

Abstract. On the basis of distance sums and their contributions to successive layers around the center of acyclic graphs (trees), around the root vertex of rooted or planted trees, a sequence of integers is obtained. By applying information-theoretic formulas, this sequence can be converted into new topological indexes, IBC and AIBC, which are much more discriminating (i.e., less degenerate) than previous centric indexes. Correlations between IBC values and normal boiling points of 73 alkanes ($C_3$-$C_9$) are satisfactory. A new coding system for rooted or planted trees has been devised.

## Introduction

In a previous paper,[1] a few "first-generation"[2] topological indexes (TIs) were proposed based on the fact that in acyclic graphs ("trees") there is a unique center consisting of one vertex, or two adjacent vertexes. Since this center is found by deleting sequentially vertexes of degree 1 (endpoints), the sequence of deleted vertexes (the "pruning" partition of the number $n$ of vertexes) afforded a simple means for obtaining the centric index B, as the sum of squares of numbers in that partition.

Despite the fact that B presents a satisfactory correlation with octane numbers of alkanes, it shares with other first generation TIs a high degree of degeneracy. The same is true for the normalized and binormalized centric indexes C and C' discussed earlier.

An attempt was made[3] for generalizing the notion of graph centers to cyclic graphs. On this basis, centric indexes for any graphs have been proposed. This provides a basis for computing centric indexes for any graph, cyclic or acyclic.

## A new centric information topological index for trees

We shall start by restricting the attention to acyclic graphs in which no vertex degree is $\geq 4$, and we shall call such graphs 4-trees.[4]

From the distance matrix, it is easy to compute for each vertex (by summing entries over rows or columns) a local vertex invariant (LOVI) which is called its <u>status</u>[5] or <u>distasum</u>.[6] The halfsum of all these LOVIs for any 4-tree affords its Wiener index (W).[7] Like other first generation TIs, the Wiener index has a fairly high degeneracy: for alkanes, two pairs of heptanes and octanes share the same W values.

In two recent papers, <u>regressive vertex degrees</u>[8] and <u>regressive distance sums</u>[9] were introduced by considering for each vertex the progressively smaller contribution from

more distant vertices, as one considers the successive "shells" of vertices around each vertex.

In this paper we apply this approach to shells around the center of the 4-tree, and we sum the distasums for vertices in each successive shell, obtaining a sequence of numbers $Q_i$. One obtains thus a vector of integer numbers which constitutes a partition of:

$$2W = \sum_i Q_i$$

This vector can be converted into a new TI for information based on center (IBC) with much lower degeneracy than B, C or C', by applying information theoretic formulas[10-15] using binary logarithms according to Shannon:[16]

$$IBC = 2W \text{ lb } 2W - \sum_i Q_i \text{ lb } Q_i$$

We devote binary logarithms by lb, and the conversion from decimal logarithms is:
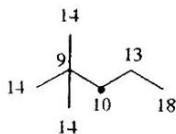
$$\text{lb } x = \frac{\log x}{\log 2} = 3.322 \log x$$

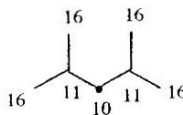The average information content based on center (AIBC) will be:

$$AIBC = IBC / W$$

We have chosen to divide by W rather than 2W, so that AIBC values are higher than 1.

Two pairs of examples shall illustrate the procedure. Distasums for two isomers of heptane are indicated on structures.

14
9  13
14  10  18
14

$Q_i = \{10,22,60\}; \; 2W = 92$

**1**, 22MMC$_5$

16   16
16  11  11  16
10

$Q_i = \{10,22,64\}; \; 2W = 96$

**2**, 24MMC$_5$

The first layer is the center itself; the second layer has contributions 9 + 13 =

11 + 11 = 22; the third and last layer sums the remaining distasums.

