

# $D_{\text{MAX}}$ Matrix Invariants as Graph Descriptors. Graphs Having the Same Balaban Index $J$

Milan Randić<sup>a</sup>, Rok Orel<sup>b</sup>, Alexandru T. Balaban<sup>c</sup>

<sup>a</sup> National Institute of Chemistry, Hajdrihova 19, Ljubljana, Slovenia, SI-1000

<sup>b</sup> XLAB, Pot za Brdom 100, Ljubljana, Slovenia

<sup>c</sup> Texas A&M University at Galveston, Department of Marine Sciences, Galveston, Texas

77553

E-mail addresses: [mrandic@msn.com](mailto:mrandic@msn.com); [rok.orel@xlab.si](mailto:rok.orel@xlab.si); [balabana@tamug.edu](mailto:balabana@tamug.edu)

(Received May 23, 2012)

## Abstract

We explore the possibility of using selective invariants of a novel distance matrix  $D_{\text{MAX}}$  of graphs to discriminate between graphs. The  $D_{\text{MAX}}$  matrix contains information only on dominant distances of a graph, which are defined as those entries in a distance matrix that are the largest in any row or a column. For a test we selected a set of the smallest trees having the same Balaban index  $J$  (six pairs of trees on  $n = 12$  vertices). The Balaban index  $J$  is one of the early topological indices which has shown considerable power of discrimination between similar graphs. We have examined ordered sequences of row sums and a new topological index  $j$ , the  $D_{\text{MAX}}$  analog to Balaban index  $J$  as invariants to be tested for their discrimination power. This novel topological index  $j$  discriminated all 12 graphs having the same index  $J$ . However, neither  $J$  nor  $j$  can discriminate the pairs of acyclic graphs constructed by Slater, which are the smallest graphs (having 18 vertices) having the same distance degree sequence. In the search for alternative conceptually and structurally simple invariants we found that the ordered sequence of row sums of terminal matrices  $T$  and  $T_{\text{MAX}}$  do differentiate the two Slater graphs, which have identical  $D_{\text{MAX}}$ . We discuss the loss of information associated with  $D_{\text{MAX}}$  and end with a brief introduction of the concept of matrix projection, which is illustrated on terminal matrices of Slater graphs.

## Introduction

Recently a novel distance matrix of graphs,  $D_{MAX}$ , was introduced [1], which was constructed from the distance matrix  $D$  of graphs so that in each row and each column it retains dominant distances, while other elements of the distance matrix are set to be zero. The dominant distances are defined here as the largest distances in each row and each column of the distance matrix of a graph. For most graphs  $D_{MAX}$  is a sparse matrix; occasionally the  $D_{MAX}$  matrix may be intermediate (half way between sparse and dense matrix), while for some graphs, such as the complete graphs  $K_n$  and star graphs, the  $D_{MAX}$  has zero elements only on the main diagonal. The  $D_{MAX}$  matrix appears to be somewhat unique, in that often for apparently similar graphs it may have widely different forms. For molecular graphs in most cases structurally similar graphs have many similar graph invariants, but this need not be the case with  $D_{MAX}$  invariants of structurally similar graphs. As a result of focusing attention only to dominant distances, as we will see, most invariants of  $D_{MAX}$  matrices are not likely to be of interest in structure-property and structure-activity studies. One may say that they betray the paradigm that holds for many molecular properties that similar structures have similar properties. The apparently chaotic nature of  $D_{MAX}$  matrices, however, may be of interest in applications to graph isomorphism testing [2-5], because of their inherent simplicity. In view of the simple conceptual and computational aspects of working with  $D_{MAX}$ , in this article we will explore the potential of a selection of invariants of  $D_{MAX}$  as preliminary tools for discrimination between graphs,

The molecular connectivity  $\chi$  (Randić index) [6] is the sum of reciprocal square roots of vertex degrees products for adjacent vertices (eq. 1).

$$\chi = \sum_{\text{edges } ij} (a_i a_j)^{-1/2} \quad (1)$$

where  $a$ 's are vertex degrees (elements of the adjacency matrix, 1 for adjacent vertices and 0 otherwise) and the summation is carried out for all edges  $ij$ .

Because vertex degrees (row sums of the adjacency matrix) for carbon chains can have only values 1, 2, 3, and 4 the molecular connectivity has a degeneracy that is not much lower than that of the Wiener index: the smallest chemical trees with degenerate  $W$  and  $\chi$  values have 7 and 8 vertices, respectively. In the search for a topological index with higher discriminating ability, Balaban [7] proposed an index  $J$  (eq. 2) based on row sums of the distance matrix, for which the only limitation is that it must be an integer.

$$J = [E/(\mu + 1)] \sum_{\text{edges } ij} (d_i d_j)^{-1/2} \quad (2)$$

where  $d_i$ 's are row sums of the distance matrix,  $E$  is the number of edges, and  $\mu$  is the cyclomatic number. With the term in square brackets, index  $J$  does not automatically increase with the graph size, yielding for an infinite linear chain the interesting value  $J = \pi$  [8].

Indeed, the degeneracy is appreciably lower: smallest chemical trees with degenerate  $J$  are six pairs of dodecane isomers to be discussed in more detail below. Further investigations of the  $J$  index showed that it can readily accommodate the presence of multiple bonding and of heteroatoms [9]. In the latter case, chemical periodicity can become a criterion by opting for parameters based on electronegativities or on atomic radii.

### The distance matrices of 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (1)

Figure 1 presents one pair of dodecane isomers with the same  $J$  index, namely 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (2), and in Table 1 one may see the distance matrix of (1), from which the row sums ("distance sums" or "distasums") and the  $J$  index can be calculated.

