The Structural Formula Version
of Graph Theory

Jan Cz. Dobrowolski [1,2]

[1]*Institute of Nuclear Chemistry and Technology, 16 Dorodna-Street,*
03-195 Warsaw, Poland
[2]*National Medicines Institute, 30/34 Chełmska-Street, 00-723 Warsaw,*
Poland

j.dobrowolski@nil.gov.pl

Abstract

The proposed Structural Formula (SF) concept is a version of Graph Theory (GT) with different kinds of vertices and edges. Within SF, any molecule depicted according to the IUPAC rules can be analyzed as it is drawn. The construction of SF requires only a slight modification of the graph definition: a family of sets of vertices and a family of sets of edges are assigned to different kinds of atoms and bonds instead of a single set of vertices and a single set of edges. To easily introduce the physical characteristics of atoms and bonds, we also include a family of weighting functions defined on families of vertices and edges. The characteristics are introduced in analyses of the SF formula through the SF incidence matrix and then, through simple equations, are transferred to other SF matrices, such as Zagreb, Randić and distance and, ultimately, SF topological indices. Finally, we show that the HOMA geometrical aromaticity index can be treated like the SF topological index.

1. Introduction

There is no chemistry without atoms. Nevertheless, a lot of molecular properties, such as number of isomers, boiling point, aromaticity and π-electron delocalization reveal strong connections to molecular topology, to such an extent that the kind of atom seems to play a secondary role [1-8]. These properties are modeled well by graph theory (GT) and its various adaptations to chemistry [e.g., 9-13]. The role of these methods for modern chemistry cannot

be overestimated. Moreover, some purely chemical inventions to graph theory are now also studied in the field of mathematics [14-23].

However, it seems that chemists' initial fascination with graph theory has worn off a bit. This is hardly surprising. One can mostly attribute this decline of interest to the development of quantum and computational chemistry [24-26]. These methods can now be used in almost every chemical laboratory and can provide justified answers to several problems that have been posed in the field of the chemical graph theory (CGT). Another reason for the approach's diminished popularity is the way CGT is developed. The discipline has become less chemical, more abstract and increasingly difficult for chemists to understand. For chemists, there is also a fundamental flaw inherent in graph theory: it does not differentiate atoms or bonds [9].

1,3-dimethylcyclohexane

cyclohexane-1,3-diol

3,5-dimethylcyclohexene

cyclohex-4-ene-1,3-diol

1,5-dimethylcyclohexa-1,3-diene

cyclohex-4-ene-1,3-dione

1,3-dimethylcyclohexa-1,3,5-triene

3*H*-pyridine-2,6-dione

m-xylene

uracil
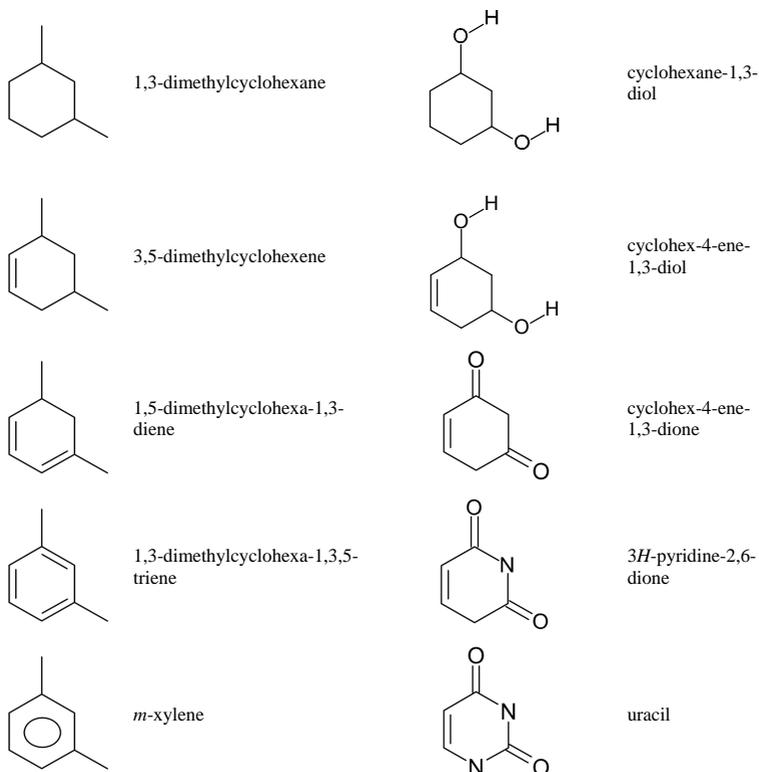
Figure 1. Structural formulae of molecules which, in the most fundamental version of the graph theory, are all represented by the top left graph of 1,3-dimethylcyclohexane.

Many modifications of GT have been proposed to counter this perceived defect. Weighted graphs [27, 28], graphs with multiple edges and loops (multigraphs) [29], complete graphs with a core electron representation [30], hypergraphs, and parameterizations of topological indices partially fill the gap between the needs of chemistry and the mathematics behind chemical graph theory. However, none of these seem to be a universal remedy to the divergence between the areas of interest in chemistry and a graph theory description of chemical structure. Even if a perfect GT description of chemical structures does not exist, searching for a better GT model of molecules still seems worthwhile.

In this paper, we analyze possible improvements in describing structural formulae with graph-theoretical methods. Structural formulae are the essence of chemical thinking and reasoning. On the other hand, they are very similar to graphs used in many fields of theoretical and applied mathematics. This similarity and the advanced level of graph theory methods have probably been the main reasons why a more adequate theory describing chemical structural formulae has not been needed for decades. Thus, GT methods have been applied to chemical problems without any special adaptation. If a problem was not satisfactorily described, one would use a "chemical parametrization" or a more sophisticated, yet known, version of the graph theory.

Yet, generalizing graph theory for a better representation of structural formulae does not seem to be that difficult. It suffices to: (1) differentiate the vertices so that they represent the whole periodic table of elements, (2) differentiate the edges so that they can flexibly represent many different kinds of bonds and (3) facilitate the introduction of chemical characteristics of atoms and bonds to the newly defined topological indices. Here, we show how to formulate such a theory that operates on common structural formulae and generates topological indices directly from them. To do this, we consider families of sets instead of the sets themselves, as well as IUPAC and some other representations of atoms and bonds. We also describe the addition and multiplication of abstract elements. Based on a so-defined SF concept, it becomes clear that, for example, the widely used HOMA geometrical aromaticity index is a topological index of a certain version of SF theory.

2. The concept

2.1.1 Structural formula

Let us first recall the definition of a graph.

Definition 1. A graph G is an ordered pair of sets $G=(V, E)$ where V is an arbitrary finite nonempty set and E is a subset of its Cartesian product $E \subseteq V \times V$.

Definition 2. A doubly-weighted graph is a quadruple $G_{f,g}=(V, E, f, g)$, where f and g are vertex and edge weighting functions $f:V \to \mathbb{R}$ and $g:E \to \mathbb{R}$, respectively, and $\mathbb{R}$ is the set of real numbers.

A structural formula can be defined similarly using families of sets instead of sets:

Definition 3. A structural formula is an ordered pair of families of sets $\mathbf{F}=(\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is an arbitrary finite family of nonempty sets and $\mathbf{E}$ is a subset of its Cartesian product $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$.

Definition 4. A doubly weighted structural formula is a quadruple $\mathbf{F}_{f,g}=(\mathbf{V}, \mathbf{E}, \mathbf{f}, \mathbf{g})$ composed of $\mathbf{V}$, $\mathbf{E}$: $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$, and two weighting functions $\mathbf{f}$ and $\mathbf{g}$ over $\mathbf{V}$ and $\mathbf{E}$, respectively, $\mathbf{f}:\mathbf{V} \to \mathbb{R}$ and $\mathbf{g}:\mathbf{E} \to \mathbb{R}$.

The family of sets $\mathbf{V}=\{V_1, V_2, \ldots, V_m\}$ is composed of sets of different vertices (representing different atoms) and the family of sets $\mathbf{E}=\{E_1, E_2, \ldots, E_n\}$ is composed of sets of different edges (representing different bonds).

The Cartesian product $\mathbf{A} \times \mathbf{B}$ of two families of sets, $\mathbf{A}=\{A_1, A_2, \ldots, A_m\}$ and $\mathbf{B}=\{B_1, B_2, \ldots, B_n\}$, is a set of Cartesian products $\mathbf{A} \times \mathbf{B} = \{A_1 \times B_1, A_1 \times B_2, \ldots, A_m \times B_n\}$. The function $\mathbf{f}:\mathbf{A} \to \mathbb{R}$ over a family of sets $\mathbf{A}$ acts on each of the sets in the family.

Remarks:

If only one set of vertices and one set of edges is considered, Definitions 3 and 4 are reduced to Definitions 1 and 2, respectively, and the SF concept is reduced to the GT approach. In contrast, applying the SF formulation allows all the drawings in Fig. 1 to be distinct SF objects.

The additional introduction of unary and m-ary relations to $\mathbf{E}$ would allow both the use of more than one type of loops on vertices and hyperedges between m-vertices. Thus, generalizing the SF concept towards a hyper structural formula approach requires only the additional assumption that $\mathbf{E} \subseteq \mathbf{V^k}$, where $k=1, 2, \ldots, m$. Loops and hyperedges can be helpful to characterize functional groups or to distinguish some atoms in the formulae.

There are two ways of modifying the vertices and edges in a graph to construct a structural formula. The first would be to declare atomic/bond symbols. The second would be assigning weights to them. Applying both approaches simultaneously may seem superfluous, but they accomplish different objectives. The former guarantees that any structural formula taken from a chemical text can be mathematically analyzed if the meaning of all elements of the chemical structural formulae is well explained. The latter allows to flexibly assign atom and bond descriptors to vertices and edges of a given type.

It seems that classical GT problems such as shortest paths, travelling salesman, vertex or edge coloring can be easily transferred to the realm of SF.

2.1.2 Path metrics

In a graph G and a structural formula F, a path p from vertex v_i to v_j is a finite sequence of edges $(E_1, E_2,… E_s)$ connecting the corresponding sequence of vertices. A path distance d_p between vertices [36] can also be defined as follows:

$$d_p(v_i, v_j) = \begin{cases} 0 & \text{for } i = j \\ 1 & \text{for } i \text{ adjacent to } j \\ s & = \sum_{i=1}^{s} d(E_i) \text{ where } (E_1, E_2,..., E_s) \text{ is a shortest path from } i \text{ to } j \end{cases}$$

Observe that d_p is non-negative, is zero for the same element, is symmetric since the shortest path from i to j and back is the same, and satisfies the triangle inequality. Thus, d_p is a metric in a graph G or a structural formula F that is called the path metric. Therefore, $(G, d_p) = (V, E, d_p)$ and $(F, d_p) = (V, E, d_p)$ are metric spaces.

