"Vasile Alecsandri" University of Bacău
Faculty of Sciences
Scientific Studies and Research
Series Mathematics and Informatics
Vol. 19 (2009), No. 2, 59 – 70

MATRIX MATHEMATICAL MODELS USED IN THE REPRESENTATION OF MOLECULAR STRUCTURES

ZOIȚA-MĂRIOARA BERINDE

Abstract. Mathematics is very useful in chemistry, among other things, to produce models. In this paper we propose a brief description of a mathematical model of the chemical structures using matrices associated to the molecular graphs. These matrices provide a source for obtaining some important molecular descriptors that can be used in QSPR (Quantitative Structure - Property Relationship) and QSAR (Quantitative Structure – Activity Relationships). Using the notions of weighted electronic distance (w.e.d.) introduced by the author in [11], we present the weighted electronic connectivity matrix (CEP), associated to a chemical graph and also illustrate the calculation technique of ZEP index.

1. INTRODUCTION

Chemistry is one of the three fundamental natural sciences, the other two being physics and biology. Mathematics is the most economical language for formulating theories in natural sciences.

Graph theory [1] is a branch of mathematics that deals with the way objects are connected. Thus, the connectivity in a system is a fundamental quality of graph theory. Graph theory is related to matrix theory, group theory, set theory, probability, combinatorics, numerical analysis and topology.

Keywords: molecular graph; electronic weighted distance; topological index
(2000) Mathematics Subjects Classification: 92E10,05C12.

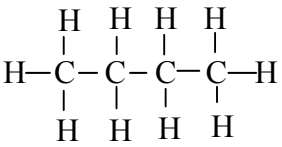
The recent years have witnessed a remarkable growth in the applications of graph theoretical principles to chemistry. In fact, graph theory serves as a mathematical model for any abstract or real chemical system involving a binary relation.

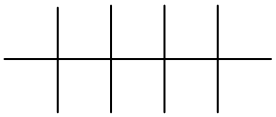
The roots of chemical graph theory, cf. [1], may be found in the work by Higgins [2], who has used first chemical graphs for representing molecules; Kopp (1842) who studied the additive properties of molecules and Crum Brown (1864) who has made use of quite a modern graphic notation for chemical compounds. Other early contributors to chemical graph theory are due to Laurent (1864) who discovered that the number of atoms with odd valency in a molecule is always even; Cayley (1874) who developed the mathematical theory of isomers (alkanes) using graphs called trees; Flavitzky (1871) who calculated the number of isomers of saturated alcohols and Sylvester (1878) who first pointed out the analogy between chemisry and algebra. Sylvester's work is especially important to us because he early expressed the belief that there is a common ground for the interchange of mathematical and chemical ideas which may lead to new development of the fields, what was the case in many instaces since his time.

2. THE GRAPH MODEL OF A MOLECULE

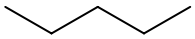
Graphs correspoding to molecules are called *chemical graphs* or *molecular graphs.* A chemical graph provides a pictorial reprsentation of the topological structure of a molecule, with vertices correspondig to individual atoms and edges depicting the valence bonds between the pairs of atoms. Chemical graphs are necessarily connected graphs.

Example

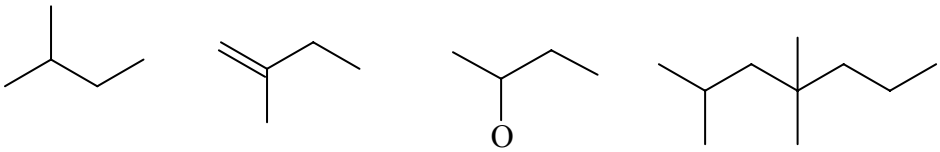




In order to simplify the handling of graphs very often are used hydrogen suppressed graphs, i.e., chemical graphs depicting only molecular skeletons without considering hydrogen atoms and their bonds.



