

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220190361>

Application of hypergraph theory in chemistry

Article in *Discrete Mathematics* · May 2001

DOI: 10.1016/S0012-365X(00)00290-9 · Source: DBLP

CITATIONS

47

READS

3,185

2 authors, including:



Elena Konstantinova

Russian Academy of Sciences

55 PUBLICATIONS 530 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Combinatorial Gray codes and Hamiltonicity of Cayley graphs [View project](#)



ELSEVIER

Discrete Mathematics 235 (2001) 365–383

DISCRETE
MATHEMATICSwww.elsevier.com/locate/disc

Application of hypergraph theory in chemistry

Elena V. Konstantinova*, Vladimir A. Skorobogatov

Sobolev Institute of Mathematics, Novosibirsk-90, 630090 790-784, Russia

Abstract

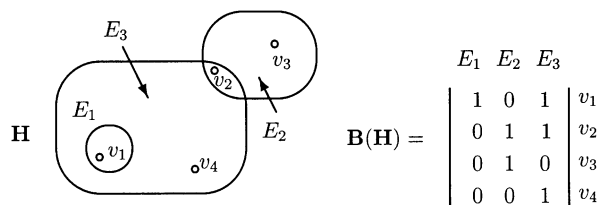
We investigate hypergraphs as mathematical model for representation of nonclassical molecular structures with polycentric delocalized bonds. The questions of identification and canonical coding of molecular hypergraphs are considered. In order to recognize isomorphic hypergraphs, the canonical forms of incidence matrices are to be found. The algorithm for construction of canonical incidence matrix of hypergraph is proposed. Some chemical problems dealing with the hypergraph theory are discussed. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

Graph representation of molecular structures is widely used in computational chemistry and theoretical chemical researches [47,10,42,4]. Trinajstić noted [47] that the roots of chemical graph theory may be found in the works by chemists of 18–19th centuries such as Higgins, Kopp, Crum Brown. First chemical graphs for representing molecules were used by them. Molecular structures are represented by graphs where vertices correspond to atoms, and edges to chemical bonds. This kind of graph, called now a molecular graph, is the object of study in the theory of ordinary graphs [25]. However, ordinary graphs do not adequately describe chemical compounds of nonclassical structure. A substantial drawback of the structure theory is the lack of a convenient representation for molecules with delocalized polycentric bonds [2]. Organometallic compounds are an example of structures that have at least one metal–carbon bond [24]. Some of the graph models used for the representation of organometallic compounds were considered in Refs. [43,46]. Sometimes, disconnected molecular graphs are used for representing ‘sandwich’ and olefin structures. This representation does not seem to be illustrative and does not allow to analyze a structure as a whole, because there are no connections between molecular subgraphs representing individual structural fragments in it. More illustrative, but still not devoid of drawbacks, are the connected molecular graphs, where all the vertices corresponding to carbon atoms are connected to the metal vertex, which corresponds to the metal atom. The degree of the metal

* Corresponding author.

E-mail address: e.konsta@math.nsc.ru (E.V. Konstantinova).

Fig. 1. Hypergraph \mathbf{H} and its incidence matrix.

vertex in this case is equal to the number of vertices connected and not necessarily equal to the valency of the metal atom. Besides, in both representations, the difference between simple covalent and polycentric bonds is obscured. Sometimes planar graphs are used to describe organometallic compounds to study the stereochemical properties. The pyramids obtained from the fragments of the complex are transformed to planar graphs. In this case, the disconnected planar graph describes the structures.

All the defects of the structure representation that were mentioned above are eliminated if the hypergraphs are used to represent the structures with delocalized polycentric bonds [29]. We consider some aspects of application of hypergraph theory in organometallic chemistry. Only unlabeled hypergraphs will be considered in this paper. Various ways of labeled hypergraph representation of molecular structures with different levels of detail and diverse types of label were considered in [30].

2. Molecular hypergraph

Let us introduce some concepts from the hypergraph theory [7,50].

Hypergraph $\mathbf{H}=(\mathbf{V},\mathbf{E})$ consists of a non-empty set of vertices $\mathbf{V}=\{v_i \mid i=1,\dots,p\}$ and a family $\mathbf{E}=\{E_j \mid j=1,\dots,q\}$ of different subsets of the set of vertices. E_j sets are called edges of a hypergraph or hyperedges. An example of hypergraph is given in Fig. 1. If $v_i \in E_j$ we say that vertex v_i is incident to edge E_j . The cardinality of the set of all edges incident to vertex v_i is called the degree of vertex v_i and is denoted as $\deg v_i$. The cardinality of the set of all vertices incident to the edge E_j of a hypergraph gives the degree of the edge E_j which is denoted as $\deg E_j$. The degrees of edges E_1, E_2 and E_3 presented in Fig. 1 are equal to one, two and three, respectively; vertex v_2 is incident to edges E_2 and E_3 ($\deg v_2=2$), and vertex v_4 is incident only to edge E_3 .

An ordinary graph is a special case of a hypergraph with degrees of all the edges equal to two.

A finite hypergraph (set \mathbf{V} and family \mathbf{E} are finite) is unambiguously defined by an incidence matrix $\mathbf{B}(\mathbf{H})=||b_{ij}||$, $i=1,\dots,p$, $j=1,\dots,q$, where

$$b_{ij} = \begin{cases} 1 & \text{if } v_i \in E_j, \\ 0 & \text{if } v_i \notin E_j. \end{cases}$$

The incidence matrix of hypergraph \mathbf{H} is also shown in Fig. 1.

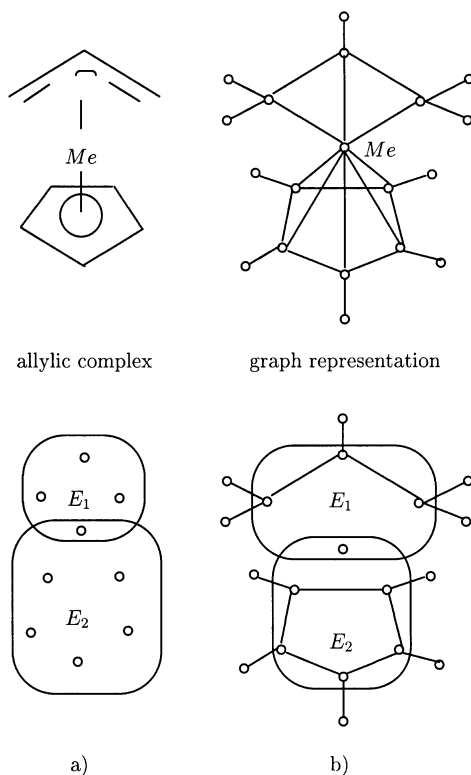


Fig. 2. Graph and hypergraph representation of allylic complex.