A molecular path of distance $\boldsymbol{d_p}(v_k, v_l)$ in G or F can be introduced based on bond lengths in the reference molecule embedded in the $\mathbb{R}^3$ space. Let M be a molecule composed of atoms $A=(A_1, A_2, …, A_n)$ and bonds $B=(B_1, B_2, …, B_m)$. Let the bond lengths $(d(B_1), d(B_2), …, d(B_m))$ be Cartesian distances in $\mathbb{R}^3$ between the bonded atoms, for instance, calculated by a structural chemistry method. In the case where only one shortest path $p=(E_1, E_2,… E_s)$ from v_k to v_l exists according to the path metric d_p, *the molecular path distance* $\boldsymbol{d_p}(v_i, v_j)$ is the sum of the bond lengths in the corresponding sequence of bonds $(B_1, B_2,… B_s)$:

$$\boldsymbol{d_p}(v_k, v_l) = \sum_{i=1}^{s} d(B_i)$$

On the other hand, when two or more the shortest paths, $p_1, p_2, …, p_t$, exist according to the path metric, $d_{p_1} = d_{p_2} = \cdots = d_{p_t}$, *the shortest molecular path distance* $\boldsymbol{d_p}$ is taken as the minimum value of $\boldsymbol{d_p}$:

$$d_p = \min_{i=1,2,..t}(d_{p_i})$$

The molecular path distance d_p is a metric. Indeed, the non-negativity, identity and symmetry properties follow the use of $\mathbb{R}^3$ Cartesian bond distances in a molecule. The triangle inequality is obvious if v_i, v_j and v_k are placed at the same path. If there are multiple paths (with no return) from v_i to v_k, the triangle inequality follows from the minimality requirement. Hence, $G=(V, E, d_p)$ and $F=(\mathbf{V}, \mathbf{E}, d_p)$ are metric spaces.

The molecular path metric, which is a special case of a weighted path metric [36], can be simply treated as a weighting function assigned to edges. Because of the important role of distance-based parameters in GT and its connection to metric topology, the path metrics defined in G and F seem to be especially interesting.

2.1.3 Atom and bond representations

To define the structural formula, a representation of atoms and bonds is essential, yet implicit. The atoms can first and foremost be represented by their IUPAC atomic symbols (accompanied by pluses, minuses, deltas, full stops, colons etc., all of them with or without integer numbers) [37]. Yet, it is also possible to introduce atom representations in many other ways, such as colors or, for example (Table 1):

(i) different graph representations of atoms $\mathbf{V}=\{G_H, G_C, G_N, G_O, \ldots\}$, including complete graphs [38, 39];

(ii) ordered pairs $(v_i, A_i) \in \mathbf{V} \square V \times A$ where v_i is a vertex and A_i is an atomic number;

(iii) vectors (ordered n-tuples) or $k \times l$ dimensional matrices in which non-zero entries identify the kind of atoms; etc.

Apart from the line or multiple lines representation of the bonds, other symbols like dotted lines, arrows, arcs inside the rings, light and dark triangles, waves etc., commonly used in chemistry [37], physics [40], and in less basic versions of graph theory [41-43] are also used in the structural formulae approach. However, natural bond representations such as: (i) overlapped atomic, hybrid or molecular orbitals; (ii) charge density contour maps, 3D surfaces or 3D diffused shapes [44]; (iii) bond critical points [45, 46], etc. can be used as well.

Table 1. A few examples of a potentially vast catalogue of possible atom representations which can be combined with the structural formula approach.

Atomic symbol	Atomic multigraph	Atomic digraph	ordered pair	ordered n-tuple
H			(v, 1)	(1,0,0,0,0,0,0,0,0,…)
C			(v, 6)	(0,0,0,0,0,1,0,0,0,…)
N			(v, 7)	(0,0,0,0,0,0,1,0,0,…)
O			(v, 8)	(1,0,0,0,0,0,0,1,0,…)

The problem of cataloguing atom and bond representations of value in the SF version of CGT can be the subject of further research. For this study, it is enough to keep in mind that for the vast majority of chemical applications, the IUPAC accepted atomic and bond symbols can also have proper mathematical meaning.

2.1.4 Example

Let us consider a neutral, closed shell, vinylacetylene molecule in its ground electronic state (Fig. 2). The molecule is well characterized [47], it is a valuable polymerization and organic chemistry reactant and was proven present in the atmosphere of Titan [48]. Assume that this molecule has single bonds between the H and C atoms and single, double, triple, or delocalized bonds between the C atoms. Two different graphical representations of vinylacetylene, Fig. 2a and 2b, can be exemplified in terms of pairs of families of sets as $\mathbf{F_a}=(\mathbf{V_a},\mathbf{E_a})$ and $\mathbf{F_b}=(\mathbf{V_b},\mathbf{E_b})$, respectively. They have identical families of sets of vertices and, although one of the sets of edges, E_1, in $\mathbf{E_a}$ and in $\mathbf{E_b}$ is identical, the two families of sets of edges are not equinumerous and $\mathbf{F_a}$ and $\mathbf{F_b}$ are not isomorphic.

(V_a, E_a):
$V_a = \{V_1, V_2\}$ $E_a = \{E_1, E_2, E_3, E_4\}$:
$V_1 = \{C_1, C_2, C_3, C_4\}$;
$V_2 = \{H_5, H_6, H_7, H_8\}$
and
$E_1 = \{H_5\text{-}C_4, H_6\text{-}C_2, H_7\text{-}C_1, H_8\text{-}C_1\}$; $E_2 = \{C_2\text{-}C_3\}$
$E_3 = \{C_1 = C_2\}$, $E_4 = \{C_3 \equiv C_4\}$

(a)

(V_b, E_b):
$V_b = \{V_1, V_2\}$; $E_b = \{E_1, E_2\}$
$V_1 = \{C_1, C_2, C_3, C_4\}$;
$V_2 = \{H_5, H_6, H_7, H_8\}$
and
$E_1 = \{H_5\text{-}C_4, H_6\text{-}C_2, H_7\text{-}C_1, H_8\text{-}C_1\}$
$E_2 = \{C_2 \approx C_3, C_1 \approx C_2, C_3 \approx C_4\}$.

(b)
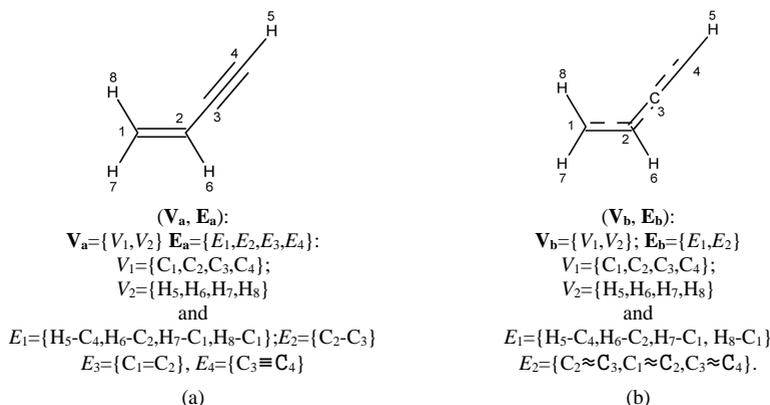
Figure 2. Drawings of vinylacetylene (but-1-en-3-yne) structural formulae showing (a) the nominal bonds and (b) the bonds indicating the π-electron delocalization. Below the drawings, the corresponding definitions in terms of families of sets of vertices and families of sets of edges are presented.

2.1.5 Weighted SF

As in the chemical weighted graph theory [e.g., 49-52], in the weighted structural formula approach, the weights can be derived from any atomic or bond properties. Formally, in SF, the weighting functions act on domains different than in GT: they act on families of sets of different elements, while in GT they act on sets of uniform vertices and uniform edges. Nevertheless, in both cases the result can be the same: we end up with vertices and edges to which some numbers are assigned. Indeed, in Table S1, the valency and partial charge may be treated either as a result of acting of a family of weighting functions $\mathbf{f} = \{f_1, f_2\}$ on SF, where f_1 and f_2 act separately on sets of C- and H-atoms, respectively. In GT, these properties are the result of the action of a sole $\mathbf{f} = \{f\}$ function on the two types of elements. The same can be said for $\mathbf{g}$ acting on different kinds of bonds.

Hence, what sense is there to introduce the modified formalism if, from the practical point of view, the result may remain the same? First, in our opinion, the SF version of CGT is worth introducing because it allows the natural inclusion of rich information about chemical composition, topology and structure contained in the structural formula. Second, within the SF concept, the use of weights for atoms and bonds is self-explanatory. Third, the well-developed GT machinery can be easily generalized to all molecules that can be drawn as structural formulae.

2.2.1 Some GT matrices

Here we focus on the practical aim of constructing "more chemical" topological indices based on structural formulae which then may be applied as molecular descriptors in structure-activity or structure-property studies. A lot of topological indices can be calculated from the appropriate matrices associated with graphs. Therefore, below we demonstrate changes in the form of some basic GT matrices when instead of a graph, a structural formula is analyzed. However, first, let us recall the basic relationships between some GT matrices:

The incidence matrix of the graph G, $\mathbb{I}(G)$, is defined as follows:

$$\mathbb{I}(G) = \|i_{mn}\|, m=1, 2, \ldots, k \text{ (vertices); } n=1, 2, \ldots, l \text{ (edges):}$$
$$i_{mn}=1 \text{ if } v_m \text{ is incident to } e_n \text{ and 0 otherwise.} \tag{1}$$

The k rows and l columns of $\mathbb{I}(G)$ correspond to G vertices and edges, respectively. There are two squares of the rectangular $\mathbb{I}(G)$ matrix:

$$\mathbb{I}^T(G)_{l\times k}\mathbb{I}(G)_{k\times l} = {}^e\mathbb{A}(G)_{l\times l} + {}^e\mathbb{V}(G)_{l\times l} \text{ (edge adjacency and valency)} \tag{2}$$

and

$$\mathbb{I}^T(G)_{k\times l}\mathbb{I}(G)_{l\times k} = {}^v\mathbb{A}(G)_{k\times k} + {}^v\mathbb{V}(G)_{k\times k}\text{(vertex adjacency and valency)} \tag{3}$$

where square $\mathbb{A}$ adjacency matrices are off-diagonal and symmetric, while square $\mathbb{V}$ valency ones are diagonal. The entries of $\mathbb{A}$ are either 1 or 0 depending on whether two edges or vertices are adjacent or not. On the other hand, the entries of $\mathbb{V}$ are either all equal to 2 (${}^e\mathbb{V}$) or equal to vertex valency (${}^v\mathbb{V}$). Since the vertex ${}^v\mathbb{A}_{k\times k}(G)$ and ${}^v\mathbb{V}_{k\times k}(G)$ matrices are used much more often, they are written without the superscript "v", in cases where it does not lead to confusion.

There are some simple equations connecting the $\mathbb{A}(G)$ and $\mathbb{V}(G)$ matrices with certain important GT matrices [54], e.g., the first and second Zagreb matrices, $\mathbb{Z}_1(G)$ and $\mathbb{Z}_2(G)$; the Randić connectivity matrix, $\mathbb{R}(G)$; the Laplacian matrix, $\mathbb{L}(G)$; and the Distance matrix, $\mathbb{D}(G)$:

$$\mathbb{Z}_1(G) = \mathbb{V}^2(G) \tag{4}$$

$$\mathbb{Z}_2(G) = \mathbb{V}(G)\mathbb{A}(G)\mathbb{V}(G) \tag{5}$$

$$\mathbb{R}(G) = \mathfrak{d}(G)\mathbb{A}(G)\mathfrak{d}(G) \tag{6}$$

$$\mathbb{L}(G) = \mathbb{V}(G) - \mathbb{A}(G) \tag{7}$$

$$\mathbb{D}(G) = \sum_{i=1}^{k} i\cdot\mathbb{A}_i(G) \tag{8}$$

where $\eth(G)=\|(v_i)^{-\frac{1}{2}}\|$ is a diagonal matrix and v_i is the valency of the vertex i. Notice, that, for instance, to calculate the Harary distance matrix [55] it is enough to replace i in the last equation by $1/i$. The adjacency matrix through 2, 3, …, k edges, $\mathbb{A}_2(G), \mathbb{A}_3(G), …, \mathbb{A}_k(G)$, has an entry equal to 1 if a respective path without repetition of length 2, 3, …, k exists and it is the shortest path between these two vertices, and 0 otherwise. If two or more different paths exist between two vertices, as in (multi)cyclic graphs, the entry remains equal to 1.

