

# CHEMICAL HYPERGRAPH THEORY

Elena Konstantinova

*Sobolev Institute of Mathematics*

*Russian Academy of Sciences*

*Novosibirsk 630090, Russia*

*e.konsta@math.nsc.ru*

This work is partially supported by Com<sup>2</sup>MaC-KOSEF, Korea.



# Preface

Graph representation of molecular structures is widely used in computational chemistry and theoretical chemical researches [2, 3, 11, 31, 43, 44, 62, 63, 68, 69]. Trinajstić noted [69] that the roots of chemical graph theory may be found in the works by chemists of 18-19 centuries such as Higgins, Kopp, Crum Brown. They have used first chemical graphs for representing molecules. Molecular structures are represented by graphs where vertices correspond to atoms, and edges to chemical bonds. This kind of a graph, called now a molecular graph, is the object of study in the theory of ordinary graphs [36]. 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 polycentric bonds [5]. Most often polycentric bonds are encountered in organometallic (sandwich-type) compounds, polycyclic conjugated molecules, benzenoid systems in particular, etc.

In 1994 the concept of a molecular hypergraph, as a mathematical tool suitable for modeling the structure of molecules with polycentric bonds was introduced [46]. Main goal of this Lecture Notes is to demonstrate advantages of an application of hypergraph theory in chemical researches.

The present Lecture Notes is organized in the following way. It consists of 6 chapters. The first chapter reports elements of organic chemistry that are needed in the rest of the Lecture Notes. The second chapter contains brief history of graph representations of molecular structures and declares problems arising from presence of polycentric bonds. In the third chapter the definition of a hypergraph is given and its ability for representing molecular structures is shown. The fourth chapter deals with the problem of hypergraph isomorphism. An

algorithm for constructing of canonical incidence matrices using König's representation of hypergraphs is presented. The fifth chapter is devoted to the problem of subgraph isomorphism. In the sixth chapter the graph and hypergraph invariants which can be use to identify molecular structures of chemical compounds are investigated.

# Contents

<b>Preface</b>	<b>iii</b>
<b>1 Elements of organic chemistry</b>	<b>1</b>
1.1 Organometallic compounds . . . . .	1
1.2 Benzenoid hydrocarbons . . . . .	3
<b>2 Graph representation of molecular structures with polycentric bonds</b>	<b>5</b>
<b>3 Hypergraphs of molecular structure</b>	<b>9</b>
3.1 Main definitions . . . . .	9
3.2 Molecular hypergraphs . . . . .	12
3.3 Examples of graph and hypergraph representation . . . . .	18
<b>4 Identification of molecular hypergraph</b>	<b>21</b>
4.1 Canonical incidence matrix of hypergraph . . . . .	22
4.2 Algorithm for canonical incidence matrix construction . . . . .	23
<b>5 The structural analysis of molecular hypergraphs</b>	<b>29</b>
<b>6 Invariants of graphs and hypergraphs</b>	<b>33</b>
6.1 Comparative analysis of invariants for organometallic compounds . . . . .	37
6.2 Comparative analysis of invariants for benzenoid hydrocarbons . . . . .	40
<b>A Metric indices</b>	<b>47</b>

<b>B Topological indices</b>	<b>49</b>
<b>C Information indices</b>	<b>51</b>
<b>Bibliography</b>	<b>52</b>

# Chapter 1

## Elements of organic chemistry

In the majority of organic molecules covalent chemical bonds exist between distinct pairs of atoms. The bonding in such a molecule is then described by classical structural formula. There are, however, quite a few examples of molecules in which several (more than two) atoms are covalently bound by means of a chemical bond, so-called *polycentric bond*. Organometallic compounds and benzenoid hydrocarbons are examples of such kind of molecules.

### 1.1 Organometallic compounds

Organometallic chemistry deals with compounds containing at least one metal-carbon ( $M-C$ ) bond [19, 37, 38, 70]. The nature and stability of this bond determine the properties of organometallic compounds. The traditional types of bonds - ionic or covalent - can only in part explain the bonding between the metal atoms and the attached organic groups or molecules called *ligands*. Also, there are types of bonding found exclusively in organometallic compounds.

One of these bonds is  $\pi$ -bond. It is formed by electron transfer from ligand to metal. Depending on their nature, such ligands can contribute a variable number of  $\pi$ -electrons: two (ethylene), three (the  $\pi$ -allyl group), four (cyclobutadiene), five (the cyclopentadienyl group), six (benzene), etc. (see Figure 1). It means that  $\pi$ -bond exist between

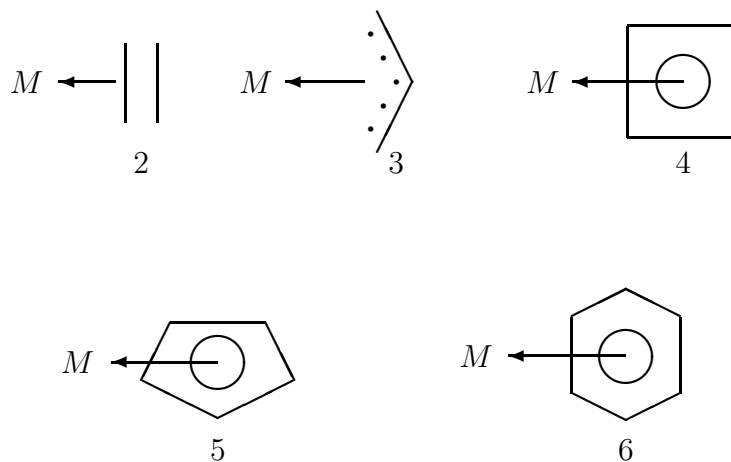


Figure 1. Ligand types (donors of 2-6 electrons)

distinct more than two atoms.

The different metal-carbon bond types and ligand types give a wide structural variety of organometallic compounds. In principle, any unsaturated or aromatic organic molecule or radical can act as a  $\pi$ -ligand. It is useful to classify the organometallic compounds according with the nature of the ligand, or, more exactly, according to the number of electrons contributed by the ligand to attaining the noble gas configuration to the central atom. In the  $\pi$ -complexes known this number varies from two to eight. More important of them are classified as follows:

- a) two-electron ligands: isolated double bonds;
- b) three-electron ligands: allyl and cyclopropenyl groups. Figure 2 presents the structural types of compounds with three-electron ligands;
- c) four-electron ligands: cyclobutadiene, butadiene, cyclopentadiene, or other molecules containing a butadiene fragment. Some typical four-electron ligands and their attachment to the metal atom are shown in Figure 3;
- d) five-electron ligands: cyclopentadienyl and other aromatic molecules;
- e) six-electron ligands: benzene and other aromatic molecules;

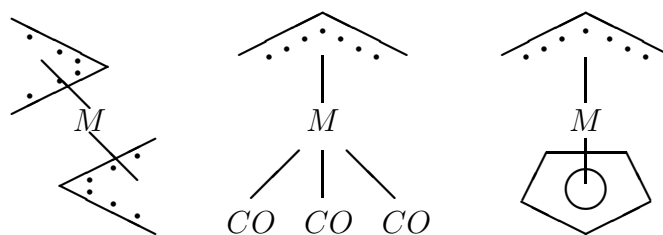


Figure 2. Structural types of compounds with three-electron ligands

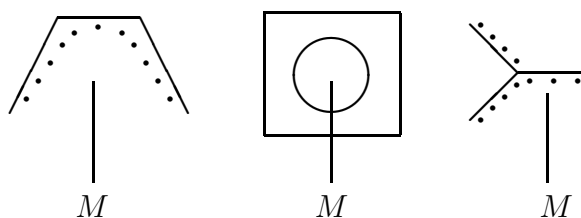


Figure 3. Four-electron ligands and their attachment to the metal atom

## 1.2 Benzenoid hydrocarbons

It is one of the standard results of theoretical organic chemistry that benzenoid hydrocarbons (and polycyclic conjugated molecules in general) have no unique classical structural formula. The conjugation models of their  $\pi$ -electrons must be visualized by means of several so-called *Kekulé* structures [22, 25, 32, 33], none of which correctly describes the actual situation. An attempt to overcome this century-long problem was proposed by Clar in the 1950's and eventually elaborated in due detail in his seminal book [18].

According to Clar, the electronic structure of benzenoid hydrocarbons is represented by, what nowadays is known as, Clar aromatic sextet formulas. These formulas contain circles drawn inside some of the hexagons of the respective molecule. The circles symbolize *aromatic sextets*, groups of six  $\pi$ -electrons, which - according the Clar theory - are dominantly located in the respective hexagons of the benzenoid molecule. The Clar aromatic sextet, located inside a given hexagon, can be understood as a six-centric chemical bond connecting the six atoms that form the respective hexagon. Two Clar aromatic sextet formulas of anthracene are shown in Figure 4.



Figure 4. Clar aromatic sextet formulas of anthracene

## Chapter 2

# Graph representation of molecular structures with polycentric bonds

The natural model of the respective molecular structure is *molecular graphs* [2, 3, 12, 69]. Molecular graphs are ones which represent the *constitution* of molecules. In these graphs vertices correspond to individual atoms and edges to chemical bonds between them. The molecular graph corresponding to propane is shown in Figure 5.

In order to simplify the handling of molecular graphs, *hydrogen-suppressed graphs*, i.e., graphs representing only molecular skeletons without hydrogen atoms and their bonds, are often used. The hydrogen-suppressed graphs are almost universally used in chemical graph theory, because the neglect of the hydrogen atoms and their bonds in most cases cannot be the cause of any ambiguity. The hydrogen-suppressed graph corresponding to propane also is given in Figure 5.

Let us consider now the graph representations of molecular structures with polycentric bonds. Some of the graph models used for representation of "sandwich" and olefin structures with polycentric bonds are given in [64, 67]. Sometimes disconnected molecular graphs are used for their representation (see Figure 6a)). 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 within it.

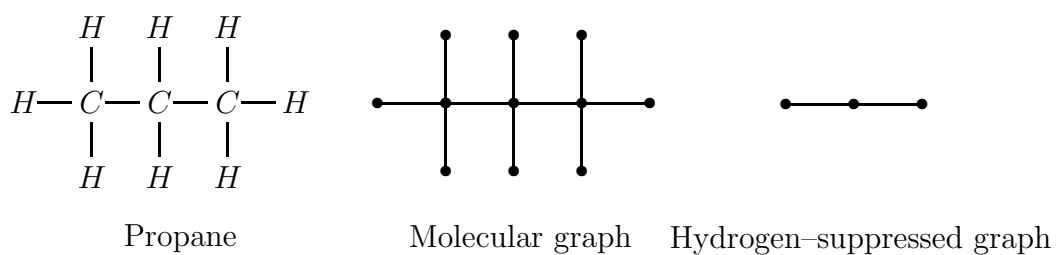


Figure 5. The molecular graphs corresponding to propane

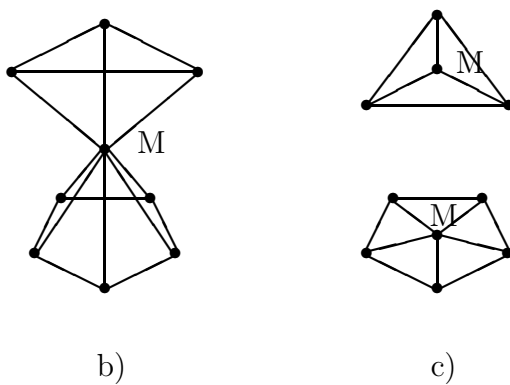
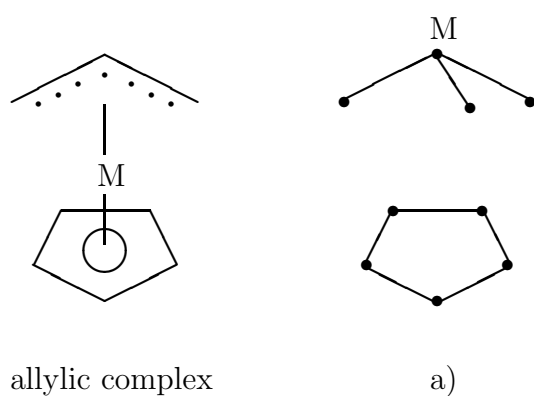


Figure 6. The hydrogen-suppressed graph representations of allylic metal-cyclopentadienyl complex

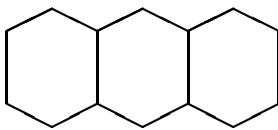


Figure 7. Graph representation of the Clar aromatic sextet formulas of anthracene

More illustrative but still not devoid of drawbacks are connected molecular graphs where all the vertices corresponding to carbon atoms are connected to the metal vertex, which corresponds to the metal atom (see Figure 6b)). The degree of the metal vertex in this case is equal to the number of vertices connected and not necessarily equal to the valency of the metal atom. For example, the degree of  $M$  vertex for graph representation in Figure 6b) is equal to eight. But the valency of the metal atom is equal to two. Besides, in both representations the difference between simple covalent and polycentric bonds is obscured. Sometimes planar graphs [8] 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 (see Figure 6c)). In this case the disconnected planar graph describes the structures. All these representations are insufficient with regard to the molecule's symmetry. They all lead to the wrong conclusion that the three carbon atoms in the  $\pi$ -allyl system are equivalent.

