MATH/CHEM/COMP 1988, Proceedings of an International Course and Conference on the Interfaces between Mathematics, Chemistry and Computer Science, Dubrovnik, Yugoslavia, 20–25 June 1988, A. Graovac (Ed.)
Studies in Physical and Theoretical Chemistry, Volume 63, pages 193–212

FORTRAN 77 COMPUTER PROGRAM FOR CALCULATING THE TOPOLOGICAL INDEX J FOR MOLECULES CONTAINING HETEROATOMS

Alexandru T. Balaban[a] and Ovidiu Ivanciuc[b]

[a]Polytechnic Institute, Organic Chemistry Department, Splaiul Independentei 313, 76206 Bucharest, Roumania

[b]Polytechnic Institute, Organic Chemistry Department, 1900 Timisoara, Roumania

A FORTRAN 77 computer program is presented, which allows the computing of the topological indices J (average distance sum connectivity) and PDS (parametrized distance sum), taking into account the presence of unsaturation, aromaticity and the chemical nature of the atoms in the molecule: JX and PDSX are based on electronegativities of heteroatoms, while JY and PDSY are based on covalent radii of heteroatoms. Graphs are encoded in an efficient and economical manner.

Molecular topology determines a large number of molecular properties, ranging from physico-chemical and thermodynamic properties to chemical reactivity and biological activity. Organic molecules are represented by hydrogen-depleted graphs depicting the covalent bonds between non-hydrogen atoms. The topology of a chemical structure can be encoded in matrix form by the use of the adjacency matrix or the distance matrix.[1]

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

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

The distance matrix of a graph G with N vertices, D(G)=D, is a square NxN symmetrical matrix, whose entries, d_{ij} are equal to the number of edges connecting vertices i and J on the shortest path between them.

The two types of matrices are interrelated and the distance matrix can be computed from the connectivity matrix using a simple algorithm based on higher powers of the adjacency matrix.[3]

From a practical point of view, a more economical way of expressing the topology of a chemical structure is represented by another class of graph invariants, usually referred to as topological indices.[4-11] A topological index (TI) is a number which characterizes the bonding topology of a molecule.

Although in their origins TI's were developed for the purpose of obtaining correlations with a variety of physico-chemical properties of chemical substances, namely QSPR (Quantitative Structure-Property Relationships), their range of applications was extended to bibliographical documentation of chemical compounds and QSAR (Quantitative Structure-Activity Relationships).

An almost general deficiency of topological indices is that they do not characterize uniquely the topology of a molecular graph, but are more or less degenerate, i.e. two or more nonisomorphic structures may lead to the same numerical value for a certain topological index. High structural selectivity of a TI is highly needed for bibliographical indexing, while in QSPR and QSAR the essential characteristic of a TI is to reflect structural information in a proper way, in connexion with the property under investigation.

A brief review of the most used TI's follows.

TI's M_1 and M_2, the Zagreb group indices,[12] are based on the adjacency matrix:

$$M_1 = \sum_{i=1}^{N} v_i^2 \qquad (2)$$

$$M_2 = \sum_{\text{edges } ij} (v_i v_j)^{1/2} \qquad (3)$$

where v_i is the degree of the vertex i in the hydrogen-suppressed graph.

The connectivity index χ, similar to the M_2 index, was introduced by Randic:[13]

$$\chi = \sum_{\text{edges } ij} (v_i v_j)^{-1/2} \qquad (4)$$

The TI Z was introduced by Hosoya:[14,15]

$$Z = \sum_{k=0}^{[N/2]} p(G,k) \qquad (5)$$

where p(G,k) is the number of ways in which k edges are chosen from

the graph G in such a way that no two of them are adjacent; N/2 in square brackets is the smallest integer not exceeding the real number in them. By definition, p(G,0)=1, while p(G,1) equals the number of edges in the molecular graph.

The path number index W, defined by Wiener,[16-18] equals the half-sum of the off-diagonal elements of the distance matrix:

$$W=1/2\sum_{i,j} d_{i,j} \qquad (6)$$

The topological index J[19,20] (average distance sum connectivity) is defined by the formula:

$$J=\frac{q}{q-p+2}\sum_{\text{edges } ij}(s_i s_j)^{-1/2} \qquad (7)$$

where the summation is over all edges e_{ij} of the molecular graph. The number of edges and vertices in the molecular graph are q and p, respectively. The distance sum s_i for vertex i represents the sum of topological distances from vertex i to all other vertices in the graph, i.e. $\sum_i d_{ij}$.

Owing to the factor q/(q-p+2), the index J does not increase automatically with the increasing number of vertices and rings, unlike most TI's. It has a very low degeneracy, as was proved analytically[21] and tested with a computer program.[22] For many infinite graphs, J attains a finite limit,[23] e.g. for an infinite linear alkane, J tends towards $J\pi=3.14159$. A mathematical study of the branching problem[24] has revealed that among all TI's under consideration, J is the only one which orders trees (molecular graphs of alkanes) up to heptane in a similar manner to Bertz's graph invariants.