The distance matrix $\mathbb{D}(G)$ plays a central role in chemical GT, as the basic topological Wiener index is calculated from it. Therefore, it is important that the $\mathbb{A}_i(G)$ adjacency matrix can be expressed using the $\mathbb{A}^i(G)$ matrix, which is the i-th power of $\mathbb{A}(G) = \mathbb{A}_1(G)$. The r,s entry of the $\mathbb{A}^i(G)$ matrix is equal to the number of paths with repetition of length i between vertices r and s. The unit entries of $\mathbb{A}^i(G)$ indicate vertices connected by a path without repetition and are in the same positions as in $\mathbb{A}_i(G)$, while all the other $\mathbb{A}^i(G)$ entries correspond to zeros in $\mathbb{A}_i(G)$:

$$\mathbb{A}_i(G) = \mathbb{A}^i(G)_{j \neq 1 \to 0} \tag{9}$$

where $j\neq1\to0$ denotes that each entry different from 1 in $\mathbb{A}^i(G)$ is set equal to 0 in $\mathbb{A}_i(G)$.

However, one can also skip the condition that in $\mathbb{A}_i(G)$ the entry is 1 if and only if, "*this is the shortest path between these two vertices*" and conserve the condition that "*if between two vertices there are two or more such paths the entry remains equal to 1*". Such "full" adjacency matrices, $^f\mathbb{A}_i(G)$, may have r, s entries equal to 1 for different $i=1,2,…,k,…,$ l, where k is the number of edges and $l \geq k$ is the longest path without repetition. Based on the full adjacency matrix without repetition, one can define the detour adjacency matrix, $^\Delta\mathbb{A}_i(G)$:

$$^\Delta\mathbb{A}_i(G)=\,^f\mathbb{A}_i(G)_{j=1\to0} \tag{10}$$

where $j=1\to0$ symbolizes that 1 entry is set 0 when it is also 1 in any matrix corresponding to a longer path.

Based on detour adjacency matrices, the detour distance matrix, $\Delta(G)$ [56, 57], can be defined as follows:

$$\Delta(G) = \sum_{i=l}^{1} i \cdot\, ^\Delta\mathbb{A}_i(G) \tag{11}$$

where $l \geq k$ and $l = k$ only for graphs of trees, and summing in the opposite direction is to emphasize that that if an entry is 1 for a greater i than for any smaller index, it is set to 0.

2.2.2 The SF matrices

Let us start from the incidence $\mathbb{I}(SF)$ matrices of the simple hydrogen cyanide, H-C≡N, and hydrogen isocyanide, H-N≡C, molecules (Fig. 3).

	H-C		C≡N				H-N		N≡C	
H	v_H	e_{HC}	0			H	v_H	e_{HN}	0	
C	v_C	e_{HC}	v_C	e_{CN}		C	0		v_C	e_{NC}
N	0		v_N	e_{CN}		N	v_N	e_{HN}	v_N	e_{NC}
	$\mathbb{I}(H\text{-}C≡N)$						$\mathbb{I}(H\text{-}N≡C)$			

Figure 3. The incidence $\mathbb{I}(SF)$ matrices of H-C≡N and H-N≡C molecules

In the above example, the SF incidence matrices are defined as follows:

$$\mathbb{I}(SF)_{k\times l} = \mathcal{V}_{k\times k} \cdot \mathbb{I}(G)_{k\times l} \cdot \mathcal{E}_{l\times l}$$
$$\mathbb{I}(SF) = \|i_{mn}\|, \; m=1, 2, \ldots, k; \; n=1, 2, \ldots, l:$$
$$\mathcal{V}_{k\times k} = \|v_m\| \text{ and } \mathcal{E}_{l\times l} = \|e_n\|$$
$$i_{mn} = v_m \circ_{SF} e_n \text{ if } v_m \text{ is incident to } e_n \text{ and 0 otherwise.}$$

(12')

where $\mathbb{I}(G)_{k\times l}$ is the incidence matrix of the graph corresponding to SF and entries v_m and e_n of the $\mathcal{V}_{k\times k}$ and $\mathcal{E}_{l\times l}$ diagonal matrices are treated as abstract elements which are multiplied according to an abstract (and not necessarily commutative) $\circ_{SF}$ operation. To make use of Eqs. (1)-(11), it is necessary to define the $v_m \circ_{SF} e_n$, $e_n \circ_{SF} v_m$, $v_s \circ_{SF} v_r$ and $e_q \circ_{SF} e_t$ operations, as well as their additions. Notice, that these definitions may be complex because the multiplication and addition depend on the atom and bond representations mentioned in section 2.1.3. However attractive, such an approach seems to go beyond the interest of the chemical graph theory and here will not be further considered.

Let us define the incidence matrix of the structural formula, $\mathbb{I}(SF)$, as follows:

$$\mathbb{I}(SF)_{k\times l} = F(v_m)_{k\times k} \cdot \mathbb{I}(G)_{k\times l} \cdot G(e_n)_{l\times l}$$
$$\mathbb{I}(SF) = \|i_{mn}\|, \; m=1, 2, \ldots, k; \; n=1, 2, \ldots, l:$$
$$F(v_m)_{k\times k} = \|f(v_m)\| \text{ and } G(e_n)_{l\times l} = \|g(e_n)\|$$
$$i_{mn} = f(v_m) \cdot g(e_n) \text{ if } v_m \text{ is incident to } e_n \text{ and 0 otherwise.}$$

(12)

where $F(v_m)_{k\times k}$ and $G(e_n)_{l\times l}$ are diagonal weighting matrices and $f(v_m)$ and $g(e_n)$ are respectively atom and bonds weights, which are values of the **f:V**→$\mathbb{R}$ and **g:E**→$\mathbb{R}$ functions.

Now, the two squares of the $\mathbb{I}(SF)$ matrix, $\mathbb{I}^T(SF)_{l\times k} \; \mathbb{I}(SF)_{k\times l}$ and $\mathbb{I}(SF)_{k\times l} \; \mathbb{I}^T(SF)_{l\times k}$, contain information on atoms and bond properties in both components: their off-diagonal adjacency and diagonal valency matrices:

$$^v\mathbb{V}_{k\times k}(SF) = \|v_{mm}\|$$
$$v_{mm} = f^2(v_m) \cdot \left[g^2(e_{ms_1}) + g^2(e_{ms_2}) + \cdots + g^2(e_{ms_l}) \right] \tag{13}$$

$$^v\mathbb{A}_{k\times k}(SF) = \|a_{mm}\|$$
$$a_{mn} = f(v_m) \cdot f(v_n) \cdot g^2(e_{mn}) \tag{14}$$

$$^e\mathbb{V}_{l\times l}(SF) = \|v_{nn}\|$$
$$v_{nn} = g^2(e_n) \cdot \left[f^2(v_{nr_1}) + f^2(v_{nr_2}) \right] \tag{15}$$

$$^e\mathbb{A}_{l\times l}(SF) = \|a_{mm}\|$$
$$a_{nm} = g(e_n) \cdot g(e_m) \cdot f^2(v_{nm}) \tag{16}$$

where s runs over l vertices linked to the m-th vertex; r indexes two vertices linked through the n-th edge; e_{mn} denotes the edge linking vertex m with n; and v_{nm} denotes the vertex through which edges n and m are adjacent.

Interestingly, assuming $g(e_i)=1$ for an arbitrary i, $^v\mathbb{A}$ and $^v\mathbb{V}$ but also $^e\mathbb{A}$ and $^e\mathbb{V}$ depend exclusively on vertex properties. And conversely, by setting $f(v_j)=1$ for an arbitrary j, $^e\mathbb{A}$ and $^e\mathbb{V}$ but also $^v\mathbb{A}$ and $^v\mathbb{V}$ depend only on edge properties. A variety of other matrices and topological indices can be obtained based on the vertex and edge $\mathbb{A}$ and $\mathbb{V}$ matrices. Thus, mixing the vertex type of matrix with the edge properties permits the construction of yet other matrices and topological indices

Matrices (13)-(16) for the structural formula of dimethyl-carbamic acid (hydrocarbon-hydrogen-depleted) molecule are labelled according to the IUPAC rules and presented in Fig. 3, where $f(e_i)$ and $g(v_j)$ are replaced by f_i and g_j, respectively.

dimethylcarbamic acid (DMCA)
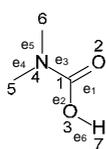
$$\mathbb{I}(SF_{DMCA})$$

$$
\begin{vmatrix}
f_1g_1 & f_1g_2 & f_1g_3 & 0 & 0 & 0 \\
f_2g_1 & 0 & 0 & 0 & 0 & 0 \\
0 & f_3g_2 & 0 & 0 & 0 & f_3g_6 \\
0 & 0 & f_4g_3 & f_4g_4 & f_4g_5 & 0 \\
0 & 0 & 0 & f_5g_4 & 0 & 0 \\
0 & 0 & 0 & 0 & f_6g_5 & 0 \\
0 & 0 & 0 & 0 & 0 & f_7g_6
\end{vmatrix}
$$

$$
\begin{vmatrix}
f^2{}_1(g^2{}_1+g^2{}_2+g^2{}_3) & f_1f_2g^2{}_1 & f_1f_3g^2{}_2 & f_1f_4g^2{}_3 & 0 & 0 & 0 \\
f_1f_2g^2{}_1 & f^2{}_2g^2{}_1 & 0 & 0 & 0 & 0 & 0 \\
f_1f_3g^2{}_2 & 0 & f^2{}_3(g^2{}_2+g^2{}_6) & 0 & 0 & 0 & f_3f_7g^2{}_6 \\
f_1f_4g^2{}_3 & 0 & 0 & f^2{}_4(g^2{}_3+g^2{}_4+g^2{}_5) & f_4f_5g^2{}_4 & f_4f_6g^2{}_5 & 0
\end{vmatrix}
$$

$$
\begin{vmatrix}
0 & 0 & 0 & f_4f_5g^2{}_4 & f^2{}_5g^2{}_4 & 0 & 0 \\
0 & 0 & 0 & f_4f_6g^2{}_5 & 0 & f^2{}_6g^2{}_5 & 0 \\
0 & 0 & f_3f_7g^2{}_6 & 0 & 0 & 0 & f^2{}_7g^2{}_6
\end{vmatrix}
$$

$$^{v}\mathbb{A}(SF_{DMCA})+{}^{v}\mathbb{V}(SF_{DMCA})$$

$$
\begin{vmatrix}
g^2{}_1(f^2{}_1+f^2{}_2) & g_1g_2f^2{}_1 & g_1g_3f^2{}_1 & 0 & 0 & 0 \\
g_1g_2f^2{}_1 & g^2{}_2(f^2{}_1+f^2{}_3) & 0 & 0 & 0 & g_2g_6f^2{}_3 \\
g_1g_3f^2{}_1 & 0 & g^2{}_3(f^2{}_1+f^2{}_4) & g_3g_4f^2{}_4 & 0 & 0 \\
0 & 0 & g_3g_4f^2{}_4 & g^2{}_4(f^2{}_4+f^2{}_5) & g_3g_5f^2{}_4 & 0 \\
0 & 0 & 0 & g_3g_5f^2{}_4 & g^2{}_5(f^2{}_4+f^2{}_6) & 0 \\
0 & g_2g_6f^2{}_3 & 0 & 0 & 0 & g^2{}_6(f^2{}_3+f^2{}_7)
\end{vmatrix}
$$

$$^{e}\mathbb{A}(SF_{DMCA})+{}^{e}\mathbb{V}(SF_{DMCA})$$

Figure 3. The incidence, edge, vertex adjacency and valency matrices for the hydrocarbon-hydrogen-depleted structural formula of the dimethylcarbamic acid molecule.

