# Topological Descriptors in Similarity Studies

**Mircea V. Diudea**

**Faculty of Chemistry and Chemical Engineering**
**Babes-Bolyai University**
**400084 Cluj, ROMANIA**
**diudea@chem.ubbcluj.ro**

# Contents

# Contents

# Contents

# Contents

# 1. DEFINITIONS

## 1.1. Layer Matrix LM

A layer of vertices located at distance ***k*** to the vertex *i* is defined as: [1-4]

$$G(i)_k = \left\{ v \middle|\ v \in V(G);\quad d_{iv} = k \right\} \tag{1}$$

Define the partition of ***G*** with respect to *i* as:

$$G(i) = \left\{ G(i)_k\ ;\ k \in [0,1,..,ecc_i] \right\} \tag{2}$$

with $ecc_i$ being the *eccentricity* of *i* (*i.e.*, the largest distance from *i* to the other vertices of *G*).

# 1. DEFINITIONS

The entries in the layer matrix (of vertex property) **LM**, is defined as:

$$[\mathbf{LM}]_{i,k} = \underset{v|d_{i,v}=k}{\Omega} p_v \qquad (3)$$

with the most used operation being the summation. The zero column is just the column of vertex properties **.** Any atomic/vertex property can be considered as *pi* and any square matrix M can be taken as *info matrix*, *i.e.*, the matrix supplying local/vertex properties as *row sum RS*, *column sum CS* or *diagonal entries* given by the *Walk* matrix[3,4] .

Layer matrix is a collection of the above defined entries:

$$\mathbf{LM} = \left\{ [\mathbf{LM}]_{i,k} ; i \in V(G); k \in [0,1,..,d(G)] \right\} \qquad (4)$$

with $d(G)$ being the diameter of the graph (*i.e.*, the largest distance in $G$).

The layer matrix of vertex labeling **LLb** (see below) is the first layer matrix to be constructed. It represents just the graph partitions with respect to all its vertices.

# 1. DEFINITIONS

## 1.2 Shell Matrix SM

Define the entries in the shell/layer matrix (of pair vertex property) **SM** as: [5]

$$[\mathbf{SM}]_{i,k} = \underset{v|d_{i,v}=k}{\Omega} [\mathbf{M}]_{i,v} \qquad (5)$$

with the most used operation being the summation.

Shell matrix is a collection of the above defined entries:

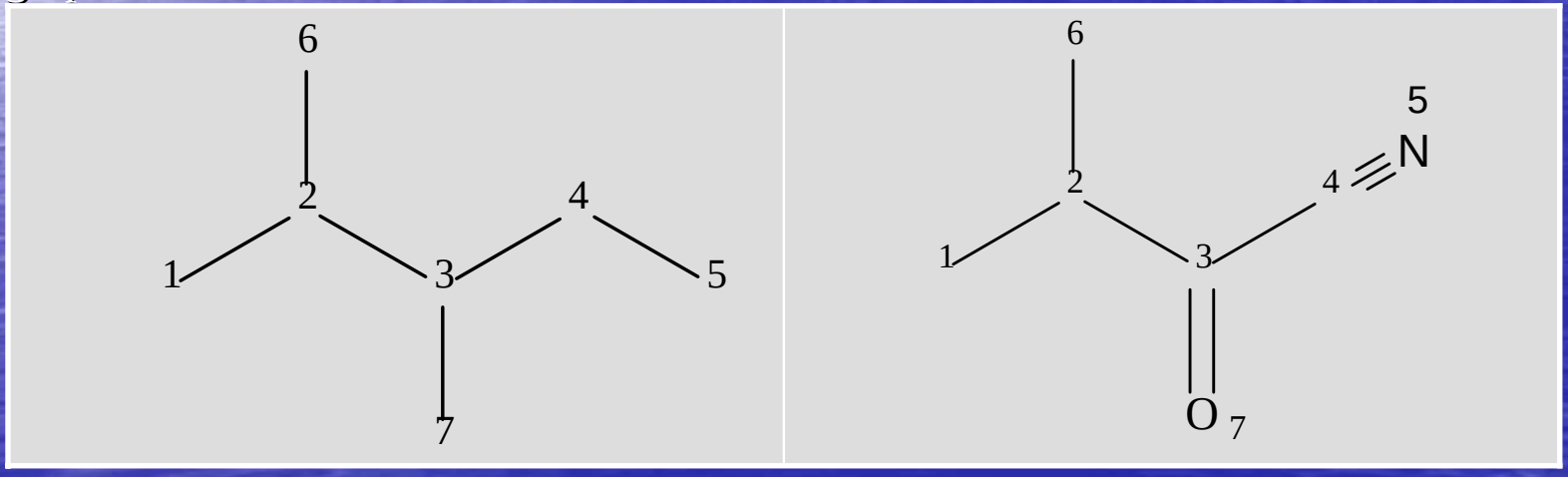
$$\mathbf{SM} = \left\{ [\mathbf{SM}]_{i,k} ; i \in V(G); k \in [0,1,..,d(G)] \right\} \qquad (6)$$

The zero column, , in case of zero diagonal square info matrix but any other vertex property (written as diagonal entries) can be considered. The above definitions hold in any graph and any square matrix.

# 1. DEFINITIONS

- **1.3. Distance Extended Properties**

- Any property can be multiplied by distance (topological or genuine one) separating vertices in the graph. The way of achieving this "extension" is different, function of matrix and operation used.
- The properties of the above defined matrices are examplified on two graphs:




$G_1$ $G_2$

- The topological descriptors are calculated by the TOPOCLUJ software package 1.0 on the following basic matrices: adjacency, connectivity, distance, 3D-distance, detour, and four types of Cluj matrices (Figure 1).

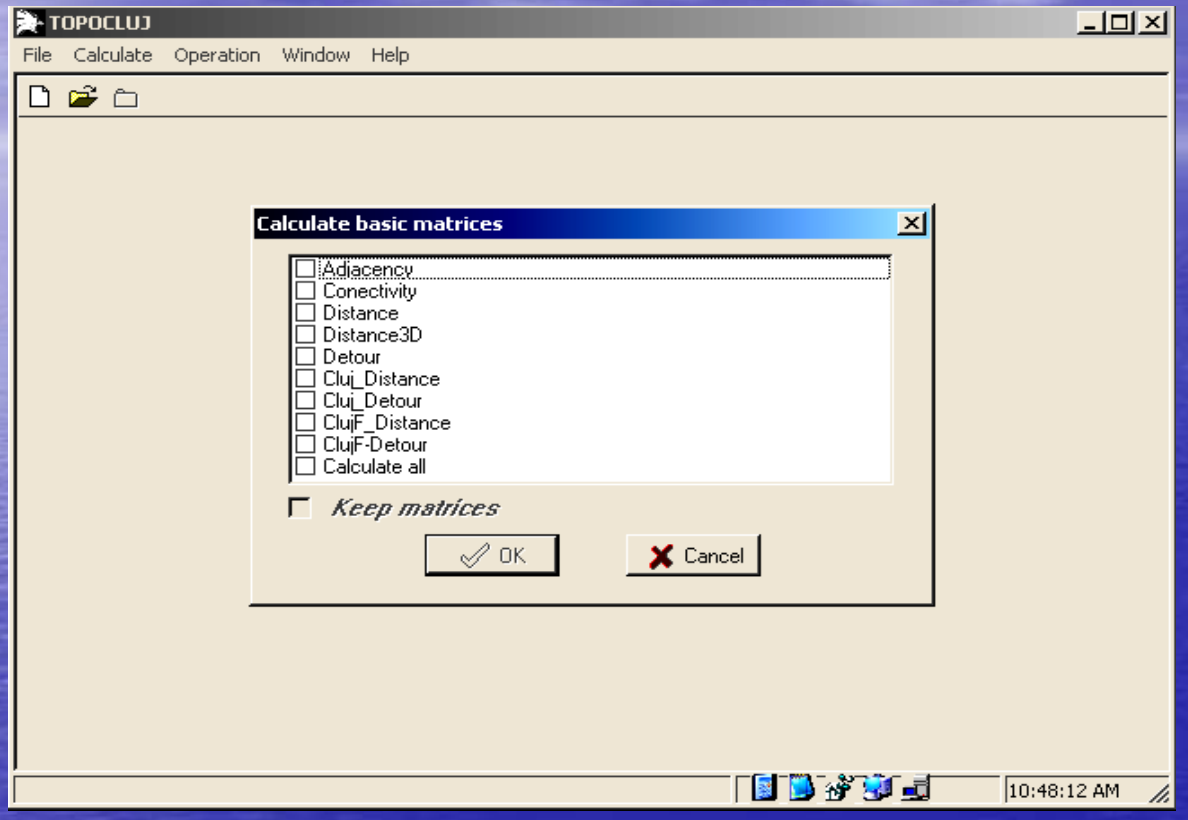



**Figure 1. Window of the basic matrices calculable by TOPOCLUJ**

# 2. PROPERTIES OF LAYER MATRICES

- ## 2.1. Layer Matrix LM Building up

- Let build up the **LLb** matrix (Table 1), by using an amended distance matrix, $\mathbf{DLb} = \mathbf{ILb} + \mathbf{D}_e$ , where **I** is the diagonal unity matrix. The partitioning of $G_1$ with respect to its vertices is obtained by collecting the labeling of vertices located at distance k from the given vertex i (cf (1)). Recall that the non-diagonal entries in the distance matrix $\mathbf{D}_e$ count the number of edges on the shortest path joining two vertices i and j, while the diagonal entries are zero.[1]

**Table 1**

| | DLb($G_1$) | | | | | | | | LLb($G_1$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | $k$ | 0 | 1 | 2 | 3 | 4 |
| 1 | *1* | 1 | 2 | 3 | 4 | 2 | 3 | | 1 | {2} | {3, 6} | {4, 7} | {5} |
| 2 | 1 | *2* | 1 | 2 | 3 | 1 | 2 | | 2 | {1, 3, 6} | {4, 7} | {5} | 0 |
| 3 | 2 | 1 | *3* | 1 | 2 | 2 | 1 | | 3 | {2, 4, 7} | {1, 5, 6} | 0 | 0 |
| 4 | 3 | 2 | 1 | *4* | 1 | 3 | 2 | | 4 | {3, 5} | {2, 7} | {1, 6} | 0 |
| 5 | 4 | 3 | 2 | 1 | *5* | 4 | 3 | | 5 | {4} | {3} | {2, 7} | {1, 6} |
| 6 | 2 | 1 | 2 | 3 | 4 | *6* | 3 | | 6 | {2} | {1, 3} | {4, 7} | {5} |
| 7 | 3 | 2 | 1 | 2 | 3 | 3 | *7* | | 7 | {3} | {2, 4} | {1, 5, 6} | 0 |

# 2. PROPERTIES OF LAYER MATRICES

- The most simple and essential is the *counting* property (*i.e.*, the existence of a vertex in a given position is counted by 1, and zero, otherwise). The corresponding layer matrix **LC** is given in Table 2. Another example is the layer matrix of partial charges **LCH**, (calculated on $G_2$).

## Table 2

| | LC($G_1$) | | | | | | LCH($G_2$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | *RS* | 0 | 1 | 2 | 3 | 4 | *RS* |
| 1 | 1 | 1 | 2 | 2 | 1 | 7 | 0.087 | 0.186 | 0.282 | -0.333 | -0.223 | 0 |
| 2 | 1 | 3 | 2 | 1 | 0 | 7 | 0.186 | 0.370 | -0.333 | -0.223 | 0 | 0 |
| 3 | 1 | 3 | 3 | 0 | 0 | 7 | 0.195 | -0.147 | -0.048 | 0 | 0 | 0 |
| 4 | 1 | 2 | 2 | 2 | 0 | 7 | 0.015 | -0.028 | -0.162 | 0.175 | 0 | 0 |
| 5 | 1 | 1 | 1 | 2 | 2 | 7 | -0.223 | 0.015 | 0.195 | -0.162 | 0.175 | 0 |
| 6 | 1 | 1 | 2 | 2 | 1 | 7 | 0.087 | 0.186 | 0.282 | -0.333 | -0.223 | 0 |
| 7 | 1 | 1 | 2 | 3 | 0 | 7 | -0.348 | 0.195 | 0.201 | -0.048 | 0 | 0 |
| *CS* | 7 | 12 | 14 | 12 | 4 | | 0 | 0.776 | 0.418 | -0.924 | -0.271 | 0 |
| *CS*·*k* | | 12 | 28 | 36 | 16 | 92[a] | | 0.776 | 0.835 | -2.771 | -1.082 | -2.241 |
| *CS*·1/*k* | | 12 | 7 | 4 | 1 | 24[b] | | 0.776 | 0.209 | -0.308 | -0.068 | 0.610 |

(a) 2x Wiener index *W*; (b) 2x Harary index *H*

# 2. PROPERTIES OF LAYER MATRICES

- Note that the rows in **LC** represent just the vertex distance degree sequences $DDS_i$ (*i.e.*, the number of vertices located at distance $k$ from $i$ )[1] while the column sums *CS* represent the graph $DDS_k$.
- **LC** is calculated when layer matrix is performed by TOPOCLUJ with no property is selected (see Figure 2).
- **LCH** is calculated by clicking on the "atomic charge" in the "Properties" second window.

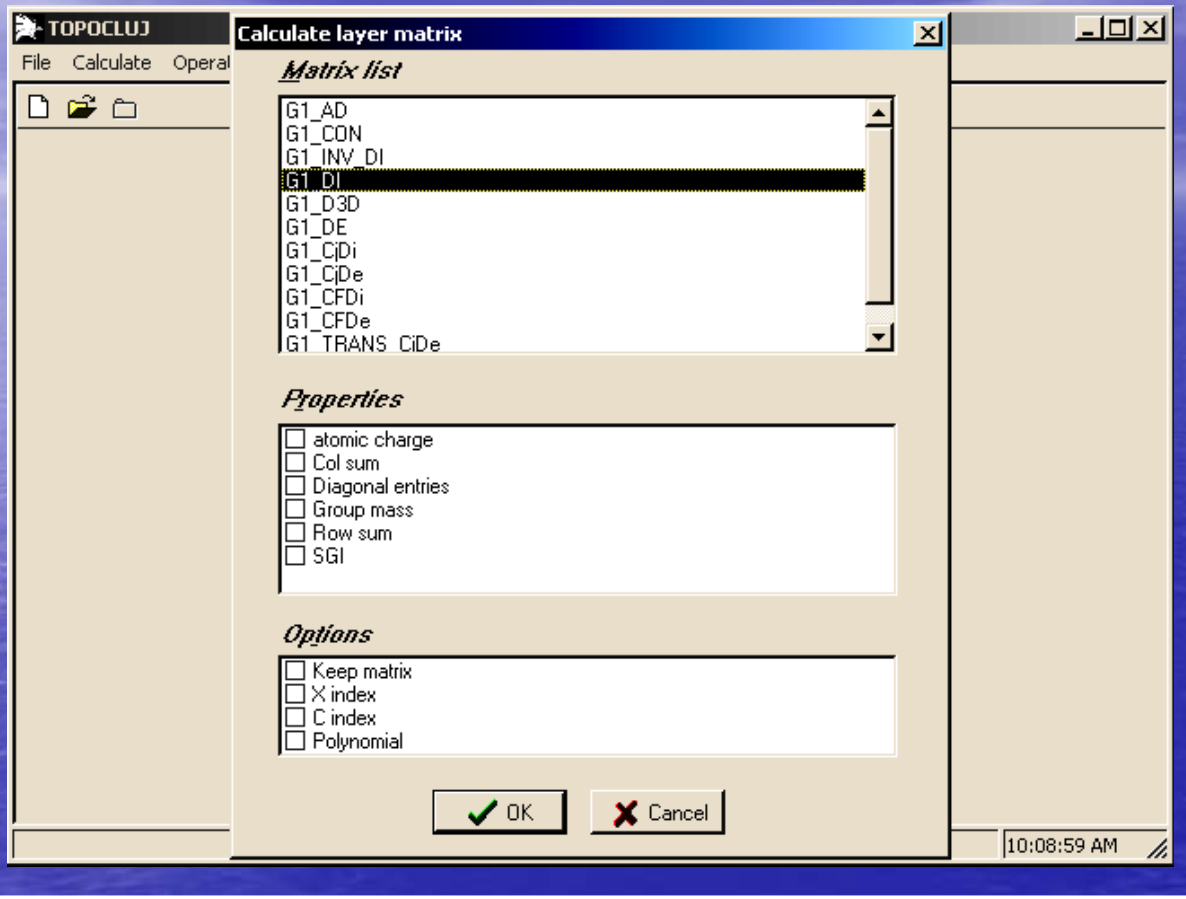



**Figure 2. Window of layer matrix calculation; properties for weighting schemes and topological indices/polynomial as options**

Within TOPOCLUJ program the partial charges are calculated as follows:

$$ch_{i,j} = \log\left[ (S_j / S_i)^{1/(d_{i,j})^2} \right] \quad (8)$$

$$ch_i = \sum_j ch_{i,j} \quad (9)$$

# 2. PROPERTIES OF LAYER MATRICES

- In the above relations, $S_i$, $S_j$ represent the Sanderson group electronegativities $SGE_s$, calculated for the hydride groups (*i.e.*, the heavy atoms with their surrounding hydrogen atoms) in the molecule.
- The log function provides the sign for the partial charge $ch_{ij}$, viewed as a distance decreasing perturbation (see also ref. 13) of the $i^{th}$ SGE produced by the atom $j$ (see the exponent, where $d_{ij}$ is the Euclidean distance separating atoms $i$ and $j$).
- The $N$ x$N$ array collecting the entries $ch_{ij}$ is the charge matrix **CH**, whose row sums $ch_i$ represent the total partial charge on hydride group/atom $i$ in the molecule (column $k$ = 0 in **LCH** - Table 2).
- Distance-extended property sum indicates, in this example, the location of the negative partial charge on the more eccentric atoms (*e.g.*, N and O).