The widely used informational TI's are those defined by Bonchev and Trinajstic:[25]

$$I_D^E=N^2 lb(N^2)-N lb(N)-\sum_{i=1} 2k_i lb(2k_i) \qquad (8)$$

$$\bar{I}_D^E=-\sum 2k_i/(N^2-N)\ lb[2k_i/(N^2-N)] \qquad (9)$$

$$I_D^W=W lb(W)-\sum k_i i\ lb(i) \qquad (10)$$

$$\bar{I}_D^W=-\sum k_i i/W\ lb(i/W) \qquad (11)$$

where k_i is the number of topological distances of length i in the molecular graph.

A comparison of the structural selectivity between six TI's : M_1, W, Z, χ, I_D^E and J was reported[26] for a fairly large sample of hydrocarbons. The chemical structures on which these indices were tested were all constitutional isomers of alkanes with 5-12 carbon atoms (659 structures) and all monocyclic and bicyclic saturated hydrocarbons with 4-8 carbon atoms (376 structures). Another comparison between these TI's consisted in the distribution of their numerical values for classes of constitutional isomers, exemplified by the 35 nonane isomers. It was concluded that J has the lowest degeneracy among the six TI's tested, but afforded less satisfactory correlations than some other TI's.

For the presence of heteroatoms in molecules , several approaches were proposed . The first and the widely used one is Randic's molecular connectivity index,[13] modified by Kier and Hall.[9,10] This method takes into account the number of valence electrons and was successfully applied in QSAR. Other methods were developed by Trinajstic and coworkers[27] and by Basak and coworkers.[28]

A different approach was proposed in order to take into account the periodicity of chemical properties for heteroatoms,[29] based on heteroatom electronegativities and on heteroatom covalent radii, respectively.

The effect of a heteroatom is characterized by its calculated relative electronegativity using a two-parametric linear equation. The two parameters for an atom i are Z_i (atomic number) and G_i (the number of the group in Mendeleev's short form of the Periodic System):

$$X_i=0.4196-0.0078Z_i+0.1567G_i \qquad (12)$$

The calculated relative electronegativity is relative to the calculated electronegativity for carbon.

The second method of caracterizing the effect of a heteroatom i is based on its relative covalent radius Y_i, based on calculated covalent radii of carbon. Parameters Y_i are obtained according to equation:

$$Y_i=1.1191+0.0160Z_i-0.0053G_i \qquad (13)$$

In order to obtain the TI J for heteroatom-containing molecules (JX and JY) the vertex distance sums s_i are multiplied by the parameters X_i or Y_i,corresponding to the chemical nature of the

atoms and with a scaling factor FX_i or FY_i, and then we apply the usual formula (7). Factors FX_i and FY_i, which in earlier papers were assumed to be 1, are introduced now for enhancing or depressing the effect of heteroatoms; for carbon, FX=FY=1.

We now define two new TI's, electronegativity- and sterically-parametrized distance sums, PDSX and PDSY, respectively:

$$PDSX = \sum_i FX_i X_i s_i$$
$$PDSY = \sum_i FY_i Y_i s_i \qquad (14)$$

These indices increase steadily with increasing number N of atoms, like most TI's, and especially like the Wiener index W. For n-pentane, isopentane and neopentane the W values are 40, 36 and 32, respectively.

COMPUTER PROGRAM JHET

In order to compute TI's JX, JY, PDSX and PDSY, the computer program JHET was developed, which takes into account the presence of unsaturation, aromaticity and the chemical nature of the atoms in the molecule.

The program, which was written in FORTRAN-77 and implemented on a CORAL 4030 computer, is intended to be user-friendly and was designed to operate in two different ways: interactively with the operator from the terminal, and through the agency of two files stored on a mass-disk. The listing of the program is presented in FIGURE 1.

A brief description of the main parameters and matrices of the program follows.

- NAT - number of atoms
- NB - number of bonds
- IBOND - bond type index, having values 1, 2, 3 and 4 for simple, double, triple and aromatic bonds, respectively
- NSB - number of supplementary bonds
- IBO - index of saturation, having value 1 if the molecular graph has only single bonds, and value 2 otherwise
- NAME(30) - array containing the name of the chemical compound
- NZ(30) - array containing the atomic number of the atoms
- ICOD(30) - the connectivity code of the molecular graph

JBO(30) - array containing the bond type index corresponding to the bonds coded in ICOD

ISB(30,2)- matrix containing the supplementary bonds; the number of the two atoms of bond i are located in the pair ISB(i,1) and ISB(i,2), respectively

A(30,30) - adjacency matrix

D(30,30) - distance matrix

PA(2,60) - matrix containing the calculated relative electronegativities and covalent radii, respectively; PA(1,Zi) contains the calculated relative electronegativity and PA(2,Zi) the calculated relative covalent radius of the element with the atomic number Zi

DS(30) - array containing the distance sums

VTIX(30) - array containing the electronegativity vertex topological indices (VTI's)

VTIY(30) - array containing the steric VTI's

JX - electronegativity-parametrized TI J

JY - sterically-parametrized TI J

SF - scaling factor

PDSX - electronegativity-parametrized distance sum

PDSY - stericaly-parametrized distance sum

First, from the terminal, the operating mode has to be chosen through the values of two indicators: INTR and IOUT. Input data can be supplied from the terminal (INTR=1) or from a datafile (INTR=2) and results are directed to the terminal (IOUT=1) or to a datafile (IOUT=2). Then the scaling factor SF is introduced.