Benzenoid hydrocarbons can be represented by ordinary graphs too. But the different Clar aromatic sextet formulas of the same conjugated molecule are presented one and the same graph. Figure 7 shows one and the same graph representation of two Clar aromatic sextet formulas of anthracene.

All the defects of the structure representation that was mentioned above are eliminated if the hypergraphs are used to represent the structures with polycentric bonds [46, 47].



## Chapter 3

# Hypergraphs of molecular structure

Before defining a molecular hypergraph, we give several notions of hypergraph theory which are necessarily for further discussion [8, 73].

### 3.1 Main definitions

*Hypergraph*  $H = (V, E)$  consists of a non-empty set of vertices  $V = \{v_i | i = 1, \dots, p\}$  and a family  $E = \{E_j | j = 1, \dots, q\}$  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 Figure 8. 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  $degv_i$ . 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  $degE_j$ . The degrees of edges  $E_1, E_2$  and  $E_3$  presented in Figure 8 are equal to one, two and three respectively; vertex  $v_2$  is incident to edges  $E_2$  and  $E_3$  ( $degv_2 = 2$ ), 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.

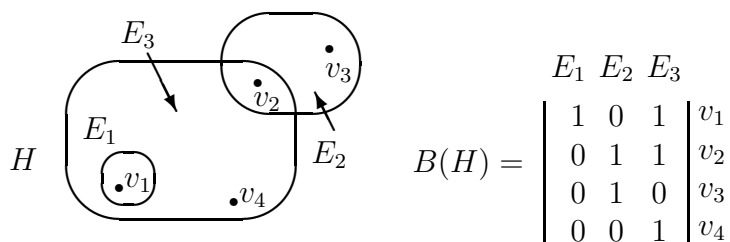
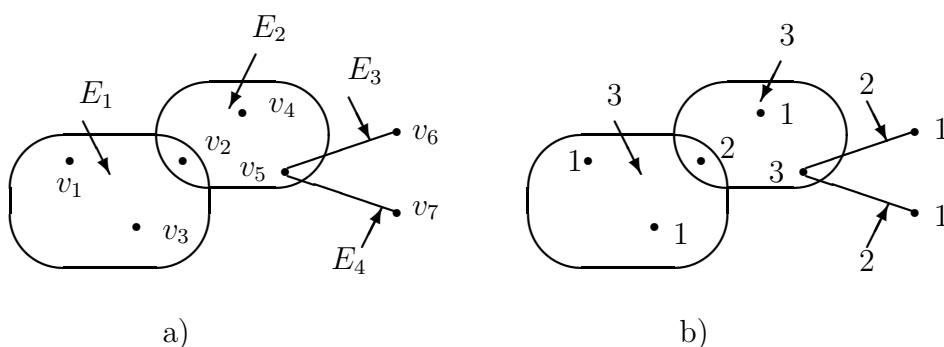
Figure 8. Hypergraph  $H$  and its incidence matrix

Figure 9. Labeled hypergraphs

A finite hypergraph (sets  $V$  and family  $E$  are finite) is unambiguously defined by an incidence matrix  $B(H) = \|b_{ij}\|$ ,  $i = 1, \dots, p, j = 1, \dots, q$ , where  $b_{ij} = 1$ , if  $v_i \in E_j$  and  $b_{ij} = 0$ , if  $v_i \notin E_j$ .

The incidence matrix of hypergraph  $H$  is also shown in Figure 8.

A hypergraph  $H = (V, E)$  is *labeled* if there exists labeling functions defined on the set of vertices  $V$  and on the set of edges  $E$ , which establish a correspondence between each vertex and each edge and some label. Let  $A(V) = \{a(v_1), \dots, a(v_p)\}$  and  $A(E) = \{a(E_1), \dots, a(E_q)\}$  be the sets of possible labels called *vertex* and *edge alphabets*. The sets  $A(V)$  and  $A(E)$  may be represented by symbolic, numerical and structural labels.

Any symbols and numbers may be elements of the vertex and edge alphabets. For example, the symbolic alphabets of the hypergraph drawn in Figure 9a) are as follows  $A(V) = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ ,

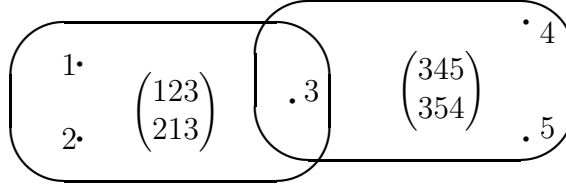


Figure 10. Structural edge labels

$A(E) = \{E_1, E_2, E_3, E_4\}$ . The degrees of vertices and edges are taken account as the corresponding labels for the same graph in Figure 9b). In this case, the vertex and edges alphabets are numerical:  $A(V) = \{1, 2, 3\}$ ,  $A(E) = \{2, 3\}$ . A hypergraph with numerical labels is also called weighted one.

Structural labels are permutations. A *permutation* of a finite set of elements  $X = \{1, \dots, n\}$  is the injective mapping of the set  $X$ . The permutation is recorded as

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix},$$

where  $i_1, i_2, \dots, i_n$  is a certain substitution of numbers such that  $\sigma(k) = i_k$ .

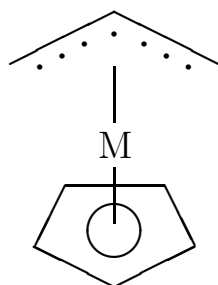
A *cyclic permutation of length  $l$*  is a permutation  $\sigma$  of a finite set  $Y = \{y_1, y_2, \dots, y_l\}$ , such that  $\sigma(y_1) = y_2, \sigma(y_2) = y_3, \dots, \sigma(y_{l-1}) = y_l, \sigma(y_l) = y_1$ . For example,  $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{pmatrix}$  is a permutation on a five-element set, and cyclic permutations are defined on the sets (1,3,4) and (2,5).

Let us give example of using permutations as hypergraph labels. Figure 10 shows a hypergraph whose set  $A(E)$  of edge labels is represented by permutations  $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$  and  $\begin{pmatrix} 3 & 4 & 5 \\ 3 & 5 & 4 \end{pmatrix}$ . For each hyperedges, permutations are defined on the set of vertices incident to the corresponding edge. The vertices of degree one make a cycle in each permutation, and the vertex 3 belonging to both hyperedges is mapped into itself. In this case, the edge labels reflect the structural peculiarities of the hypergraph.

### 3.2 Molecular hypergraphs

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 polycentric bonds and hyperedges  $E \in E(H)$  with  $\text{deg}E = 2$  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.

Molecular hypergraphs as well as ordinary molecular graphs can represent molecular structures with different levels of details. Let us consider the example of six molecular hypergraphs (Figure 11) of allylic complex shown earlier in Figure 6. It is known that the allylic ligand has an open chain of three  $sp^2$ - hybridized carbon atoms, each having a  $\pi$ -electron available for metal–ligand bond formation [37]. Figure 11 a) shows a molecular hypergraph where the polycentric bond in allyl is represented by the edge  $E_1$ ,  $\text{deg}E_1 = 3$ , and the bond between allyl and metal is represented by the edge  $E_2$ ,  $\text{deg}E_2 = 4$ . Similarly, edges  $E_3$  and  $E_4$  represent  $\pi$ -bond in cyclopentadienyl group and the bond between ligand and metal, respectively. All polycentric bonds of allylic complex are presented in this molecular hypergraph. It is also possible to consider a molecular hypergraph which has hyperedges corresponding only to metal–ligand bonds in the molecular structure (Figure 11 b)). In both cases hydrogens are suppressed. Figure 11 c) shows molecular hypergraph where polycentric bonds and covalent carbon-carbon and carbon-hydrogen bonds are represented. As it seems, Figure 11 c) gives an adequate hypergraph representation of allylic complex, and Figure 11 a),b) are misleading with respect to symmetry. Until now the unlabeled molecular hypergraphs were considered only. Figure 11 d)-f) shows the labeled hypergraphs of allylic complex.



allylic complex

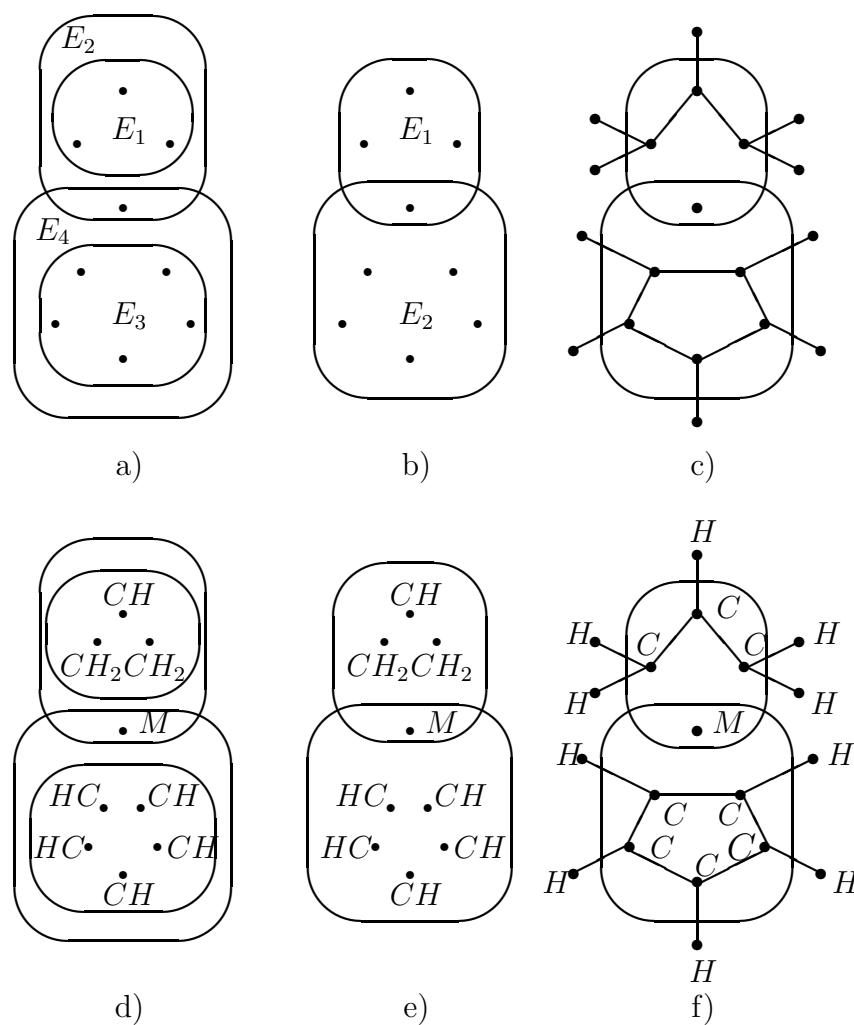


Figure 11. Molecular unlabeled (a-c) and labeled (d-f) hypergraphs representing allylic complex with different degrees of details