A hypergraph $H = (V, E)$ is a *molecular hypergraph* if it represents molecular structure F where vertices $v \in V(H)$ correspond to individual atoms, hyperedges $E \in E(H)$ with degrees greater than 2 correspond to delocalized polycentric bonds and hyperedges $E \in E(H)$ with $\deg E = 2$ correspond to simple covalent bonds. Hyperedges of degree two will be shown as ordinary edges to stress the difference between simple covalent bonds and polycentric bonds.

Fig. 2 illustrates the difference between graphs and hypergraphs in the representation of the same molecular structure of allylic complex. In the case of an ordinary graph, the metal–ligand bonds in allylic complex are shown by eight edges connecting the vertex representing metal Me with the other eight vertices. This drawback is eliminated when allylic complex is represented by hypergraph. Hyperedges E_1 and E_2 represent π -bonds between metal and ligands (cases a) and b)). Moreover, the representation of allylic complex as a hypergraph makes it possible to visually illustrate the difference between sigma (carbon–carbon and carbon–hydrogen) bonds and π -bonds (case b)). In the hypergraph representations, the valency of the metal atom is being ‘conserved’, because the degree of metal vertex in the hypergraph is equal to two. The degree of the same vertex of the graph is equal to eight. Note also that valencies of carbon atoms

are being conserved under the hypergraph representation. As it seems, in Fig. 2b is given an adequate hypergraph representation of an allylic complex.

We have to note that the delocalized bonds which can be described by hyperedges in a molecular hypergraph are not found in organometallic compounds alone. The same bonds can be also found in conjugated molecules [16]. A special symbol is introduced to denote the six π -electrons — a circle inside a hexagon. Single hyperedge of the unlabeled hypergraph is incident to all the six vertices and represents the bonding of all six π -electrons in the benzene. The valency of a carbon atom is being conserved in this case.

3. Identification of molecular hypergraph

The problem, known as the identification problem, is to decide whether two given hypergraphs are isomorphic or not. Two hypergraphs are *isomorphic* if there exists a mapping conserving the one-to-one correlation between the sets of vertices and edges of these hypergraphs.

Most conventional way to define molecular graph G is to determine the adjacency matrix $A(G) = ||a_{ij}||$, $i, j = 1, \dots, p$, where $a_{ij} = 1$ if vertices v_i and v_j are adjacent and $a_{ij} = 0$ otherwise. The adjacency matrix characterizes the graphs completely. In order to recognize isomorphic graphs the canonical forms of adjacency matrices are to be found. The canonical adjacency matrix is defined on the basis of the maximal matrix. For two binary matrices of the same dimension, the greater matrix is the one, in which the first row that differs from the corresponding row of the second matrix is lexicographically greater. The maximal of all possible matrices is called canonical matrix. A number of methods of defining the canonical form of an adjacency matrix are described in [1,17,39,27,26,32].

The incidence matrix is convenient for the computer representation of hypergraphs. Let us define the canonical incidence matrix of hypergraph. The canonical $(p \times q)$ -incidence matrix $B^*(H)$ of hypergraph H is the maximal of all possible $(p! \cdot q!)$ incidence matrices. The definition of a canonical incidence matrix for a hypergraph is also valid for a graph [5,36]. It is evident that the canonical incidence matrix is unique and can be used for determining the canonical form of the hypergraph. The hypergraph defined by canonical incidence matrix is the canonical form of a hypergraph. The numbering of vertices and edges of a canonical form of a hypergraph is called canonical. The canonical form of the hypergraph for hydrogen-deleted structure of allylic complex (Fig. 2) and the canonical incidence matrix are shown in Fig. 3.

There are works [38,8] dealing with the theoretic aspects of isomorphism problems for hypergraph. We propose the algorithmic approach to solve this problem. The canonical incidence matrix may be obtained on the basis of a simple algorithm that uses the representation of hypergraph $H = (V, E)$ as a bipartite graph $K(H) = (V, E, Y)$ (further also referred to as König's representation or König's graph [50]) with the set

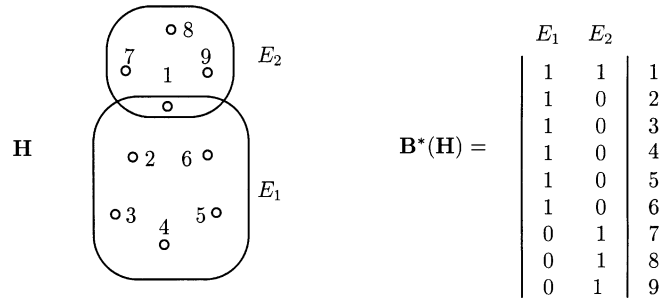


Fig. 3. Canonical form and canonical incidence matrix of a hypergraph of an allylic complex.

of vertices $V \cup E$, $|V \cup E| = p + q = s$, and the set of edges Y , $|Y| = \sum_{j=1}^q \deg E_j$. Sets V and E are called parts of a graph: V -part and E -part. Vertices $v_i \in V$ and $E_j \in E$ in $K(H)$ are adjacent if and only if vertex v_i and edge E_j are incident in H .

Any finite hypergraph has König's representation, and vice versa, any König's graph is a representation of some finite hypergraph H and defines it unambiguously [50]. Hence, the study of the properties of hypergraph can be reduced to the study of analogous properties of König's representation.

In particular, the canonical incidence matrix of hypergraph H can be obtained from the canonical adjacency matrix of König's representation $K(H)$ of hypergraph H .

Let $A(K(H)) = ||a_{ij}||$, $i, j = 1, \dots, s$, be adjacency matrix of graph $K(H)$, where $a_{ij} = 1$ if vertices v_i and v_j are adjacent and $a_{ij} = 0$ otherwise. The canonical $(s \times s)$ -adjacency matrix $A^*(K(H))$ of the graph $K(H)$ is the maximal of all possible $s!$ adjacency matrices.

Beside the adjacency matrix of a bipartite graph, the reduced adjacency matrix $R(K(H)) = ||r_{ij}||$, $i = 1, \dots, p$, $j = 1, \dots, q$, is also considered, where $r_{ij} = 1$ if vertices v_i and E_j are adjacent and $r_{ij} = 0$ otherwise.