For the following chemical compounds having the structures:
CH_3-$CH(CH_3)$-CH_2-CH_3; H_2C=$C(CH_3)$-CH_2-CH_3; H_3C-$CH(OH)$-CH_2-CH_3; CH_3-$CH(CH_3)$-CH_2-$C(CH_3)_2$-CH_2-CH_2-CH_3
and called 2-methyl-butane, 2-methyl-1-butene, 2-butanol, 2,4,4-trimethyl-heptane the corresponding (hydrogen suppressed) chemical graphs G.1 , G.2 , G.3, G.4 are the following ones.



G.1 G.2 G.3 G.4

A chemical graph enables to make useful predictions about the physical and chemical properties of molecules. To this end the chemical graph is represented by several topological matrices: adjacency matrix, distace matrix, connectivity matrix etc.[3] From this matrix one extracts several informations to produce topological indices.

3. THE TOPOLOGICAL MATRIX MODEL OF A MOLECULE

The mathematical modelling of the chemical structures using matrices associated to the molecular graphs has developed rapidly, as these matrices provide a source for obtaining some important molecular descriptors, see [3]-[4].

A topological matrix is a mathematical expression that leads itself to a computation that yields a number or a set of numbers that are characterstic of the graph from which they are derived.

The most frequently matrix used in such computations is the adjacency matrix of the graph G, $A = A(G)$. **The adjacency matrix** of a graph G = (V(G), E(G)), where V(G) denotes the set of vertices and E(G) denotes the set of edges of the graph, is a quadratic and symmetric $n \times n$ matrix, whose entries a_{ij} are defined by:

$$\begin{cases} a_{ij} = 1, \ if \ \ i \neq j \ \ and \ (i, j) \in E(G) \\ a_{ij} = 0, \ if \ \ i = j \ \ or \ (i, j) \notin E(G) \end{cases} \tag{1}$$

and the matrix A(G) will be:

$$\mathrm{A(G)} = \left\{ a_{ij}; i, j \in V(G) \right\}. \tag{2}$$

The most important tolological indices related to adjacency matrix [1] of a molecular graphs is:

a. The total adjacency index defined as

$A' = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}$, where N is the number of vertices of the graph;

b. The Zagreb Group Indices:

$M_1 = \sum_{i=1}^{N} v_i^2; \qquad M_2 = \sum_{all\ edges} \left(v_i \times v_j \right);$ where v_i is the degree of the vertex *i;*

c. The Randić connectivity indices:

$$\chi = \sum_{all\ edges} \left(v_i \times v_j \right)^{-\frac{1}{2}}; \qquad \chi_c = \sum_{paths} \left(v_i, v_j, ..., v_k \right)^{-\frac{1}{2}}$$

The models based on the adjacency have a drawback: the matrix A does not consider the multiple bonds exting into the molecule. In order to reproduce the type of the existing multiple bonds, an other matrix, that is, **the connectivity matrix C(G)** is usually considered, whose entries are defined by

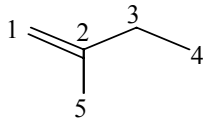
$$\begin{cases} c_{ij} = b_{ij}, \ if \ \ i \neq j \ \ and \ (i, j) \in E(G) \\ c_{ij} = 0, \ if \ \ i = j \ \ or \ (i, j) \notin E(G) \end{cases} \tag{3}$$

and the matrix C(G) is then:

$$\mathrm{C(G)} = \left\{ c_{ij}; i, j \in V(G) \right\} \tag{4}$$

Here b_{ij} denotes the conventional order of the multiple bonds: 1 for single bonds, 2 for double bonds, 3 for triple bonds and 1.5 for aromatic bonds.

