# Cluj Matrix Invariants†

Mircea V. Diudea

Department of Chemistry, "Babes-Bolyai" University, 3400 Cluj, Romania

Received June 19, 1996⊗

The newly proposed matrix, $CJ_u$ (unsymmetric Cluj matrix), is examplified for cycle-containing structures. Its relation with the matrix $SZ_u$ (unsymmetric Szeged matrix) is discussed both as definition and related indices. The derived "Cluj" numbers are compared to the Wiener matrix- and Szeged matrix-derived numbers and tested for discriminating and correlating ability on selected sets of graphs.

## INTRODUCTION

The problem of distances in a graph, **G**, is one of the most studied questions, both from theoretical point of view and applications (the reader can consult two recent reviews[1,2]). It is connected to the Wiener number,[3] ***W***, or "the path number" as denominated by its initiator.

In acyclic structures, Wiener number and its extension, hyper-Wiener number,[4] can be defined as edge/path contributions, $I_{e/p}$, to the global number, $I$

$$I = I(G) = \sum_{e/p} I_{e/p} = \sum_{e/p} N_{i,\ e/p}\, N_{j,\ e/p} \qquad (1)$$

where $N_i$ and $N_j$ represent the number of vertices lying on the two sides of the edge/path, **e/p** (having the endpoints **i** and **j**). Edge/path contributions, $I_{e/p}$ are just the entries in the Wiener matrices,[5,6] $W_e$ and $W_p$. Thus, $I$ can be calculated by

$$I = (1/2) \sum_{\mathbf{i}} \sum_{\mathbf{j}} [W_{e/p}]_{i\,j} \qquad (2)$$

Within this paper, numbers (i.e., topological indices) are denoted by the corresponding matrix (italicized) symbols, whereas matrices (and their entries) by special right letters. The reason for such a notation came out from the aim to suggest the different graph-theoretical properties collected by different matrices, which provide identical Wiener-type numbers (see below).

As suggested in eqs 1 and 2, $I$ can be defined either on edge (i.e. **(i,j)**∈**E(G)**, where **E(G)** is the set of edges in graph) or on path (i.e., **(i,j)**∈**P(G)**, with **P(G)** being the set of paths in graph). Thus, when defined on edge, the index is a Wiener, $W_e$, number; when defined on path, it is a hyper-Wiener, $W_p$, number (also denoted ***WW*** and ***R***—see refs 5 and 6). The numbers $W_e$ and $W_p$ count all external paths passing through the two endpoints of all edges/paths **(i,j)**.

Attempts have been made[7−9] to extend the definition (1) to cycle-containing structures, such as

$$I_e = (1/2) \sum_{\mathbf{i}} \sum_{\mathbf{j}} C^e_{ij}/C_{ij} \qquad (3)$$

where $C_{ij}$ is the number of the shortest paths between vertices **i** and **j**, and $C^e_{ij}$ denotes the number of those shortest paths between **i** and **j** which contain the edge **e**. No convincing definition of $I_p$ contributions have been offered by Lukovits and Linert[9] (see Discussion).

A challenge definition[10,11] of Wiener-type numbers is supplied by the distance matrix, D

$$I = (1/2) \sum_{\mathbf{i}} \sum_{\mathbf{j}} [D_{e/p}]_{ij} \qquad (4)$$

where $D_e$ is just the classical D matrix, whereas $D_p$ is the "distance path" matrix.[10] The Wiener number, $D_e$, and the hyper-Wiener number, $D_p$, count all "internal" paths of length |**e/p**| contained into all shortest paths, **(i,j)**, in a graph;[10,12] |**e**| ≤ |**p**| ≤ |**(i,j)**|. Note that, in acyclic structures $D_e = W_e$ and $D_p = W_p$. $D_p$ matrix is defined[10] as

$$[D_p]_{ij} = \binom{[D_e]_{ij} + 1}{2} \qquad (5)$$

and since it is derived on the $D_e$ matrix, the relation (4) is valid both for acyclic and cyclic structures. The $D_p$ matrix is illustrated in Figure 1, for the graph 1 (1-methylcyclohexane).

One of the Wiener analogue numbers is the so called "Szeged" number, ***SZ***, patterned by Gutman.[13−18] It is defined according to eq 1 but the sets $N_i$ and $N_j$ are so defined that eq 1 becomes valid both for acyclic and cycle-containing graphs

$$N_{i,e/p} = |\{\mathbf{v}|\mathbf{v}\in\mathbf{V(G)};\quad [D_e]_{iv} < [D_e]_{jv}\}| \qquad (6)$$

$$N_{j,e/p} = |\{\mathbf{v}|\mathbf{v}\in\mathbf{V(G)};\quad [D_e]_{jv} < [D_e]_{iv}\}| \qquad (7)$$

Thus, $N_i$ and $N_j$ denote the cardinality of the sets of vertices closer to **i** and to **j**, respectively; vertices equidistant to **i** and **j** are not counted.

The Szeged numbers can be also calculated from the Szeged matrices,[19] $SZ_{e/p}$, by applying the Wiener operator (i.e., the half sum of the matrix entries) by analogy to relation (2)

$$I = (1/2) \sum_{\mathbf{i}} \sum_{\mathbf{j}} [SZ_{e/p}]_{ij} \qquad (8)$$

When defined on edge, the index is the Szeged, $SZ_e$; when defined on path, the index is the hyper-Szeged,[19] $SZ_p$.

The Szeged index parallels,[13,17] in some classes of graphs, the Wiener index: in tree graphs, $SZ_e(T) = W_e(T) = D_e(T)$; in complete graphs, $SZ_e(K_N) = D_e(K_N) = N(N-1)/2$; $N > 2$; in simple cycles, $SZ_e(C_{N,even}) = 2D_e(C_{N,even})$. There exist cyclic graphs,[13] for which $SZ_e = D_e$. The hyper-Szeged index is different from the hyper-Wiener index, both in acyclic and cyclic structures: $SZ_p \neq W_p \neq D_p$. For odd $N$-membered cycles, $SZ_p(C_{N,odd}) = N(N-1)^3/8$, whereas for

† This work is dedicated to my city, Cluj, and its wonderful people.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1997.

