

COMPUTER PROGRAM FOR TOPOLOGICAL INDEX J  
( AVERAGE DISTANCE SUM CONNECTIVITY )

Alexandru T. Balaban<sup>a</sup> and Petru Filip<sup>b</sup>

<sup>a</sup> The Polytechnic Institute - TCH, Department of Organic  
Chemistry, Spl. Independentei 313, Bucharest, Roumania

<sup>b</sup> Center of Organic Chemistry - ICECHIM, Bucharest, Roumania

(Received: February 1984)

Abstract

A computer program was developed for calculating the highly discriminating distance-based topological index J on the basis of vertex adjacency data. This topological index can take into account the presence of unsaturation or aromaticity in hydrocarbons or hydrocarbon radicals (fragments). A new definition is proposed for the topological index of a univalent hydrocarbon radical, corresponding to a rooted graph, and an alternative approach is considered for olefinic or acetylenic unsaturation. Problems raised by heteroatoms are also discussed.

### Introduction

Topological indices (TI's) convert chemical structures into numerical values which can be used for quantitative correlations with physical, chemical or biochemical properties. The reverse process is, however, impossible, i.e. topological indices are not in one-to-one correspondence with chemical structures : two or more structures may correspond to the same topological index which in this case is degenerate.

The forty odd topological indices so far proposed have been reviewed.<sup>1-7</sup> So far, all topological indices proposed till now take into account only constitutional isomerism, ignoring stereoisomerism. Among those which have proved to be most successful in quantitative structure-activity relationships, especially for drug design, Randić's index  $\chi$  is the first to deserve mention.<sup>6,8</sup> Wiener's index  $w$  is historically the first TI;<sup>9</sup> it allowed a good correlation with boiling points of alkanes in isomeric series. Centric indices<sup>10,11</sup> as well as the mean square topological distance<sup>12</sup> afford good correlations with octane numbers in isomeric series of heptanes and octanes. Information TI's developed by Bonchev and Trinajstić,<sup>13-15</sup> and more recently by Bertz<sup>16</sup> provide increased discrimination, i.e. lower degeneracy.

### Topological index J

A new, highly discriminating topological index  $J$ , average distance sum connectivity, was recently proposed.<sup>12,17</sup>

It is based on topological distances in graphs. Chemical structures possessing covalent bonds correspond to graphs : atoms are the vertices (points) of the graph, and covalent bonds are the edges (lines) of the graph. As customary in organic chemistry, hydrogen atoms are ignored so that hydrocarbons correspond to hydrogen-depleted graphs whose vertices represent carbon atoms and whose edges represent carbon-carbon bonds : single bonds for saturated hydrocarbons represented by simple graphs, or multiple bonds for unsaturated hydrocarbons represented by multigraphs. The topological distance between two vertices is the number of edges along the shortest path between these vertices. In acyclic system (trees) the path is unique, but in cyclic systems two vertices may be connected by more than one path. Two vertices at distance one are said to be adjacent.

The adjacency matrix  $A$  of a graph with  $n$  vertices is a square symmetrical  $n \times n$  matrix whose entries are 1 for adjacent vertices and 0 otherwise. The distance matrix  $D$  is also a square symmetrical  $n \times n$  matrix, but the entries are the topological distances. In both cases the main diagonal has only zeroes and the entries 1 are the same, but all other entries are different for  $A$  and  $D$  : in  $A$  they are zeroes, while in  $D$  they are numbers larger than one. The sum of entries for column or row  $i$  in  $A$  is called the degree of vertex  $i$  ( $v_i$ ). The sum of entries for column or row  $i$  in  $D$  is occasionally called distance degree, but we prefer to call it the distance sum of vertex  $i$  ( $s_i$ ). Both  $v_i$  and  $s_i$  are graph invariants which do not depend on how the graph is drawn or on how the vertices are numbered.

The cyclomatic number of a connected graph with  $n$  vertices and  $q$  edges (in this case multiple edges are counted as a simple edge) is  $\mu$  :

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

This cyclomatic number is the number of rings in the classical definition of polycyclic compounds according to the Baeyer, IUPAC and Chemical Abstracts nomenclature systems.

By definition, the new topological index is

$$J = \frac{q}{\mu + 1} \sum_{\text{edge}} (s_i s_j)^{-1/2} \quad (2)$$

for all edges having  $i$  and  $j$  as endpoints (the sum is over all edges).

There is a limited analogy between formula (2) and the definition of the Randić index  $\chi$  by formula (3) :

$$\chi = \sum_{\text{edge}} (v_i v_j)^{-1/2} \quad (3)$$

However, instead of using products of vertex degrees  $v_i v_j$  (degrees can have only the values 1,2,3 or 4 for hydrocarbons, totalling ten possible combinations) index  $J$  uses products of distance sums  $s_i s_j$  leading to a double infinity of combinations : in addition, the normalizing factor before the sum of formula (2) ensures that  $J$  does not increase automatically with  $n$  like most TI's, but reflects only the "topological shape", i.e. the degree of branching. Indeed, for some infinite graphs consisting of linearly repeating units,  $J$  has finite

values ( $\pi$  for infinite polyethene,  $1.5\pi$  for polypropene, etc. while for highly branched infinite graphs  $J$  can be  $\omega$ <sup>12,17</sup>)

Computer program HYDROCARBON 1 - J

For saturated systems formula (2) affords in a straightforward manner the value of  $J$ . In order to facilitate the calculation from adjacency data, we have developed a computer program written in BASIC which was implemented on a Hewlett-Packard 9830A desk-top computer.