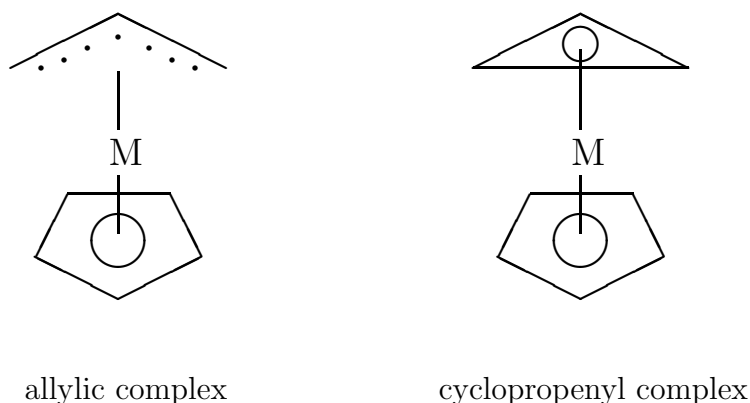
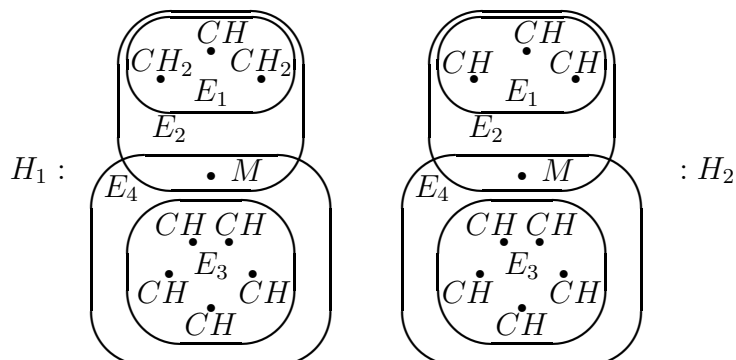


Figure 12. Allylic and cyclopropenyl complexes

Therefore, there are the number of hypergraph representations with the different degrees of details for one and the same molecular structures. Now let us consider the various ways of hypergraph representations with the different types of labels for allylic and cyclopropenyl complexes given in Figure 12 [48]. It is known that the density of three  $\pi$ -electrons available for metal–ligand bond formation is distributed between three carbon atoms in the allylic and cyclopropenyl ligands. In allylic complex, the carbon atoms of the  $\pi$ -allyl group form an open chain, whereas in cyclopropenyl complex, the carbon atoms form a three-membered ring. These chemical differences can be reflected in hypergraphs.

*The first way.* Molecular structures are represented by the labeled hypergraphs whose vertex and edge labels are elements of the symbolic alphabets.

Figure 13 shows molecular hypergraphs  $H_1$  and  $H_2$  taking into account all polycentric bonds of allylic and cyclopropenyl complexes. The polycentric bonds in allyl and cyclopropenyl ligands are represented by hyperedges  $E_1$ ,  $\text{deg}E_1 = 3$ , in hypergraphs  $H_1$  and  $H_2$ . The polycentric bonds between the allylic ligand and metal and between the cyclopropenyl ligand and metal are represented by hyperedges  $E_2$ ,  $\text{deg}E_2 = 4$ , in hypergraphs  $H_1$  and  $H_2$ . Analogously, hyperedges  $E_3$ ,  $\text{deg}E_3 = 5$ , and  $E_4$ ,  $\text{deg}E_4 = 6$ , describe the  $\pi$ -bonds in the cyclopentadienyl ligands and the metal–ligand bonds. The vertex and edges alphabets of hypergraph  $H_1$  representing allylic complex look like  $A(V(H_1)) = \{CH, CH_2, M\}$  and  $A(E(H_1)) = \{E_1, E_2, E_3, E_4\}$ .



$$A(V(H_1)) = \{CH, CH_2, M\} \quad A(V(H_2)) = \{CH, M\}$$

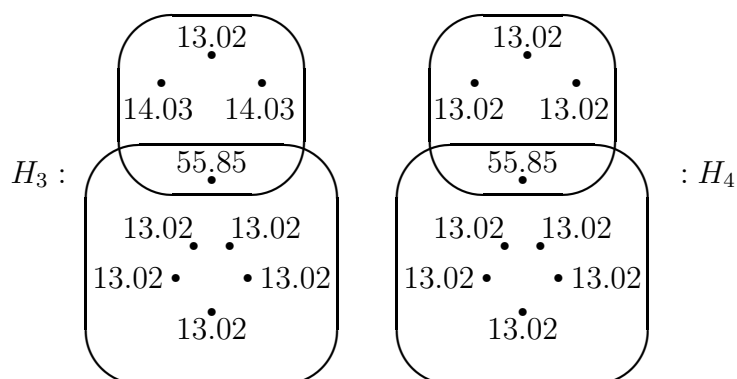
$$A(E(H_1)) = A(E(H_2)) = \{E_1, E_2, E_3, E_4\}$$

Figure 13. Hypergraphs with the symbolic vertex and edge labels representing allylic and cyclopropenyl complexes given in Figure 12

The vertex and edges alphabets of hypergraph  $H_2$  representing cyclopropenyl complex look like  $A(V(H_2)) = \{CH, M\}$  and  $A(E(H_2)) = \{E_1, E_2, E_3, E_4\}$ . The vertex labels are presented by symbols of functional chemical groups. It is easy to see that the vertex alphabets are the different for these hypergraphs.

*The second way.* Molecular structures are represented by the weighted hypergraphs, i.e., by labeled hypergraphs with numerical labels.

Figure 14 shows molecular hypergraphs  $H_3$  and  $H_4$  which contain only hyperedges representing the metal–ligand bonds in molecular structures. The total atomic masses of elements comprising functional groups are used as numerical labels of vertices. For example, the label of the vertex  $v$  corresponding to  $CH$ -group is defined as  $a(v(CH)) = m_C + m_H = 12.01 + 1.01 = 13.02$ , where  $m_C$  and  $m_H$  are the atomic masses of carbon and hydrogen. The metal atom may be represented by iron, whose atomic mass is 55.85.



$$A(V(H_3)) = \{13.02, 14.03, 55.85\} \quad A(V(H_4)) = \{13.02, 55.85\}$$

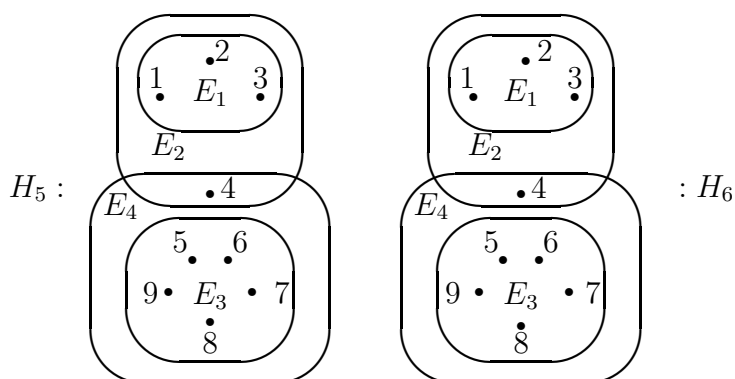
Figure 14. Hypergraphs with the numerical vertex labels representing allylic and cyclopropenyl complexes given in Figure 12

This way of hypergraph representation of molecular structures disregards the structure of ligands in the complexes. Yet it gives unambiguous descriptions, because the incidence matrices of the hypergraphs differ:

$$B(H_3) = \begin{bmatrix} 14.03 \\ 13.02 \\ 14.03 \\ 55.85 & 55.85 \\ & 13.02 \\ & 13.02 \\ & 13.02 \\ & 13.02 \\ & 13.02 \end{bmatrix} \quad B(H_4) = \begin{bmatrix} 13.02 \\ 13.02 \\ 13.02 \\ 55.85 & 55.85 \\ & 13.02 \\ & 13.02 \\ & 13.02 \\ & 13.02 \\ & 13.02 \end{bmatrix}$$

This example demonstrates the possibility of hypergraph modeling for molecular structure representation with the different degrees of accuracy.

*The third way.* Molecular structures are represented by the labeled hypergraphs with the structural labels.



$$A(E(H_5)) = \left\{ \begin{pmatrix} (123) \\ (233) \end{pmatrix}, \begin{pmatrix} (1234) \\ (4444) \end{pmatrix}, \begin{pmatrix} (56789) \\ (67895) \end{pmatrix}, \begin{pmatrix} (456789) \\ (444444) \end{pmatrix} \right\}$$

$$A(E(H_6)) = \left\{ \begin{pmatrix} (123) \\ (231) \end{pmatrix}, \begin{pmatrix} (1234) \\ (4444) \end{pmatrix}, \begin{pmatrix} (56789) \\ (67895) \end{pmatrix}, \begin{pmatrix} (456789) \\ (444444) \end{pmatrix} \right\}$$

Figure 15. Hypergraphs with the structural edge labels representing allylic and cyclopropenyl complexes given in Figure 12

Figure 15 gives the hypergraph  $H_5$  and  $H_6$ , whose edge labels are permutations and transformations. A *transformation* is a mapping of the set into itself, i.e., various elements of the set may be mapped into one and the same element. This distinguishes a transformation from a permutation. Each transformation presented in Figure 15 is defined on a set of vertices incident to the given hyperedge. So the transformation for the edge  $E_2$  is defined on a set of vertices  $\{1, 2, 3, 4\}$ , and the transformation for the edge  $E_4$  is defined on a set of vertices  $\{4, 5, 6, 7, 8, 9\}$ . The hyperedges  $E_2$  in the both hypergraphs are represented by the label  $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 4 & 4 & 4 \end{pmatrix}$  emphasizing that all carbon atoms of allyl and cyclopropenyl ligands donate their electrons to form the metal–ligand bond. The labels of hyperedges  $E_4$  are defined analogously. In the both hypergraphs the hyperedges  $E_3$  are labeled by the cyclic permutation on a set of vertices  $\{5, 6, 7, 8, 9\}$ , reflecting

the structure of cyclopentadienyl. The cyclic permutation on a set of vertices  $\{1, 2, 3\}$  is used as a label of hyperedge  $E_1$  in the hypergraph  $H_6$  to represent the bond structure in cyclopropenyl whose three atoms form a ring. The transformation  $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 3 \end{pmatrix}$  is used as a label of hyperedge  $E_1$  in the hypergraph  $H_5$  representing the allylic ligand since allyl is a chain of three  $sp^2$ -hybridized carbon atoms.

This example vividly illustrates the relationship between the structure of chemical compounds and the labels used in molecular hypergraphs. In particular, the labels of hyperedges  $E_2$  and  $E_4$  adequately reflect the essence of the metal–ligand bonds, in which ligands donate their electrons to form these bonds.

### 3.3 Examples of graph and hypergraph representation

Now we will give three examples to illustrate the difference between graphs and hypergraphs in the representation of the same molecular structures.

Figure 16 illustrates the difference between molecular 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  $M$  with the other eight vertices. This drawback is eliminated when allylic complex is represented by a hypergraph. Hyperedges  $E_1$  and  $E_2$  represent  $\pi$ -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 covalent (carbon-carbon and carbon-hydrogen) bonds and  $\pi$ -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.