S0095-2338(96)00365-4 CCC: $14.00 

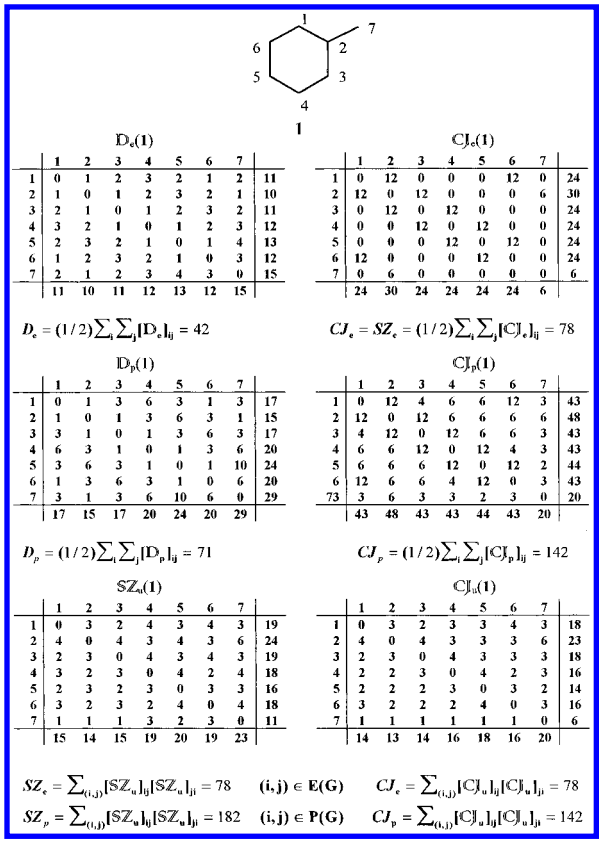



**Figure 1.** Matrices $D_e$, $D_p$, $CJ_e$, $CJ_p$, $SZ_u$, and $CJ_u$ and derived indices.

even ones, $SZ_p(C_{N,even}) = N(N^3-3N^2+6N-4)/8$. Both $SZ_e$ and $SZ_p$ correlate[19] with some physico-chemical properties (i.e., boiling points, chromatographic retention index, enthalpies of formation, etc.).

In this paper, the applicability of the $CJ_u$ matrix[12] is illustrated in cycle-containing structures. The Cluj indices, calculated from this matrix, are compared to the numbers derived on $D_e$, $D_p$, $SZ_e$, and $SZ_p$ matrices and then tested for correlating and discriminating ability on selected sets of graphs.

## CLUJ MATRICES AND DERIVED NUMBERS

In a previous paper,[12] a square matrix, $CJ_u$ (Cluj unsymmetric), was defined by following the principle of "single endpoint characterization of a path"

$$[CJ_u]_{ij} = N_{i,(i,j)} \tag{9}$$

$$N_{i,(i,j)} = \mathbf{max}|\{v|v\in V(G);\quad [D_e]_{iv} < [D_e]_{jv};\quad (i,v)\cap(i,j) = \mathbf{max}\{i\};|(i,j)| = \mathbf{min}\}|$$

It collects the vertices lying closer to the focused vertex **i** but out of the shortest path **(i,j)** or, in other words, the "external" paths on the side of **i**, which include the path **(i,j)**. In the case of two or more shortest paths **(i,j)** (i.e., two or more sets {**v**}) are encountered, the maximal cardinality of the sets {**v**} is taken, by definition (9). The above definition is valid both for acyclic and cyclic graphs. Figure 1 illustrates the $CJ_u$ matrix for the graph 1. For trees, $CJ_u$ was illustrated in ref 12. One can see that it is a square, unsymmetric (except some regular graphs) of dimensions **N*N**. Note that the subscript **u** comes from the "unsymmetric characterization of path".