It is evident that matrices $R(K(H))$ and $B(H)$ are defined the same way. If, analogous to the canonical incidence matrix $B^*(H)$ of hypergraph H , one defines the canonical reduced adjacency matrix $R^*(K(H))$ for bipartite graph $K(H)$, the following theorem is true:

Theorem 1. $R^*(K(H)) = B^*(H)$.

The proof follows from the definition of the canonical incidence matrices $B^*(H)$ and $R^*(K(H))$.

Hence, the search of canonical incidence matrix $B^*(H)$ is restricted to the search of canonical reduced adjacency matrix $R^*(K(H))$. The last one can be obtained from matrix $A^*(K(H))$ according to the following theorems:

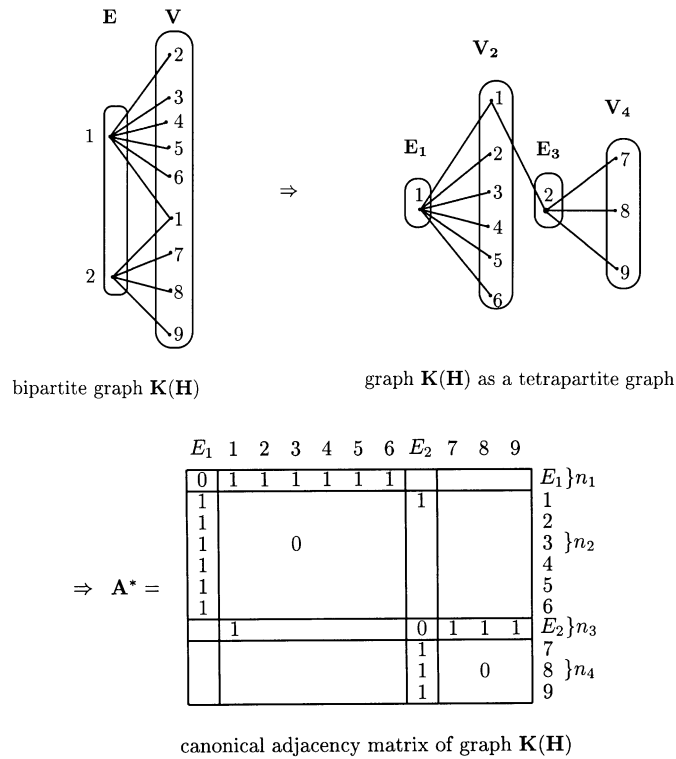


Fig. 4. König's representation of hypergraph \mathbf{H} from Fig. 3 and its canonical adjacency matrix.

Theorem 2. *The canonical adjacency matrix of König's representation of hypergraph \mathbf{H} has a symmetrical block-diagonal form:*

$$\mathbf{A}^*(\mathbf{K}(\mathbf{H})) = \begin{array}{c} \begin{array}{cccccc} n_1 & n_2 & n_3 & & n_{t-1} & n_t \\ \begin{array}{c} 0 \\ A_{21} \end{array} & A_{12} & & \cdots & & \\ A_{21} & 0 & A_{23} & \cdots & & \\ & A_{32} & 0 & \cdots & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & \cdots & 0 & A_{t-1t} \\ & & & \cdots & A_{tt-1} & 0 \end{array} & \begin{array}{c} n_1 \\ n_2 \\ n_3 \\ \\ n_{t-1} \\ n_t \end{array} \end{array}$$

where rows (columns) of non-zero matrices $A_{n_i n_{i+1}}$ ($A_{n_{i+1} n_i}$), $i = 1, \dots, t-1$, are defined only by vertices from V -part, or only by vertices from E -part, and if $n_i = |V_i|$, where V_i is the subset of the set of vertices, $V_i \subset V$, then $n_{i+1} = |E_{i+1}|$, where E_{i+1} is the subset of the set of hyperedges, $E_{i+1} \subset E$, and vice versa.

Proof. Let the canonical numbering of the vertices of graph $\mathbf{K}(\mathbf{H})$ be known and the first numbers in the canonical numbering belong to vertices from the set V_1 , belonging

to V -part. (Note that it is just an assumption only. Actually, the canonical numbering of $K(H)$ can start as with the V -part and with the E -part). According to the definition, the two vertices in graph $K(H)$ are adjacent if and only if they belong to different parts. Hence the vertices of the E -part constitute the set of vertices E_2 adjacent to the vertices from the set V_1 and possess the numbers that follow the ones for vertices from the set V_1 in the canonical numbering. For the vertices of the set E_2 the vertices from the V -part are adjacent and so on. Thus the graph $K(H)$ in accordance with canonical numbering can be presented as a t -partite graph where just the vertices from adjacent parts are adjacent. So nonzero elements in the matrix $A^*(K(H))$ are defined only by vertices of the two adjacent parts $V_i \subset V$ and $E_{i+1} \subset E$, i.e. $A_{n_i n_{i+1}} \neq 0 (A_{n_{i+1} n_i} \neq 0)$, where $n_i = |V_i|$ and $n_{i+1} = |E_{i+1}|$. The theorem is proved. \square

The example in Fig. 4 illustrates the Theorem 2.

Let us call the i th row (i th column) of blocks in matrix $A^*(K(H))$ — i -block row (i -block column). Let all i -block rows (columns), where $i=2k+1$, $k=0, 1, \dots$, contain numbers of all vertices from the V -part (the E -part). Then in accordance with Theorem 2, all i -block rows (columns) where $i=2k$, $k=1, 2, \dots$, contain numbers of all vertices from the E -part (the V -part). To obtain a reduced $(p \times q)$ -adjacency matrix we have to eliminate all even-block (odd-block) rows and all odd-block (even-block) columns from $A^*(K(H))$. The resulting matrix $R(A^*(K(H)))$ will have one of the following forms:

$$R(A^*(K(H))) = \begin{array}{c|cccc|c} & n_2 & n_4 & n_6 & n_{t-3} & n_{t-1} & \\ \hline A_{12} & 0 & 0 & \cdots & & & n_1 \\ A_{32} & A_{34} & 0 & \cdots & & & n_3 \\ & A_{54} & A_{56} & \cdots & & & n_5 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & \cdots & A_{t-2t-3} & A_{t-2t-1} & n_{t-2} \\ & & & \cdots & 0 & A_{tt-1} & n_t \end{array}$$

$$R(A^*(K(H))) = \begin{array}{c|ccccc|c} & n_1 & n_3 & n_5 & n_{t-2} & n_t & \\ \hline A_{21} & A_{23} & 0 & \cdots & & & n_2 \\ 0 & A_{43} & A_{45} & \cdots & & & n_4 \\ & 0 & A_{65} & \cdots & & & n_6 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & \cdots & A_{t-3t-2} & 0 & n_{t-3} \\ & & & \cdots & A_{t-1t-2} & A_{tt-1} & n_{t-1} \end{array}$$

Theorem 3. $R(A^*(K(H))) = R^*(K(H))$.