IBC(1) = 92 lb 92-10 lb 10-22 lb 22-60 lb 60 = 114.43; IABC = 114.43 / 92 = 2.49

IBC(2) = 96 lb 96-10 lb 10-22 lb 22-64 lb 64 = 116.82; IABC = 116.82 / 96 = 2.43

It can be seen that the ordering of **1** vs. **2** becomes reversed for the two TIs.

17
15  13
21  11  23
17
17
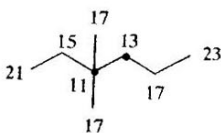
$Q_i = \{24,66,44\}; \; 2W = 134$

**3**, 33MMC$_6$

18
16  12
22  12  22
16
18
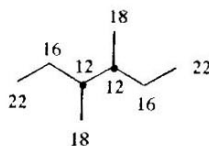
$Q_i = \{24,68,44\}; \; 2W = 136$

**4**, 34MMC$_6$

In the pair of octane isomers, we have a bicenter which constitutes the first layer: 11 + 13

= 12 + 12 = 24; for the second layer one sums distasums of vertices adjacent to the

bicenter vertices; for the third and last layer, the distasums of the remaining vertices are

considered.

Table 1 presents the new TIs denoted by IBC and AIBC for all constitutional

isomers of alkanes with 3-8 atoms, as well as their boiling points at normal pressure.

Table 2 presents the same data for all isomers of nonane. Boiling point data are from Needham, Wei and Seybold's paper.[17]

The column $N_i$ indicates the pruning partition starting from the center outwards. Taking the order of its entries into account, it offers slightly better discrimination than indexes B or C, but it still leads to considerable degeneracy.

It can be observed that, unlike W and the previous centric indices B, C and C', there are no degeneracies of IBC and AIBC for the alkanes in Tables 1 or 2.

When the first numbers of the $Q_i$ sequence are equal for two or more alkanes (such as for entries 10 and 11, 15 and 16, 17 and 19, 25-27, 28 and 31, 29 and 30, 33 and 35, as well as for the triplet 32, 34, and 36 in Table 1), this indicates that the numbers of vertices in the first shell around the center are equal, and that pruning partitions $N_i$ are identical.

It can be seen that the ordering in Table 1, according to decreasing values of IBC, is not identical to that induced by Wiener's index W; this is evidenced by entries 16-19 in the heptane isomer series, and by many other entries in the octane or nonane isomer series. The intermolecular ordering of isomeric alkanes has been discussed earlier by several authors.[18-23]

It can be noted from Table 1 that there exist differences in the intermolecular ordering between IBC and AIBC, namely in increased apparent degree of branching (i.e. a later place among isomers) for 2-methylalkanes, for 2,3-dimethylbutane, 2,4-dimethylpentane and 2,5-dimethylhexane. Some properties such as boiling points at normal pressure are markedly lowered by such branching at next-to-marginal positions.

A few comments are necessary concerning Table 2 which includes data for nonane isomers. Again, no degeneracy was observed, but when both W and the $N_i$ sequence are

