

Molecular Hypergraphs: The New Representation of Nonclassical Molecular Structures with Polycentric Delocalized Bonds

Elena V. Konstantinova* and Vladimir A. Skorobogatov

Institute of Mathematics, Russian Academy of Sciences, Siberian Division, 630090 Novosibirsk, Russia

Received June 27, 1994[®]

The hypergraph model for presentation of nonclassical molecular structures with polycentric delocalized bonds is proposed here. The questions of identification and canonical coding of molecular hypergraphs are considered. Some topological indices are generalized for hypergraphs. The comparative analysis of graph and hypergraph models indices for organometallic molecular structures is proposed.

1. INTRODUCTION

Graph presentation of molecular structures is widely used in computational chemistry and theoretical chemical research.^{1,2} Molecular structures are represented by graphs where vertices correspond to atoms and edges to chemical bonds. This kind of graph, called a molecular graph,¹ is the object of study in the theory of ordinary graphs.³ However ordinary graphs do not adequately describe chemical compounds of nonclassical structure. As noted in ref 4 a substantial drawback of the structure theory is the lack of a convenient representation for molecules with delocalized polycentric bonds. Organometallic compounds are one example of such structures.⁵

Organometallic chemistry deals with compounds that have at least one metal-carbon (M-C) bond. This bond may be a simple covalent bond, as for example in $\text{Pb}(\text{C}_2\text{H}_5)_4$, but also may be an ionic bond, as for example in $\text{Na}^+\text{C}_2\text{H}_5^-$, or a π -bond, as for example in ferrocene, where iron has bonds with cyclopentadienyls, which are five-electron ligands. In principle any nonsaturated or aromatic organic molecule or radical can be a ligand in metal-carbon bond.

Just the same molecule may play roles of different ligands. The variety of M-C bond types and ligands leads to the appearance of a considerable range of organometallic compounds and provides a basis for classification of these compounds.⁵ In Figure 1, examples of molecular structures for compounds with three-electron ligands are given.

Let us consider some of the graph models used for representation of such compounds. In refs 6 and 7 examples of representation for "sandwich" and olefin structures by disconnected molecular graphs (see Figure 2a) are described. Representation 2a does not seem to be illustrative and does not allow us to analyze a structure as a whole, because there are no connections between molecular subgraphs representing individual structural fragments within it.

More illustrative but still not devoid of drawbacks are connected molecular graphs where all the vertices corresponding to carbon atoms are connected to the M vertex, which corresponds to a metal atom (Figure 2b). If $\text{M} = \text{Fe}$, Ni the valency of the metal atom is equal two. The degree of the M vertex in this case is equal to the number of vertices connected and not necessarily equal to the valency of the M atom. Besides, in both representations the difference between simple covalent and polycentric bonds is obscured.

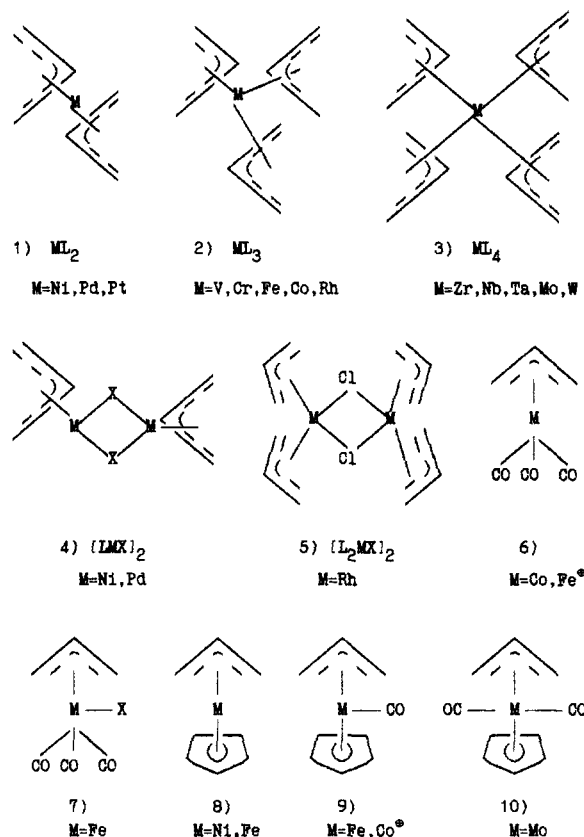


Figure 1. Molecular structures of compounds with three-electron ligands.

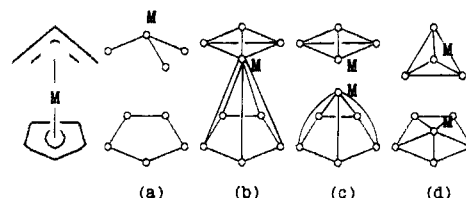


Figure 2. Different ways of representing an allylic metal-cyclopentadienyl complex hydrogen-deleted structure.