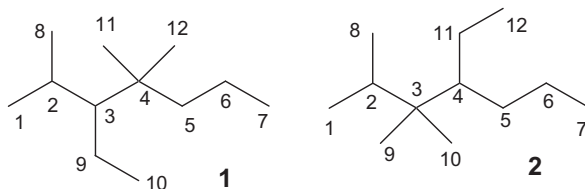


Fig. 1. Molecular graphs and labeling of carbon atoms in 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (2)

Table 1. Upper matrix: The distance matrices of 3-ethyl-2,4,4-trimethylheptane (**1**), its row sums (RS) and the reciprocal square roots of the row sums ( $1/\sqrt{RS}$ ). Lower matrix: same data for 4-ethyl-2,3,3-trimethylheptane (**2**)

	1	2	3	4	5	6	7	8	9	10	11	12	RS	$1/\sqrt{RS}$
1	0	1	2	3	4	5	6	2	3	4	4	4	38	0.1622
2	1	0	1	2	3	4	5	1	2	3	3	3	27	0.1925
3	2	1	0	1	2	3	4	2	1	2	2	2	19	0.2294
4	3	2	1	0	1	2	3	3	2	3	1	1	22	0.2132
5	4	3	2	1	0	1	2	4	3	4	2	2	28	0.1890
6	5	4	3	2	1	0	1	5	4	5	3	3	36	0.1667
7	6	5	4	3	2	1	0	6	5	6	4	4	46	0.1474
8	2	1	2	3	4	5	6	0	3	4	4	4	38	0.1622
9	3	2	1	2	3	4	5	3	0	1	3	3	30	0.1826
10	4	3	2	3	4	5	6	4	1	0	4	4	40	0.1581
11	4	3	2	1	2	3	4	4	3	4	0	2	32	0.1768
12	4	3	2	1	2	3	4	4	3	4	2	0	32	0.1768

	1	2	3	4	5	6	7	8	9	10	11	12	RS	$1/\sqrt{RS}$
1	0	1	2	3	4	5	6	2	3	3	4	5	38	0.1622
2	1	0	1	2	3	4	5	1	2	2	3	4	28	0.1890
3	2	1	0	1	2	3	4	2	1	1	2	3	22	0.2132
4	3	2	1	0	1	2	3	3	2	2	1	2	22	0.2132
5	4	3	2	1	0	1	2	4	3	3	2	3	28	0.1890
6	5	4	3	2	1	0	1	5	4	4	3	4	36	0.1667
7	6	5	4	3	2	1	0	6	5	5	4	5	46	0.1474
8	2	1	2	3	4	5	6	0	3	3	4	5	38	0.1622
9	3	2	1	2	3	4	5	3	0	2	3	4	32	0.1768
10	3	2	1	2	3	4	5	3	2	0	3	4	32	0.1768
11	4	3	2	1	2	3	4	4	3	3	0	1	30	0.1826
12	5	4	3	2	3	4	5	5	4	4	1	0	40	0.1581

In Table 2 one sees for each C–C bond in the degenerate isomer pair **1** and **2** the bond contribution  $(d_i d_j)^{-1/2}$  in the column labeled  $J$ .

Table 2. Bond contributions to two graphs having the same Balaban  $J$  index: 3-ethyl-2,4,4-trimethylheptane (**1**) and 4-ethyl-2,3,3-trimethylheptane (**2**), and the corresponding index based on  $D_{\text{MAX}}$  matrix (column  $j$ )

<b>1</b>			<b>2</b>		
C—C Bond	$J$	$j$	C—C Bond	$J$	$j$
1—2	0.03122	0.08771	1—2	0.03122	0.09759
2—3	0.04452	0.22361	2—3	0.04452	0.22361
2—8	0.03122	0.08775	2—8	0.03122	0.08607
3—4	0.04891	0.14434	3—4	0.04891	0.16667
3—9	0.04189	0.09129	3—9	0.03769	0.22361
4—5	0.04029	0.08334	3—10	0.03769	0.22361
4—11	0.03769	0.07217	4—5	0.04029	0.11785
4—12	0.03769	0.07217	4—11	0.04189	0.09623
5—6	0.03150	0.07454	5—6	0.0315	0.1118
6—7	0.02457	0.03938	6—7	0.02457	0.04823
9—10	0.02887	0.03582	11—12	0.02887	0.07544
Sum	0.39836	1.01209	Sum	0.39836	1.47068

The remaining five pairs of dodecane isomers with pairwise identical  $J$  indices (**3–12**) are shown in Fig. 2. Thus, for all the twelve graphs in Figures 1 and 2, the Balaban, Wiener and Randić indices are pairwise the same.

In Table 1 we showed the distance matrix of 3-ethyl-2,4,4-trimethylheptane (**1**) and its row sums. In the last column of Table 1 are shown the reciprocal square root values of the row sums, the quantities used in the construction of the Balaban index  $J$ . Index  $J$  is bond additive, with bond contributions being defined as the product of the reciprocal square root values of the row sums of end vertices of each bond. In Table 2 in the column  $J$  we showed for 3-ethyl-2,4,4-trimethylheptane (**1**) the bond contributions that when added give  $J$ , which is equal to 0.3984.

The same value is obtained for 4-ethyl-2,3,3-trimethylheptane (**2**), the bond contributions of which are shown for the column  $J$  in the right-hand part of Table 2. A close look at Table 2 shows that all bond terms which contribute to the construction of  $J$  are the same in both molecules even though the methyl and ethyl groups in the two molecules are situated at different locations. In the next section we will consider the  $D_{\text{MAX}}$  matrices of 3-ethyl-2,4,4-trimethylheptane (**1**) and 4-ethyl-2,3,3-trimethylheptane (**2**).

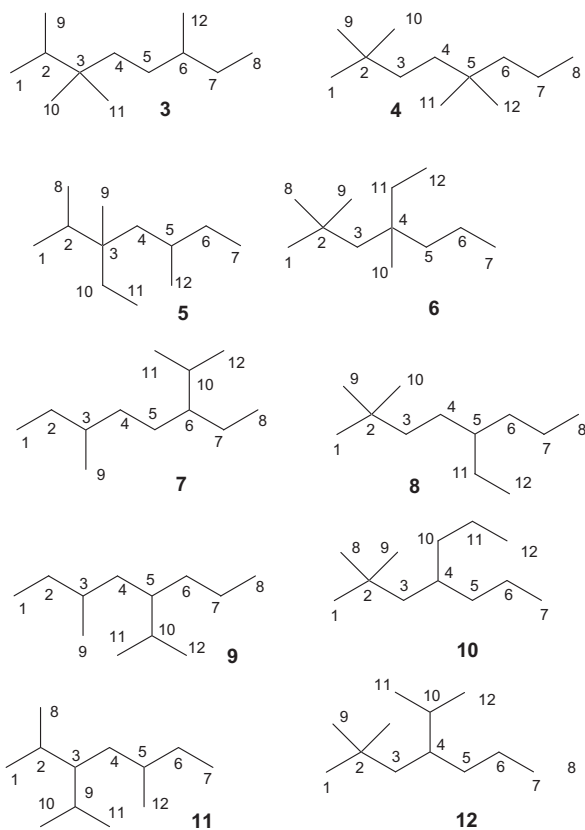


Fig. 2. The remaining isomers of *n*-dodecane (3–12) having pairwise the same Balaban index *J*,