By multiplying *CS* by *k* (or 1/*k*) distance-extended properties are obtained.

In the above example, Wiener[6-8] *W* and Harary[9-11] *H* indices are thus obtained (twice value, in Table 2).

Reciprocal matrices are calculable by the window "Mathematical operations", and next "Reciprocal".

Topological indices ***I*** are calculated as the ***half sum of the entries*** in a square matrix (*i.e.,* the matrix supplying the property *p*) :

$$I = \sum_{i<j} [\mathbf{M}]_{i,j} = (1/2) \cdot \mathbf{u} \cdot \mathbf{M} \cdot \mathbf{u}^{\mathrm{T}} \qquad (7)$$

with **u** and **uT** being the unity vector (of dimension ***N***, *i.e.*, the number of vertices in *G*) and its transpose, respectively.[12]

# 2. PROPERTIES OF LAYER MATRICES

- 2.2. LM of Symmetric Square Matrices
- Layer matrix of walk degrees $\mathbf{L}^{e}\mathbf{W}$. The walk degree (of length *e*) equals the row sum in the adjacency matrix **A** (raised at power *e*). For *e* = 1 one recovers the classical vertex degree (*i.e.*, vertex valency).[3,4]
- Entries in **A** matrix equal unity when two vertices *i* and *j* are connected by an edge and zero otherwise.
- The distance extended property is now the valency[14] and the valency-distance index was patterned by several authors[12,15-18] (see footnote, Table 3). In matrix terms, the Cramer product $\mathbf{AD}_{e}$ is calculated.

## Table 3

| | $L^1W$ ($G_1$) | | | | | | $L^2W$ ($G_1$) | | | | | | $L^3W$ ($G_1$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | *RS* | 0 | 1 | 2 | 3 | 4 | *RS* | 0 | 1 | 2 | 3 | 4 | *RS* |
| 1 | 1 | 3 | 4 | 3 | 1 | 12 | 3 | 5 | 9 | 7 | 2 | 26 | 5 | 12 | 17 | 14 | 4 | 52 |
| 2 | 3 | 5 | 3 | 1 | 0 | 12 | 5 | 12 | 7 | 2 | 0 | 26 | 12 | 22 | 14 | 4 | 0 | 52 |
| 3 | 3 | 6 | 3 | 0 | 0 | 12 | 6 | 12 | 8 | 0 | 0 | 26 | 12 | 26 | 14 | 0 | 0 | 52 |
| 4 | 2 | 4 | 4 | 2 | 0 | 12 | 4 | 8 | 8 | 6 | 0 | 26 | 8 | 16 | 18 | 10 | 0 | 52 |
| 5 | 1 | 2 | 3 | 4 | 2 | 12 | 2 | 4 | 6 | 8 | 6 | 26 | 4 | 8 | 12 | 18 | 10 | 52 |
| 6 | 1 | 3 | 4 | 3 | 1 | 12 | 3 | 5 | 9 | 7 | 2 | 26 | 5 | 12 | 17 | 14 | 4 | 52 |
| 7 | 1 | 3 | 5 | 3 | 0 | 12 | 3 | 6 | 9 | 8 | 0 | 26 | 6 | 12 | 20 | 14 | 0 | 52 |
| ***CS*** | 12 | 26 | 26 | 16 | 4 | | 26 | 52 | 56 | 38 | 10 | | 5 | 108 | 112 | 74 | 18 | |
| ***CS·k*** | | 26 | 52 | 48 | 16 | 142[a] | | 52 | 112 | 114 | 40 | 318 | | 108 | 224 | 222 | 72 | 626 |
| ***CS·1/k*** | | 26 | 13 | 5.3 | 1 | 45.3 | | 52 | 28 | 12.7 | 2.5 | 95.2 | | 108 | 56 | 24.7 | 4.5 | 193.2 |

(a) 2x*I* (*I* = Degree-Distance Index; Ivanciuc, Dobrynin, Schultz, Estrada - see text)

# 2. PROPERTIES OF LAYER MATRICES

- The walk matrix **W** (Figure 3), running the $^{e}$WM *algorithm* (Diudea, Topan and Graovac[4]), allows to elude the raising of a square matrix at a power ***e***.
- It evaluates a (topological) property of a vertex ***i***, by iterative summation of the first neighbors contributions.
- The algorithm is extended to account for general graphs (with loops and multiple bonds):

# ${}^{e}$WM *algorithm*

**M** is any square matrix, **I** is the unity diagonal matrix and
*Li* is the number of loops attached to the atom *i*.

$$\mathbf{M}+\mathbf{I}={}^{e}\mathbf{W}_{M};\ e=0 \tag{10}$$

$$[{}^{e+1}\mathbf{W}_{M}]_{ii}=2L_i[{}^{e}\mathbf{W}_{M}]_{ii}+\sum_{j\neq i}([\mathbf{M}]_{ij}[{}^{e}\mathbf{W}_{M}]_{jj}) \tag{11}$$

$$[{}^{e}\mathbf{W}_{M}]_{jj}=2L_j[{}^{e-1}\mathbf{W}_{M}]_{jj}+\sum_{k\neq j}([\mathbf{M}]_{jk}[{}^{e-1}\mathbf{W}_{M}]_{kk}) \tag{12}$$

$$[{}^{e+1}\mathbf{W}_{M}]_{ij}=[{}^{e}\mathbf{W}_{M}]_{ij}=[\mathbf{M}]_{ij} \tag{13}$$

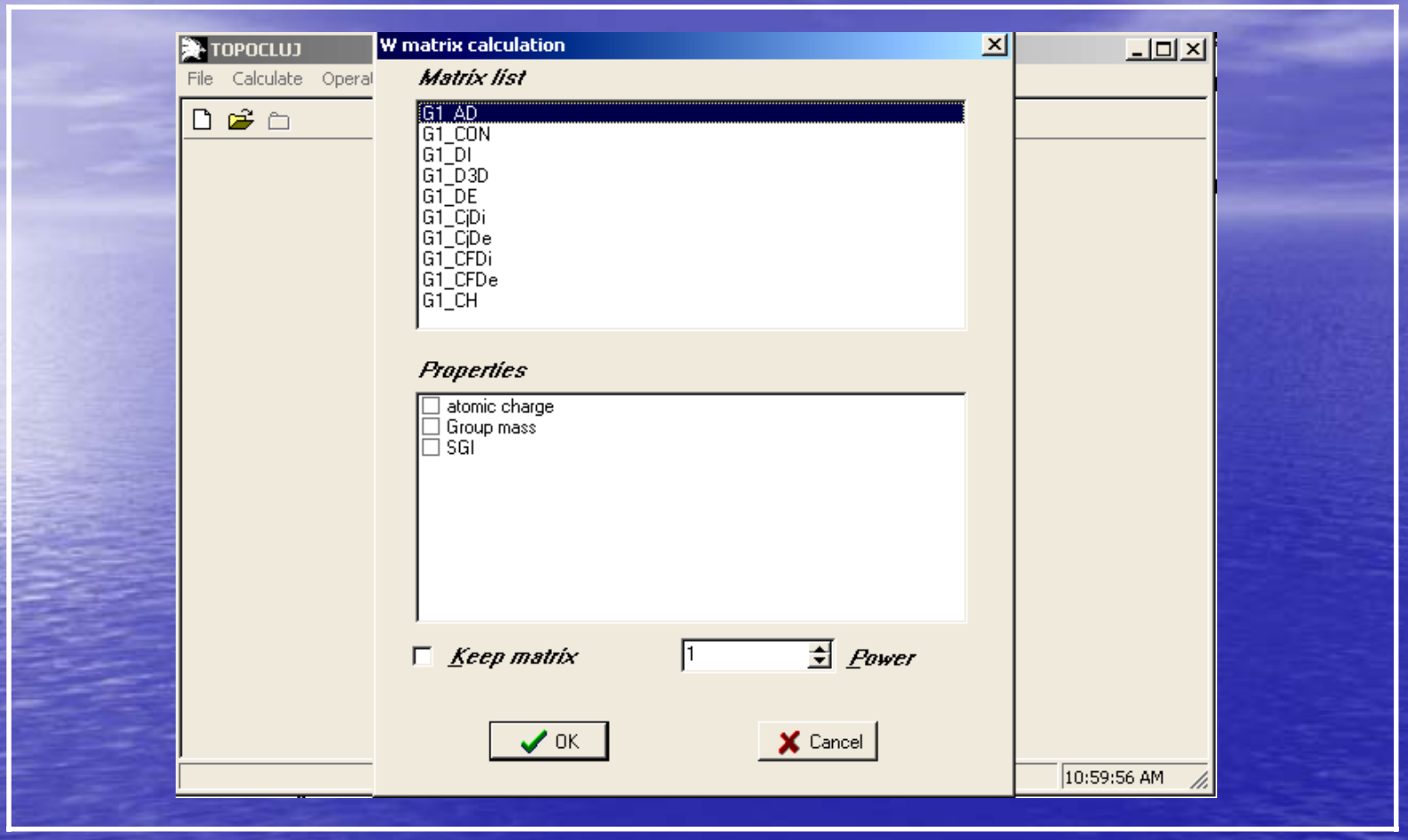



**Figure 3. Window of W matrix calculation (weighted by a chosen property**

# ${}^{e}$WM *algorithm*

The algorithm starts with the diagonal entries $[{}^{e}WM]_{ii} = 1$. In each of the following steps $[{}^{e}WM]_{ii}$ become the row sums $RS_i$ of the matrix **M** raised at a power *e*, $\mathbf{M}^{e}$:

$$[{}^{e}\mathbf{W}_{M}]_{ii} = \sum_{j} [\mathbf{M}^{e}]_{ij} = {}^{e}w_{\mathbf{M},i} \qquad (14)$$

They represent *walk degrees*, ${}^{ew}M_i$ , weighted by the property collected in **M**. The sum of all diagonal entries in ${}^{e}\mathbf{W}_{M}$ is twice the global graph invariant ${}^{e}WM$ , called *molecular walk count*:

$$\sum_{i} {}^{e}w_{M,i} = 2\,{}^{e}W_{M} = 2\,{}^{e}W_{M}(G) \qquad (15)$$

In the above, **M** = **A** or **CON**. When **M** = **D**, then ${}^{e}WM$ defines the Wiener number of rank *e* .[27]

# 2. PROPERTIES OF LAYER MATRICES

- To account for the general graphs, the algorithm needs the specification of the atoms bearing loops (Figure 4)
- . It is input as a binary numbers column under the name "aproperty aloop" in the file of properties "name.prp" provided by the TOPOCLUJ program.
- In this case, the proper matrix is the connectivity matrix **CON**, whose non-diagonal entries are just the conventional bond orders: 0, 1, 2, 3 and 1.5, for non-bonding, single, double, triple and aromatic bonds.
- If the graph has no loops, the classical $^{e}WM$ algorithm is recovered

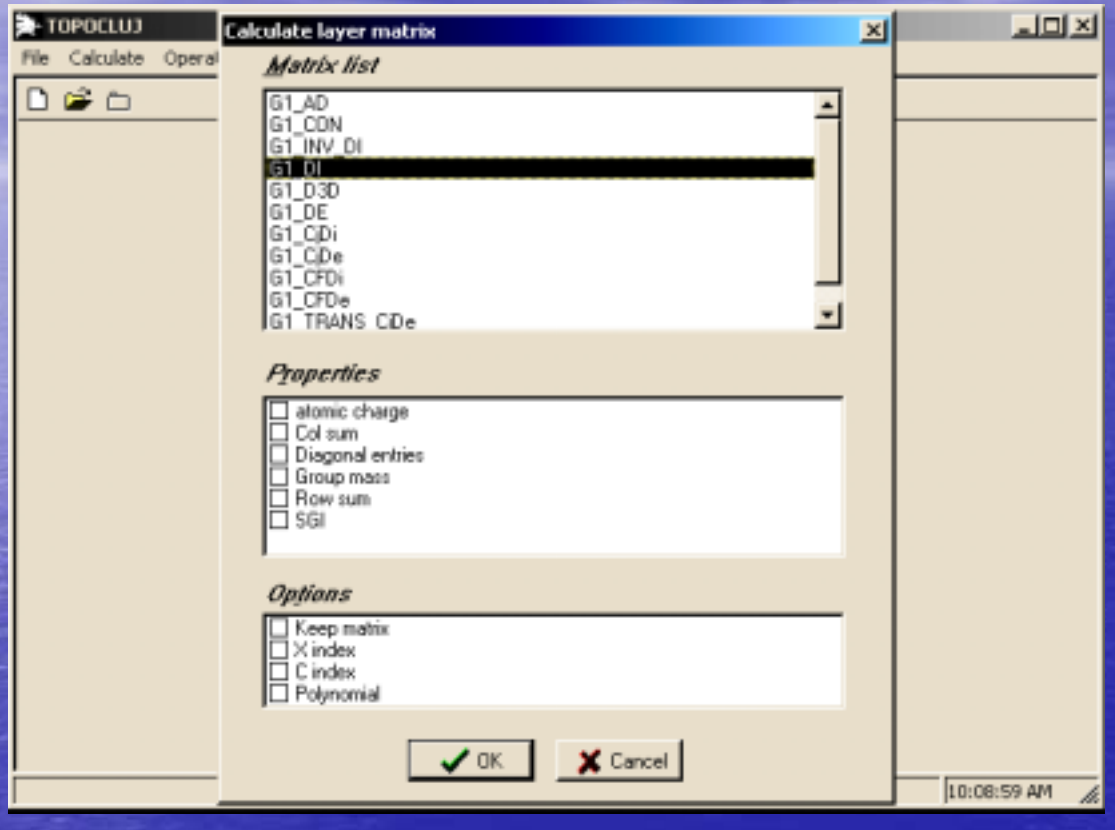



**Figure 4. Window of local properties ("aproperty", in a "name.prp" output file)**

**Examples** of ${}^{e}W_{M}$ are given in Tables 4 for $G_2$, once for no loops and next for loops at atoms 5 and 7 (representing heteroatoms with unshared electrons in the loops). The parameter $e$ is 1 and 2, respectively.

**Table 4.1. Walk Count in the general Graph $G_2$**

| | $L^1W$ (CON($G_2$)) | | | | | | | $L^1W$ (CON_loop($G_2$)) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 3 | 1 | 0 | 0 | 1 | 0 | 1 | 3 | 1 | 0 | 0 | 1 | 0 |
| 3 | 0 | 1 | 4 | 1 | 0 | 0 | 2 | 0 | 1 | 4 | 1 | 0 | 0 | 2 |
| 4 | 0 | 0 | 1 | 4 | 3 | 0 | 0 | 0 | 0 | 1 | 4 | 3 | 0 | 0 |
| **5** | 0 | 0 | 0 | 3 | **3** | 0 | 0 | 0 | 0 | 0 | 3 | **5** | 0 | 0 |
| 6 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| **7** | 0 | 0 | 2 | 0 | 0 | 0 | **2** | 0 | 0 | 2 | 0 | 0 | 0 | **4** |
| | Sum of diagonal elements = 18 | | | | | | | Sum of diagonal elements = 22 | | | | | | |

## Table 4.2

| | $L^2W(CON(G_2))$ | | | | | | | $L^2W(CON\_loop(G_2))$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 6 | 1 | 0 | 0 | 1 | 0 | 1 | 6 | 1 | 0 | 0 | 1 | 0 |
| 3 | 0 | 1 | 11 | 1 | 0 | 0 | 2 | 0 | 1 | 15 | 1 | 0 | 0 | 2 |
| 4 | 0 | 0 | 1 | 13 | 3 | 0 | 0 | 0 | 0 | 1 | 19 | 3 | 0 | 0 |
| **5** | 0 | 0 | 0 | 3 | **12** | 0 | 0 | 0 | 0 | 0 | 3 | **22** | 0 | 0 |
| 6 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 0 |
| **7** | 0 | 0 | 2 | 0 | 0 | 0 | **8** | 0 | 0 | 2 | 0 | 0 | 0 | **16** |
| | **Sum of diagonal elements = 56** | | | | | | | **Sum of diagonal elements = 84** | | | | | | |

# Distance (edge) matrix $D_e$

The property in column $k = 0$ is the row sum in the distance matrix $RS(D_e)$ and the corresponding **LM** is called *layer of distance sum* $LD_eS$.