1. New centric indexes for isomers    (TI's for alkanes $C_3$-$C_8$).

| No. | Alkane | $Q_i$ | $N_i$ | 2W | IBC | AIBC | bp (°C) Exptl. | bp (°C) Calc. |
|---|---|---|---|---|---|---|---|---|
| | Propane: | | | | | | | |
| 1 | $C_3$ | 2,6 | 1,2 | 8 | 6.49 | 1.62 | -42.07 | -24.33 |
| | Butanes: | | | | | | | |
| 2 | $C_4$ | 8,12 | 2,2 | 20 | 19.42 | 1.94 | -0.50 | 1.10 |
| 3 | $2MC_3$ | 3,15 | 1,4 | 18 | 11.70 | 1.30 | -11.73 | 0.18 |
| | Pentanes: | | | | | | | |
| 4 | $C_5$ | 6,14,20 | 1,2,2 | 40 | 57.61 | 2.88 | 36.07 | 29.22 |
| 5 | $2MC_4$ | 11,25 | 2,3 | 36 | 31.96 | 1.78 | 27.85 | 26.42 |
| 6 | $22MMC_3$ | 4,28 | 1,4 | 32 | 17.39 | 1.09 | 9.50 | 24.74 |
| | Hexanes: | | | | | | | |
| 7 | $C_6$ | 18,22,30 | 2,2,2 | 70 | 108.67 | 3.11 | 68.74 | 58.02 |
| 8 | $2MC_5$ | 8,18,38 | 1,2,3 | 64 | 85.95 | 2.67 | 60.27 | 55.94 |
| 9 | $3MC_5$ | 7,29,26 | 1,3,2 | 62 | 58.49 | 1.89 | 63.28 | 53.19 |
| 10 | $23MMC_4$ | 14,44 | 2,4 | 58 | 46.24 | 1.60 | 57.99 | 51.89 |
| 11 | $22MMC_4$ | 14,42 | 2,4 | 56 | 45.44 | 1.62 | 49.74 | 51.80 |
| | Heptanes: | | | | | | | |
| 12 | $C_7$ | 12,26,32,42 | 1,2,2,2 | 112 | 210.71 | 3.76 | 98.43 | 89.10 |
| 13 | $2MC_6$ | 23,27,54 | 2,2,3 | 104 | 153.66 | 2.95 | 90.05 | 85.50 |
| 14 | $3MC_6$ | 21,42,37 | 2,3,2 | 100 | 152.93 | 3.06 | 91.85 | 85.45 |
| 15 | $3EC_5$ | 9,36,51 | 1,3,3 | 96 | 128.19 | 2.67 | 93.48 | 83.54 |
| 16 | $23MMC_5$ | 9,36,47 | 1,3,3 | 92 | 124.45 | 2.71 | 89.78 | 83.23 |
| 17 | $24MMC_5$ | 10,22,64 | 1,2,4 | 96 | 116.82 | 2.43 | 80.50 | 82.60 |

Table 1 (continued)

| No. | Alkane | $Q_i$ | $N_i$ | 2W | IBC | AIBC | bp (°C) Exptl. | bp (°C) Calc. |
|-----|--------|-------|-------|-----|-------|------|-------|-------|
| 18 | 33MMC$_5$ | 8,48,32 | 1,4,2 | 88 | 116.35 | 2.64 | 86.06 | 82.56 |
| 19 | 22MMC$_5$ | 10,22,60 | 1,2,4 | 92 | 114.43 | 2.49 | 79.20 | 82.39 |
| 20 | 223MMMC$_4$ | 17,67 | 2,5 | 84 | 61.04 | 1.45 | 80.88 | 77.33 |
|    | Octanes: | | | | | | | |
| 21 | C$_8$ | 32,36,42,54 | 2,2,2,2 | 164 | 323.28 | 3.94 | 125.67 | 116.87 |
| 22 | 2MC$_7$ | 15,32,38,73 | 1,2,2,3 | 158 | 284.12 | 3.60 | 117.65 | 116.00 |
| 23 | 4MC$_7$ | 13,49,38,50 | 1,3,2,2 | 150 | 279.49 | 3.73 | 117.71 | 115.86 |
| 24 | 3MC$_7$ | 14,30,58,50 | 1,2,3,2 | 152 | 279.23 | 3.67 | 118.93 | 115.85 |
| 25 | 3EC$_6$ | 26,50,68 | 2,3,3 | 144 | 214.16 | 2.97 | 118.53 | 113.16 |
| 26 | 24MMC$_6$ | 26,51,65 | 2,3,3 | 142 | 212.31 | 2.99 | 109.43 | 113.06 |
| 27 | 23MMC$_6$ | 26,50,64 | 2,3,3 | 140 | 209.70 | 3.00 | 115.61 | 112.92 |
| 28 | 25MMC$_6$ | 28,32,88 | 2,2,4 | 148 | 203.93 | 2.76 | 109.10 | 112.60 |
| 29 | 34MMC$_6$ | 24,68,44 | 2,4,2 | 136 | 199.69 | 2.94 | 117.73 | 112.36 |
| 30 | 33MMC$_6$ | 24,66,44 | 2,4,2 | 134 | 197.67 | 2.89 | 118.26 | 112.25 |
| 31 | 22MMC$_6$ | 28,29,79 | 2,2,4 | 136 | 190.42 | 2.80 | 106.84 | 111.82 |
| 32 | 3E2MC5$_5$ | 11,43,80 | 1,3,4 | 134 | 169.69 | 2.53 | 115.65 | 110.50 |
| 33 | 3E3MC$_5$ | 10,58,60 | 1,4,3 | 128 | 168.58 | 2.63 | 118.26 | 110.42 |
| 34 | 234MMMC$_5$ | 11,43,76 | 1,3,4 | 130 | 166.66 | 2.56 | 113.47 | 110.29 |
| 35 | 233MMMC$_5$ | 10,58,56 | 1,4,3 | 124 | 164.09 | 2.65 | 114.76 | 110.11 |
| 36 | 223MMMC$_5$ | 11,43,72 | 1,3,4 | 126 | 163.49 | 2.59 | 109.84 | 110.07 |
| 37 | 224MMMC$_5$ | 12,26,94 | 1,2,5 | 132 | 148.47 | 2.25 | 99.24 | 108.99 |
| 38 | 2333MMMMC$_4$ | 20,96 | 2,6 | 116 | 76.91 | 1.33 | 106.47 | 102.81 |