For unsaturated and aromatic systems, two approaches are possible. The first<sup>12,17</sup> is to take into consideration conventional bond orders  $b$  ( $b = 1, 2$  and  $3$  for single, double and triple bonds, respectively, and  $b = 1.5$  for aromatic bonds); then the entry in the distance matrix for two adjacent multiply-bonded vertices with bond order  $b$  between them is  $1/b$ .

Distance sums  $s_1$  are then calculated by taking into account such non-integer topological distances. In the application of formula (2) multiple bonds are treated in this case as a single bond when calculating  $q$ .

As examples, all pentene  $C_5H_{12}$  (1 - 2) and pentene  $C_5H_{10}$  (4 - 12) isomeric structures will be discussed in Figures 1 and 2 ; as mentioned in the introduction stereoisomerism will be ignored. In all these cases, the cyclomatic number is  $\mu = 0$ .

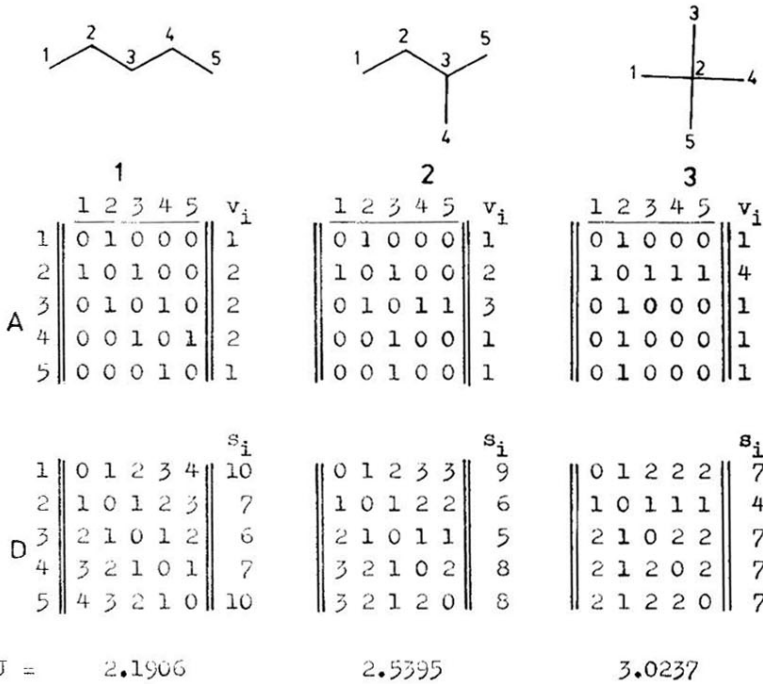


Fig. 1. Adjacency and distance matrices for  $\underline{1} - \underline{3}$  as well as corresponding graph invariants  $v_i$  and  $s_i$ .

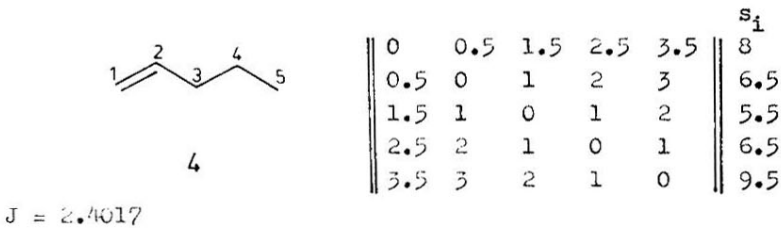
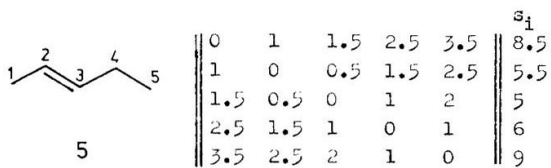
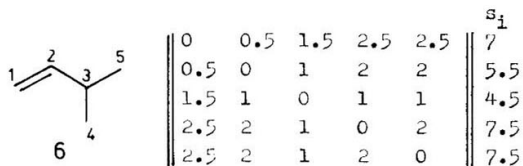


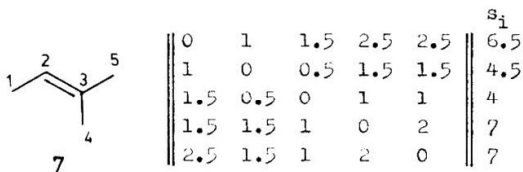
Fig. 2. Distance matrices and distance sums  $s_i$  for  $\underline{4} - \underline{12}$



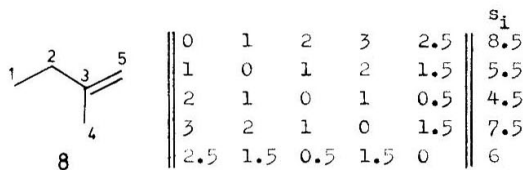
J = 2.6224



J = 2.8257



J = 3.1943



J = 2.8474

Fig. 2 (Continued)