**Table 5**

| | $LD_eS(G_1)$ | | | | | | | $D_e(G_1)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | *RS* | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 15 | 10 | 24 | 26 | 17 | 92[a] | | 0 | 1 | 2 | 3 | 4 | 2 | 3 | 15 |
| 2 | 10 | 39 | 26 | 17 | 0 | 92 | | 1 | 0 | 1 | 2 | 3 | 1 | 2 | 10 |
| 3 | 9 | 36 | 47 | 0 | 0 | 92 | | 2 | 1 | 0 | 1 | 2 | 2 | 1 | 9 |
| 4 | 12 | 26 | 24 | 30 | 0 | 92 | | 3 | 2 | 1 | 0 | 1 | 3 | 2 | 12 |
| 5 | 17 | 12 | 9 | 24 | 30 | 92 | | 4 | 3 | 2 | 1 | 0 | 4 | 3 | 17 |
| 6 | 15 | 10 | 24 | 26 | 17 | 92 | | 2 | 1 | 2 | 3 | 4 | 0 | 3 | 15 |
| 7 | 14 | 9 | 22 | 47 | 0 | 92 | | 3 | 2 | 1 | 2 | 3 | 3 | 0 | 14 |
| *CS* | 92[a] | 142 | 176 | 170 | 64 | - | *CS* | 15 | 10 | 9 | 12 | 17 | 15 | 14 | 92[a] |
| *CS·k* | | 142 | 352 | 510 | 256 | 1260[b] | | | | | | | | | |
| *CS·1/k* | | 142 | 88 | 56.67 | 16 | 302.67 | | | | | | | | | |

(a) 2x*W*; (b) $u(D_eD_e)uT$

# 2. PROPERTIES OF LAYER MATRICES

- The distance extended property is just the (topological) distance. In matrix terms, it means the Cramer product $\mathbf{D}_e\mathbf{D}_e$ (see footnote b, Table 5 and Figure 5).
- Similar calculations cam be performed on the detour matrix $\Delta_e$ (with the non-diagonal entries counting the number of edges on the longest path joining two vertices *i* and *j*, while the diagonal entries are zero).[9,28]

# Distance and Detour (path) matrices $\mathbf{D}_p$ and $\Delta_p$

**The path-defined distance and detour matrices are calculable by the "Combinatorial matrix" command (Figure 5):[9]**

$$[\mathbf{D}_p]_{ij} = \begin{cases} N_{p,(i,j)}; (i,j) \in D(G), & if \quad i \neq j \\ 0 \quad if \quad i = j \end{cases} \tag{16}$$

$$[\mathbf{\Delta}_p]_{ij} = \begin{cases} N_{p,(i,j)}; (i,j) \in \Delta(G), & if \quad i \neq j \\ 0 \quad if \quad i = j \end{cases} \tag{17}$$

$$N_{p,(i,j)} = \binom{[\mathbf{M}_e]_{ij} + 1}{2} = (1/2)\left[([\mathbf{M}_e]_{ij})^2 + [\mathbf{M}_e]_{ij}\right]; \quad \mathbf{M} = \mathbf{D}; \mathbf{\Delta} \tag{18}$$

where $N_{p,(i,j)}$ is the number of all internal paths[29] of length $1 \le |p| \le |(i,j)|$ included in the path $(i,j)$.

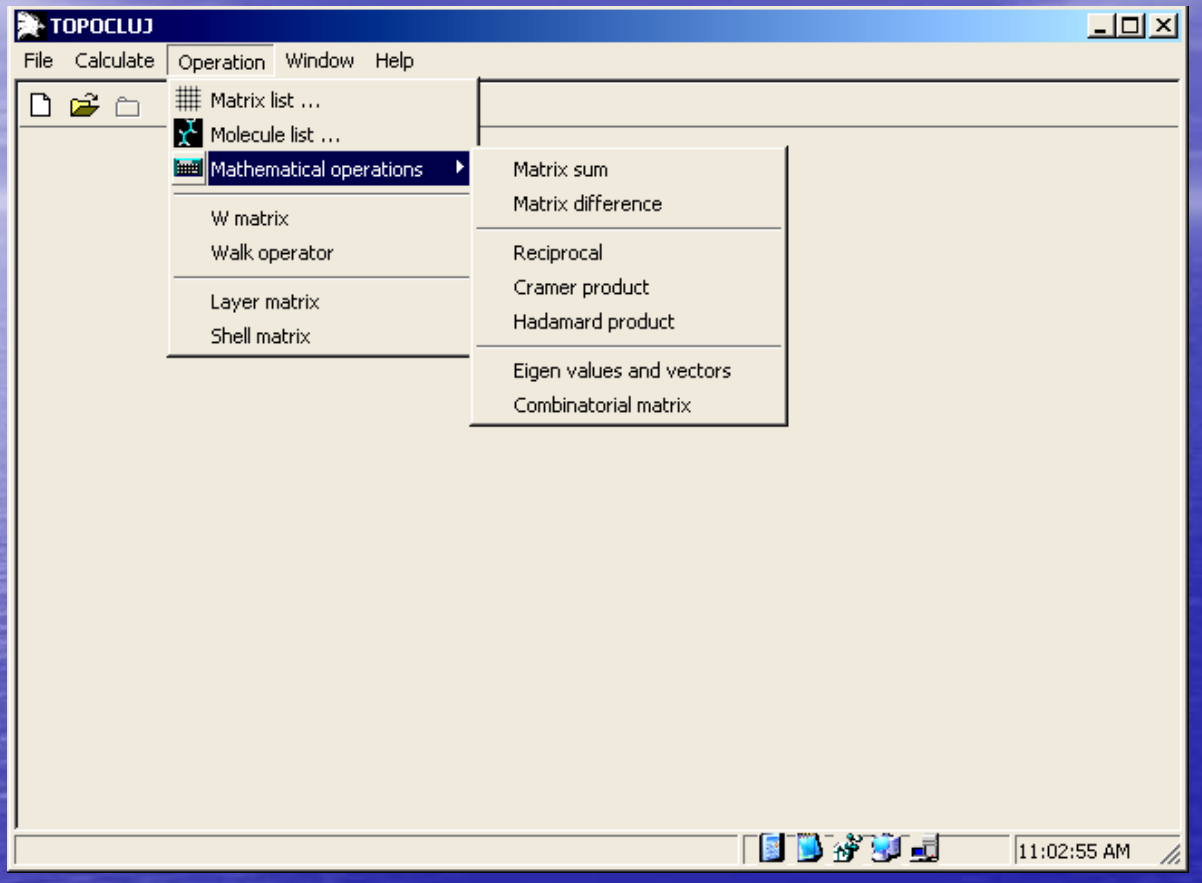



**Figure 5. Window of the basic algebraic matrix operations**

# 2. PROPERTIES OF LAYER MATRICES

## 2.3. LM of Unsymmetric Square Matrices

An unsymmetric matrix, say the product matrix $AD_e$, can participate either by its *RS* or *CS*, thus providing two different LMs (Table 6 and below).

Extension by distance is equivalent to the Cramer multiplication *to the left* by $D_e$ (of the matrix product $AD_e$ - see footnote a, Table 6).

# PROPERTIES OF LAYER MATRICES

**Table 6**

| | LAD_*RS*($G_1$) | | | | | | AD($G_1$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | *RS* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 10 | 39 | 46 | 35 | 12 | 142 | 1 | 0 | 1 | 2 | 3 | 1 | 2 | 10 |
| 2 | 39 | 56 | 35 | 12 | 0 | 142 | 4 | 3 | 4 | 7 | 10 | 4 | 7 | 39 |
| 3 | 36 | 74 | 32 | 0 | 0 | 142 | 7 | 4 | 3 | 4 | 7 | 7 | 4 | 36 |
| 4 | 26 | 48 | 48 | 20 | 0 | 142 | 6 | 4 | 2 | 2 | 2 | 6 | 4 | 26 |
| 5 | 12 | 26 | 36 | 48 | 20 | 142 | 3 | 2 | 1 | 0 | 1 | 3 | 2 | 12 |
| 6 | 10 | 39 | 46 | 35 | 12 | 142 | 1 | 0 | 1 | 2 | 3 | 1 | 2 | 10 |
| 7 | 9 | 36 | 65 | 32 | 0 | 142 | 2 | 1 | 0 | 1 | 2 | 2 | 1 | 9 |
| *CS* | 142 | 318 | 308 | 182 | 44 | | *CS* | 24 | 14 | 12 | 18 | 28 | 24 | 22 | 142 |
| *CS*·*k* | | 318 | 616 | 546 | 176 | 1656[a] | | | | | | | | |
| *CS*·1/*k* | | 318 | 154 | 60.67 | 11 | 543.67 | | | | | | | | |

(a) $\mathbf{u}(\mathbf{D}_e(\mathbf{AD}_e)\mathbf{u}^T$

# 2.3. LM of Unsymmetric Square Matrices.

## Walk operator[14,19,22] $\mathbf{W}_{(M1,M2,M3)}$

**It is defined as (see Figure 6):[19]**

$$[\mathbf{W}_{(\mathbf{M}_1,\mathbf{M}_2,\mathbf{M}_3)}]_{ij} = {}^{[\mathbf{M}_2]_{ij}} W_{\mathbf{M}_1,i}[\mathbf{M}_3]_{ij} = [RS((\mathbf{M}_1)^{[\mathbf{M}_2]_{ij}})]_i[\mathbf{M}_3]_{ij} \qquad (19)$$

where $W_{\mathbf{M}1,i}$ is the walk degree of elongation $[\mathbf{M}_2]_{ij}$, of the vertex $i$, weighted by the property collected in matrix $\mathbf{M}_1$ (*i.e.*, the $i^{th}$ row sum of the matrix $\mathbf{M}_1$, raised to power $[\mathbf{M}_2]_{ij}$). The diagonal entries are zero.

This matrix, that mixes three square matrices, is a true matrix operator (see below).

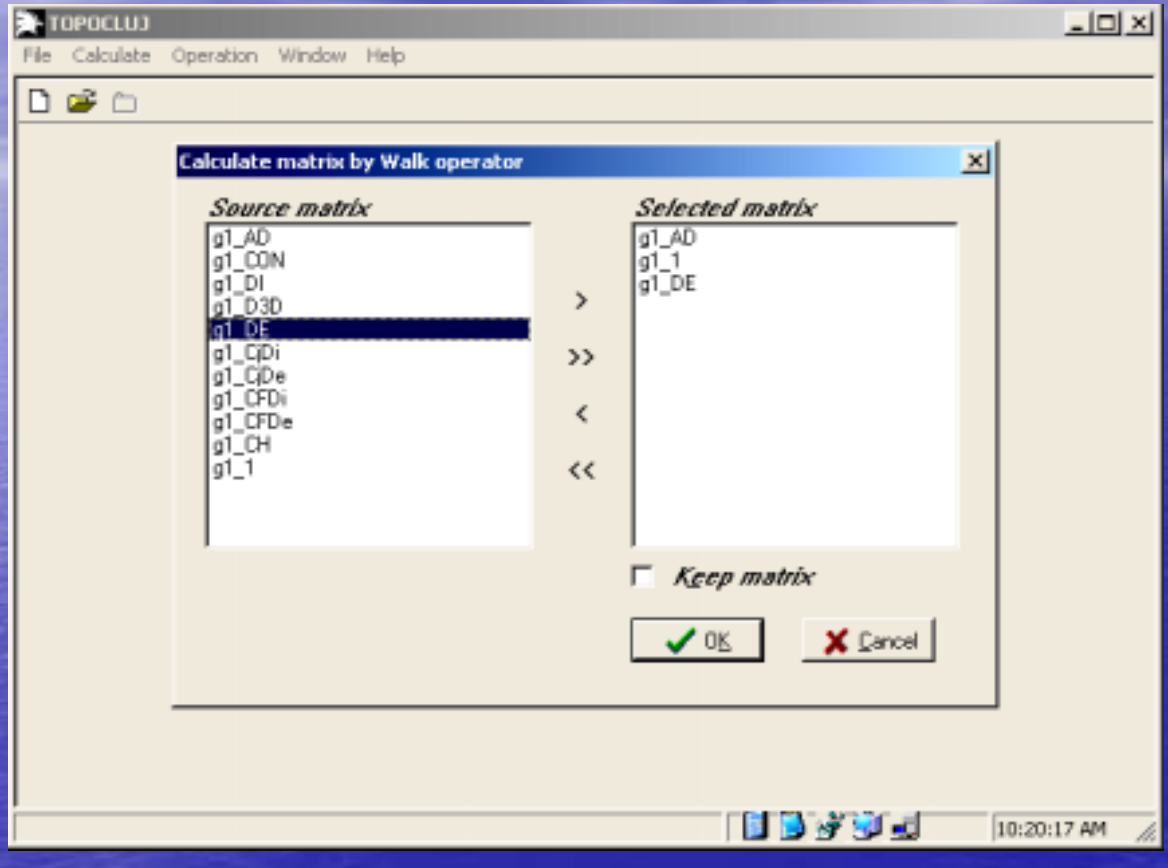



**Figure 6. Window of the Walk operator**

# 2. PROPERTIES OF LAYER MATRICES

- Within the $\mathbf{W}_{(M1,M2,M3)}$ matrix operator, the Hadamard algebra works.
- The *CS* recovers the *RS*/*CS* of the corresponding Cramer matrix product $\mathbf{M}_1\mathbf{M}_3$
- The *RS* equals the pairwise Hadamard product of the two properties collected in $\mathbf{M}_1$ and $\mathbf{M}_2$. The last result has no correspondence in the Cramer algebra.
- Examples are given for $\mathbf{M}_1 = \mathbf{A}$ and $\mathbf{M}_3 = \mathbf{D}_e$ in Tables 7 and 8 (see also footnote a, Table 3).