le 2. New centric indexes for isomers of nonanes.

| No. | Nonanes | $Q_i$ | $N_i$ | 2W | IBC | AIBC | Boiling point (°C) | |
|-----|---------|-------|-------|-----|-----|------|-------|-------|
| | | | | | | | Exptl. | Calc. |
| 1 | $C_9$ | 20,42,48,58,72 | 1,2,2,2,2 | 240 | 532.67 | 4.44 | 150.80 | 136.69 |
| 2 | $2MC_8$ | 39,43,51,95 | 2,2,2,3 | 228 | 433.02 | 3.80 | 143.26 | 140.46 |
| 3 | $4MC_8$ | 35,65,51,65 | 2,3,2,2 | 216 | 423.34 | 3.92 | 142.48 | 140.65 |
| 4 | $3MC_8$ | 27,41,77,65 | 2,2,3,2 | 210 | 391.46 | 3.73 | 144.18 | 141.04 |
| 5 | $4EC_7$ | 15,56,73,60 | 1,3,3,2 | 204 | 375.10 | 3.68 | 141.20 | 141.11 |
| 6 | $25MMC_7$ | 17,36,69,86 | 1,2,3,3 | 208 | 371.95 | 3.58 | 136.00 | 141.12 |
| 7 | $3EC_7$ | 17,36,67,88 | 1,2,3,3 | 208 | 371.24 | 3.57 | 143.00 | 141.12 |
| 8 | $27MMC_7$ | 16,59,44,85 | 1,3,2,3 | 204 | 369.09 | 3.62 | 133.50 | 141.12 |
| 9 | $23MMC_7$ | 17,36,67,84 | 1,2,3,3 | 204 | 366.20 | 3.59 | 140.50 | 141.12 |
| 10 | $26MMC_7$ | 18,38,44,116 | 1,2,2,4 | 216 | 364.84 | 3.98 | 135.21 | 141.11 |
| 11 | $34MMC_7$ | 15,56,67,58 | 1,3,3,2 | 196 | 362.48 | 3.70 | 140.60 | 141.11 |
| 12 | $35MMC_7$ | 16,34,94,60 | 1,2,4,2 | 204 | 357.66 | 3.51 | 136.00 | 141.09 |
| 13 | $22MMC_7$ | 18,38,44,108 | 1,2,2,4 | 208 | 357.48 | 3.44 | 132.69 | 141.09 |
| 14 | $44MMC_7$ | 14,76,44,58 | 1,4,2,2 | 192 | 348.20 | 3.63 | 135.20 | 141.04 |
| 15 | $33MMC_7$ | 16,34,88,58 | 1,2,4,2 | 196 | 347.32 | 3.54 | 137.30 | 141.03 |
| 16 | $224MMC_6$ | 31,60,96 | 2,3,4 | 188 | 280.12 | 2.98 | 126.54 | 139.76 |
| 17 | $4E2MC_6$ | 31,59,106 | 2,3,4 | 196 | 278.67 | 2.84 | 133.80 | 139.71 |
| 18 | $3E4MC_6$ | 29,80,79 | 2,4,3 | 188 | 275.63 | 2.93 | 140.40 | 139.62 |