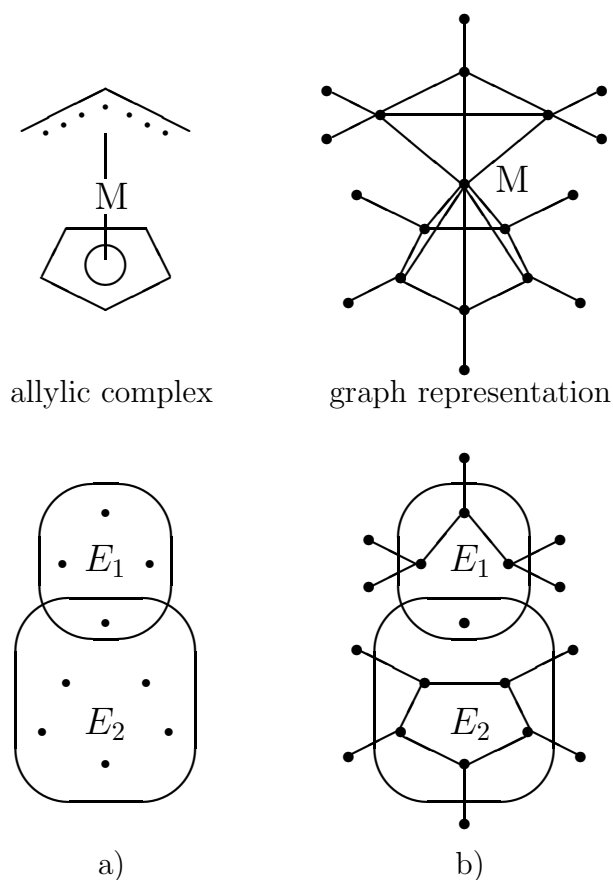


Figure 16. Graph and hypergraph (a, b)) representations of allylic complex

Another example given in Figure 17 illustrates the difference between molecular graphs and hypergraphs in the representation of benzene. Benzene is a conjugated molecule [18]. A special symbol is introduced to denote the six  $\pi$ -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  $\pi$ -electrons in benzene. The fundamental principle of the tetravalency of carbon is fulfilled in this case because the valency of a carbon atom is being conserved.

Hypergraphs appear to be convenient for a description of Clar aromatic sextet formulas of conjugated molecules. In this case hyperedges correspond to aromatic sextets. Figure 18 shows two different

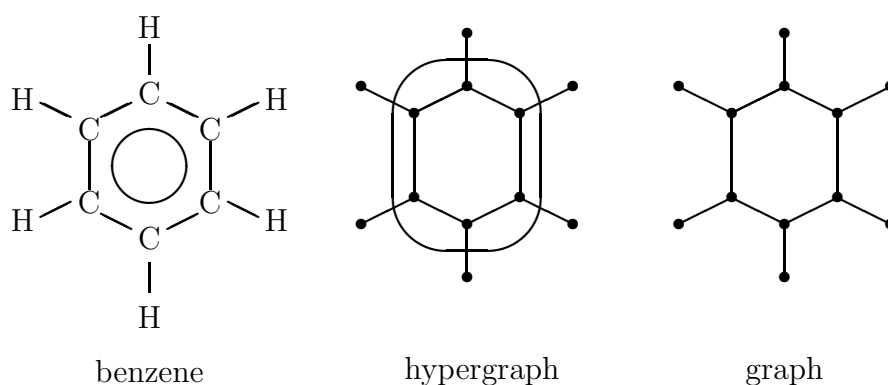


Figure 17. Hypergraph and graph benzene representations including hydrogens

hypergraph representations for two formulas of anthracene given in Figure 4. An attempt to describe these two structures by an ordinary graphs produces one and the same graph shown in Figure 7.

Thus the concept of a molecular hypergraph was introduced in this chapter. 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 various ways for hypergraph representation of molecular structures with the different degrees of detail and different types of labels were also considered. The choice of a representation is defined by the conditions of a considered problem. Combined use of these ways is possible. The accuracy of hypergraph description of molecular structures may be important for establishing the structure–property–activity relationships [41, 42].

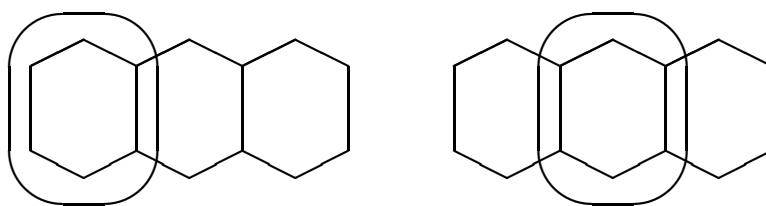


Figure 18. Hypergraph representation of Clar aromatic sextet formulas of anthracene

## Chapter 4

# Identification of molecular hypergraph

The problem, known as the identification problem, is to decide whether two given graphs or hypergraphs are isomorphic or not. Two graphs or hypergraphs are *isomorphic* if there exists a mapping conserving the one-to-one correlation between the sets of vertices and edges of these graphs or 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 that one, where 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 for defining the canonical form of an adjacency matrix are described in [1, 20, 39, 40, 51, 58].

The incidence matrix is convenient for the hypergraph computer representation. Figure 19 gives three examples of hypergraphs and their incidence matrices for the four-electron ligands given in Figure 3.

Let us define the canonical incidence matrix of a hypergraph [8, 73].

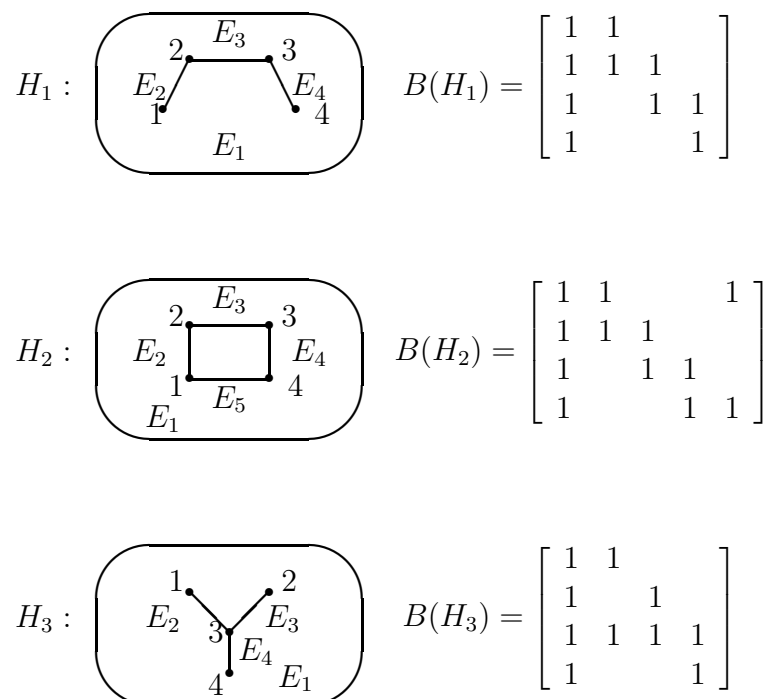


Figure 19. Hypergraphs representing the four-electron ligands (see Figure 3) and their incidence matrices

## 4.1 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 [6, 56]. It is evident that the canonical incidence matrix is unique and can be used for determining the canonical form of a 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-suppressed structure of allylic complex (Figure 6) and the canonical incidence matrix are shown in Figure 20.

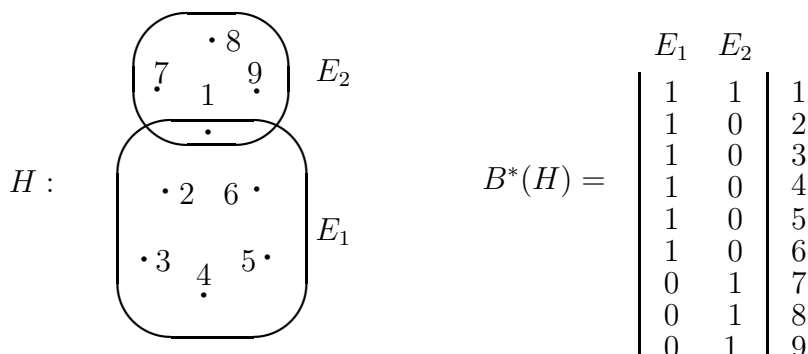


Figure 20. Canonical form and canonical incidence matrix of a hypergraph of allylic complex

There are works [9, 60] dealing with the theoretic aspects of isomorphism problems for hypergraphs. We propose the algorithmic approach to solve this problem.

## 4.2 Algorithm for canonical incidence matrix construction

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 [73]) with the set of vertices  $V \cup E$ ,  $|V \cup E| = p + q = s$ , and the set of edges  $Y$ ,  $|Y| = \sum_{j=1}^q \text{deg}E_j$ .

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 [73]. 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, analogously 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.

**Theorem 2** *The canonical adjacency matrix of König's representation of hypergraph  $H$  has a symmetrical block-diagonal form:*

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

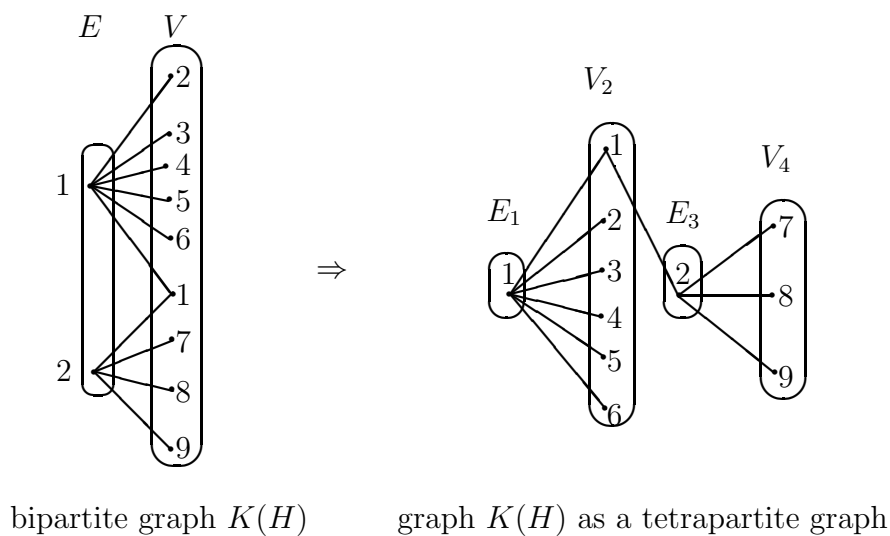
where rows (columns) of non-zero matrices

$$A_{n_i n_{i+1}}(A_{n_{i+1} n_i}), i = 1, \dots, t - 1,$$

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 subset of the set of hyperedges,  $E_{i+1} \subset E$ , and vice versa.

Proof: Let the canonical numbering of the vertices of graph  $K(H)$  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. Really the canonical numbering of  $K(H)$  can start as with the  $V$ -part as 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 only the vertices from adjacent parts are adjacent. So non-zero 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 proven.  $\square$

The example in Figure 21 illustrates the Theorem 2.



$\Rightarrow A^* =$

	$E_1$	1	2	3	4	5	6	$E_2$	7	8	9		
0	1	1	1	1	1	1	1	1				$E_1\}n_1$	
1		0						1					1
1		0											2
1		0											3
1		0											4
1		0											5
1		0											6
1	1	0						0	1	1	1		$E_2\}n_3$
		0						1					7
		0						1	0				8
		0						1					9

canonical adjacency matrix of graph  $K(H)$

Figure 21. König's representation of hypergraph  $H$  from Figure 20 and its canonical adjacency matrix

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 resulted matrix  $R(A^*(K(H)))$  will have one of the following forms:

$$R(A^*(K(H))) = \begin{array}{c} \begin{array}{cccccc} & n_2 & n_4 & n_6 & & n_{t-3} & n_{t-1} \\ \hline A_{12} & 0 & 0 & \cdots & & & \\ A_{32} & A_{34} & 0 & \cdots & & & \\ \hline & A_{54} & A_{56} & \cdots & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ \hline & & & \cdots & A_{t-2t-3} & A_{t-2t-1} \\ & & & \cdots & 0 & A_{tt-1} \\ \hline \end{array} & \begin{array}{l} n_1 \\ n_3 \\ n_5 \\ n_{t-2} \\ n_t \end{array} \\ \\ \begin{array}{c} \begin{array}{cccccc} & n_1 & n_3 & n_5 & & n_{t-2} & n_t \\ \hline A_{21} & A_{23} & 0 & \cdots & & & \\ 0 & A_{43} & A_{45} & \cdots & & & \\ \hline & 0 & A_{65} & \cdots & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ \hline & & & \cdots & A_{t-3t-2} & 0 \\ & & & \cdots & A_{t-1t-2} & A_{t-1t} \\ \hline \end{array} & \begin{array}{l} n_2 \\ n_4 \\ n_6 \\ n_{t-3} \\ n_{t-1} \end{array} \end{array}$$

**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  $A^*(K(H))$ , what 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 proven.  $\square$

Let us note that Figure 20 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  $K(H)$  (Figure 21) by the method discussed above.