# $\mathbf{W}_{(M1,M2,M3)}$ matrix operator and the Hadamard algebra

In case of matrix $\mathbf{W}_{(M1,1,M3)}$, (with $\mathbf{M}_2 = \mathbf{1}$, *i.e.*, the matrix having all non-diagonal entries unity) the meaning of their vector sums is given[22] in eqs. (20) to (22).

$$CS(\mathbf{W}_{(M1,1,M3)}) = CS(\mathbf{M}_1\mathbf{M}_3) \quad (20)$$

$$CS(\mathbf{W}_{(M3,1,M1)}) = RS(\mathbf{M}_1\mathbf{M}_3) \quad (21)$$

$$RS(\mathbf{W}_{(M1,1,M3)}) = RS(\mathbf{W}_{(M3,1,M1)}) = RS(\mathbf{M}_1) \bullet RS(\mathbf{M}_3) \quad (22)$$

# Layer Matrices of Unsymmetric $W_{(M1,M2,M3)}$

**Table 7**

| | $LW_{(A,1,De)}\_RS(G_1)$ | | | | | | $W_{(A,1,De)}(G_1)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | *RS* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS*[c] |
| 1 | 15 | 30 | 42 | 38 | 17 | 142 | 0 | 1 | 2 | 3 | 4 | 2 | 3 | 15 |
| 2 | 30 | 57 | 38 | 17 | 0 | 142 | 3 | 0 | 3 | 6 | 9 | 3 | 6 | 30 |
| 3 | 27 | 68 | 47 | 0 | 0 | 142 | 6 | 3 | 0 | 3 | 6 | 6 | 3 | 27 |
| 4 | 24 | 44 | 44 | 30 | 0 | 142 | 6 | 4 | 2 | 0 | 2 | 6 | 4 | 24 |
| 5 | 17 | 24 | 27 | 44 | 30 | 142 | 4 | 3 | 2 | 1 | 0 | 4 | 3 | 17 |
| 6 | 15 | 30 | 42 | 38 | 17 | 142 | 2 | 1 | 2 | 3 | 4 | 0 | 3 | 15 |
| 7 | 14 | 27 | 54 | 47 | 0 | 142 | 3 | 2 | 1 | 2 | 3 | 3 | 0 | 14 |
| *CS* | 142 | 280 | 294 | 214 | 64 | | *CS*[b] | 24 | 14 | 12 | 18 | 28 | 24 | 22 | 142 |
| *CS*·*k* | | 280 | 588 | 642 | 256 | 1766[a] | | | | | | | | |
| *CS*·1/*k* | | 280 | 147 | 71.33 | 16 | 514.33 | | | | | | | | |

(a) $\mathbf{u}(\mathbf{D}_e(\mathbf{W}_{(A,1,De)}))\mathbf{u}T$; (b) $CS(\mathbf{W}_{(A,1,De)}) = CS(\mathbf{AD}_e)$; (c) $RS(\mathbf{W}_{(A,1,De)}) = RS(\mathbf{A}) \bullet RS(\mathbf{D}_e)$

# Layer Matrices of Unsymmetric $W_{(M1,M2,M3)}$

**Table 8**

| | $LW_{(A,1,De)}$–$CS(G_1)$ | | | | | | $W_{(De,1,A)}(G_1)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | *RS* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 24 | 14 | 36 | 40 | 28 | 142 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 15 |
| 2 | 14 | 60 | 40 | 28 | 0 | 142 | 10 | 0 | 10 | 0 | 0 | 10 | 0 | 30 |
| 3 | 12 | 54 | 76 | 0 | 0 | 142 | 0 | 9 | 0 | 9 | 0 | 0 | 9 | 27 |
| 4 | 18 | 40 | 36 | 48 | 0 | 142 | 0 | 0 | 12 | 0 | 12 | 0 | 0 | 24 |
| 5 | 28 | 18 | 12 | 36 | 48 | 142 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 17 |
| 6 | 24 | 14 | 36 | 40 | 28 | 142 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 15 |
| 7 | 22 | 12 | 32 | 76 | 0 | 142 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 14 |
| $CS^c$ | 142 | 212 | 268 | 268 | 104 | | $CS^b$ 10 | 39 | 36 | 26 | 12 | 10 | 9 | 142 |
| *CS*·*k* | | 212 | 536 | 804 | 416 | 1968[a] | | | | | | | | |
| *CS*·1/*k* | | 212 | 134 | 89.33 | 26 | 461.33 | | | | | | | | |

(a) $\mathbf{u}(\mathbf{D}^e(\mathbf{D}_e\mathbf{A})\mathbf{u}T$; (b) $CS(\mathbf{W}(\mathbf{D}_e,\mathbf{1},\mathbf{A})) = RS(\mathbf{AD}_e)$; (*c*) $CS(\mathbf{LW}(\mathbf{A},\mathbf{1},\mathbf{D}_e)) = CS(\mathbf{L}(\mathbf{AD}_e))$

# 3. SHELL MATRICES

- **3.1. SM of Basic Square Matrices**
- Let us consider the behavior of the basic symmetric adjacency **A** and distance $\mathbf{D}_e$ matrices. Since the first matrix is based on the connectivity relation, the corresponding **SM** will have only one column,

  at $k = 1$ (along with the non-related $k = 0$ column).

**SM** are calculable by TOPOCLUJ, see Figure 7.

As a general property of **SM**s , *RS* vector equals *RS* in the info matrix **M**.

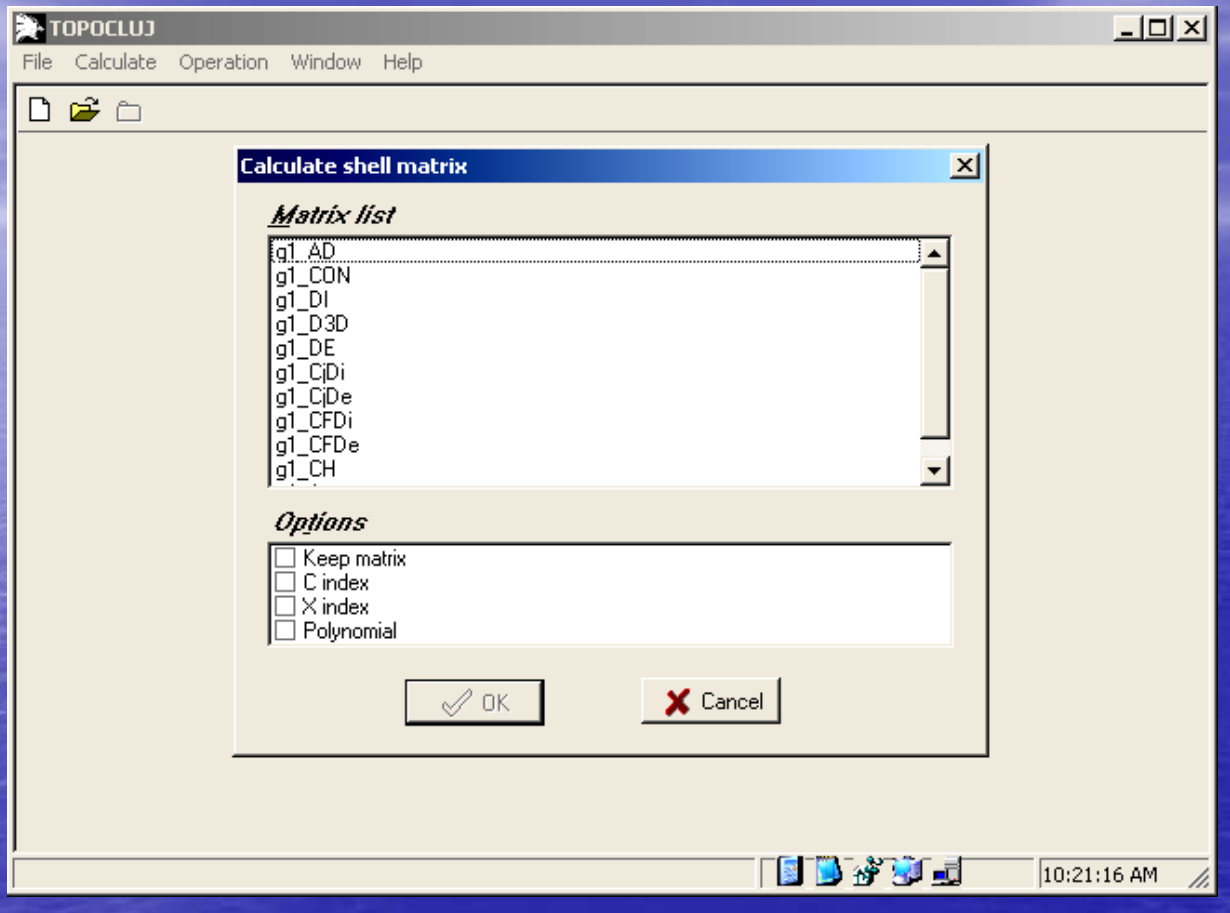



**Figure 7. Window of shell matrix calculation**

# 3.1. SM of Basic Square Matrices

## Table 9

| | $SA(G_1)$ | | | | | $\sum_{k=1}^{d(G)}$ | $SD_e(G_1)$ | | | | | $\sum_{k=1}^{d(G)}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | | 0 | 1 | 2 | 3 | 4 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 4 | 6 | 4 | 15 |
| 2 | 1 | 3 | 0 | 0 | 0 | 3 | 1 | 3 | 4 | 3 | 0 | 10 |
| 3 | 1 | 3 | 0 | 0 | 0 | 3 | 1 | 3 | 6 | 0 | 0 | 9 |
| 4 | 1 | 2 | 0 | 0 | 0 | 2 | 1 | 2 | 4 | 6 | 0 | 12 |
| 5 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 6 | 8 | 17 |
| 6 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 4 | 6 | 4 | 15 |
| 7 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 4 | 9 | 0 | 14 |
| CS | 7 | 12 | | | | 12[a] | 7 | 12 | 28 | 36 | 16 | 92[b] |
| *CS·k* | | 12 | | | | 12 | | 12 | 56 | 108 | 64 | 240[c] |
| *CS·1/k* | | 12 | | | | 12 | | 12 | 14 | 12 | 4 | 42 |

**(a) 2x *e*; (b) 2x*W* ; (c) $u(D_e \bullet D_e)u^T$**

# 3.1. SM of Basic Square Matrices

- In case of $\mathbf{SD}_e$, the distance extended property is again the distance but now the extension is equivalent to the Hadamard product $\mathbf{D}_e \cdot \mathbf{D}_e$ .
- Recall that this kind of matrix product:[23]
  $[\mathbf{M}_a \cdot \mathbf{M}_b]_{ij} = [\mathbf{M}_a]_{ij} [\mathbf{M}_b]_{ij}$ , operating within SM matrices, is basically different from the Cramer algebra working in case of **LM** matrices (see above).
- The Wiener matrices $W_e$ and $W_p$ are defined only on trees.[24,25] They are calculable as the symmetric Cluj matrices (see the next section).

# 3. SHELL MATRICES

- **3.2. SM of Cluj Matrices**
- The unsymmetric Cluj matrix **UCJ** is defined on the distance concept[26,27] (for other extensions see refs. [28-31]). In trees, its non-diagonal entries count the paths going to *j* through *i*, the number of which equals that of vertices located closer to *i* than to *j*. It is just the cardinality of the set of vertices/atoms obeying such condition;
- Since in cycle-containing graphs more than one path could join *i* and *j* (and thus providing different sets $V_{i,p(i,j)_k}$ of vertices, referred to *i*, with respect to the path $p(i,j)_k$ the entries will be the maximal cardinality value, over all the paths $p(i,j)_k$ :

# 3.2. SM of Cluj Matrices

$$V_{i,p(i,j)_k} = \{v \mid v \in V(G); d_{iv} < d_{jv}; \quad p(i,v)_h \text{ I } p(i,j)_k = \{i\}; \quad \}; h,k = 1,2,\ldots$$

(23)

$$[\mathbf{UCJ}]_{ij} = \max_{k=1,2,\ldots} \left| V_{i,p(i,j)_k} \right|$$

(24)

**The diagonal entries are zero. The Cluj matrices are defined in any connected graph and is worthy to note that they were designed as an extension of the Wiener matrices (see below).**

**Any other property (*e.g.*, a property specifying the chemical nature of vertices) and any mathematical operation (other than summation) can be considered in view of defining the entries in the Cluj matrices of property.[9,39] We limit here to the counting property, so that the entries will be integers.**

# SM of Cluj Matrices

The graph-theoretical Cluj matrices are calculable as "basic matrices" by TOPOCLUJ.

**Table 10**

| | SUCJ($G_1$) | | | | | $\sum_{k=1}^{d(G)}$ | | UCJ($G_1$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 1 | 1 | 2 | 2 | 1 | 6 | | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 6 |
| 2 | 1 | 15 | 6 | 3 | 0 | 24 | | 6 | 0 | 3 | 3 | 3 | 6 | 3 | 24 |
| 3 | 1 | 15 | 13 | 0 | 0 | 28 | | 4 | 4 | 0 | 5 | 5 | 4 | 6 | 28 |
| 4 | 1 | 8 | 4 | 4 | 0 | 16 | | 2 | 2 | 2 | 0 | 6 | 2 | 2 | 16 |
| 5 | 1 | 1 | 1 | 2 | 2 | 6 | | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 6 |
| 6 | 1 | 1 | 2 | 2 | 1 | 6 | | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 6 |
| 7 | 1 | 1 | 2 | 3 | 0 | 6 | | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 6 |
| *CS* | 7 | 42 | 30 | 16 | 4 | 92[a] | *CS* | 15 | 10 | 9 | 12 | 17 | 15 | 14 | 92 |
| *CS·k* | | 42 | 60 | 48 | 16 | 166[b] | | | | | | | | | |
| *CS·1/k* | | 42 | 15 | 5.33 | 1 | 63.33 | | | | | | | | | |

(a) 2x Wiener index *W*; (b) 2x hyper-Wiener index *WW*

# SM of Cluj Matrices

- The column sums in the shell matrix **SUCJ** represent just the distance-based decomposition of the Cluj property.
- The distance-extended property represents Cluj fragments and the involved equivalent matrix operation is **UCJ** • $\mathbf{D}_e$. Note that the Hadamard product is symmetric (*i.e.*, the same matrix product is obtained by operating both to the left and to right side).
- The extension by the reciprocal distance is important in modeling intra or intermolecular phenomena, which show decreasing values as distance between the involving partners increases.
- In trees, the distance extended Cluj fragments equal to the hyper-Wiener index contributions (see footnote b).
- Also note that the product

  **UCJ** • $3\mathbf{D}\mathbf{D}_e$

  leads to a property extended by the Euclidean distance.

# SM of Cluj Matrices

The row sum *RS* and column sum *CS* in **UCJ** are related to those in the basic matrices as:

$$RS(\mathbf{UCJ}) = RS(\mathbf{W}_e) \quad (25)$$

$$CS(\mathbf{UCJ}) = CS(\mathbf{D}_e) \quad (26)$$

The *RS* and *CS* in distance extended matrix **D_UCJ** are related to the above basic matrices, defined for all vertex pairs:

$$RS(\mathbf{D\_UCJ}) = RS(\mathbf{W}_p) \quad (27)$$

$$CS(\mathbf{D\_UCJ}) = CS(\mathbf{D}_p) \quad (28)$$

In (25) and (27) $\mathbf{W}_e$ and $\mathbf{W}_p$ represent the Wiener matrices defined on edge and path, respectively.

Matrices **D_UCJ** and $\mathbf{D}_p$ are illustrated in Table 11.

# SM of Cluj Matrices

**Table 11**

| | D_UCJ ($G_1$) | | | | | | | | D$_p$ ($G_1$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 0 | 1 | 2 | 3 | 4 | 2 | 3 | 15 | 0 | 1 | 3 | 6 | 10 | 3 | 6 | 29 |
| 2 | 6 | 0 | 3 | 6 | 9 | 6 | 6 | 36 | 1 | 0 | 1 | 3 | 6 | 1 | 3 | 15 |
| 3 | 8 | 4 | 0 | 5 | 10 | 8 | 6 | 41 | 3 | 1 | 0 | 1 | 3 | 3 | 1 | 12 |
| 4 | 6 | 4 | 2 | 0 | 6 | 6 | 4 | 28 | 6 | 3 | 1 | 0 | 1 | 6 | 3 | 20 |
| 5 | 4 | 3 | 2 | 1 | 0 | 4 | 3 | 17 | 10 | 6 | 3 | 1 | 0 | 10 | 6 | 36 |
| 6 | 2 | 1 | 2 | 3 | 4 | 0 | 3 | 15 | 3 | 1 | 3 | 6 | 10 | 0 | 6 | 29 |
| 7 | 3 | 2 | 1 | 2 | 3 | 3 | 0 | 14 | 6 | 3 | 1 | 3 | 6 | 6 | 0 | 25 |
| *CS* | 29 | 15 | 12 | 20 | 36 | 29 | 25 | 166[a] | *CS* | 29 | 15 | 12 | 20 | 36 | 29 | 25 | 166[a] |

(a) 2x hyper-Wiener index *WW*

# SM of Cluj Matrices

- Despite the different distribution of numbers (due to the reversed meaning: each entry count the paths going to *i* through *j*), the *CS* in the shell matrix **SUCJ$^T$** of Cluj matrix transpose (Table 12) are identical to those in **SUCJ** matrix (which is an expected result).

# SM of Cluj Matrices

**Table 12**

| | $S(UCJ)^T,(G_1)$ | | | | | $\sum_{k=1}^{d(G)}$ | | $(UCJ)^T\,(G_1)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* | 0 | 1 | 2 | 3 | 4 | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| ¥1*k* | 1 | 6 | 5 | 3 | 1 | 15 | | 0 | 6 | 4 | 2 | 1 | 1 | 1 | 15 |
| 2 | 1 | 6 | 3 | 1 | 0 | 10 | | 1 | 0 | 4 | 2 | 1 | 1 | 1 | 10 |
| 3 | 1 | 6 | 3 | 0 | 0 | 9 | | 1 | 3 | 0 | 2 | 1 | 1 | 1 | 9 |
| 4 | 1 | 6 | 4 | 2 | 0 | 12 | | 1 | 3 | 5 | 0 | 1 | 1 | 1 | 12 |
| 5 | 1 | 6 | 5 | 4 | 2 | 17 | | 1 | 3 | 5 | 6 | 0 | 1 | 1 | 17 |
| 6 | 1 | 6 | 5 | 3 | 1 | 15 | | 1 | 6 | 4 | 2 | 1 | 0 | 1 | 15 |
| 7 | 1 | 6 | 5 | 3 | 0 | 14 | | 1 | 3 | 6 | 2 | 1 | 1 | 0 | 14 |
| *CS* | 7 | 42 | 30 | 16 | 4 | 92[a] | *CS* | 6 | 24 | 28 | 16 | 6 | 6 | 6 | 92[a] |
| *CS·k* | | 42 | 60 | 48 | 16 | 166[b] | | | | | | | | | |
| *CS·1/k* | | 42 | 15 | 5.33 | 1 | 63.33 | | | | | | | | | |

(a) 2x*W*; (b) 2x*WW*

# Simmetric Cluj Matrices

The symmetric matrix can be obtained from **UCJ** matrix by the Hadamard product with its transpose:

$$\mathbf{CJ} = \mathbf{UCJ} \bullet \mathbf{UCJ}^{\mathbf{T}} \qquad (29)$$

In trees, the following relations hold:

$$\mathbf{CJ} = \mathbf{CJ}_p = \mathbf{W}_p \qquad (30)$$

$$\mathbf{CJ}_e = \mathbf{W}_e = \mathbf{CJ}_p \bullet \mathbf{A} \qquad (31)$$

# Simmetric Cluj Matrices

- The *CS* in the shell matrix **SCJ** represent contributions to the hyper-Wiener global index (see - Table 13 - footnote a ).
- The corresponding extension by distance is equivalent to $\mathbf{CJ}_p \bullet \mathbf{D}_e$ (and also $\mathbf{W}_p \bullet \mathbf{D}_e$).
- In cycle-containing graphs, no relation exists between the Cluj matrices and Wiener matrices (the last ones being not defined in such graphs).