If input data are provided from the terminal, the following routine has to be performed for every molecular structure:

- the name of the compound, NAME, is introduced. If the set of molecular structures is exhausted, a return command (<CR>) must be given at this stage and the program stops.
- the number of atoms, NAT, and the number of bonds, NB, are introduced.
- the vertices of the input structure are arbitrarly labelled with numbers between 1 and NAT, and for each bond between vertices I and J the bond order IBOND is supplied in a DO cycle running over the NB bonds.
- in another DO cycle, the atomic number Z of every atom is introduced.

FIGURE 2 presents the output of the program JHET with input

data supplied from the terminal.

Alternatively, if data are provided from a datafile, the program opens the file JHIN.DAT, where structural information must be previously introduced. A somewhat different way of coding is used, in comparison with the one used if input data are provided from the terminal. This is a more compact and space saving procedure of coding a molecular graph.

If the molecular graph is a polycyclic one, we define the NSB quantity, i.e. the minimum number of bonds by whose removal the graph is transformed into a tree. It is easy to observe that NSB equals the cyclomatic number of the graph, i.e. the number of rings in the definition of polycyclic compounds according to the IUPAC and Chemical Abstracts nomenclature systems.

The molecular structure is transformed into an arbitrary spanning tree by removing NSB bonds. An arbitrary vertex is defined to be the focal vertex of the tree and is numbered with 1. The rest of NAT-1 vertices are numbered in such a way that on every path of the tree if $d(1,J1)<d(1,J2)$ then $J1<J2$.

The tree is coded by a sequence of numbers, in which at the position i appears the number of the vertex closer to the focal one to which is bonded atom i+1. Thus, the sequence begins with the digit 1, the vertex to which atom bearing number 2 is bonded, then the second digit indicates the vertex bonded to atom number 3, closer to the focal point,and so on.[30,31]

Thus, every polycyclic compound can be coded by a tree-code and a sequence of NSB edges which transform the tree into the original polycyclic graph.

We have to note that this code is not unique, but depends on the vertex defined to be the focal one, and on the selection of the spanning tree. If we impose a restrictive condition, namely that the sequence of digits, read as a number in a NAT-1 base, is the minimal one, we obtain a unique code for a given tree. It is important to note that this condition is not restrictive for the input of the program.

Taking into account the above way of coding a molecular structure, the input information read from the file JHIN.DAT is contained in the following items:

-item 1 contains the name of the compound, NAME, in FORMAT 30A1.

-item 2 contains the values for three indicators: NAT (number of atoms), NSB (number of supplementary bonds reforming the original graph from the spanning tree) and IBO (index of bonds),

in FORMAT 3I2.

-item 3 contains the code of the spanning tree of the graph, ICOD, in FORMAT 30I2.

-item 4 contains the array JBO in FORMAT 30I1; JBO(I) contains the bond index of the bond I, between vertices I+1 and ICOD(I). If IBO=1, this item is missing.

-item 5 contains matrix ISB in pairs ISB(I,1) and ISB(I,2) corresponding to the vertices forming the supplementary bond I, in FORMAT 30I2. If NSB=0, this item is absent.

-item 6 contains the bond indices of the supplementary bonds, in FORMAT 30I1, read in the array JBO. If NSB=0 or IBO=1, this item is missing.

-item 7 contains the array NZ in FORMAT 30I2.

Based on the input data, the program contructs the adJacency matrix A and the initial distance matrix, whose entries are 1/b if between atoms i and j there exists a bond with the bond order b, otherwise d(i,J) equals 0. For bonds in aromatic rings, b=1.5.

Then subroutine DIST computes the distance matrix D, and subroutine SJH computes TI's JX, JY, PDSX and PDSY. If IOUT=1, the values of the TI's are displayed on the terminal. In the other case, when IOUT=2, the output is directed to the file JHOUT.DAT and the program writes an item containing NAME, JX, JY, PDSX and PDSY in FORMAT 1X,30A1, 4(2X,F14.6).

FIGURE 3 contains datafile JHIN.DAT in which input data is stored. Datafile JHOUT.DAT containing output data in a test run is presented in FIGURE 4.