**Table 1.** Boiling Points[46] and Wiener-Type Indices for Some Cycloalkanes

| no. | graph[a] | BP | *NC* | $D_e$ | $D_p$ | $SZ_e$ | $SZ_p$ | $CJ_p$ |
|---|---|---|---|---|---|---|---|---|
| 1 | C4 | 13.1 | 4 | 8 | 10 | 16 | 18 | 18 |
| 2 | 11MC3 | 21 | 5 | 15 | 20 | 15 | 24 | 24 |
| 3 | EC3 | 35.9 | 5 | 17 | 26 | 17 | *40* | 32 |
| 4 | MC4 | 40.5 | 5 | 16 | 23 | 28 | 41 | 37 |
| 5 | C5 | 49.3 | 5 | 15 | 20 | 20 | *40* | 40 |
| 6 | 112MC3 | 56.5 | 6 | 26 | *39* | 26 | 59 | 49 |
| 7 | 123MC3 | 66 | 6 | 27 | *42* | 27 | 63 | 54 |
| 8 | EC4 | 70.7 | 6 | 29 | 49 | 45 | 92 | 73 |
| 9 | MC5 | 71.8 | 6 | 26 | *39* | 33 | 79 | 71 |
| 10 | C6 | 80.7 | 6 | 27 | *42* | 54 | 105 | 90 |
| 11 | PC4 | 110 | 7 | 48 | 94 | 68 | 181 | 132 |
| 12 | 11MC5 | 88.9 | 7 | 39 | 61 | 48 | 133 | 105 |
| 13 | 12MC5 | 91.9[b] | 7 | 40 | 64 | 49 | 137 | 109 |
| 14 | 13MC5 | 91.7[b] | 7 | 41 | 68 | 51 | 147 | 119 |
| 15 | MC6 | 100.9 | 7 | 42 | 71 | 78 | 182 | 142 |
| 16 | C7 | 117 | 7 | 42 | 70 | 63 | 189 | 154 |
| 17 | 112MC5 | 114 | 8 | 56 | 92 | 67 | 218 | 150 |
| 18 | 113MC5 | 105 | 8 | *58* | 100 | 71 | 240 | 170 |
| 19 | 123MC5 | 115 | 8 | *58* | 99 | 70 | 230 | 164 |
| 20 | 1M2EC5 | 124 | 8 | *61* | 110 | 72 | 247 | 178 |
| 21 | 1M3EC5 | 121 | 8 | 63 | 119 | 76 | 270 | 199 |
| 22 | PC5 | 131 | 8 | 67 | 135 | 78 | *296* | 215 |
| 23 | IPC5 | 126.4 | 8 | *62* | 114 | 73 | 278 | 186 |
| 24 | 11MC6 | 119.5 | 8 | 59 | 103 | 104 | 284 | 197 |
| 25 | 12MC6 | 123.4[b] | 8 | 60 | 106 | 106 | *296* | 202 |
| 26 | 13MC6 | 124.5[b] | 8 | *61* | 110 | 108 | *296* | 211 |
| 27 | 14MC6 | 120 | 8 | *62* | 115 | 110 | 310 | 220 |
| 28 | EC6 | 131.8 | 8 | *64* | 122 | 109 | 308 | 226 |
| 29 | MC7 | 134 | 8 | *61* | 109 | 88 | 306 | 225 |
| 30 | C8 | 146 | 8 | *64* | 120 | 128 | 364 | 288 |
| 31 | 1123MC5 | 132.7 | 9 | 78 | 137 | 93 | 358 | 222 |
| 32 | 113MC6 | 136.6 | 9 | 82 | 152 | 140 | *447* | 285 |
| 33 | 124MC6 | 136 | 9 | *84* | 160 | *144* | 474 | 296 |
| 34 | 135MC6 | 138.5 | 9 | *84* | 159 | *144* | *447* | 291 |
| 35 | 1M2EC6 | 151 | 9 | 86 | 167 | *142* | 460 | 300 |
| 36 | 1M3EC6 | 149 | 9 | *88* | *176* | 146 | 481 | 322 |
| 37 | PC6 | 154 | 9 | 94 | 203 | 148 | 508 | 352 |
| 38 | IPC6 | 146 | 9 | *88* | *176* | *142* | 487 | 313 |
| 39 | EC7 | 163.5 | 9 | *88* | 175 | 121 | 484 | 337 |
| 40 | C9 | 170 | 9 | 90 | 180 | *144* | 576 | 450 |
| 41 | 1M2IPC6 | 171 | 10 | 114 | 231 | 180 | 693 | 401 |
| 42 | 1M3IPC6 | 167.5 | 10 | 117 | 245 | 186 | 739 | 436 |
| 43 | 13EC6 | 170.5 | 10 | 121 | 263 | 192 | 735 | 467 |
| 44 | PC7 | 183.5 | 10 | 124 | 277 | 163 | 757 | 503 |
| 45 | C10 | 201 | 10 | 125 | 275 | 250 | 945 | 705 |

[a] M = methyl; E = ethyl; P = propyl; IP = isopropyl; Cn = n-membered cycle. [b] Values for the *trans*-isomer.

The $CJ_u$ matrix allows the construction of symmetric $CJ_e$ and $CJ_p$ matrices by relation

$$[CJ_{e/p}]_{ij} = [CJ_u]_{ij}\,[CJ_u]_{ji} \tag{10}$$

and the derivation of two Wiener-type indices, by applying either the Wiener operator on the symmetric matrices or an "orthogonal" operator on the unsymmetric matrix $CJ_u$

$$I = (1/2)\sum_i\sum_j[CJ_{e/p}]_{ij} = \sum_{(i,j)}[CJ_u]_{ij}[CJ_u]_{ji} \tag{11}$$

When defined on edge, the index $CJ_e$ shows the following relations with the Wiener and Szeged numbers: $CJ_e(T) = SZ_e(T) = W_e(T) = D_e(T)$ and $CJ_e(C) = SZ_e(C) \neq D_e(C)$. When defined on path, $CJ_p(T) = W_p(T) = D_p(T) \neq SZ_p(T)$ and $CJ_p(C) \neq D_p(C) \neq SZ_p(C)$. Despite the formal similarity between $CJ_p$ and $SZ_p$ indices (compare eqs 6−8 with eqs 9−11), the supplementary condition, $(i,v)\cap(i,j) = \mathbf{max}\{i\}$

(see eq 9), detaches the two hyper-indices (and the corresponding matrices). Values of the above mentioned indices in a set of cycloalkanes are listed in Table 1.

## CLUJ INDEX $CJ_P$ IN SIMPLE CYCLES

By analogy to the hyper-Wiener numbers, $W_p$ and $D_p$, the hyper-Cluj index, $CJ_p$, can be decomposed as

$$CJ_p = CJ_e + CJ_\Delta \tag{12}$$

It is obvious that $CJ_\Delta$ is obtained when $I$ is defined (cf. eq 11) on paths larger than 1

$$CJ_\Delta = \sum_{(i,j)}[CJ_u]_{ij}[CJ_u]_{ji}; \quad |(i,j)| > 1 \tag{13}$$

As above mentioned, in cyclic graphs $CJ_e$ equals $SZ_e$, and it is given by relations

$$CJ_e\,(C_{N,odd}) = N(N-1)^2/4 \tag{14}$$

$$CJ_e\,(C_{N,even}) = N^3/4 \tag{15}$$

$CJ_\Delta$ can be calculated both for odd and even $N$-membered cycles by

$$CJ_\Delta\,(C_N) = zN\left(\frac{N-z}{2}\right)^2 + 2N\sum_{x=1}^{x_{max}-1}\left(\frac{N-z}{2}-x\right)^2 + \left[\left(\frac{N-z}{2}\right)\mathbf{mod}2\right]N\left(\frac{N-z}{2}-x_{max}\right)^2 + (1-z)\left(\frac{N-z}{2}\right)\left(\frac{N-z}{2}-x_{max}\right)^2; \quad x_{max} = [(N-z)/4]; \quad z = N \bmod 2 \tag{16}$$