Sometimes planar graphs⁸ are used to describe organometallic compounds to study the stereochemical properties. The pyramids obtained from the fragments of the complex (see Figure 2c) are transformed to planar graphs (see Figure 2d). In this case the disconnected planar graph describes the structures. All representations (2a-d) are insufficient with regard to the molecule's symmetry. They all lead to

[®] Abstract published in *Advance ACS Abstracts*, December 15, 1994.

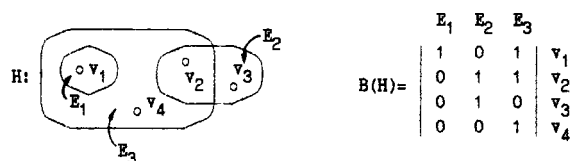


Figure 3. Hypergraph H and its incidence matrix.

the wrong conclusion that the three C atoms in the π -allyl system are equivalent.

All the defects of the structure representation that was mentioned above are eliminated if the hypergraphs are used to represent the structures with delocalized polycentric bonds.

2. HYPERGRAPHS OF MOLECULAR STRUCTURES

Let us introduce some concepts from the hypergraph theory.^{9,10}

Hypergraph $H = (V, E)$ consists of a nonempty 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 3. If v_i belongs to 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 Figure 3 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 (sets V and family E are finite) is unambiguously defined by an incidence matrix $B(H) = \|b_{ij}\|$, 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 H is also shown in Figure 3.

Two hypergraphs H^1 and H^2 are isomorphic if there exists a mapping conserving the one-to-one correlation between the sets of vertices and edges of these graphs.

2.1. 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 two correspond to delocalized polycentric bonds, and hyperedges $e \in E(H)$ with $\deg e = 2$ correspond to simple covalent bonds.

A molecular 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. 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 detail. Let us consider the example of six molecular hypergraphs (Figure 4) of an allylic complex shown earlier (Figure 2). It is known that the allylic ligand has an open chain of three sp^2 -hybridized carbon atoms, each having a

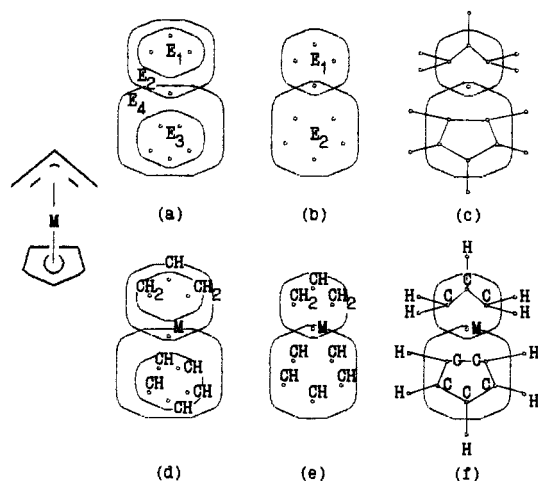


Figure 4. Molecular unlabeled (a-c) and labeled (d-f) hypergraphs representing an allylic complex with different degrees of detail; in (c-f) hydrogens are included, but in the other representations they are suppressed.

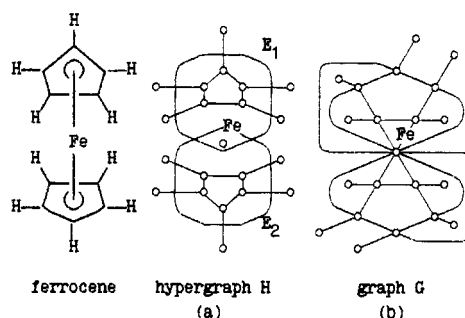


Figure 5. Hypergraph and graph ferrocene representation, including hydrogens.

π -electron available for metal-ligand bond formation.⁵ Figure 4a shows a molecular hypergraph where the delocalized bond in the allyl is represented by the edge E_1 , $\deg E_1 = 3$, and the bond between the allyl and the metal is represented by the edge E_2 , $\deg E_2 = 4$. Similarly, edges E_3 and E_4 represent π -bonds in cyclopentadienyl groups and the bond between the ligand and the metal. All delocalized polycentric bonds of the allylic complex are present 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 4b). In both cases hydrogens are suppressed. Figure 4c shows a molecular hypergraph where delocalized and σ carbon-carbon and carbon-hydrogen bonds are represented. As it seems, in Figure 4c is given an adequate hypergraph representation of an allylic complex, and Figure 4a,b are misleading with respect to symmetry. Until now the unlabeled molecular hypergraphs were considered only. Figure 4d-f shows the labeled hypergraphs of an allylic complex.