Proof. According to the definition, the canonical reduced adjacency matrix of a bipartite graph $K(H)$ is the maximal of all possible $(p! \cdot q!)$ matrices. Let us show that the matrix $R(A^*(K(H)))$ meets the condition of maximum. This follows from the properties of canonical matrix $A^*(K(H))$ which is maximal and from the method

of construction of matrix $R(A^*(K(H)))$. Suppose that $R(A^*(K(H)))$ is not maximal, i.e. there exists a reduced $(p \times q)$ -matrix $R(A'(K(H)))$ which has a lexicographically greater row in the first pair of different rows. The way of construction of the reduced adjacency matrix implies that the corresponding adjacency matrix $A'(K(H))$ should also have a lexicographically greater row than $A^*(K(H))$, which contradicts the fact that $A^*(K(H))$ is the maximal matrix. Thus, $R(A^*(K(H)))$ is the maximal and $R(A^*(K(H))) = R^*(K(H))$. The theorem is proved.

Let us note that Fig. 3 presents the canonical incidence matrix of hypergraph H that coincides with the canonical reduced adjacency matrix $R^*(K(H))$ obtained from the canonical adjacency matrix $A^*(K(H))$ of graph $K(H)$ (Fig. 4) by the method discussed above.

Thus, the search algorithm for the canonical incidence matrix of a hypergraph includes the following steps:

- (1) construction of König's graph $K(H)$ for hypergraph H ;
- (2) search of canonical adjacency matrix for graph $K(H)$;
- (3) construction of reduced canonical adjacency matrix for graph $K(H)$.

The evaluation of complexity of the algorithm is determined by the complexity of the algorithm for the search for canonical adjacency matrix of graph. Note that the canonization of bipartite graphs with parts N and M on the base of their reduced adjacency matrices is more effective in comparison with canonization based on adjacency matrices because the first approach at worst case requires $(N! \cdot M!)$ operations while the second one requires $(N + M)!$ operations. The effective algorithms for canonization of adjacency matrices have been realized in practice [1,17]. Algorithms for canonization of incidence matrices [5,36] that are known to the authors are enumerative and the estimations of complexity for the given algorithms are either not presented [5], or not allow to estimate the running effectivity of an algorithm in practice [36].

4. The structural analysis of molecular hypergraphs

This kind of analysis allows to investigate molecular hypergraphs through their common subgraphs. Let us consider the König's representation of hypergraphs when the subgraph isomorphism problem is solved. All graph algorithms may be used in this case. So we discuss the problem of finding common subgraphs of two hypergraphs in the terms of ordinary graph. In the chemical compounds databases, the problem arises as one of testing whether a new compound is a known compound (graph isomorphism) or subcompound (subgraph isomorphism) or contains known subcompounds (common subgraphs) [18,40,13].

Let $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$ be hypergraphs and $K(H_1)$ order $(p_1 + q_1)$ and $K(H_2)$ order $(p_2 + q_2)$ be their König's representation. Denote $K(H_1)$ as G_1 and $K(H_2)$ as G_2 .

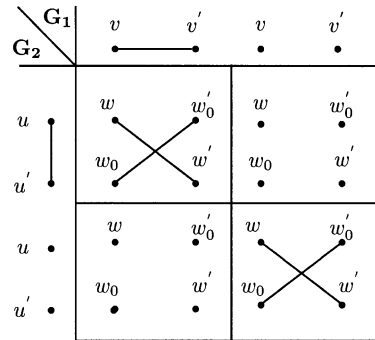


Fig. 5. The table representation of a correspondence graph.

The subgraph isomorphism problem is as follows. Given two graphs $G_1, m = |V(G_1)|$ and $G_2, n = |V(G_2)|, m \leq n$, find all the subgraphs of G_2 which are isomorphic G_1 . A subgraph $S = (V(S), E(S))$ of $G = (V, E)$ is a graph S such that $V(S) \in V$ and $E(S) = E \cap (V(S) \times V(S))$.

Given two graphs G_1 and G_2 , if there exists a subgraph of order k S^1 of G_1 isomorphic to a subgraph S^2 of G_2 , the pair of isomorphic subgraphs (S^1, S^2) is called a common subgraph of order k of G_1 and G_2 . A common subgraph (S^1, S^2) is maximal if there is no common subgraph (S_i^1, S_j^2) such that S^1 is a subgraph of S_i^1 and S^2 is a subgraph of S_j^2 .

There are two approaches to solve the subgraph isomorphism problem. The first approach is based on the following two-step procedure:

- (1) find sets of subgraphs of order k of G_1 and G_2 .
- (2) test all pairs subgraphs for isomorphism.

Step (2) can be realized, for example, by a trivial algorithm known as vertex re-ordering algorithm [17], which is based on vertex permutations and adjacency matrix comparisons. If this algorithm is used, finding common subgraphs of order k requires $m!n!/k!(m-k)!(n-k)!$ matrix comparisons. Returning to hypergraphs this algorithm requires $(p_1 + q_1)!(p_2 + q_2)!/k!(p_1 + q_1 - k)!(p_2 + q_2 - k)!$, where $p_1(p_2)$ and $q_1(q_2)$ is the number of vertices and edges in hypergraph $H_1(H_2)$.

The second approach is based on the properties of a correspondence graph $L(G_1, G_2)$ of given graphs $G_1 = (V, E)$ and $G_2 = (U, X)$ [33,15,6]. This graph $L = (W, Y)$ has a set of vertices $W = \{w = (v, u), v \in V, u \in U\}$ and a set of edges $Y = \{(w, w'), \text{ such that } \{(v, v') \in E, (u, u') \in X\} \text{ or } \{(v, v') \notin E, (u, u') \notin X\}\}$ (see Fig. 5).