### The $D_{MAX}$ matrices of 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (1)

If in Table 1 one selects only the dominant distances, which are the distances which are maximal in each row and each column of the distance matrix, and set all other matrix elements to be zero one obtains of  $D_{MAX}$  matrix of 3-ethyl-2,4,4-trimethylheptane (1), which is shown in Table 3. In Table 4, for comparison we show the  $D_{MAX}$  matrix of 4-ethyl-2,3,3-

trimethylheptane (**2**). One can immediately see that some row sums of the two matrices are different, those of **2** being visibly smaller than the corresponding row sums of **1**. This makes it less likely, or unlikely, that their sum and other quantities constructed from the row sum will be the same for the two molecules. Indeed, as one can see from the last row in Table 2, the graph invariant  $j$ , which is analogous to the Balaban index  $J$ , but is obtained from  $D_{MAX}$ , rather than the distance matrix, equals 1.0121 for **1** and 1.4707 for **2**. These row sums are of considerably different magnitude. This illustrates the fact that even though the two molecules are fairly similar, some properties (invariants) derived from  $D_{MAX}$  matrices can be quite different. This is welcome news, because it suggests that the  $D_{MAX}$  matrix may be promising tool for graph isomorphism testing.

Table 3. The  $D_{MAX}$  matrix of 3-ethyl-2,4,4-trimethylheptane (**1**), its row sums and the reciprocal square roots of the row sums

	1	2	3	4	5	6	7	8	9	10	11	12	RS	1/√RS
1	0	0	0	3	4	5	6	0	0	0	4	4	26	0.1961
2	0	0	0	0	0	0	5	0	0	0	0	0	5	0.4472
3	0	0	0	0	0	0	4	0	0	0	0	0	4	0.5000
4	3	0	0	0	0	0	3	3	0	3	0	0	12	0.2887
5	4	0	0	0	0	0	0	4	0	4	0	0	12	0.2887
6	5	0	0	0	0	0	0	5	0	5	0	0	15	0.2582
7	6	5	4	3	0	0	0	6	5	6	4	4	43	0.1525
8	0	0	0	3	4	5	6	0	0	0	4	4	26	0.1961
9	0	0	0	0	0	0	5	0	0	0	0	0	5	0.1826
10	0	0	0	3	4	5	6	0	0	0	4	4	26	0.1961
11	4	0	0	0	0	0	4	4	0	4	0	0	16	0.2500
12	4	0	0	0	0	0	4	4	0	4	0	0	16	0.2500

As we see  $J$  and  $j$  invariants provide one such straightforward route for testing graph isomorphism. Since 3-ethyl-2,4,4-trimethylheptane (**1**) and 2,3,3-trimethyl-3-ethylheptane (**2**) have several row sums of  $D_{MAX}$  that do not match the sequences of their ordered row sums are also of interest to examine as an additional property of  $D_{MAX}$  to test graphs for isomorphism. Below we show the ordered sequences of row sums for the two molecules:

3-ethyl-2,4,4-trimethylheptane (**1**): (43, 26, 26, 26, 16, 16, 15, 12, 12, 5, 5, 4)

4-ethyl-2,3,3-trimethylheptane (**2**): (43, 27, 21, 15, 12, 10, 9, 8, 5, 5, 5, 4)

The row sum can be easily calculated and are thus accessible to everyone. If all the entries in the above sequences are added and divided by two one obtains the sum of matrix elements of  $D_{MAX}$  matrix above the main diagonal. This quantity is analogous to the well known

graph invariant, the Wiener number  $W$  [10-21] when one considers the distance matrix, which, as Hosoya was first to show [22] can be obtained by summing the matrix elements above the main diagonal of  $D$ . For the above two molecules  $W$  is necessarily the same, because their row sums have the same matching numerical values, being 194. However, the corresponding quantity based on the  $D_{MAX}$  matrix, which we will denote as  $w$ , are different, being 103 and 83 for (1) and (2), respectively. This reflects the fact that the average non-zero matrix element of  $D_{MAX}$  matrices for 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (2) are different, being 4.2917 and 4.3158, respectively. Hence, in this case already  $w$  differentiates the two isomers of decane.

When one considers CC bond types of the two molecules according to vertex degrees of endpoints for each edge, one can see that they are also different:

3-ethyl-2,4,4-trimethylheptane (1):  $2(1,2) + 2(1,3) + 2(1,4) + (2,2) + (2,3) + (2,4) + (3,3) + (3,4)$

4-ethyl-2,3,3-trimethylheptane (2):  $2(1,2) + 2(1,3) + 2(1,4) + (2,2) + 2(2,3) + 2(3,3)$

Here the symbol  $(m,n)$  signifies a CC bond between carbon atoms having  $m$  and  $n$  adjacent carbon atoms. Because molecules have different bond types, many topological indices for the two molecules, including the connectivity  $\chi$  index, may be different. So it may be surprising that Balaban's index  $J$  (which generally has a much better discriminatory power than many simple topological indices such as  $W$  or  $\chi$ ) does not discriminate between (1) and (2). The reason for this is that  $J$  is based on row sums, and whenever summation is used in construction of invariants, there is loss of information.