It can be observed from FIGURE 4 that if the connectivity of the heteroatom increases from 1 to 4, JX and PDSX decrease markedly whereas JY and PDSY increase slightly in the following three series: butylamine - N-methylpropylamine - diethylamine; isobutylamine - sec-butylamine - N-methylisopropylamine - N,N-dimethylethylamine; tert-butylamine - tetramethylammonium. On the other hand, on comparing primary amines one can observe that JX and JY increase whereas PDSX and PDSY decrease with increasing branching.

These observations agree with earlier data on the heteroatom-modified index J, and on the Wiener index. Indeed, the above three series have PDSY values which are close to the Wiener index for n-pentane, isopentane and neopentane, respectively; for PDSX the values have a slightly larger spread. The above three hydrocarbons also served for comparing other TI's.[32,33]

```
C     ************************************************************
      PROGRAM JHET
C     ************************************************************
C
C                  PROGRAM JHET
C
C       THIS  PROGRAM  CALCULATES  THE  TOPOLOGICAL  INDICES
C       J  ( AVERAGE  DISTANCE  SUM   CONNECTIVITY )   AND
C       PDS ( PARAMETRIZED DISTANCE SUM ), TAKING INTO ACCOUNT
C       :   THE PRESENCE  OF  UNSATURATION,  AROMATICITY  AND
C       THE  CHEMICAL  NATURE  OF  THE ATOMS  IN THE MOLECULE,
C       ACCORDING  TO  ELECTRONEGATIVITIES  OR  COVALENT RADII
C
C       REFERENCES :
C
C           1. A.T. Balaban, Chem. Phys. Lett., 80, 399 (1982).
C           2. A.T. Balaban, Pure Appl. Chem., 55, 189 (1983).
C           3. A.T. Balaban and P. Filip, Math. Chem., 14, 163 (1984).
C           4. A.T. Balaban, Math. Chem., 21, 115 (1986).
C
C
C       FIRST, FROM TERMINAL, TWO INDICATORS AND THE SCALING FACTOR
C       ARE SUPPLIED :
C
C         INTR=1 : DATA IS SUPPLIED FROM THE TERMINAL
C             =2 : DATA IS SUPPLIED FROM A DATAFILE
C
C         IOUT=1 : RESULTS ARE DIRECTED TO THE TERMINAL
C             =2 : RESULTS ARE DIRECTED TO A DATAFILE
C
C         SF     : SCALING FACTOR   FORMAT : F10.3
C
C       INPUT DESCRIPTION :
C
C         IF INTR=1 THE FOLLOWING DATA ARE PROVIDED FROM TERMINAL :
C
C         NAME(30) : THE NAME OF THE COMPOUND
C
C         NAT      : NUMBER OF ATOMS    FORMAT : I2
C                    THE UPPER VALUE OF NAT IS 30
C
C         NB       : NUMBER OF BONDS
C
C         BONDING DATA : I , J , IBOND
C
C                        I AND J ARE THE NUMBERS OF THE ATOMS FROM THE
C                        TWO ENDS OF THE BOND
C
C                        IBOND=1 : SINGLE BOND
C                             =2 : DOUBLE BOND
C                             =3 : TRIPLE BOND
C                             =4 : AROMATIC BOND
C
C         Z DATA : THE ATOMIC NUMBER Z IS INTRODUCED FOR EVERY ATOM.
C
C         AFTER  FINISHING ONE MOLECULAR STRUCTURE COMPUTATION, THE
C         USER IS ASKED IF ANOTHER RUN IS TO BE PERFORMED. IF THE
```

FIGURE 1. Program JHET

```
C       ANSWER IS 'N', THE PROGRAM STOPS AND IF IT IS 'Y' THE
C       PROGRAM SEEKS THE NEXT MOLECULAR STRUCTURE
C
C       IF INTR=2 INPUT DATA ARE READ FROM THE FILE  JHIN.DAT
C
C       NAME(30) : THE NAME OF THE COMPOUND
C
C       NAT,NSB,IBO : FORMAT 3I2
C             NAT - NUMBER OF ATOMS
C             NSB - NUMBER OF SUPPLEMENTARY BONDS (CYCLOMATIC NUMBER)
C             IBO - INDEX OF SATURATION
C             IBO=1 : THE COMPOUND HAS ONLY SINGLE BONDS
C                =2 : THE COMPOUND HAS DOUBLE, TRIPLE OR AROMATIC
C                     BONDS
C
C       ICOD(30) : THE CONNECTIVITY CODE FOR THE SPANNING TREE OF THE
C                  MOLECULAR GRAPH, IN FORMAT 30I2
C
C       JBO(30)  : ARRAY CONTAINING THE BOND TYPE INDEX CORRESPON-
C                  DING TO THE BONDS CODED IN ICOD, IN FORMAT 30I1
C                  IF IBO=1, THIS ITEM IS MISSING
C
C       ISB(30,2): MATRIX CODING THE NSB BONDS. A BOND IS CODED BY
C                  A PAIR ISB(I,1) AND ISB(I,2), GIVING THE
C                  NUMBERS OF THE TWO ATOMS OF THE BOND IN
C                  FORMAT 30I2
C                  IF NSB=0, THIS ITEM IS MISSING
C
C       JBO(30)  : ARRAY CONTAINING THE BOND TYPE INDEX CORRESPON-
C                  DING TO THE BONDS CODED IN ISB, IN FORMAT 30I1
C                  IF IBO=1 OR NSB=0 THIS ITEM IS MISSING
C
C       NZ(30)   : ARRAY CONTAINING THE ATOMIC NUMBER OF THE
C                  ATOMS, IN FORMAT 30I2
C
C       AFTER  FINISHING ONE MOLECULAR STRUCTURE COMPUTATION, THE
C       PROGRAM AUTOMATICALLY LOOKS TO SEE IF DATA FOR ANOTHER
C       MOLECULE FOLLOWS
C
C   OUTPUT DESCRIPTION :
C
C       JX     : ELECTRONEGATIVITY TOPOLOGICAL INDEX J
C
C       JY     : STERIC TOPOLOGICAL INDEX J
C
C       PDSX   : ELECTRONEGATIVITY PARAMETRIZED DISTANCE SUM
C
C       PDSY   : STERICALLY PARAMETRIZED DISTANCE SUM
C
C       IF IOUT=2, RESULTS ARE TYPED IN FILE JHOUT.DAT IN AN
C       ITEM CONTAINING NAME, JX, JY, PDSX AND PDSY IN
C       FORMAT 1X,30A1,4(2X,F14.6)
C
C   WORK MATRICES :
C
C       A(30,30)   - ADJACENCY MATRIX
C
C       D(30,30)   - DISTANCE MATRIX
C
```