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

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

The evaluation of the algorithm complexity is determined by the complexity of the algorithm for the canonical adjacency matrix searching of ordinary graphs. 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, 20]. Algorithms for canonization of incidence matrices [6, 56] that are known to the author are enumerative and the estimations of complexity for the given algorithms are either not presented [6], or not allow to estimate the running effectivity of an algorithm in practice [56].

## Chapter 5

# 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) [13, 21, 59].

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  $G_2$ .

The subgraph isomorphism problem states as follows. Given two graphs  $G_1, m = |V(G_1)|$  and  $G_2, n = |V(G_2)|, m \leq n$ , find all 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) to find sets of subgraphs of order  $k$  of  $G_1$  and  $G_2$ .
- 2) to test all pairs subgraphs for the isomorphism.

The step 2) can be realized, for example, by a trivial algorithm known as vertex reordering algorithm [20], which is based on vertex permutations and adjacency matrix comparisons. If this algorithm is used, finding common subgraphs of order  $k$  requires  $\frac{m!n!}{k!(m-k)!(n-k)!}$  matrix comparisons. Returning to hypergraphs this algorithm requires  $\frac{(p_1+q_1)!(p_2+q_2)!}{k!(p_1+q_1-k)!(p_2+q_2-k)!}$ , 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  $G_2 = (U, X)$  [7, 17, 52]. 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 Figure 22).

The main property of a correspondance graph is the following one: each clique of graph  $L(G_1, G_2)$  corresponds to the some common subgraph of  $G_1$  and  $G_2$  [52]. So to find the common subgraph of  $G_1$  and  $G_2$  it is enough 1) to have the correspondance graph  $L(G_1, G_2)$  and 2) to derive all cliques in  $L(G_1, G_2)$ .

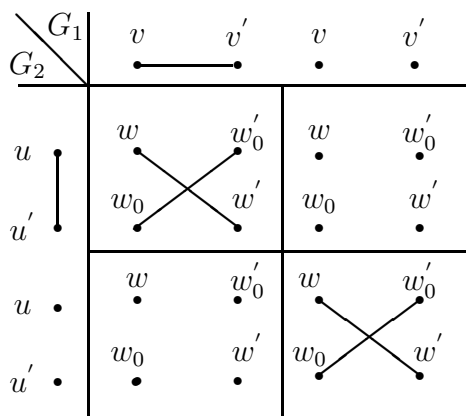


Figure 22. The table representation of a correspondence graph

The complexity of the subgraph isomorphism problem is equivalent to the complexity of the clique finding problem [30]. To find all cliques, the graph recursive analysis technique is used [10, 16, 23]. 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 finished when the some rule of stop is fulfilled.

The theoretical estimate of complexity for the cliques finding algorithm is established as  $O(l^2 \cdot 3^{\frac{l}{3}})$ , where  $l$  is the order of a correspondance graph [10]. This estimate is achieved for the  $k$ -partite graphs  $K_{3,3,\dots,3}$  that are known as the clique extremal graphs [53]. Returning to a hypergraph this algorithm requires  $((p_1+q_1+p_2+q_2)^2 \cdot 3^{\frac{(p_1+q_1+p_2+q_2)}{3}})$ . In practice algorithms of the first [20] and second [10, 16] approaches are realized.



## Chapter 6

# Invariants of graphs and hypergraphs

To identify molecular structures of chemical compound the molecular graph and hypergraph invariants, called topological indices [3, 11, 69], can 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  $H$  is equal 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**  $d_G(u, v) \geq d_H(u, v)$ .

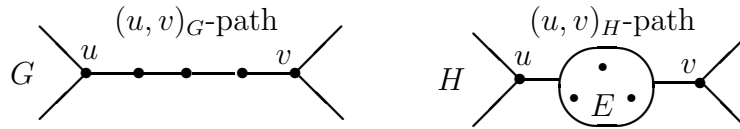
Proof: Let us consider two cases.

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

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

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.

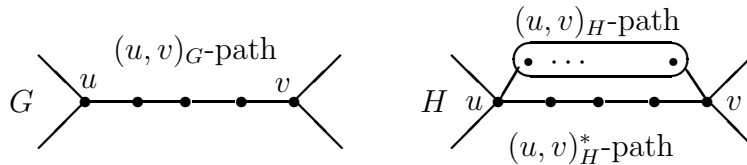
Suppose that the same set of vertices belongs to both paths:



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

$$d_H(u, v) = d_G(u, v) - \deg E + 2, \text{ i.e.} \\ d_H(u, v) < d_G(u, v)$$

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

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

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

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 1,2 and 3 it follows that

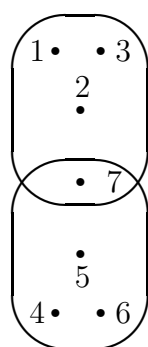
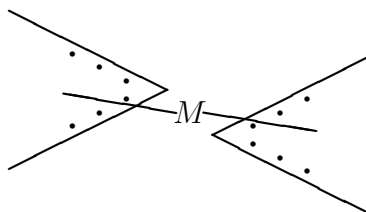
$$d_H(u, v) < d_G(u, v)$$

The Theorem is proven.  $\square$

The Theorem 4 implies that for any  $F$  represented by  $G$  and  $H$  the following inequality holds  $W(G) > W(H)$ , where  $W = \frac{1}{2} \sum_{u,v \in V} d(u, v)$  is the first mathematical invariant reflecting the topological structure of a molecular graph, called *the Wiener index* [71]. The example of obtaining the values of the Wiener index for graph and hypergraph representations of the molecular structure with three-electron ligands presented in Figure 2 is shown in Figure 23. Hydrogens are suppressed in these representations.

The same inequalities hold for some another metric and topological indices presented in Appendix A and Appendix B. For example,  $rad(G) > rad(H)$  and  $diam(G) > diam(H)$ , where  $rad = \min_{v \in V} \max_{u \in V} d(u, v)$  is the radius and  $diam = \max_{v \in V} \max_{u \in V} d(u, v)$  is the diameter [36].

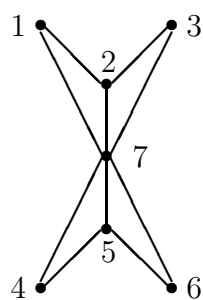
A few other topological indices are likely to reveal similar regularities. The behavior on the whole set of graphs and hypergraphs cannot be estimated for all topological indices.

hypergraph  $H$ 

$$\Rightarrow D(H) = \begin{array}{c|ccccccc|c} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \Sigma \\ \hline 0 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 9 \\ 1 & 0 & 1 & 2 & 2 & 2 & 2 & 1 & 9 \\ 1 & 1 & 0 & 2 & 2 & 2 & 2 & 1 & 9 \\ 2 & 2 & 2 & 0 & 1 & 1 & 1 & 1 & 9 \\ 2 & 2 & 2 & 1 & 0 & 1 & 1 & 1 & 9 \\ 2 & 2 & 2 & 1 & 1 & 0 & 1 & 1 & 9 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 6 \end{array}$$

Distance matrix

$$\Rightarrow W(H) = (9 + 9 + 9 + 9 + 9 + 9 + 6)/2 = 30$$

graph  $G$ 

$$\Rightarrow D(G) = \begin{array}{c|ccccccc|c} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \Sigma \\ \hline 0 & 1 & 2 & 2 & 2 & 2 & 2 & 1 & 10 \\ 1 & 0 & 1 & 2 & 2 & 2 & 2 & 1 & 9 \\ 2 & 1 & 0 & 2 & 2 & 2 & 2 & 1 & 10 \\ 2 & 2 & 2 & 0 & 1 & 2 & 2 & 1 & 10 \\ 2 & 2 & 2 & 1 & 0 & 1 & 1 & 1 & 9 \\ 2 & 2 & 2 & 2 & 1 & 0 & 1 & 1 & 10 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 6 \end{array}$$

Distance matrix

$$\Rightarrow W(G) = (10 + 9 + 10 + 10 + 9 + 10 + 6)/2 = 32$$

Figure 23. The example of obtaining the values of the Wiener index for hypergraph and graph representations of the molecular structure with three-electron ligands; hydrogens are suppressed in these representations

## 6.1 Comparative analysis of invariants for organometallic compounds

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 widespread structural types presented in Figure 24 (methyl  $M1$ , allyl  $M2$ , isobutyl  $M3$ , phenyl  $M4$ , cyclohexyl  $M5$ , ferrocenyl  $M6$ , benchtorenlyl  $M7$ , allylferrotricarboxylidide  $M8$ ) 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 [47, 49].

Three groups of invariants presented in Appendixes 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 [57, 65]. Topological indices are well-known in

Table 1. The widespread substituents  $X$

N	$X$	N	$X$
1	$R-H$	16	$R-COOC_2H_5$
2	$R-CH_3$	17	$R-COCH_3$
3	$R-C_2H_5$	18	$R-CH_2OH$
4	$R-(CH_2)_2CH_3$	19	$R-N(CH_3)_2$
5	$R-CH(CH_3)_2$	20	$R-CH(OH)CH_3$
6	$R-(CH_2)_3CH_3$	21	$R-CH(OH)C(CH_3)_3$
7	$R-C(CH_3)_3$	22	$R-CH_2C_6H_5$
8	$R-CN$	23	$R-COC_6H_5$
9	$R-CH=CH_2$	24	$R-C_6H_{11}$
10	$R-COOH$	25	$R-C(CH_3)=CH_2$
11	$R-OCOH$	26	$R-CH=CHCH_3$
12	$R-COOCH_3$	27	$R-CH=CHCOC_6H_5$
13	$R-OCOCH_3$	28	$R-COOC_6H_5$
14	$R-COH$	29	$R-OCOC_6H_5$
15	$R-CH=CHC_6H_5$	30	$R-(CH_2)_4COOH$