Now, the vertex and edge first Zagreb SF matrices, $^{v}\mathbb{Z}_1$ and $^{e}\mathbb{Z}_1$, have the following forms:

$$^{v}\mathbb{Z}_1(SF) = \|v^2_{mm}\|$$
$$v^2_{mm} = f^4_m\left(g^2_{ms_1} + g^2_{ms_2} + \cdots + g^2_{ms_l}\right)^2 \tag{17}$$

$$^{e}\mathbb{Z}_1(SF) = \|v^2_{nn}\|$$
$$v^2_{nn} = g^4_n\left(f^2_{nr_1} + f^2_{nr_2}\right)^2 \tag{18}$$

where the simplified index notation is similar to that in Fig 3. Because in Eq. (17) and (18) only powers of the valency matrix entries are present, it seems that these two Zagreb matrices do not introduce more information than do the $^{v}\mathbb{V}$ and $^{e}\mathbb{V}$ valency matrices.

The vertex and edge second Zagreb SF matrices, $^{v}\mathbb{Z}_2$ and $^{e}\mathbb{Z}_2$, have the following forms:

$$^{v}\mathbb{Z}_2(SF) = \left\| {}^{v}z_{2,mn}\right\|$$
$$^{v}z_{2,mn} = {}^{v}v_m \, {}^{v}a_{mn} \, {}^{v}v_n$$
$$^{v}z_{2,mn} = \left[f^2_m\left(g^2_{ms_1} + g^2_{ms_2} + \cdots + g^2_{ms_l}\right)\right]\left[f_m f_n g^2_{mn}\right]\left[f^2_n\left(g^2_{nt_1} + g^2_{nt_2} + \cdots + g^2_{nt_l}\right)\right]$$
$$^{v}z_{2,mn} = f^3_m f^3_n g^2_{mn}\left(g^2_{ms_1} + g^2_{ms_2} + \cdots + g^2_{ms_l}\right)\left(g^2_{nt_1} + g^2_{nt_2} + \cdots + g^2_{nt_l}\right) \tag{19}$$

$$^{e}\mathbb{Z}_2(SF) = \left\| {}^{e}z_{2,nm}\right\|$$
$$^{e}z_{2,nm} = {}^{e}v_n \, {}^{e}a_{nm} \, {}^{e}v_m$$
$$^{e}z_{2,nm} = \left[g^2_n\left(f^2_{nr_1} + f^2_{nr_2}\right)\right]\left[g_n g_m f^2_{nm}\right]\left[g^2_m\left(f^2_{nq_1} + f^2_{nq_2}\right)\right]$$
$$^{e}z_{2,nm} = g^3_n g^3_m f^2_{nm}\left(f^2_{nr_1} + f^2_{nr_2}\right)\left(f^2_{nq_1} + f^2_{nq_2}\right) \tag{20}$$

where the indices have analogous meanings to those used for (13)-(16).

The Randić matrix can be defined analogously to the second Zagreb matrix using the $\mathfrak{d}(SF)=\|(v_i)^{-\frac{1}{2}}\|$ matrix instead of the valency matrix. Thus:

$$^{v}\mathbb{R}(SF) = \|\,^{v}r_{mn}\|$$

$$^{v}r_{mn} = (\,^{v}v_{mm})^{-\frac{1}{2}}(\,^{v}a_{mn})(\,^{v}v_{nn})^{-\frac{1}{2}} =$$

$$^{v}r_{mn} = \left[f_m^2\left(g_{ms_1}^2 + g_{ms_2}^2 + \cdots + g_{ms_l}^2\right)\right]^{-\frac{1}{2}}\left[f_m f_n g_{mn}^2\right]\left[f_n^2\left(g_{nt_1}^2 + g_{nt_2}^2 + \cdots + g_{nt_l}^2\right)\right]^{-\frac{1}{2}} \tag{21}$$

$$^{v}r_{mn} = f_m^2 f_n^2 g_{mn}^2\left[\left(g_{ms_1}^2 + g_{ms_2}^2 + \cdots + g_{ms_l}^2\right)\left(g_{nt_1}^2 + g_{nt_2}^2 + \cdots + g_{nt_l}^2\right)\right]^{-\frac{1}{2}}$$

$$^{e}\mathbb{R}(SF) = \|\,^{e}r_{nm}\|$$

$$^{e}r_{nm} = (\,^{e}v_{nn})^{-\frac{1}{2}}(\,^{e}a_{nm})(\,^{e}v_{mm})^{-\frac{1}{2}}$$

$$^{e}r_{nm} = \left[g_n^2\left(f_{nr_1}^2 + f_{nr_2}^2\right)\right]^{-\frac{1}{2}}\left[g_n g_m f_{nm}^2\right]\left[g_m^2\left(f_{nq_1}^2 + f_{nq_2}^2\right)\right]^{-\frac{1}{2}} \tag{22}$$

$$^{e}r_{nm} = g_n^2 g_m^2 f_{nm}^2\left[\left(f_{nr_1}^2 + f_{nr_2}^2\right)\left(f_{nq_1}^2 + f_{nq_2}^2\right)\right]^{-\frac{1}{2}}$$

Expressions for the Laplacian matrices, $^{v}\mathbb{L}(SF)$ and $^{e}\mathbb{L}(SF)$, are simple (7) and require no comments.

On the other hand, calculation of the SF distance matrices, $^{v}\mathbb{D}(SF)$ and $^{e}\mathbb{D}(SF)$, requires determining the subsequent $^{v}\mathbb{A}_i(SF)$ and $^{e}\mathbb{A}_j(SF)$ matrices and then summation as in (9):

$$\mathbb{D}(SF) = \sum_{i=1}^{k} i \cdot \mathbb{A}_i(SF) \tag{23}$$

dimethylcarbamic acid (DMCA) $\qquad$ $\mathbb{V}(SF_{DMCA})f(SF_{DMCA})\mathbb{I}(G_{DMCA})$
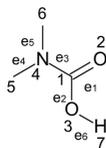
$$\begin{vmatrix} v_1f_1 & v_1f_1 & v_1f_1 & 0 & 0 & 0 \\ v_2f_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & v_3f_3g_2 & 0 & 0 & 0 & v_3f_3 \\ 0 & 0 & v_4f_4 & v_4f_4 & v_4f_4 & 0 \\ 0 & 0 & 0 & v_5f_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & v_6f_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & v_7f_7 \end{vmatrix}$$

$$\begin{vmatrix} 0 & \alpha_1\alpha_2f_1f_2g_1^2 & \alpha_1\alpha_3f_1f_3g_2^2 & \alpha_1\alpha_4f_1f_4g_3^2 & 0 & 0 & 0 \\ \alpha_1\alpha_2f_1f_2g_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1\alpha_3f_1f_3g_2^2 & 0 & 0 & 0 & 0 & 0 & \alpha_3\alpha_7f_3f_7g_6^2 \\ \alpha_1\alpha_4f_1f_4g_3^2 & 0 & 0 & 0 & \alpha_4\alpha_5f_4f_5g_4^2 & \alpha_4\alpha_6f_4f_6g_5^2 & 0 \\ 0 & 0 & 0 & \alpha_4\alpha_5f_4f_5g_4^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_4\alpha_6f_4f_6g_5^2 & 0 & 0 & 0 \\ 0 & 0 & \alpha_3\alpha_7f_3f_7g_6^2 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$^{v}\mathbb{Z}_2(SF_{DMCA})$ if $\alpha_m = {}^{v}v_{mm}$ or $^{v}\mathbb{R}(SF_{DMCA})$ $\alpha_m = {}^{v}\delta_m$

$$\begin{vmatrix} 0 & {}^{e}\alpha_1{}^{e}\alpha_2g_1g_2f^2{}_1 & {}^{e}\alpha_1{}^{e}\alpha_3g_1g_3f^2{}_1 & 0 & 0 & 0 \\ {}^{e}\alpha_1{}^{e}\alpha_2g_1g_2f^2{}_1 & 0 & 0 & 0 & 0 & {}^{e}\alpha_2{}^{e}\alpha_6g_2g_6f^2{}_3 \\ {}^{e}\alpha_1{}^{e}\alpha_3g_1g_3f^2{}_1 & 0 & 0 & {}^{e}\alpha_3{}^{e}\alpha_4g_3g_4f^2{}_4 & 0 & 0 \\ 0 & 0 & {}^{e}\alpha_3{}^{e}\alpha_4g_3g_4f^2{}_4 & 0 & {}^{e}\alpha_3{}^{e}\alpha_5g_3g_5f^2{}_4 & 0 \\ 0 & 0 & 0 & {}^{e}\alpha_3{}^{e}\alpha_5g_3g_5f^2{}_4 & 0 & 0 \\ 0 & {}^{e}\alpha_2{}^{e}\alpha_6g_2g_6f^2{}_3 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$^e\mathbb{Z}_2(SF_{\text{DMCA}}) \text{ if } \alpha_n = {}^ev_{nn} \text{ or } {}^e\mathbb{R}(SF_{\text{DMCA}}) \; \alpha_n = {}^e\delta_{nn}$$

$$
\begin{bmatrix}
0 & f_1f_2g_1^2 & f_1f_3g_1^2 & f_1f_4g_3^2 & 2f_1f_4g_3^2f_4f_5g_4^2 & 2f_1f_4g_2^2f_4f_6g_5^2 & 2f_1f_3g_2^2f_3f_7g_6^2 \\
f_1f_2g_1^2 & 0 & 2f_1f_3g_2^2f_1f_3g_2^2 & 2f_1f_3g_2^2f_1f_4g_3^2 & 3f_1f_3g_2^2f_1f_4g_3^2f_4f_5g_4^2 & 3f_1f_3g_2^2f_1f_4g_3^2f_4f_6g_5^2 & 3f_1f_3g_2^2f_1f_3g_2^2f_3f_7g_6^2 \\
f_1f_3g_2^2 & 2f_1f_3g_2^2f_1f_3g_2^2 & 0 & 2f_1f_3g_2^2f_1f_4g_3^2 & 3f_3f_5g_4^2f_1f_4g_3^2f_3f_7g_2^2 & 3f_3f_6g_5^2f_1f_4g_3^2f_3f_7g_2^2 & f_3f_7g_6^2 \\
f_1f_4g_3^2 & 2f_1f_3g_2^2f_1f_4g_3^2 & 2f_1f_3g_2^2f_1f_4g_3^2 & 0 & f_4f_5g_4^2 & f_4f_6g_5^2 & 4f_5f_4^2f_1^2f_3^2f_1^2f_3^2g_2^2g_3^2g_4^2g_2^2 \\
2f_1f_4g_3^2f_4f_5g_4^2 & 3f_1f_3g_2^2f_1f_4g_3^2f_4f_5g_4^2 & 3f_3f_5g_4^2f_1f_4g_3^2f_3f_7g_2^2 & f_4f_5g_4^2 & 0 & 2f_4f_5g_4^2f_4f_6g_5^2 & 4f_6f_4^2f_1^2f_3f_1^2f_3g_2^2g_3^2g_2^2 \\
2f_1f_3g_2^2f_3f_7g_6^2 & 3f_1f_3g_2^2f_1f_3g_2^2f_3f_7g_6^2 & f_3f_7g_6^2 & & 3f_1f_3g_2^2f_1f_3g_2^2f_3f_7g_6^2 & 4f_5f_4^2f_1^2f_3f_1^2f_3g_2^2g_3^2g_4^2g_2^2 & 4f_6f_4^2f_1^2f_3f_1^2f_3g_2^2g_3^2g_5^2g_2^2 & 0
\end{bmatrix}
$$