Figure 1 shows the adjacency and connectivity matrices for the graph G.2 (2-methyl-1-butene):



$$C(G.2) = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}; \quad A(G.2) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

;

Fig. 1. The adjacency and connectivity matrix for 2-metyl-1-butene

Another important matrix in molecular topology is **the distance matrix** [1]**,** denoted by D or D(G). Its entries, usually denoted by d_{ij} , are equal to the number of edges connecting the vertices i and j on the shortest path between them. Thus, d_{ij} are defined as follows:

$$\begin{aligned} &d_{ij} = 0 \text{ if } i = j \\ &d_{ij} \text{ is the topological distance between } i \text{ and } j \text{ .} \end{aligned} \text{ ,} \tag{5}$$

and then distace matrix is

$$\mathrm{D(G)} = \left\{ d_{ij}; i, j \in V(G) \right\} \tag{6}$$

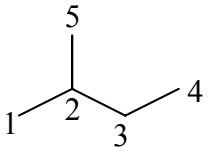
The distance marix collects a richer information than the adjacency matrix. The last one may be regarded as a special distance matrix, containing only the distances equal to the unity.

It is important to stress on the fact that the first topological index, introduced by Wiener [5] in 1947, was obtained from the distance matrix by the formula:

$$W = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}\ ; \tag{7}$$

and has been used for correlating thermodynamic properties of acyclic hydrocarbons.

Figure 2 illustrates the distance matrix of the graph G.1 (2-methyl-butane):



$$D(G.1) = \begin{bmatrix} 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 3 \\ 2 & 1 & 2 & 3 & 0 \end{bmatrix}$$

Fig. 2. The distance matrix for 2-methyl butane

Both adjacency and distance matrices in their standard definitions [1] are not able to be applied to molecular graphs containing multiple bonds and/or heteroatoms. As a consequence, distinct molecular graphs as 2-methyl-butene (G.2) and 2-butanol (G.3) are represented by the same matrices A and D as the 2-methyl-butane (G.1)

The definition of the adiacency matrix has been, consequenntly, extended [6] in order to cover all possible molecular structures. The efforts made in order to obtain similar extensions for the distance matrix has been so far limitated [1] to prticular classes of compounds. An important achievement in this direction was obtained by Balaban for multigraphs [7], Barysz et comp [8], Estrada [9] and Randić [10].

By replacing the usual topological distances from the adjacency matrix by the so called weighted electronic distances (w.e.d.), the author obtained a new **connectivity matrix, CEP** [11], [12], used for the construction of some topological indices [11], [13].

Such a topological index is the ZEP index, which was shown to well correlate to several physical properties, as for example, the boiling point (BP) and the water solubility (log(1/S) for aliphatic alcohols, the molar volume for alkanes etc. [13]-[17]

The main merit of the new concept of weighted connectivity matrix, defined in Berinde [11], is that it does reproduce all information in the molecule related to the adjacency and distance. The CEP matrix is able to detect not only the presence of multiple bonds and heteroatoms, but it is also able to reproduce the information related to the neighborhood of bonds.

For a molecule having N atoms, whose graphs is G = (V(G)), E(G)), the weighted electronic connectivity matrix CEP(G) is a symmetric and quadratic $N \times N$ matrix given by:

$$CEP(G) = \{[CEP]_{ij}; i, j \in V(G)\} \tag{8}$$

where $V(G)$ is the set of vertices of the molecular graph *G*, and $E(G)$ is the set of edges, and

$CEP_{ij} = w.e.d\ (i, j)$, if $i \neq j$ and $(i,j) \in E(G)$ and $CEP_{ij} = 0$, (9)

otherwise, where w.e.d.*(i,j)* denotes the weighted electronic distance between the atoms (vertices) *i* and *j*.

The concept of weighted electronic distance, w.e.d., was defined in [11] by the following formula:

$$w.e.d.(i, j) = \frac{1}{b_r} \cdot \frac{Z_i^{'} + Z_j^{'}}{v_i \cdot v_j} \tag{10}$$

where b_r is the bond weight with values: *1*, for single bond, *2* for double bond, *3* for triple bond and 1.5 for aromatic bond, like in Barysz and al. [8];

– v_i, v_j denote the degrees of the vertices *i* and *j*, respectively;

– Z_i' denotes the formal degree of vertex *i*, and is defined by

$$Z_i' = Z_i \cdot v_i \tag{11}$$

– Z_i denotes the order number of atom *i* in Mendeleev's periodic system.