Table 2. (continued)

| No. | Nonanes | Q | $N_t$ | 2W | IBC | AIBC | Boiling point (°C) | |
| | | | | | | | Exptl. | Calc. |
|---|---|---|---|---|---|---|---|---|
| 19 | 235MMMC$_6$ | 31,59,102 | 2,3,4 | 192 | 275.08 | 2.87 | 131.34 | 139.60 |
| 20 | 3E2MC$_6$ | 31,58,103 | 2,3,4 | 192 | 274.27 | 2.86 | 138.00 | 139.57 |
| 21 | 3E3MC$_6$ | 29,77,78 | 2,4,3 | 184 | 270.66 | 2.942 | 140.60 | 139.45 |
| 22 | 233MMMC$_6$ | 29,79,76 | 2,4,3 | 184 | 270.62 | 2.941 | 137.68 | 139.45 |
| 23 | 234MMMC$_6$ | 29,80,75 | 2,4,3 | 184 | 270.54 | 2.940 | 139.00 | 139.45 |
| 24 | 223MMMC$_6$ | 31,58,95 | 2,3,4 | 184 | 266.86 | 2.90 | 133.60 | 139.32 |
| 25 | 233MMMC$_6$ | 29,77,74 | 2,4,3 | 180 | 265.62 | 2.95 | 137.68 | 139.28 |
| 26 | 225MMMC$_6$ | 33,37,126 | 2,2,5 | 196 | 254.14 | 2.593 | 124.08 | 138.86 |
| 27 | 334MMMC$_6$ | 27,98,51 | 2,5,2 | 176 | 246.95 | 2.81 | 140.46 | 138.57 |
| 28 | 33EEC$_5$ | 12,68,96 | 1,4,4 | 176 | 223.74 | 2.54 | 146.17 | 137.52 |
| 29 | 3E23MMC$_5$ | 12,68,92 | 1,4,4 | 172 | 220.19 | 2.56 | 142.00 | 137.34 |
| 30 | 2334MMMMC$_5$ | 12,68,88 | 1,4,4 | 168 | 216.52 | 2.577 | 141.55 | 137.15 |
| 31 | 3E24MMC$_5$ | 13,50,117 | 1,3,5 | 180 | 214.41 | 2.38 | 136.73 | 137.04 |
| 32 | 3E22MMC$_5$ | 13,50,113 | 1,3,5 | 176 | 211.89 | 2.41 | 133.83 | 136.91 |
| 33 | 2234MMMMC$_5$ | 13,50,109 | 1,3,5 | 172 | 209.29 | 2.43 | 133.02 | 136.77 |
| 34 | 2233MMMMC$_5$ | 12,58,84 | 1,4,4 | 154 | 199.35 | 2.589 | 140.27 | 136.22 |
| 35 | 2244MMMMC$_5$ | 14,30,132 | 1,2,6 | 176 | 182.49 | 2.07 | 122.28 | 135.20 |

identical (namely entries 21-23), the IBC and AIBC values appear very close to each other, therefore AIBC values were computed with four significant digits; also in the case of entries 26, 30, and 34 it was necessary to have the same number of digits for discriminating the AIBC values.