Table 4. The  $D_{MAX}$  matrix of 4-ethyl-2,3,3-trimethylheptane (2), its row sums and the reciprocal square roots of the row sums

	1	2	3	4	5	6	7	8	9	10	11	12	RS	1/√RS
1	0	0	0	3	4	5	6	0	0	0	4	5	27	0.2182
2	0	0	0	0	0	0	5	0	0	0	0	0	5	0.4472
3	0	0	0	0	0	0	4	0	0	0	0	0	4	0.5000
4	3	0	0	0	0	0	3	3	0	0	0	0	9	0.3333
5	4	0	0	0	0	0	0	4	0	0	0	0	8	0.3536
6	5	0	0	0	0	0	0	5	0	0	0	0	10	0.3162
7	6	5	4	3	0	0	0	6	5	5	4	5	43	0.1525
8	0	0	0	3	4	5	6	0	0	0	4	5	27	0.1925
9	0	0	0	0	0	0	5	0	0	0	0	0	5	0.4472
10	0	0	0	0	0	0	5	0	0	0	0	0	5	0.4472
11	4	0	0	0	0	0	4	4	0	0	0	0	12	0.2887
12	5	0	0	0	0	0	5	5	0	0	0	0	15	0.2582

### Dodecane isomers having the same $J$

Figures 1 and 2 illustrate all dodecane isomers having the same  $J$ . Only the pair of dodecane isomers of Fig. 1 (considered in the previous section) has the same alkyl groups off-the-main chain, though located at different sites. The other five pairs of molecules (Fig. 2) have *different* alkyl groups along the main chain of the molecule. In a way they are more diverse and less similar, which should make it easier to differentiate them, and one may therefore expect that they will have fewer common invariants. It is therefore of interest to see how different are their sequences of ordered row sums and whether the sum of entries in  $D_{MAX}$  for pairs of these molecule lead to different  $w$  values. In Table 5 we collected information on the sequences of row sums and their  $w$  values.

Observe that already the leading row sum in all cases, except for the pair of Fig. 1 (row 3 and row 4 of Table 5) are different, indicating the larger differences among the molecules constituting pairs of graphs having the same  $J$  index than those found in 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (2). Observe also how in some cases variations of  $w$  (1/2 row sums) between the isomers having the same  $J$  can vary quite widely, the smallest difference being 9 and the largest 87, for the pair (7, 8) and the pair (9, 10) graphs of Table 5, respectively. All this is indicative that the  $D_{MAX}$  matrix and its various invariants can be useful as an additional instrument in testing of graphs for isomorphism, including the invariants considered here: ordered row sums and  $j$ . However, without further

investigations of  $D_{MAX}$  and its invariants on a wider class of graphs it would be premature to speculate on the potential of  $D_{MAX}$  and its invariants as tools for graph isomorphism tests. Recall that there are graphs for which the isomorphism test is very tedious. Such are, for instance, Mathon graphs [23,24], which are of interest in communication networks. Of interest here are primarily molecular graphs; their size and complexities are moderate, although occasionally they can be high. It is of interest to search for sets of invariants of molecular graphs, which can be used in simple procedures for the isomorphism problem.

Table 5. Row sums for the six pairs of trees of Fig. 1 and Fig. 2 having the same indices  $J$

Row →	1	2	3	4	5	6	7	8	9	10	11	12	$w$	$j$
<b>1</b>	26	5	4	12	12	15	43	26	5	26	16	16	103	1.1963
<b>2</b>	27	5	4	9	8	10	43	27	5	5	12	15	85	1.4583
<b>3</b>	28	6	5	4	8	10	12	41	28	6	6	12	83	1.4419
<b>4</b>	32	6	5	4	12	15	18	36	32	32	15	15	111	1.0908
<b>5</b>	23	5	4	12	12	15	40	23	5	5	23	15	91	1.3681
<b>6</b>	31	5	4	12	12	15	43	31	31	16	16	20	118	1.0064
<b>7</b>	42	18	15	12	4	5	6	28	18	6	28	28	105	1.1972
<b>8</b>	33	6	5	4	12	15	18	36	33	33	15	18	114	1.0738
<b>9</b>	39	6	5	4	4	5	6	28	6	5	6	6	60	1.9762
<b>10</b>	33	10	8	15	16	20	45	33	33	16	20	45	147	0.6758
<b>11</b>	23	5	4	15	16	20	41	23	5	23	23	20	109	1.1606
<b>12</b>	32	5	4	12	12	15	44	32	32	16	20	20	122	0.9859

One such set of few molecular descriptors is the quartet: Balaban index  $J$  obtained from the distance matrix; the ordered row sums ( $\Sigma$ ) of the distance matrix; the modified Balaban index  $j$  of the  $D_{MAX}$  matrix; and the ordered row sums ( $\sigma$ ) of the  $D_{MAX}$  matrix. Whether this will be true or not will depend on how well these four descriptors differentiate “difficult” graphs for isomorphism tests. In the next section we will consider one such pair of difficult trees, the Slater graphs of Fig. 3.

## Slater graphs

Slater graphs, the smallest pair of trees having the same distance degree sequences. The distance degree sequence of a vertex is the ordered list for the number of vertices at increasing distances (0, 1, 2, 3, etc.) from the selected vertex. Thus for the graphs at left in

Fig. 3 the distance degree sequence for the four leftmost terminal vertices is:  $(1, 1, 1, 4, 7, 4)$ . The first entry by definition is equal 1. The rest of the sequence shows, that the selected vertices have one adjacent vertex and one vertex at distance two, four vertices at distances three and five and seven vertices at distance four. When such sequences for all vertices of a graph are ordered lexicographically one obtains the distance degree sequences of a graph.

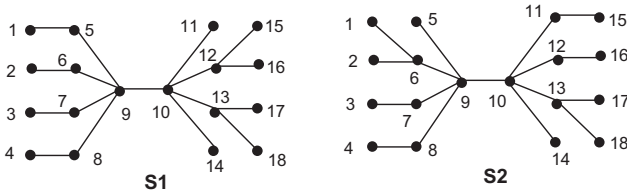


Fig. 3. Slater graphs, **S1** and **S2**, the smallest pair of trees having the same distance degree sequences.

In Table 6 for both Slater graphs we have listed all distance sequences for all six classes of symmetry distinct vertices, while the numbers of vertices having the same sequence are shown as exponents. There may be not many such acyclic graphs, particularly among acyclic graphs for which the distance degree sequences are very often different. Clearly the Balaban index  $J$  cannot discriminate the two graphs. However, as we will see, neither can the corresponding novel invariants  $j$  (the analog of the Balaban index  $J$  but based on the  $D_{MAX}$  matrix rather than the distance matrix) nor can the ordered row sums  $ORS$  (the sum of elements above the main diagonal of  $D_{MAX}$  matrix).