2.2. Examples of Molecular Structure Presentation. Now we will give examples to illustrate the difference between graphs and hypergraphs in the representation of the same molecular structures.

In case of an ordinary graph, the metal-ligand bonds in ferrocene are shown by ten edges connecting the vertex representing Fe with the other ten vertices (Figure 5b). This drawback is eliminated when ferrocene is represented by a hypergraph (Figure 5a). Hyperedges E_1 and E_2 represent π -bonds between iron and ligands. Moreover, the repre-

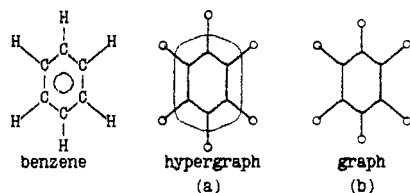


Figure 6. Hypergraph and graph benzene representation, including hydrogens.

sensation of ferrocene as a hypergraph makes it possible to visually illustrate the difference between σ (carbon–carbon and carbon–hydrogen) bonds and π -bonds. In the molecular hypergraph representing ferrocene the valency of the iron atom is being “conserved”, because the degree of Fe vertex in the hypergraph is equal to two. The degree of the same vertex of the graph G is equal to ten (Figure 5b). Note also that valencies of carbon atoms are being conserved under the hypergraph representation. Another example shows how the hypergraph representation preserves valency: molecular hypergraph of an allylic complex (Figure 4c) has vertices with degrees corresponding to the valencies of metal and carbon atoms. In the case of an ordinary graph (Figure 2), the valency of the metal atom is not being conserved.

We have to note that not only organometallic compounds contain delocalized bonds which can be described by hyperedges in a molecular hypergraph. The same bonds can also be found in conjugated molecules.¹¹ For example, Figure 6 shows the hypergraph (a) and the graph (b) representation of benzene. 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 six vertices and represents the bonding of all six π -electrons in the benzene. The fundamental principle of the tetravalency of carbon is fulfilled in this case.

3. IDENTIFICATION OF MOLECULAR STRUCTURES

The canonical representation is necessary for computer processing of molecular structures. The most conventional way to define molecular graphs is to determine an adjacency matrix which characterizes the graphs completely. In order to recognize isomorphic graphs the canonical forms of adjacency matrices are to be found. A number of methods of defining the canonical form of an adjacency matrix are described in refs 12–16. Concerning the hypergraphs computer representation the incidence matrix is especially convenient.

3.1. Canonical Incidence Matrices of Molecular Hypergraphs. 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*.

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.^{17,18}

It is evident that the canonical incidence matrix is unique and can be used for determining the canonical form of the hypergraph. The canonical form of a hypergraph is the hypergraph defined by canonical incidence matrix. The numbering of vertices and edges of a canonical form of a hypergraph is called canonical. In Figure 7 the canonical

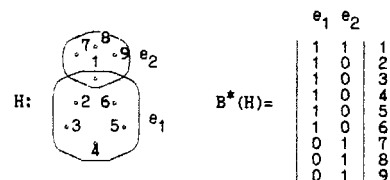
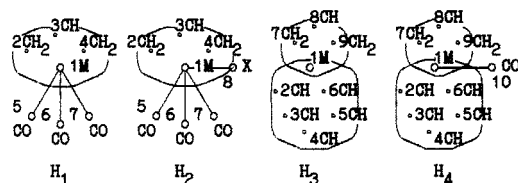


Figure 7. Canonical form and canonical incidence matrix of a hypergraph of an allylic complex.



$C(H_1) = (1Me, 2CH_2, 3CH, 4CH_2) (1, 500) (1, 600) (1, 700)$
 $C(H_2) = (1Me, 2CH_2, 3CH, 4CH_2) (1, 500) (1, 600) (1, 700) (1, 8X)$
 $C(H_3) = (1Me, 2CH, 3CH, 4CH, 5CH, 6CH) (1, 7CH_2, 8CH, 9CH_2)$
 $C(H_4) = (1Me, 2CH, 3CH, 4CH, 5CH, 6CH) (1, 7CH_2, 8CH, 9CH_2) (1, 1000)$

Figure 8. Canonical forms and codes of labeled molecular hypergraphs of some allylic compounds.

form of the hypergraph for the hydrogen–deleted structure of the allylic complex (Figure 4) and the canonical incidence matrix are shown.

We denote *canonical code* $C(H)$ of a hypergraph H in the form of q lists of sets $E_j = \{v_i | i=1 \dots q, i \in \{1 \dots p\}\}$ of labeled or nonlabeled vertices, ordered according to row numbers of the canonical incidence matrix. Each vertex label is present in the canonical code just once.

