Graph and hypergraph models of molecular structure: A comparative analysis of indices

Article *in* Journal of Structural Chemistry · November 1998

DOI: 10.1007/BF02903615

2 authors, including:

Elena Konstantinova
Russian Academy of Sciences
69 PUBLICATIONS **732** CITATIONS

SEE PROFILE

Journal of Structural Chemistry, Vol. 39, No. 6, 1998

GRAPH AND HYPERGRAPH MODELS OF MOLECULAR STRUCTURE: A COMPARATIVE ANALYSIS OF INDICES

E. V. Konstantinova and V. A. Skoroboratov

UDC 519.17.547.25:547.1

Eight series of molecular structures represented by molecular hypergraphs and molecular graphs are considered. A comparative analysis is performed for 23 integral topological and information theoretical indices for the graph and hypergraph models. For each series, the sensitivity is studied on both models, correlations are established for indices inside the models, and indices not correlating with each other are found. The results of this analysis showed that the sensitivity of most indices is higher in the hypergraph model. The total number of noncorrelated indices also increases in the latter model.

INTRODUCTION

Graph-theoretical concepts are widely used in computer-aided and theoretical chemical research [1, 2]. Molecular structures are represented as graphs whose vertices denote atoms and edges correspond to chemical bonds. This is called a molecular graph [1], which is an object of investigation in ordinary graph theory [3]. A new model based on hypergraph theory [5] was suggested in [4] to represent molecular structures with delocalized multicenter bonds. In contrast to the graph model, the hypergraph model represents not only edges of degree 2 corresponding to ordinary covalent bonds (Fig. 1*b*) but also edges of higher degrees corresponding to delocalized multicenter bonds (Fig. 1*a*) and called hyperedges. Figure 1*a* depicts the molecular hypergraph H of ferrocene in which the hyperedges E_1 and E_2 define the π-bonds between the iron atom and the ligand. A hypergraph representation clearly demonstrates the difference between the ordinary carbon–hydrogen covalent bonds and the π-bonds. Moreover, the formal valence of the metal atom is preserved (in Fig. 1, the vertex Fe has degree 2 in the hypergraph H and 10 in the graph G).

The differences between the graph and hypergraph representations of molecular structures are reflected in

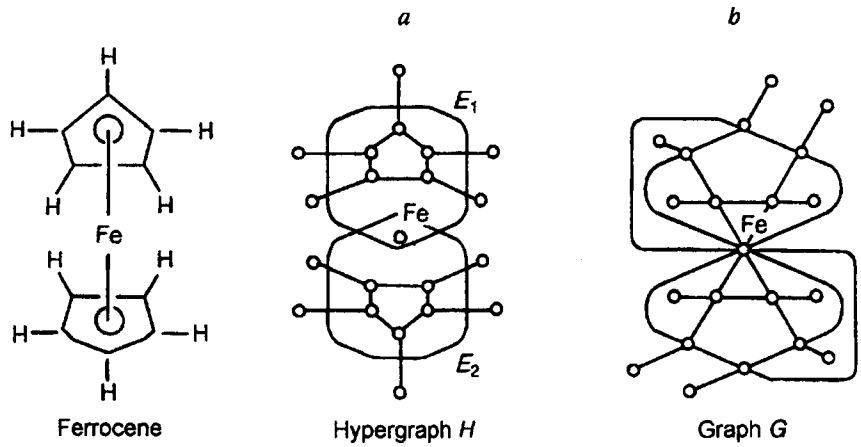

Fig. 1. Hypergraph (*a*) and graph (*b*) representations of ferrocene.

Institute of Mathematics, Siberian Branch, Russian Academy of Sciences (Novosibirsk). Translated from *Zhurnal Strukturnoi Khimii*, Vol. 39, No. 6, pp. 1163-1171, November-December, 1998. Original article submitted

 0022-4766/98/3906-0958$20.00

the values of topological indices. This paper gives the results of a comparative analysis of the values of integral topological and information theoretical indices for 8 series of molecular systems R–X, where R are widespread structural types (linear, branched, cyclic, aromatic, heterocyclic, organometal):

No.	Name	R,
1	Methyl	CH_3–,
2	*iso*-Butyl	$(CH_3)_3C$–,
3	Allyl	CH_2CHCH_2–,
4	Phenyl	C_6H_5–,
5	Cyclohexyl	C_6H_{11}–,
6	Ferrocenyl	$C_5H_5FeC_5H_4$–,
7	Benchrotrenyl	$(CO)_3CrC_6H_5$–,
8	Allylirontricarbonyl iodide	$(CO)_3IFeCH_2CHCH_2$–,

and X are widespread substituents:

No.	X	No.	X,
1	–H	16	$-COOC_2H_5$,
2	$-CH_3$	17	$-COCH_3$,
3	$-C_2H_5$	18	$-CH_2OH$,
4	$-(CH_2)_2CH_3$	19	$-N(CH_3)_2$,
5	$-CH(CH_3)_2$	20	$-CH(OH)CH_3$,
6	$-(CH_2)_3CH_3$	21	$-CH(OH)C(CH_3)_3$,
7	$-C(CH_3)_3$	22	$-CH_2C_6H_5$,
8	–CN	23	$-COC_6H_5$,
9	$-CH=CH_2$	24	$-C_6H_{11}$,
10	–COOH	25	$-C(CH_3)=CH_2$,
11	–OCOH	26	$-CH=CHCH_3$,
12	$-COOCH_3$	27	$-CH=CHCOC_6H_5$,
13	$-OCOCH_3$	28	$-COOC_6H_5$,
14	–COH	29	$-OCOC_6H_5$,
15	$-CH=CHC_6H_5$	30	$-(CH_2)_4COOH$,

Unlabeled graphs and hypergraphs are discussed. Figure 2 presents an example of graph and hypergraph representation for the structural formula of benzoylferrocene. For each series, the sensitivities of indices are studied

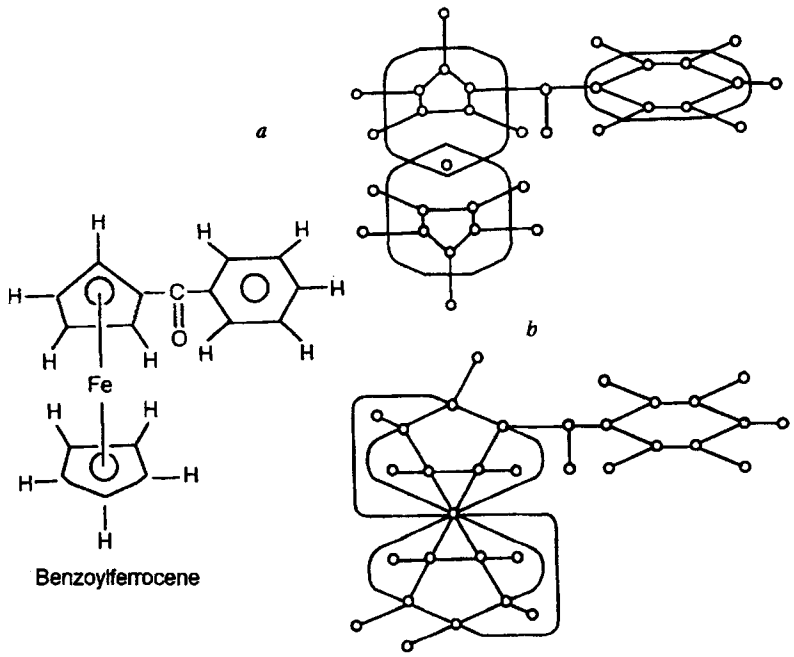

Fig. 2. Hypergraph (*a*) and graph (*b*) representations of benzoylferrocene.

on both models, correlations are established for indices within the models, and mutually noncorrelating indices are revealed.

For the tabulated data below, symbols G and H denote the graph and hypergraph models, respectively, and the following notation is used for series of molecular systems (each series containing 30 structures):

M_1 – methane derivatives
M_2 – isobutane derivatives
M_3 – allyl derivatives
M_4 – benzyl derivatives
M_5 – cyclohexane derivatives
M_6 – ferrocene derivatives
M_7 – benzenechromotricarbonyl derivatives
M_8 – allylirontricarbonyl iodide derivatives

1. TOPOLOGICAL AND INFORMATION INDICES

The sample of invariants used to compare the models consists of 23 integral indices based on a distance in an unlabeled graph (hypergraph). The distance $d(u, v)$ between the vertices u, v of the graph G (hypergraph H) is the length of the shortest ordinary (u, v)-chain. The length of any hyperedge in H is 1, i.e., all the vertices incident to the same hyperedge lie distance 1 apart. For example, the distance between any vertices of the hyperedge E_1 (E_2) of H represented in Fig. 1*a* equals 1. In the graph model, the distance between the analogous vertices will be either 1 or 2.