The formal degree Z_i' represents a local vertex invariant (LOVI) in the molecular graph, while w.e.d. (i,j) represents a local edge invariant (LOEI).

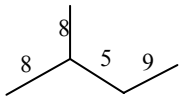
The sum of all entries on the *i*-th row in the CEP matrix is denoted by SEP_i :

$$\cdot \quad SEP_i = \sum_{j=1}^{n} [CEP]_{ij}, \ i = 1, 2, \ldots, n \tag{12}$$

The ZEP index is defined similarly to Randić connectivity indices, as the sum of root square of SEP_i :

$$ZEP = \sum_{i=1}^{n} (SEP_i)^{1/2} \tag{13}$$

In Figure 3 is given the corresponding CEP matrix for the hydrogen-deleted graph G.1 of 2-methyl-butane and is illustrated the calculation technique of ZEP index:



G.1

$$CEP(G.1) = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} SEP_i \\ \left[\begin{array}{ccccc} 0 & 8 & 0 & 0 & 0 \\ 8 & 0 & 5 & 0 & 8 \\ 0 & 5 & 0 & 9 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \end{array}\right] \begin{array}{c} 8 \\ 21 \\ 14 \\ 9 \\ 8 \end{array} \end{array}$$

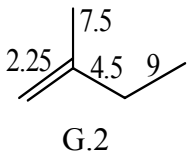
$$ZEP(G.1) = 2 \cdot 8^{\frac{1}{2}} + 21^{\frac{1}{2}} + 14^{\frac{1}{2}} + 9^{\frac{1}{2}} = 16.98$$

Figure 3. The CEP matrix for 2-methyl-butane and the calculation of ZEPindex

The *ZEP* index can be calculated direcly using the weighted electronic distance w.e.d. (i,j) on the graph without a matrix representation, considering instead of SEP_i, the sum of w.e.d.'s extended to all incident edges to i.

The w.e.d. for the graph of 2-methyl-1-butene, (G.2) its CEP matrix, and the calculation of *ZEP* index is given in Figure 4.

SEP_i,



G.2

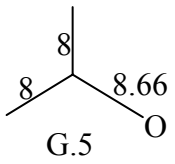
$$CEP(G.2) = \begin{bmatrix} 0 & 2.25 & 0 & 0 & 0 & 2.25 \\ 2.25 & 0 & 4.5 & 0 & 7.5 & 14.25 \\ 0 & 4.5 & 0 & 9 & 0 & 13.5 \\ 0 & 0 & 9 & 0 & 0 & 9 \\ 0 & 7.5 & 0 & 0 & 0 & 7.5 \end{bmatrix}$$

$$ZEP(G.2) = 2.25^{\frac{1}{2}} + 14.25^{\frac{1}{2}} + 13.5^{\frac{1}{2}} + 9^{\frac{1}{2}} + 7.5^{\frac{1}{2}} = 14.687$$

Figure 4. The computing of the ZEP index for graph G.2 representing the skeleton of 2 – methyl-1-butene.

Figure 5 illustrates the CEP matrix for a structure with heteroatoms:

SEP_i



G.5

$$CEP(G.5) = \begin{bmatrix} 0 & 8 & 0 & 0 & 8 \\ 8 & 0 & 8 & 8.66 & 24.66 \\ 0 & 8 & 0 & 0 & 8 \\ 0 & 8.66 & 0 & 0 & 8.66 \end{bmatrix}$$

$$ZEP(G.5) = 2 \cdot 8^{\frac{1}{2}} + 24.66^{\frac{1}{2}} + 8.66^{\frac{1}{2}} = 13.565$$

Figure 5. The CEP matrix and the calculation technique of ZEP index for the hydrogen-suppressed graph of 2-propanol (G.5)

4. CONCLUSIONS

The chemical structure of a chemical substance can be modelled using molecular graphs, topological matrices and topological index.