Canonical forms of labeled hypergraphs of four allylic compounds (see Figure 1) and their canonical codes are shown in Figure 8. The hypergraphs shown in Figure 8 contain only those hyperedges that correspond to the metal–ligand bonds. The bonds within the ligand itself are not considered for given presentation.

Canonical codes may be obtained on the basis of a simple method of construction of canonical incidence matrix for a hypergraph.

3.2. Algorithm for Construction of Canonical Incidence Matrices. The algorithm 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*¹⁰) with the set of vertices $V \cup E$, $|V \cup E| = p + q = s$, and the set of edges Y , $|Y| = \sum_{j=1}^q \deg E_j$. 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.¹⁰ 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 an adjacency matrix of graph $K(H)$, where $a_{ij} = 1$ if vertices i and 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 $p!$ adjacency matrices.

Besides the adjacency matrix of a bipartite graph, the reduced adjacency matrix $R(K(H)) = \|r_{ij}\|$, $i=1 \dots p$, $j=1 \dots q$,

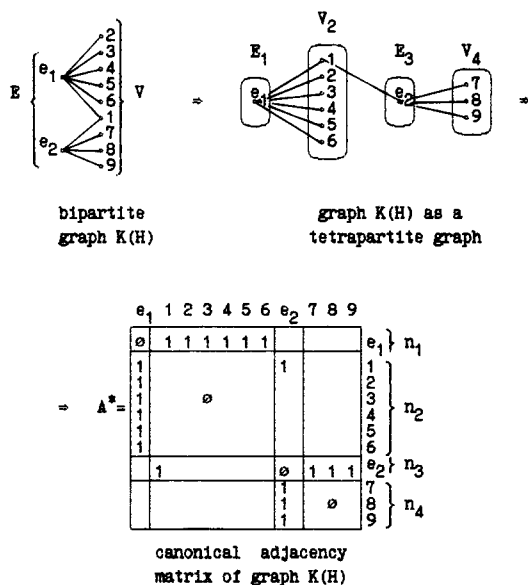


Figure 9. König's representation $K(H)$ of hypergraph H from Figure 7 and its canonical adjacency matrix.

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 is to define the canonical reduced adjacency matrix $R^*(K(H))$ for bipartite graph $K(H)$, the following proposition is true.

Proposition 1. $R^*(K(H)) = B^*(H)$. The proof follows from the definition of the canonical incidence matrices $R^*(K(H))$ and $B^*(H)$.

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 propositions.

Proposition 2. The canonical adjacency matrix of König's representation, of hypergraph H has a symmetrical block-diagonal form:

$$A^*(K(H)) = \begin{matrix} & \begin{matrix} n_1 & n_2 & n_3 & \dots & n_{t-1} & n_t \end{matrix} \\ \begin{matrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_{t-1} \\ n_t \end{matrix} & \begin{bmatrix} \emptyset & A_{12} & \emptyset & \dots & \emptyset & \emptyset \\ A_{21} & \emptyset & A_{23} & \dots & \emptyset & \emptyset \\ \emptyset & A_{32} & \emptyset & \dots & \emptyset & \emptyset \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \emptyset & \emptyset & \emptyset & \dots & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \dots & A_{t-1,t} & \emptyset \\ \emptyset & \emptyset & \emptyset & \dots & \emptyset & A_{t,t} \end{bmatrix} \end{matrix}$$

where rows (columns) of nonzero 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 the V -part. (Note that it is just an assumption only. Really the canonical numbering of $K(H)$ can start with the V -part or 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 proposition is proven.

The example in Figure 9 illustrates the proposition 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 proposition 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:

$$R(A^*(K(H))) = \begin{matrix} & \begin{matrix} n_2 & n_4 & n_6 & \dots & n_{t-3} & n_{t-1} \end{matrix} \\ \begin{matrix} n_1 \\ n_3 \\ n_5 \\ \vdots \\ n_{t-2} \\ n_t \end{matrix} & \begin{bmatrix} A_{12} & \emptyset & \emptyset & \dots & \emptyset & \emptyset \\ A_{32} & A_{34} & \emptyset & \dots & \emptyset & \emptyset \\ \emptyset & A_{54} & A_{56} & \dots & \emptyset & \emptyset \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \emptyset & \emptyset & \emptyset & \dots & A_{t-2,t-3} & A_{t-2,t-1} \\ \emptyset & \emptyset & \emptyset & \dots & \emptyset & A_{t,t-1} \end{bmatrix} \end{matrix}$$

$$R(A^*(K(H))) = \begin{matrix} \begin{matrix} n_1 & n_3 & n_5 & \dots & n_{t-2} & n_t \end{matrix} \\ \begin{matrix} n_2 \\ n_4 \\ n_6 \\ \vdots \\ n_{t-3} \\ n_{t-1} \end{matrix} & \begin{bmatrix} A_{21} & A_{23} & \emptyset & \dots & \emptyset & \emptyset \\ \emptyset & A_{43} & A_{45} & \dots & \emptyset & \emptyset \\ \emptyset & \emptyset & A_{65} & \dots & \emptyset & \emptyset \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \emptyset & \emptyset & \emptyset & \dots & A_{t-3,t-2} & \emptyset \\ \emptyset & \emptyset & \emptyset & \dots & A_{t-1,t-2} & A_{t-1,t} \end{bmatrix} \end{matrix}$$