FIGURE 1. (continued)

```
C       PA(2,60)   - HETEROATOM PARAMETERS MATRIX. PA(1,ZI)
C                    CONTAINS THE CALCULATED RELATIVE ELECTRO-
C                    NEGATIVITY AND PA(2,ZI) CONTAINS THE
C                    CALCULATED RELATIVE COVALENT RADIUS OF THE
C                    ELEMENT WITH THE ATOMIC NUMBER ZI
C
        CHARACTER*1 NAME(30),ANS
        DIMENSION A(30,30),D(30,30)
        DIMENSION PA(2,60),NZ(30)
        DIMENSION ICOD(30),JBO(30),ISB(30,2)
        REAL JX,JY
        COMMON A,D,PA,NZ
        PA(1,5)=0.851
        PA(2,5)=1.038
        PA(1,6)=1.000
        PA(2,6)=1.000
        PA(1,7)=1.149
        PA(2,7)=0.963
        PA(1,8)=1.297
        PA(2,8)=0.925
        PA(1,9)=1.446
        PA(2,9)=0.887
        PA(1,14)=0.937
        PA(2,14)=1.128
        PA(1,15)=1.086
        PA(2,15)=1.091
        PA(1,16)=1.235
        PA(2,16)=1.053
        PA(1,17)=1.384
        PA(2,17)=1.015
        PA(1,33)=0.946
        PA(2,33)=1.379
        PA(1,34)=1.095
        PA(2,34)=1.341
        PA(1,35)=1.244
        PA(2,35)=1.303
        PA(1,52)=0.954
        PA(2,52)=1.629
        PA(1,53)=1.103
        PA(2,53)=1.591
10      TYPE 100
        READ(5,101,ERR=10)INTR
        IF(INTR.NE.1.AND.INTR.NE.2)GO TO 10
11      TYPE 102
        READ(5,101,ERR=11)IOUT
        IF(IOUT.NE.1.AND.IOUT.NE.2)GO TO 11
12      TYPE 121
        READ(5,122,ERR=12)SF
        IF(INTR.EQ.2)OPEN (1,FILE='JHIN.DAT',STATUS='OLD')
        IF(IOUT.EQ.2)THEN
        OPEN (2,FILE='JHOUT.DAT',STATUS='NEW')
        WRITE(2,120)
        ENDIF
1       CONTINUE
        DO 9 I=1,30
        NAME(I)=' '
        DO 9 J=1,30
        A(I,J)=0.
```