The main property of the correspondence graph is that every clique of graph $L(G_1, G_2)$ corresponds to some common subgraph of G_1 and G_2 [33]. So to find the common subgraph of G_1 and G_2 it is enough

- (1) to have the correspondence graph $L(G_1, G_2)$ and
- (2) to derive all cliques in $L(G_1, G_2)$.

The complexity of the subgraph isomorphism problem is equivalent the complexity of the clique finding problem [22]. To find all cliques, the graph recursive analysis technique is used [19,14,9]. The graph recursive analysis is defined as the following succession of procedures. The set of subgraphs is derived for given graph by the some rule. Then, for every subgraph its set of subgraphs is derived by the same rule and so on. The algorithm is completed when the some rule of stop is fulfilled.

The theoretical estimate of complexity for the cliques finding algorithm is established as $O(l^2 \cdot 3^{l/3})$, where l is the order of a correspondence graph [9]. This estimate is achieved for the k -partite graphs $K_{3,3,\dots,3}$ that are known as the clique extremal graphs [34]. Returning to a hypergraph, this algorithm requires $((p_1 + q_1 + p_2 + q_2)^2 \times 3((p_1 + q_1 + p_2 + q_2)/3))$. In practice, algorithms of the first [17] and second [14,9] approaches are realized.

5. Invariants of graphs and hypergraphs

To identify molecular structures of chemical compound, the molecular graph and hypergraph invariants, called topological indices [47,10,4], could be used too. Topological indices are designed basically by transforming a molecular graph into a number. The difference between graph and hypergraph representations of molecular structures affects the values of topological indices. The results of comparative analysis of some topological indices based on distances within a graph are listed below. The *distance* $d(u, v)$ between vertices u, v , in graph G (hypergraph H), is the length of the shortest walk (the path) that connects vertices u and v . The length of any hyperedge in H equals to one, i.e. all the vertices that are incident to the same hyperedge are elongated from each other at the distance one.

Let $G = (V, E_G)$ and $H = (V, E_H)$ be the graph and hypergraph representations of a molecular structure F with delocalized polycentric bonds, where $V(G) = V(H)$. Then for $u, v \in V(G) = V(H)$ the following theorem is true.

Theorem 4. $d_G(u, v) \geq d_H(u, v)$.

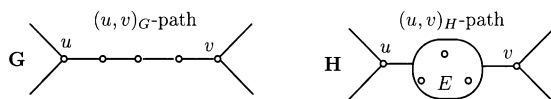
Proof. Let us consider several different cases.

Case 1: The shortest (u, v) -walk in graph G and hypergraph H consists only edges of degree two. In this case, paths have the same length as G and H , describe the same molecular structure F on the same set of vertices and as (u, v) -walks are the shortest. Thus, we have

$$d_G(u, v) = d_H(u, v).$$

Case 2: The shortest $(u, v)_G$ -walk in graph G consists only edges of degree two, and the shortest $(u, v)_H$ -walk in hypergraph H contains at least one edge E of degree more than two.

Suppose that the same set of vertices belongs to both paths:



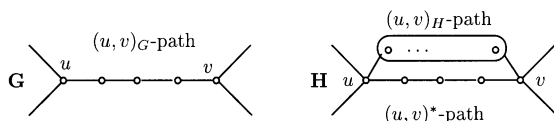
As $\deg E > 2$, the distance between all vertices belonging to the same edge is equal to one and

$$d_H(u, v) = d_G(u, v) - \deg E + 2,$$

i.e.

$$d_H(u, v) < d_G(u, v).$$

Now suppose that (u, v) -paths contain different sets of vertices:



Let $(u, v)^*$ be the path in hypergraph H , containing the same vertices as (u, v) -path in graph G . Then due to the cases considered above, either

$$d_G(u, v) = d_H(u, v)^* \quad (1)$$

or

$$d_G(u, v) > d_H(u, v)^*. \quad (2)$$

But as $(u, v)_H$ -path is the shortest one in H , the following inequality holds:

$$d_H(u, v) < d_H(u, v)^*. \quad (3)$$

Then from Eqs. (1)–(3) it follows that

$$d_H(u, v) < d_G(u, v).$$

The theorem is proved. \square

Theorem 4 implies that for any F represented by G and H the following inequalities hold: $W(G) > W(H)$, $\text{rad}(G) > \text{rad}(H)$, $\text{diam}(G) > \text{diam}(H)$, where W is the first mathematical invariant reflecting the topological structure of a molecular graph, called the Wiener index [48], rad is the radius and diam is the diameter [25]. Formulas for indices are presented in the appendix.

A few other topological indices are likely to reveal similar regularities. The behavior on the whole set of graphs and hypergraphs cannot be estimated for all topological indices.

In this section we present the results of the comparative analysis of graph and hypergraph invariants for 8 series of molecular structures of type $R - X$, where R is the

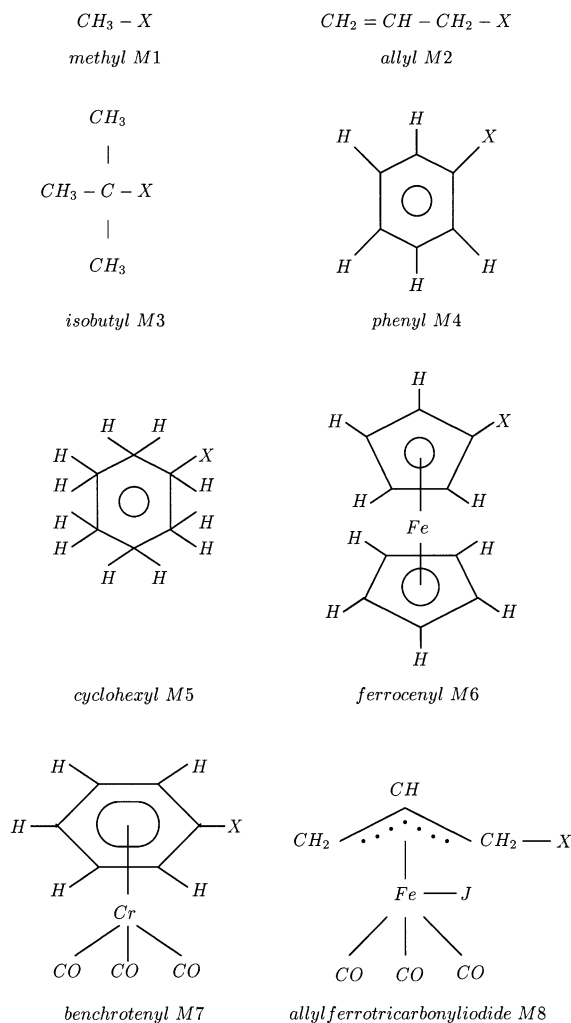


Fig. 6. The widespread structural types.

widespread structural types presented in Fig. 6 and X is the widespread substituents presented in Table 1. We investigate the same invariants for the same molecular structures represented by graphs and hypergraphs. All graph invariants are generalized for hypergraphs [29,31].