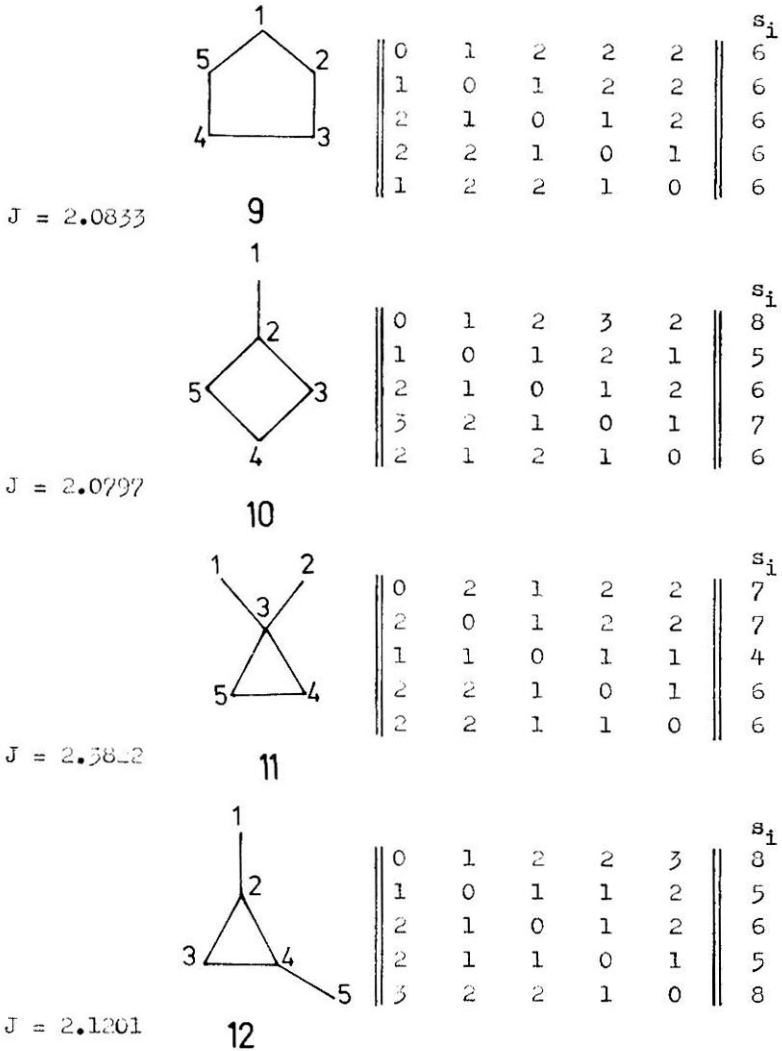


Fig. 2 (Continued)



It can be seen that  $J$  increases with graph branching, and that  $J$  is larger for systems with "internal" double bonds. The listing of the computer program HYDROCARBON 1-J which incorporates this approach is presented in Figures 3 and 4. The input for calculating the topological index  $J$  consists of adjacency data recorded on magnetic tape with the aid of a small interactive routine which is listed in Fig. 3. In order to enable the calculation of  $J$  for large numbers of different structures without computer-operator interaction, the input data are most conveniently recorded on magnetic tape. The adjacency data are inputted as pairs of bonded atoms, and bond type. For the HYDROCARBON 1-J program (Fig. 4), one starts from a null matrix with an order equal to the number  $N_9$  of carbon atoms. For obtaining the adjacency matrix, the program (statements 150-220) replaces the corresponding zeroes by the bond type for each pair of bonded atoms. The adjacency matrix is printed if the corresponding option (statement 30) is chosen; in the printed adjacency matrix, topological distances are presented as reciprocals of the code for bond type. Notation for parameters is presented in Table 1.

Afterwards, the distance matrix is obtained (statements 270-450). The entries of the distance matrix for bonded atoms are obtained by considering the non-zero terms of the printed adjacency matrix. Then the topological distances between the non-bonded atoms are computed on all possible paths, the lowest value being retained for each atom, and are optionally printed as terms of the distance matrix (statements 460-550).

Finally the cyclomatic number and the topological

Table 1

Notation for program parameters and variables

Program	Notation	Parameter
INPUT	X1	number of adjacencies
	N9	number of carbon atoms
	I9	structure identification number
	X[I,1] X[I,2] } X[I,3]	numbering of bonded atoms
		bond type
	F1, F9	first and last data file number, respectively
INPUT plus HYDROCARBON	K2	option parameter for printing: only adjacency matrix (K2 = 1); only distance matrix (K2 = 2); both matrices (K2 = 3); no matrix (K2 = 0)
	A [I,K]	adjacency matrix with bond types
	D [I,K]	distance matrix
	S [I]	distance sum for atom I
	I8	cyclomatic number of graph
	J1	topological index J
INPUT HYDROCARBON plus RADICAL	S0	minimum distance sum in the graph
	N5	numbering of rooted vertex

```
10 COM X1,N9,I9,XS[15,3]
20 DISP "NUMBER OF ATOMS  =";
30 INPUT N9
40 DISP "NUMBER OF BONDS  =";
50 INPUT X1
60 DISP "FIRST DATA FILE NUMBER ";
70 INPUT F1
80 DISP "STRUCTURE IDENTIFICATION NUMBER ";
90 INPUT I9
100 FOR I=1 TO X1
110 DISP "INPUT      ";I;" BOND ";
120 INPUT X[I,1],X[I,2]
130 DISP I;" BOND TYPE ";
140 INPUT X[I,3]
150 NEXT I
160 STORE DATA F1
170 PRINT
180 PRINT "STRUCTURE "I9,LIN2,"BONDS ",LIN2
190 FOR I=1 TO X1
200 PRINT X[I,1];"-";X[I,2];" TYPE ";X[I,3]
210 NEXT I
220 PRINT
230 DISP "DO YOU HAVE MORE DATA (1=YES)";
240 INPUT W9
250 IF W9=0 THEN 280
260 F1=F1+1
270 GOTO 80
280 END
```