Table 6. Distance sequences for the six classes of symmetry distinct vertices in Slater graphs

Sequence	Vertices in $S_1$	Vertices in $S_2$
$(1, 1, 1, 4, 7, 4)^4$	1, 2, 3, 4	3, 4, 15, 16
$(1, 2, 4, 7, 4)^4$	5, 6, 7, 8	7, 8, 11, 12
$(1, 5, 8, 4)^2$	9, 10	9, 10
$(1, 1, 4, 8, 4)^2$	11, 14	5, 14
$(1, 3, 4, 6, 4)^2$	12, 13	6, 13
$(1, 1, 2, 4, 6, 4)^2$	15, 16, 17, 18	1, 2, 17, 18

In Table 7 we show the  $D_{MAX}$  matrix of Slater graph  $S_1$ . Observe that even though two graphs have different distance matrices the  $D_{MAX}$  matrix is the same for both graphs. This happens because the two graphs differ in distances which are not dominant, and which are therefore eliminated in construction of  $D_{MAX}$  matrix. The pair of Slater graphs is therefore a counterexample to expectations that  $D_{MAX}$  matrix may be unique to a graph and that graphs can always be reconstructed from  $D_{MAX}$ . From Table 7 one can easily find the sequence of ordered row sums (for both Slater graphs) is:

$$(39, 39, 39, 39, 39, 39, 39, 39, 39, 16, 16, 16, 16, 16, 16, 16, 16, 12, 12).$$

Table 7.  $D_{MAX}$  matrix of Slater graph  $S_1$  and  $S_2$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	0	0	0	0	0	0	0	0	3	4	4	4	4	5	5	5	5
2	0	0	0	0	0	0	0	0	0	3	4	4	4	4	5	5	5	5
3	0	0	0	0	0	0	0	0	0	3	4	4	4	4	5	5	5	5
4	0	0	0	0	0	0	0	0	0	3	4	4	4	4	5	5	5	5
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	4
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	4
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	4
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	4
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	3	3
10	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	5	5	5	5	4	4	4	4	3	0	0	0	0	0	0	0	0	0
16	5	5	5	5	4	4	4	4	3	0	0	0	0	0	0	0	0	0
17	5	5	5	5	4	4	4	4	3	0	0	0	0	0	0	0	0	0
18	5	5	5	5	4	4	4	4	3	0	0	0	0	0	0	0	0	0

Thus the ordered row sums are also of no use for graphs having the same distance degree sequences. Here obviously we have a problem. The question is whether we can find *simple* conceptual and computational solutions to this problem that can supply additional invariants capable of discriminating between graphs having the same distance degree sequences? By *simple* we consider here solutions that can be obtained using pencil, paper, and a hand calculator, as has been the case with the hitherto considered quartet of invariants  $(J, \Sigma, j, \sigma)$ .

We will present in the next section one such solution, which follows from similar consideration of terminal matrix of acyclic graphs, which, as follows from a theorem of Zarestkii [25], is not accompanied with loss of information on graphs, although the matrix considered contains only information of distances between terminal vertices.

## Terminal Matrix $T$ and Terminal Matrix $T_{MAX}$

Terminal matrices have received limited attention in chemistry, most though in mathematical chemistry [26-30], although they may have some a wider applications in mathematical modeling in chemistry. Thus, it was shown that the motor and research octane numbers (MON and RON, respectively) for all isomers of heptane and octane revealed better correlation with the *endpoint mean square distances* obtained only from terminal vertices than the corresponding ones involving all graph vertices [31].

From the distance matrix  $D$  of an acyclic graph it is straightforward to arrive at the terminal matrix  $T$ , simply by ignoring all rows and columns that do not belong to terminal vertices. In the case of Table 1, which is the distance matrix of 3-ethyl-2,4,4-trimethylheptane (**1**), one should erase rows and columns 2 – 6, and 9, which belong to secondary, tertiary and quaternary carbon atoms and obtain a 6×6 matrix. In Table 9 we show terminal matrices for the two dodecane isomers of Fig. 1, merely to illustrate that terminal matrices for similar graphs, such as are molecular graphs of 3-ethyl-2,4,4-trimethylheptane (**1**) and 4-ethyl-2,3,3-trimethylheptane (**2**), can be visibly different, similarly to matrices involving dominant distances of graphs.

In the case of Slater graphs  $S_1$  and  $S_2$ , terminal matrices  $T$  and  $T_{MAX}$  are of size 10×10 as shown in Table 9.

For the two matrices the ordered row sums are:

$$S_1 = (40, 40, 40, 40, 36, 36, 36, 36, 30, 30)$$

$$S_2 = (37, 37, 37, 37, 35, 35, 35, 35, 31, 31),$$

which are clearly different. It is interesting that already the average matrix elements for the two graphs are also different, being 3.64 and 3.58, respectively, or alternatively the sum of elements above the main diagonal (corresponding to the Wiener index  $W$  of the distance matrix), which are 182 and 179, respectively are also different.

Table 8. Terminal matrices T (top) and terminal matrices having only dominant distances  $T_{MAX}$  (bottom) for 3-ethyl-2,4,4-trimethylheptane (1) and 4-ethyl-2,3,3-trimethylheptane (2)

$S_1$							$S_2$						
	1	7	8	10	11	12		1	7	8	9	10	12
1	0	6	2	4	4	4	1	0	6	2	3	3	5
7	6	0	6	6	4	4	7	6	0	1	2	2	4
8	2	6	0	4	4	4	8	2	1	0	3	3	5
10	4	6	4	0	4	4	9	3	2	3	0	2	4
11	4	4	4	4	0	2	10	3	2	3	2	0	4
12	4	4	4	4	2	0	12	5	4	5	4	4	0
	1	7	8	10	11	12		1	2	3	4	5	6
1	0	6	0	0	4	4	1	0	6	0	0	0	0
7	6	0	6	6	4	4	2	6	0	0	0	0	0
8	0	6	0	0	4	4	3	0	0	0	0	0	5
10	4	6	0	0	4	4	4	0	0	0	0	0	4
11	4	4	4	4	0	0	5	0	0	0	0	0	4
12	4	4	4	4	0	0	6	0	0	5	4	4	0

Finally, the same is true for the  $T_{MAX}$  matrices of Slater graphs  $S_1$  and  $S_2$ . The ordered row sums are:

$$s_1 = (28, 28, 28, 28, 20, 20, 20, 20, 16, 16)$$

$$s_2 = (24, 24, 24, 24, 24, 24, 24, 24, 16, 16).$$

As one can see in this case the average matrix elements and the sum of elements above the main diagonal are the same for both matrices, even though the ordered row sum sequences are different.