By replacing the usual topological distance between two atoms linked covalently with the weighted electronic distance, we obtain a new matrix representation of the molecular structure that reproduces much more accurately the connectivities between atoms.

Besides the capacity of differentiating the multiple bonds [17] and those containing heteroatoms [12], [14], w.e.d. has the merit of distinguishing even the bonds of the same type between two identical atoms, depending on their connectivities. No other system of representing the chemical bonds has, as far as we know, this power of differentiating the chemical bonds according to the context in which they are in the molecule.

REFERENCES

1. N. Trinajstić, **"Chemical Graph Theory "**, 2nd revised edition, vol. I, CRC Press, Boca Raton, Florida, 1992

2. W. A. Higgins, **"Comparative view of the phlogistic and anti-phlogistic theories"**, Murray, London, 1789

3. M. Diudea, O. Ivanciuc**, Topologie moleculară,** Comprex Cluj-Napoca, 1995 (*in Romanian*)

4. N. Trinajstić, ***"Chemical Graph Theory ",*** 2nd revised edition, vol.II, CRC Press, Boca Raton, Florida, 1992

5. H. Wiener, **Structural Determination of Paraffins Boiling Points**, *J. Amer. Chem. Soc*. 1947, 69, 17-20

6. H. D. Rouvray, **"Chemical Applications of Graph Theory"**, Edited by A. T. Balaban, Academic Press, London, 1976, Chapter 7

7. A.T. Balaban**, Applications of graph theory in chemistry,** *J. Chem*. *Inf. Comput. Sci*., 1985, *25*, 334-343

8. M. Barysz, G. Jashari, R.S. Lall, V.K. Srivastaya, N. Trinajstić**, On the Distance Matrix of Molecules Containing Heteroatoms**, in "*Chemical Applications of Graph Theory and Topology*", King, R.B. (Ed), Elsevier, Amsterdam, 1983, 222-230

9. E. Estrada**, Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A Novel Topological Index Related to Molar Volume**, *J. Chem. Inf. Comput. Sci*., 1995, *35*, 701-707
10. M. Randić**, Novel Graph Theoretical Approach to Heteroatoms in Quantitative Structure-Activity Relationships**, *Chemometrics Intel. Lab. Syst*., 1991, *10*, 213-227
11. Z. Berinde**, "Applications of molecular topology in the study of physico-chemical properties of organic compounds***",*** Cub Press 22, Baia Mare, 2001 *(in Romanian)*
12. Z. Berinde, **Consideraţii privind modelarea matricială a compuşilor halogenaţi**, *Revista de. Chimie.* (Bucureşti), 2001, *52*, No. 12, 788-792
13. Z. Berinde, **Un nou indice topologic de tip Randič**, *Revista de Chimie* (Bucureşti), 53, nr. 1, 2002, 812-816
14. Z. Berinde, **Vertex- and edge-weighted molecular graphs for amines**, *Revue Roumaine de Chimie*, 2006, *51*, No. 11, 1131-1135
15. Z. Berinde, **QSPR modelling of molar volume of alkanes using the ZEP topological index**, *Creative Math. and Inf.* 2008, 17**,** No. 3, 308-312
16. Z. Berinde, C.Drinkal, **On a local invariant for modelling molecular double bonds graphs**, *Chem. Bull. „Politehnica" Univ.* (Timişoara), vol. 49 (63), 1-2, 2004, 8-10
17. Z. Berinde, C.Drinkal, **Molecular modelling of molar refraction for alkanes**, *Chem. Bull. „Politehnica" Univ.* (Timişoara), vol. 49 (63), 1-2, 2004, 4-7

Department of Chemistry and Biology,
North University of Baia Mare, Romania
Victoriei 76 4800 Baia Mare, ROMANIA
e-mail : zoitaberinde@ubm.ro
zoita_berinde@yahoo.com