Fig. 3. Subroutine for input data of program  
HYDROCARBON 1-J (in BASIC)

```
10 GOM X1,N9,I9,XS[15,3]
20 DIM AS[15,15],DS[15,15],SS[15]
30 DISP "TYPE A,D,A+D (1,2,3) ";
40 INPUT K2
50 DISP "INPUT FIRST DATA FILE NUMBER ";
60 INPUT F1
70 DISP "INPUT LAST DATA FILE NUMBER ";
80 INPUT F9
90 LOAD DATA F1
100 IF I9#0 THEN 120
110 STOP
120 REDIM A[N9,N9],D[N9,N9]
130 PRINT LIN5,"STRUCTURE "I9
140 MAT A=ZER
150 FOR I=1 TO X1
160 A[X[I,1],X[I,2]]=X[I,3]
170 NEXT I
180 FOR I=1 TO N9
190 FOR J=I+1 TO N9
200 A[J,I]=A[I,J]
210 NEXT J
220 NEXT I
230 IF K2=2 OR K2=0 THEN 270
240 PRINT LIN2,"ADJACENCY MATRIX      "
250 K1=1
260 %SUB 710
270 MAT D=ZER
280 FOR I=1 TO N9
290 FOR J=I+1 TO N9
300 IF A[I,J]=0 THEN 320
310 D[I,J]=D[J,I]=1/A[I,J]
320 NEXT J
330 NEXT I
340 FOR I=1 TO N9
350 FOR J=1 TO N9
360 IF I=J THEN 440
370 FOR L=1 TO N9
380 IF I=L OR L=J THEN 430
390 IF D[I,J]=0 OR D[L,J]=0 THEN 430
400 D1=D[I,J]+D[J,L]
410 IF D[I,L]#0 AND D1 >= D[I,L] THEN 430
420 D[I,L]=D[L,I]=D1
430 NEXT L
440 NEXT J
450 NEXT I
460 FOR I=1 TO N9
470 S[I]=0
480 FOR J=1 TO N9
490 S[I]=S[I]+D[I,J]
500 NEXT J
510 NEXT I
520 IF K2<2 THEN 560
530 PRINT LIN2,"DISTANCE MATRIX      "
```

Fig. 4. Program HYDROCARBON 1-J (in BASIC)

```
540 K1=2
550 GO SUB 710
560 J1=0
570 FOR I=1 TO N9
580 FOR J=I+1 TO N9
590 IF A[I,J]=0 THEN 610
600 J1=J1+1/SQR(S[I]*S[J])
610 NEXT J
620 NEXT I
630 I8=X1-N9+1
640 J1=X1*J1/(I8+1)
650 PRINT LIN2,"TOPOLOGICAL INDEX J="J1
660 PRINT "CYCLOMATIC NUMBER = "I8
670 F1=F1+1
680 IF F1>F9 THEN 700
690 GOTO 90
700 END
710 PRINT
720 FOR I=1 TO N9
730 IF I#1 THEN 750
740 WRITE (15,770)" ";
750 WRITE (15,770)I;
760 NEXT I
770 FORMAT F6.0
780 FORMAT F6.2
790 IF K1=1 THEN 810
800 WRITE (15,770)" SI";
810 PRINT
820 PRINT
830 FOR I=1 TO N9
840 WRITE (15,770)I;
850 FOR J=1 TO N9
860 GOTO K1 OF 870,980
870 IF A[I,J]=0 OR A[I,J]=1 THEN 960
880 A2=2*A[I,J]-2
890 GOTO A2 OF 900,920,920,940
900 WRITE (15,770)" 2/3";
910 GOTO 990
920 WRITE (15,770)" 1/2";
930 GOTO 990
940 WRITE (15,770)" 1/3";
950 GOTO 990
960 WRITE (15,770)A[I,J];
970 GOTO 990
980 WRITE (15,780)D[I,J];
990 NEXT J
1000 IF K1=1 THEN 1020
1010 WRITE (15,780)S[I];
1020 PRINT
1030 PRINT
1040 NEXT I
1050 RETURN
1060 END
```

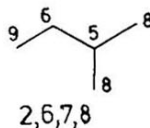
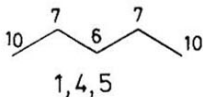
Fig. 4 (Continued)

index J are computed and printed (statements 560-670). The adjacency matrix and the distance matrix are printed with a subroutine which appears in statements 720-1060.