It may be observed from Table 2 that the intermolecular ordering of nonanes induced by IBC values, which was adopted for Tables 1 and 2, is similar for both these tables: it places 4-methylheptane and 4-methyloctane before the 3-methyl isomers, and leaves as the last entry in Table 2 the alkane with two t-butyl groups at each end (which happens to have also the lowest normal boiling point).

### Information TIs for cyclic graphs

By generalizing the notion of graph center, as indicated earlier,[3] one can apply information theoretic formulas to the sequence $Q_i$ formed by summing distasums for shells around the generalized center.  Details for applying this idea will be presented in a future paper.

### Correlation of IBC values with normal boiling points of all 73 alkanes $C_3$-$C_9$

For validating the new centric TI denoted by IBC, we report the results of a correlation with boiling points at normal pressure for all 73 alkanes with 3-9 carbon atoms. The following equation was obtained, using also the number N of carbon atoms:

$$bp = -97.75 \ (\pm 6.60) + 0.1242 \ (\pm 0.0045) \ IBC -$$

$$- 0.000168 \ (\pm 0.000072) \ IBC^2 + 23.875 (\pm 1.419) \ N,$$

with the following statistical parameters:  $r^2 = 0.9732$, root mean sqare error 6.78, and F = 835.23.  The two most serious outliers are propane and neopentane.

From the data in Table 2, one can see that all entries from 4 to 15 have calculated boiling points around 141°, although their IBC values range from 347 to 391. This is due to the parabolic nature of the above equation which flattens for IBC values around 400.

### Information topological indices for rooted trees

Rooted trees have one vertex distinguished from the other ones; when this vertex is an endpoint, we have a planted tree.

In chemistry, rooted or planted trees can symbolize substituents; the root indicates the point of attachment to the core of the molecule. In geographical applications of graph theory, rooted trees may symbolize river basins with the root indicating the end of the watercourse, e.g. an estuary. If one assigns directions on edges of planted/rooted trees from the endpoints towards the root, then rooted trees can model acyclic directed graphs (digraphs).

If, instead of calculating distances from each vertex to all other vertices, we compute distances from the root vertex, we can proceed in a similar fashion as for 4-trees, and apply similar information-theoretic formulas.

Two examples shall illustrate the procedure.



5                                         6
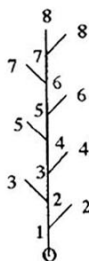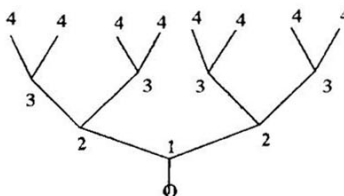
$Q_i(5) = \{4,6,8,10,12,14,16\}$; $2W = 70$     $Q_i(6) = \{4,12,32\}$; $2W = 48$

We illustrate two planted trees **5** and **6** with 17 vertices each, but we ignore the first shells and we start with the shell grouping vertices at distance two from the root.

Then we obtain:

For (**5**):     IBC = 70lb70-4lb4-6lb6-8lb8-10lb10-12lb12-14lb14-16lb16 = 187.95

AIBC = 2x187.95/70 = 5.36

For (**6**):     IBC = 48lb48-4lb4-12lb12-32lb32 = 57.05

AIBC = 2x57.05/48 = 2.38

## New coding systems for rooted trees

By analogy with ideas developed previously for coding the structure of alkanes,[19] one can extend this coding to rooted trees. The basic idea consists in obtaining a canonical labeling of vertices according to a set of rules, and then to specify sequentially for each vertex the label of the forerunner vertex, i.e. the unique adjacent vertex with lower numbering.

The rules for vertex numbering are as follows:[19]

1. Number 1 corresponds to the vertex of highest degree ($V_1$) which is closest to the root vertex.

2. Numbers 2, 3, ..., $V_1$ + 1 are assigned to vertexes adjacent to vertex 1 in the order of (i) decreasing vertex degrees, and (ii) increasing remoteness from the root vertex, in this priority order.