All the indices under analysis are defined on ordinary graphs and generalized for hypergraphs. Among the indices are pure graph invariants and the topological and information theoretical indices well known in the chemical literature. Below we give formulas for calculating indices. The first five indices are based on the notion of eccentricity of a graph vertex $e(v) = \max\limits_{u \in V} d(u, v)$ [3].

1. Radius of a graph [3]

$$r(G) = \min_{v \in V} e(v).$$

2. Diameter of the graph [3]

$$D(G) = \max_{v \in V} e(v).$$

3. Average eccentricity of vertices in the graph [6]

$$e_{av}(G) = \frac{1}{p} e(G).$$

4. Eccentricity of the graph [3]

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

5. Eccentricity deviation of the graph [5]

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

The next seven indices use the notions of the distance of a graph vertex $D(v) = \sum\limits_{u \in V} d(u, v)$ and the distance of a graph $D(G) = \frac{1}{2} \sum\limits_{v \in V} D(v)$ [7, 8].

6. Mean distance of vertices [6]

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

7. Minimal distance (unipolarity) [8]

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

8. Distance deviation of the graph (centralization) [8]

$$\Delta G^* = 2D(G) - pD^*.$$

9. Mean distance deviation of the graph [6]

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

10. Dispersion of the graph [9]

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

11. Compactness of the graph [10]

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

12. Variation of the graph [8]

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

The following five indices were chosen from the set of topological indices widely used in chemical investigations.

13. Wiener index [11]

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

14. Polarity index [12]

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

where $d_{3,i}$ are the distances in the graph which equal 3.

15. Rms distance index [13]

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

where g_i is the number of pairs of vertices distance i apart.

16. Distance index [14]

$$GDI(G) = \sum_{i=1}^{\max} g_i^2.$$

17. Complete adjacency index [14]

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

where a_{ij} is the adjacency matrix element.

The information theoretical indices are represented by six indices.

18. Information theoretical Wiener index [15]

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

19. Autometricity index H_1, introduced in [16] taking into account the partitioning of the set of graph vertices into autometricity classes. An autometricity class is a set of vertices having the same vector, $\lambda_{i1}, ..., \lambda_{id(G)}$, where λ_{ij} is the number of vertices lying at a distance j from the i vertex. Suppose p_i is the number of vertices in the ith ($i = 1, ..., N$) autometricity class and $P = \sum_{i=1}^{N} p_i$ is the number of vertices in the graph. Then

$$H_1(G) = -\sum_{i=1}^{N} \frac{p_i}{p} \log_2 \frac{p_i}{p}.$$

20, 21, 22. Information integral indices $H_\lambda(G)$, $H_d(G)$, $H_A(G)$, defined in [17] based on the information vertex indices $H_\lambda(i)$, $H_d(i)$, $H_A(i)$ calculated by the formulas

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

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

$$H_A(i) = -\sum_{i=1}^{\deg(i)} \left(a_{ij} \Big/ \sum_{j=1}^{\deg(i)} a_{ij} \right) \log_2 \left(a_{ij} \Big/ \sum_{j=1}^{\deg(i)} a_{ij} \right).$$

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

Then the integral information indices are defined as follows:

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

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

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

23. The information analog of the mean distance of vertices [16]

$$H_{av}(G) = \log_2 D_{av}(G).$$

For all the series under analysis, the indices were evaluated for the graph and hypergraph models. The results of the comparative analysis of the sensitivities and correlation dependences of the indices in each model are presented in the following sections.

2. SENSITIVITY OF INDICES

The sensitivity of a topological index I is a measure of its ability to recognize nonisomorphic graphs (hypergraphs) [18]. Theoretical estimation of I on the set of all graphs (hypergraphs) is rather complex. Therefore, in practice one uses an estimate S of the sensitivity I on a fixed set M of nonisomorphic graphs (hypergraphs):

$$S = \frac{N - N_1}{N},$$

where N_I is the number of degeneracies of I on the set $M, N = |M|$.

Let us compare the sensitivities of the 23 indices chosen above for the models in question on the sets $M_1, ..., M_8$. Table 1 lists the results for all the series of compounds. As can be seen from the table, the sensitivity of the same index may be higher on the set of graphs for one series and on the set of hypergraphs for another. Let $I_G(M)$ and $I_H(M)$ denote the sensitivity I of an index on the set of structures M for the graph and hypergraph models, respectively. Then, for example, for the autometricity index H_1 (No. 18 in Table 1), we have $I_G(M_4) > I_H(M_4)$ (0.8 > 0.733) but $I_G(M_7) < I_H(M_7)$ (0.733 < 0.867).