Computer program HYDROCARBON 2-J

A second approach for systems with multiple bonds is to calculate distance sums as integer values ignoring the presence of multiple bonds. Then formula (2) is applied separately for each edge (considering a double bond as two edges and a triple bond as three edges); also in calculating  $q$  in this case each edge is counted separately, even for multiple bonds. This second approach has not been mentioned earlier.

We shall exemplify again with the same structures of the isomeric pentenes 4 - 3. The distance sums for 4 and 5 are those calculated for the saturated congener 1, whereas for 6 - 3 the distance sums are those for 2 :



The resulting J values are :

$$\begin{aligned}
 \underline{4} : J &= 5(3/\sqrt{70} + 2/\sqrt{42}) &= 3.3359 \\
 \underline{5} : J &= 5(2/\sqrt{70} + 3/\sqrt{42}) &= 3.5098 \\
 \underline{6} : J &= 5(2/\sqrt{54} + 1/\sqrt{30} + 2/\sqrt{40}) &= 3.8548 \\
 \underline{7} : J &= 5(1/\sqrt{54} + 2/\sqrt{30} + 2/\sqrt{40}) &= 4.0873 \\
 \underline{8} : J &= 5(1/\sqrt{54} + 1/\sqrt{30} + 3/\sqrt{40}) &= 3.8855
 \end{aligned}$$

It can be observed that the first and second approach for unsaturation give similar ordering of isomers, but not identical numerical results : marginal double bonds lead to smaller J values than internal double bonds, as can be seen by comparing 5 with 4, or 7 with 6 and 8.

The program HYDROCARBON 2-J is presented in Fig. 5.

A drawback of the second approach is that for aromatic systems such as ortho-xylene, different Kekule structures lead to different J values; therefore for aromatic systems, the first approach (HYDROCARBON 1-J) is to be preferred.

#### Computer program RADICAL 1-J

Most biological correlations involve various molecular fragments or radicals attached to one and the same skeleton. So far, no satisfactory rules were devised for TI's or radicals. We present in the present paper an approach to this problem and a computer program implementing the underlying idea.

A molecular (constitutional) graph, e.g. an alkane, can give rise to as many alkyl radicals as the number of constitutionally non-equivalent carbon atoms ; in graph-theoretical terminology, a given graph can give rise to several rooted graphs, where the root vertex corresponds to the carbon atom which has lost a hydrogen on forming the alkyl radical.

For calculating J for a radical we first obtain the s values for the graph, then according to formula (3) we assign to the root vertex a distance sum which is ten times smaller

```
10 COM X1,N9,I9,XS[15,3]
20 DIM AS[15,15],DS[15,15],SS[15]
30 DISP "TYPE A,D,A+D (1,2,3) ";
40 INPUT K2
50 DISP "INPUT FIRST DATA FILE NUMBER ";
60 INPUT F1
70 DISP "INPUT LAST DATA FILE NUMBER ";
80 INPUT F9
90 LOAD DATA F1
100 IF I9#0 THEN 120
110 STOP
120 REDIM A[N9,N9],D[N9,N9]
130 PRINT LIN5,"STRUCTURE "I9
140 MAT A=ZER
150 FOR I=1 TO X1
160 A[X[I,1],X[I,2]]=X[I,3]
170 NEXT I
180 FOR I=1 TO N9
190 FOR J=I+1 TO N9
200 A[J,I]=A[I,J]
210 NEXT J
220 NEXT I
230 IF K2=2 THEN ?70
240 PRINT LIN2,"ADJACENCY MATRIX      "
250 K1=1
260 GOSUB 710
270 MAT D=ZER
280 FOR I=1 TO N9
290 FOR J=I+1 TO N9
300 IF A[I,J]=0 THEN 320
310 D[I,J]=D[J,I]=1
320 NEXT J
330 NEXT I
340 FOR I=1 TO N9
350 FOR J=1 TO N9
360 IF I=J THEN 440
370 FOR L=1 TO N9
380 IF L=I OR L=J THEN 430
390 IF D[I,J]=0 OR D[L,J]=0 THEN 430
400 D1=D[I,J]+D[J,L]
410 IF D[I,L]#0 AND D1 >= D[I,L] THEN 430
420 D[I,L]=D[L,I]=D1
430 NEXT L
440 NEXT J
450 NEXT I
460 FOR I=1 TO N9
470 S[I]=0
480 FOR J=1 TO N9
490 S[I]=S[I]+D[I,J]
500 NEXT J
510 NEXT I
```

Fig. 5. Program HYDROCARBON 2-J (in BASIC)



```
520 IF K2<2 THEN 560
530 PRINT LIN2,"DISTANCE MATRIX      "
540 K1=2
550 GOSUB 710
560 J1=0
570 FOR I=1 TO N9
580 FOR J=I+1 TO N9
590 IF A[I,J]=0 THEN 610
600 J1=J1+A[I,J]/SQR(S[I]*S[J])
610 NEXT J
620 NEXT I
630 I8=X1-N9+1
640 J1=X1*J1/(I8+1)
650 PRINT LIN2,"TOPOLOGICAL INDEX  J="J1
660 PRINT "CYCLOMATIC NUMBER    = "I8
670 F1=F1+1
680 IF F1>F9 THEN 700
690 GOTO 90
700 END
710 PRINT
720 FOR I=1 TO N9
730 IF I#1 THEN 750
740 WRITE (15,770)"      ";
750 WRITE (15,770)I;
760 NEXT I
770 FORMAT F6.0
780 FORMAT F6.2
790 IF K1=1 THEN 810
800 WRITE (15,770)"      SI";
810 PRINT
820 PRINT
830 FOR I=1 TO N9
840 WRITE (15,770)I;
850 FOR J=1 TO N9
860 GOTO K1 OF 870,890
870 WRITE (15,770)A[I,J];
880 GOTO 900
890 WRITE (15,780)D[I,J];
900 NEXT J
910 IF K1=1 THEN 930
920 WRITE (15,780)S[I];
930 PRINT
940 PRINT
950 NEXT I
960 RETURN
970 END
```