3. Numbers $V_1$ + 2, $V_1$ + 3, ..., $V_1$ + $V_2$ (where $V_2$ is the degree of vertex 2) are assigned to vertexes adjacent to vertex 2, with the same priority order as under item 2.

4. The same procedure is repeated for vertices adjacent to vertex 3, then for those adjacent to vertex 4, until all vertexes have been labeled with numbers 1 to $n$, where $n$ + 1 is the total number of vertexes (the root vertex remains without a number).

The graph code (GC) results from listing the label of the forerunner vertex of each vertex starting with vertex 4. (Since both vertex 2 and vertex 3 are always attached to vertex labeled 1, this information would be redundant.)

This code may be converted into a difference code (DC or differcode) by conserving the first digit of the GC, and by subtracting each label from the following one.

The two rooted trees which were used for illustrating the preceding section will again serve as examples.

Canonical numbering:
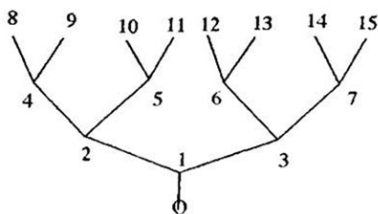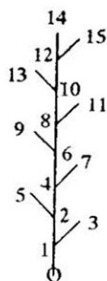


| Vertex: | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GC: | 2 | 2 | 4 | 4 | 6 | 6 | 8 | 8 | 10 | 10 | 12 | 12 |
| DC: | 2 | 0 | 2 | 0 | 2 | 0 | 2 | 0 | 2 | 0 | 2 | 0 |

| Vertex: | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GC: | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 |
| DC: | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

The resulting differcode may be then read as a number, and from it one can easily retrieve the GC and the structure of the rooted tree. For purposes of reducing the number of digits in the differcode, which for rooted 4-trees usually does not contain digits larger than 4, one can read this number in base 5 and convert it into a decimal number and one can take its decimal logarithm. If, as seen in examples (A) and (B), we deal with rooted or planted 3-trees, then normally the highest digit in the differcode is 2, so that one can read this number in base 3 and proceed likewise.

Three examples follow:
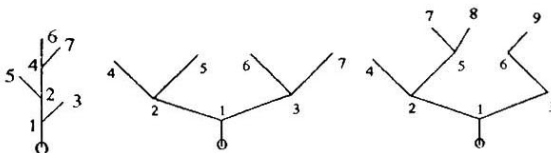


|  Vertex: | 4 5 6 7 | 4 5 6 7 | 4 5 6 7 8 9 |
|---|---|---|---|
| GC: | 2 2 4 4 | 2 2 3 3 | 2 2 3 4 4 6 |
| DC: | 2 0 2 0 | 2 0 1 0 | 2 0 1 1 0 2 |
| | | | |
| Base 3 | $2 \times 3^3 + 2 \times 3 =$ | $2 \times 3^3 + 1 \times 3 =$ | $2 \times 3^5 + 1 \times 3^3 + 1 \times 3^2 + 2 =$ |
| Base 10 | $= 54 + 6 = 60$ | $= 54 + 3 = 57$ | $= 486 + 27 + 9 + 2 = 524$ |

Of course, if one has large rooted (or planted) trees, and if one wishes to retrieve their structure, it is necessary to conserve all the necessary information for recovering the integers DC and GC from the different conversions into various bases and logarithms.

Herndon and Leonard[24] had previously published a code for chemical graphs which could also serve as a TI. In our case, the logarithm of DC can serve for a similar purpose, but it shares with the analog from the 4-tree case several drawbacks which were discussed earlier.[19] We advocate, therefore, for rooted or planted trees the use of IBC and AIBC values as TIs for QSAR/QSPR, and of GC or DC as codes for structure input and retrieval.