Proposition 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 $p! \times q!$ 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 proposition is proved.

Figure 10 presents the canonical reduced adjacency matrix $R^*(K(H))$ obtained from the canonical adjacency matrix $A^*(K(H))$ of graph $K(H)$ (Figure 9) by the method discussed above. Easy to see that in accordance with proposition 1 we have the equality $R^*(K(H)) = B^*(H)$ that is just true (see Figure 7).

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 $K(H)$; and

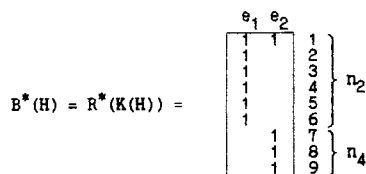


Figure 10. Canonical reduced adjacency matrix.

(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 of canonical adjacency matrix. Note that the canonization of bipartite graphs with parts N and M on the basis 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.¹² Algorithms for canonization of incidence matrices^{17,18} that are known to the authors are enumerative, and the estimations of complexity for the given algorithms are either not presented¹⁷ or not allowed to estimate the running effectivity of an algorithm in practice.¹⁸

4. TOPOLOGICAL INDICES OF GRAPHS AND HYPERGRAPHS

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 $G = (V, E)$ and $H = (V, E)$ 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 proposition is true.

Proposition 4.

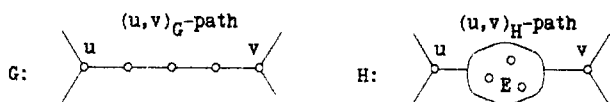
$$d_G(u,v) \geq d_H(u,v)$$

Proof. Let us consider several different cases. Case 1. The shortest (u,v) -walk in graph G and hypergraph H consists only of edges of degree two. In this case paths have the same length as G , and H describes the same molecular structure F on the same set of vertices and (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 of edges of degree two, and the shortest $(u,v)_H$ -walk in hypergraph H contains at least one edge E of a 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

Table 1. Formulas for Metric, Topological, and Information Indices^a

no.	name	formula
1	radius ³	$r = \min_v e(v)$
2	diameter ³	$d = \max_v e(v)$
3	av vertex eccentricity ¹⁹	$e_{av} = 1/p e$
4	eccentricity ¹⁹	$e = \sum_v e(v)$
5	eccentric ¹⁹	$\Delta G = 1/p \sum_v e(v) - e_{av} $
6	av vertex distance ¹⁹	$D_{av} = 2W/p$
7	unipolarity ¹⁹	$D^* = \min_v d(v)$
8	centralization ¹⁹	$\Delta G^* = 2W - pD^*$
9	mean distance deviation ¹⁹	$\Delta D = 1/p \sum_v D(v) - D_{av} $
10	dispersion ²³	$m_2^* = \min_v 1/p \sum_u [d(v,u)]^2$
11	compactness ²⁴	$\mu = 4/(p(p-1)) W$
12	variation ²⁵	$\text{var} = \max_v [d(v) - D^*]$
13	Wiener index ²⁶	$W = 1/2 \sum_{u,v} d(u,v)$
14	polarity no. ²⁷	see ref 27
15	mean-squares distance index ²⁸	$D^2 = (\sum_{i=1}^{\max} g_i^2 / \sum_{i=1}^{\max} g_i)^{1/2}$
16	graph distance index ²⁹	$GDI = \sum_{i=1}^{\max} (g_i)^2$
17	total adjacency index ²⁹	$A = \sum_{i,j} a_{ij}$
18	autometricity index ³⁰	see ref 30
19	information Wiener index ²²	$I_D^W = 1/2 \sum_{i=1}^{\max} (g_i i) / W \log_2 (g_i i) / W$
20	information Index H_3^{30}	$H_3 = \log_2 D_{av}$
21	information Index H_A^{31}	see Ref 31
22	information Index H_d^{31}	see Ref 31
23	information Index H_A^{31}	see Ref 31