For the series under analysis, in the hypergraph model the sensitivities are generally higher than those in the graph model. This is confirmed by the percent of indices whose sensitivity in the graph model is smaller than, or greater than, or equals the sensitivity of these indices in the hypergraph model.

Sensitivity	%
$I_G < I_H$	20.7
$I_G > I_H$	10.9
$I_G = I_H$	68.4

TABLE 1. Sensitivity of Indices

Index	M_1		M_2		M_3		M_4	
	G	H	G	H	G	H	G	H
1	0.067	0.033	0	0.033	0.067	0.033	0	0.033
2	0.067	0.1	0.033	0.067	0.067	0.1	0.067	0.1
3	0.767	0.7	0.867	0.8	1	1	0.933	0.867
4	0.867	0.933	0.933	0.867	0.867	0.867	0.8	0.867
5	0.867	0.867	1	0.933	1	1	0.933	0.933
6	0.933	0.933	1	1	1	1	0.933	0.933
7	0.8	0.8	0.8	0.8	0.8	0.733	0.733	0.733
8	0.8	0.8	0.933	0.933	0.8	0.8	0.933	0.933
9	0.867	0.867	1	1	1	1	0.933	0.933
10	0.933	0.933	1	1	1	1	0.933	0.933
11	0.933	0.933	1	1	1	1	0.933	0.933
12	0.433	0.433	0.567	0.4	0.567	0.567	0.4	0.3
13	0.933	0.933	1	1	1	1	0.933	0.933
14	0.5	0.533	0.5	0.533	0.5	0.533	0.3	0.6
15	0.933	0.933	1	1	1	1	0.933	0.933
16	0.933	0.933	1	1	1	1	0.933	0.933
17	0.233	0.367	0.233	0.367	0.233	0.367	0.233	0.367
18	0.667	0.667	0.733	0.733	0.8	0.8	0.8	0.733
19	0.933	0.933	1	1	1	1	0.933	0.933
20	0.933	0.933	1	1	1	1	0.933	0.933
21	0.933	0.933	1	1	1	1	0.933	0.933
22	0.933	0.933	1	1	1	1	1	1
23	0.6	0.6	0.8	0.733	0.6	0.6	0.6	0.6

TABLE 1 (Continued)

Index	M_5		M_6		M_7		M_8	
	G	H	G	H	G	H	G	H
1	0	0.033	0.067	0.033	0.067	0.033	0.067	0.033
2	0.067	0.1	0.067	0.1	0.067	0.1	0.067	0.1
3	0.933	1	1	0.933	0.933	0.933	0.933	1
4	0.867	0.933	1	1	1	1	1	1
5	0.933	1	1	1	1	0.933	1	1
6	1	1	1	1	1	1	1	1
7	0.733	0.8	0.733	0.8	0.733	0.8	0.667	0.767
8	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1
12	0.467	0.3	0.5	0.533	0.5	0.5	0.567	0.433
13	1	1	1	1	1	1	1	0.933
14	0.5	0.533	0.5	0.6	0.5	0.633	0.6	0.6
15	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1
17	0.233	0.367	0.233	0.367	0.233	0.367	0.233	0.367
18	0.8	0.8	0.8	0.8	0.733	0.867	0.733	0.733
19	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1
21	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1
23	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6

The number of indices whose sensitivity is higher in the hypergraph model is nearly doubled with respect to the number of indices whose sensitivity is higher in the graph model. When the sensitivity of an index in the two models is the same, structure ordering according to the values of this index may vary with model. These differences may prove useful for investigating structure–property correlations.