Three groups of invariants presented in the appendix are considered: metric, topological and information indices. Metric indices of graphs have been studied in many papers of graph theory. A bibliography on the subjects is given in [44,37]. Topological indices are well known in mathematical chemistry and are used for representing the molecular structure and for finding the quantitative structure-property and structure-activity relationships [47]. Information indices [10,12] are constructed for various matrices (layer

Table 1
The widespread substituent X

<i>N</i>	X	<i>N</i>	X
1	R–H	16	R–COOC ₂ H ₅
2	R–CH ₃	17	R–COCH ₃
3	R–C ₂ H ₅	18	R–CH ₂ OH
4	R–(CH ₂) ₂ CH ₃	19	R–N(CH ₃) ₂
5	R–CH(CH ₃) ₂	20	R–CH(OH)CH ₃
6	R–(CH ₂) ₃ CH ₃	21	R–CH(OH)C(CH ₃) ₃
7	R–C(CH ₃) ₃	22	R–CH ₂ C ₆ H ₅
8	R–CN	23	R–COC ₆ H ₅
9	R–CH=CH ₂	24	R–C ₆ H ₁₁
10	R–COOH	25	R–C(CH ₃)=CH ₂
11	R–COH	26	R–CH=CHCH ₃
12	R–COOCH ₃	27	R–CH=CHCOC ₆ H ₅
13	R–OCOCH ₃	28	R–COOC ₆ H ₅
14	R–COH	29	R–OCOC ₆ H ₅
15	R–CH=CHC ₆ H ₅	30	R–(CH ₂) ₄ COOH

matrix, distance matrix, adjacency matrix) and also for some topological indices such as the Wiener index.

The sensitivity of indices for both models are investigated and the noncorrelating indices are revealed.

The sensitivity of topological index *I* is the measure of its ability to distinguish the nonisomorphic graphs by the given topological index *I*. The theoretical evaluation of sensitivity *I* on the set of all graphs is too difficult, therefore, the evaluation *S* of sensitivity *I* on the fixed set *M* of nonisomorphic graphs is used [11,41] by formula

$$S = (N - N_I)/N,$$

where *N_I* is the number of degeneracies *I* on the set *M*, *M* = *N*. The sensitivity of all 23 indices from the appendix for both models are calculated. The results have shown that for the series of considered molecular structures the indices sensitivity for the hypergraph model are two times higher than those for the graph model.

The indices correlations for the models considered are also investigated. Let us define the noncorrelating indices as those with the correlation *r* ranging from 0 to 0.9. All noncorrelating indices were found for the graph and hypergraph models. Solid lines presented in Fig. 7 correspond to the percentage of noncorrelating indices for different models. Fig. 7 shows that the number of noncorrelating indices for the hypergraph model is greater than that of the graph one, for all molecular structures series. For example, on the set M8, this number exceeds the same parameter for the graph model by 1.5 times as large.

So the comparative analysis of indices for both models of eight series of nonclassical molecular structures shows that the hypergraph model gives a higher accuracy of

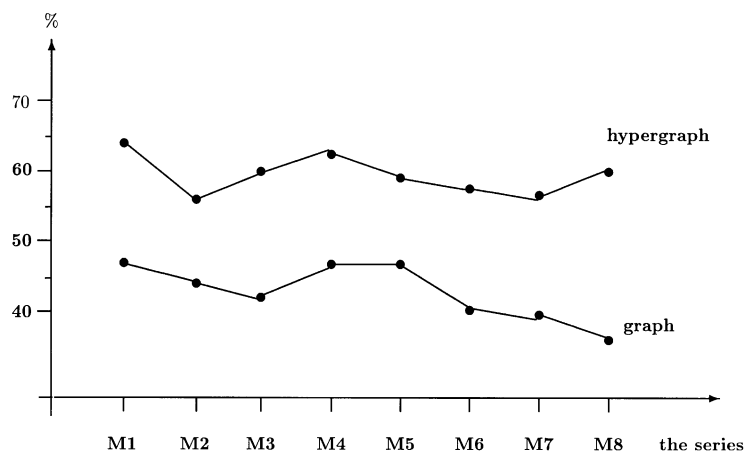


Fig. 7. The percentage of noncorrelating indices for different models.

molecular structure description: the higher is the accuracy of the model, the greater is the diversity of the behavior of its invariants.

6. Conclusion

By now many chemists have realized the usefulness of graph theory for chemistry. In the present work, it is proposed to use the hypergraph theory in chemistry. In fact, it is tempting to use hypergraph (a graph, one edge of which at least is incident to more than two vertices) for description of a molecule in which more than two atoms are bonded by one and the same bond, such as the one in an allylic complex. Such cases cause difficulties in traditional chemical graph theory. The defects peculiar for ordinary molecular graphs are absent in molecular hypergraphs and moreover the hypergraph representation facilitates the task of comparing the ordinary molecular structures with the structures containing polycentric bonds.

The concept of a molecular hypergraph was introduced and the ways of its implementation were shown for organometallic (sandwich-type) compounds having polycentric bonds. As was mentioned above, bonds of the same kind occur also in polycyclic conjugated molecules. The new approach of hypergraph theory application to the Clar aromatic sextet theory was considered in [23]. It was shown that the Clar hypergraph has properties quite different from the traditional molecular graph.

The use of hypergraphs for representation of the molecular structure of nonclassical compounds and polycyclic conjugated molecules reveals a new area of research where the results of chemical application of graph theory are generalized and also allows to expand the range of hypergraph theory application.

Appendix

A.1. Metric indices

The first five invariants are based on the vertex eccentricity $e(v) = \max_{u \in V} d(u, v)$ [25].