## Conclusions

By summing distasums of vertices in shells at progressively increasing distances around the unique graph center of trees or around the root vertex of planted/rooted trees, one obtains a partition $Q_i$ of the integer number 2W, where W is the Wiener index of the graph. On applying information-theoretic formulas to this sequence $Q_i$, it is possible to obtain new centric topological indexes IBC and AIBC with low degeneracy, which order isomeric alkanes in a slightly different manner from previously proposed TIs and from each other. The index IBC presents satisfactory correlation with normal boiling points of alkanes.

A new coding system for rooted/planted trees has been devised by analogy with a previous coding system for alkanes.

Extension of these ideas to cyclic graphs, ti graphs with heteroatoms and/or multiple bonds and applications of the new TIs and codes for other types of QSAR/QSPR will be presented in future papers.

## Acknowledgements

## References

1. A.T. Balaban, Theor. Chim. Acta, 1979, 53, 355.

2. A.T. Balaban, J. Chem. Inf. Comput. Sci., 1992, 32, 23.

3. D. Bonchev, A.T. Balaban and M. Randić, Int. J. Quantum. Chem., 1981, 19, 61; 1982, 22, 441; D. Bonchev, A.T. Balaban and O. Mekenyan, J. Chem. Inf. Comput. Sci., 1980, 20, 106; D. Bonchev, O. Mekenyan and A. T. Balaban, ibid., 1989, 29, 91; A.T. Balaban, D. Bonchev and W.A. Seitz, J. Mol. Struct. (Theochem), 1993, 280, 253.

4. A.T. Balaban, J.W. Kennedy and L.V. Quintas, J. Chem. Educ., 1988, 65, 304.

5. F. Buckley and F. Harary, "Distance in Graphs," Addison-Wesley Pub. Co., 1989.

6. A.T. Balaban, Pure Appl. Chem., 1983, 55, 199.

7. H. Wiener, J. Amer. Chem. Soc., 1947, 69, 17.

8. M.V. Diudea, O. Minailiuc and A.T. Balaban, J. Comput. Chem., 1991, 12, 527.

9. A.T. Balaban and M.V. Diudea, J. Chem. Inf. Comput. Sci., 1993, 33, 421.

10. D. Bonchev and N. Trinajstić, J. Chem. Phys., 1977, 67, 4517.

11. D. Bonchev, "Information-Theoretic Indices for Characterization of Chemical Structures," Research Studies Press, Chichester, 1983.

12. S.C. Basak, G.J. Niemi and G.D. Veith, J. Math. Chem., 1991, 7, 243; Idem, in "Computational Graph Theory" (ed. D.H. Rouvray), Nova Science Publishers, New York, 1990, p. 235.

13. S.C. Basak, Med. Sci. Res., 1987, 15, 605.

14. C. Raychaudhury, S.K. Ray, J.J. Ghosh, A.B. Roy, and S.C. Basak, J. Comput. Chem., 1984, 5, 581.

15. A.T. Balaban and T.S. Balaban, J. Math. Chem., 1991, 8, 383.

16. C.E. Shannon, Bell Syst. Tech. J., 1948, 27, 379.

17. D.E. Needham, I.C. Wei and P.G. Seybold, J. Amer. Chem. Soc., 1988, 110, 4186.

18. D.H. Rouvray, J. Molec. Struct. (Theochem), 1989, 185, 187.

19. A.T. Balaban, J. Molec. Struct. (Theochem), 1988, 165, 243.

20. S.H. Bertz, Discr. Appl. Math., 1988, 19, 65; reprinted in "Applications of Graphs in Chemistry and Physics" (eds. J.W. Kennedy and L.W. Quintas), North Holland, Amsterdam, 1988.

21. A.T. Balaban, J. Chem. Inf. Comput. Sci., (in press).

22. M. Randic, J. Chem. Educ., 1992, 69, 712 and furhter reference therein.

23. S.C. Grossman, Int. J. Quantum Chem., 1985, 28, 1.

24. W.C. Herndon and J.E. Leonard, Inorg. Chem., 1983, 22, 554.