Hence, if we increase the quartet of invariants  $(J, \Sigma, j, \sigma)$  to a quintet:  $(J, \Sigma, j, \sigma, T\Sigma)$ , where  $T\Sigma$  is the ordered row sum of the terminal matrix and  $T_{MAX}$  matrix. Of course, the first index which is different suffices to discriminate between graphs, and it is only when the four quartet indices are all same in tow graphs that  $T\Sigma$  should be considered. The  $T_{MAX}$  matrix offers also additional invariants that can be kept as a reserve if one finds all five descriptors

to be the same. Alternatively, in these, what one can expect rare occasions, one can always resort to canonical numbering of vertices, such as labeling of vertices that make the adjacency matrix, when its rows are red from left to right and from top to bottom, the smallest binary number [32,33].

Table 9. Terminal matrix T of Slater graphs  $S_1$  (top left) and  $S_2$  (top right) and  $T_{MAX}$  (bottom left and bottom right), respectively

	1	2	3	4	11	14	15	16	17	18
1	0	4	4	4	4	4	5	5	5	5
2	4	0	4	4	4	4	5	5	5	5
3	4	4	0	4	4	4	5	5	5	5
4	4	4	4	0	4	4	5	5	5	5
11	4	4	4	4	0	2	3	3	3	3
14	4	4	4	4	2	0	3	3	3	3
15	5	5	5	5	3	3	0	2	4	4
16	5	5	5	5	3	3	2	0	4	4
17	5	5	5	5	3	3	4	4	0	2
18	5	5	5	5	3	3	4	4	2	0

	1	2	3	4	5	14	15	16	17	18
1	0	2	4	4	3	4	5	5	5	5
2	2	0	4	4	3	4	5	5	5	5
3	4	4	0	4	3	4	5	5	5	5
4	4	4	4	0	3	4	5	5	5	5
5	3	3	3	3	0	3	4	4	4	4
14	4	4	4	4	3	0	3	3	3	3
15	5	5	5	5	4	3	0	4	4	4
16	5	5	5	5	4	3	4	0	4	4
17	5	5	5	5	4	3	4	4	0	2
18	5	5	5	5	4	3	4	4	2	0

	1	2	3	4	11	14	15	16	17	18
1	0	0	0	0	4	4	5	5	5	5
2	0	0	0	0	4	4	5	5	5	5
3	0	0	0	0	4	4	5	5	5	5
4	0	0	0	0	4	4	5	5	5	5
11	4	4	4	4	0	0	0	0	0	0
14	4	4	4	4	0	0	0	0	0	0
15	5	5	5	5	0	0	0	0	0	0
16	5	5	5	5	0	0	0	0	0	0
17	5	5	5	5	0	0	0	0	0	0
18	5	5	5	5	0	0	0	0	0	0

	1	2	3	4	5	14	15	16	17	18
1	0	0	0	0	0	4	5	5	5	5
2	0	0	0	0	0	4	5	5	5	5
3	0	0	0	0	0	4	5	5	5	5
4	0	0	0	0	0	4	5	5	5	5
5	0	0	0	0	0	0	4	4	4	4
14	4	4	4	4	0	0	0	0	0	0
15	5	5	5	5	4	0	0	0	0	0
16	5	5	5	5	4	0	0	0	0	0
17	5	5	5	5	4	0	0	0	0	0
18	5	5	5	5	4	0	0	0	0	0

Looking at the two matrices at bottom part of Table 9 one can even without calculating their various invariants establish that the matrices are fundamentally different and that they belong to different graphs, as is outlined in the next section.

### Matrix Projections

Comparison of matrices is rather difficult, because such matrices can belong to the same graph the vertices of which are labeled differently. However, in some cases, as will be seen in the next section, matrices can be associated with distinct connectivity. Let us introduce a novel concept: *Matrix Projection*, which we define as follows:

*Definition:* Matrix projection is a binary matrix obtained from a non-binary matrix in which all non-zero matrix elements are replaced by 1 (one).

Hence, adjacency matrix of any graph can be viewed as its own projection. The projection of any matrix having all matrix elements, except on the main diagonal, zero is trivially the adjacency matrix of the complete graph having the number of vertices same as the number of rows and columns in the matrix. The projection of  $T_{MAX}$  matrices of the two Slater graphs are shown in Table 10.

Table 10 Projection matrices of  $T_{MAX}$  matrices of Slater graphs  $S_1$  and  $S_2$

$S_1$											$S_2$											
	1	2	3	4	11	14	15	16	17	18		1	2	3	4	5	14	15	16	17	18	
1	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
2	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
3	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
4	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
11	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

It is apparent that the projection of  $S_1$  is the adjacency matrix of complete bipartite graph  $K_{4,6}$  while the projection of  $S_2$  is the adjacency matrix of bipartite graph  $(K_{5,5} - e)$  obtained from  $K_{5,5}$  in which one edge has been deleted, both shown in Fig. 4. We can refer to the two graphs of Fig. 4 as  $T_{MAX}$  projection of Slater graphs. It remains to be investigated how useful may be the concept of matrix projection in other applications.