# SM of Symmetric Cluj Matrices

**Table 13**

| | SCJ($G_1$) | | | | | | CJ($G_1$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | $\sum_{k=1}^{d(G)}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 1 | 6 | 5 | 3 | 1 | 15 | 0 | 6 | 4 | 2 | 1 | 1 | 1 | 15 |
| 2 | 1 | 24 | 9 | 3 | 0 | 36 | 6 | 0 | 12 | 6 | 3 | 6 | 3 | 36 |
| 3 | 1 | 28 | 13 | 0 | 0 | 41 | 4 | 12 | 0 | 10 | 5 | 4 | 6 | 41 |
| 4 | 1 | 16 | 8 | 4 | 0 | 28 | 2 | 6 | 10 | 0 | 6 | 2 | 2 | 28 |
| 5 | 1 | 6 | 5 | 4 | 2 | 17 | 1 | 3 | 5 | 6 | 0 | 1 | 1 | 17 |
| 6 | 1 | 6 | 5 | 3 | 1 | 15 | 1 | 6 | 4 | 2 | 1 | 0 | 1 | 15 |
| 7 | 1 | 6 | 5 | 3 | 0 | 14 | 1 | 3 | 6 | 2 | 1 | 1 | 0 | 14 |
| *CS* | 7 | 92 | 50 | 20 | 4 | 166[a] | *CS* | 15 | 36 | 41 | 28 | 17 | 15 | 14 | 166 |
| *CS·k* | | 92 | 100 | 60 | 16 | 268[b] | | | | | | | | |
| *CS·1/k* | | 92 | 25 | 6.67 | 1 | 124.67 | | | | | | | | |

(a) 2x hyper Wiener index *WW*; (b) 2x Tratch index *Tr*

# Shell matrix of the walk operator $W_{(A,1,De)}$

**Table 14**

| | $SW_{(A,1,De)}(G_1)$ | | | | | | $W_{(A,1,De)}(G_1)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | $\sum_{k=1}^{d(G)}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | *RS* |
| 1 | 1 | 1 | 4 | 6 | 4 | 15 | 0 | 1 | 2 | 3 | 4 | 2 | 3 | 15 |
| 2 | 1 | 9 | 12 | 9 | 0 | 30 | 3 | 0 | 3 | 6 | 9 | 3 | 6 | 30 |
| 3 | 1 | 9 | 18 | 0 | 0 | 27 | 6 | 3 | 0 | 3 | 6 | 6 | 3 | 27 |
| 4 | 1 | 4 | 8 | 12 | 0 | 24 | 6 | 4 | 2 | 0 | 2 | 6 | 4 | 24 |
| 5 | 1 | 1 | 2 | 6 | 8 | 17 | 4 | 3 | 2 | 1 | 0 | 4 | 3 | 17 |
| 6 | 1 | 1 | 4 | 6 | 4 | 15 | 2 | 1 | 2 | 3 | 4 | 0 | 3 | 15 |
| 7 | 1 | 1 | 4 | 9 | 0 | 14 | 3 | 2 | 1 | 2 | 3 | 3 | 0 | 14 |
| *CS* | 7 | 26 | 52 | 48 | 16 | 142[a] | *CS* | 24 | 14 | 12 | 18 | 28 | 24 | 22 | 142 |
| *CS·k* | | 26 | 104 | 144 | 64 | 338[b] | | | | | | | | |
| *CS·1/k* | | 26 | 26 | 16 | 4 | 72[c] | | | | | | | | |

(a) $u(AD_e)u^T$; (b) $u(A(D_e \bullet D_e))u^T$; (c) $u(A(D_e \bullet RD_e))u^T$

# 4. INDICES CALCULATED ON LAYER AND SHELL MATRICES

## 4.1. Indices of centrality, *C*(*M*); M = LM/SM:

$$C(M)_i = \left[ \sum_{k=1}^{ecc} \left( [\mathbf{M}]_{ik}^{2k} \right)^{1/(ecc)^2} \right]^{-1} \quad \textbf{(32)}$$

$$C(M) = w \sum_i C(M)_i \quad \textbf{(33)}$$

where *ecc* is the maximal distance in *G* (*i.e.*, maxmax *d*(*i*,*k*)) and *w* is a weighting factor

# INDICES CALCULATED ON LAYER AND SHELL MATRICES

## 4.1. Indices of centrocomplexity, $X(M)$; M = LM/SM:

$$X(M)_i = [\sum_{j=0}^{ecc_i} M_{ij} * 10^{-z_i}]^{-1} * t_i \qquad (34)$$

$$X(M) = w \sum_i X(M)_i \qquad (35)$$

An example given in Table 15, for the layer matrix **LCON** (with $p_i$ being the row sum *RS* in the connectivity matrix **CON** and $w = 1$).

# INDICES CALCULATED ON LAYER MATRICES

**Table 15**

**LCON_*RS*($G_2$); $w = 1$**

| *i* ¥ *k* | 0 | 1 | 2 | 3 | 4 | $C_i$ | $X_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 5 | 6 | 3 | 0.1579 | 1.3563 |
| 2 | 3 | 6 | 6 | 3 | 0 | 0.2312 | 3.6630 |
| 3 | 4 | 9 | 5 | 0 | 0 | 0.3557 | 4.9500 |
| 4 | 4 | 7 | 5 | 2 | 0 | 0.2458 | 4.7520 |
| 5 | 3 | 4 | 4 | 5 | 2 | 0.1711 | 3.4452 |
| 6 | 1 | 3 | 5 | 6 | 3 | 0.1579 | 1.3563 |
| 7 | 2 | 4 | 7 | 5 | 0 | 0.2153 | 2.4750 |
| Sum: | | | | | | 1.5349 | 21.9978 |

# 5. POLYNOMIALS CALCULATED ON LAYER MATRICES

## 5.1. Characteristic Polynomial

The characteristic polynomial of a matrix is defined (see refs. 9, 40) by relation:

$$Ch(G, M, x) = \det[x\mathbf{I} - \mathbf{M}(G)] \qquad \textbf{(36)}$$

with **I** being the unit matrix of a pertinent order and **M** is a square matrix. The polynomial roots are just the eigenvalues of the matrix **M**. The string of the decreasing values of the eigenvalues is called the spectrum of **M**.

The coefficients $a_k$ of the characteristic polynomial (**M** = **A**, in which case the matrix symbol is omitted) of order ***N*** are calculable from the graph *G* on *N* vertices:

$$Ch(G, x) = \sum_{k=0}^{N} a_k(G) \cdot x^{N-k} \qquad \textbf{(37)}$$

# POLYNOMIALS CALCULATED ON LAYER MATRICES

- Relation (37) was discovered independently by Sachs, Harary, Milić, Spialter, *etc*.[9]
- It makes use of the *Sachs graphs*, contained as subgraphs in *G*.
- More efficient are the numerical methods of linear algebra, such as the well-known recursive algorithms of Le Verier, Frame, or Fadeev.[40-42]

# POLYNOMIALS CALCULATED ON LAYER MATRICES

## 5.2. Distance Property Polynomial

A *distance property polynomial* was defined[13] as:

$$P(G,M,p,x) = \sum_{k=0}^{d(G)} p(G,M,k) \cdot x^k \qquad (38)$$

with $p(G,M,0) = P(G) = \sum_i p_i$ and $p_i$ being a vertex property.

In relation (38), $p(G,M,k)$ is twice the contribution to the global (molecular) property $P(G)$ of the vertex pairs located at distance $k$ to each other, in the graph $G$.

The summation runs from zero to $d(G)$, which is the *diameter* of $G$ or the longest distance in $G$.

# POLYNOMIALS CALCULATED ON LAYER MATRICES

- When the local property $p_i$ = 1 (*i.e.*, the vertex count), $p(G,k)$ denotes the number of pair vertices separated by distance $k$ in $G$, and the classical Hosoya polynomial[44] (more exactly twice this polynomial) is recovered. In this case, $p(G,0) = N$ , where $N$ stands for the number of vertices in the hydrogen depleted molecular graph.
- The polynomial coefficients $p(G,M,k)$ are calculable as the column sums in the layer matrices LM and SM.
- Thus, $P(G,M,p,x)$ is written as $P(G,LM,p,x)$ or $P(G,SM,p,x)$ . For example, $P(G,SUCJ,ch,x)$ reads: the polynomial of the shell of unsymmetric Cluj matrix, calculated by partial charges. When $p_i$ is 1, it can be omitted.

# POLYNOMIALS CALCULATED ON LAYER MATRICES

## 5.3. Global Descriptors from Polynomials

1. The sum of absolute values of the polynomial coefficients (see the Hosoya's Z-counting polynomial[43] and also ref. 45):

$$SumP(G,x) = \sum_k \left| a_k(G,x) \right| \qquad (40)$$

2. The maximal and minimal values of a spectrum (see the first eigenvalue of **A**, proposed by Lovasz and Pellikan[46] for characterizing the branching of a graph, and also ref. 47):

$$MaxSpP(G,x) \quad ; \quad MinSpP(G,x) \qquad (41)$$

Examples are given in Table 16.

# POLYNOMIALS CALCULATED ON LAYER MATRICES

3. A *distance-extended property* can be calculated by evaluating the *first derivative* of the polynomial, for $x = 1$:

$$P'(G,M,p,1) = \sum_{k=1}^{d(G)} k \cdot p(G,M,k) = D\_P(G) \qquad \textbf{(39)}$$

In case $p_i = 1$, is just the Wiener index.

The property $p_i$ can be taken either as a crude property (*i.e.*, the column zero in **LM**) or within some weighting scheme (*i.e.*, transformed by the sequence: **W**-operator $\mathbf{W}_{(M1,M2,M3)}$, **W(M)** matrix, **LM**/**SM**), as shown above.

Any square matrix can be used as an info matrix for the layer matrices, thus resulting an unlimited number of property polynomials.

## POLYNOMIALS CALCULATED ON LAYER MATRICES

**Table 16.**

**Distance Matrix D, Shell of Distance Matrix SD, Polynomials, Roots, MaxSp*P*(G,x) and MinSp*P*(G,x), and Sum*P*(G,x) (in bold).**

| | | | | | | | | $D_e$ | | | | | | | $SD_e$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | $x_i$ | $k$ | 0 | 1 | 2 | 3 | 4 | Real $x_i$ |
| 1 | 0 | 1 | 2 | 3 | 4 | 2 | 3 | **13.635** | | 1 | 1 | 4 | 6 | 4 | **-0.044** |
| 2 | 1 | 0 | 1 | 2 | 3 | 1 | 2 | -0.432 | | 1 | 3 | 4 | 3 | 0 | -0.044 |
| 3 | 2 | 1 | 0 | 1 | 2 | 2 | 1 | -0.665 | | 1 | 3 | 6 | 0 | 0 | -0.813 |
| 4 | 3 | 2 | 1 | 0 | 1 | 3 | 2 | -1.309 | | 1 | 2 | 4 | 6 | 0 | **-0.813** |
| 5 | 4 | 3 | 2 | 1 | 0 | 4 | 3 | -2.000 | | 1 | 1 | 2 | 6 | 8 | |
| 6 | 2 | 1 | 2 | 3 | 4 | 0 | 3 | -3.006 | | 1 | 1 | 4 | 6 | 4 | |
| 7 | 3 | 2 | 1 | 2 | 3 | 3 | 0 | **-6.223** | | 1 | 1 | 4 | 9 | 0 | |
| Sum*P*(G,x) | | | | | | | | 6041 | | 7 | 12 | 28 | 36 | 16 | 99 |

$$Ch(G,D,x) = x^7-120x^5-752x^4-1840x^3-2080x^2-1056^x-192$$

$$P(G,SD,x) = 16x^4+36x^3+28x^2+12x+7$$

# 6. ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

- *Similarity* expresses the relatedness of two molecules, with a large number if their molecular descriptions are closely related and with a number going to zero in case they are unrelated.
- *Dissimilarity* also expresses the relatedness of two molecules, but the number goes to zero when their molecular descriptions are closely related.
- A dissimilarity measure is the well-known *chemical distance*.
- Similarity and dissimilarity are both included in the more general term *proximity*. Several proximity numbers are known but we limit here to some more utilised ones.

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## 6.1 Similarity Measures

***Distance indices*** usually assume a Minkowski metric within an *m*-dimensional space:[1,40,41]

$$D(x,y) = \left[ \sum_{i=1}^{m} \left| x_i - y_i \right|^z \right]^{1/z} \qquad (42)$$

**where $x = (x_1, x_2, \ldots x_m)$ and $y = (y_1, y_2, \ldots y_m)$ are the two structures of *m* points. Such indices are extensively used owing to their geometrical interpretation:**

**-when $z = 1$, the city-block distance (Tanimoto or the Manhattan distance )**

**-when $z = 2$, the Euclidean distance results.**

# *Similarity Measures*

## Similarity indices:

**1. Carhart** index:[4]

$$SIM(A,B) = \sum_k \min(I_{Ak}, I_{Bk}) / 0.5[\sum_k I_{Ak} + \sum_k I_{Bk}] \quad \textbf{(43)}$$

where $I_{Ak}$ are numbers describing molecule A and summation runs over all the $k$ descriptors taken in work.

SIM(A,B) takes values from 0.0 (nothing in common) to 1.0 (identity).

# *Similarity Measures*

## Similarity indices:

**2. Good** index[5,6] (linear form)

$$SIM(A,B)_k = 1 - \left( \frac{|I_{Ak} - I_{Bk}|}{\max|I_{Ak}, I_{Bk}|} \right) \qquad \textbf{(44)}$$

taking values in the range [0,1].

Considering all the *nk* descriptors, the index will be:

$$SIM(A,B) = \frac{1}{w_k n} \sum_{k=1}^{n} w_k SIM(A,B)_k \qquad \textbf{(45)}$$

where *wk* is the weighting factor for the descriptor *k*.

# *Similarity Measures*

## Similarity indices:

**3. Cluj-Simil Index** (a modiffication of the Good index):

$$SIM(A,B)_k = 1 - \left( \frac{\left| I_k^2(A) - I_k^2(B) \right|}{I_k^2(A) + I_k^2(B)} \right) \qquad \textbf{(46)}$$

Comercial programs for similarity search: CONCORD (University of Texas at Austin and TRIPOS Associates), ChemModel (Chemical Design Ltd.), ALLADIN , etc.

# *Similarity Measures*

Any proximity measure is a metric if it satisfies the relations: [3]

**(1)** $D(x,y) = 0$ **for** $x = y$**;**

**(2)** $D(x,y) = D(y,x)$ and (47)

**(3)** $D(x,z) \leq D(x,y) + D(y,z)$

Similarity procedures provide partitioning of sets of molecules into disjoint subsets or clusters based on their similarity.

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## 6.2. Similarity Search by Cluj-Simil Program

- Similarity search, using topological descriptors proved to be very useful in mining large databases.
- There exist global properties (i.e., properties involving the whole molecule) that do not need the speciffication of a substructure responsible for them (in the opposition to some particular biological activities).
- Many physico-chemical properties (of industrial importance), such as: boiling point, enthalpy of formation, octane number, viscosity, etc. can be estimated by using appropriate correlation equations, calibrated on well-defined training sets.
- A property similarity procedure looks for similar molecular descriptors in lists of molecules in a database.

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

- The software **Cluj-Simil, Rel. 1.0** is written (in Delphi 4.0 language) for similarity search in large databases.