Fig. 5 (Continued)

than the minimal distance sum among all graph vertices. Then we apply formula (2).

$$s(\text{root}) = s(\text{min})/10 \quad (3)$$

For example, for the 1-pentyl, 2-pentyl and 3-pentyl radicals derived from 1 the J values thus computed are 3.6643, 4.8365 and 4.8598 respectively. For the 2-methyl-1-butyl, 2-methyl-2-butyl, 2-methyl-3-butyl and 2-methyl-4-butyl radicals derived from 2 the J values are 4.4369, 6.8537, 5.6803 and 4.3046 respectively.

The program RADICAL 1-J (Fig. 6) has a similar structure to that of program HYDROCARBON 1-J, excepting statements 560-730 wherein the smallest distance sum is found and used according to formula (3) for obtaining the distance sum of the rooted vertex. The numbering of the rooted vertex is asked by the program in the same sequence of statements. The program allows the calculation of indices J for several radicals obtained from the same graph by changing the rooted vertex (statement 860 which asks if a loop is wanted).

Unlike program HYDROCARBON 1-J, program RADICAL 1-J uses input data from a data block (statement 1290) which is easily appended at the end of the program on the HP-9830A desk-top computer.

#### Presence of heteroatoms

The presence of heteroatoms is not taken into account in most topological indices ; a few TI's, however, do account for heteroatoms.<sup>1,2,6</sup>

```
10 DIM AS[15,15],DS[15,15],SS[15]
20 DISP "TYPE A,D,A+D (1,2,3) ";
30 INPUT K2
40 DISP "ORDER OF GRAPH ";
50 INPUT N9
60 DISP "NUMBER OF ADJACENCIES ";
70 INPUT X1
80 REDIM A[N9,N9],D[N9,N9]
90 READ I9
100 IF I9#0 THEN 120
110 STOP
120 PRINT LIN5,"STRUCTURE "I9
130 MAT A=ZER
140 FOR I=1 TO X1
150 READ I1,J1,I9
160 A[I1,J1]=I9
170 NEXT I
180 FOR I=1 TO N9
190 FOR J=I+1 TO N9
200 A[J,I]=A[I,J]
210 NEXT J
220 NEXT I
230 IF K2=2 OR K2=0 THEN 270
240 PRINT LIN2,"ADJACENCY MATRIX "
250 K1=1
260 GOSUB 930
270 MAT D=ZER
280 FOR I=1 TO N9
290 FOR J=I+1 TO N9
300 IF A[I,J]=0 THEN 320
310 D[I,J]=D[J,I]=1/A[I,J]
320 NEXT J
330 NEXT I
340 FOR I=1 TO N9
350 FOR J=1 TO N9
360 IF I=J THEN 440
370 FOR L=1 TO N9
380 IF I=I OR I=J THEN 430
390 IF D[I,J]=0 OR D[I,J]=0 THEN 430
400 D1=D[I,J]+D[J,L]
410 IF D[I,L]#0 AND D1 >= D[I,L] THEN 430
420 D[I,L]=D[I,L]=D1
430 NEXT L
440 NEXT J
450 NEXT I
```

Fig. 6. Program RADICAL 1-J (in BASIC)

```
460 FOR I=1 TO N9
470 S[I]=0
480 FOR J=1 TO N9
490 S[I]=S[I]+D[I,J]
500 NEXT J
510 NEXT I
520 IF K2<2 THEN 560
530 PRINT LIN2,"DISTANCE MATRIX      "
540 K1=2
550 GO TO 930
560 S0=100
570 FOR I=1 TO N9
580 IF S[I] >= S0 THEN 600
590 S0=S[I]
600 NEXT I
610 DISP "ROOTED VERTEX      ";
620 INPUT N5
630 S1=S[N5]
640 S[N5]=S0/10
650 PRINT
660 PRINT
670 PRINT "ROOTED VERTEX      = "N5
680 PRINT
690 PRINT "      I      S(I) "
700 PRINT
710 FOR I=1 TO N9
720 WRITE (15,730)I,S[I]
730 FORMAT F4.0,F8.2
740 NEXT I
750 J1=0
760 FOR I=1 TO N9
770 FOR J=I+1 TO N9
780 IF A[I,J]=0 THEN 800
790 J1=J1+1/SQR(S[I]*S[J])
800 NEXT J
810 NEXT I
820 I8=X1-N9+1
830 J1=X1*J1/(I8+1)
840 PRINT LIN2,"TOPOLOGICAL INDEX      J="J1
850 PRINT "CYCLOMATIC NUMBER      = "I8
860 DISP "DO YOU CHANGE THE ROOT (1=YES)";
870 INPUT W9
880 IF W9#1 THEN 910
890 S[N5]=S1
900 GO TO 610
910 GO TO 20
920 END
```