$$^v\mathbb{D}(SF_{\text{DMCA}})$$

$$
\begin{bmatrix}
0 & g_1g_2f_1^2 & g_1g_3f_1^2 & 2g_1g_3f_1^2g_3g_4f_4^2 & 2g_1g_3f_1^2g_3g_5f_4^2 & 2g_1g_2f_1^2g_2g_6f_3^2 \\
g_1g_2f_1^2 & 0 & g_2g_3f_1^2 & 2g_2g_3f_1^2g_3g_4f_4^2 & 2g_2g_3f_1^2g_3g_5f_4^2 & 2g_2g_6f_3^2 \\
g_1g_3f_1^2 & g_2g_3f_1^2 & 0 & g_3g_4f_4^2 & g_3g_5f_4^2 & 2g_2g_3f_1^2g_2g_6f_3^2 \\
2g_1g_3f_1^2g_3g_4f_4^2 & 2g_2g_3f_1^2g_3g_4f_4^2 & g_3g_4f_4^2 & 0 & g_3g_5f_4^2 & 3g_3g_4f_4^2g_2g_3f_1^2g_2g_6f_3^2 \\
2g_1g_3f_1^2g_3g_5f_4^2 & 2g_2g_3f_1^2g_3g_5f_4^2 & g_3g_5f_4^2 & g_3g_5f_4^2 & 0 & 3g_3g_5f_4^2g_2g_3f_1^2g_2g_6f_3^2 \\
2g_1g_2f_1^2g_2g_6f_3^2 & g_2g_6f_3^2 & 2g_2g_3f_1^2g_2g_6f_3^2 & 3g_3g_4f_4^2g_2g_3f_1^2g_2g_6f_3^2 & 3g_3g_5f_4^2g_2g_3f_1^2g_2g_6f_3^2 & 0
\end{bmatrix}
$$

$$^e\mathbb{D}(SF_{\text{DMCA}})$$

Figure 4. The second Zagreb and Randić vertex matrices and the vertex and edge distance matrices for the hydrocarbon-hydrogen-depleted structural formula of the dimethylcarbamic acid molecule.

It can be useful to know that the $(^va_2)_{ik}$ entry in $^v\mathbb{A}_2(SF)$ has the following form: $(f_ig_p^2f_j)(f_jg_r^2f_k) = (f_if_j^2f_k)(g_p^2g_r^2)$, which indicates that the path from the vertex i to the vertex k goes through the vertex j by the edges p (from i to j) and r (from j to k). Similarly, the $(^va_3)_{il}$ entry in $^v\mathbb{A}_3(SF)$ corresponding to the path $i{\to}j{\to}k{\to}l$ through the edges $p{\to}r{\to}s$ has the following form: $(f_ig_p^2f_j)(f_jg_r^2f_k)(f_kg_s^2f_l) = (f_if_j^2f_k^2f_l)(g_p^2g_r^2g_s^2)$.

Analogously, the $(^ea_2)_{ps}$ entry in $^e\mathbb{A}_2(SF)$ eqals: $(g_pf_i^2g_r)(g_rf_j^2g_s) = (g_pg_r^2g_s)(f_i^2f_j^2)$ (the path $p{\to}r{\to}s$ by the vertices i and j), and $(^ea_3)_{pt}$ in $^e\mathbb{A}_3(SF)$ is as follows: $(g_pf_i^2g_r)(g_rf_j^2g_s)(g_sf_k^2g_t) = (g_pg_r^2g_s^2g_t)(f_i^2f_j^2f_k^2)$ (path $p{\to}r{\to}s{\to}t$ through $i{\to}j{\to}k$).

Yet, the following question still remains unanswered: what if different paths connect the same two points i and k but go through different vertices j' and j''? We can see this phenomenon in cyclic structures, such as the paths between vertices 1 and 3 in cyclobutane: $1{\to}2{\to}3$ and $1{\to}4{\to}3$. In unweighted graphs, the corresponding entry is simply 1, but for SF we would obtain two different matrices describing the same structural formula. Therefore, we propose that for structural formulae of a cyclic molecule, the $x_{ij}(SF)$ entries in the $\mathbb{X}(SF)$ matrix can be understood as arithmetic means over all paths of the same length:

$$x_{ij}(SF) = \frac{1}{r}\sum_{p=1}^{r} x_{ij}^p(SF) \tag{24}$$

where r is the number of different paths of the same length from i to j. Assume that in edge matrices, we understand a path as a way through subsequent vertices.

2.2.3 The SF topological indices

For graphs several topological indices are obtained as half of the sum of the off-diagonal elements of topological square symmetric matrix Đ (Croatian D with a stroke) $p(Đ(G))$ [54]. For any square symmetric $Đ_{n\times n}$ matrix of the structural formula, such topological indices $p(Đ(SF))$ can be defined analogously:

$$p(Đ(SF)) = \frac{1}{2} \sum_{i=1, j=1, i\neq j}^{n} đ_{ij}(SF) \tag{25}$$

Indeed, the Wiener, Detour, Zagreb 2, Randić connectivity and Laplacian topological indices of a structural formula can be obtained from distance and the other mentioned SF matrices. However, three versions of the weighted incidence SF matrix (12) can be constructed, be it vertex or edge type of the matrix. In consequence, for each pair of atom and bond properties, three versions of the topological indices can be formulated: (i) the atomic-property-weighted index; (ii) the bond-property-weighted index; and (iii) the mixed atomic-and-bond-properties-weighted index. Thus for instance, three weighted distance vertex matrices can be written and three appropriately weighted Wiener indices can be calculated. The same holds true for the edge distance matrix. Moreover, for the Zagreb 2 and Randić indices, Eq. (19)-(22), even more complex indices may be calculated.

Table 3. The second Zagreb, Randić and Wiener vertex (v) and edge (e) indices for the hydrocarbon-hydrogen-depleted structural formula of dimethylcarbamic acid (DMCA), isobutyric acid (IBA) and two representations of vinyl acetylene (VA) molecules using NBO partial charges, q (e), and bond lengths, R (Å), Wiberg atom indices (a) and bond orders (bo) (Table S1) as the weights for atoms and bonds based on NBO calculations at the B3LYP/6-31G** level (Table S2).

index	DMCA		IBA		index	VA1		VA2	
	v	e	v	e		v	e	v	e
$Z_2(q,R)$	-17.329	26.879	-9.363	17.519	$Z_2(a,bo)$	3850240	573440	2670179.9	649548.9
$R(q,R)$	-0.752	3.742	-0.768	9.699	$R(a,bo)$	1.984	1.008	1.410	1.006
$D(q,R)$	-2.370	20.712	4.252	14.141	$D(a,bo)$	449248	3152	416634.8	3430.2

As an example, observe the Zagreb 2, Randić and Wiener vertex and edge indices determined for the hydrocarbon-hydrogen-depleted structural formulae of the dimethyl carbamic acid, isobutyric acid and vinyl acetylene (Table 3) calculated based on NBO partial charges, bond lengths, atom indices and bond orders (Tables S1 and S2). An inspection of Table 3 indicates that:

    (i)   molecules exhibiting the same graphs but different structural formulae have significantly different SF topological indices;

    (ii) both the vertex and edge topological indices significantly differentiate the structural formulae;

    (iii) the values of the indices, including the Wiener index, can be both negative or positive.

Basic knowledge of computational and structural chemistry also indicates that the SF topological indices depend on:

    (iv) the method with which weighting parameters were obtained; and thus

    (v) the conformation or the physical state in which they were estimated.

The latter two remarks provoke the question of whether SF topological indices can still be called "topological" if they depend on geometrical and/or physical parameterization. Note, however, that SF are objects in spaces without a scalar product or orthogonality: neither angles nor solid angles are necessary to calculate basic topological indices. Thus the spaces used in the SF approach, may, at least, not be standard Cartesian spaces. Moreover, the influence of the physical property of the weight on SF topological index can be minimized by the weight standardization, or by using parameters of the most stable molecular form, etc. Nevertheless eventually, the answer to the above question is authoritative: they are calculated in the same way as GT topological indices and thus it is practical to call them "topological".

On the other hand, why introduce all those SF topological indices if the parametrization requires some physical or quantum chemical studies? Would it not be more natural to analyze the problem based solely on unprocessed quantum chemical data? In our opinion, the SF topological indices are worth calculating for the same reasons as the ordinary GT topological indices: they can be calculated quickly and can introduce molecular characteristics complementary to the common quantum or physical chemistry data. They diversify the pool of potentially useful independent variables in QSAR/QSPR. Thus, it is likely that using them allows one to better model some specific molecular properties with the SF approach than with only GT methods or quantum-chemistry data. However, the selection of molecular

characteristics useful in serious QSAR or QSPR modelling should be preceded by a careful introductory inspection of experimental and SF data.

2.2.4 The HOMA index

Aromaticity is a concept extensively referred to in chemical graph theory [e.g., 58-62] and in physical organic chemistry [e.g., 63-68]. It seems that its best definition is enumerative and comprised of four criteria: 1. energetic, 2. geometric, 3. magnetic, and 4. chemical reactivity [63]. The second most used index after NICS (magnetic aromaticity) [64, 66, 69] is probably the HOMA (geometric aromaticity) index [64, 65, 70, 71]. It is defined as follows:

$$HOMA = 1 - \frac{\alpha}{n} \sum_{i=1}^{n} (R_i - R_{opt})^2 \qquad (26)$$

where R_i and R_{opt} stand for distances of the *i*-th ring bond in the analyzed structure and the reference optimal bond in benzene (1.388 Å), and *n* is the number of CC bonds in the ring, whereas $\alpha = 257.7$ Å^{-2} is a normalization factor guaranteeing that the unitless HOMA index of an aromatic compound approaches 1 and that of its Kekulé non-aromatic structure approaches 0.

Recently we have shown that the HOMA index expresses much more than only aromaticity, because it reasonably characterizes cyclic molecules regardless of whether they are aromatic, non-aromatic, unsaturated or saturated [70]. We named this generalized aromaticity property the *savoricity*. Moreover, we have also demonstrated that HOMA valuably characterizes acyclic structures, whether delocalized or not [71, 72].

Hence, HOMA is a useful, angle-independent, general geometrical index. It can be used to any molecule if the HOMA parametrization for heteroatoms is introduced [73, 74]. However, it becomes useless as a GT topological index, because for any ordinary graph $R_i = R_{opt} = 1$ and HOMA=1. Yet, HOMA does not equal 1 for SF with differentiated atoms and/or bonds. Such a SF is a metric space with the molecular path metric (Section 2.1.2). It is worth adding that not only bond length can constitute the molecular path metric. In Ref. 70, we demonstrated that HOMA can be treated as a function, HOMA($\cdot$), of other bond characteristics such as, for example, electron density (ρ) or its Laplacian in the bond critical point. Thus, HOMA(R), HOMA(ρ), and others may provide precious information and a clear distinction between structures for which pure topological data are degenerated.

2.2.5 An alternative to the presented SF approach

Here, the SF version of GT is constructed based on the modified incidence matrix Eq. (12). After this adjustment no further intervention into the SF matrices is made. Although the incidence matrix contains all information about the described graph, it has been relatively infrequently exploited in the chemical graph theory [75]. Therefore, Jerzy Ciosłowski, the first reader of this paper, rightly noticed that grounding the SF concept on a modified adjacency matrix would probably be more intuitive for the chemical graph theorists community. Indeed, the CGT reasoning usually begins with presentation of the adjacency matrix. Moreover, for the π-electronic systems, the adjacency matrix looks like the Hückel hamiltonian matrix [76-79]. In such SF_A approach, the atom properties would be present at diagonal of the $\mathbb{V}_A$ valency matrix and the bond properties would be located at the off-diagonal elements of the $\mathbb{A}_A$ adjacency matrix. All the next steps can be identical as described in equations (4)-(11). As a side effect, the entries in the $\mathbb{A}_A$ matrix would not be the second powers of the incidence matrix elements which could simplify further expressions. This is indeed good and natural way to construct the SF concept. Nevertheless, we also think that the SF_A approach has some deficiencies or, at least, the approach presented in this paper exhibits some advantages over the SF_A one.