The Program has 3 modules:

- Topological index calculation (Indices)
- Similaritaty search (Simil)
- Statistical analysis (Statistics)

# Cluj-Simil Program

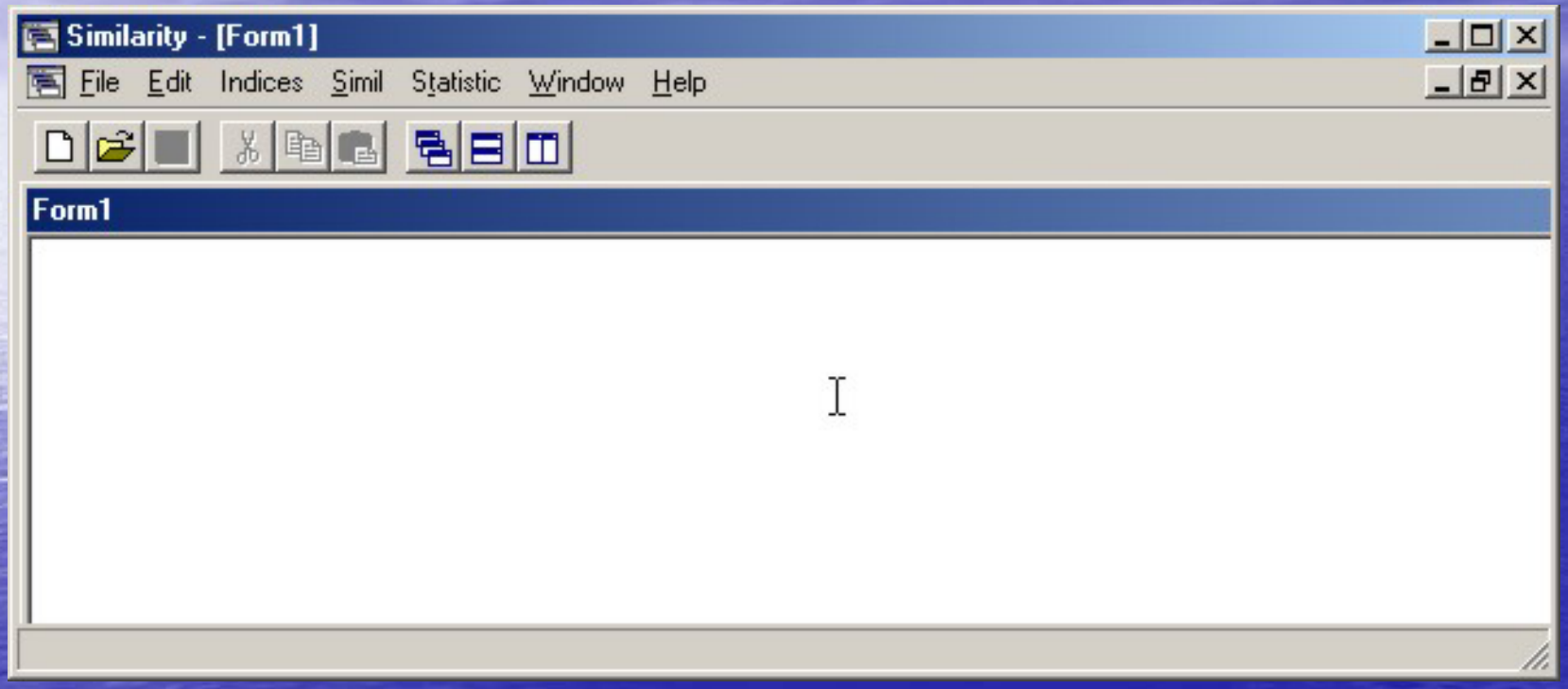
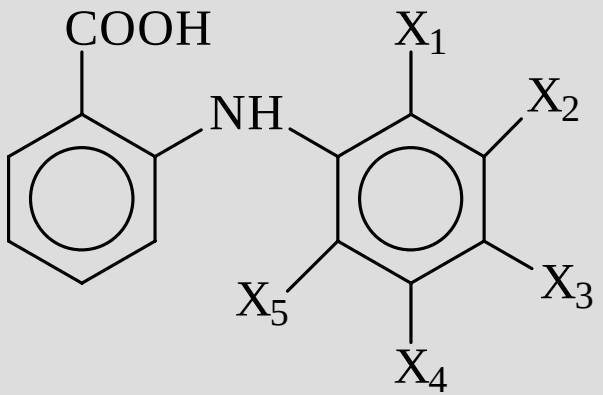



**Figure 8. The main window of the Cluj-Simil program**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

Similarity search by our program, in a set of 100 anthranylic acids ($G_3$), showing aniinflamatory activity (Table 17):[45-47]




**$G_3$**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

The biological activitaty *A* was calculated from the Minimal Effective Dose MED (mg/kgbody) by formula:

$$A = \log\left(k / MED\right); \qquad k = 4000 \qquad \textbf{(48)}$$

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| Molecule | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | MED | $A_{obs}$ |
|---|---|---|---|---|---|---|---|
| 1 | H | H | H | H | H | 200 | 1.3010 |
| 2 | H | $CF_3$ | H | H | H | 3.3 | 3.0835 |
| 3 | H | $CH_3$ | H | H | H | 100 | 1.6021 |
| 4 | H | Cl | H | H | H | 25 | 2.2041 |
| 5 | H | $NH_2$ | H | H | H | 400 | 1.0000 |
| 6 | H | $OCH_3$ | H | H | H | 50 | 1.9031 |
| 7 | H | $SO_2N(CH_3)_2$ | H | H | H | 50 | 1.9031 |
| 8 | H | $COCH_3$ | H | H | H | 200 | 1.3010 |
| 9 | H | $N(CH_3)_2$ | H | H | H | 100 | 1.6021 |
| 10 | H | H | Cl | H | H | 200 | 1.3010 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11 | H | $C_4H_9$ | H | H | H | 200 | 1.3010 |
| 12 | H | CN | H | H | H | 25 | 2.2041 |
| 13 | H | $C_3H_7$ | H | H | H | 50 | 1.9031 |
| 14 | H | $SCH_3$ | H | H | H | 100 | 1.6021 |
| 15 | H | $NO_2$ | H | H | H | 100 | 1.6021 |
| 16 | H | $OC_2H_5$ | H | H | H | 100 | 1.6021 |
| 17 | H | Br | H | H | H | 50 | 1.9031 |
| 18 | H | $C_2H_5$ | H | H | H | 25 | 2.2041 |
| 19 | Cl | H | H | H | H | 50 | 1.9031 |
| 20 | $CH_3$ | H | H | H | H | 200 | 1.3010 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 21 | H | H | $CH_3$ | H | H | 400 | 1.0000 |
| 22 | Cl | H | Cl | H | H | 100 | 1.6021 |
| 23 | H | Cl | Cl | H | H | 100 | 1.6021 |
| 24 | $CH_3$ | $CH_3$ | H | H | H | 10.4 | 2.5850 |
| 25 | $CH_3$ | $CF_3$ | H | H | H | 1 | 3.6021 |
| 26 | $CH_3$ | $SO_2N(CH_3)_2$ | H | H | H | 6.2 | 2.8097 |
| 27 | $CH_3$ | $NH_2$ | H | H | H | 50 | 1.9031 |
| 28 | $CH_3$ | $N(CH_3)_2$ | H | H | H | 6.2 | 2.8097 |
| 29 | $CH_3$ | Cl | H | H | H | 5.3 | 2.8778 |
| 30 | $CH_3$ | $OCH_3$ | H | H | H | 6.2 | 2.8097 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 31 | H | $CF_3$ | H | $CF_3$ | H | 100 | 1.6021 |
| 32 | Br | $CF_3$ | H | H | H | 1.6 | 3.3979 |
| 33 | Br | Br | H | H | H | 3.1 | 3.1107 |
| 34 | H | $CH_3$ | H | $CH_3$ | H | 100 | 1.6021 |
| 35 | Cl | H | H | H | $CH_3$ | 12.5 | 2.5051 |
| 36 | Br | CN | H | H | H | 1.5 | 3.4260 |
| 37 | F | Cl | H | H | H | 3.1 | 3.1107 |
| 38 | H | Cl | H | Cl | H | 50 | 1.9031 |
| 39 | Cl | Cl | H | H | H | 2.1 | 3.2798 |
| 40 | $CH_3$ | $NO_2$ | H | H | H | 3.1 | 3.1107 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 41 | $CH_3$ | CN | H | H | H | 3.1 | 3.1107 |
| 42 | $CH_3$ | $C_2H_5$ | H | H | H | 3.1 | 3.1107 |
| 43 | Cl | H | H | H | Cl | 3.1 | 3.1107 |
| 44 | Cl | $CH_3$ | H | H | H | 6.2 | 2.8097 |
| 45 | Cl | H | H | Cl | H | 12.5 | 2.5051 |
| 46 | $CH_3$ | H | H | H | $CH_3$ | 50 | 1.9031 |
| 47 | $CH_3$ | H | H | $CH_3$ | H | 200 | 1.3010 |
| 48 | H | $CH_3$ | $CH_3$ | H | H | 200 | 1.3010 |
| 49 | $CH_3$ | H | $CH_3$ | H | H | 400 | 1.0000 |
| 50 | $CH_3$ | $SO_2N(CH_3)_2$ | H | H | Cl | 0.7 | 3.7570 |

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## TABLE 17.
## ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 51 | Cl | Cl | H | Cl | H | 3.1 | 3.1107 |
| 52 | H | Cl | Cl | Cl | H | 200 | 1.3010 |
| 53 | $CH_3$ | $CH_3$ | H | $CH_3$ | H | 25 | 2.2041 |
| 54 | $CH_3$ | H | $CH_3$ | $CH_3$ | H | 100 | 1.6021 |
| 55 | H | Cl | $CH_3$ | Cl | H | 100 | 1.6021 |
| 56 | $CH_3$ | H | $CH_3$ | H | $CH_3$ | 400 | 1.0000 |
| 57 | Cl | $SO_2N(CH_3)_2$ | H | H | Cl | 1.3 | 3.4881 |
| 58 | Cl | $OCH_3$ | H | H | Cl | 0.3 | 4.1249 |
| 59 | $CH_3$ | Br | H | H | $CH_3$ | 1.6 | 3.3979 |
| 60 | Cl | CN | H | H | Cl | 1.6 | 3.3979 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 61 | $CH_3$ | Cl | H | H | Cl | 3.1 | 3.1107 |
| 62 | $CH_3$ | Cl | H | H | $CH_3$ | 0.4 | 4.0000 |
| 63 | Cl | $OC_2H_5$ | H | H | Cl | 0.8 | 3.6990 |
| 64 | $CH_3$ | $COCH_3$ | H | H | $CH_3$ | 0.9 | 3.6478 |
| 65 | $CH_3$ | $N(CH_3)_2$ | H | H | $CH_3$ | 1.6 | 3.3979 |
| 66 | $C_2H_5$ | $NO_2$ | H | H | $C_2H_5$ | 12.5 | 2.5051 |
| 67 | $NH_2$ | Cl | H | H | $CH_3$ | 25 | 2.2041 |
| 68 | $CH_3$ | $CH_3$ | H | Cl | H | 25 | 2.2041 |
| 69 | $CH_3$ | CN | H | H | $CH_3$ | 0.4 | 4.0000 |
| 70 | $CH_3$ | $SCH_3$ | H | H | $CH_3$ | 0.4 | 4.0000 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 71 | $CH_3$ | $NO_2$ | H | H | Cl | 1.6 | 3.3979 |
| 72 | $CH_3$ | $C_3H_7$ | H | H | $CH_3$ | 6.2 | 2.8097 |
| 73 | $C_2H_5$ | $SO_2N(CH_3)_2$ | H | H | $C_2H_5$ | 12.5 | 2.5051 |
| 74 | $C_2H_5$ | $COCH_3$ | H | H | $C_2H_5$ | 25 | 2.2041 |
| 75 | Cl | H | $CF_3$ | H | Cl | 0.8 | 3.6990 |
| 76 | $CH_3$ | $SO_2N(CH_3)_2$ | H | H | $CH_3$ | 0.5 | 3.9031 |
| 77 | $CH_3$ | $NH_2$ | H | H | Cl | 6.2 | 2.8097 |
| 78 | $CH_3$ | $CH_3$ | H | H | Cl | 12.5 | 2.5051 |
| 79 | Cl | Cl | H | H | $CH_3$ | 0.8 | 3.6990 |
| 80 | Cl | H | $C_2H_5$ | H | Cl | 0.8 | 3.6990 |

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

TABLE 17.
ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 81 | Cl | H | Cl | Cl | H | 400 | 1.0000 |
| 82 | Cl | Cl | Cl | H | H | 200 | 1.3010 |
| 83 | Cl | H | Cl | H | Cl | 100 | 1.6021 |
| 84 | $NH_2$ | $CH_3$ | H | H | $CH_3$ | 25 | 2.2041 |
| 85 | $CH_3$ | $CH_3$ | H | H | $CH_3$ | 6.2 | 2.8097 |
| 86 | Cl | $CH_3$ | H | H | $CH_3$ | 3.1 | 3.1107 |
| 87 | $CH_3$ | Cl | H | $CH_3$ | H | 1.6 | 3.3979 |
| 88 | $CH_3$ | $C_2H_5$ | H | H | $CH_3$ | 1.6 | 3.3979 |
| 89 | $CH_3$ | $NH_2$ | H | H | Cl | 1.3 | 3.4881 |
| 90 | $CH_3$ | $SO_2CH_3$ | H | H | $CH_3$ | 0.6 | 3.8239 |

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## TABLE 17.
## ANTHRANYLIC ACIDS WITH ANTIINFLAMATORY ACTIVITY

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 91 | Cl | $N(CH_3)_2$ | H | H | Cl | 0.6 | 3.8239 |
| 92 | $CH_3$ | $SOCH_3$ | H | H | $CH_3$ | 0.5 | 3.9031 |
| 93 | Cl | Cl | Cl | H | $CH_3$ | 12.5 | 2.5051 |
| 94 | $CH_3$ | $CH_3$ | H | $CH_3$ | $CH_3$ | 100 | 1.6021 |
| 95 | Cl | Cl | Cl | H | Cl | 12.5 | 2.5051 |
| 96 | Cl | $CH_3$ | Cl | H | Cl | 12.5 | 2.5051 |
| 97 | Cl | Cl | Cl | Cl | H | 100 | 1.6021 |
| 98 | Cl | Cl | H | Cl | Cl | 1.6 | 3.3979 |
| 99 | Cl | Cl | Cl | Cl | Cl | 25 | 2.2041 |
| 100 | $CH_3$ | $CH_3$ | Cl | $CH_3$ | Cl | 100 | 1.6021 |

## Table 18. Literature Data:[47]

| No. | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $A_{calc}$ | | $A_{obs}$ |
|---|---|---|---|---|---|---|---|---|
| | | | | | | P | C | |
| 1 | $CH_3$ | $NH_2$ | H | H | $CH_3$ | 2,121 | 2,784 | 1,903 |
| 2 | $CH_3$ | $OCH_3$ | H | H | Cl | 3,551 | 3,925 | 3,824 |
| 3 | $CH_3$ | $CF_3$ | H | H | $CH_3$ | 4,121 | 4,156 | 3,699 |
| 4 | $CH_3$ | $NO_2$ | H | H | $CH_3$ | 3,308 | 2,457 | 3,398 |
| 5 | $CH_3$ | Cl | H | H | $NH_2$ | 2,409 | 2,731 | 2,810 |
| 6 | Cl | CN | H | H | $CH_3$ | 3,839 | 4,759 | 3,903 |
| 7 | Cl | $N(CH_3)_2$ | H | H | $CH_3$ | 3,625 | 3,463 | 3,398 |
| 8 | $CH_3$ | Cl | H | H | Cl | 3,161 | 3,448 | 3,699 |
| 9 | Cl | Cl | H | $CH_3$ | H | 2,891 | 2,290 | 2,810 |
| 10 | $CH_3$ | Cl | H | Cl | H | 2,728 | 2,634 | 3,398 |

**The prediction quality by PRECLAV ($R^2$ = 0.7026, $r_k$ = 0,7333) was considered as "good" while those supplied by CODESSA ($R^2$ = 0,4254, $r_k$ = 0,6444) as "satisfactory".**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

- **6.3. Cluj-Simil Algorithm.**[48]

- 1. **Separate** the initial set in **two subsets:** the ***training*** set and the ***prediction*** set. (In our example, the 100 molecules were partitioned in 1:1 ratio).
- 2. **Calculate topological descriptors** for all molecules of the two sets.
- 3. **Clustering the training set** in ***m*** (here ***m*** **=** 4) *clusters*,
- with *optional leaders* and *similarity criteria*, **as follows:**

- a. leader with $A_{obs} \sim 1$ (molecule # 5)
- b. leader with $A_{obs} \sim 2$ (molecule # 6)
- c. leader with $A_{obs} \sim 3$ (molecule # 2)
- d. leader with $A_{obs} \sim 4$ (molecule #70)

- A unique chriterion was used in our example: index **DSI**.