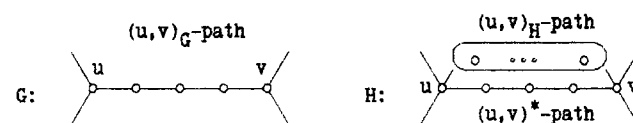
^a The following designations are used in the table: $e(v) = \max_u d(u,v)$ is the vertex eccentricity; $d(v) = \sum_u d(u,v)$ is the vertex distance; g_i is a number of vertex pairs at a distance i from each other; a_{ij} is the adjacency matrix element.

to the same edge is equal to one and

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

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

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 $(u,v)_G$ -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

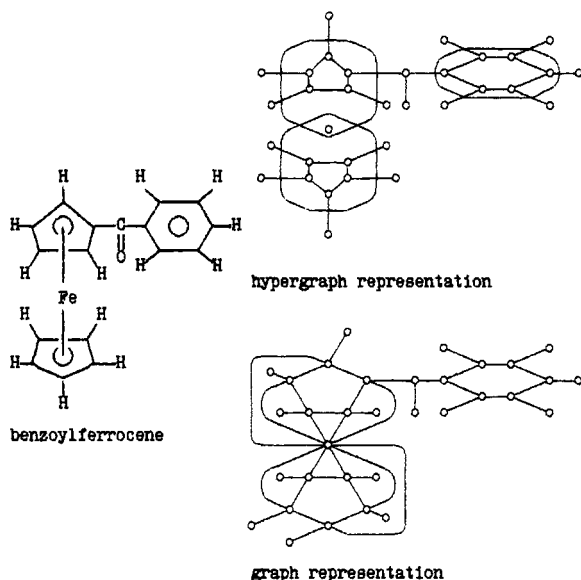
$$d_H(u,v) < d_G(u,v)$$

The proposition is proven.

Proposition 4 implies that for any F represented by G and H the following inequalities hold: $W(G) > W(H)$, $r(G) > r(H)$, $d(G) > d(H)$, where W is the Wiener index,²⁶ r is the radius,³ and d is the diameter³ (see Table 1).

Table 2. Formulas of Substituents for Ferrocene Derivatives

no.	substituents R	no.	substituents R
1	-H	16	-COOC ₂ H ₅
2	-CH ₃	17	-COCH ₃
3	-C ₂ H ₅	18	-CH ₂ OH
4	-(CH ₂) ₂ CH ₃	19	-N(CH ₃) ₂
5	-CH(CH ₃) ₂	20	-CH(OH)CH ₃
6	-(CH ₂) ₃ CH ₃	21	-CH(OH)C(CH ₃) ₂
7	-C(CH ₃) ₃	22	-CH ₂ C ₆ H ₅
8	-CN	23	-COC ₆ H ₅
9	-CHCH ₂	24	-C ₆ H ₁₁
10	-COOH	25	-C(CH ₃)CH ₂
11	-COCH ₃	26	-(CH ₂) ₂ CH ₃
12	-COOCH ₃	27	-(CH ₂) ₂ COC ₆ H ₅
13	-OCOCH ₃	28	-COOC ₆ H ₅
14	-COH	29	-OCOC ₆ H ₅
15	-(CH ₂) ₂ C ₆ H ₅	30	-(CH ₂) ₄ COOH

**Figure 11.** Hypergraph and graph representation of benzoylferrocene.

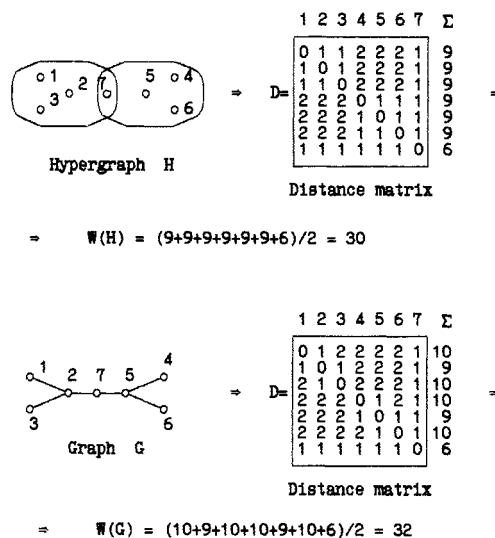
A few other topological indices are likely to reveal similar regularities. The behavior on the whole set of graphs and hypergraphs cannot be estimated for all topological indices.

The comparative analysis of 23 metric, topological, and information indices of graph and hypergraph models for 30 ferrocene derivatives CpFeC₅H₄R, where R is a series of substituents given in Table 2, is below submitted. Graph and hypergraph ferrocene and benzene representations are shown in Figures 5 and 6. For example, in Figure 11 is given the structural formula of benzoylferrocene and its graph and hypergraph representation. The unlabeled molecular graphs and hypergraphs are used here.