1. The radius [25]

$$\text{rad} = \min_{v \in V} e(v).$$

2. The diameter [25]

$$\text{diam} = \max_{v \in V} e(v).$$

3. The average vertex eccentricity [44]

$$e_{\text{av.}} = \frac{1}{p} e,$$

where p is a number of graph vertices.

4. The eccentricity [44]

$$e = \sum_{v \in V} e(v).$$

5. The eccentric [44]

$$\Delta G = \frac{1}{p} \sum_{v \in V} |e(v) - e_{\text{av.}}|.$$

Next seven invariants are based on the vertex distance $D(v) = \sum_{u \in V} d(u, v)$ and the graph distance $D = \frac{1}{2} \sum_{v \in V} D(v)$ [24,43].

6. The average vertex distance [44]

$$D_{\text{av.}} = \frac{2D}{p}.$$

7. The unipolarity [44]

$$D^* = \min_{v \in V} D(v).$$

8. The centralization [44]

$$\Delta^* = 2D - pD^*.$$

9. The mean distance deviation [44]

$$\Delta D = \frac{1}{p} \sum_{v \in V} |D(v) - D_{\text{av.}}|.$$

10. The dispersion [35]

$$m_2^* = \min_{v \in V} \left(\frac{1}{p} \sum_{u \in V} [d(u, v)]^2 \right).$$

11. The compactness [20]

$$\mu = \frac{4}{p(p-1)}D.$$

12. The variation [21]

$$\text{var} = \max_{v \in V} (D(v) - D^*).$$

A.2. Topological indices

13. The Wiener index [48]

$$W = \frac{1}{2} \sum_{u,v \in V} d(u,v).$$

14. The polarity number [49]

$$P = \frac{1}{2} \sum_{i=1}^p d_{3,i},$$

where $d_{3,i}$ is a distance of length three.

15. The mean square distance index [3]

$$D^2 = \left(\frac{\sum_{i=1}^{\max} g_i i^2}{\sum_{i=1}^{\max} g_i} \right)^{1/2},$$

where g_i is a number of vertex pairs at a distance i from each other.

16. The distance index [42]

$$\text{GDI} = \sum_{i=1}^{\max} g_i^2.$$

17. The total adjacency index [42]

$$A = \sum_{i,j=1}^p a_{ij},$$

where a_{ij} is the adjacency matrix element.

A.3. Information indices

18. The Information Wiener Index [12]

$$I_D^W = \frac{1}{2} \sum_{i=1}^{\max} \frac{g_i i}{W} \log_2 \frac{g_i i}{W}.$$

19. The autometricity index [45]

$$H_1 = - \sum_{i=1}^N \frac{p_i}{p} \log_2 \frac{p_i}{p},$$

where p_i is the number of vertices in the i th ($i = 1, \dots, N$) class of autometricity and $p = \sum_{i=1}^N p_i$ is the total number of vertices in graph. Class autometricity is the set of vertices with the same vector $\lambda_{i1}, \dots, \lambda_{id}$, where λ_{ij} is the number vertex at a distance j from vertex i .

20,21,22. The information integral indices H_λ , H_d , H_A are defined on the basis of the information vertex indices $H_\lambda(i)$, $H_d(i)$, $H_A(i)$ [28]

$$H_\lambda(i) = - \sum_{j=1}^{e(i)} \frac{\lambda_{ij}}{p} \log_2 \frac{\lambda_{ij}}{p},$$

$$H_d(i) = - \sum_{j=1}^p \frac{d(i,j)}{D(i)} \log_2 \frac{d(i,j)}{D(i)},$$

$$H_A(i) = - \sum_{j=1}^{\deg(i)} \frac{a_{ij}}{\sum_{j=1}^{\deg(i)} a_{ij}} \log_2 \frac{a_{ij}}{\sum_{j=1}^{\deg(i)} a_{ij}},$$

where $\deg(i)$ is the degree of vertex i .

Then, the integral information indices take the forms:

$$H_\lambda = \sum_{i=1}^p H_\lambda(i),$$

$$H_d = \sum_{i=1}^p H_d(i),$$

$$H_A = \sum_{i=1}^p H_A(i).$$

23. The information index of average vertex distance [45]

$$H_{av.} = \log_2 D_{av.}$$

References

- [1] V.A. Arlasarov, I.I. Zuev, A.V. Uskov, I.A. Faradzhev, Algorithm for transformation of finite nonoriented graphs to canonical form, *Zn. Vychisl. Mat. Mat. Fis.* 14 (1974) 737–743 (in Russian).
- [2] E.V. Babaev, Principles of Symmetry and Systemic Approach in Chemistry, Nauka, Moscow, 1987, pp. 30–54 (in Russian).
- [3] A.T. Balaban, Topological indices based on topological distances in molecular graphs, *Pure Appl. Chem.* 55 (1983) 199–206.
- [4] A.T. Balaban, Applications of graph theory in chemistry, *J. Chem. Inform. Comput. Sci.* 25 (1985) 334–343.
- [5] A.M. Baraev, I.A. Faradzhev, in: I.A. Faradzhev (Ed.), Algorithmic Studies in Combinatorics, Nauka, Moscow, 1978 (in Russian).
- [6] H.G. Barrow, R.M. Burstall, Subgraph isomorphism, matching relational structures and maximal cliques, *Inform. Process. Lett.* 4 (4) (1976) 83–84.