Making $N = 2k + z$ and $(N-z)/2 \bmod 2 = r$, the evaluation of sum in eq 16 leads to

$$CJ_\Delta = \frac{7}{6}k^4 + \left(\frac{7}{3}z + \frac{1}{2}r - \frac{11}{4}\right)k^3 + \left(z^2 - 2z + \frac{1}{4}zr - \frac{2}{3}\right)k^2 + \left(\frac{1}{2}zr - \frac{5}{6}z + \frac{1}{2}r - \frac{3}{4}\right)k + \frac{1}{4}zr - \frac{1}{2}z \tag{17}$$

By detailing the dependency of $CJ_\Delta$ by the **Nmod4** (see eq 16), four cases are to be considered, as Table 2 shows.

Thus eq 17 can be written (function of **k** and **N**, respectively) as

$$CJ_\Delta\,(a) = k^2(14k^2-27k+4)/12 = N^2(7N^2-27N+8)/96 \tag{18}$$

$$CJ_\Delta\,(b) = k(2k+1)(7k^2-3k+2)/12 = N(N-1)(7N^2-20N+21)/96 \tag{19}$$

$$CJ_\Delta\,(c) = k(14k^3-27k^2+4k-3)/12 = N(7N^3-27N^2+8N-12)/96 \tag{20}$$

$$CJ_\Delta\,(d) = (2k+1)(7k^3-3k^2-k-3)/12 = N(N-3)(7N^2-6N+11)/96 \tag{21}$$

Keeping in mind eq 12 and making the appropriate substitutions, the following relations for $CJ_p$ index can be obtained

$$CJ_p\,(a) = k^2(14k^2-3k+4)/12 = N^2(7N^2-3N+8)/96 \tag{22}$$

$$CJ_p(b) = k(2k+1)(7k+2)(k+1)/12 = N(N-1)(N+1)(7N-3)/96 \tag{23}$$

$$CJ_p(c) = k(2k-1)(7k^2+2k+3)/12 = N(N-1)(7N^2+4N+12)/96 \tag{24}$$

$$CJ_p(d) = (2k+1)(k+1)(7k^2+2k-3)/12 = N(N+1)(7N^2-10N-9)/96 \tag{25}$$

**Table 2.** Parameters of Eq 17, by Detailing **Nmod4**

| case | N | k | z | r |
|---|---|---|---|---|
| a | 4p | 2p | 0 | 0 |
| b | 4p+1 | 2p | 1 | 0 |
| c | 4p+2 | 2p+1 | 0 | 1 |
| d | 4p+3 | 2p+1 | 1 | 1 |

**Table 3.** Statistics of Single Variable Regression: **BP** vs Topological Indices (Table 1)

| | index | *r* | *s* | cv (%) | *F* |
|---|---|---|---|---|---|
| **1** | $D_e$ | **0.950** | **13.75** | **11.87** | **399.5** |
| **2** | $D_p$ | **0.929** | **16.29** | **14.07** | **272.2** |
| **3** | $SZ_e$ | **0.916** | **17.68** | **15.26** | **224.8** |
| **4** | $SZ_p$ | **0.916** | **17.68** | **15.26** | **224.7** |
| **5** | $CJ_p$ | **0.920** | **17.29** | **14.93** | **237.0** |
| **6** | $NC$ | **0.969** | **10.88** | **9.40** | **663.7** |
| **7** | ln $D_e$ | **0.974** | **10.03** | **8.66** | **788.5** |
| **8** | ln $D_p$ | **0.976** | **9.54** | **8.24** | **875.8** |
| **9** | ln $SZ_e$ | **0.967** | **11.28** | **9.74** | **614.58** |
| **10** | ln $SZ_p$ | **0.986** | **7.39** | **6.38** | **1489.7** |
| **11** | ln $CJ_p$ | **0.991** | **5.93** | **5.12** | **2333.7** |
| **12** | ln $NC$ | **0.965** | **11.59** | **10.00** | **580.5** |