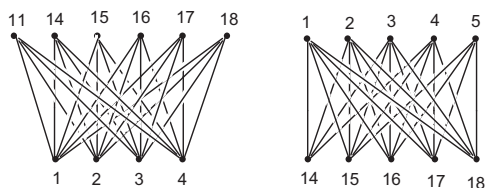


Fig. 4. The two bipartite graphs representing the  $T_{MAX}$  projection of Slater graphs  $S_1$  and  $S_2$

## Concluding Remarks

What makes  $D_{MAX}$  attractive, in comparison to some more elaborate molecular descriptors is its conceptual elegance and simplicity and the relative ease of construction, which border with the elegance and the simplicity of some computational approaches to molecules and molecular properties that have been characterized as offering solution by using pencil and paper, such as is the recent approach to calculations of ring currents in polycyclic conjugated hydrocarbons [34, 35], using the concept of conjugated circuits [36-44], which produce numerical results that closely parallel *ab initio* computations or ring currents [45-50].

It is not that one is striving for simple isomorphism test that will be absolute, but rather one is trying to arrive at simple isomorphism test that offer *prescreening* and would eliminate need for detailed and elaborate tests on large number of cases. In that respect  $D_{MAX}$  matrix and its selected invariants appears promising. It therefore is of interest to see the discriminatory power of the index  $j$ , which combined with index  $J$ , may suffice to resolve isomorphism issue for most cases of molecular graphs.

Finally, we may add that the notion of  $D_{MAX}$  is general enough that it may equally apply to other matrices of interest in chemistry [51-53], starting with other distance-related matrices, such as the Resistance-distance [54, 55], the Line distance [56, 57],  $N$ -Dimensional distance matrix [58], Path matrices [59], and various non-distance matrices.

## Acknowledgments

Rok Orel would like to thank for the financial support the European Union, European Social Fund. Milan Randić would like to thank the Laboratory for Chemometric, National Institute of Chemistry, Ljubljana, Slovenia, for hospitality and partial financial support. This work has been supported in part by the Ministry of Science and Higher Education, of Republic of Slovenia under Research Grant P1017.

## References

- [1] M. Randić,  $D_{\text{MAX}}$  novel distance matrix, *MATCH Commun. Math. Comput. Chem.* **70** (2013) 221–238.
- [2] R. C. Read, D. C. Corneil, The graph isomorphism disease, *J. Graph Theory* **1** (1977) 339–363.
- [3] B. Weisfeiler, *On Construction and Identification of Graphs*, Springer, Berlin, 1976.
- [4] G. Gati, Further annotated bibliography on the isomorphism disease, *J. Graph Theory* **3** (1979) 95–109.
- [5] X. Liu, D. J. Klein, The graph isomorphism problem, *J. Comput. Chem.* **12** (1991) 1243–1251.
- [6] M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
- [7] A. T. Balaban, Highly discriminating distance-based topological index, *Chem. Phys. Lett.* **89** (1982) 399–404.
- [8] A. T. Balaban, L. V. Quintas, The smallest graphs, trees and 4-trees with degenerate topological index  $J$ , *MATCH Commun. Math. Comput. Chem.* **14** (1983) 213–233.
- [9] A. T. Balaban, Topological indices based on topological distances in molecular graphs, *Pure Appl. Chem.* **55** (1983) 199–206.
- [10] O. Ivanciuc, T. Ivanciuc, A. T. Balaban, Design of topological indices. Part 10. Parameters based on electronegativity and covalent radius for the computation of molecular graph descriptors for heteroatom-containing molecules, *J. Chem. Inf. Comput. Sci.* **38** (1998) 395–401, and further references therein.
- [11] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.
- [12] A. Graovac, T. Pisanski, On the Wiener index of a graph, *J. Math. Chem.* **8** (1991) 53–62.
- [13] M. Randić, Novel molecular descriptor for structure-property studies, *Chem. Phys. Lett.* **211** (1993) 478–483.
- [14] I. Gutman, Y. N. Yeh, S. L. Lee, Y. L. Luo, Some recent results in the theory of the Wiener number, *Indian J. Chem.* **32A** (1993) 651–661.
- [15] I. Gutman, A formula for the Wiener number of trees and its extension to graphs containing cycles, *Graph Theory Notes New York* **27** (1994) 9–15.
- [16] I. Lukovits, W. Linert, A novel definition of the hyper-Wiener index for cycles, *J. Chem. Inf. Comput. Sci.* **34** (1994) 899–902.
- [17] P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. A. Dobrynin, I. Gutman, G. Dömötör, The Szeged index and an analogy with the Wiener index, *J. Chem. Inf. Comput. Sci.* **35** (1995) 547–550.
- [18] D. J. Klein, I. Lukovits, I. Gutman, On the definition of the hyper-Wiener index for cycle-containing structures, *J. Chem. Inf. Comput. Sci.* **35** (1995) 50–52.
- [19] A. A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: theory and applications, *Acta Appl. Math.* **66** (2001) 211–249.
- [20] M. Randić, On generalization of Wiener index for cyclic structures, *Acta Chim. Slov.* **49** (2002) 483–496.
- [21] T. Pisanski, M. Randić, Use of the Szeged index and the revised Szeged index for measuring network bipartivity, *Discr. Appl. Math.* **158** (2010) 1936–1944.