3. CORRELATION DEPENDENCES OF INDICES

Let us study the correlation dependences of indices inside the models. For each series of structures, correlations between the indices of the graph and hypergraph models were examined in terms of a linear relationship of $y = a + bx$ type. In the plot of Fig. 3, for each series of structures, the dots mark the values corresponding to the percent of the mutually noncorrelated (for $0 \leq r < 0.9$) pairs of indices within the graph and hypergraph models. As

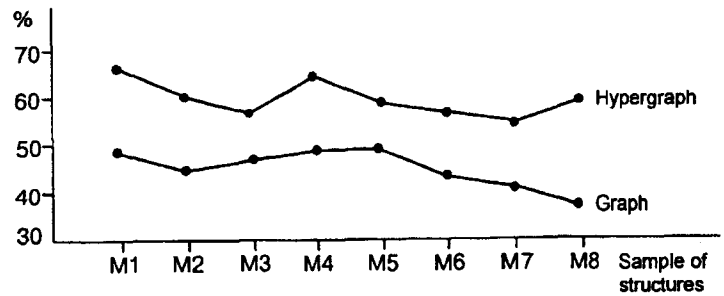

Fig. 3. Percent of pairs of indices with a correlation coefficient r = (0-9).

can be seen in the plot, for all series, the number of noncorrelated pairs of indices is greater in the hypergraph model. For example, on the set M_8 this number is 1.5 times higher than the number in the graph model.

Among all indices in each model one can separate out the indices having the correlation coefficients $r < 0.9$ (at least 90% of the total number of indices). The scheme below presents the serial numbers of such indices for each set of structures:

Set of structures	G	H
M_1	18, 19	5, 6, 18, 19, 23
M_2	18, 19	6, 18, 19, 23
M_3	18, 19	18, 19, 23
M_4	18, 19	6, 17, 18, 19, 23
M_5	18, 19	5, 18, 19, 23
M_6	19	18, 19, 23
M_7	18, 19	18, 19, 23
M_8	19	18, 19, 23

Such indices are encountered more often in the hypergraph model.

CONCLUSIONS

Thus the comparative analysis of topological and information indices for the 8 series of molecular structures showed that the sensitivity of most indices is higher in the hypergraph model. In the same model, we also observe an increase in the total number of noncorrelated indices. This indicates that the hypergraph model gives a higher accuracy of molecular structure description: the higher the accuracy of the model, the greater the diversity of the behavior of its invariants. On the other hand, the nonlabeled molecular graphs and hypergraphs reflect only the specific features of the framework of a molecular structure. Probably, the results may be improved if we consider labeled molecular hypergraphs [19]. Consideration of the chemical peculiarities of atoms and bonds in this case makes the model more accurate. Using a particular model must rely on the conditions and requirements of the problem being solved.

REFERENCES

1. N. Trinajstic, *Chemical Graph Theory*, CRC Press, Boca Raton, FL (1983).
2. D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press, Chichester (1983).
3. F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA (1969).
4. E. V. Konstantinova and V. A. Skorobogatov, *J. Chem. Inf. Comp. Sci.*, **35**, 472-478 (1995).
5. C. Berge, *Graphes et hypergraphes*, Dunod, Paris (1970).
6. V. A. Skorobogatov and P. V. Khvorostov, *Vychisl. Sist. (Novosibirsk)*, No. 91, 3-20 (1981).
7. R. C. Entriger, D. E. Jackson, and D. A. Snyder, *Czechoslovak Math. J.*, **26**, No. 2, 283-297 (1976).
8. T. Hevic and N. P. Gledic, in: *Mathematics in Sociology* [in Russian], Moscow (1977), pp. 151-169.
9. O. Ore, *Graph Theory* [in Russian], Nauka, Moscow (1968).
10. I. K. Doyle and I. E. Graver, *Discrete Math.*, **17**, No. 2, 147-155 (1977).
11. H. Wiener, *J. Am. Chem. Soc.*, **69**, 17-20 (1947).
12. H. Wiener, *J. Chem. Phys.*, **52**, 425-430 (1948).
13. A. T. Balaban, *Pure Appl. Chem.*, **55**, 199-206 (1983).
14. D. H. Rouvray, in: *Chemical Applications of Topology and Graph Theory. Studies in Physical and Theoretical Chemistry*, **28**, 159-177 (1983).
15. D. Bonchev and N. Trinajstic, *J. Chem. Phys.*, **67**, 4517-4533 (1977).
16. V. A. Skorobogatov, E. V. Konstantinova, Yu. S. Nekrasov, et al., *Math. Chem. (MATCH)*, No. 26, 215-228 (1991).

17. E. V. Konstantinova and V. A. Paleev, *Vychisl. Sist. (Novosibirsk)*, No. 136, 38-48 (1991).
18. D. Bonchev, O. Mekenyan, and N. Trinajstic, *J. Comput. Chem.*, **2**, No. 2, 127-148 (1981).
19. E. V. Konstantinova and V. A. Skorobogatov, *Zh. Strukt. Khim.*, **39**, No. 2, 328-337 (1998).