First, by modifying the incidence matrix in our SF version, we obtain not only the vertex but also the edge adjacency and valency matrices. In contrast, in the SF_A construction, obtaining the edge matrices needs new definitions of the diagonal elements in the ${}^e\mathbb{V}_A$ matrix and the off-diagonal elements of the ${}^e\mathbb{A}_A$ ones. Despite at the moment only little attention is being paid to the edge-matrices-derived topological indices, this may be changed since within the SF approach such indices can be better differentiated (Table 3). Notice that obtaining the rectangular incidence matrix through an operation similar to a square root of $\mathbb{A}_A$, and indirectly generate the edge variants of the adjacency matrix, is not easy. Such an operation is not unequivocal although the algorithms for similar procedure are developed quite some time ago [80-82]. Second, using two different diagonal weighting matrices $f(v_m)$ and $g(e_n)$ to define the SF incidence matrix makes it possible to easily generate a large number of closely related, but different, matrices and topological indices. This may significantly improve the pool of topological descriptors for use in QSAR/QSPR analyses. Third, in the SF_A construction, the atom and bond properties are separated in the $\mathbb{V}_A$ and $\mathbb{A}_A$ matrices, respectively, whereas in the SF version presented here, the $\mathbb{V}$ matrix elements contain some contributions from bonds formed by given atoms and the $\mathbb{A}$ matrix elements contain some contributions from atoms linked by the bonds.

Eventually, both approaches are slightly different but seem to be good enough to introduce the structural formula version of the graph theory. After all, good codes for calculating the SF topological indices can help in forgetting about some inconveniences of the formalism behind the approaches and in focusing on examining potential profits that can be earned from using new descriptors. Series of computational experiments and successful applications of topological indices obtained within one or the other formalism can promote one approach over the other.

3. Conclusions

The aim of this paper has been to better connect graph theory to chemistry. It is desirable that in such a modified GT, any molecule considered in chemistry can be studied without loss of information about atoms and bonds present in its structural formula and topological indices can be calculated directly from the formula. To do this we proposed the Structural Formula (SF) approach in which different kinds of vertices and edges are possible. Only a slight modification of GT is sufficient to construct SF: the SF is a pair (**V**, **E**) where **V** is a family of sets of vertices and **E** is a family of sets of edges instead of single V and E sets in definition of a graph. The weighting functions defined on the families of sets allow for easy introduction of atoms and bonds properties to the SF incidence matrix. Then, through simple equations, the chemical information is transferred to other SF matrices such as Zagreb, Randić and distance, and to the corresponding SF topological indices. Finally, we show that the HOMA geometrical aromaticity index can be treated as the SF topological index.

In the SF concept, there is room to color vertices and edges or to use other representations of vertices and/or edges. The SF concept can further be modified. For example, it is easy to include unary and *n*-ary relations between vertices or use hypervertices and obtain a hyper structural formula approach or include a chirality property within a SF. Families of sets can also be replaced by multisets. Finally, it seems that classical GT problems, such as shortest paths, travelling salesman, vertex or edge coloring, can be easily transferred to the SF framework.

References

[1] Cayley, On the mathematical theory of isomers, *Philosoph. Mag.* **67** (1874) 444–446.

[2] W. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.

[3] H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.

[4] M. Randić, On characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.

[5] I. Gutman, B. Ruscić, N. Trinajstić, C. F. Wilcox, Graph theory and molecular orbitals. XII. Acyclic polyenes, *J. Chem. Phys.* **62** (1975) 1692–1704.

[6] F. Harary, E. M. Palmer, R. W. Robinson, R. C. Read, Pólya's contribution to chemical enumeration, in: A. T. Balaban (Ed.), *Chemical Applications of Graph Theory*, Academic Press, London, 1976, pp. 11–24.

[7] I. Gutman, M. Milun, N. Trinajstić, Graph theory and molecular orbitals. 19. Nonparametric resonance energies of arbitrary conjugated systems, *J. Am. Chem. Soc.* **99** (1977) 1692–1704.

[8] M. Randić, Aromaticity of polycyclic conjugated hydrocarbons, *Chem. Rev.* **103** (2003) 3449–3605.

[9] I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.

[10] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 1992.

[11] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley, Weinheim, 2009.

[12] S. C. Basak, G. Restrepo, J. L. Villaveces (Eds.), *Advances in Mathematical Chemistry and Applications*, Bentham, Sharjah, 2014.

[13] R. Chauvin, C. Lepetit, B. Silvi, E. Alikhani (Eds.), *Applications of Topological Methods in Molecular Chemistry*, Springer, Switzerland, 2016.

[14] L. H. Clark, J. W. Moon, On the general Randić index for certain families of trees, *Ars Comb.* **54** (2000) 223–235.

[15] Y. Hou, On acyclic systems with minimal Hosoya index, *Discr. Appl. Math.* **119** (2002) 251–257.

[16] A. A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: Theory and applications, *Acta Appl. Math.* **66** (2001) 211–249.

[17] C. Delorme, O.Favaron, D.Rautenbach, On the Randić index, *Discr. Math.* **257** (2002) 29–38.

[18] G. G. Cash, Relationship between the Hosoya polynomial and the hyper-Wiener

index, *Appl. Math. Lett.* **15** (2002) 893–895.

[19]   H. Wang, The extremal values of the Wiener index of a tree with given degree sequence, *Discr. Appl. Math.* **156** (2008) 2647–2654.

[20]   S. Wagner, I. Gutman, Maxima and minima of the Hosoya index and the Merrifield-Simmons index, a survey of results and techniques, *Acta Appl. Math.* **112** (2010) 323–346.

[21]   L. Feng, A. Ilić, Zagreb, Harary and hyper-Wiener indices of graphs with a given matching number, *Appl. Math. Lett.* **23** (2010) 943–948.

[22]   M. Knor, P. Potočnik, R. Škrekovski, Relationship between the edge-Wiener index and the Gutman index of a graph, *Discr. Appl. Math.* **167** (2014) 197–201.

[23]   H. Lei, H. Yang, Bounds for the sum-Balaban index and (revised) Szeged index of regular graphs, *Appl. Math. Comput.* **268** (2015) 1259–1266.

[24]   L. Piela, *Ideas of Quantum Chemistry*, Elsevier, Amsterdam, 2013.

[25]   T. Helgaker, W. Klopper, D. P. Tew, Quantitative quantum chemistry, *Mol. Phys.* **106** (2008) 2107–2143.

[26]   R. A. Friesner, Ab initio quantum chemistry: Methodology and applications, *Proc. Natl Acad. Sci. USA* **102** (2005) 6648–6653.

[27]   O. E. Polansky, Graphs in chemistry, *MATCH Commun. Math. Comput. Chem.* **1** (1975) 183–196.

[28]   A. Graovac, O. E. Polansky, N. Trinajstić, N. Tyutyulkov, Graph theory in chemistry II. Graph-theoretical description of heteroconjugated molecules, *Z. Naturforsch. A* **30** (1975) 1696–1699.

[29]   J. Aihara, General rules for constructing Hückel molecular orbital characteristic polynomials, *J. Am. Chem. Soc.* **98** (1976) 6840–6844.

[30]   I. Gutman, S. Bosanac, Topological studies on heteroconjugated molecules. The stability of alternant systems with one heteroatom, *Chem. Phys. Lett.* **43** (1976) 371–373.

[31]   I. Gutman, Topological studies on heteroconjugated molecules – Alternant systems with one heteroatom, *Theor. Chim. Acta* **50** (1979) 287–297.

[32]   L. H. Hall, B. Mohney, L. B. Kier, The electrotopological state: an atom index for QSAR, *Quant. Struct. Activ. Rel.* **10** (1991) 43–51.

[33]   B. Ren, New atom-type-based AI topological indices: Application to QSPR studies of aldehydes and ketones, *J. Comp. Aided Mol. Des.* **17** (2003) 607–620.

[34]   L. Pogliani, Core electrons and hydrogen atoms in chemical graph theory, *J. Math. Chem.* **43** (2008) 1233–1255.

[35]   L. Pogliani, From molecular connectivity indices to semiempirical connectivity terms: recent trends in graph theoretical descriptors, *Chem. Rev.* **100** (2000) 3827–3858.

[36]   M. M. Deza, E. Deza, *Encyclopedia of Distances*, Springer, Berlin, 2009.

[37]   A.D. McNaught, A. Wilkinson, *IUPAC. Compendium of Chemical Terminology* (the "Gold Book"), Blackwell, Oxford, 1997.

[38]   L. Pogliani, Model of the physical properties of halides with complete graph-based indices, *Int. J. Quant. Chem.* **102** (2005) 38–52.

[39]   L. Pogliani, A natural graph-theory model for partition and kinetic coefficients, *New J. Chem.* **29** (2005) 1082–1088.

[40]   D. Kaiser, Physics and Feynman's diagrams, *Am. Sci.* **93** (2005) 156–165.

[41]   F. Harary, R. Z. Norman, D. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs*, Wiley, New York, 1965.

[42]   J. M. Gilbert, Strategies for multigraph edge coloring, *Johns Hopkins Appl. Techn. Digest* **23** (2002) 187–201.

[43]   C. Berge, *Hypergraphs*, North–Holland, Amsterdam, 1989.

[44]   D. Stalke, Charge density and chemical bonding, in: D. Mingos, P. Michael (Eds.), *The Chemical Bond I: 100 Years Old and Getting Stronger,* Springer, Berlin, 2016, pp. 57–88.

[45]   R. F. W. Bader, A quantum theory of molecular structure and its applications, *Chem. Rev.* **91** (1991) 893–928.

[46]   H. Jacobsen, Hypovalency – a kinetic-energy density description of a 4c–2e bond, *Dalton Trans.* (2009) 4252–4258.

[47]   E. Tørneng, C. J. Nielsen, P. Klaeboe, H. Hopf, H. Priebe, The i.r., Raman and microwave spectra of 1-butene-3-yne (vinylacetylene) and 1-butene-3-yne-4d, *Spectrochim. Acta A* **36** (1980) 975–987.

[48]   Y. S. Kim, R. I. Kaiser, An infrared spectroscopic study of amorphous and crystalline ices of vinylacetylene and implications for Saturn's satellite Titan, *Astrophys. J. Suppl. Ser.* **181** (2009) 543–547.

[49]   E. Estrada, L. A. Montero, Bond order weighted graphs in molecules as structure-property indices, *Mol. Eng.* **2** (1993) 363–373.

[50]   E. Estrada, Three-dimensional molecular descriptors based on electron charge density weighted graphs, *J. Chem. Inf. Comput. Sci.* **35** (1995) 708–713.

[51]   O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, A. T. Balaban, Comparison of weighting schemes for molecular graph descriptors: application in quantitative structure-retention relationship models for alkylphenols in gas-liquid chromatography, *J. Chem. Inf. Comput. Sci.* **40** (2000) 732–743.

[52] C. F. Matta, I. Sumar, R. Cook, P. W. Ayers, Localization-delocalization matrices and electron density-weighted adjacency/connectivity matrices: A bridge between the quantum theory of atoms in molecules and chemical graph theory, in: R. Chauvin, C. Lepetit, B. Silvi, E. Alikhani (Eds), *Applications of Topological Methods in Molecular Chemistry*, Springer, Switzerland, 2016, pp. 53–88.