## CORRELATING AND DISCRIMINATING TESTS

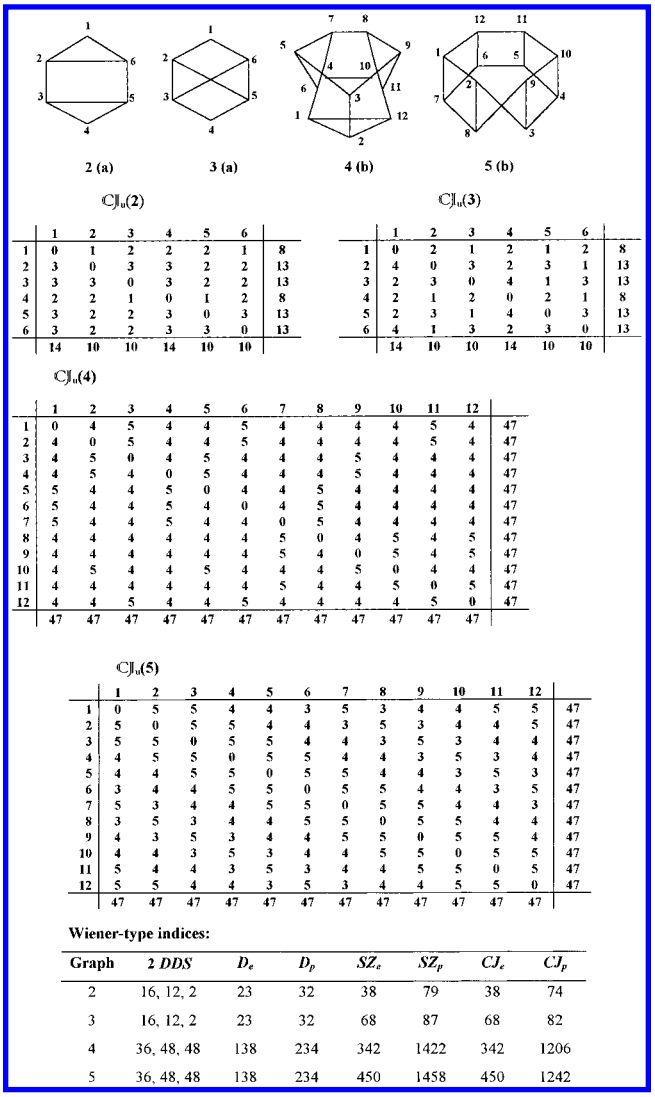
Boiling point, **BP**, is one of the most common physicochemical properties which characterizes a chemical compound. It is directly related to other physical quantities, such as critical temperature or enthalpy of vaporization. Usually, **BP** is easily to measure. However, there are compounds whose vapors are very aggressive or toxic, or compounds which are simply unavailable. In such cases, a procedure which estimates **BP** from the chemical structure is welcome.

A set of 45 cycloalkanes (Table 1) was chosen for testing the correlating ability of some Wiener-type indices ($CJ_p$ included) with **BP**. Statistics of single variable regression are listed in Table 3.

Besides the correlating ability, a second major quality required for a topological index is the discrimination of nonisomorphic isomers. Discriminating ability is useful in chemical documentation (storage and retrieval of chemical structures).[20−22]

A crude evaluation of the discriminating ability of the herein discussed indices can be made from the values of indices included in Table 1: $CJ_p$ shows no degeneracy (the degenerate values are italicized).

A systematic search, including the calculation of the sensitivity, **S** (calculated as the ratio of the number of distinct values over the whole number of investigated structures),[23,24] of these indices on the set of all cycloalkane isomers having ten vertices/atoms and 3−10 membered cycles (439 structures) is presented in ref 25. That study indicated that the




| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 2 | 2 | 2 | 1 | 8 |
| 2 | 3 | 0 | 3 | 3 | 2 | 2 | 13 |
| 3 | 3 | 3 | 0 | 3 | 2 | 2 | 13 |
| 4 | 2 | 2 | 1 | 0 | 1 | 2 | 8 |
| 5 | 3 | 2 | 2 | 3 | 0 | 3 | 13 |
| 6 | 3 | 2 | 2 | 3 | 3 | 0 | 13 |
| | 14 | 10 | 10 | 14 | 10 | 10 | |

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 2 | 1 | 2 | 1 | 2 | 8 |
| 2 | 4 | 0 | 3 | 2 | 3 | 1 | 13 |
| 3 | 2 | 3 | 0 | 4 | 1 | 3 | 13 |
| 4 | 2 | 1 | 2 | 0 | 2 | 1 | 8 |
| 5 | 2 | 3 | 1 | 4 | 0 | 3 | 13 |
| 6 | 4 | 1 | 3 | 2 | 3 | 0 | 13 |
| | 14 | 10 | 10 | 14 | 10 | 10 | |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 4 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 4 | 47 |
| 2 | 4 | 0 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 4 | 47 |
| 3 | 4 | 5 | 0 | 4 | 5 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 47 |
| 4 | 4 | 5 | 4 | 0 | 5 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 47 |
| 5 | 5 | 4 | 4 | 5 | 0 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 47 |
| 6 | 5 | 4 | 4 | 5 | 4 | 0 | 4 | 5 | 4 | 4 | 4 | 4 | 47 |
| 7 | 5 | 4 | 4 | 5 | 4 | 4 | 0 | 5 | 4 | 4 | 4 | 4 | 47 |
| 8 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 0 | 4 | 5 | 4 | 5 | 47 |
| 9 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 0 | 5 | 4 | 5 | 47 |
| 10 | 4 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 5 | 0 | 4 | 4 | 47 |
| 11 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 4 | 5 | 0 | 5 | 47 |
| 12 | 4 | 4 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 0 | 47 |
| | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 5 | 5 | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 5 | 5 | 47 |
| 2 | 5 | 0 | 5 | 5 | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 5 | 47 |
| 3 | 5 | 5 | 0 | 5 | 5 | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 47 |
| 4 | 4 | 5 | 5 | 0 | 5 | 5 | 4 | 4 | 3 | 5 | 3 | 4 | 47 |
| 5 | 4 | 4 | 5 | 5 | 0 | 5 | 5 | 4 | 4 | 3 | 5 | 3 | 47 |
| 6 | 3 | 4 | 4 | 5 | 5 | 0 | 5 | 5 | 4 | 4 | 3 | 5 | 47 |
| 7 | 5 | 3 | 4 | 4 | 5 | 5 | 0 | 5 | 5 | 4 | 4 | 3 | 47 |
| 8 | 3 | 5 | 3 | 4 | 4 | 5 | 5 | 0 | 5 | 5 | 4 | 4 | 47 |
| 9 | 4 | 3 | 5 | 3 | 4 | 4 | 5 | 5 | 0 | 5 | 5 | 4 | 47 |
| 10 | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 5 | 5 | 0 | 5 | 5 | 47 |
| 11 | 5 | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 5 | 5 | 0 | 5 | 47 |
| 12 | 5 | 5 | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 5 | 5 | 0 | 47 |
| | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | |