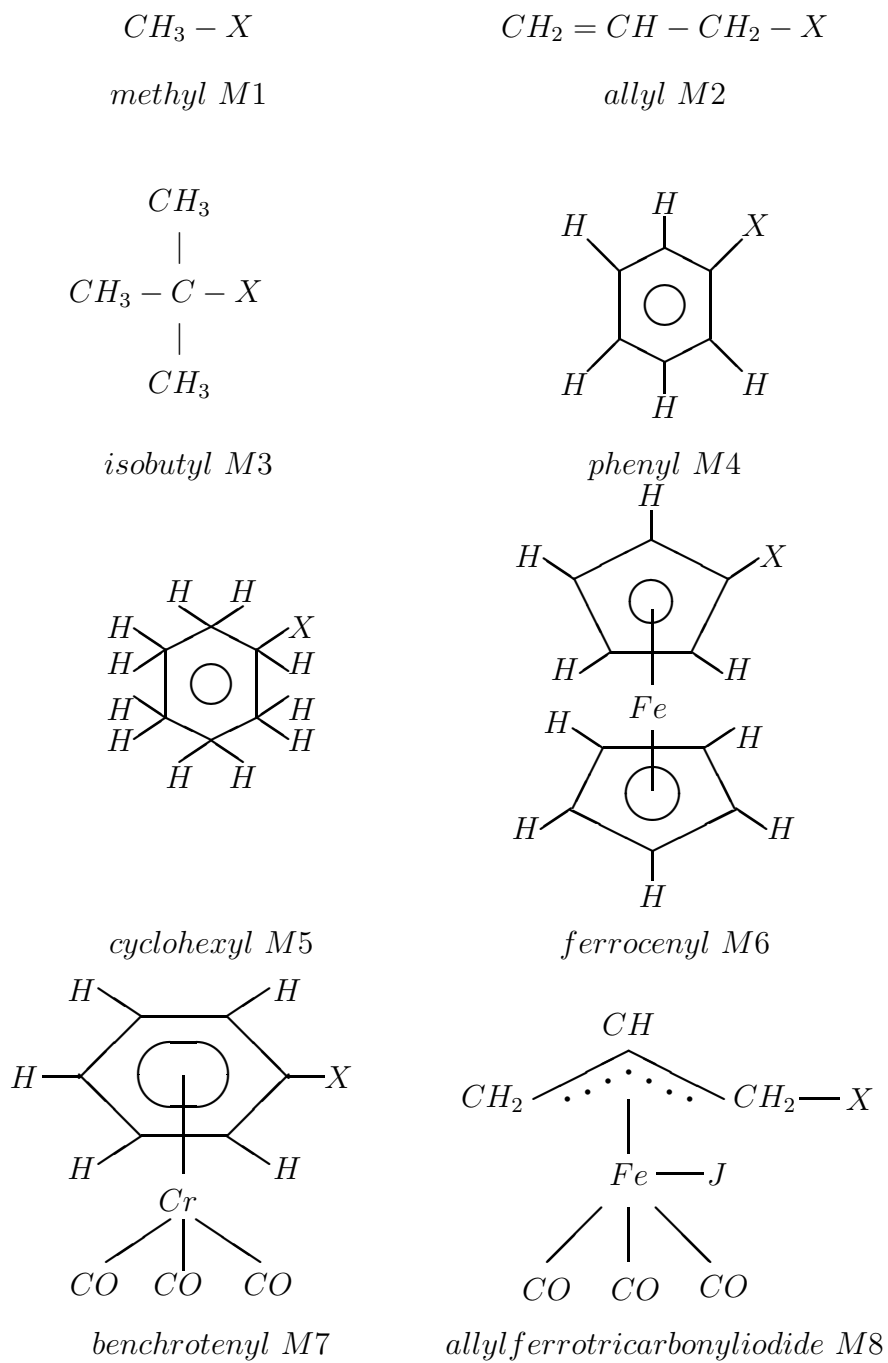


Figure 24. The widespread structural types

mathematical chemistry and used for representation of molecular structure and for finding of the quantitative structure-property and structure-activity relationships [69]. Information indices [11, 14] are constructed for various matrices (layer 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 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 [15, 61] 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 Appendixes for both models are calculated.

The results have shown that for considered series of molecular structures the indices sensitivity for the hypergraph model are higher than ones for the graph model. The data submitted below show the quantity of indices, the sensitivity  $S_G$  of which in the graph model is less, more, or equal to the sensitivity  $S_H$  of the indices in the hypergraph model:

Sensitivity	%
$S_G < S_H$	21.7
$S_G > S_H$	8.7
$S_G = S_H$	69.6

The indices correlations for considered models are also investigated. Let us define the noncorrelating indices as ones 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 Figure 25 correspond to the percentage of noncorrelating indices for different models. The Figure 25 shows that the number of noncorrelating indices for the hypergraph model is greater than for the graph one for all molecular structures series. For example on the set  $M8$  this number exceeds the same parameter for the graph model 1.5 times as large.

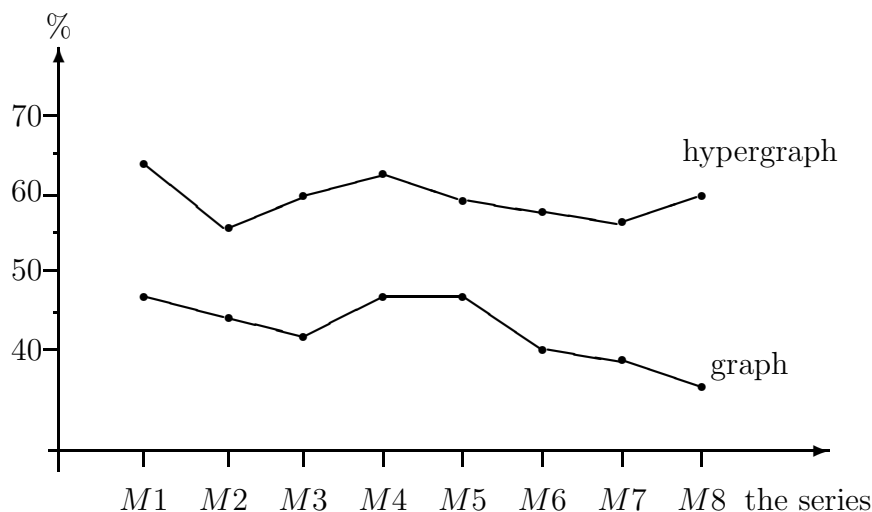


Figure 25. The percentage of noncorrelating indices for different models

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

## 6.2 Comparative analysis of invariants for benzenoid hydrocarbons

The results of comparative analysis of the invariants values for fully-benzenoid hydrocarbons (fully-benzenoid systems) are presented in this section [35]. Benzenoid systems in which there are no double bonds in the Clar aromatic sextet formulas, i.e., in which all  $\pi$ -electrons belong to aromatic sextets, are called *fully-benzenoid* [32]. One example of fully-benzenoid system is shown in Figure 26.

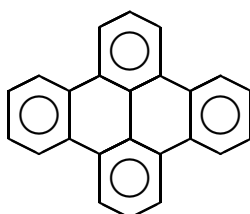


Figure 26. Clar aromatic sextet formula of dibenzo[*fg,op*]naphthalene

The following three topological indices were investigated: the Wiener index  $W$ , the mean square distance index  $SDI$ , and the vertex distance index  $VDI$  (formulas 13, 15 and 16 in Appendix B). The Wiener index is the oldest and one of the most thoroughly studied topological indices (for review see [24, 26, 29, 34, 50, 54]). The indices  $SDI$  and  $VDI$ , proposed by Rouvray [62], reflect the structural features other than  $W$ . It is known that both  $W$  and  $SDI$  decrease with increasing branching of the molecular skeleton, the  $VDI$ -index exhibits an opposite, decreasing trend. As was mentioned above  $W(G) > W(H)$  (see the beginning of the section 6). Taking into account Theorem 4 and the formulas of indices  $SDI$  and  $VDI$ , we immediately arrive at the following inequalities:  $SDI(G) > SDI(H)$  and  $VDI(G) < VDI(H)$ . The values of these indices were calculated for graphs and hypergraphs representing the 28 fully-benzenoid molecules with 11 or less hexagons whose structures are given in Figure 27. For the class of these molecules a very good linear correlations were found:

$$W(H) = (0.649 \pm 0.006) W(G) + b_W \quad ; \quad r = 0.9989$$

$$SDI(H) = (0.67 \pm 0.02) SDI(G) + b_{SDI} \quad ; \quad r = 0.992$$

$$VDI(H) = (1.97 \pm 0.04) VDI(G) + b_{VDI} \quad ; \quad r = 0.995,$$

where  $r$  stands for the correlation coefficient. The calculated values of the parameters  $b_W$ ,  $b_{SDI}$  and  $b_{VDI}$  are  $4 \pm 20$ ,  $-0.2 \pm 0.1$  and  $400 \pm 2300$ , respectively, and therefore it is reasonable to set them equal to zero in the above expressions.

For the class of the 16 benzenoid systems with two or three aromatic sextets (hyperedges), which are not fully-benzenoid, but which have a unique Clar aromatic sextet formula we found

$$W(H) = (0.72 \pm 0.01) W(G) + b_W \quad ; \quad r = 0.998$$

$$SDI(H) = (0.66 \pm 0.05) SDI(G) + b_{SDI} \quad ; \quad r = 0.96$$

$$VDI(H) = (1.54 \pm 0.03) VDI(G) + b_{VDI} \quad ; \quad r = 0.997$$

with  $b_W = -30 \pm 10$ ,  $b_{SDI} = 0.05 \pm 0.2$  and  $b_{VDI} = 1000 \pm 300$ , which should be neglected as well.

Now let us denote by  $\Delta W$ ,  $\Delta SDI$  and  $\Delta VDI$  the differences  $W(G) - W(H)$ ,  $SDI(G) - SDI(H)$  and  $VDI(H) - VDI(G)$ , respectively. It is clear all these three differences are necessarily positive-valued.