- [7] C. Berge, *Graphes et Hypergraphes*, Dunod, Paris, 1970, 502p.
- [8] C. Berge, *Isomorphism Problems for Hypergraphs*, Lecture Notes in Mathematics, Vol. 411, Springer, Berlin, 1974, pp. 1–12.
- [9] Yu. E. Bessonov, V.A. Skorobogatov, On graph recursive analysis, *Vychisl. Systemy* 85 (1981) 3–20 (in Russian).
- [10] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press: Chichester, 1983, 249pp.
- [11] D. Bonchev, O. Mekenyan, N. Trinajstić, Isomer discrimination by topological information approach, *J. Comput. Chem.* 2 (1981) 127–148.
- [12] D. Bonchev, N. Trinajstić, Information theory, distance matrix and molecular branching, *J. Chem. Phys.* 67 (1977) 4517–4533.
- [13] D. Bonchev, N. Trinajstić, Chemical information theory: structural aspects, *Internat. J. Quantum. Chem. Symp.* 18 (1982) 463–480.
- [14] C. Bron, J. Kerbosh, Finding all cliques of an undirected graph, *Comm. Assoc. Comput. Mach.* 16 (9) (1973) 575–577.
- [15] D.E. Chahraman, A.K.C. Wong, T. Au, Graph optimal monomorphism algorithms, *IEEE Trans. Systems Man Cybernet.* 10 (4) (1980) 181–197.
- [16] E. Clar, *The Aromatic Sextet*, Wiley, London, 1972, 120p.
- [17] D.G. Corneil, C.C. Gotlieb, An efficient algorithm for graph isomorphism, *J. Assoc. Comput. Mach.* 17 (1) (1970) 51–64.
- [18] R.D. Cramer, G. Redl, C.E. Berkoff, Substructural analysis, A novel approach to the problem of drug design, *J. Med. Chem.* 17 (5) (1974) 533–535.
- [19] S.R. Das, C.L. Sheng, Z. Chen, An algorithm for finding all maximal complete subgraphs and an estimate of the order of computational complexity, *Comput. Electron. Eng.* 5 (4) (1978) 365–368.
- [20] J.K. Doyle, J.E. Graver, Mean distance in graph, *Discrete Math.* 17 (1977) 147–154.
- [21] R.C. Entringer, D.E. Jackson, D.A. Snyder, Distance in graphs, *Czechoslovak Math. J.* (1976) Vol. 26, N(2), pp. 283–297.
- [22] M.R. Garey, D.S. Johnson, *Computers and Intractability, a Guide to the Theory of NP-Completeness*, Bell Laboratories, Murray Hill, NJ, 1979, 420p.
- [23] I. Gutman, E.V. Konstantinova, V.A. Skorobogatov, Molecular hypergraphs and clar structural formulas of benzenoid hydrocarbons, *Models in Chemistry* 136(5–6) (1999) 539–548.
- [24] I. Haiduc, J.J. Zuckerman, *Basic Organometallic Chemistry*, de Gruyter, Berlin, 1985, 488p.
- [25] F. Harary, *Graph Theory*, Addison-Wesley, London, 1971, 294p.
- [26] B.R. Heap, The production of graphs by computer, in: R.C. Read (Ed.), *Graph Theory and Computing*, Academic Press, New York, 1972, pp. 47–62.
- [27] J.B. Hendrickson, A.G. Toczek, Unique numbering and cataloguing of molecular structures, *J. Chem. Inform. Comput. Sci.* 23 (1983) 171–177.
- [28] E.V. Konstantinova, A.A. Paleev, Sensitivity of topological indices of polycyclic graphs, *Vychisl. Sistemy* 136 (1990) 38–48 (in Russian).
- [29] E.V. Konstantinova, V.A. Skorobogatov, Molecular hypergraphs: the new representation of nonclassical molecular structures with polycentric delocalized bonds, *J. Chem. Inform. Comput. Sci.* 35 (3) (1995) 472–478.
- [30] E.V. Konstantinova, V.A. Skorobogatov, Molecular structures of organoelement compounds and their representation by the labeled molecular hypergraphs, *Zh. Strukt. Khim.* 39 (2) (1998) 328–337.
- [31] E.V. Konstantinova, V.A. Skorobogatov, Graph and hypergraph models of molecular structure: a comparative analysis of indices, *Zh. Strukt. Khim.* 39 (6) (1998) 958–966.
- [32] V. Kvasnicka, J. Pospichal, Canonical indexing and constructive enumeration of molecular graphs, *J. Chem. Inform. Comput. Sci.* 30 (1990) 99–105.
- [33] G. Levi, A note of the derivation of maximal common subgraphs of two directed or undirected graphs, *Calcolo* 5 (9) (1972) 341–352.
- [34] J. Moon, L. Moser, On cliques in graphs, *Israel J. Math.* 3 (1) (1965) 23–28.
- [35] O. Ore, *Theory of Graphs*, American Mathematical Society, Colloquim Publications, Providence, RI, 1962, Vol. 38, 330p.
- [36] M.L. Overton, A. Proskurowski, Canonical incidence matrices of graphs, *BIT* 19 (1979) 271–273.
- [37] J. Plesnik, On the sum of all distances in a graph or digraph, *J. Graph Theory* 8 (1984) 1–21.

- [38] R. Rado, Isomorphisms between hypergraphs, New Direct, Theory Graphs, Proc. Third Ann. Arbor Conf., Univ. Michigan 1971, 207–237 (1973).
- [39] M. Randić, On the recognition of identical graphs representing molecular topology, *J. Chem. Phys.* 60 (1974) 3920–3928.
- [40] M. Randić, On canonical numbering of atoms in a molecule and graph isomorphism, *J. Chem. Inform. Comput. Sci.* 17 (3) (1977) 171–180.
- [41] M. Razinger, J.R. Chretien, J.K. Dubois, Structural selectivity of topological indices in alkane series, *J. Chem. Inform. Comput. Sci.* 25 (1985) 23–27.
- [42] D.H. Rouvray, Should we have designs on topological indices?, in: R.B. King (Ed.), *Chemical Applications of Topology and Graph Theory*, Studies in Physical and Theoretical Chemistry, Vol. 28, Elsevier, Amsterdam, 1983, pp. 159–177.
- [43] G.B. Shulpin, *The World of Unusual Molecules*, Mir, Moscow, 1986, 174pp (in Russian).
- [44] V.A. Skorobogatov, A.A. Dobrynin, Metric analysis of graphs, *Math. Chem. (MATCH)* 23 (1988) 105–151.
- [45] V.A. Skorobogatov, E.V. Konstantinova, Yu.S. Nekrasov, Yu.N. Sukharev, E.E. Tepfer, On the correlation between the molecular information topological and mass-spectra indices of organometallic compounds, *Math. Chem. (MATCH)* 26 (1991) 215–228.
- [46] I.V. Stankevitch, in: N.S. Zefirov, C.I. Kuchanov (Eds.), *Application of Graphs in Chemistry*, Nauka, Novosibirsk, 1988, pp. 7–69 (in Russian).
- [47] N. Trinajstić, *Chemical Graph Theory*, 2nd revised edition, CRC Press, Boca Raton, FL, 1992, 147p.
- [48] H. Wiener, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.* 69 (1947) 17–20.
- [49] H. Wiener, Vapor pressure — temperature relationships among the branched paraffin hydrocarbons, *J. Chem. Phys.* 52 (1948) 425–430.
- [50] A.A. Zykov, Hypergraphs, *Usp. Math. Nauk* 29(6) (1974) 89–154 (in Russian).