**Wiener-type indices:**

| Graph | 2 *DDS* | $D_e$ | $D_p$ | $SZ_e$ | $SZ_p$ | $CJ_e$ | $CJ_p$ |
|---|---|---|---|---|---|---|---|
| 2 | 16, 12, 2 | 23 | 32 | 38 | 79 | 38 | 74 |
| 3 | 16, 12, 2 | 23 | 32 | 68 | 87 | 68 | 82 |
| 4 | 36, 48, 48 | 138 | 234 | 342 | 1422 | 342 | 1206 |
| 5 | 36, 48, 48 | 138 | 234 | 450 | 1458 | 450 | 1242 |

**Figure 2.** Pairs of graphs showing identical distance degree sequence, ***DDS***: (a) irregular graphs and (b) regular graphs; $CJ_u$ matrices and Wiener-type indices.

sensitivity of $CJ_p$ to distinguish among the above mentioned isomers is about 0.476. This value is very close to that shown by $SZ_p$ (0.483) and superior to the sensitivity of the Wiener, $D_e$ (0.116), and hyper-Wiener, $D_p$ (0.339), indices.

The cycloalkane isomers were generated by the program FRAGGEN, written in Turbo Pascal.

In an additional test, two pairs of graphs,[26,27] showing degeneracy of several topological indices (e.g., those based on distances in graph) were selected (Figure 2). As it can be seen, their distance degree sequences, **DDS**, and the Wiener numbers are pairwise identical. Conversely, the Cluj (and Szeged) indices discriminates between the nonisomorphic isomers of each pair; $CJ_e$ equals $SZ_e$ in each graph, whereas $CJ_p$ is different from $SZ_p$.

Note that $CJ_u$ matrix is unsymmetric for the regular graph 4 but symmetric for 5, which is a symmetric regular graph. For nonregular graphs, $CJ_u$ is always unsymmetric.

## DISCUSSION

Wiener-type indices are perhaps the most studied topological descriptors out of more than 100 devised indices, owing to their easy calculation and structural interpretation. Equation 1 is one of the simplest relations, allowing both manual and computer operation. The meaning of the quantities $N_i$ and $N_j$ differs within different variants of the Wiener number: they count vertices, lying on the two sides of the edge/path **e/p** (with **i** and **j** as endpoints), connected in the case of Wiener, $W_e$, and hyper-Wiener, $W_p$, numbers but not necessarily connected, in the case of the Szeged indices.[16]

In Wiener indices, $I$ counts (cf. eq 1) external paths passing through the two endpoints of all edges/paths **(i,j)**. In Szeged indices, $I$ counts only vertices, lying closer to one or another endpoint of the edge/path **(i,j)**.

When defining its hyper-Wiener, Randić[4] has given a manual calcuation: cut the path **(i,j)** and then multiply the number of vertices of the two resulting disconnected subgraphs. The global number is obtained by summing the contributions $I_p$ for all paths in graph. This is equivalent (at least in the opinion of ref 9) to the replacement of the path **(i,j)** by the edge **(i,j)** in a reduced graph. This artefact has no consequences in trees, where the number of external paths is counted only with respect to the endpoints **i** and **j**. Conversely, in cycles, C, the edge contribution in the reduced cycle, $I_e(C')$ is not equivalent to the path contribution in the original cycle, $I_p(C)$. Although, in acyclic structures, the global number of internal and external paths[28] (i.e., the quantities $W_p$ and $D_p$) is identical, only the internal path calculation (i.e., the quantity $D_p$ ) is consistent both in acyclic and cyclic structures.[10,28] It is not surprising that the values for the hyper-Wiener number, $W_p$, calculated by the procedures of Randić[5] and Lukovits[9] are different from one another and from the $D_p$ values.[10]

Szeged index[13] is a more rational application of eq 1 in cycles. The elimination of the vertices equidistant to **i** and **j** makes superfluous the checking equality of the internal and external path counting. The immediate consequence is that the Szeged indices[13,19] do not count paths but vertices (even unconnected !) lying closer to one or another endpoint of the edge/path **(i,j)**. This result is more pregnant in the hyper-Szeged index,[19] where some vertices closer say to **i** are internal with respect to the path **(i,j)**.

The principle[10] of "single endpoint characterization of a path" could encourage us to check further the equality of the internal and external path counting, in cyclic graphs. However, no values identical to $D_p$ values were obtained. Given this result, we adopted the idea of Gutman[13] to count the vertices lying closer to the focused point, **i**. The additional condition: the vertices {**v**}, closer to **i**, must be external with respect to the path **(i,j)** involves that the $CJ_p$ counts external paths (as $W_p$ does) originating in the *vicinity* of the two endpoints of the given path **(i,j)**. Thus, $CJ_p$ is a hybrid between hyper-Wiener and hyper-Szeged numbers; it equals the $W_p$ number in trees but is different both from $W_p$ and $SZ_p$ in cyclic structures.

Note that Tratch et al.[29] have recently proposed an "extended distance matrix", E, whose entries are defined as the product of the distance $\mathbf{D_{ij}}$ and a multiplier, $\mathbf{m_{ij}}$, which is the number of paths in the graph of which the path **(i,j)** is a subgraph. In acyclic structures, the multipliers $\mathbf{m_{ij}}$ equal the entries $[W_p]_{ij}$ in the Wiener matrix. Thus, E can be obtained as the matrix pairwise (Hadamard)[30] product: $E = \mathbf{Had}(D_e W_p)$. In cycle containing graphs, $\mathbf{m_{ij}}$ is very close to entries in the Cluj and Szeged matrices. Distance-extended Szeged and Cluj matrices were developed in two new papers.[31,32]