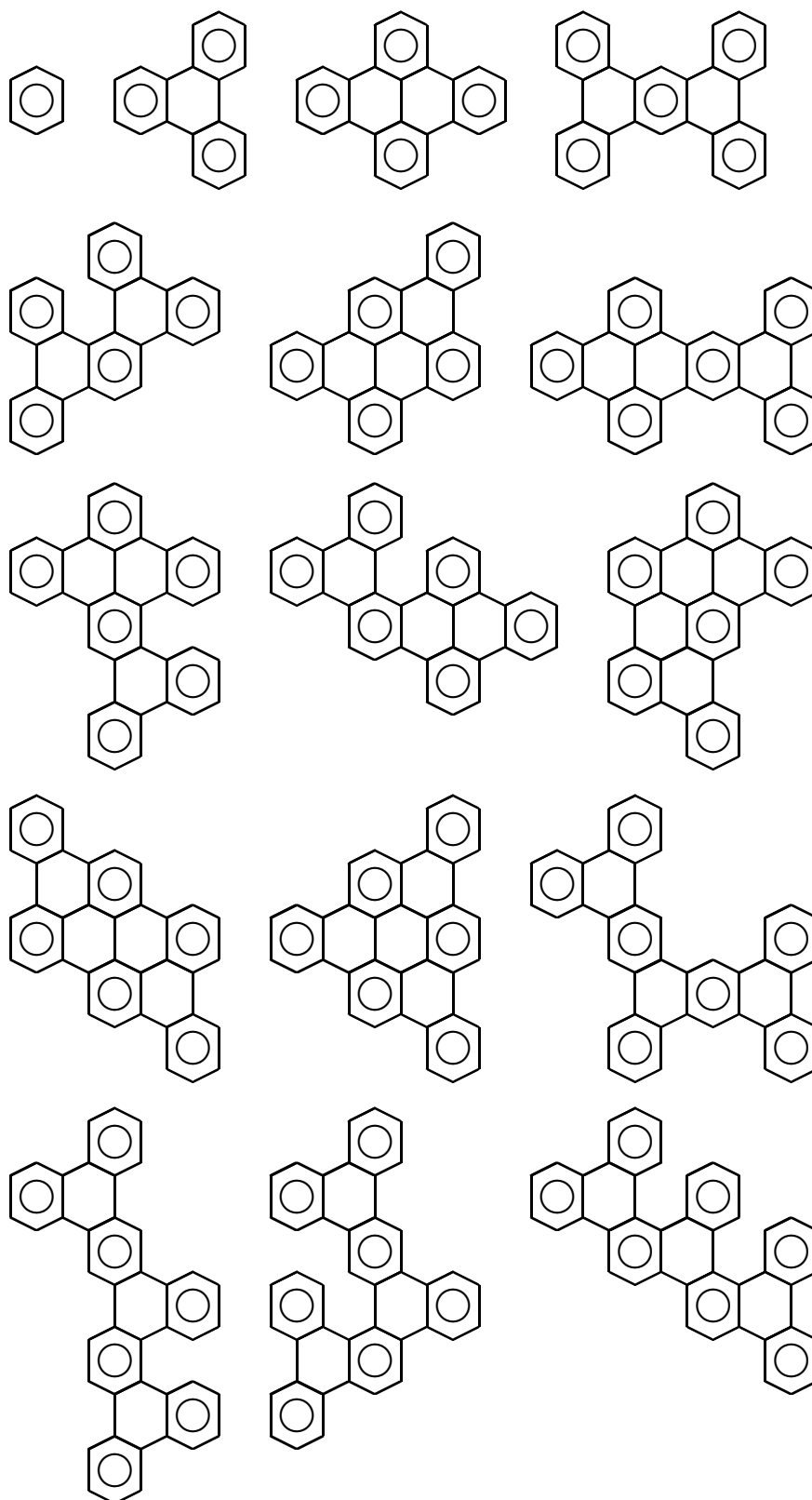


Figure 27. The set of fully-benzenoid molecules

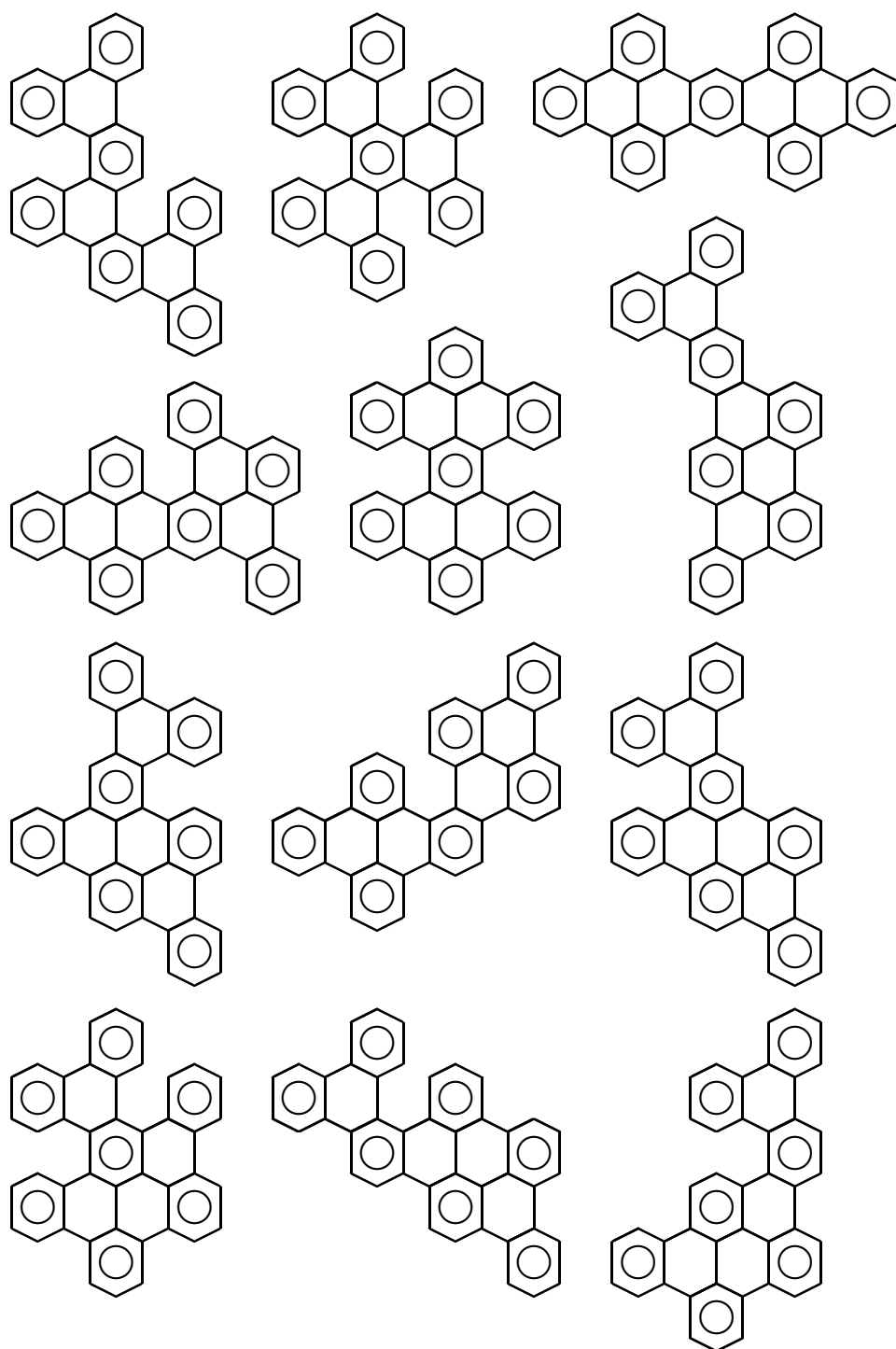


Figure 27. The set of fully-benzenoid molecules (continuation)

The next two properties of topological indices were obtained also.

**Property 1.** The values of  $\Delta W$ ,  $\Delta SDI$  and  $\Delta VDI$  rapidly increase with the increasing size (number of vertices and/or number of aromatic sextets) of the benzenoid molecule and vary to a much lesser extent among benzenoid isomers.

The examples illustrating the size-dependency of  $\Delta W$ ,  $\Delta SDI$  and  $\Delta VDI$  and their relatively low variation in the case of isomers (molecules F3–F7) are shown in Figure 28.

As another illustration of Property 1 we give the formulas for  $W(G)$ ,  $W(H)$  and  $\Delta W$  for the homologous series whose first members are the fully-benzenoid molecules F1 ( $x=0$ ), F2 ( $x=1$ ), F3 ( $x=2$ ) and F8 ( $x=3$ ), see Figure 28:

$$W(G) = 96x^3 + 216x^2 + 246x - 45 \quad \text{for } x \geq 1$$

$$W(H) = 48x^3 + 192x^2 + 78x + 15 \quad \text{for } x \geq 0$$

$$\Delta W = 48x^3 + 24x^2 + 168x + 60 \quad \text{for } x \geq 1.$$

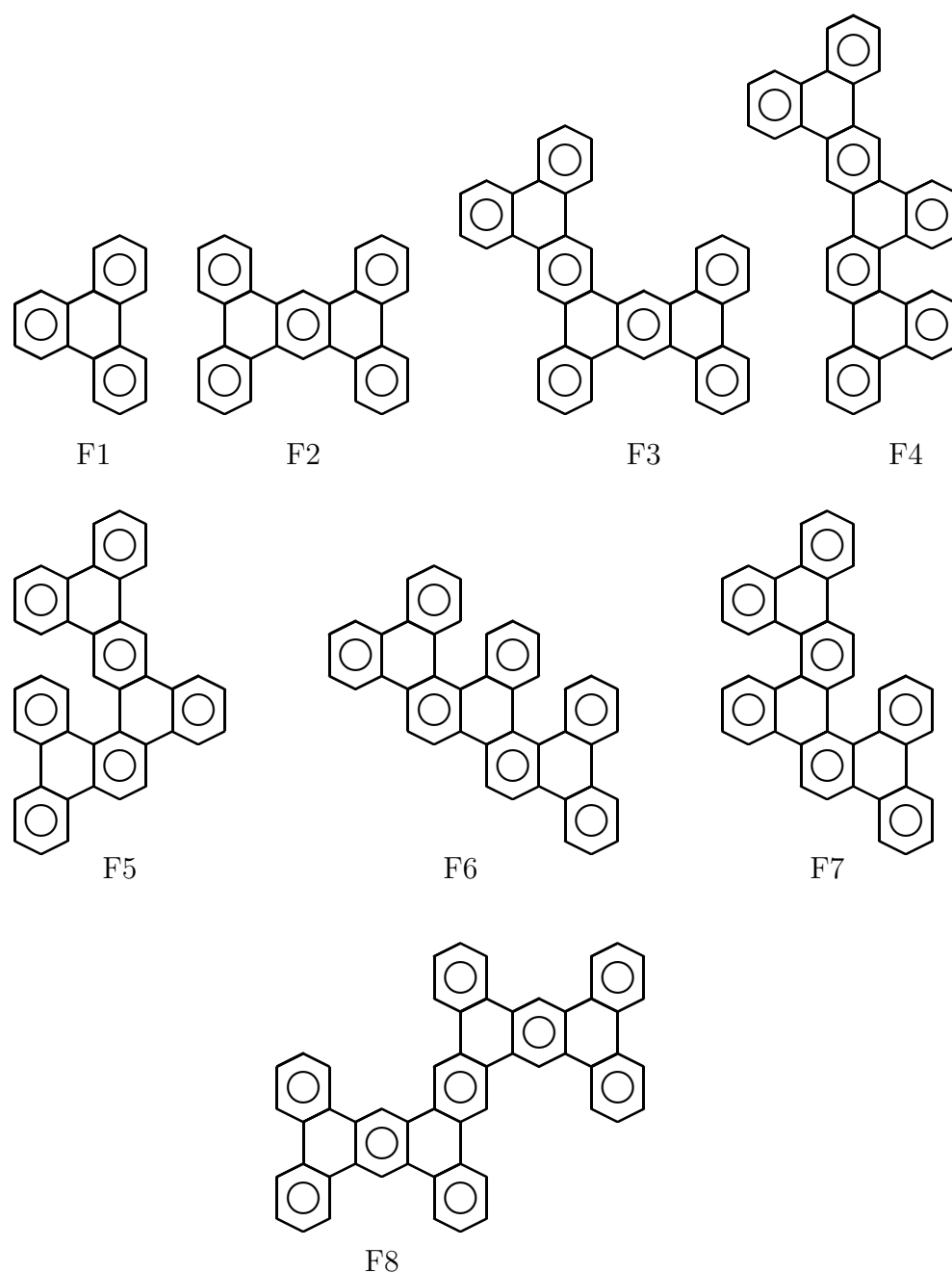
We see that  $\Delta W$  increases as the third power of the number of aromatic sextets (i.e., hyperedges).

In a sharp contrast to Property 1 it was found that the  $\Delta$ -values are almost independent of the number of hexagons.

**Property 2.** For benzenoid molecules with a fixed number of aromatic sextets, but with different number of hexagons, the values of  $\Delta W$ ,  $\Delta SDI$  and  $\Delta VDI$  are nearly constant.

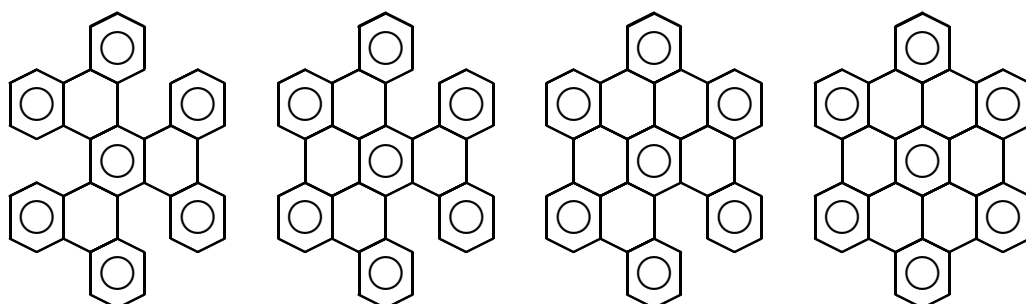
This property is illustrated by the data given in Figure 29. In this picture you can see the examples showing the almost-independence of  $\Delta W$ ,  $\Delta SDI$  and  $\Delta VDI$  of the number of hexagons, provided the number of aromatic sextets is fixed.

So the presented results show that the Clar hypergraph has properties quite different from the traditional molecular graph. These differences are best manifested in the case of topological indices. The Clar aromatic sextet theory was introduced without a theoretical foundation and without any mathematical formalism. Its remarkable success



	F1	F2	F3	F4	F5	F6	F7	F8
$\Delta W$	180	756	1956	1884	1740	1812	1668	4068
$\Delta SDI$	1.35	2.03	2.66	2.59	2.32	2.49	2.22	3.38
$\Delta VDI$	4770	24004	70952	69870	66176	68320	64994	159940

Figure 28. The benzenoid molecules and their  $\Delta W$ -,  $\Delta SDI$ -,  $\Delta VDI$ -values (the illustration of property 1)



$\Delta W$	1524	1508	1492	1476
$\Delta SDI$	2.00	1.99	1.97	1.96
$\Delta VDI$	96372	88116	84600	85824

Figure 29. The benzenoid molecules and their  $\Delta W$ -,  $\Delta SDI$ -,  $\Delta VDI$ -values (the illustration of property 2)

in describing, rationalizing and predicting experimental facts is not fully understood. By pointing out a hitherto overlooked “hypergraph connection” we hope to make a step forward towards constructing (or reconstructing) the mathematics of Clar theory.

# Appendix A

## Metric indices

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

1. The radius [36]

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

2. The diameter [36]

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

3. The average vertex eccentricity [65]

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

where  $p$  is a number of graph vertices.