[53] K. B. Wiberg, Application of the Pople-Santry-Segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane, *Tetrahedron* **24** (1968) 1083–1096.

[54] J. C. Dobrowolski, The chiral graph theory, *MATCH Commun. Math. Comput. Chem.* **73** (2015) 347–374.

[55] D. Plavšić, S. Nikolić, N. Trinajstić, Z. Mihalić, On Harary index for the characterization of chemical graphs, *J. Math. Chem.* **12** (1993) 235–250.

[56] D. Amić, N. Trinajstić, On the detour matrix, *Croat. Chem. Acta* **68** (1995) 53–62.

[57] N. Trinajstić, S. Nikolić, B. Lucić, D. Amić, Z. Mihalić, The detour matrix in chemistry, *J. Chem. Inf. Comput. Sci.* **37** (1997) 631–638.

[58] I. Gutman, Total π-electron energy of benzenoid hydrocarbons, *Top. Curr. Chem.* **162** (1992) 30–63.

[59] M. Randić, Aromaticity of polycyclic conjugated hydrocarbons, *Chem. Rev.* **103** (2003) 3449–3605.

[60] I. Lukovits, A. Graovac, E. Kálmán, G. Kaptay, P. Nagy, S. Nikolić, J. Sytchev, N. Trinajstić, Nanotubes: Number of Kekulé structures and aromaticity, *J. Chem. Inf. Comput. Sci.* **43** (2003) 609–614.

[61] P. Schwerdtfeger, L. N Wirz, J. Avery, The topology of fullerenes, *WIREs Comput. Mol. Sci.* **5** (2015) 96–145.

[62] J. Aihara, Graph theory of ring-current diamagnetism, *Bull. Chem. Soc. Jpn.* **91** (2018) 274–303.

[63] T. M. Krygowski, H. Szatylowicz, Aromaticity: what does it mean? *Chem. Texts* **1** (2015) 1–12.

[64] M. K. Cyrański, T. M. Krygowski, A. R. Katritzky, P. von Ragué Schleyer, To what extent can aromaticity be defined uniquely? *J. Org. Chem.* **67** (2002) 1333–1338.

[65] T. M. Krygowski, M. K. Cyrański, Structural aspects of aromaticity, *Chem. Rev.* **101** (2001) 1385–1420.

[66] Z. Chen, C. S. Wannere, C. Corminboeuf, R. Puchta, P. von Ragué Schleyer, Nucleus-independent chemical shifts (NICS) as an aromaticity criterion, *Chem. Rev.* **105** (2005) 3842–3888.

[67] E. Steiner, P. W. Fowler, A. Soncini, L. W. Jenneskens, Current-density maps as probes of aromaticity: global and Clar π ring currents in totally resonant polycyclic aromatic hydrocarbons, *Faraday Discuss.* **135** (2007) 309–323.

[68]   P. W. Fowler, C. M. Gibson, D. E. Bean, Writing with ring currents: selectively hydro-genated polycyclic aromatics as finite models of graphene and graphane, *Proc. R. Soc. A* **470** (2013) #0617.

[69]   J. C. Dobrowolski, P. F. J. Lipiński, On splitting of the NICS(1) magnetic aromaticity index, *RSC Adv.* **6** (2016) 23900–23904.

[70]   S. Ostrowski, J. Cz. Dobrowolski, What does the HOMA index really measure? *RSC Adv.* **4** (2014) 44158–44161.

[71]   J. C. Dobrowolski, S. Ostrowski, On the HOMA index of some acyclic and conducting systems, *RSC Adv*. **5** (2015) 9467–9471.

[72]   J. C. Dobrowolski, P. F. J. Lipiński, S. Ostrowski, M. H. Jamróz, J. E. Rode, The influence of the position of a chiral substituent on undecathiophene chain. A DFT study, *Synth. Met.* **242** (2018) 73–82.

[73]   T. M. Krygowski, M. K. Cyrański, Separation of the energetic and geometric contributions to the aromaticity. Part IV. A general model for the π-electron systems, *Tetrahedron* **52** (1996) 10255–10264.

[74]   K. K. Zborowski, I. Alkorta, J. Elguero, L. M. Proniewicz, HOMA parameters for the boron–boron bond: How the introduction of a BB bond influences the aromaticity of selected hydrocarbons, *Struct. Chem.* **24** (2013) 543–548.

[75]   D. Janežič, A. Miličević, S. Nikolić, N. Trinajstić, *Graph Theoretical Matrices in Chemistry*, Univ. Kragujevac, Kragujevac, 2007, pp. 50–61.

[76]   E. Hückel, Grundzüge der Theorie ungesättigter und aromatischer Verbindungen, *Z. Elektrochem.* **43** (1937) 752–788.

[77]   H. H. Günthard, H. Primas, Zusammenhang von Graphentheorie und MO-Theorie von Molekeln mit Systemen konjugierter Bindungen, *Helv.Chim. Acta* **39** (1956)1645–1653.

[78]   A. J. Sadlej, *Elementary methods of quantum chemistry*, PWN, Warsaw, 1966, (in Polish).

[79]   D. H. Rouvray, Graph theory in chemistry, *R. Inst. Chem. Rev.* **4** (1971) 173–195.

[80]   R. Penrose, A generalized inverse for matrices, *Proc. Camb. Phil. Soc.* **51** (1955) 406–413.

[81]   M. Warmus, *Generalized Inverses of Matrices*, PWN, Warsaw, 1972, (in Polish).

[82]   M. Warmus, *Vectors and Matrices*, PWN, Warsaw, 1981, (in Polish).

APPENDIX

DMCA   IBA
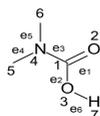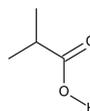
$$
\begin{bmatrix}
1.145 & 1.281 & 1.276 & 0 & 0 & 0 \\
-0.819 & 0 & 0 & 0 & 0 & 0 \\
0 & -1.016 & 0 & 0 & 0 & -0.720 \\
0 & 0 & -0.671 & -0.717 & -0.716 & 0 \\
0 & 0 & 0 & -0.699 & 0 & 0 \\
0 & 0 & 0 & 0 & -0.697 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.497
\end{bmatrix}
\qquad
\begin{bmatrix}
1.014 & 1.137 & 1.272 & 0 & 0 & 0 \\
-0.730 & 0 & 0 & 0 & 0 & 0 \\
0 & -0.98 & 0 & 0 & 0 & -0.701 \\
0 & 0 & -0.546 & -0.554 & -0.549 & 0 \\
0 & 0 & 0 & -1.052 & 0 & 0 \\
0 & 0 & 0 & 0 & -1.051 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.488
\end{bmatrix}
$$

$I(SF)(q,R)$

$$
\begin{bmatrix}
4.580 & -0.938 & -1.301 & -0.856 & 0 & 0 & 0 \\
-0.938 & 0.671 & 0 & 0 & 0 & 0 & 0 \\
-1.301 & 0 & 1.551 & 0 & 0 & 0 & -0.358 \\
-0.856 & 0 & 0 & 1.477 & 0.501 & 0.499 & 0 \\
0 & 0 & 0 & 0.501 & 0.489 & 0 & 0 \\
0 & 0 & 0 & 0.499 & 0 & 0.486 & 0 \\
0 & 0 & -0.358 & 0 & 0 & 0 & 0.247
\end{bmatrix}
\qquad
\begin{bmatrix}
3.939 & -0.740 & -1.114 & -0.695 & 0 & 0 & 0 \\
-0.740 & 0.533 & 0 & 0 & 0 & 0 & 0 \\
-1.114 & 0 & 1.452 & 0 & 0 & 0 & -0.342 \\
-0.695 & 0 & 0 & 0.906 & 0.583 & 0.577 & 0 \\
0 & 0 & 0 & 0.583 & 1.107 & 0 & 0 \\
0 & 0 & 0 & 0.577 & 0 & 1.105 & 0 \\
0 & 0 & -0.342 & 0 & 0 & 0 & 0.238
\end{bmatrix}
$$

$^{v}A(SF)\ (q,R)+{}^{v}V(SF)\ (q,R)$

$$
\begin{bmatrix}
1.982 & 1.467 & 1.461 & 0 & 0 & 0 \\
1.467 & 2.673 & 1.635 & 0 & 0 & 0.732 \\
1.461 & 1.635 & 2.078 & 0.481 & 0.480 & 0 \\
0 & 0 & 0.481 & 1.003 & 0.513 & 0 \\
0 & 0 & 0.480 & 0.513 & 0.998 & 0 \\
0 & 0.732 & 0 & 0 & 0 & 0.765
\end{bmatrix}
\qquad
\begin{bmatrix}
1.561 & 1.153 & 1.290 & 0 & 0 & 0 \\
1.153 & 2.253 & 1.446 & 0 & 0 & 0.687 \\
1.290 & 1.446 & 1.916 & 0.302 & 0.300 & 0 \\
0 & 0 & 0.302 & 1.414 & 0.304 & 0 \\
0 & 0 & 0.300 & 0.304 & 1.406 & 0 \\
0 & 0.687 & 0 & 0 & 0 & 0.730
\end{bmatrix}
$$

$^{e}A(SF)\ (q,R)+{}^{e}V(SF)\ (q,R)$

$$
\begin{bmatrix}
0 & -2.882 & -9.242 & -5.789 & 0 & 0 & 0 \\
-2.882 & 0 & 0 & 0 & 0 & 0 & 0 \\
-9.242 & 0 & 0 & 0 & 0 & 0 & -0.137 \\
-5.789 & 0 & 0 & 0 & 0.362 & 0.359 & 0 \\
0 & 0 & 0 & 0.362 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.359 & 0 & 0 & 0 \\
0 & 0 & -0.136 & 0 & 0 & 0 & 0
\end{bmatrix}
\qquad
\begin{bmatrix}
0 & -1.553 & -6.372 & -2.482 & 0 & 0 & 0 \\
-1.553 & 0 & 0 & 0 & 0 & 0 & 0 \\
-6.372 & 0 & 0 & 0 & 0 & 0 & -0.118 \\
-2.482 & 0 & 0 & 0 & 0.584 & 0.578 & 0 \\
0 & 0 & 0 & 0.584 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.578 & 0 & 0 & 0 \\
0 & 0 & -0.118 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$^{v}Z_2(SF)\ (q,R)$

$$
\begin{bmatrix}
0 & 7.772 & 6.017 & 0 & 0 & 0 \\
7.772 & 0 & 9.082 & 0 & 0 & 1.497 \\
6.017 & 9.082 & 0 & 1.002 & 0.995 & 0 \\
0 & 0 & 1.002 & 0 & 0.514 & 0 \\
0 & 0 & 0.995 & 0.514 & 0 & 0 \\
0 & 1.497 & 0 & 0 & 0 & 0
\end{bmatrix}
\qquad
\begin{bmatrix}
0 & 4.055 & 3.859 & 0 & 0 & 0 \\
4.055 & 0 & 6.242 & 0 & 0 & 1.131 \\
3.859 & 6.242 & 0 & 0.819 & 0.808 & 0 \\
0 & 0 & 0.819 & 0 & 0.605 & 0 \\
0 & 0 & 0.808 & 0.605 & 0 & 0 \\
0 & 1.131 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$^{e}Z_2(SF)\ (q,R)$

Figure S1. Comparison of incidence, adjacency, valency and second Zagreb matrices for hydrocarbon-hydrogen-depleted structural formula of the dimethylcarbamic acid (DMCA) and isobutyric acid (IBA) based on NBO partial charge q and bond distance R.