As it can be seen from Table 3, none of the investigated indices correlates satisfactorily with the boiling points of the selected set of cycloalkanes. The correlation increases if a logarithmic scale is used for the topological descriptors (Table 3, entries 7−12). This is more pregnant for the Szeged indices and maximal in the case of $\boldsymbol{CJ_p}$ (from $r = 0.920$; $s = 17.29$, for $\boldsymbol{CJ_p}$, to $r = 0.991$; $s = 5.93$ for $\mathbf{ln}\boldsymbol{CJ_p}$). Observe the drop (about three times !) in standard error of estimate, *s*. Interpretation of this statistical parameter needs, however, the knowledge of the values of the investigated property (**BP** in this case). It is not more necessary when the coefficient of variance, **cv** is used (Table 3); this parameter, calculated as the ratio of *s* over the mean value of the given property, allows an immediate evaluation of the quality of the regression equation. **Cross validation** test (let 1/3 out) showed a correlation coefficient of 0.983 for the equation: $\mathbf{BP} = \mathbf{a} + \mathbf{b}\ \mathbf{ln}\ \boldsymbol{CJ_p}$ indicating for it a good predictibility.

The good correlation of **BP** with the number of carbon atoms, ***NC***, is an expression of the molecular "size" contribution to that property. The size and "shape" contributions can be clearly seen when the Wiener index is decomposed as:

$$\mathbf{W_e} = \binom{\mathbf{N}}{\mathbf{2}} + \binom{\mathbf{N}}{\mathbf{3}}$$

in *n*-alkanes. The first term is the size contribution, whereas the second term is the shape contribution. In trees, the number of triplets (counted for each branching point) has to be substracted from the second term.[12,33] All topological descriptors herein discussed include both a size and a shape contribution. To resume, the "crude" contribution of ***NC*** to the correlation can be (sometimes) overpassed by using a logarithmic scale for indices (see Table 3).

Discriminating ability among nonisomorphic isomers is a desiderate not easily reached. Many efforts have been devoted to the finding of a topological index able to characterize a molecule uniquely. The competition started by the Randić's molecular **ID** (identification) number,[20] defined as the sum of all weighted paths in a molecule. Since then, various weighted paths,[34−36] walks,[37,38] or distances[39] have been considered for devising new **ID** numbers. They have been tested on larger and larger sets of chemical structures.

Very recently, a new scheme for weighting edges in a graph, involving layer matrices and powers of an extended adjacency matrix, has been proposed by Hu and Xu.[40,41] The resulting **EAID** number appears to be the highest discriminating index tested so far (no degeneracy on over 4 000 000 chemical structures).

It was not our intention to propose an **ID** number here; our Cluj indices are simple descriptors which can be compared only to simple indices, such as Wiener,[3] Szeged,[13,19] or Balaban[42] ***J*** index (the last one showing the first degenerate values in dodecanes).

Our study[25] showed that the $\boldsymbol{H_{CJp}}$ index (calculated from the reciprocal Cluj entries $1/[CJ_u]_{ij}$) is the *only nondegenerate index* among other 12 simple indices, ***J*** included. This fact indicates that, in calculating $\boldsymbol{CJ_p}$, no *assignment degeneracy*[43] occurs; the degeneracy is induced only at the *operational stage* (simple summation of local values). Further development of Cluj and Szeged fragmental indices and the corresponding distance-extended indices will offer even more discriminating molecular descriptors.[44]

Analytical relations for calculating the $\boldsymbol{CJ_p}$ index in simple cycles were derived by using the MAPLE. V algebraic program. A more complete list of analytical relations for the herein discussed indices is presented in a work dealing with indices of reciprocal properties (or Harary indices).[25]

## CONCLUSIONS

The principle of single endpoint characterization of a path enables the construction of (in general) unsymmetric matrices.[10] Among them, the Cluj matrix, $CJ_u$, includes the information of at least two different symmetric square matrices (e.g., $W_e$ and $D_e$, in acyclic graphs). It can be used as a basis for constructing Wiener-type indices[12] as well as Schultz-type indices.[31,45] The novel hyper-index, $\boldsymbol{CJ_P}$, shows good corelation with the boiling points of a selected set of cycloalkanes.. It is also a promising tool in separating nonisomorphic isomers, particularly as the reciprocal property index, $\boldsymbol{H_{CJp}}$, which within the set of all cycloalkanes with ten carbon atoms and 3−10 membered cycles was the most discriminating among 13 indices tested. Additional studies in these topics are in progress.

## ACKNOWLEDGMENT

The author wishes to thank to the referee for his valuable comments and to professor Bazil Pârv, Department of Mathematics, "Babeş-Bolyai" University of Cluj, for computer assistance. This work is supported by the Grant A 17.4/979 of the Roumanian Ministry of Education.

## REFERENCES AND NOTES

(1) Gutman, I.; Yeh, Y. N.; Lee, S. L.; Luo, Y. L. Some Recent Results in the Theory of the Wiener Number. *Indian J. Chem.* **1993**, *32A*, 651−661.
(2) Nikolić, S.; Trinajstić, N.; Mihalić, Z. The Wiener Index: Development and Applications. *Croat. Chem. Acta.* **1995,** *68*, 105−129.
(3) Wiener, H. Structural Determination of Paraffin Boiling points. *J. Am. Chem. Soc.* **1947,** *69*, 17−20.
(4) Randić, M. Novel Molecular Descriptor for Structure-Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478−483.
(5) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener Matrix: Source of Novel Graph Invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709−716.
(6) Randić, M; Guo, X.; Oxley; T.; Krishnapriyan, H.; Nayor, L. Wiener Matrix Invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361−367.
(7) Lukovits, I.; Gutman, I. Edge-Decomposition of the Wiener Number. *Commun. Math. Comput. Chem.* (*MATCH)* **1994,** *31***,** 133−144.
(8) Lukovits, I. An Algorithm for Computation of Bond Contributions of the Wiener Index. *Croat. Chem. Acta* **1995**, *68*, 99−103.
(9) Lukovits, I.; Linert, W. A Novel Definition of the Hyper-Wiener Index for Cycles. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 899−902.
(10) Diudea, M. V. Walk Numbers $^{e}W_M$ : Wiener-Type Numbers of Higher Rank. *J. Chem. Inf. Comput. Sci.* **1996,** *36*, 535−540.
(11) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971,** *44*, 2332−2339.
(12) Diudea, M. V. Wiener and Hyper-Wiener Numbers in a Single Matrix. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 833−836.
(13) Gutman, I. A Formula for the Wiener Number of Trees and Its Extension to Graphs Containing Cycles. *Graph Theory Notes New York* **1994**, *27*, 9−15.
(14) Dobrynin, A. A.; Gutman, I. On a Graph Invariant Related to the Sum of All Distances in a Graph. *Publ. Inst. Math. (Beograd)* **1994**, *56*, 18−22.
(15) Dobrynin, A. A.; Gutman, I.; Dömötör, G. A Wiener-Type Graph Invariant for Some Bipartite Graphs. *Appl. Math. Lett.* **1995**, *8*, 57−62.