Three groups of indices presented in Table 1 are considered: metric (nos. 1–12), topological (nos. 13–17), and information indices (nos. 18–23). Metric indices of graphs have been studied in many papers of graph theory. A bibliography on the subject is given in refs 19,20. Topological indices are well-known in mathematical chemistry and are used for representation of molecular structure and for finding the quantitative structure–property and structure–activity relationships.²¹ Information indices^{2,22} are constructed for various matrices (layer matrix (no. 21), distance matrix (no. 22), adjacency matrix (no. 23)) and also for some topological indices such as the Wiener index (no. 19). The expressions for calculation of indices are presented in Table 1 only if the formula is short. Otherwise the reference is indicated.

Table 3. Quantity of Indices with 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

**Figure 12.** The example of obtaining the values of the Wiener index for hypergraph and graph representations of the first molecular structure shown in Figure 1; hydrogens are suppressed in these representations.

The sensitivity of indices for both models is 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 difficult enough: therefore the evaluation S of sensitivity I on the fixed set M of nonisomorphic graphs is used^{32,33} by formula $S = (N - N_I)/N$, where N_I is the number of degeneracies I on the set M .

The analysis has shown that for considered series of molecular structures the indices sensitivities of the hypergraph model is higher than ones on the graph model. The data, submitted in the Table 3, shown in percentage of the quantity of indices, the sensitivity S_G of which in the graph model less, more and coincides with the sensitivity S_H of these indices in the hypergraph model.

The correlation dependences of indices within considered models are investigated. We shall consider that the correlation ratio r of noncorrelating indices is under limits of $0 < r < 0.9$. The share of pairs of indices with the mentioned correlation ratio in the hypergraph model is equal to 58.4%, and for the graph model is equal to 40.8%. Thus the number of noncorrelating indices is greater for the hypergraph model than for the graph one.

Now let us consider the change of values of topological indices nos. 13 and 15 (see Table 1) across the set of hydrogen-deleted structures including ten complexes shown in Figure 1 (one structure of each type), represented by graphs and hypergraphs. The example of obtaining the values of the Wiener index for graph and hypergraph representations of the first molecular structure shown in Figure 1 is given in Figure 12. Values of the Wiener index and the mean-squares distance index $D^{(2)}$ for graphs G and

Table 4. Ordering of Molecular Structures in Accordance with the Values of Indices W and $D^{(2)}$

no.	$W(G)$	$W(H)$	no.	$D^{(2)}(G)$	no.	$D^{(2)}(H)$
1	32	30	1	1.6036	8	1.5000
6	34	33	8	1.6583	1	1.5119
7	47	46	6	1.6903	9	1.5916
8	57	51	9	1.7127	6	1.6475
9	74	68	2	1.7320	10	1.6569
2	75	72	7	1.7423	2	1.6733
10	93	87	10	1.7529	7	1.7113
4	100	98	3	1.7974	3	1.7541
3	136	132	4	2.4766	4	2.4494
5	300	296	5	2.7386	5	2.7203

hypergraphs H are listed in Table 4. The table displays the values of the first and the second indices across the set of graphs which exceed the values of the indices across the set of hypergraphs. However, in the case of the Wiener index, the results of ordering based on the values of the index coincides for sets of graphs and hypergraphs, but they are considerably different for $D^{(2)}$ index on the same sets. The hypergraph representation used in the last example seems to have the lead detail of all the possible ones, but nevertheless it allows investigation of the differences of the topological indices values. The differences may happen to be important for the study of structure–property correlations.³⁴

5. CONCLUSION

The new representation of molecular structures with polycentric delocalized bonds is proposed in this work. 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 is introduced, and the ways of its implementation are shown in several examples. A method for the construction of the canonical codes for molecular hypergraphs is also presented. The canonical coding of hypergraphs may be used to obtain formal description of nonclassical molecular structures for databases processing and solving problems of constructing their nomenclature.

A study of topological indices of molecular graphs and molecular hypergraphs of the same compounds was held. The comparative analysis of indices for graph and hypergraph models for 30 ferrocene derivatives shows that an increase of the general number of noncorrelating indices on the hypergraph model is observed. The study has also shown that ordering of structures relative to some indices may differ substantially. The differences can be used as additional information when predicting properties of nonclassical compounds from their structure.

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

ACKNOWLEDGMENT

This work was supported by the Russian Fund of Fundamental Investigations through Grant No. 93-03-18657. The authors are grateful to the referees who suggested some valuable improvements for contents of this paper. The authors are also grateful to Professor Yu. S. Nekrasov

(Nesmeyanov Institute of Organoelement Compounds) for important information and discussion.