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

- **4. Calculate regression equations** (mono or bivariate) **in the clusters of training set.**
- Here, the modeling descriptors were different from the similarity criterion: namely, Cluj indices, centrality indices and composite indices calculated by the aid of Walk Matrix operator. Some of these descriptor take into account the heteroatom and multple bond present in molecule.
- Regressions were calculated by MATLAB: STATISTIC TOOLBOX, REGRESS.
- **5. Clustering the prediction set.** The step 2 is repeated, with the same leaders and similarity criterion as used for the training set.
- **6. Predict the property *A* by the aid of regression equations, calibrated in step 4.**
- **7. Validate the algorithm** (in case the activity in the predicting set was known).
- Reslts of the application of the above algorithm are given in Tables 19-22 (only the prediction step).

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## 6.4. Study of the Case: Antiinflamatory Activity of Anthranylic Acids

***Cluster* (a); *Leader #5*.**

THE EQUATION USED FOR PREDICTION (TABLE 19) IS THAT CALIBRATED IN THE TRAINING CLUSTER (A):

$$\lg Med_{pr} = -1.4 \cdot 10^{-3} \cdot CjDiIP + 3.3 \cdot 10^{-3} \cdot CFDeIP \qquad \textbf{(49)}$$

$n = 7$; $R^2 = 0.9991$

where ***CJDiIP*** and ***CFDeIP*** are the theoretical hyper-indices ***CJDi*** and ***CFDe*** and $n$ is the number of structures in the cluster.

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

Table 19. Prediction in Cluster (a); Leader #5

| Molecule | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $A_{obs}$ | $A_{calc}$ |
|---|---|---|---|---|---|---|---|
| 21 | H | H | $CH_3$ | H | H | 1.000 | 1.207 |
| 3 | H | $CH_3$ | H | H | H | 1.602 | 1.739 |
| 37 | F | Cl | H | H | H | 3.111 | 2.825 |
| 38 | H | Cl | H | Cl | H | 1.903 | 1.866 |
| 22 | Cl | H | Cl | H | H | 1.602 | 2.211 |
| 23 | H | Cl | Cl | H | H | 1.602 | 1.540 |
| 39 | Cl | Cl | H | H | H | 3.280 | 2.825 |

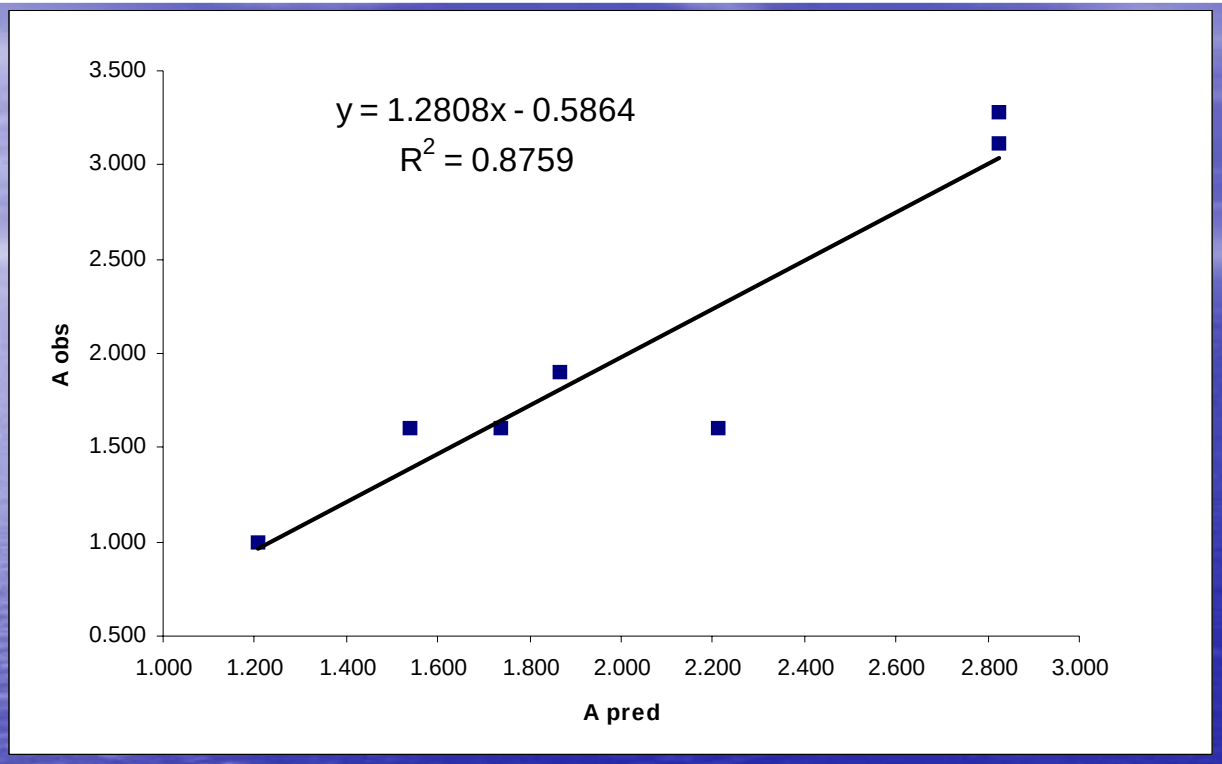



**Figure 9. The plot $A_{obs}$ *vs.* $A_{calc}$ (cf. eq. 49)**

**The quality of prediction is good (R2 = 0.8759), as can be seen in the plot $A_{obs}$ *vs.* $A_{calc}$ (Figure 9). This fortunate case, when the activity of the structures in the prediction set is known, can be considered the validation step of the algorithm.**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## Prediction step

***Cluster* (b); *Leader* #6.**

THE EQUATION USED FOR PREDICTION (TABLE 20) IS THAT CALIBRATED IN THE TRAINING CLUSTER (B):

$$\lg Med_{pr} = 69.5201 \cdot CCH - 6 \cdot 10^{-3} \cdot XCH \qquad \textbf{(50)}$$

**n = 8; $R^2$ = 0.9813**

**where *CCH* and *XCH* are the centrality and centrocomplexity indices, respectively, weighted by the partial charges of atoms or of *hydride groups*.**

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

Table 20. Prediction in Cluster (b); Leader #6

| Molecule | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $A_{obs}$ | $A_{calc}$ |
|---|---|---|---|---|---|---|---|
| 14 | H | $SCH_3$ | H | H | H | 1.602 | 1.578 |
| 27 | $CH_3$ | $NH_2$ | H | H | H | 1.903 | 1.908 |
| 47 | $CH_3$ | H | H | $CH_3$ | H | 1.301 | 1.778 |
| 9 | H | $N(CH_3)_2$ | H | H | H | 1.602 | 1.697 |
| 8 | H | $COCH_3$ | H | H | H | 1.301 | 1.564 |
| 67 | $NH_2$ | Cl | H | H | $CH_3$ | 2.204 | 2.370 |
| 68 | $CH_3$ | $CH_3$ | H | Cl | H | 2.204 | 2.117 |
| 78 | $CH_3$ | $CH_3$ | H | H | Cl | 2.505 | 2.761 |

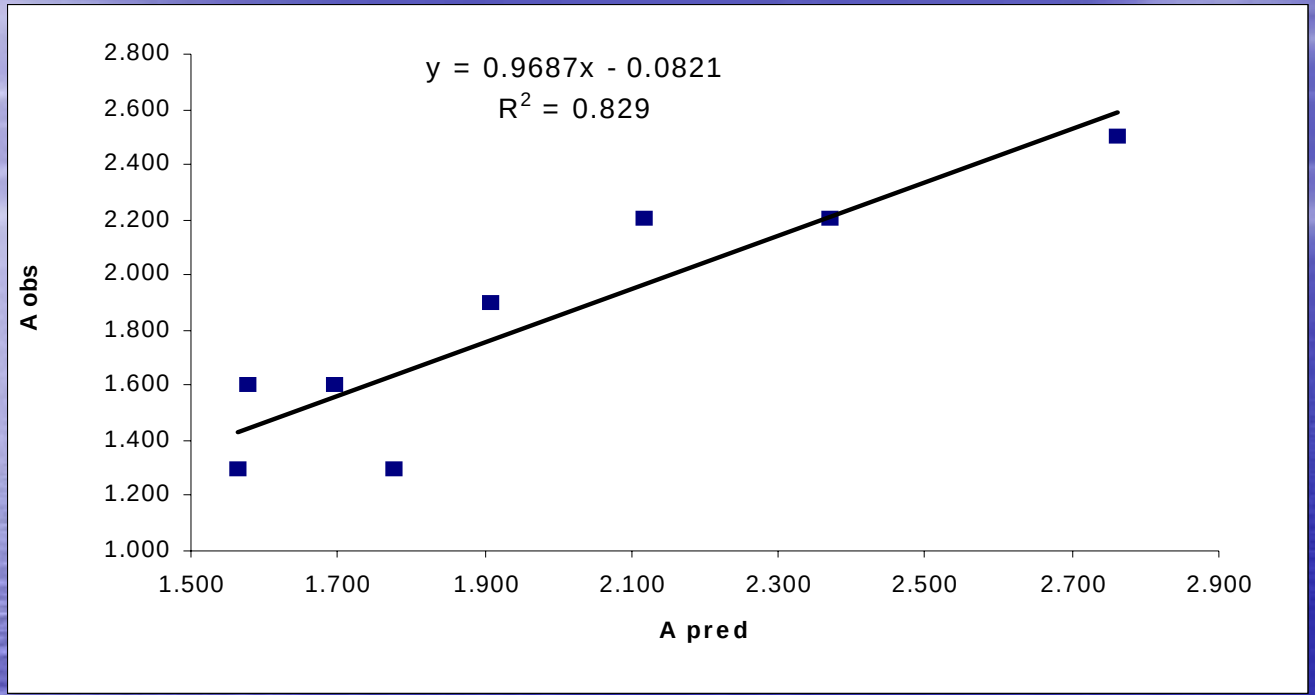



**Figure 10. The plot $A_{obs}$ *vs.* $A_{calc}$ (cf. eq. 50)**

**The quality of prediction is again good ($R^2$ = 0.829 - Figure 10).**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## Prediction step

***Cluster* (c); *Leader* #2.**

THE EQUATION USED FOR PREDICTION (TABLE 21) IS THAT CALIBRATED IN THE TRAINING CLUSTER (C):

$$\lg Med_{pr} = 3.46 \cdot 10^{-2} \cdot CjDeIE - 5.4 \cdot 10^{-3} \cdot W_{CjD,Di,A} \qquad (51)$$

$$n = 10;\ R^2 = 0.9272$$

**where *CJDeIE* is the Cluj Detour index (calculated on edges) while $W_{CJDiDiA}$ is the index obtained by the matrix operator $W_{(M1,M2,M3)}$ with $M_1$ = CJD (matrix Cluj-Distance), $M_2$ = Di (distance matrix) and $M_3$ = A (adjacency matrix).**

# Similarity by Cluj-Simil Program

Table 21. Prediction in Cluster (c); Leader #2

| Molecule | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $A_{obs}$ | $A_{calc}$ |
|---|---|---|---|---|---|---|---|
| 95 | Cl | Cl | Cl | H | Cl | 2.505 | 2.451 |
| 93 | Cl | Cl | Cl | H | $CH_3$ | 2.505 | 2.451 |
| 96 | Cl | $CH_3$ | Cl | H | Cl | 2.505 | 2.451 |
| 80 | Cl | H | $C_2H_5$ | H | Cl | 3.699 | 3.472 |
| 88 | $CH_3$ | $C_2H_5$ | H | H | $CH_3$ | 3.398 | 3.686 |
| 53 | $CH_3$ | $CH_3$ | H | $CH_3$ | H | 2.204 | 2.002 |
| 85 | $CH_3$ | $CH_3$ | H | H | $CH_3$ | 2.810 | 3.061 |
| 77 | $CH_3$ | $NH_2$ | H | H | Cl | 2.810 | 3.061 |
| 55 | H | Cl | $CH_3$ | Cl | H | 1.602 | 1.120 |
| 52 | H | Cl | Cl | Cl | H | 1.301 | 1.120 |

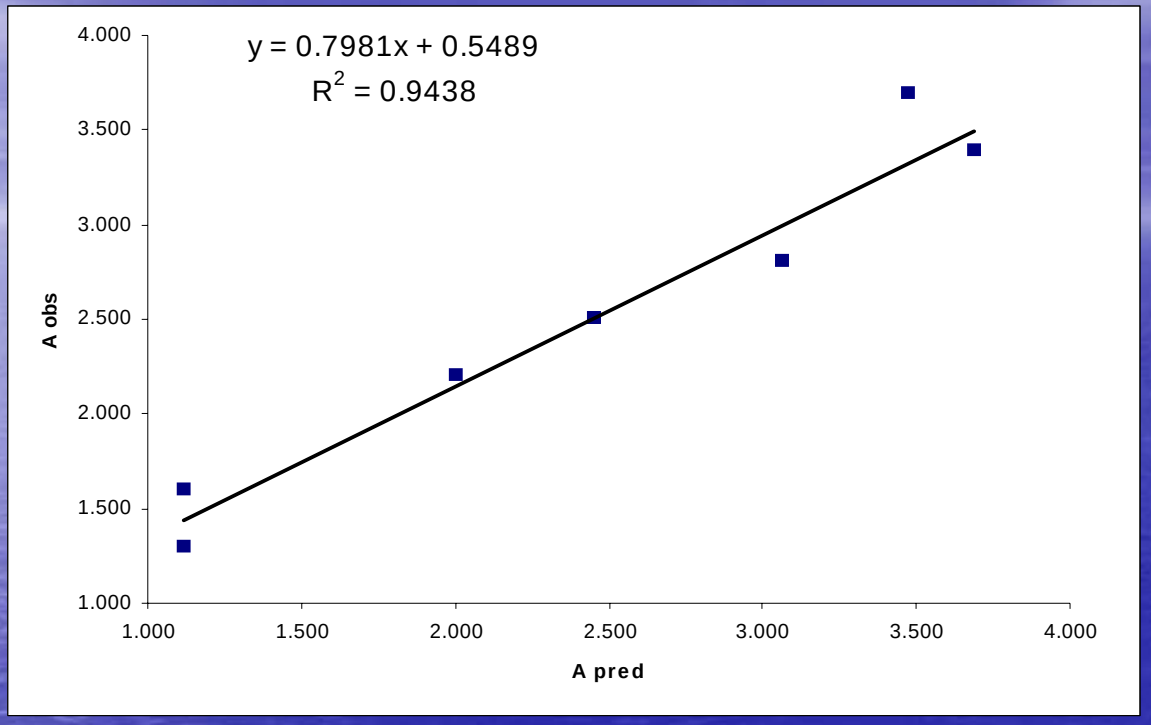



**Figure 11. The plot $A_{obs}$ *vs.* $A_{calc}$ (cf. eq. 51)**

**The quality of prediction is the best, in our experiment: $R^2 = 0.9438$ (Figure 11).**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

## Prediction step

***Cluster* (d); *Leader* #70.**

THE EQUATION USED FOR PREDICTION (TABLE 21) IS THAT CALIBRATED IN THE TRAINING CLUSTER (D):

$$\lg Med_{pr} = 3.13 \cdot 10^{-2} \cdot CjDeIE - 4.8 \cdot 10^{-3} \cdot W_{CjD,Di,A} \qquad \textbf{(52)}$$

$$\boldsymbol{n = 11;\ R^2 = 0.9596}$$

**where *CJDeIE* is the Cluj Detour index (calculated on edges) while $W_{CJDiDiA}$ is the index obtained by the matrix operator $W_{(M1,M2,M3)}$ with $M_1$ = CJD (matrix Cluj-Distance), $M_2$ = Di (distance matrix) and $M_3$ = A (adjacency matrix).**

## ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

### Table 22. Prediction in Cluster (d); Leader #70

| Molecule | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $A_{obs}$ | $A_{calc}$ |
|---|---|---|---|---|---|---|---|
| 69 | $CH_3$ | CN | H | H | $CH_3$ | 4.000 | 3.826 |
| 80 | Cl | H | $C_2H_5$ | H | Cl | 3.699 | 3.633 |
| 93 | Cl | Cl | Cl | H | $CH_3$ | 2.505 | 2.709 |
| 96 | Cl | $CH_3$ | Cl | H | Cl | 2.505 | 2.709 |
| 95 | Cl | $CH_3$ | Cl | H | Cl | 2.505 | 2.709 |
| 40 | $CH_3$ | $NO_2$ | H | H | H | 3.111 | 3.856 |
| 2 | H | $CF_3$ | H | H | H | 3.084 | 3.328 |
| 92 | $CH_3$ | $SOCH_3$ | H | H | $CH_3$ | 3.903 | 4.191 |
| 53 | $CH_3$ | $CH_3$ | H | $CH_3$ | H | 2.204 | 2.240 |
| 85 | $CH_3$ | $CH_3$ | H | H | $CH_3$ | 2.810 | 3.192 |
| 68 | $CH_3$ | $CH_3$ | H | Cl | H | 2.204 | 2.240 |