(16) Dobrynin, A. A.; Gutman, I.; Solving a Problem Connected with Distances in Graph. *Graph Theory Notes New York* **1995**, *28*, 21−23.
(17) Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A. A.; Gutman, I.; Dömötör, G. The Szeged Index and an Analogy with the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547−550.
(18) Gutman, I.; Klavžar, S. An Algorithm for the Calculation of the Szeged Index of Benzenoid Hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1011−1014.
(19) Diudea, M. V.; Minailiuc, O. M.; Katona, G.; Gutman, I. Szeged Matrices and Related Numbers. *Commun. Math. Comput. Chem.* (*MATCH)* **1997**, *35*, 000−000.
(20) Randić, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.
(21) Burden, F. R. Molecular Identification Number for Substructure Search. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 225−227.
(22) Elk, S. B.; Gutman, I. Further Properties Derivable from the Matula Numbers of an Alkane. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 54−57.
(23) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Isomer Discrimination by Topological Information Approach, *J. Comput. Chem.* **1981**, 2, 127−148.
(24) Razinger, M.; Chrétien, J. R.; Dubois, J. K. Structural Selectivity of Topological Indices in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23−27.
(25) Diudea, M. V. Indices of Reciprocal Properies or Harary Indices. *J. Chem. Inf. Comput. Sci.* To be published.
(26) Skorobogatov, V. A.; Dobrynin, A. A. Metric Analysis of Graphs. *Commun. Math. Comput. Chem.* (*MATCH)* **1988**, *23*, 105−151.
(27) Diudea, M. V. Layer Matrices in Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1064−1071.
(28) Klein, D. J.; Lukovits, I.; Gutman, I. On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50−52.
(29) Tratch, S. S.; Stankevitch, M. I.; Zefirov, N. S. Combinatorial Models and Algorithms in Chemistry. The Expanded Wiener Number - a Novel Topological Index. *J. Comput. Chem.* **1990**,*11*, 899−908.
(30) Horn, R. A.; Johnson, C. R. *Matrix Analysis*; Cambridge Univ. Press, Cambridge, 1985.
(31) Diudea, M. V. Unsymmetric Matrix $CJ_u$: Source of Various Graph Invariants. *Commun. Math. Comput. Chem.* (*MATCH)* **1997**, *35*, 000−000.
(32) Diudea, M. V.; Parv, B.; Topan, M. Derived Szeged and Cluj Indices. *J. Serb. Chem. Soc.* **1997**, *62*, 000−000.
(33) Gutman, I. Calculating the Wiener Number: The Doyle-Graver Method. *J. Serb. Chem. Soc.* **1993,** *58***,** 745−750.
(34) Randić, M. Molecular ID Numbers: by Design. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 134−136.
(35) Balaban, A. T. Numerical Modelling of Chemical Structures: Local Graph Invariants and Topological Indices. In *Graph Theory and Topology*; King, R. B., Rouvray, D. H., Eds.; Elsevier: Amsterdam, 1987; pp 413−415.
(36) Hall, L. H.; Kier, L. B. Determination of Topological Equivalence in Molecular Graphs from the Topological State. *Quant. Struct.-Act. Relat.* **1990**, *9*, 115−131.
(37) Knop, J. V.; Müller, W. R.; Szymaski, K.; Trinajstić, N. On the Identification Numbers for Chemical Structures. *Int. J. Quantum Chem.*; *Quantum Chem. Symp.* **1986**, *20*, 173−183.
(38) Müller, W. R.; Szymaski, K.; Knop, J. V.; Mihalić, Z.; Trinajstić, N. The Walk ID Numbers Revisited. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 231−233.
(39) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 3. New Identification Numbers for Chemical Structures: MINID and MINSID. *Croat. Chem. Acta* **1996**, *69*, 9−16.
(40) Hu, C. Y.; Xu, L. Algorithm for Computer Perception of Topological Symmetry. *Anal. Chem. Acta* **1994**, *295*, 127−134.
(41) Hu, C. Y.; Xu, L. On Highly Discriminating Molecular Topological Index. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 82−90.
(42) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(43) Filip, P. A.; Balaban, T. S.; Balaban, A. T. A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlation Ability. *J. Math. Chem.* **1987**, *1*, 61−83.
(44) Diudea, M. V.; Minailiuc, O.; Gutman, I. Fragmental Cluj and Szeged Indices. Manuscript in preparation.
(45) Diudea, M. V.; Randić, M. Unsymmetric Square Matrices and Schultz-Type Numbers. *J. Chem. Inf. Comput. Sci.* To be submitted.
(46) Beilsteins Handbook of Organic Chemistry, Band 1 and 5; Springer-Verlag: Berlin, 1925.

CI960365W