Fig. 6 (Continued)

```
930 PRINT
940 FOR I=1 TO N9
950 IF I#1 THEN 970
960 WRITE (15,990)" ";
970 WRITE (15,990)I;
980 NEXT I
990 FORMAT F6.0
1000 FORMAT F6.2
1010 IF K1=1 THEN 1030
1020 WRITE (15,990)" S I";
1030 PRINT
1040 PRINT
1050 FOR I=1 TO N9
1060 WRITE (15,990)I;
1070 FOR J=1 TO N9
1080 GOTO K1 OF 1090,1200
1090 IF A[I,J]=0 OR A[I,J]=1 THEN 1180
1100 A2=2*A[I,J]-2
1110 GOTO A2 OF 1120,1140,1140,1160
1120 WRITE (15,990)" 2/3";
1130 GOTO 1210
1140 WRITE (15,990)" 1/2";
1150 GOTO 1210
1160 WRITE (15,990)" 1/3";
1170 GOTO 1210
1180 WRITE (15,990)A[I,J];
1190 GOTO 1210
1200 WRITE (15,1000)D[I,J];
1210 NEXT J
1220 IF K1=1 THEN 1240
1230 WRITE (15,1000)S[I];
1240 PRINT
1250 PRINT
1260 NEXT I
1270 RETURN
1280 END
1290 DATA 171,1,2,1,2,3,1,3,4,1,4,5,1
```

Fig. 6 (Continued)

For distance sums in heteroatom-containing molecules a formula (4) based on the atomic number  $Z$  of the heteroatom was proposed by Trinajstić and coworkers.<sup>18</sup> The off-diagonal elements  $d_{ij}$  are calculated normally, but the diagonal elements  $d_{ii}$  are

$$d_{ii} = 1 - \frac{6}{Z_i} \quad (4)$$

where  $Z_i$  is the atomic number of the heteroatom (provided that  $Z_i \geq 6$ ). The idea is extended<sup>18</sup> to determine multiple bond orders  $b_{ij}$  (involving atoms  $i$  and  $j$  and having  $Z_i$  and  $Z_j \geq 6$ ) according to formula (5).

$$b_{ij} = b \frac{Z_i Z_j}{36} \quad (5)$$

Then one can use this bond order  $b_{ij}$  for multigraphs instead of the conventional bond order  $b$  (as indicated above for orders  $b = 1, 1.5, 2$  and  $3$  involving only carbon atoms).

On trying to explain the physical meaning of most topological indices a relationship was found with molecular volume: for several TI's, correlation with the molar refraction evidenced that 70-85% of the explained variance is due to the molecular volume.<sup>19,20</sup> Since atomic volumes, like electronegativities and all chemical properties, show a periodical, rather than a monotonous variation versus the atomic number  $Z$ , we believe that a more accurate way of expressing the presence of heteroatoms ought to take into account this periodical variation.

In a future joint paper with Professor Trinajstić we shall indicate a more elaborate approach than formulas (4) and (5) for taking into account the presence of heteroatoms, on

the basis of the above idea.

### Conclusions

By means of the computer programs presented here, even with a relatively small and slow computer such as the Hewlett-Packard 9830A, it is possible to obtain rapidly for fairly complicated graphs the topological index  $J$  from input data which give only adjacencies and, for multigraphs, corresponding bond orders according to an arbitrary numbering of the graph vertices.

Two different methods for obtaining  $J$  of multigraphs ( HYDROCARBON 1- $J$  and HYDROCARBON 2- $J$  ) were presented, the latter being based on a new approach.

A similar program enabling one to compute  $J$  for molecular fragments ( RADICAL 1- $J$  ) was presented; it is based on a new idea, namely to correlate such fragments or radicals with rooted graphs, and to assign to the rooted vertex a lower distance sum than the minimum distance sum for the corresponding non-rooted graph.

Figures 7 and 8 present the FORTRAN-IV version of the program HYDROCARBON 1- $J$  ; analogous versions of the other programs presented here are available on request.

```
DIMENSION IX(15,3),A(15,15),D(15,15),S(15)
IIN= input channel label
IOUT= output channel label
1000 READ(IIN,100) ICONT,ILEG,NDIM,IDNT
100  FORMAT(4I3)
      IF(ILEG,LE.0) STOP
      WRITE(IOUT,200) IDNT
200  FORMAT(//," STRUCTURE      ",I3)
      DO 1 I=1,ILEG
1     READ(IIN,101)(IX(I,J),J=1,3)
      DO 2 I=1,NDIM
      S(I)=0.0
      DO 2 J=1,NDIM
      A(I,J)=0.0
2     D(I,J)=0.0
      DO 3 I=1,ILEG
      I1=IX(I,1)
      J1=IX(I,2)
      GO TO (4,5,6,6),IX(I,3)
4     AO=1.
      GO TO 7
5     AO=1./1.5
      GO TO 7
6     AO=1./(FLOAT(IX(I,3))-1.)
7     A(I1,J1)=AO
3     A(J1,I1)=AO
      IF(ICONT.EQ.0.OR.ICONT.EQ.2)GOTO 8
      WRITE(IOUT,201)
201  FORMAT(//," ADJACENCY MATRIX  ",/)
      CALL TTMAT(NDIM,A,S,I,IOUT)
8     DO 9 I=1,NDIM
      DO 9 J=1,NDIM
9     D(I,J)=A(I,J)
      DO 11 I=1,NDIM
      DO 11 J=1,NDIM
      IF(I.EQ.J) GO TO 11
      DO 10 K=1,NDIM
```