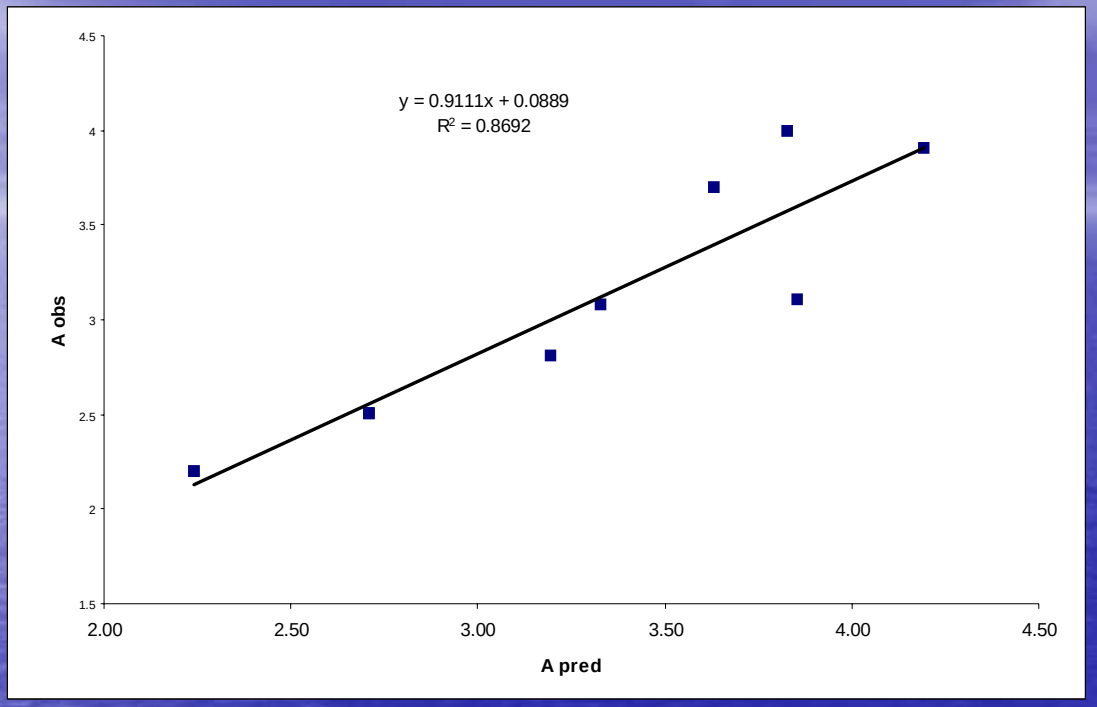



**Figure 12. The plot $A_{obs}$ *vs.* $A_{calc}$ (cf. eq. 52)**

**The quality of prediction is again good ($R^2$ = 0.8692 - Figure 12)**

# ACTIVITY PREDICTION BY CLUJ-SIMIL PROGRAM

- A reunion of the above predictions is given in Figure 13.
- The global correlation : $A_{pred}$ *vs.* $A_{obs}$ is pretty good ($R^2 = 0.8647$), in comparison to the literature data ($R^2 < 0.7100$).[47]
- Among the 50 structures in the prediction set 42 were correctly classified.

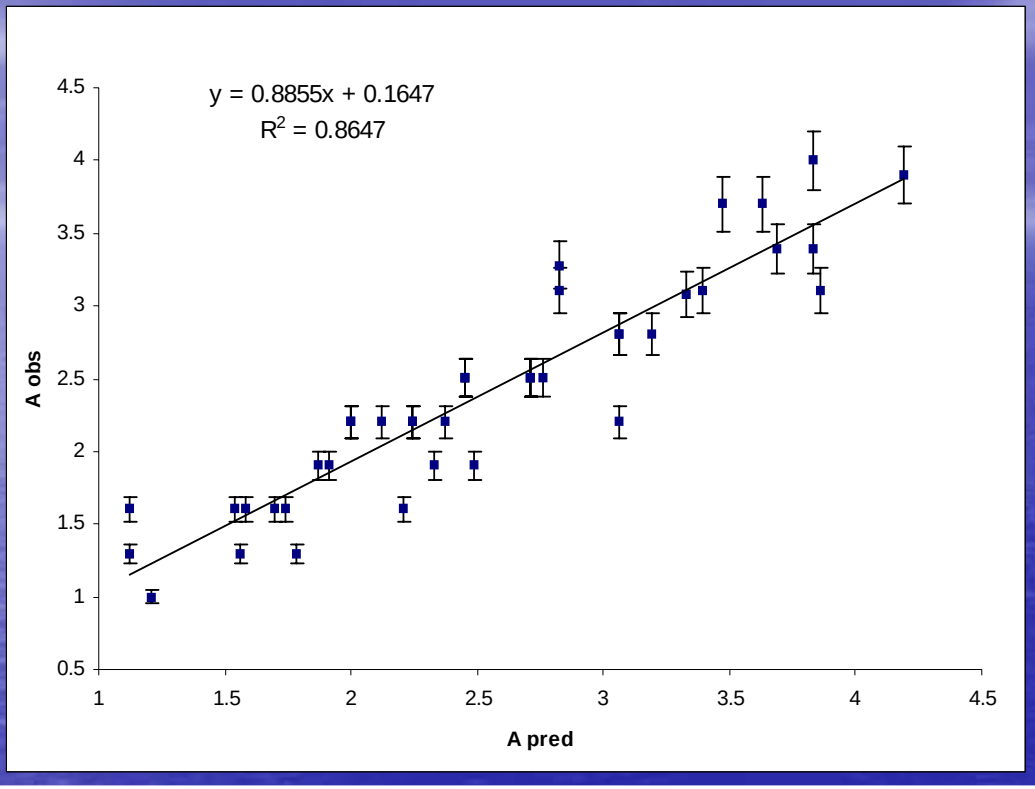



**Figure 13. The plot $A_{obs}$ *vs.* $A_{calc}$ (cf. eqs. 49-52)**

# 7. CONCLUSIONS

- Layer matrices have gained a constant interest among the chemical theorists, both in correlating studies and graph discriminating analysis. It is noteworthy that the super-index EATI, authored by Hu and Xu[49] is based on a layer matrix originating at the TOPOGROUP CLUJ.
- The shell matrix proposed by Randić[50] in 2001 is a ***LeW*** matrix, published by us[3,4] since 1994. We recognize here the priority of the Novosibirsk group, guided by Skorobogatov.[51]
- Our aim was to contribute, along with the well-known commercial programs: CODESSA,[52] MOLCONN Z,[53] DRAGON,[54] TOSS MODE[55] or POLLY,[56] by TOPOCLUJ software package[2] (which comes from an experience of two decades of our group)[1,57] to the increase of use of the topological descriptors in modeling physico-chemical and biological properties.

# 7. CONCLUSIONS

- Description of molecules by the aid of topological indices: DSI, Cluj (CjDi, CjDe, CFDi, CFDe), $W_{(M1,M2,M3)}$, indices of centricity and centrocomplexity (weighted by partial charges of atoms and hydride groups) is adequate for the main purpose: the similarity search and QSAR.
- Clustering by the criterion: DSI (that weights the atom valency by group electronegativities of Sanderson type) is in agreement with the chemical percepts, in the set of Anthranylic acids.
- Ecuations calibrated within the training set showed good statistics, that recommend them for the prediction set.

# 7. CONCLUSIONS

- **The predicted values** for the biological activity of the structures in the prediction set are **pertinent**: the plot $A_{pred}$ ***vs.*** $A_{obs}$ showed a correlation: $R^2 = 0.8647$, which overpassed the results reported in literature.[9] **Among the 50 structures in the prediction set 42 were correctly classified.**
- Our procedure can be useful in **Combinatorial Chemistry**, for fast clustering, according to the similarity vs. selected leaders, and predicting (non-measured) molecular properties.

# 8. REFERENCES


- 1. M.V. Diudea; I. Gutman; L. Jäntschi, *Molecular topology*. NOVA, Huntington, New York, 2001.
- 2. M. V. Diudea, and O. Ursu, *TOPOCLUJ* Copyright Babes-Bolyai Univ. Cluj , 2002.
- 3. M. V. Diudea, M. V. Layer matrices in molecular graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1064-1071.
- 4. M.V. Diudea, M. I. Topan, and A. Graovac, Layer matrices of walk degrees. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1072-1078.
- 5. M. V. Diudea, Cluj polynomials. *Commun, Math. Chem*. (*MATCH*) (in press).
- 6. H. Wiener, Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
- 7. M. V. Diudea and I. Gutman, Wiener-type topological indices. *Croat. Chem. Acta,* **1998**, *71*, 21-51.
- 8. M. V. Diudea, Wiener and hyper-Wiener numbers in a single matrix. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 833-836.
- 9. O. Ivanciuc, T. S. Balaban, and A. T. Balaban, Reciprocal distance matrix, related local vertex invariants and topological indices, *J. Math Chem.* **1993**, *12*, 309-318.
- 10. D. Plavšić, S. Nikolić, N. Trinajstić, and Z. Mihalić, On the Harary index for the characterization of chemical graphs, *J. Math. Chem.* **1993**, *12*, 235-250.

- 11. M. V. Diudea, Indices of reciprocal property or Harary indices, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 292-299.
- 12. E. Estrada, L. Rodriguez, A. Gutierrez, Matrix algebraic manipulation of molecular graphs.1. Distance and vertex -adjacency matrices. *MATCH, Commun, Math. Chem*. **1997**, *35*, 145-156.
- 13. L. B. Kier and L. H. Hall, *Molecular structure description: the electrotopological state*. Acad. Press, San Diego 1999.
- 14. M. V. Diudea, Valencies of property, *Croat. Chem. Acta*, **1999**, *72*, 835-851.
- 15. H. P. Schultz, Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 227-228.
- 16. O. Ivanciuc, Design of Topological Indices. Part 14. Distance-valency matrices and structural descriptors for vertex-and edge-weighted molecular graphs. *Rev. Roum.. Chim.,* **2000**, *45*, 587-596.
- 17. A. Dobrynin, A. A. Kochetova, Degree distance of a graph: a degree analogue of the Wiener index. *J. Chem. Inf. Comput. Sci*. **1994**, **34**, 1082-1086.
- 18. E. Estrada, L. Rodriguez, *MATCH, Commun, Math. Chem*. , **35** (1997) 157-167.
- 19. M. V. Diudea, Walk numbers *e*W*M* : Wiener numbers of higher rank. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 535-540.
- 20. F. Harary , *Graph theory*, Addison−Wesley, Reading, M.A., 1969.

- 21. D. J. Klein, I. Lukovits, and I. Gutman, On the definition of the hyper-Wiener index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50-52.
- 22. M. V. Diudea and M. Randić, Matrix operator, W(M1,M2,M3) and Schultz-type numbers. *J. Chem. Inf. Comput. Sci.* 1997, *37*, 1095-1100.
- 23. R. A. Horn and C. R. Johnson, *Matrix analysis*; Cambridge Univ. Press, Cambridge, **1985**.
- 24. M. Randić, X. Guo, T. Oxley, and H. Krishnapriyan, Wiener matrix: source of novel graph invariants, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 700-716.
- 25. M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, and L. Naylor, Wiener matrix invariants, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361-367.
- 26. M. V. Diudea, Cluj matrix invariants. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 300-305.
- 27. M. V. Diudea, Cluj Matrix **CJu** : source of various graph descriptors, *Commun. Math. Comput. Chem.* (*MATCH*), **1997**, *35*, 169-183.
- 28. M. V. Diudea, B. Parv, and I. Gutman, Detour-Cluj matrix and derived invariants. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1101-1108.
- 29. I. Gutman and M. V. Diudea, Defining Cluj matrices and Cluj matrix invariants. *J. Serb. Chem. Soc*. **1998**, *63*, 497-504.
- 30. M. V. Diudea, G. Katona, I. Lukovits, and N. Trinajstić, Detour and Cluj-detour indices. *Croat. Chem. Acta*, **1998**, *71*, 459-471.

- 31. L. Jäntschi, G. Katona, and M. V. Diudea, Modeling molecular properties by Cluj indices. *MATCH, Commun. Math. Comput. Chem.*, **2000**, *41*, 151-188.
- 32. M. V. Diudea and O. Ivanciuc, *Molecular topology*, COMPREX, Cluj, 1995.
- 33. P. S. Dwyes, *Linear computations*, Wiley, N. Y. 1951.
- 34. D. K. Fadeev and I. S. Sominskii, *Problems in higher algebra*, Freeman, San Francisco, 1965.
- 35. H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Japan*, **1971**, *44*, 2332-2339.
- 36. H. Hosoya, On some counting polynomials in chemistry. *Discrete Appl. Math.*, 1988, *19*, 239-257.
- 37. O. Ivanciuc, Design of topological indices. Part 18. Modeling the physical properties of alkanes with molecular graph descriptors derived from the Hosoya operator. *Rev. Roum. Chim.* **2001**, *46*, 129-141.
- 38. L. Lovasz and J. Pelikan, On the eigenvalues of trees. Period. Math. Hung. **1973**, *3*, 175-182.
- 39. M. V. Diudea, O. Ivanciuc, S. Nikolić, N.Trinajstić, Matrices of Reciprocal Distance. Polynomials and Derived Numbers. *MATCH, Commun. Math. Comput. Chem*, **1997**, *35*, 41-64.
- 40. Randić, M. Design of Molecules with Desired Properties. A Molecular Similarity Approach to Property Optimization. In: Johnson, M. A.; Maggiora, G. M. Eds. *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990, Chap. 5, pp. 77-145.

- 41. Maggiora, G. M.; Johnson, M. A. Introduction to Similarity in Chemistry. In: Johnson, M. A.; Maggiora, G. M. Eds. *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990, Chap. 1, pp. 1-13.
- 42.Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Deffinition and Application. *J. Chem. Phys. Comput. Sci.* **1985**, *25*, 64-73.
- 43. Good, A. C. *J. Mol. Graphics*, **1992**, *10*, 114.
- 44. Son, S. H.; Han, Ch. K.; Ahn, S. K.; Yoon, J. H.; No, K. T. Development of Tree-Dimensional Descriptors Represented by Tensors: Free Energy of Hydration Density Tensor.
  J. Chem. Inf. Comput. Chem. 1999, 39, 601-609.
- 45. Dearden, J. C., Duffy, J. C., *J. Pharm. Pharmacol*., **1993**, *45*, 1142.
- 46. Duffy, J. C., Dearden, J. C., Rostron, C., *J. Pharm. Pharmacol*., **1996**, *48*, 883.
- 47. Tarko, L. Proprietatile moleculelor care nu exista., Ed. Zecasin, 2000.
- 48. O. Ursu, G. Katona, M. V. Diudea, Activity prediction by Cluj-Simil program, *Rev. Roum. Chim.* 2003, *48*(4), 321-330.
- 49. C. Y. Hu and L. Xu, On Highly Discriminating Molecular Topological Index, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 82-90
- 50. M. Randić, Graph valence shells as molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2001**,
- 51. V. A. Skorobogatov AND A. A. Dobrynin, Metric analysis of graphs. *MATCH ,Commun, Math. Chem*., **1988**, *23*, 105-151

- 52. A. R Katritzky, V. S. Lobanov, and M. Karelson, *CODESSA reference manual*. Ver. 2.0, Univ. Florida, 1996.
- 53. *MOLCONN-Z*, Hall associates consulting. Quincy, MA, 02170, USA.
- 54. R. Todeschini and V. Consonni, *Handbook of molecular descriptors*. Wiley-VCH, **2000**; Dragon 1.1 software package.
- 55. Y. Gutierrez, E. Estrada, L. Rodriguez, I. Munoz, *TOSS_MODE*, ver. 2.5 for Windows, Central Univ. Las Villas, Santa Clara, Cuba, 1995.
- 56. S. C. Basak, D. K. Harris, V. R. Magnuson, *POLLY* 2.3, Copyright Univ. Minnesota, 1988.
- 57. M. V. Diudea, O. Ursu, Layer matrices and distance property descriptors. *Indian J. Chem.*, 42*A*, **2003**, 1283-1294.