FIGURE 1. (continued)

```
      D(I,J)=0.
9     CONTINUE
      IF(INTR.EQ.1)THEN
      TYPE *,' '
      TYPE 103
      READ(5,104)NQ,(NAME(I),I=1,NQ)
13    TYPE 105
      READ(5,106,ERR=13)NAT
      IF(NAT.GT.30)THEN
      TYPE 220
      GO TO 1
      ENDIF
14    TYPE 107
      READ(5,106,ERR=14)NB
      TYPE *,' '
      DO 2 K=1,NB
15    TYPE 108,K
      READ(5,112,ERR=15)I,J,IBOND
      IF(IBOND.LT.0.OR.IBOND.GT.4)GO TO 15
      IF(IBOND.EQ.4)THEN
      D(I,J)=1./1.5
      ELSE
      D(I,J)=1./FLOAT(IBOND)
      ENDIF
      D(J,I)=D(I,J)
      A(I,J)=1.
      A(J,I)=1.
2     CONTINUE
      TYPE *,' '
      DO 3 I=1,NAT
16    TYPE 109,I
      READ(5,106,ERR=16)NZ(I)
3     CONTINUE
      ELSE
      KODER=0
      READ(1,104,END=999)NQ,(NAME(I),I=1,NQ)
      READ(1,106)NAT,NSB,IBO
      IF(NAT.GT.30)THEN
      KODER=1
      TYPE 201,(NAME(I),I=1,NQ)
      TYPE 220
      ENDIF
      NAT1=NAT-1
      READ(1,110)(ICOD(I),I=1,NAT1)
      IF(IBO.EQ.1)THEN
      DO 8 I=1,NAT1
      JBO(I)=1
8     CONTINUE
      ELSE
      READ(1,111)(JBO(I),I=1,NAT1)
      ENDIF
      DO 4 I=1,NAT1
      I1=I+1
      ICO=ICOD(I)
      IF(JBO(I).EQ.4)THEN
      D(I1,ICO)=1./1.5
      ELSE
      D(I1,ICO)=1./FLOAT(JBO(I))
```

FIGURE 1. (continued)

```
        ENDIF
        D(ICO,I1)=D(I1,ICO)
        A(I1,ICO)=1.
        A(ICO,I1)=1.
4       CONTINUE
        IF(NSB.EQ.0)GO TO 17
        READ(1,110)(ISB(I,1),ISB(I,2),I=1,NSB)
        IF(IBO.EQ.1)THEN
        DO 7 I=1,NSB
        JBO(I)=1
7       CONTINUE
        ELSE
        READ(1,111)(JBO(I),I=1,NSB)
        ENDIF
        DO 5 I=1,NSB
        I1=ISB(I,1)
        I2=ISB(I,2)
        IF(JBO(I).EQ.4)THEN
        D(I1,I2)=1./1.5
        ELSE
        D(I1,I2)=1./FLOAT(JBO(I))
        ENDIF
        D(I2,I1)=D(I1,I2)
        A(I1,I2)=1.
        A(I2,I1)=1.
5       CONTINUE
17      READ(1,110)(NZ(I),I=1,NAT)
        NB=NAT+NSB-1
        IF(KODER.EQ.1)GO TO 1
        ENDIF
        CALL DIST(NAT,NB)
        CALL SJH(NAT,NB,SF,JX,JY,PDSX,PDSY)
        IF(IOUT.EQ.1)THEN
        TYPE 201,(NAME(I),I=1,NQ)
        TYPE 202,JX,JY,PDSX,PDSY
        ELSE
        WRITE(2,203)(NAME(I),I=1,30),JX,JY,PDSX,PDSY
        ENDIF
        IF(INTR.EQ.1)THEN
20      TYPE 115
        READ(5,116)ANS
        IF(ANS.EQ.'Y')GO TO 1
        IF(ANS.EQ.'N')GO TO 999
        GO TO 20
        ELSE
        GO TO 1
        ENDIF
999     CONTINUE
        IF(INTR.EQ.2)CLOSE (1)
        IF(IOUT.EQ.2)CLOSE (2)
        STOP
100     FORMAT('$DATA IS SUPPLIED FROM   : TERMINAL-1 DATAFILE-2 : ')
101     FORMAT(I1)
102     FORMAT('$RESULTS ARE DIRECTED TO : TERMINAL-1 DATAFILE-2 : ')
121     FORMAT('$SCALING FACTOR SF = ')
122     FORMAT(F10.3)
120     FORMAT(37X,'JX',14X,'JY',12X,'PDSX',12X,'PDSY'//)
103     FORMAT('$NAME OF THE COMPOUND : ')
```