4. The eccentricity [65]

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

5. The eccentric [65]

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

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

6. The average vertex distance [65]

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

7. The unipolarity [65]

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

8. The centralization [65]

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

9. The mean distance deviation [65]

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

10. The dispersion [55]

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

11. The compactness [27]

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

12. The variation [28]

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

# Appendix B

## Topological indices

13. The Wiener index [71]

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

14. The polarity number [72]

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

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

15. The mean square distance index [4]

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

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

16. The vertex distance index [62]

$$VDI = \sum_{i=1}^{max} g_i^2$$

17. The total adjacency index [62]

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

where  $a_{ij}$  is the adjacency matrix element.



# Appendix C

## Information indices

18. The information Weiner index [14]

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

19. The autometricity index [66]

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

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  $\lambda_{ij}$  is the number vertices at a distance  $j$  from vertex  $i$ .

20,21,22. The information integral indices  $H_\lambda, H_d, H_A$  are defined on the basis of the information vertex indices  $H_\lambda(i), H_d(i), H_A(i)$  [45]

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

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

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

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 [66]

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

# Bibliography

- [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. i mat.fis.*, **14** (1974) 737–743. (Russian).
- [2] A.T. Balaban, (ed.), *Chemical Applications of Graph Theory*, Academic Press, London, 1976.
- [3] A.T. Balaban, Applications of Graph Theory in Chemistry, *J. Chem. Inf. Comput. Sci.*, **25** (1985) 334–343.
- [4] A.T. Balaban, Topological Indices Based on Topological Distances in Molecular Graphs, *Pure Appl. Chem.*, **55** (1983) 199–206.
- [5] E.V. Babaev, In *Principles of Symmetry and Systemic Approach in Chemistry*, Nauka, Moscow, 1987, 30–54 (Russian).
- [6] A.M. Baraev, I.A. Faradzhev, In *Algorithmic Studies in Combinatorics*, I.A. Faradzhev, (ed.), Nauka, Moscow, 1978 (Russian).
- [7] H.G. Barrow, R.M. Burstall, Subgraph Isomorphism, Matching Relational Structures and Maximal Cliques, *Information Processing Letters*, **4** (1976) 83–84.
- [8] C. Berge, *Graphes et Hypergraphes*, Dunod, Paris, 1970.
- [9] C. Berge, Isomorphism Problems for Hypergraphs, *Lect. Notes Math.*, **411** (1974) 1–12.
- [10] Yu.E. Bessonov, V.A. Skorobogatov, On Graph Recursive Analysis, *Vychisl. Systemy*, **85** (1981) 3–20 (Russian).

- [11] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press, Chichester, 1983.
- [12] D. Bonchev, D.H. Rouvray, (eds.), *Chemical Graph Theory: Introduction and Fundamentals*, Abacus Press/Gordon & Breach Science Publishers, New York, 1991.
- [13] D. Bonchev, N. Trinajstić, Chemical Information Theory: Structural Aspects, *Intern. J. Quantum. Chem. Symp.*, **18** (1982) 463–480.
- [14] D. Bonchev, N. Trinajstić, Information Theory, Distance Matrix and Molecular Branching, *J. Chem. Phys.*, **67** (1977) 4517–4533.
- [15] D. Bonchev, O. Mekenyan, N. Trinajstić, Isomer Discrimination by Topological Information Approach, *J. Comp. Chem.*, **2** (1981) 127–148.
- [16] C. Bron, J. Kerbosh, Finding All Cliques of an Undirected Graph, *Comm. Assoc. Comput. Mach.*, **9** (1973) 575–577.
- [17] D.E. Chahraman, A.K.C. Wong, T. Au, Graph Optimal Monomorphism Algorithms, *IEEE Transaction on Systems, Man and Cybernetics*, **4** (1980) 181–197.
- [18] E. Clar, *The Aromatic Sextet*, J. Wiley, London, 1972.
- [19] J.P. Collman, L.S. Hegedus, *Principles and Applications of Organotransition Metal Chemistry*, J. Wiley, London, 1980.
- [20] D.G. Corneil, C.C. Gotlieb, An Efficient Algorithm for Graph Isomorphism, *J. Assoc. Comput. Mach.*, **1** (1970) 51–64.
- [21] R.D. Cramer, G. Redl, C.E. Berkoff, Substructural Analysis. A Novel Approach to the Problem of Drug Design, *J. Med. Chem.*, **5** (1974) 533–535.
- [22] S.J. Cyvin, I. Gutman, *Kekulé Structures in Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1988.

- [23] 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. and Elec. Eng.*, **4** (1978) 365–368.
- [24] M.V. Diudea, I. Gutman, Wiener-type Topological Indices, *Croat. Chem. Acta.*, **71** (1998) 21–51.
- [25] J.R. Dias, *Handbook of Polycyclic Hydrocarbons. Part A. Benzenoid Hydrocarbons*, Elsevier, Amsterdam, 1987.
- [26] A.A. Dobrynin, I. Gutman, Wiener Index for Trees and Hexagonal Systems, *Diskretn. Anal. Issled. Oper. Ser.2*, **2** (1998) 34–60 (Russian).
- [27] J.K. Doyle, J.E. Graver, Mean Distance in Graph, *Discrete Math.*, **17** (1977) 147–154.
- [28] R.C. Entringer, D.E. Jackson, D.A. Snyder, Distance in Graphs, *Czechoslovak Math.J.*, **2** (1976) 283–297.
- [29] R.C. Entringer, Distance in Graphs: Trees, *J. Combin. Math. Combin. Comput.*, **24** (1997) 65–84.
- [30] M.R. Garey, D.S. Johnson, *Computers and Intractability, a Guide to the Theory of NP-Completeness*, Bell Laboratories Murray Hill, New Jersey, 1979.
- [31] I. Gutman, O.E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.
- [32] I. Gutman, S.J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1989.
- [33] I. Gutman, S.J. Cyvin, (eds.), *Advances in the Theory of Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1990.
- [34] I. Gutman, Y.N. Yeh, S.L. Lee, Y.L. Luo, Some Recent Results in the Theory of the Wiener Number, *Indian J. Chem.*, **32A** (1993) 651–661.

- [35] I. Gutman, E.V. Konstantinova, V.A. Skorobogatov, Molecular Hypergraphs and Clar Structural Formulas of Benzenoid Hydrocarbons, *Acta Chimia Hungaria - Models in Chemistry*, **5–6** (1999) 539–548.
- [36] F. Harary, *Graph Theory*, Addison-Wesley Publishing Company, London, 1971.
- [37] I. Haiduc, J.J. Zuckerman, *Basic Organometallic Chemistry*, Walter de Gruyter, Berlin, New York, 1985.
- [38] F.R. Hartley, *Elements of Organometallic Chemistry*, The Chemical Society, London, 1974.
- [39] J.B. Hendrickson, A.G. Toczko, Unigue Numbering and Cataloguing of Molecular Structures, *J.Chem.Inf.Comput.Sci.*, **23** (1983) 171–177.
- [40] B.R. Heap, The Production of Graphs by Computer, In *Graph Theory and Computing*, Academic Press, New York, 1972, 47–62.
- [41] L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, London, 1976.
- [42] L.B. Kier, L.H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, 1986.
- [43] R.B. King, (ed.), *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983.
- [44] R.B. King, D.H. Rouvray, (eds.), *Graph Theory and Topology in Chemistry*, Elsevier, Amsterdam, 1987.
- [45] E.V. Konstantinova, A.A. Paleev, Sensitivity of Topological Indices of Polycyclic Graphs, *Vychisl. Sistemy*, **136** (1990) 38–48 (Russian).
- [46] E.V. Konstantinova, V.A. Skorobogatov, Molecular Hypergraphs, *Vychisl. Sistemy*, **151** (1994) 55–79. (Russian)

- [47] E.V. Konstantinova, V.A. Skorobogatov, Molecular Hypergraphs: The New Representation of Nonclassical Molecular Structures with Polycentric Delocalized Bonds, *J.Chem.Inf.Comput.Sci.*, **3** (1995) 472–478.
- [48] E.V. Konstantinova, V.A. Skorobogatov, Molecular Structures of Organoelement Compounds and Their Representation by the Labeled Molecular Hypergraphs, *Zh.Strukt.Khim.*, **2** (1998) 328–337.
- [49] E.V. Konstantinova, V.A. Skorobogatov, Graph and Hypergraph Models of Molecular Structure: a Comparative Analysis of Indices, *Zh.Strukt.Khim.*, **6** (1998) 958–966.
- [50] E.V. Konstantinova, M. Diudea, The Wiener Polynomial Derivatives and Other Topological Indices in Chemical Researches, *Croat. Chem. Acta*, **73** 2 (2000) 383–403.
- [51] V. Kvasnicka, J. Pospichal, Canonical Indexing and Constructive Enumeration of Molecular Graphs, *J.Chem.Inf.Comput.Sci.*, **30** (1990) 99–105.
- [52] G. Levi, A Note of the Derivation of Maximal Common Subgraphs of Two Directed or Undirected Graphs, *Calcolo*, **9** (1972) 341–352.
- [53] J. Moon, L. Moser, On Cliques in Graphs, *Israel J. Math.*, **1** (1965) 23–28.
- [54] S. Nikolić, N. Trinajstić, Z. Mihalić, The Wiener Index: Developments and Applications, *Croat. Chem. Acta*, **68** (1995) 105–129.
- [55] O. Ore, *Theory of Graphs*, American Mathematical Society, Colloquium Publications, 1962.
- [56] M.L. Overton, A. Proskurowski, Canonical Incidence Matrices of Graphs, *BIT*, **19** (1979) 271–273.
- [57] J. Plesnik, On the Sum of All Distances in a Graph or Digraph, *J.Graph Theory*, **8** (1984) 1–21.

- [58] M. Randić, On the Recognition of Identical Graphs Representing Molecular Topology, *J.Chem.Phys.*, **60** (1974) 3920–3928.
- [59] M. Randić, On Canonical Numbering of Atoms in a Molecule and Graph Isomorphism, *J. Chem. Inf. Comp. Sci.*, **3** (1977) 171–180.
- [60] R. Rado, Isomorphisms Between Hypergraphs, *New Direct. Theory Graphs*, **2** (1973) 207–237.
- [61] M. Razinger, J.R. Chretien, J.K. Dubois, Structural Selectivity of Topological Indices in Alkane Series, *J.Chem.Inform.Comp.Sci.*, **25** (1985) 23–27.
- [62] D.H. Rouvray, Should We Have Designs on Topological Indices? *Chemical Applications of Topology and Graph Theory. Studies in Physical and Theoretical Chemistry*, King R.B. (ed.), Elsevier: Amsterdam, **28** (1983) 159–177.
- [63] D.H. Rouvray, (ed.), *Computational Chemical Graph Theory*, Nova Science Publishers, New York, 1990.
- [64] G.B. Shulpin, *The World of Unusual Molecules*, Mir, Moscow, 1986 (Russian).
- [65] V.A. Skorobogatov, A.A. Dobrynin, Metric Analysis of Graphs, *Math.Chem. (MATCH)*, **23** (1988) 105–151.
- [66] 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.
- [67] I.V. Stankevitch, In *Application of Graph Theory in Chemistry*, N.S. Zefirov, C.I. Kuchanov, (eds.), Nauka, Novosibirsk, 1988, 7–69 (Russian).
- [68] N. Trinajstić, (ed.), *Mathematics and Computational Concepts in Chemistry*, Ellis Horwood, Chichester, 1986.

- [69] N. Trinajstić, *Chemical Graph Theory*, 2nd revised ed; CRC Press, Boca Raton, FL, 1992.
- [70] M. Tsutsui, Y. Ishii, H. Yaozeng, (eds.), *Fundamental Research in Organometallic Chemistry*, Van Nostrand Reinhold, New York, 1982.
- [71] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.*, **69** (1947) 17–20.
- [72] H. Wiener, Vapor Pressure–Temperature Relationships Among the Branched Paraffin Hydrocarbons, *J. Chem. Phys.*, **52** (1948) 425–430.
- [73] A.A. Zykov, Hypergraphs, *Usp.Math.Nauk*, **6** (1974) 89–154 (Russian).