Fig. 7. The FORTRAN-IV  
version of the program  
HYDROCARBON 1-J



```
IF(K.EQ.I.OR.K.EQ.J) GO TO 10
IF(D(I,J).EQ.0.0.OR.D(K,J).EQ.0.0) GO TO 10
D1=D(I,J)+D(J,K)
IF(D(I,K).NE.0.0.AND.D1.GE.D(I,K)) GO TO 10
D(I,K)=D1
D(K,I)=D1
10 CONTINUE
11 CONTINUE
DO 12 I=1,NDIM
DO 12 J=1,NDIM
12 S(I)=S(I)+D(I,J)
IF(ICONT.LT.2) GO TO 13
WRITE(IOUT,202)
202 FORMAT("//," DISTANCE MATRIX ",/")
CALL TIMAT(NDIM,D,S,2,IOUT)
13 AJ1=0.0
DO 14 I=1,NDIM
I1=I+1
DO 14 J=I1,NDIM
IF(A(I,J).EQ.0.0) GO TO 14
AJ1=AJ1+1./SQRT(S(I)*S(J))
14 CONTINUE
ICYCLE=ILEG - NDIM + 1
AJ1=ILEG*AJ1/(ICYCLE+1)
WRITE(IOUT,203)AJ1
203 FORMAT("//," TOPOLOGICAL INDEX J =",F8.4)
WRITE(IOUT,204)ICYCLE
204 FORMAT("/," CYCLOMATIC NUMBER =",I3)
GO TO 1000
END
```

Fig. 7 (Continued)

```

SUBROUTINE TIMAT(N,A,S,KL,IOUT)
DIMENSION A(N,N),S(N)
WRITE(IOUT,100)(I,I=1,N)
100 FORMAT(/,"      ",15I8)
DO 1 I=1,NDIM
  1 WRITE(IOUT,101)I,(A(I,J),J=1,NDIM)
101 FORMAT(/,I4,15F8.3)
  IF(KL.EQ.1) GO TO 2
  WRITE(IOUT,102)(S(J),J=1,NDIM)
102 FORMAT(/," SI ",15F8.3)
  2 RETURN
END
```

Fig. 8. The FORTRAN-IV version of the  
subroutine for input data of  
program HYDROCARBON 1-J

References

1. A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, in "Steric Effects in Drug Design", Topics Curr. Chem. 114, 21 (1983) (eds. M. Charton and I. Motoc).
2. A.T. Balaban, A. Chiriac, I. Motoc and Z. Simon, "Steric Fit in QSAR", Lecture Notes in Chem. No. 15, Springer, Berlin, 1980, p. 22.
3. N. Trinajstić, "Chemical Graph Theory", The CRC Press, Boca Raton, Florida, 1983, p. 105.
4. 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.
5. D.H. Rouvray, Math. Chem. 1, 125 (1975).
6. L.B. Kier and L.H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.
7. A. Sablić and N. Trinajstić, Acta Pharm. Jugosl. 31, 189 (1981).
8. M. Randić, J. Amer. Chem. Soc., 97, 6609 (1975) ; Int. J. Quantum Chem., Symp., 5, 245 (1978).
9. H. Wiener, J. Amer. Chem. Soc., 69, 17, 2636 (1947) ; J.Chem. Phys., 15, 766 (1947); J. Chem. Phys., 52, 425, 1082 (1948).
10. A.T. Balaban, Theoret. Chim. Acta, 53, 355 (1979).
11. D. Bonchev, A.T. Balaban and M. Randić, Int. J. Quantum Chem., 19, 61 (1981).
12. A.T. Balaban, Pure Appl. Chem., 55, 199 (1983).
13. D. Bonchev and N. Trinajstić, J. Chem. Phys. 67, 4517 (1977)
14. D. Bonchev, Int. J. Quantum Chem. Symp., 12, 293 (1978)

15. D. Bonchev, "Theoretic Information Indices for Characterization of Chemical Structures", Wiley - Research Studies, Chichester, 1983.
16. S.H. Bertz, J. Amer. Chem. Soc., 103, 3599 (1981) ; J. Chem. Soc. Chem. Comm., 819 (1981).
17. A.T. Balaban, Chem. Phys. Lett., 89, 399 (1982).
18. M. Barysz, G. Jashari, R.S. Dall, V.K. Srivatsva and N. Trinajstić, Proc. Symp. "Chemical Applications of Topology and Graph Theory", Athens, Georgia, April 17-21, 1983 (R.B. King, ed.), Elsevier, to appear.
19. I. Motoc and A.T. Balaban, Rev. Roum. Chim., 26, 593 (1981)
20. I. Motoc, A.T. Balaban, O. Mekenyan and D. Bonchev, Math. Chem., 12, 369 (1982).