FIGURE 1. (continued)

```
104     FORMAT(Q,30A1)
105     FORMAT('$NUMBER OF ATOMS : ')
106     FORMAT(3I2)
107     FORMAT('$NUMBER OF BONDS : ')
108     FORMAT('$ BOND : ',I2,'  I , J , IBOND : ')
109     FORMAT('$ ATOM  : ',I2,'  Z = ')
110     FORMAT(30I2)
111     FORMAT(30I1)
112     FORMAT(3I3)
115     FORMAT('$DO YOU WANT ANOTHER RUN? [Y/N] : ')
116     FORMAT(A1)
201     FORMAT(//10X,'NAME : ',30A1)
202     FORMAT(/5X,'   JX=',F14.6,5X,'    JY=',F14.6/
     15X,' PDSX=',F14.6,5X,' PDSY=',F14.6//)
203     FORMAT(1X,30A1,4(2X,F14.6))
220     FORMAT('   WARNING  !!!  N > 30 !')
        END
C   ************************************************************
        SUBROUTINE DIST(NAT,NB)
C   ************************************************************
C
C       SUBROUTINE DIST COMPUTES THE DISTANCE MATRIX OF THE
C       MOLECULAR GRAPH
C
        DIMENSION A(30,30),D(30,30)
        DIMENSION PA(2,60),NZ(30)
        COMMON A,D,PA,NZ
        DO 51 I=1,NAT
        DO 52 J=1,NAT
        IF(I.EQ.J)GO TO 52
        DO 53 K=1,NAT
        IF(K.EQ.I.OR.K.EQ.J)GO TO 53
        IF(D(I,J).EQ.0..OR.D(K,J).EQ.0.)GO TO 53
        DD=D(I,J)+D(J,K)
        IF(D(I,K).NE.0..AND.DD.GE.D(I,K))GO TO 53
        D(I,K)=DD
        D(K,I)=DD
53      CONTINUE
52      CONTINUE
51      CONTINUE
        RETURN
        END
C   ************************************************************
        SUBROUTINE SJH(NAT,NB,SF,JX,JY,PDSX,PDSY)
C   ************************************************************
C
C       SUBROUTINE SJH COMPUTES TOPOLOGICAL INDICES
C       JX, JY, PDSX AND PDSY
C
C   DS(I)    :  CONTAINS THE DISTANCE SUM OF ATOM I
C
C   VTIX(I)  :  CONTAINS THE VERTEX PARAMETER OF
C               ATOM I, BASED ON THE RELATIVE ELECTRONEGATIVITY
C
C   VTIY(I)  :  CONTAINS THE VERTEX PARAMETER OF
C               ATOM I, BASED ON THE RELATIVE COVALENT RADII
C
```

FIGURE 1. (continued)

```
      DIMENSION A(30,30),D(30,30)
      DIMENSION PA(2,60),NZ(30),DS(30),VTIX(30),VTIY(30)
      COMMON A,D,PA,NZ
      REAL JX,JY
      DO 9 I=1,30
9     DS(I)=0.
      DO 10 I=1,NAT
      DO 10 J=1,NAT
      IF(I.EQ.J)GO TO 10
      DS(I)=DS(I)+D(I,J)
10    CONTINUE
      JX=0.
      JY=0.
      PDSX=0.
      PDSY=0.
      DO 11 I=1,NAT
      IND=NZ(I)
      IF(IND.EQ.6)THEN
      VTIX(I)=DS(I)*PA(1,IND)
      VTIY(I)=DS(I)*PA(2,IND)
      ELSE
      VTIX(I)=SF*DS(I)*PA(1,IND)
      VTIY(I)=SF*DS(I)*PA(2,IND)
      ENDIF
11    CONTINUE
      DO 12 I=1,NAT
      PDSX=PDSX+VTIX(I)
      PDSY=PDSY+VTIY(I)
      DO 12 J=I+1,NAT
      IF(A(I,J).EQ.0.)GO TO 12
      JX=JX+1./SQRT(VTIX(I)*VTIX(J))
      JY=JY+1./SQRT(VTIY(I)*VTIY(J))
12    CONTINUE
      CYCL=FLOAT(NB)/FLOAT(NB-NAT+2)
      JX=JX*CYCL
      JY=JY*CYCL
      RETURN
      END
```

FIGURE 1. (continued)

```
 DATA IS SUPPLIED FROM   : TERMINAL-1 DATAFILE-2 : 1
 RESULTS ARE DIRECTED TO : TERMINAL-1 DATAFILE-2 : 2
 SCALING FACTOR SF = 1.

 NAME OF THE COMPOUND : ISOBUTYLAMINE
 NUMBER OF ATOMS : 5
 NUMBER OF BONDS : 4

  BOND :  1  I , J , IBOND : 1,2,1
  BOND :  2  I , J , IBOND : 2,3,1
  BOND :  3  I , J , IBOND : 3,4,1
  BOND :  4  I , J , IBOND : 3,5,1

  ATOM  :  1  Z = 7
  ATOM  :  2  Z = 6
  ATOM  :  3  Z = 6
  ATOM  :  4  Z = 6
  ATOM  :  5  Z = 6

        NAME : ISOBUTYLAMINE

      JX=       2.503020          JY=       2.549807
    PDSX=      37.341003        PDSY=      35.667000

 DO YOU WANT ANOTHER RUN? [Y/N] : N
```

FIGURE 2. Output data of program JHET with input data supplied from the terminal.

```
BUTYLAMINE
 5 0 1
 1 2 3 4
 7 6 6 6 6
sec-BUTYLAMINE
 5 0 1
 1 2 3 2
 7 6 6 6 6
ISOBUTYLAMINE
 5 0 1
 1 2 3 3
 7 6 6 6 6
tert-BUTYLAMINE
 5 0 1
 1 2 2 2
 7 6 6 6 6
N-METHYLPROPYLAMINE
 5 0 1
 1 2 3 4
 6 7 6 6 6
DIETHYLAMINE
 5 0 1
 1 2 3 4
 6 6 7 6 6
N-METHYLISOPROPYLAMINE
 5 0 1
 1 2 3 3
 6 7 6 6 6
N,N-DIMETHYLETHYLAMINE
 5 0 1
 1 2 3 3
 6 6 7 6 6
TETRAMETHYLAMMONIUM
 5 0 1
 1 2 2 2
 6 7 6 6 6
```

FIGURE 3. File JHIN.DAT used as input data for program in a work-session using datafiles.