- [22] H. Hosoya, Topological index. A newly proposed quantity characterizing topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.
- [23] R. Mathon, Sample graphs for isomorphism testing, *Proc. 9th S-E. Conf. Combinatorics, Graph Theory and Computing*, 1980.
- [24] R. Mathon, On self-complementary strongly regular graphs, *Discr. Math.* **69** (1988) 263–281.
- [25] K. A. Zaretskii, Constructing a tree on the basis of a set of distances between the hanging vertices, *Uspekhi Mat. Nauk* **20** (1965) 90–92 (in Russian).
- [26] M. Randić, J. Zupan, D. Vikić-Topić, Graphical representation of proteins by star-like graphs, *J. Mol. Graphics Model.* **26** (2007) 290–305.
- [27] B. Horvat, T. Pisanski, M. Randić, Terminal polynomials and star-like graphs, *MATCH Commun. Math. Comput. Chem.* **60** (2008) 493–512.
- [28] E. A. Smolenskii, E. V. Shuvalova, L. K. Maslova, I. V. Chuvaeva, M. S. Molchanova, Reduced matrix of topological distances with a minimum number of independent parameters: distance vectors and molecular codes, *J. Math. Chem.* **45** (2009) 1004–1020.
- [29] I. Gutman, B. Furtula, M. Petrović, Terminal Wiener index, *J. Math. Chem.* **46** (2009) 522–531.
- [30] I. Gutman, B. Furtula, A survey on terminal Wiener index, in: I. Gutman, B. Furtula (Eds.), *Novel Molecular Structure Descriptors – Theory and Applications*, Univ. Kragujevac, Kragujevac, 2010, pp. 173–190.
- [31] A. T. Balaban, Topological indices based on topological distances in molecular graphs, *Pure Appl. Chem.* **55** (1983) 199–206.
- [32] M. Randić, On the recognition of identical graphs representing molecular topology, *J. Chem. Phys.* **60** (1974) 3920–3928.
- [33] M. Randić, On canonical numbering of atoms in a molecule and graph isomorphism, *J. Chem. Inf. Comput. Sci.* **17** (1977) 171–180.
- [34] M. Randić, Graph theoretical approach to  $\pi$ -electron currents in polycyclic conjugated hydrocarbons, *Chem. Phys. Lett.* **500** (2010) 123–127.
- [35] M. Randić, M. Novič, M. Vračko, D. Vukičević, D. Plavšić,  $\pi$ -electron currents in polycyclic conjugated hydrocarbons: Coronene and its isomers having five and seven member rings, *Int. J. Quantum Chem.* **112** (2012) 972–985.
- [36] M. Randić, Conjugated circuits and resonance energies of benzenoid hydrocarbons, *Chem. Phys. Lett.* **38** (1976) 68–70.
- [37] M. Randić, Aromaticity and conjugation, *J. Am. Chem. Soc.* **99** (1977) 444–450.
- [38] M. Randić, A graph theoretical approach to conjugation and resonance energies of hydrocarbons, *Tetrahedron* **33** (1977) 1905–1920.
- [39] The first publication on resonance graphs was by W. Gründler, *Wiss. Z. Univ. Halle* **31** (1982) 97. We thank Professor I. Gutman (Kragujevac, Serbia) who drew our attention to this paper.
- [40] M. Randić, D. J. Klein, S. El-Basil, P. Calkins, Resonance in large benzenoid hydrocarbons, *Croat. Chem. Acta* **69** (1996) 1639–1660.
- [41] M. Randić, Resonance in catacondensed benzenoid hydrocarbons, *Int. J. Quantum Chem.* **63** (1997) 585–600.
- [42] M. Randić, Aromaticity of polycyclic conjugated hydrocarbons, *Chem. Rev.* **103** (2003) 3449–3605.

- [43] M. Randić, A. T. Balaban, D. Plavšić, Applying the conjugated circuits method to Clar structures of [*n*]phenylenes for determining resonance energies, *Phys. Chem. Chem. Phys.* **13** (2011) 20644–20648.
- [44] A. T. Balaban, M. Randić, Structural approach to aromaticity and local aromaticity in conjugated polycyclic systems, in: M. H. Putz (Ed.), *Carbon Bonding and Structures, Advances in Physics and Chemistry*, Springer, Dordrecht, 2011, pp. 159–204.
- [45] A. T. Balaban, D. E. Bean, P. W. Fowler, Patterns of ring current in coronene isomers, *Acta Chem. Slov.* **57** (2010) 507–512.
- [46] E. Steiner, P. W. Fowler, A. Soncini, L. W. Jenneskens, Current–density maps as probes of aromaticity: Global and Clar ring currents in totally resonant polycyclic aromatic hydrocarbons, *Faraday Disc.* **135** (2007) 309–323.
- [47] P. W. Fowler, E. Steiner, Pseudo–currents: rapid and accurate visualisation of ring currents in conjugated hydrocarbons, *Chem. Phys. Lett.* **364** (2002) 259–266.
- [48] E. Steiner, P. W. Fowler, Patterns of ring currents in conjugated molecules: A few–electron model based on orbital contributions, *J. Phys. Chem.* **105** (2001) 9553–9562.
- [49] E. Steiner, P. W. Fowler, L. W. Jenneskens, Counter–rotating ring currents in coronene and corannulene, *Angew. Chemie Int. Ed.* **40** (2001) 362–366.
- [50] E. Steiner, P. W. Fowler, Four– and two–electron rules for diatropic and paratropic ring currents in monocyclic  $\pi$  systems, *Chem. Commun.* (2001) 2220–2221.
- [51] Z. Mihalić, D. Veljan, D. Amić, S. Nikolić, D. Plavšić, N. Trinajstić, The distance matrix in chemistry, *J. Math. Chem.* **11** (1992) 223–258.
- [52] Z. Mihalić, S. Nikolić, N. Trinajstić, Comparative study of molecular descriptors derived from the distance matrix, *J. Chem. Inf. Comput. Sci.* **32** (1992) 28–36.
- [53] D. Janežič, A. Miličević, S. Nikolić, N. Trinajstić, *Graph Theoretical Matrices in Chemistry*, Univ. Kragujevac, Kragujevac, 2007.
- [54] D. J. Klein, M. Randić, Resistance distance, *J. Math. Chem.* **12** (1993) 81–95; M. Randić, X. F. Guo, T. Oxley H. Krishnapriyan, L. C. Naylor, Wiener matrix invariants, *J. Chem. Inf. Comput. Sci.* **34** (1994) 361–367.
- [55] D. Babić, D. J. Klein, I. Lukovits, S. Nikolić, N. Trinajstić, Resistance–distance matrix: A computational algorithm and its applications, *Int. J. Quantum Chem.* **90** (2002) 166–176.
- [56] G. Jaklič, T. Pisanski, M. Randić, On description of biological sequences by spectral properties of line distance matrix, *MATCH Commun. Math. Comput. Chem.* **58** (2007) 301–307.
- [57] M. Randić, J. Zupan, T. Pisanski, On representation of DNA by line distance matrix, *J. Math. Chem.* **43** (2008) 674–692.
- [58] M. Randić, T. Pisanski, M. Novič, D. Plavšić, Novel graph distance matrix, *J. Comput. Chem.* **31** (2010) 1832–1841.
- [59] M. Randić, X. F. Guo, S. Bobst, Use of path matrices for a characterization of molecular structures, in: P. Hansen, P. Fowler, M. Zheng (Eds.), *Discrete Mathematical Chemistry*, American Math. Soc., Providence, 2000, pp. 305–322.