REFERENCES AND NOTES

- (1) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; Vol. 1, Chapter 3.
- (2) Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures*; Research Studies Press: Chichester, 1983.
- (3) Harary, F. *Graph Theory*; Addison-Wesley Publishing Company: London, 1969.
- (4) Babaev, E. V. In *Principles of Symmetry and Systemic Approach in Chemistry*; Nauka: Moscow, 1987; p 30 (in Russian).
- (5) Haiduc, I.; Zuckerman, J. J. *Basic Organometallic Chemistry*; Walter de Gruyter: Berlin, New York, 1985.
- (6) Stankevitch, I. V. In *Application of Graphs in Chemistry*; Zefirov, N. S., Kuchanov, C. I., Eds.; Nauka: Novosibirsk, 1988; p 7 (in Russian).
- (7) Shulpin, G. B. *The World of Unusual Molecules*; Nauka: Moscow, 1986 (in Russian).
- (8) Berge, C. *Theorie des Graphes et ses Applications*; Dunod: Paris, 1958.
- (9) Berge, C. *Graphes et Hypergraphes*; Dunod: Paris, 1970.
- (10) Zykov, A. A. Hypergraphs. *Usp. Math. Nauk* **1974**, *6*, 89–154 (in Russian).
- (11) Clar, E. *The Aromatic Sextet*; Wiley: London, 1972.
- (12) Arlasarov, V. A.; Zuev, I. I.; Uskov, A. V.; Faradzhev, I. A. Algorithm for Transformation of Finite Nonoriented Graphs to Canonical Form. *Zh. Vychisl. Mat. Mat. Fis.* **1974**, *14*, 737–743 (in Russian).
- (13) Randić, M. On the Recognition of Identical Graphs Representing Molecular Topology. *J. Chem. Phys.* **1974**, *60*, 3920–3928.
- (14) Hendrickson, J. B.; Toczko, A. G. Unique Numbering and Cataloguing of Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171–177.
- (15) Heap, B. R. The Production of Graphs by Computer. In *Graph Theory and Computing*; Academic Press: New York, 1972; pp 47–62.
- (16) Kvasnicka, Vladimir; Pospichal, Juri. Canonical Indexing and Constructive Enumeration of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 99–105.
- (17) Baraev, A. M.; Faradzhev, I. A. In *Algorithmic Studies in Combinatorics*; Faradzhev, I. A., Ed.; Nauka: Moscow, 1978 (in Russian).
- (18) Overton, M. L.; Proskurowski, A. Canonical Incidence Matrices of Graphs. *BIT* **1979**, *19*, 271–273.
- (19) Skorobogatov, V. A.; Dobrynin, A. A. Metric Analysis of Graphs. *Math. Chem. (MATCH)* **1988**, *23*, 105–151.
- (20) Plesnik, J. On the Sum of All Distances in a Graph or Digraph. *J. Graph Theory* **1984**, *8*, 1–21.
- (21) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; Vol. 2, Chapter 4.
- (22) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533.
- (23) Ore, O. *Theory of Graphs*; American Mathematical Society: Colloquium Publications: 1962; Vol. 38.
- (24) Doyle, J. K.; Graver, J. E. Mean Distance in Graph. *Discrete Math.* **1977**, *17*, 147–154.
- (25) Entiger, R. C.; Jackson, D. E.; Snyder, D. A. Distance in Graphs. *Czechoslovak Math. J.* **1976**, *26*, 283–297.
- (26) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- (27) Wiener, H. Vapor pressure-temperature Relationships among the Branched Paraffin Hydrocarbons. *J. Chem. Phys.* **1948**, *52*, 199–206.
- (28) Balaban, A. T. Topological Indices based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199–206.
- (29) Rouvray, D. H. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Eds.; Elsevier: Amsterdam, 1983; p 159.
- (30) Skorobogatov, V. A.; Konstantinova, E. V.; Nekrasov, Yu. S.; Sukharev, Yu. N.; Tepfer, E. E. On the Correlation between the Molecular Information Topological and Mass Spectra Indices of Organometallic Compounds. *Math. Chem. (MATCH)* **1991**, *26*, 215–228.
- (31) Konstantinova, E. V.; Paleev, A. A. Sensitivity of Topological Indices of Polycyclic Graphs. *Vychisl. Systemy* **1990**, *136*, 38–48 (in Russian).
- (32) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Isomer Discrimination by Topological Information Approach. *J. Comput. Chem.* **1981**, *2*, 127–148.
- (33) Razinger, M.; Chretien, J. R.; Dubois, J. K. Structural Selectivity of Topological Indices in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23–27.
- (34) Randić, M.; Kraus, G. A.; Džonova-Jerman-Blažić, B. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Eds.; Elsevier: Amsterdam, 1983; p 192.