NAME	JX	JY	PDSX	PDSY
BUTYLAMINE	2.158535	2.199708	41.490002	39.630001
sec-BUTYLAMINE	2.497108	2.551574	37.192001	35.704002
ISOBUTYLAMINE	2.503020	2.549897	37.341003	35.667000
tert-BUTYLAMINE	2.973001	3.038101	33.043003	31.741001
N-METHYLPROPYLAMINE	2.117126	2.211453	41.043003	39.741001
DIETHYLAMINE	2.107793	2.214101	40.894001	39.778000
N-METHYLISOPROPYLAMINE	2.454025	2.563795	36.894001	35.778000
N,N-DIMETHYLETHYLAMINE	2.405682	2.577507	36.745003	35.815002
TETRAMETHYLAMMONIUM	2.820856	3.081257	32.596001	31.852001

FIGURE 4. File JHOUT.DAT containing output data of the program JHET in a work-session using datafiles.

REFERENCES

1. D.H. Rouvray, in "Chemical Applications of Graph Theory", Ed. A.T. Balaban, Academic Press, New York, 1976, p. 175.
2. J.J. Sylvester, Am. J. Math. 1, 64 (1874).
3. M. Barysz, D. Plavsic and N. Trinajstic, Math. Chem. 19, 89 (1986).
4. 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.
5. A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Top. Curr. Chem. 114, 21 (1983), Eds. M. Charton and I. Motoc, Springer, Berlin.
6. D.H. Rouvray, Stud. Phys. Theor. Chem. 28, 159 (1983), Ed. R.B. King, Elsevier, Amsterdam.
7. N. Trinajstic, "Chemical Graph Theory", CRC Press, Boca Raton, 1983, vol. 2, chapter 4.
8. D. Bonchev, "Theoretic Information Indices for Characterization of Chemical Structures", Research Studies Press, Chichester, 1983.
9. L.B. Kier and L.H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.
10. L.B. Kier and L.H. Hall, "Molecular Connectivity in Structure-Activity Analysis", Research Studies Press, Letchworth, 1986.
11. M.I. Stankevich, I.V. Stankevich and N.S. Zefirov, Usp. Khim. 57, 337 (1988).
12. I. Gutman, B. Ruscic, N. Trinajstic and C.F. Wilcox, Jr., J. Chem. Phys. 62, 3339 (1975).
13. M. Randic, J. Am. Chem. Soc. 97, 6609 (1975).
14. H. Hosoya, Bull. Chem. Soc. Japan 44, 2332 (1971).
15. H. Hosoya, K. Kawasaki and K. Mizutani, Bull. Chem. Soc. Japan 45, 3415 (1972).
16. H. Wiener, J. Am. Chem. Soc. 69, 17, 2636 (1947).
17. H. Wiener, J. Chem. Phys. 15, 766 (1947).
18. H. Wiener, J. Chem. Phys. 52, 425, 1082 (1948).
19. A.T. Balaban, Chem. Phys. Lett. 80, 399 (1982).
20. A.T. Balaban, Pure Appl. Chem. 55, 189 (1983).
21. A.T. Balaban and L.V. Quintas, Math. Chem. 14, 213 (1983).

22. A.T. Balaban and P. Filip, Math. Chem. 16, 163 (1984).
23. A.T. Balaban, N. Ionescu-Pallas and T.-S. Balaban, Math. Chem. 17, 121 (1985).
24. S.H. Bertz, Discrete Appl. Math. 19, 65 (1988).
25. D. Bonchev and N. Trinajstic, J. Chem. Phys. 67, 4517 (1977).
26. M. Razinger, J.R. Chretien and J.E. Dubois, J. Chem. Inf. Comput. Sci. 25, 23 (1985).
27. M. Barysz, G. Jashari, R.S. Lall, V.K. Srivastava and N. Trinajstic, Stud. Phys. Theor. Chem. 28, 222 (1983), Ed. R.B. King, Elsevier, Amsterdam.
28. S.C. Basak, D.P. Gieschen, V.R. Magnuson and D.K. Harris, ICRS Med. Sci. 10, 619 (1982).
29. A.T. Balaban, Math. Chem. 21, 115 (1986).
30. A.T. Balaban, J. Mol. Struct. (THEOCHEM), 165, 243 (188)
31. A.T. Balaban, T.-S. Balaban, P. Filip and O. Ivanciuc, to be published.
32. P.A. Filip, T.-S. Balaban and A.T. Balaban, J. Math. Chem. 1, 61 (1987).
33. A.T. Balaban, in "Graph Theory and Topology in Chemistry", Eds. R.B. King and D.H. Rouvray, Elsevier, Amsterdam, 1987, p.159.