DMCA

IBA
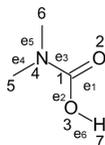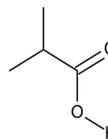
$$
\begin{matrix}
0 & -0.535 & -0.488 & -0.329 & 0 & 0 & 0 \\
-0.535 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.488 & 0 & 0 & 0 & 0 & 0 & -0.578 \\
-0.329 & 0 & 0 & 0 & 0.589 & 0.589 & 0 \\
0 & 0 & 0 & 0.589 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.589 & 0 & 0 & 0 \\
0 & 0 & -0.578 & 0 & 0 & 0 & 0
\end{matrix}
\qquad
\begin{matrix}
0 & -0.511 & -0.466 & -0.368 & 0 & 0 & 0 \\
-0.511 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.466 & 0 & 0 & 0 & 0 & 0 & -0.582 \\
-0.368 & 0 & 0 & 0 & 0.582 & 0.577 & 0 \\
0 & 0 & 0 & 0.582 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.577 & 0 & 0 & 0 \\
0 & 0 & -0.582 & 0 & 0 & 0 & 0
\end{matrix}
$$

$${}^{v}R(SF)\,(q,R)$$

$$
\begin{matrix}
0 & 0.637 & 0.720 & 0 & 0 & 0 \\
0.637 & 0 & 0.694 & 0 & 0 & 0.512 \\
0.720 & 0.694 & 0 & 0.333 & 0.333 & 0 \\
0 & 0 & 0.333 & 0 & 0.513 & 0 \\
0 & 0 & 0.333 & 0.513 & 0 & 0 \\
0 & 0.512 & 0 & 0 & 0 & 0
\end{matrix}
\qquad
\begin{matrix}
0 & 2.163 & 2.231 & 0 & 0 & 0 \\
2.163 & 0 & 3.005 & 0 & 0 & 0.882 \\
2.231 & 3.005 & 0 & 0.497 & 0.493 & 0 \\
0 & 0 & 0.497 & 0 & 0.428 & 0 \\
0 & 0 & 0.493 & 0.428 & 0 & 0 \\
0 & 0.882 & 0 & 0 & 0 & 0
\end{matrix}
$$

$${}^{e}R(SF)\,(q,R)$$

$$
\begin{matrix}
0 & -0.938 & -1.301 & -0.856 & -0.858 & -0.854 & 0.932 \\
-0.938 & 0 & -0.480 & -1.149 & -0.863 & -0.860 & 0.938 \\
-1.301 & -0.480 & 0 & 2.227 & 1.674 & 1.667 & -0.358 \\
-0.856 & -1.149 & 2.227 & 0 & 0.501 & 0.499 & -1.196 \\
-0.858 & -0.863 & 1.674 & 0.501 & 0 & 0.500 & -0.799 \\
-0.854 & -0.860 & 1.667 & 0.499 & 0.500 & 0 & -0.796 \\
0.932 & 0.938 & -0.358 & -1.196 & -0.799 & -0.796 & 0
\end{matrix}
$$

$$
\begin{matrix}
0 & -0.740 & -1.114 & -0.695 & -0.810 & -0.802 & 0.762 \\
-0.740 & 0 & 1.649 & 1.029 & 0.900 & 0.890 & -0.528 \\
-1.114 & 1.649 & 0 & 1.548 & 1.354 & 1.340 & -0.342 \\
-0.695 & 1.029 & 1.548 & 0 & 0.583 & 0.577 & -0.794 \\
-0.810 & 0.900 & 1.354 & 0.583 & 0 & 0.673 & -0.617 \\
-0.802 & 0.890 & 1.340 & 0.577 & 0.673 & 0 & -0.611 \\
0.762 & -0.528 & -0.342 & -0.794 & -0.617 & -0.611 & 0
\end{matrix}
$$

$${}^{v}D(SF)\,(q,R)$$

$$
\begin{matrix}
0 & 1.467 & 1.461 & 1.405 & 1.403 & 2.148 \\
1.467 & 0 & 1.635 & 1.573 & 1.570 & 0.732 \\
1.461 & 1.635 & 0 & 0.481 & 0.480 & 2.394 \\
1.405 & 1.573 & 0.481 & 0 & 0.513 & 1.727 \\
1.403 & 1.570 & 0.480 & 0.513 & 0 & 1.723 \\
2.148 & 0.732 & 2.394 & 1.727 & 1.723 & 0
\end{matrix}
\qquad
\begin{matrix}
0 & 1.153 & 1.290 & 0.779 & 0.774 & 1.584 \\
1.153 & 0 & 1.446 & 0.873 & 0.868 & 0.687 \\
1.290 & 1.446 & 0 & 0.302 & 0.300 & 1.987 \\
0.779 & 0.873 & 0.302 & 0 & 0.304 & 0.900 \\
0.774 & 0.868 & 0.300 & 0.304 & 0 & 0.894 \\
1.584 & 0.687 & 1.987 & 0.900 & 0.894 & 0
\end{matrix}
$$

$${}^{e}D(SF)\,(q,R)$$

Figure S2. Comparison of vertex and edge Randić and distance matrices for hydrocarbon-hydrogen-depleted structural formula of the dimethylcarbamic acid (DMCA) and isobutyric acid (IBA) based on NBO partial charge (q) and bond distance (R).

VA1   VA2

$$
\begin{vmatrix} 8 & 0 & 0 \\ 8 & 4 & 0 \\ 0 & 4 & 12 \\ 0 & 0 & 12 \end{vmatrix}
\qquad
\begin{vmatrix} 7.367 & 0 & 0 \\ 7.431 & 4.452 & 0 \\ 0 & 4.505 & 11.227 \\ 0 & 0 & 11.067 \end{vmatrix}
$$

I(*SF*)(a,bo)

$$
\begin{vmatrix} 64 & 64 & 0 & 0 \\ 64 & 80 & 16 & 0 \\ 0 & 16 & 160 & 144 \\ 0 & 0 & 144 & 144 \end{vmatrix}
\qquad
\begin{vmatrix} 54.273 & 54.744 & 0 & 0 \\ 54.744 & 75.040 & 20.056 & 0 \\ 0 & 20.056 & 146.341 & 124.249 \\ 0 & 0 & 124.249 & 122.478 \end{vmatrix}
$$

^{v}A(*SF*) (a,bo)+^{v}V(*SF*) (a,bo)

$$
\begin{vmatrix} 128 & 32 & 0 \\ 32 & 32 & 48 \\ 0 & 48 & 288 \end{vmatrix}
\qquad
\begin{vmatrix} 109.492 & 33.083 & 0 \\ 33.083 & 40.115 & 50.578 \\ 0 & 50.578 & 248.524 \end{vmatrix}
$$

^{e}A(*SF*) (a,bo)+^{e}V(*SF*) (a,bo)

$$
\begin{vmatrix} 0 & 327680 & 0 & 0 \\ 327680 & 0 & 204800 & 0 \\ 0 & 204800 & 0 & 3317760 \\ 0 & 0 & 3317760 & 0 \end{vmatrix}
\qquad
\begin{vmatrix} 0 & 222952.9 & 0 & 0 \\ 222952.9 & 0 & 220243.5 & 0 \\ 0 & 220243.5 & 0 & 2226983.5 \\ 0 & 0 & 2226983.5 & 0 \end{vmatrix}
$$

^{v}Z$_2$(*SF*) (a,bo)

$$
\begin{vmatrix} 0 & 131072 & 0 \\ 131072 & 0 & 442368 \\ 0 & 442368 & 0 \end{vmatrix}
\qquad
\begin{vmatrix} 0 & 145309.527 & 0 \\ 145309.527 & 0 & 504239.35 \\ 0 & 504239.35 & 0 \end{vmatrix}
$$

^{e}Z$_2$(*SF*) (a,bo)

$$
\begin{vmatrix} 0 & 0.894 & 0 & 0 \\ 0.894 & 0 & 0.141 & 0 \\ 0 & 0.141 & 0 & 0.949 \\ 0 & 0 & 0.949 & 0 \end{vmatrix}
\qquad
\begin{vmatrix} 0 & 0.858 & 0 & 0 \\ 0.858 & 0 & 0.191 & 0 \\ 0 & 0.191 & 0 & 0.361 \\ 0 & 0 & 0.361 & 0 \end{vmatrix}
$$

^{v}R(*SF*) (a,bo)

$$
\begin{vmatrix} 0 & 0.500 & 0 \\ 0.500 & 0 & 0.508 \\ 0 & 0.508 & 0 \end{vmatrix}
\qquad
\begin{vmatrix} 0 & 0.499 & 0 \\ 0.499 & 0 & 0.507 \\ 0 & 0.507 & 0 \end{vmatrix}
$$

^{e}R(*SF*) (a,bo)

$$
\begin{vmatrix} 0 & 64 & 2048 & 442368 \\ 64 & 0 & 16 & 4608 \\ 2048 & 16 & 0 & 144 \\ 442368 & 4608 & 144 & 0 \end{vmatrix}
\qquad
\begin{vmatrix} 0 & 54.744 & 2195.892 & 409256 \\ 54.744 & 0 & 20.056 & 4983.876 \\ 2195.892 & 20.056 & 0 & 124.249 \\ 409256 & 4983.876 & 124.249 & 0 \end{vmatrix}
$$

^{v}D(*SF*) (a,bo)

$$
\begin{vmatrix} 0 & 32 & 3072 \\ 32 & 0 & 48 \\ 3072 & 48 & 0 \end{vmatrix}
\qquad
\begin{vmatrix} 0 & 33.083 & 3346.544 \\ 33.083 & 0 & 50.578 \\ 3346.544 & 50.578 & 0 \end{vmatrix}
$$

^{e}D(*SF*) (a,bo)

Figure S3. Comparison of vertex and edge Randić and distance matrices for hydrocarbon-hydrogen-depleted two structural formulae of the vinyl acetylene (VA) based on NBO calculated atom index (a) and bond order (bo) (Table 2).

Table S1. Example of weighting functions for structural formula of vinylacetylene based on NBO calculations at the B3LYP/6-31G** level. The Wiberg atom index and bond order are sums of square of corresponding elements of the density matrix in NAO basis [53].

atom	valency		partial NBO charge		bond	multiplicity	
	nominal	Wiberg atom index	summed into C-atoms	NBO		nominal	Wiberg bond order
C_1	4	3.906	0.070	-0.381	$C_1=C_2$	2	1.886
C_2	4	3.940	-0.049	-0.301	C_2-C_3	1	1.130
C_3	4	3.987	-0.053	-0.053	$C_3\equiv C_4$	3	2.816
C_4	4	3.930	0.033	-0.211	C_1-H_8	1	0.929
H_5	1	0.942	0.000	0.244	C_1-H_7	1	0.931
H_6	1	0.939	0.000	0.252	C_2-H_6	1	0.889
H_7	1	0.951	0.000	0.225	C_4-H_5	1	0.926
H_8	1	0.951	0.000	0.226			

Table S2. The NBO partial charges (e) and bond lengths (Å) as the weighs for atoms and bonds for hydrocarbon-hydrogen-depleted structural formula of the dimethylcarbamic acid (DMCA) and isobutyric acid (IBA) molecules based on NBO calculations at the B3LYP/6-31G** level.

atom	NBO partial charge		bond	bond length	
	DMCA	IBA		DMCA	IBA
C1	0.937	0.837	C1=O2	1.222	1.212
O2	-0.670	-0.602	C1-O3	1.367	1.359
O3	-0.743	-0.721	C1-N4(C4)	1.362	1.520
N4(C4)	-0.493	-0.359	N4-C5(C4-C5)	1.454	1.543
C5	-0.481	-0.682	N4-C6(C4-C6)	1.453	1.530
C6	-0.480	-0.687	O3-H7	0.969	0.972
H7	0.513	0.502			